SYM-AM-19-030

Proceedings
of the
Sixteenth Annual Acquisition Research Symposium

Thursday Sessions
Volume II

Acquisition Research:
Creating Synergy for Informed Change

May 8–9, 2019

Published: April 30, 2019

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

The research presented in this report was supported by the Acquisition Research Program of the Graduate School of Business & Public Policy at the Naval Postgraduate School.

To request defense acquisition research, to become a research sponsor, or to print additional copies of reports, please contact any of the staff listed on the Acquisition Research Program website (www.acquisitionresearch.net).

Table of Contents

SYM-AM-19-060

Proceedings
of the
Sixteenth Annual
Acquisition Research Symposium

Thursday Sessions
Volume II

Acquisition Research:
Creating Synergy for Informed Change

May 8–9, 2019

Published: April 30, 2019

THIS PAGE INTENTIONALLY LEFT BLANK

Welcome: James Greene, RADM, USN (Ret.), Chair of Acquisition, Acquisition Research Program

James Greene, RADM, USN (Ret.)— RADM Greene has been the Chair of Acquisition, Acquisition Research Program, Professor of the Practice of Acquisition at the Naval Postgraduate School (NPS) since 2003. Before serving at NPS, RADM Greene was an independent consultant focusing on defense industry business development strategy and execution (for both the public and private sectors), minimizing life cycle costs through technology applications, alternative financing arrangements for capital-asset procurement, and "red-teaming" corporate proposals for major government procurements.

RADM Greene served as the Assistant Deputy Chief of Naval Operations (Logistics) in the Pentagon from 1991–1995. As Assistant Deputy, he provided oversight, direction and budget development for worldwide U.S. Navy logistics operations. He facilitated depot maintenance, supply chain management, base/station management, environmental programs and logistic advice, and support to the Chief of Naval Operations. Some of his focuses during this time were leading Navy-wide efforts to digitize all technical data (and, therefore, reduce cycle-time) and to develop and implement strategy for procurement of eleven Sealift ships for the rapid deployment forces. He also served as the Senior Military Assistant to the Under Secretary of Defense (Acquisition) from 1987–1990; as such, he advised and counseled the Under Secretary in directing the DoD procurement process.

From 1984–1987, he was the project manager for the AEGIS project. This was the DoD's largest acquisition project, with an annual budget in excess of $5 billion/year. The project provided oversight and management of research, development, design, production, fleet introduction, and full life cycle support of the entire fleet of AEGIS cruisers, destroyers, and weapons systems through more than 2,500 industry contracts. From 1980–1984, he served as director, committee liaison, Office of Legislative Affairs, followed by a tour as the executive assistant to the Assistant Secretary of the Navy (Shipbuilding and Logistics). From 1964–1980, RADM Greene served as a Surface Warfare Officer in various duties, culminating in Command-at-Sea. His assignments included numerous wartime deployments to Vietnam as well as to the Indian Ocean and the Persian Gulf. RADM Greene received a BS in electrical engineering from Brown University in 1964; he earned an MS in electrical engineering and an MS in business administration from the Naval Postgraduate School in 1973.

Keynote Speaker: Ms. Allison F. Stiller, Principal Civilian Deputy to the Assistant Secretary of the Navy Research, Development and Acquisition

Ms. Allison F. Stiller—Ms. Stiller's responsibilities include oversight and policy for Navy and Marine Corps research, development, and acquisition programs for shipbuilding, aviation, space, weapon systems, and communication systems. This portfolio includes oversight of more than 100,000 people and an annual budget in excess of $50 billion as well as hundreds of technical development, procurement, and sustainment programs for the Department of the Navy. Ms. Stiller also leads the Department's Acquisition Corps.

Ms. Stiller entered the Senior Executive Service in January 2004. She has spent nearly 30 years in the Department of Defense's acquisition community. Prior to her current position, Ms. Stiller served as the Deputy Assistant Secretary of the Navy for Ship Programs. In this capacity, she was responsible for executive oversight of all naval shipbuilding and associated weapon systems programs, major ship conversions, and nuclear ship refuelings, as well as the maintenance, modernization and disposal of in-service ships.

Ms. Stiller has served in a number of shipbuilding acquisition positions throughout her career where she led in the development and procurement of complex ship programs in multiple phases of the acquisition life cycle. She served as Deputy Program Manager for the design, development, acquisition, and fleet introduction of amphibious ships and landing craft. She also served on the Navy staff where her responsibilities included oversight of amphibious and auxiliary ship construction and conversion programs, as well as overall shipbuilding industrial base matters. Early in her career, Ms. Stiller also served on the Virginia Class and SEAWOLF Class attack submarine programs in the design and early construction phases of both programs.

Ms. Stiller holds a BS in Systems Engineering from the University of Virginia and a MS in Engineering Management from Virginia Tech. She is also a graduate of the JFK School of Government's Senior Executive Fellows Program at Harvard University and the Defense Systems Management College.

Throughout her career, Ms. Stiller's leadership and performance has been recognized by numerous awards including the Presidential Rank Award (Distinguished and Meritorious) and the Department of the Navy's Superior and Meritorious Civilian Service awards.

Panel 14. The Section 809 Panel Report on Findings—What You Should Know

Thursday, May 9, 2019	
9:05 a.m. – 10:15 a.m.	**Chair: David Drabkin,** Chair, Section 809 Panel **Panelists:** **Al Burman,** Commissioner, Section 809 Panel **Joseph W. Dyer, VADM, USN (Ret.),** Commissioner, Section 809 Panel **Charlie E. Williams, Jr.,** Commissioner, Section 809 Panel

David Drabkin—Dave is the Chairman of the Section 809 Panel, www.section809panel.org, Administrator, Council of Defense and Space Industries Association (CODSIA), www.codsia.org, as well as Director, Government Contract Advisory, Dixon Hughes Goodman, LLP. He is a member of the Bar of the Commonwealth of Virginia and a Vice Chair of two committees of the Public Contract Law Section of the American Bar Association. Dave is a Fellow and a member of the Board of Advisors of the National Contract Management Association (NCMA). He also serves as a guest lecturer at various Universities including GW Law School. Dave is a member of the Board of Directors of the Procurement Roundtable (PRT), the Public Contracting Institute, and CASI Global Alliance Inc. www.drabkinandassociatesllc.com He is a member of the Board of advisors of Dustoff Technologies, LLC.

Al Burman—Allan V. Burman, Ph.D. is President of Jefferson Solutions (Solutions), the government consulting practice of the Jefferson Consulting Group. Under his leadership, Solutions provides analysis, evaluation, program management and acquisition assistance and assessment services to many government departments and agencies. Dr. Burman advises firms, Congressional committees, federal and state agencies and international bodies on acquisition matters. He has testified before Congress over forty-five times on a variety of management issues during his career.

Prior to joining the firm, Dr. Burman served in policy positions in the Office of the Secretary of Defense and in the White House's Office of Management and Budget (OMB). As Administrator for Federal Procurement Policy in OMB, a Senate-confirmed position, Dr. Burman authored policy letters requiring use of "performance-based acquisition" for services contracting and use of past performance in evaluating contractor proposals. Both of these documents reinforce a shift in federal management practices from an emphasis on procedure to a focus on outcomes.

Dr. Burman is Chairman of the Procurement Round Table, a Fellow of the National Academy of Public Administration (NAPA), a Fellow and Member of the Board of Advisors of the National Contract Management Association (NCMA), a SAGE at the Partnership for Public Service and an honorary member of the National Defense Industrial Association (NDIA). He is an adjunct professor at George Mason University and at the International Law Institute. He is also a Commissioner on the Section 809 Panel aimed at streamlining and making more effective Defense acquisition policies and operations.

In 2009 he received a Federal 100 award in recognition of his contributions to the Federal information technology community and in 2018 received NCMA's Lifetime Achievement Award for exceptional and noteworthy contributions to the contract management field, the Association's highest honor.

Dr. Burman holds a Ph.D. from The George Washington University, a master's degree from Harvard University, was a Fulbright Scholar at the Institute of Political Studies, University of Bordeaux, Bordeaux, France, and graduated Summa Cum Laude, Phi Beta Kappa from Wesleyan University, Middletown, Connecticut.

Joseph W. Dyer, VADM, USN (Ret.)—Mr. Dyer is an independent consultant in the technology and defense markets. He was an executive at iRobot Corporation serving consecutively as President, corporate Chief Operating Officer, and Chief Strategy Officer. On active duty, VADM Dyer served as Commander of the Naval Air Systems Command, Commander of the Naval Air Warfare Center Aircraft Division, and as the Chief Engineer for Naval Aviation. He is a graduate of North Carolina State University with a B.S. in chemical engineering and the Naval Postgraduate School, Monterey, California with an M.S. in financial management. He is an elected Fellow in the National Academy of Public Administration and the Society of Experimental Test Pilots.

Charlie E. Williams, Jr.—As President of CWilliams LLC, Charlie Williams provides strategic advice and support to the defense industry and government officials in all matters of acquisition planning, contract formation and general contract management processes. In August 2016, Mr. Williams was appointed as a Commissioner on the Section 809 Panel, which was chartered by Congress to streamline and codify defense acquisition. Mr. Williams serves on the National Contract Management Association's Board of Directors and is the President for July 2018 - June 2019. He is also a Director on the Procurement Roundtable and a member of the Defense Acquisition University's Board of Visitors. Among other previous positions, Mr. Williams served as Director, Defense Contract Management Agency, and Deputy Assistant Secretary of the Air Force for Contracting.

Panel 15. Using Data Analytics to Improve Acquisition Performance

Thursday, May 9, 2019	
10:30 a.m. – 11:45 a.m.	**Chair: Lorna Estep,** Executive Director, Air Force Installation and Mission Support Center ***Product Life-Cycle Management for Early Acquisition Programs*** Lawrence Uchmanowicz, Siemens Government Technologies ***Predicting Federal Contractor Performance Issues Using Data Analytics*** David Gill, Internal Revenue Service; Rene Rendon and William Muir, Naval Postgraduate School ***Opportunities for Improved Acquisition Information Management in the Emerging Acquisition Environment*** Jeff Drezner and Megan McKernan, RAND Corporation

Lorna Estep—Lorna B. Estep, a member of the Senior Executive Service, is the Executive Director, Air Force Installation and Mission Support Center, Joint Base San Antonio-Lackland, Texas. She directs the management of human and financial resources in a single Air Force enterprise, utilizing a $6.4 billion annual budget to provide installation and mission support capabilities to 77 Air Force installations, 10 major commands and two direct reporting units. The center also serves as the parent organization for 10 detachments co- located at each major command and six primary subordinate units including the Air Force Civil Engineer Center, Air Force Financial Management Center of Expertise, Air Force Financial Services Center, Air Force Installation Contracting Agency, Air Force Security Forces Center and Air Force Services Activity.

Ms. Estep started her career as a Navy logistics management intern. She has directed the Joint Center for Flexible Computer Integrated Manufacturing, was the first program manager for Rapid Acquisition of Manufactured Parts, and has served as Technical Director of Information Technology Initiatives at the Naval Supply Systems Command. In these positions she has developed logistics programs for the Department of Defense, implemented one of the first integrated and agile data-driven manufacturing systems, and directed the development of complex technical data systems for the Navy.

As the Director of Joint Logistics Systems Center, Ms. Estep carried out the duties of a commanding officer for a major subordinate command. In addition, she acted as the Logistics Community Manager, an emerging organization to coordinate and implement the revised Defense Department logistics strategy for achieving Joint Vision 2010 through modern information techniques and processes. She has also served as Chief Information Officer for the Naval Sea Systems Command in Arlington, Va.; Executive Director of Headquarters Materiel Systems Group at Wright-Patterson AFB; Deputy Director for Logistics Readiness at the Pentagon; Executive Director, Air Force Global Logistics Support Center; and Deputy Director, Logistics, for Air Force Material Command. Prior to her current assignment she was the Director of Resource Integration, Deputy Chief of Staff for Logistics, Engineering and Force Protection, Headquarters U.S. Air Force, Washington, D.C.

Product Life-Cycle Management for Early Acquisition Programs

Lawrence Uchmanowicz—Siemens Government Technologies, Inc.
[larry.uchmanowicz@siemensgovt.com]

Abstract

The Department of Defense (DoD) understands the value of the "Digital Thread" that is created during the pre-Milestone A & B activities, but how does one capture and store this information for later use? Up until now, most of this information was collected by program support functions and stored in various stove-piped systems, making retrieval difficult at best, impossible at worst. The U.S. Air Force has created a template for capturing, managing, and controlling this early acquisition data in a Product Lifecycle Management (PLM suite), bringing the same engineering rigor to early acquisition data as to later engineering and technical data and contract deliverables. The Air Force and Siemens funded a trial project to create an "Early Acquisition" PLM-in-a-box template, using Siemens Teamcenter PLM suite, for the purpose of proving out the ability to push the digital thread backward into the DoD 5000 pre-Milestone A & B activities. The goal was a reusable template that can be used for any program of any size. The template produced was adopted by two programs within the Air Force (Ground-Based Strategic Deterrent and Long-Range Standoff) and partially by one program in the Navy (Stingray). The Air Force will use this model to capture and reuse pre-Milestone A & B data for future programs.

Acquisition inside the DoD has been moving rapidly to suppliers relying exclusively on Digitized, Engineering Model-based designs.[1] This causes some friction inside the DoD as many organizations are not set up to receive this digitized data and use it across the acquisition to sustainment process. The reduction of this gap in capability between the vendor and the DoD customer has been recognized as one of the biggest drivers of readiness in the coming decades (DoD, 2009). The DoD Systems Engineering Forum has developed and maintains an Acquisition Modeling and Simulation (M&S) Master Plan (DoD, 2009) with five major objectives:

- Provide Necessary Policy & Guidance
- Enhance the Technical Framework for M&S
- Improve M&S Capabilities
- Improve M&S Use
- Shape the Workforce

Although these objectives are designed to enhance the warfighter's capabilities for current platforms, it also has a fit into the Early Acquisition (pre-Milestone B) process. Chang and Modigliani (2017) pointed out that today, acquisition professionals are expected to tailor

[1] There is some confusion about what constitutes an Engineering Model versus DoD Architectural Framework (DoDAF) model. Table 2 lists the engineering artifact created by manufacturers in the Digital Design Process. DoDAF Models are used for systems design (e.g., a Data Model).

the DoDI 5000.02 on their own. This can be compared to “handing them a map and telling them to figure out the best way to drive from New York City to Los Angeles. If this is their first time traveling this route, it would take a lot of time to study the map, plan the route, talk to others about shortcuts, and encounter traffic and detours along the way. Perhaps they will reach their final destination, but not without wasting significant time and fuel.” To meet this challenge, the U.S. Air Force Life Cycle Management Center worked with the National Center for Manufacturing Sciences (NMCS) to create a workable concept to move Digitalization and Product Life-Cycle Management (PLM) earlier into the acquisition process (Lilu & Uchmanowicz, 2015). Figure 1 shows current uses of PLM inside the Air Force and where it can be used as a support mechanism earlier in the process.

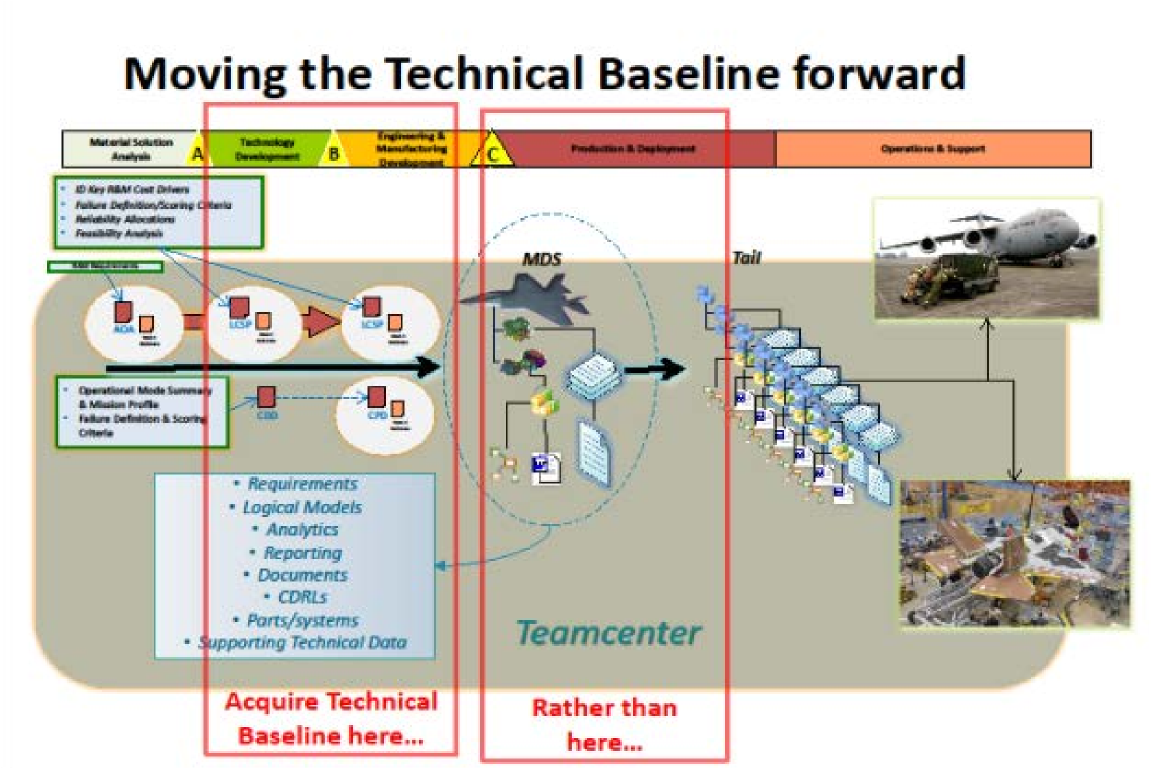


Figure 1. **PLM Overlay on DoD-5000.3 Milestones**

PLM

Product Life-Cycle Management Capability Initiative

Product Life-Cycle Management Capability Initiative (PLM-CI) is part of the Air Force (AF) Logistics Information Technology (Log IT) modernization effort. Specifically, PLM-CI is an effort to deliver an Enterprise Defense Business System (DBS) chartered to improve AF logistics and engineering through the life cycle of a product. Specifically, improvement must address common access by logistics and engineering communities to Product Life-Cycle Information (PLI). It must also provide accurate storage and quick retrieval of unclassified PLI to support efficient configuration management, integrated engineering processes across and between program offices, timely responses to customer requests for engineering and related technical assistance, and effective engineering analysis activities. These improvements should be focused on upstream acquisition activities and downstream sustainment engineering processes that impact supply and maintenance customer support (Lilu & Uchmanowicz, 2015).

PLI

PLI[2] encompasses two areas. First, PLI is defined as engineering specific stock listed and non–stock listed master product data information that includes engineering managed items, drawings and geometry, Sustainment Bills of Material (BOM), Technical Orders, maintenance specific data (e.g., master configurations, maintenance requirements, process orders), supply data (Part Master and planning BOM), Military Specification/Standard and other product specific documents requiring configuration control, and data for other functional activities including engineering assistance requests, purchasing or acquisition of parts. Second, logisticians and system/sustainment engineers produce specific product support data (i.e., PLI) as part of life-cycle planning and execution processes during acquisition of weapon systems, end items, support equipment and/or modifications (Lilu & Uchmanowicz, 2015).

The DoD logistics and engineering communities lack a standardized and integrated method of accessing PLI, managing configuration control of PLI, synchronizing changes among PLI, and sharing the PLI with downstream consumers (e.g., maintenance, planning). This results in unplanned, manual intervention of limited manpower resources on activities to create, maintain, and update product information before use (Lewis & Dwyer, 2018). Key downstream impacts include degraded planning and maintenance functions, excess inventory costs, delayed weapon system availability, lengthy repair cycles, and increased customer wait times. The absence of a single, authoritative source for engineering data also leads to inefficiencies in gathering technical information, developing and employing analytical tools, conducting analyses, and reporting/storing outcomes. The process then becomes one of time-consuming research and frequent work-arounds required to support Operational Safety, Suitability, and Effectiveness (OSS&E) assurance.

Requirements

The scope of this project was to describe and configure a PLM prototype for United States Air Force (USAF) early acquisition program activities from DoD 5000.02 pre-Milestone A up to Milestone B process (NMCS, 2017). This includes all contract deliverables in the Technical Maturity Readiness Review (TMRR).[3]

In today's Department of Defense (DoD), there is a growing need for the services to own the technical baseline. In the past, much of the technical data required to keep a weapon system or support system operational was maintained and updated by the Original Equipment Manufacturer (OEM). This was considered a best practice, and in many cases, still is. But with the advent of new technologies, and the considerably longer predicted life cycles of existing and future platforms, the DoD realizes that information gathered early in the acquisition process will be the foundation of a robust digital thread that will grow throughout the system's life cycle. Lewis and Dwyer state that to drive achievement of these objectives as rapidly and economically as possible, we organizations (i.e., Army Futures

---

[2] PLI is defined as "life-cycle logistics planning data, part items, bills of material, geometry (models and drawings), product structure and technical order data" (Lilu & Uchmanowicz, 2015).
[3] TMRR usually involves two vendors; the design used in this development can support *n* vendors (for example, an engine fly off involving three or more teams).

Command, etc.) must become a “digital data driven organization leveraging a modern PLM platform to reap the benefits of rapid, accurate collaboration across the … Department of Defense” (Lewis & Dwyer, 2018).

In the Air Force, this is referred to as “Owning the Technical Baseline,” and leadership within the Air Force Materiel Command (AFMC) realizes that in order to capture this data at any level, requires tools and processes to capture, store, and analyze data received early in the process (AFLCMC, 2016). Specific requirements for this initiative include:

- Baseline of all Request for Proposal (RFP) data including requirements and program documents
- Receipt and review of CDRLs/data
- Configuration data management change processes
- Population of weapon system data
- Enablement of DOORS integration
- CAD Models and other MBSE Model integration

New Approach to Acquisition

The scope of this project was to describe and configure a PLM prototype for USAF early acquisition program activities from DoD 5000.02 pre-Milestone A up to Milestone B process (NMCS, 2017). As Chang and Modigliani (2017) pointed out, tailoring of acquisition models helps to focus programs on their particular core elements. Acquisition professionals can navigate the acquisition life cycle faster, by leveraging the best practices and exemplar strategies of many previous programs. Siemens and the Air Force (AF) exercised their commercial expertise and practices, such as the AdvantEdge Delivery methodology (Siemens PL Software, 2016), to deliver a Commercial off the Shelf (COTS) solution, “PLM-in-a-Box—Early Acquisition Edition” (Lilu & Uchmanowicz, 2015). The team partnered with NMCS using an available Other Transaction Authority (OTA), specifically the Commercial Technologies for Maintenance Activities (CTMA). This is a public/private partnership that uses Agile approaches to solve government problems. The result would be a template solution containing the virtual machine application copy and supporting documentation to enable deployment to a host environment for PLM on-boarding and program management office.

The team designed and deployed the solution to the Ground Based Strategic Deterrent Program (GBSD) program office in October 2017, in support of their ongoing TMRR activities in DoD 5000.02 Milestone B. The scope of this effort included planned and documented collaboration between the GBSD program office, the U.S. Air Force Product Life Cycle Management Capability Initiative (PLM CI) effort, NMCS, and Siemens Government Technologies. The solution definition included program management needs in the areas of Requirements Management processes, Documents Management Processes, CDRL deliverables and Acceptance Processes, Engineering Change Processes, Asset Configurations and Analytics (NMCS, 2016).

The Air Force required models that produce performance results used to validate weapon system specifications to be part of technical baselines for TMRR (U.S. Air Force, 2016). Table 2 lists the model types that needed to perform TMRR. This data is used during analysis to execute the models/tools (i.e., inputs) defined in its analytical architecture, as well as the outputs generated during model/tool execution such that the government can regenerate the data (U.S. Air Force, 2016).

Execution

Phased Approach

The joint Siemens/U.S. Air Force team executed the project using Siemens' Advantedge™ Agile Methodology. This consisted of 10 "sprints" over the course of a year to get capability to the user for testing and acceptance. Each of these Phases lasted approximately three weeks, with one of those weeks a workshop to make sure both the configurators and functional users were agreeing to the solution that would meet requirements.[4] Each of these sprints generally corresponded to a capability in the PLN suite (Requirements Management, Contract Data Management, CDRL Management, Document Management, etc.). The GBSD program office provided detailed requirements documentation of what data elements were required and sample workflows for the configurators to use.

Results

At the end of each sprint, the configuration would consist of updates to the data model by extension or renaming, a workflow where necessary, and an updated Business Modeler Integrated Data Environment (BMIDE) image. This would then be handed over to the functional team for user testing. Any issues or changes would be agreed upon before the next sprint would start. In practice, each sprint took much less time than planned, and the only delays were the availability of functional subject matter experts. At the end of the last sprint, a fully operational PLM system was built, tested, and working.

Deliverables

The final deliverables for this project consisted of a portable BMIDE image template for reuse by any program, and an Advantedge™ template for other programs to follow if any changes to the base template were made.[5]

Follow-On Programs

As of today, this template is in use in part on the U.S. Navy Stingray PLM instance and the Long-Range Standoff (LRSO) program who is also undergoing TMRR. Standing up a fully configured PLM suite is not an easy undertaking, and the ability to save resources and time down to two to four weeks instead of six to 10 months is noteworthy.

Availability of Template and BMIDE Image

As the NMS Charter requires any work performed under the CTMA OTA, this template is available for use across any DoD program. Currently the templates and images are under the control for USAF AFLCMC/HI organization located at Wright-Patterson AFB, OH.

---

[4] Note that this was a configuration exercise, not software development, Teamcenter© was already installed and serviceable, meeting all U.S. Air Force Approved Product List (APL) requirements. The installation process lasted three days.

[5] The GBSD BMIDE Image did have specific naming conventions applied to it at the program's request. Subsequent programs will no doubt want to do the same. This is fairly straightforward and expected.

Conclusion

The U.S. Air Force intends to adopt the "PLM-in-a-Box, Early Acquisition" Template across all new programs. The project came in ahead of schedule and under budget. As a set template is built and configured, it will save programs time and money to get acquisition data under configuration control and promises quicker reviews. There are, however, hurdles to overcome. First, bringing engineering rigor to the TMRR process is a big organizational change management issue, which also comes with a large training requirement. Second, latency and bandwidth issues in the DoD networks need to be considered.[6] In the end, the AF realizes that the transformation of the acquisition process is not only possible but that it can be achieved at the program level and at an affordable cost.

References

Chang, S., & Modigliani, P. (2017, March–April). Proactively tailored acquisition models. *Defense AT&L*.

Cummins, K. (2018). Beyond the FAR: Growth of OTAs changing procurement landscape. Professional Services Council. Retrieved from https://www.pscouncil.org/a/Content/2018

DoD Systems Engineering Forum. (2009). Acquisition modeling and simulation master plan (AMSP). Retrieved from http://www.acq.osd.mil/sse/dte/initiat_ms.html

George, D., & Nance, B. (2016). *Teamcenter™ deployment architecture and infrastructure design* (Deployment guide).

Lewis, S., & Dwyer, J. (2018). *Digitalization drives readiness* (Siemens point paper).

Lilu, D., & Uchmanowicz, L. (2015). *USAF product life-cycle management (PLM) for integrating program requirements into an asset's life cycle* (Concept paper). National Center for Manufacturing Sciences (NCMS).

National Center for Manufacturing Sciences (NMCS). (2017). *U.S. Air Force product life-cycle management (PLM) for integrating maintenance and sustainment program requirements through an asset's life cycle* (Final report).

Siemens PL Software. (2016). *AdvantEdge™ familiarization* (PowerPoint slides).

Siemens PL Software. (2016). AdvantEdge™ schedule (MS project plan).

U.S. Air Force. (2016). *Ground-based strategic deterrent (GBSD) technology maturation and risk reduction (TMRR)* (Statement of Work).

U.S. Air Force Life Cycle Management Center (AFLCMC). (2016). *Ground-based strategic deterrent (GBSD) technology maturation and risk reduction (TMRR)* (Statement of work).

[6] These are being addressed inside the U.S. Air Force by means of a Common Computing Environment based on Commercial GovCloud offerings.

Appendix

Table 1. Teamcenter© Product List

Product #	Product Name
TC030109	Teamcenter Requirements Integrator User
TC10101	Teamcenter Author
TC010231	Change Management User
TC030301	Schedule Manager User
TC030233	Contract Data Management User
TC030101	Requirements Manager User
NX13100	NX Mach 3 Product Design
NX30120	NX Viewer
TC20615	Visualization Professional
TC1DOTC	Teamcenter Deployment
TC20505	Reporting and Business Analytics
TC030107	Teamcenter Requirements Integrator/RIF/ReqIF Interface

Table 2. GBSD MBSE Model Types

Model	Description
Coordinate Systems (Frames)	A partial list of coordinate systems is provided here for reference use with the models described below. The coordinate systems (frames) listed are only partially defined and the ultimate coordinate systems required for the WS is not limited to this list. Vehicle Reference Frame – A Cartesian coordinate frame defined with the x axis pointing aft ward along the missile center line. The origin is on the missile center line and relative to a consistent and non-changing location (i.e., located at 1000” forward of the missile aft skirt edge. Vehicle Flight Control Frame – A Cartesian coordinate frame defined with the x axis pointing forward parallel to the missile center line and the origin located at the center of gravity (CG) location. This frame moves with the missile CG location during operation. Vehicle Aerodynamic Frame – A Cartesian coordinate frame defined with the x axis pointing forward along the missile center line. The origin is located at the aerodynamic moment reference point.

Propulsion Model	The propulsion system model contains all necessary elements to fully describe the boost and post-boost systems. This includes the axial thrust, action time, mass expulsion rates, including time scaling relationships of these parameters for the solid propellant rocket motors. Must also include all reaction control systems and engines used in the post-boost stage.
Mass Properties Model	The mass properties model describes the mass properties of individual components and assemblies with corresponding CG location, moments-of-inertia, and products-of-inertia (Mass Moments of Inertia [MOI], MOI, tensor). It describes how those mass properties change as a function of time via table lookup or equivalent. The Center of Gravity (CG) locations in x, y, and z is defined in Vehicle Reference Frame. The MOI tensor defined about Vehicle Flight Control Frame. The mass property data includes the GBSD operational configuration as well as GBSD test flight configurations (includes mass due to test instrumentation, etc.) This model will include weapon system growth allowance as well as baseline mass properties.
Aerodynamics Model	The aerodynamics model provides the data necessary to define inflight aerodynamic forces and moments. The aerodynamic forces define using axial force coefficients in the Vehicle Aerodynamic Frame. The aero moment coefficients follow the right-hand rule for each axis in the Vehicle Aerodynamic Frame.
GN&C Model Flight Mission Model	The GN&C model: · Provides a detailed description and derivations of all navigation, steering, guidance, and control law logic necessary to calibrate, align, and fly the missile · Models the plant (physics) and control loops of the platform mechanization (including actuators and sensors) · Models the plant and control loops of the inertial sensors For trajectory optimization, the Flight Mission Model includes the trajectory assumptions and constraints governing the mission control logic for the timing of events such as such as staging initiation, jettisons, and other events. This model also includes analysis parameters related to all trajectory shaping assumptions, constraints, and rules, which may include, but are not limited to, staging dynamic pressure constraints, shroud jettison dynamic pressure constraints, attitude rate constraints during trajectory events, azimuth direction, altitude at launch, V-gamma reentry constraints, and other data that affects range/payload performance.

Thrust Control System Model	The thrust control system model provides the control dynamics for all thrust control elements (such as gimbals and jets) which, in conjunction with the GN&C model, are sufficient to reproduce in-flight dynamics.
Separation/Staging Model	The staging/separation model describes the effect of staging separation between each of the boost and post-boost stages and the shroud. The model also includes interstage skirt jettison timing as applicable.
Dynamics Model	The dynamics model contains all necessary data to perform and simulate structural dynamics analyses including loads and control bending modes. Includes the files associated with the software used such as ANSYS, NASTRAN, solid model files, etc.
RS/RV Models	RS/RV models, which include separation, reentry, spin-up, aerodynamics, and all sub-models relevant to reentry and reentry accuracy performance.
Post-Boost Prototype Model	This model represents the system that will be demonstrated by the post-boost prototype. This model identifies and predicts the performance of the prototype and will be used to assess performance following testing.
Propellant Residual Model	The propellant residual model defines the equations and data necessary to predict residual fuel at the end of the final boost stage (boost and post-boost phase) allocated for perturbation reserves and performance margins.
Parameter Perturbation Model	The parameter perturbation describes all missile system parameters that are necessary for a Monte Carlo evaluation including statistical distributions, means, and variation parameters.
WS Solid Model	The AVE solid model includes the geometry and mass properties for all major system components. This includes locations and orientations of sensors and separation planes. The Ground Segment solid model includes a representation of the preliminary design with focus on the most impactful design elements (i.e., major changes from existing LFs and LCs).
AVE Structural Models	Models (including the files associated with the software used such as ANSYS, NASTRAN, etc.) used for structural and/or thermal analysis of AVE structure elements during, but not limited to, AVE on-alert status in the LF, AVE fly-out, and AVE in flight.
Launch Systems Structural Models	Models (including the files associated with the software used such as ANSYS, NASTRAN, etc.) used for structural analysis for the reuse of the existing facility including any modifications/additions to it. These models also include those for analysis of new LS structures and the MSS interface to the LF infrastructure.

Launch Systems Power Budget Models and HVAC Models	Models that enable calculation and simulation of LS power budget analyses to ensure that power demands can be accommodated by backup and emergency power systems. Include HVAC models that show LS mechanical systems meet thermal needs of the weapon system.
RAM Model	RAM model as described in 3.2.8.3
WS Cost Model	Model includes all estimated life-cycle costs (i.e., Acquisition, deployment and O&S) for the entire WS. The math model includes uncertainty bounds, cost estimating methodologies and relationships.
WS Deployment Model	Model supporting the results of the analysis conducted in 3.2.19.1
WS Survivability Model	Model includes survivability estimates for pre-, trans- and post-attack for the Command and Launch Systems and all phases of flight for the AVE (i.e., boost, mid-course, terminal).
WSC2 Communication Models	All math models required to assess WSC2 communications effectiveness against the *WSS* (may include but not limited to responsiveness analysis, link budget analysis, System-generated Electromagnetic Pulse analysis, etc.) These models include simulations of communications between ALSC-R and NC3 with other WSC2 elements.

Figure 2. **Sample Business Use Case**

Predicting Federal Contractor Performance Issues Using Data Analytics

David Gill—is a Supervisory IT Specialist responsible for a team of CORs who manage information technology contracts. Gill has worked at the IRS since 2006 in various roles such as Contracting Officer, Procurement Analyst, and manager for tax fraud data analytics. He has a master's in business administration and bachelor's in political science from University System of Maryland schools. [david.gill@irs.gov]

William A. Muir—PhD, is an Assistant Professor at the Naval Postgraduate School, Graduate School of Business and Public Policy. His research focuses on public-sector supply management, productivity and efficiency of logistics organizations, and inventory dynamics. His research has appeared in the *Journal of Business Logistics*, *Journal of Purchasing and Supply Management*, and the *Journal of Public Procurement*. [wamuir1@nps.edu]

Rene G. Rendon—is a nationally recognized authority in the areas of supply management, contract management, and project management. Dr. Rendon is currently on the faculty of the United States Naval Postgraduate School, where he teaches in the MBA and Master of Science programs. Prior to his appointment at the Naval Postgraduate School, he served for more than 22 years as an acquisition and contracting officer in the United States Air Force, retiring at the rank of lieutenant colonel. His Air Force career included assignments as a warranted contracting officer for the Peacekeeper ICBM, Maverick Missile, C-20 (Gulfstream IV), and the F-22 Raptor. [rgrendon@nps.edu]

Abstract

The purpose of this research is to evaluate the degree to which predictive modeling techniques can enhance the quality of contractor source selection decisions. Use risk indicators created from existing publicly available contracting datasets to predict which contractors are most likely to perform successfully. Examples of risk indicators are quantitative measurements of contractor dollar velocity, instability in federal contract business, and level of experience in performing similarly sized contracts. Examine how big data analytics can be used to augment traditional source selection techniques such as proposal evaluation and past performance/responsibility checks.

Introduction

A primary goal of public-sector contracting, and more broadly, public procurement policy, is to ensure value in the use of public funds (Dimitri, 2013; Hawkins et al., 2016; Rendon & Rendon, 2016). The concept of value creation in business-to-government exchange, while latent and challenging to assess, has taken on an increased importance in the policies surrounding public procurement (e.g., Kendall, 2015; Weichert, 2019), just as it has in the management of industrial supply chains (Hendricks & Singhal, 2003; Ketchen & Hult, 2007). However, despite the ubiquity of contracting in the public and private sectors, organizations both public and private struggle to effectively and efficiently contract out for goods and services, often failing to achieve full value for their contract dollars (Rich, 2018). In extreme cases, contractual risks and hazards may lead to severe post-award issues, such as contract failure (Rendon et al., 2014) and contract termination (Davison & Sebastian, 2006, 2009). Understanding these cases is important not only because severe contractual issues (e.g., contract failure) jeopardize value and performance of taxpayer funds but because they may diminish an agency's ability to execute critical programs and deliver governmental services core to agencies' missions.

The purpose of this research is to uncover antecedents and to develop a *predictive* model of severe contractual performance issues, such as those leading to contract failure, in transactions between federal agencies and their suppliers. Existing empirical research into this topic has been limited and has largely focused on transaction-level factors leading to contract performance problems and/or contract termination. In contrast, this research examines factors at the firm-level, and utilizing data now publically available on contractor performance and integrity issues, proposes a predictive model through the application of random decision forests, a machine learning technique. To do this, firm-level antecedents are collected from multiple, publicly available data sources, including the Federal Procurement Data System (FPDS) and the System for Award Management (SAM). Incidents of contract failure are identified within the Federal Awardee Performance and Integrity Information System (FAPIIS). The resulting random forest model exhibits excellent classification performance, as measured by out-of-bag error rate and provides information on the relative importance of firm-level factors.

The remainder of the paper is structured as follows. Within the following section, we provide a brief review of the literature on contract failure in public-sector procurement. Next, we describe the data and our modeling approach, followed by the results of our analysis. The final section provides a discussion of the findings and provides several recommendations.

Prior Research

There has been only limited research into the factors leading to severe performance issues in public-sector procurement, and more specifically, contract failure and the dissolution of government-supplier relationships. Of the prior work in this area, most research has tended to focus on the analysis of transaction-level factors and how these factors correlate to post-award contractual issues. Davison and Sebastian (2009) explored associations between product/service type and the occurrences of severe problems in contract administration, finding that performance delays were the most prevalent problem encountered by contract administrators, with problems arising from other forms of risk—proposal risk, surety and liability risk, contractual risk, and price risk—varying based on the class of goods or services under contract. Rendon et al. (2014) similarly performed an examination at the transaction level, investigating contract failure rates under defense services contracts. The authors found significant differences in failure rates based on contract type and value, but—in contrast to the findings of Davison and Sebastian (2009)—did not find the rates to differ by service type, nor did they find rates to differ by level of competition (Rendon et al., 2014). In a later analysis, Dixon et al. (2015) extended the work of Rendon et al. (2014), in part, by applying techniques common to predictive analytics (logistic regression, decision-tree analysis, neural networks) to uncover the determinants of contractor performance ratings on defense services contracts. Among their findings, Dixon et al. (2015) identified a positive relationship between the workload of contract administration personnel and the likelihood of contract failure, such that failures appear to become more likely to occur as workload increases. Along those lines, prior research (e.g., Brown & Potoski, 2003) has also emphasized the importance of post-award management activities, such as monitoring, arguing that risks of contract failure will increase if these activities are under-resourced. Most recently, Liebman and Mahoney (2017) examined contractor performance on major, public-sector information technology contracts and found performance to be lower on end-of-year (i.e., last week) purchases. Interestingly, the authors found that, upon deeper analysis, these overall performance differences appear to be most strongly driven by an individual component: perceptive evaluations made by agency chief information officers (CIOs). However, these evaluations are not linked directly to incidents of contract failure.

Our examination focuses heavily on one of the most severe cases of failure—also representing the most frequent entry in FAPIIS—the government's termination of a contractor for default of the contractor (or for cause). A termination for default is defined here as the government's exercise of a contractual right to terminate a contract, or some portion thereof, due to the failure of a contractor to perform its contractual obligations (James, 1963).[1] In general, the termination of federal contractors for default involves such "serious consequences for a contractor, they are considered drastic sanctions that should be imposed or sustained only for good grounds and on solid evidence" (GAO, 2008). Yet, terminations for default (and cause) are not uncommon in public procurement. For instance, in 1994, the Government Accountability Office noted that the General Services Administration (GSA) terminated "hundreds" of contracts annually as a result of contractor default (GAO, 1994). As of 2015, terminations for default and for cause, and other severe issues (e.g., instances of defective pricing, subcontractor non-payment) are reported to the FAPIIS (2 CFR § 200.340), with records currently numbering in the thousands. FAPIIS records remain active for five years, during which time agencies are required to review and consider information contained in the system prior to making a contract award over the simplified acquisition threshold (41 U.S.C. 2313(d)(3)). The government's acquisition policy states that contracting officers will consider information in FAPIIS when determining responsibility of a prospective contractor, and separately, when evaluating the past performance of offerors during source selections (48 CFR § 9.104-6). Accordingly, a primary intent of the government's policy (and the resulting FAPIIS system) is to provide acquisition officials across the government insight into the performance and integrity of suppliers; the effect is to broaden knowledge and, potentially, the impact of severe contractor performance and integrity issues beyond that of the individual transaction. In many cases, inclusion of a severe issue in FAAPIS may lead toward ultimate dissolution of the government's relationship with a supplier (e.g., as it applies to new contract awards). The purpose of this study is to add to the existing literature regarding contract failure in public sector procurement through the development of a predictive model for severe performance issues.

Methodology

Introduction

As the intent of this research is to generate a predictive model of performance issues, we rely on techniques from the machine learning statistical tradition, namely the random forest modeling technique (Breiman, 2001). The random forest is a supervised, tree-based prediction strategy that seeks to reduce overfitting and improve generalizability through aggregated estimates from an ensemble of trees. Model performance (e.g., predictive accuracy) is measured utilizing out-of-bag (OOB) estimates which, according to Breiman (2001), eliminates the need for reserving some portion of the data for cross-

---

[1] Similarly, FAPIIS defines a *termination for cause* as the "exercise of the Government's right by a contracting officer to completely or partially terminate a contract if the Contractor fails to comply with any contract terms and conditions, or fails to provide the Government, upon request, with adequate assurances of future performance. Terminations for Cause are similar to Terminations for Default, but are applicable to contracts awarded using commercial procedures."

validation, while reducing bias. Applications of random forests can be readily found within the scientific literature and across numerous fields. Their popularity is owed, in part, to the ability of the random forest to handle high-dimensional data with relatively few observations, while providing measures for the relative importance of variables (Grömping, 2009).

Sample

Observations of contract terminations and other severe issues were obtained from FAPIIS, which contains reports on terminations and other severe contract issues that have occurred over a five-year period (early 2014 to early 2019). As the unit of analysis for this research is the firm, and since multiple records can exist in FAPIIS for a single firm, we identify for each firm the first date that the firm was entered into FAPIIS. We uniquely identify firms by their DUNS number, resulting in 1,602 distinct entities. After joining FAPIIS records with data from SAM and data from the FPDS, and then removing entities with missing or incomplete data, we are left with a sample of 780 firms. We then pair this sample with a random sample of 780 firms who did not have severe issues reported in FAPIIS during this period, thus creating a balanced dataset. Accordingly, our final sample size is 1,560 firms—a size which is comparable to sample sizes used in prior research within this area (e.g., Dixon et al., 2015).

Response Variable

The response variable in our analysis is a binary indicator of a severe contractual performance or integrity issue, as reported/indicated in FAPIIS. Of the 780 firms in our dataset with this indicator, the first entry in FAPIIS for the majority (734 firms, or 94%) was a termination for default or cause. As previously mentioned, FAPIIS defines termination for default as exercising the government's right to completely or partially terminate a contract because of the contractor's actual or anticipated failure to perform its contractual obligations. Termination for cause is the commercial item contract version of a default termination. For the remaining 46 firms, reasons for inclusion in FAPIIS included Department of Defense Determination of Contractor Fault, Non-Responsibility Determination, and Subcontractor Payment Issues.

Explanatory Variables

We identify and utilize several firm-level features (i.e., variables) to predict occurrences of severe performance issues. Longitudinal data on a firm's contractual relationship with the federal government is obtained from the FPDS, using federal-wide data starting with Fiscal Year 2009. Summary information on total contract obligations for five recent years is shown in Figure 1. We begin our analysis with data at Fiscal Year 2009 as implementation of the Federal Funding Accountability and Transparency Act of 2006 resulted in significant improvements to the quality of contract metadata contained in FPDS (Lewis, 2017). For firms experiencing a severe performance issue, FPDS data used to calculate measures extends through the period prior to a severe issue.[2] We account for time-series components of a firm's contractual exchanges with the federal government in three ways. First, we include a variable *obligations_mean* to account for the average level

---

[2] Here, "period prior" means the starting period of the analysis (Fiscal Year 2009) through the fiscal year preceding a firm's first record in FAPIIS. Our measures do not include the fiscal year of the severe issue and associated report in FAPIIS, as contractual deobligations are likely to accompany terminations and thus may confound relationships under study.

(amount in dollars) of annualized business the entity engaged in with federal agencies. Second, we include a variable *obligations_growth* to account for the change in annual obligations over the period of analysis, operationalized as the sum of first differences of the time series data. Third, to account for stability (or variability/volatility) in exchanges with federal agencies, we take the standard deviation of first differences of annual obligations, following Doboeck et al. (2009). We label this variable *obligations_variability*, given that higher values on the measure reflect a greater degree of variability. Next, we account the level of diversification in a firm's federal clientele and in the industries that it operates in within its capacity as a federal contractor. We operationalize diversification in clientele as the pooled mean of the annualized count of distinct federal agencies that the firm conducted business with over the period of analysis. Agencies are identified using the Contracting Agency Code in FPDS; we refer to the resulting variable as *diversification_agencies*. Similarly, we operationalize diversification in industries that a firm operates in (within its capacity as a federal contractor) as the pooled mean of the annualized count of distinct North American Industry Classification System (NAICS) codes as reported in FPDS; we refer to the resulting variable as *diversification_industries*. Following the finding of Rendon et al. (2014) that contract failure rates were highest among competitively awarded contracts, we include a measure, *competition*, reflecting the average number of offers agencies received in response to solicitations for contracts awarded to the firm during the period of analysis.

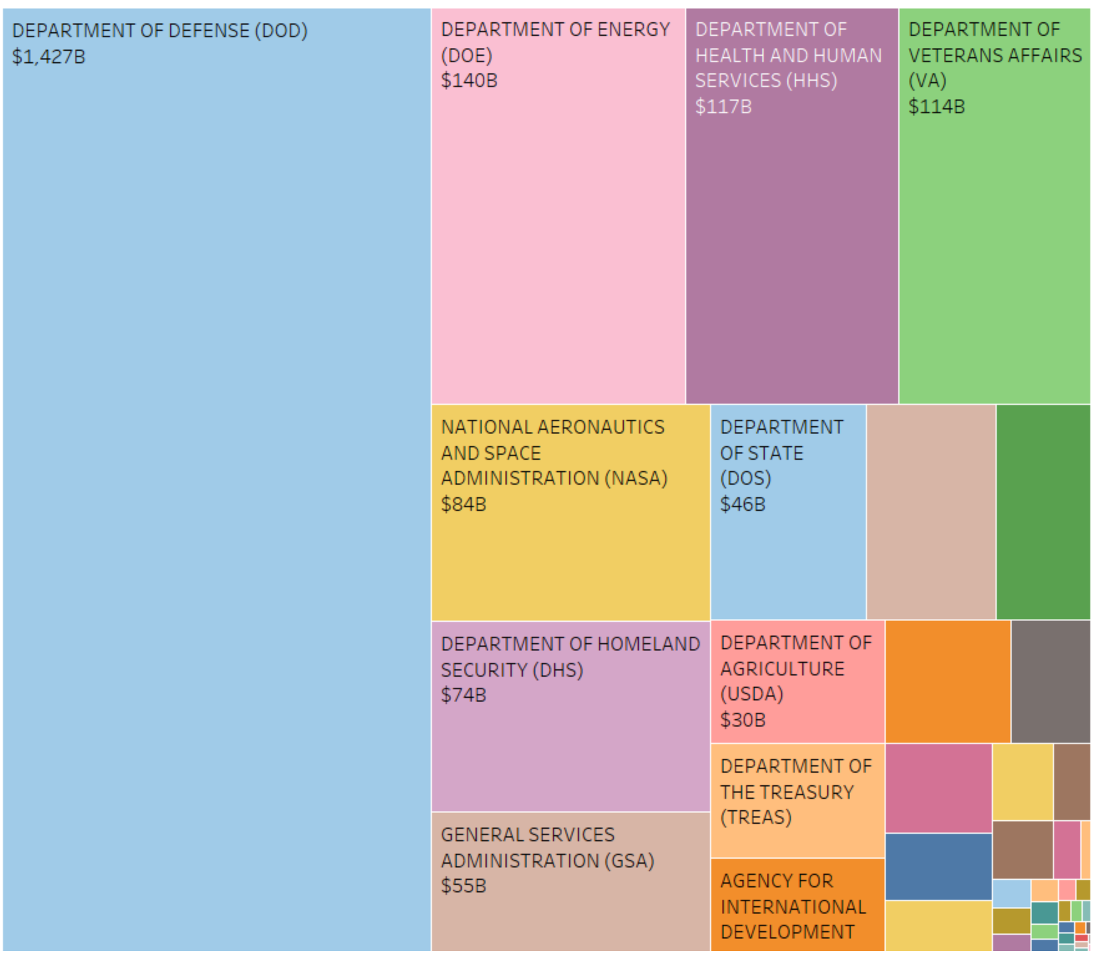


Figure 1. **Five Years of Contract Obligations (Fiscal Years 2014–2018)**

Next, we obtained entity information on the firms from SAM. We account for a firm's age in two ways, first, as the amount of time elapsed, in days, between the firm's business start date and the period prior to a severe issue (if any). SAM describes their data on Business State Date as follows: "The date the entity was started or acquired." We refer to this measure as *days_since_bus_start*. Second, we account for a firm's tenure in the federal market, *days_since_registration*, measured as the number of days between a firm's registration as a federal contractor and the period prior to a severe issue (if any). SAM describes their data on Registration Date as follows: "The date the initial entity registration was submitted, this date will not change." Lastly, we account for firm's corporate structure, using the Corporate Structure Code field in SAM; SAM defines this field as follows: "The structure of the entity as defined by the IRS, as a code." Of the total 1,560 firms, 70.32% (1097) were corporations, 8.65% (135) were partnerships, 6.03% (94) were sole proprietorships, 3.46% (54) were tax-exempt corporations, 2.44% (38) were international organizations, and 9.10% (142) fell into other categories of corporate structure.

Results

A random forest model was estimated in *R* using the *randomForest* package (Liaw & Weiner, 2002). Two parameters are primarily of interest. First, we set the model parameter corresponding to the number of predictor variables (*p*) sampled at each split to a value of three, which is equivalent to the square root of *p*; this heuristic has been found to be optimal in several empirical studies and, accordingly, is seen as a reasonable default (Strobl et al., 2008). We also observed poorer predictive performance at higher and lower values. Second, we set the number of trees at 4,096. There is no scientific standard for the number of trees to grow in a random forest; however, up to a point, the addition of trees will improve predictive performance and the interpretation of variable importance measures (Strobl et al., 2009). We observed that meaningful improvements to model performance were not realized beyond this point (Figure 2).

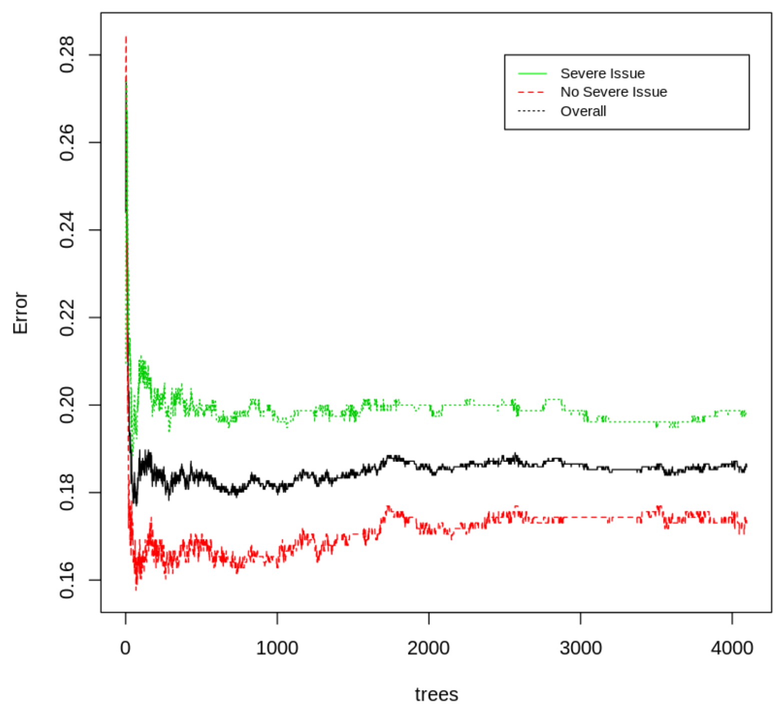


Figure 2. **Effect of Number of Trees on Model Error Rate**

Estimation of the random forest model resulted in an overall classification performance of 81.41% (i.e., 18.59% misclassification), based on the out-of-bag error rate. The model performed marginally better at classifying firms who did not experience severe issues during the period of analysis (false positive rate of 17.31%) than it did at classifying firms who did (false negative rate of 19.87%). Variable importance measures are provided within Table 1, and importance for each variable is assessed by the associated decrease in node impurities, as measured by the Gini index. Higher values of Gini importance reflect greater importance. As seen in Figure 3, the tenure of firms (days since business start and days since registration as a contractor) was the most important variable in the model, followed by industrial diversification, the average level of competition firms faced on government contracts that they won, the diversification of firms' federal clientele (agencies), the level, growth and variability of firms' business with the federal government, and, lastly, the corporate structure of firms in the sample.

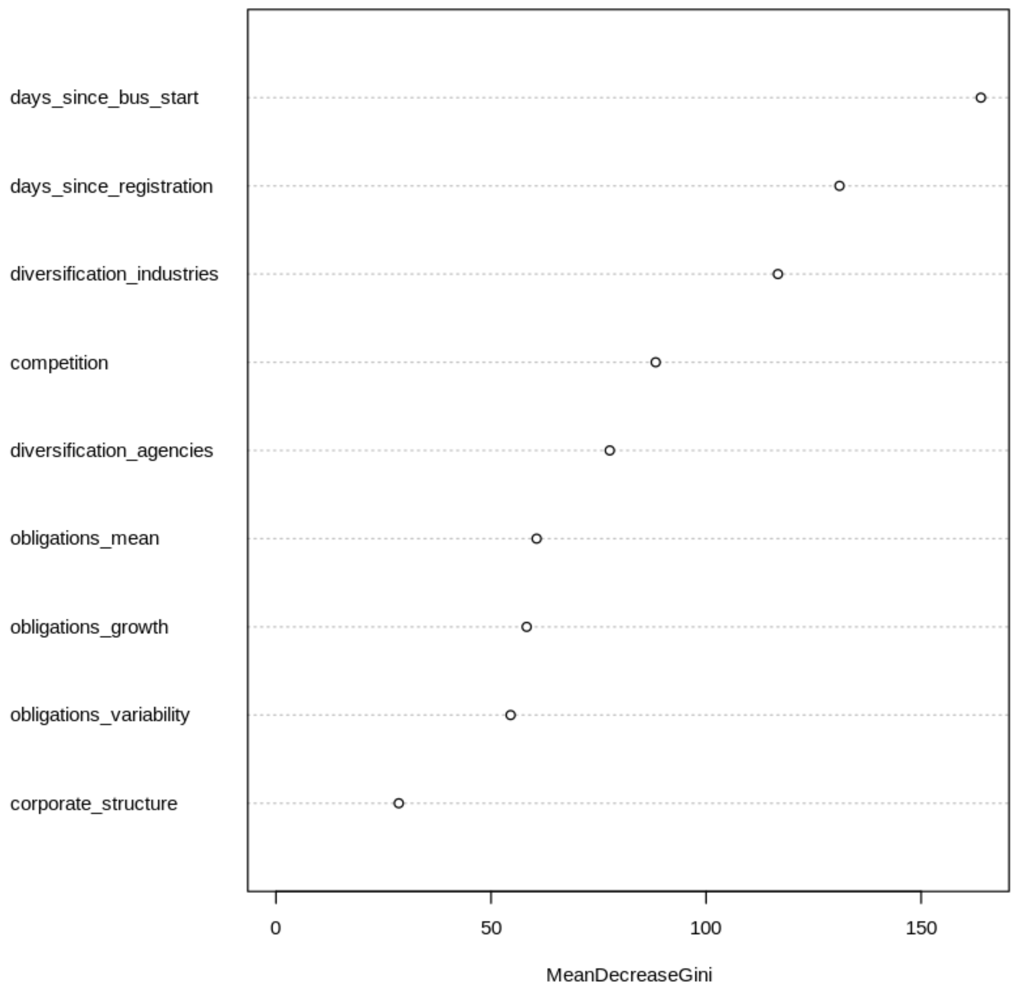


Figure 3. **Gini Importance of Variables**

Discussion

Each year the federal government receives a large quantity of contractor offers—proposals or other vendor-supplied information (e.g., oral presentations or product demonstrations). These proposals receive carefully written evaluations by government technical experts and contracting officials. Proposals and other vendor-supplied information contain a wealth of information on technical approaches to meeting specific agency mission needs. Further, proposals often contain corporate experience and past performance information. Contracting officers use platforms such as the Contractor Performance Assessment Rating System (CPARS) and FAPIIS to assess the responsibility of prospective contractors and to evaluate past performance. While both platforms are valuable, each has significant gaps. For instance, CPARS past performance narratives are sometimes not completed, have inconsistent information quality, and have a significant time lag (due to the annual evaluation cycle). FAPIIS lists contractors that experienced an adverse termination but does not track all types of contractor performance issues—only those that tend to be severe in nature. Further, both CPARS and FAPIIS are lagging indicators of performance issues. However, our results show some of the most severe contract performance issues might be predicted using publicly available data alone, as classification performance for the random forest model exceeded 80% using only nine variables. While we are careful not to suggest that a statistical model should be used in isolation in source selections, model estimates may very well serve as additional information that prompt deeper research and analysis, or when considered in concert with all other information, help to form an assessment of a prospective contractor's likely future performance.

In general, the results of this analysis suggest that a data-driven predictive modeling approach can be used to correctly identify (classify) a sizeable percentage of contractors who will later go on to experience a severe performance or integrity issue. Given that agency resources are finite, and given that a primary purpose of the public procurement is to ensure value for taxpayer dollars, predicting performance issues early—especially prior to the award of a contract—is important. Again, our analysis highlights the potential role that data analytics may serve to inform and even augment efforts by contracting officials. For one, data analytics can inform contracting officers with a picture of a vendor's overall federal business. The presence of similar, successful past contracts may evidence capability to perform on future requirements of the same type and scale. The absence of similar, federal prime contracts would likewise indicate that more information may be needed to make a determination of contractor responsibility. Further, analytics can help answer often-opaque questions regarding a contractor's ability to comply with the required delivery schedules, taking into consideration existing commercial and governmental business commitments. For example, a recent and sizable (even anomalous) increase in contract awards might prompt scrutiny regarding a contractor's capacity to handle additional volume.

Table 1. Notional Application of Data Analytics for Evaluation of Corporate Experience

Example Evaluation Factor for Corporate Experience	**Analytics-Enhanced Evaluation Factor of Corporate Experience**
The Offeror shall provide at least two, but no more than three examples of relevant and recent contracts performed. “Recent” is defined as a contract performed within the last three (3) years from the submission deadline. If a contract is ongoing, it must be at least one year into performance by the submission deadline. “Relevant” is defined as a contract that is of similar size, scope, and complexity to the requirements as set forth in this solicitation. A minimum of one (1) contract shall be the experience of the Offeror performing as a prime. The other contract(s) may be experience of the Offeror performing as a subcontractor, or the experience of a proposed subcontractor. Experience where the Offeror performed as a prime will be considered most favorably.	The Government will evaluate offerors experience on federal prime contracts of a similar size, scope, and complexity. Dollar amounts, the type of work, data derived risk metrics, and any other contract data available may be considered in the evaluation of experience. Offerors may submit a supplemental corporate experience narrative describing experience on subcontracts, non-federal contracts, and/or providing additional information regarding prime contract experience.

A primary recommendation is for the development of an open platform for *analytics* to support procurement decisions by federal agencies and their acquisition workforce. Along these lines, Executive Order 13859 (Trump, 2019) states that the government should prioritize the development of open data models and reduce barriers to their deployment. Contracting officers would benefit from having a website that summarizes data and risk characteristics of a vendor. Currently, USASpending.gov does allow viewing dollars by year and the top five transactions for specific contractors (recipients). That said, the below prototype shows the potential for more robust information that could be evaluated during a source selection. Notably, the vendor shown below experienced a rapid, five-fold increase in dollars during Fiscal Year 2018. Further, notice the two dollars (each representing a different agency that obligated over $10 million). The dots for the large dollar obligations are separated from small actions by quite a lot of whitespace (this is a relatively small vendor that suddenly received much larger than typical contracts).

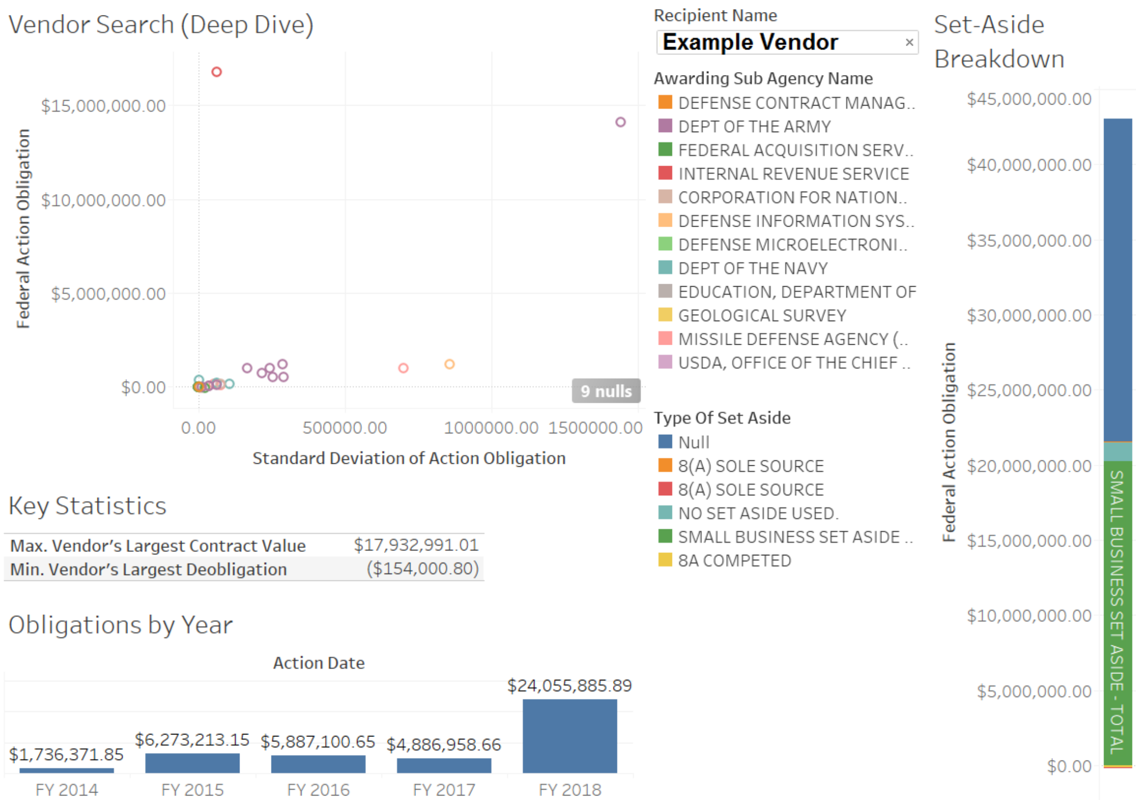


Figure 4. **Prototype of a Procurement Analytics Platform**

Limitations

We would be remiss not to acknowledge that limitations exist in our study and analysis. First, as the focus of our research is on firm-level characteristics, factors at other levels may influence the likelihood of contract failures. For instance, at the agency level, increases in an agency's capability and capacity to manage (i.e., administer) contracts and monitor contractor performance should decrease the likelihood of contract failures (Brown & Potoski, 2003). Further, evidence suggests that macroeconomic factors can exert a strong influence on the behavior of parties in buyer–supplier relationships (Krause & Ellram, 2014). We are unaware, to date, of research that has investigated how contract outcomes in the public sector might be influenced by between or within-industry variation (e.g., those relating to cyclical nature of the economy). More broadly, the literature recognizes that numerous forces come to bear on the effectiveness of public procurement, including market forces, internal forces, legal forces, social and economic forces, and forces internal to governments (Thai, 2001).

An additional limitation of our research involves the generalizability of findings. Given that our objective was to develop a *predictive* model, we sought to generalize to future times. We selected a model technique—random decision forests—as to minimize overfitting to the data (e.g., fitting to sample-specific idiosyncrasies in the data set). However, we are unable to state with certainty that the relationships uncovered in the data analysis are truly time-invariant.

Lastly, our choice of statistical technique comes with several trade-offs. One primary benefit of the approach is that it is able to account for non-linearities in the relationships between our explanatory variables and the response. However, this also comes at a cost of interpretability, as an easily interpretable coefficient reflecting the direction and magnitude of a relationship, such as what one might obtain in log-odds form from a logistic regression, is not directly estimated. As such, while we are able to assess the overall performance of the model and even assess the relative importance of variables in the model, we are not able to readily express and interpret relationships as might be accomplished after estimation of a parametric regression model.

Future Research

This research effort represents an initial, and very limited, investigation into the feasibility of predicting severe, future contractual issues through the analysis of open data. As such, it represents a first iteration, albeit one that suggests a high degree of promise. Future work should expand the variables and features under analysis (e.g., to include transaction-level variables, to include data on the economy) and explore alternative parametric and non-parametric modeling techniques.

References

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Brown, T. L., & Potoski, M. (2003). Managing contract performance: A transaction costs approach. *Journal of Policy Analysis and Management*, *22*(2), 275–297.

Davison, B., & Sebastian, R. J. (2009). The relationship between contract administration problems and contract types. *Journal of Public Procurement, 9*(2), 261–285.

Dimitri, N. (2013). ‘Best value for money’ in procurement. *Journal of Public Procurement*, *13*(2)*,* 149–175

Doboeck, P. R., Montpetit, M. A., Bergeman, C., & Boker, S. M. (2009). Using derivative estimates to describe intraindividual variability at multiple time scales. *Psychological Methods*, *14*(4), 367–386.

GAO. (1994). *Supply contract terminations: GSA is missing opportunities to recover costs from vendor default* (GGD-94-137). Washington, DC: Author.

GAO. (2008). *Termination costs are generally not compelling reason to continue programs or contracts that otherwise warrant ending* (GAO-08-379). Washington, DC: Author.

Grömping, U. (2009). Variable importance assessment in regression: Linear regression versus random forest. *The American Statistician, 63*(4), 308–319.

Hawkins, T. G., Yoder, C., & Gravier, M. J. (2016). Federal bid protests: Is the tail wagging the dog? *Journal of Public Procurement*, *16*(2), 152–190.

Hendricks, K. B., & Singhal, V.R. (2003). The effect of supply chain glitches on shareholder wealth. *Journal of Operations Management, 21*, 501–522.

James, E. P. (1963). Termination for default and for convenience of the government. *Boston College Law Review*, *5*(1), 65–78.

Kendall, F. (2015). *Implementation directive for Better Buying Power 3.0—Achieving dominant capabilities through technical excellence and innovation* [Memorandum]. Washington, DC: Office of the Undersecretary of Defense for Acquisition, Technology, and Logistics.

Ketchen, D. J., & Hult, G. T. M. (2007). Bridging organization theory and supply chain management: The case of best value supply chains. *Journal of Operations Management*, *25*, 573–580.

Lewis, G. H. (2017). Effects of federal socioeconomic contracting preferences. *Small Business Economics*, *49*(4), 763–783.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News 2*(3), 18–22.

Rendon R. G., Apte, U., & Dixon, M. (2014, February). *Services supply chain in the Department of Defense: Drivers of success in services acquisition* (NPS-CM-14-001). Monterey, CA: Naval Postgraduate School.

Rendon, J. M., & Rendon, R. G. (2016). Procurement fraud in the U.S. Department of Defense: Implications for contracting processes and internal controls. *Managerial Auditing Journal*, *31*(6/7), 748–767.

Rich, B. (2018, February). How AI is changing contracts. *Harvard Business Review*. Retrieved from https://hbr.org/2018/02/how-ai-is-changing-contracts

Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*(1), 307.

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, *14*(4), 323–348.

Thai, K. V. (2001). Public procurement re-examined. *Journal of Public Procurement*, *1*(1), 9–50.

Trump, D. J. (2019, February). *Executive order on maintaining American leadership in artificial intelligence* (Executive Order 13859). Washington, DC: Office of the President.

Weichert, M. M. (2019, March). *Category management: Making smarter use of common contract solutions and practices* (Memorandum M-19-13). Washington, DC: Office of Management and Budget.

Panel 16. Acquisition Reform—Better Schedule Estimating

Thursday, May 9, 2019	
10:30 a.m. – 11:45 a.m.	**Chair: Vice Admiral David Lewis, USN,** Director, Defense Contract Management Agency ***Making Time From Data: Toward Realistic Acquisition Schedule Estimates*** Charles Pickar, Raymond Franck, and Gregory Hildebrandt, Naval Postgraduate School ***A Comparative Analysis of Advanced Methodologies to Improve the Acquisition of Information Technology for Optimal Risk Mitigation and Decision Support Systems to Avoid Cost and Schedule Overruns*** Thomas Housel and Johnathan Mun, Naval Postgraduate School ***System Maturity Estimation During Program Execution*** John Kamp, Kaena Point Consulting LLC

Vice Admiral David Lewis, USN—Vice Adm. David H. Lewis, a native of the state of Washington, graduated in 1979 from the University of Nebraska with a Bachelor of Science in Computer Science and was commissioned through the Naval Reserve Officers Training Corps program. He graduated from the Naval Postgraduate School in 1988 with a Master of Science in Computer Science.

At sea, Lewis served aboard USS Spruance (DD 963) as communications officer, where he earned his surface warfare qualification; USS Biddle (CG 34) as fire control officer and missile battery officer; and USS Ticonderoga (CG 47) as combat systems officer. His major command assignment was Aegis Shipbuilding Program manager in Program Executive Office Ships, where he led the delivery of seven DDG 51 class ships and procured another 10 ships.

At shore, Lewis' assignments include assistant chief of staff for maintenance and engineering; commander, Naval Surface Forces; the Navy secretariat staff; Naval Sea Systems Command staff; Aegis Shipbuilding Program Office; supervisor of shipbuilding, Bath, Maine; and Readiness Support Group, San Diego.

Upon selection to flag rank in 2009, Lewis served as vice commander, Naval Sea Systems Command and then served four years as program executive officer, Ships, where he directed the delivery of 18 ships and procurement of another 51 ships. From 2014-2017 he served as commander, Space and Naval Warfare Systems Command where he led a global workforce of 10,300 civilian and military personnel who design, develop and deploy advanced communications and information capabilities. In May 2017 he assumed command as the director, Defense Contract Management Agency, Fort Lee, Virginia.

Lewis' personal awards include the Distinguished Service Medal (two awards), Legion of Merit (four awards), Meritorious Service Medal (three awards), Navy and Marine Corps Commendation Medal (two awards), Navy and Marine Corps Achievement Medal, and various service and unit awards.

Making Time From Data: Toward Realistic Acquisition Schedule Estimates

Charles Pickar—is a member of the NPS faculty where he teaches project management, defense acquisition, and systems engineering. Before joining NPS, he led the Applied Systems Engineering Program Area at the Johns Hopkins University Applied Physics Laboratory. He is a retired Army officer with extensive experience in the U.S. defense industry, including director and VP levels at Lockheed Martin, Northrop Grumman, and SAIC. He is the current Chair of the Systems Education Technical Committee of the IEEE Systems Council. His research and published work focus on applying systems engineering and system dynamics analytical approaches to defense acquisition problems. [ckpickar@nps.edu]

Raymond Franck, Brig Gen, USAF (Ret.)—retired from the faculty of the Graduate School of Business & Public Policy (GSBPP), Naval Postgraduate School (NPS) in 2012. He retired from the Air Force in 2000 in the grade of Brigadier General. His active duty career included a number of operational tours staff positions, and head of the Department of Economics and Geography, USAF Academy. His published work includes a number of journal articles and research reports on military innovation and defense acquisition management. [cfranck215@aol.com]

Abstract

This paper continues a research agenda started in 2016 with an aim of more realistic acquisition program scheduling estimates, especially for the development (SSD) phase. We discuss acquisition management as a system, and its execution (especially with respect to schedule) from the perspective of Systems Dynamics (SD). We then present two episodes from F-35 program history. We then essay an integration of the SD method with these episodes using Cooper's (1998) failure modes. Finally, we present a discussion of system performance as a potential metric for schedule estimation and analysis (through schedule estimating relationships.)

Introduction

This paper is the *fourth* in this series of investigations into identifying both alternatives to the way we do schedule estimation (process), and the schedule dynamics that impact weapons system development execution (effects). It builds on the research agenda proposed by Franck et al. in 2016 and furthered in Franck et al. in 2017 and 2018 (Franck, Hildebrandt, & Udis, 2016, 2017, 2018). The goal of this ongoing project is to examine weapons systems development schedules to both identify current state and contributing causes of schedule estimating difficulties and suggest ways to more accurately predict development duration.

Before proceeding, it is worthwhile recap the genesis of the three previous research efforts. The original intent, unchanged, is to pursue a research agenda aimed at producing more accurate schedule estimates with a focus on major defense acquisition programs. The original research questions included the following:

- What is the current state of schedule estimation and control? What's needed?
- Where are the gaps?
- How can operational performance metrics better capture contemporary operations?
- What model(s) best capture the trade-offs among program cost and schedule, as well as operational capability of fielded equipment? Can those models give

insight into “troubled programs,” with difficulties in cost, schedule, and performance?

- Analyze previous case studies (e.g., from Kennedy School of Government) for insights into program schedule drivers.
- What estimating relationships best capture time to field new hardware? What schedule drivers are generally most important?
- Based on available data, formulate and empirically test models with hypothesized schedule drivers.
- Formulate and test prediction markets for cost and schedule problems.

While many of these questions have been considered, we have not yet been able to fully answer them. This paper continues the quest to better understand the schedule estimation process and why, after so much research and practice, we still have not come to terms with accurately estimating and executing development schedules. These are some interim findings from the past three years:

1. Data science, analysis, and empirical models show the type of analysis that can be accomplished using Selected Acquisition Reports (SAR) data.
2. The mining and analysis of acquisition data helps to identify reasons for schedule delays. The reasons (Schedule Delay Factors SDF) inform planners and schedulers on additional activities and sources of delays that must be considered in schedule planning and execution.
3. Systems Dynamics and other network models that include program schedules as an integral part of the modeled acquisition process have value in explaining the nature of schedule delays.
4. Exploration of more sophisticated mathematical models that interpret the causal structure associated with program schedule achievement show promise but need more work.

Why should we care about schedule delays? The primary reason is the impact on the warfighter. Systems scheduled to reach or provide initial operational capability that are delayed by years or even decades impact the DoD’s ability to fulfill its ultimate mission of protecting the country. Contractors care about delays because delays contribute to cash flow problems, and ultimately future contracts. Taxpayers care because delays not only can ultimately increase the cost of the development but may also result in canceled program and money wasted (Stumpf, 2000).

Exploring the Concept of Schedule

Review of the literature and discussion with defense acquisition scholars and practitioners interested in schedules reveals a fundamental distinction in the concept of weapons system development schedules. The first group focuses on the time it takes to develop a weapon system (Drezner & Smith, 1990; Pugh, 1987; Rothman, 1987; Tyson et al., 1989; Van Atta et al., 2015). This is the most prevalent research focus driven by the concern in the length of time necessary to field systems. This emphasis identifies schedule as a problem of technology maturity, cost overruns, cost estimating, budget formulation, and the time it takes to deliver weapons to the field. One of the aims of this aspect of schedule research is identify ways to reduce the time necessary to field systems.

The second interest and the one pursued in this research agenda asks the question, why did it take so long? This approach, focused on the mechanics of the system development, explores the issues of realism in creating and executing weapons system development schedules. For schedule creation, we focus on the schedule development process, task duration estimation, and the fundamentals of the Critical Path and Program Evaluation Review Technique (CPM/PERT). For schedule execution, we examine the reasons the established schedule is overrun. Instead, we concentrate on the challenges of bureaucracy, high-tech, technological complexity, and maturity and ultimately accept that serendipity has a role to play in the development of advanced weapons systems. Thus, we accept the fact that acquisition programs take longer to complete. Instead, we are interested in examining the details and decisions of weapon system development, and how those details and decisions can affect the dynamics reflected in program execution length.[1]

In order to effectively examine the creation and execution of schedule, we use three main approaches. The first is a systems approach emphasizing the dynamics of both schedule creation and execution. This systems approach is based in part on the idea that planning, scheduling, and project execution must be examined as a system—that the project or program does not consist of separate and unrelated variables (Senge, 2006). The second approach uses the case study approach. Because of its interest and size, our current efforts examine the F-35. Our case study approach uses a mixed-methods analysis using data, interviews, and qualitative analysis of program reports. Finally, we have been examining schedule through a quantitative approach through earned value management.

Systems, Complexity, and Schedules

A critical point to be made when discussing weapons system development is that the act of development, that which we call a program is actually a system. A system consists of activities or parts that interact to produce something. A system uses inputs and operating through constraints and mechanisms to produce an output. An effective way to visualize a system is by using an IDEF model. IDEF (**I**ntegrated Computer-Aided [ICAM] **DEF**inition) was developed by the U.S. Air Force in 1973. IDEF was derived from a well-established modeling language, the Structured Analysis and Design Technique (SADT) (Marca & McGowan, 2005). IDEF is useful when exploring the activities of a system by identifying what functions are performed (inputs and outputs), what is needed to perform those functions (controls), and who or what is performing those functions (Mechanisms). Figure 1 shows an elementary model of a weapons system development project as a system.

Figure 1 is almost deceptive in its simplicity until one considers the volatile mix of the variables named. Inputs to the system include warfighter needs effectively translated into valid requirements. Controls or constraints include Congressional oversight and funding, as well as the constant challenge of shifting priorities. Acquisition and engineering personnel provide the mechanism for the process of development to actually occur. The output is the completed weapon system delivered to the warfighter. While easily diagrammed, no one would argue that this process is not a complex undertaking. And, while we are only examining a part of this system, no discussion on creating or executing weapons systems development schedules would be complete without considering the complexity involved.

[1] While the field of schedule development also includes operational research approaches to schedule development and estimation (e.g., Van de Vonder, Demeulemeester, & Herroelen, 2007; Vandevoorde & Vanhoucke, 2006), this aspect is not included in our study.

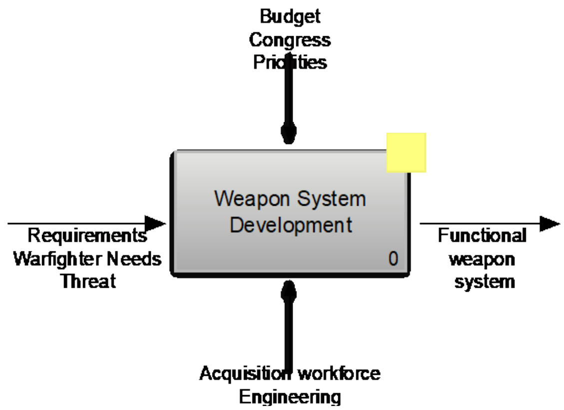


Figure 1. **Weapon System Development as System**

Because complexity science is, well, complex, we limit this discussion of complexity to three recognized types, structural complexity, detail complexity, and dynamic complexity (Dörner, 1990; Perrow, 1999; Senge, 2006; Williams, 2002). The first type, structural complexity is a construct developed by Williams that effectively captures the later classification of detail and dynamic complexity and includes the idea of uncertainty as a complexity contributor. Figure 2 shows a modified structural complexity construct (Williams, 2002). The revised graphic acknowledges the Williams' structural complexity and uncertainty, but suggests that decision dynamics is a more suitable result of uncertainty.

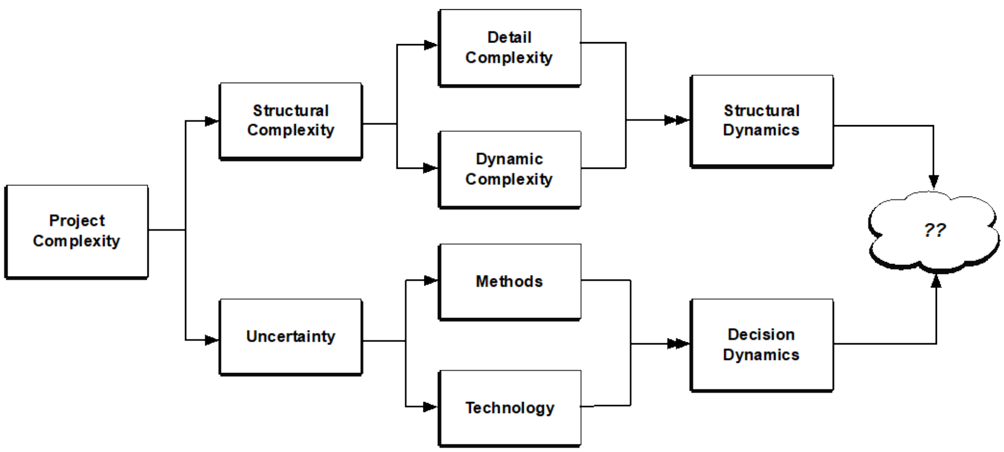


Figure 2. **Complexity Model**
(Adapted from Williams, 2002)

Detail complexity is about the size, scope, and/or the amounts of "things" in a system. It is concerned with the number and differentiation of the quantities of parts, dollars, pages in a contract, subsystems, or the size of a system, in other words, the number of variables (Baccarini, 1996). Detail complexity can often be overwhelming, but that is caused by the sheer number of elements one has to consider. Detail complexity is also the most familiar and thus addressable of these two forms of complexity because detail complexity can be captured in a spreadsheet.

Dynamic complexity is about interdependence and interrelationships and the feedback loops of various events of the development (Dörner, 1990). It is dynamic complexity that is central to the idea of schedule. We find dynamic complexity in "situations where cause and effect are subtle and where the effects over time of interventions are not obvious" (Senge, 2006, p. 70). Dynamic complexity is one of the greatest challenges that

PMs have to overcome. It is insidious in its effect because the results of dynamic complexity are not immediately apparent. Time is a critical factor in dynamic complexity:

> We rarely have trouble dealing with configurations in space. If we're not entirely sure of what we're looking at, we can take another look and resolve our uncertainty. We can normally look at forms in space again and again and in this way precisely determine their particular configuration. That is not true of configurations in time. A time configuration is available for examination only in retrospect. (Dörner, 1997, p. 100)

Managers in every industry make decisions and expect to see quick results of those decisions. In fact, this almost immediate feedback has become central to the U.S. stock market, for example. Market and industry analysts drive investors to expect to see the results of decisions often within the next quarter. However, dynamic systems and the associated complexity may or may not react in defined time frames. In reality, "*Conventional forecasting, planning and analysis methods* are not equipped to deal with dynamic complexity" [emphasis added] (Senge, 2006, p. 70).

A major manifestation of dynamic complexity is the time frame. The greatest threat to the success of a system development is not a quick, single catastrophic act, but instead the slow, almost imperceptible changes in the system that result from PM decisions (Senge, 2006). In fact, many PMs will not see the effects of their decisions before they move on to another position. This is the end state of decision dynamics.

And the problem continues because we learn best from experience. This is the benefit of experiential learning whether it is part of a curriculum or a result of on-the-job training. However, we rarely directly experience the consequences of many of our most important decisions (Senge, 2006, p. 30). This idea of project dynamics is one the DoD tends to ignore, but one we will continue to explore to better understand and explain how we can build and execute better weapons system development schedules.

Schedule Processes as System

Project planning is a well-defined and generally well understood process detailed in both DoD and the Project Management Institute (PMI) documents. Figure 3 shows a modified version of the generally accepted schedule development process from activity definition to execution. The work breakdown structure (WBS) identifies the tasks necessary for system development. WBS feeds these tasks into the scheduling process by providing activity definitions. The activity definition part of schedule estimation focuses on those activities defined by the WBS. If an activity is not named in the WBS, it is not included. This requires consideration of those tasks/activities that may not have a direct link to engineering tasks but are still essential to system development. The activities include other events, such as those imposed by the customer, in this case the DoD.

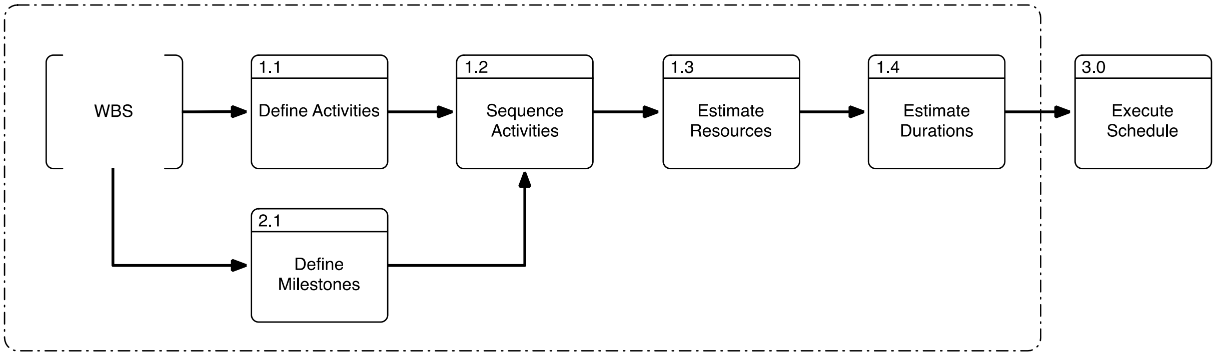


Figure 3. **Schedule Planning Process**

Activity sequencing is the process of sequencing the tasks identified through the work breakdown structure. The project planning team determines the logical sequence of tasks necessary to develop the system. At the same time, the planning team identifies those activities dependent on other activities (e.g., activities that can't start until another is finished). Correct sequencing drives efficiency in execution. However, scheduling decisions from activity definition to execution depend on the recognition and an appreciation of schedule factors that are often beyond the traditional scheduling considerations. This is reflected in the box, Function 2.1. For example, a WBS will often identify testing as an activity required to be performed many times during a development as initial assemblies are completed through integration of those assemblies into a component or subsystem. The WBS will also identify contractor reviews of testing results. However, the WBS cannot identify management attention manifested as questions to be answered (contractor and government) on the testing and potential retesting (rework), as well as emphasis on reviews that may occur if problems are identified, wherever they occurred.

This is where an appreciation of the project dynamics, the associated dynamic complexity, and ways of addressing dynamic complexity including system dynamics can be useful. While the normal scheduling process focuses on the actual tasks related to the completion of the development, system dynamics allows the addition of other, recognized relationships and their effects to the basic schedule. This allows the program manager to better anticipate potential problems.

Schedule Dynamics

Figure 4 is a graphical representation of the DoD "triple constraint" of cost, schedule, and performance. The goal on each axis is to move to the center, the cloud that depicts system completion. The red marks show the incremental attainment of the various targets of cost, schedule, and performance. The bi-directional arrows indicate the "one step forward, two steps back" progress often seen in system developments. For example, the contract point is a critical, established event, but one that is often revisited in the course of a system development. Cost is re-evaluated, performance is re-assessed, and schedules are redone. The dynamic changes occur in both directions representing the idea that the dynamics of the development consists of both success and failure (as measured through cost, schedule, and performance)—a back and forth.

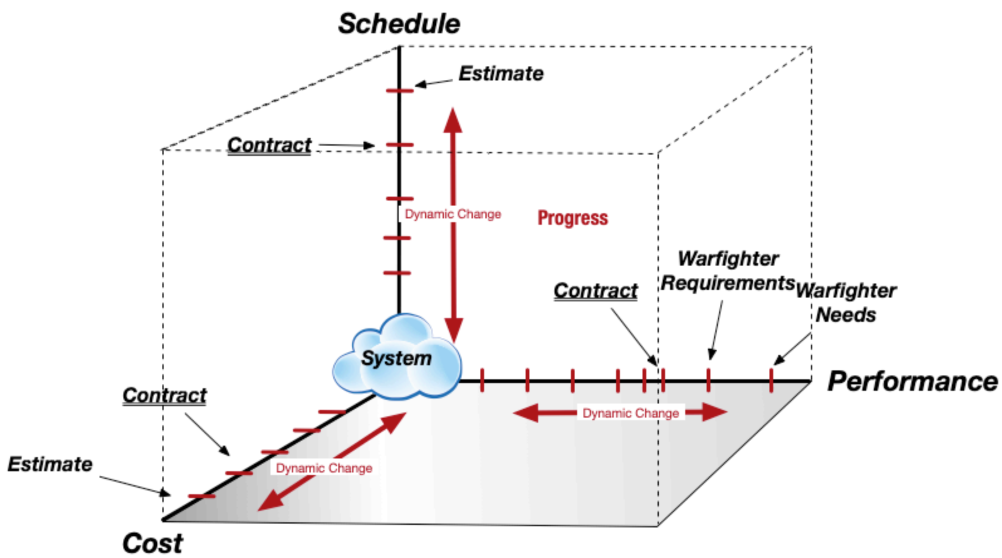


Figure 4. **The Dynamic Environment of Cost, Schedule, and Performance**

While Figure 4 emphasizes the dynamic nature of the entire project planning and execution process, it is at best a simplistic view of an extremely dynamic process. The current test of the whether the schedule was planned correctly and executed flawlessly is whether the cost, schedule, and performance axes are addressed and kept moving towards system completion. Unfortunately, in the defense world, we focus almost exclusively on cost, and to a lesser degree performance, while ignoring for the most part, the impact on and of the schedule. We have discussed this emphasis on cost and performance in previous papers.

Simply stated, the planning process—focused on cost, schedule, and performance—is itself a dynamic system. The activities on these three axes (and within the system that is the development project) change on their own through the dynamic processes of the development effort. In the execution process, the activities on these three axes are also changing. This movement creates time pressure forcing PMs to act, often with incomplete and/ or imperfect information. They can't wait to act before making a decision as failure to act, also has dynamic consequences.

> We cannot content ourselves with observing and analyzing situations at any single moment but must instead try to determine where the whole system is heading over time for many people, certainly those associated with weapon systems development, this is an extremely difficult task. (Dörner, 1997)

Consideration of the dynamic nature of the project/program management system is a question of the program manager's perspective, both government and industry contractor, and what is being measured. Anyone that has experienced a program review knows the focus is on quantitative metrics. Using an Integrated Master Schedule (IMS) and other quantitative tools including earned value, the review focuses on how we are performing to schedule and cost. We measure schedule and cost efficiency using accurate and extremely precise measures such as the cost and schedule performance index (CPI & SPI). Unfortunately, this accuracy can be misleading in light of the actual dynamics that are likely occurring. Culturally, we tend to accept metrics and computed numbers over real life. In fact, we often distort our view of real life because of computed interpretations.

The quantitatively measured progress of system development is potentially overestimated because of the focus on quantitative metrics at the expense of what is actually happening in a development (Cooper & Mullen, 1993). Specifically, the difference between the actual progress in a development effort and the actual completion rates can and often are very different. Those with project management experience will always recall the development project that slowly progresses until the "last 10%," which then takes an inordinate amount of time. That last 10% is most often due to rework, whether is a software development, hardware development, or an integration activity.

Unfortunately, the accuracy and precision afforded by our quantitative focus become accepted as the "ground truth," which leads to some of the problems we discuss in the F-35 development (Hennessy, 1996). Basically, we have created an illusion of accuracy and understanding that is not real. Further, this illusion can also affect our risk assessment, sometimes leading to false conclusions. That is not accurate. We frequently tend to ask questions focused on uncertainties. And we address the uncertainties through mathematical models based on deterministic statistical probabilities that fail to account for the exponential effects of interdependencies.

> Many projects fail to deliver against their targets because conventional project management techniques are failing to cope with the project's dynamic environment, complex interactions and the multitude of "soft"/people issues. (Mawby, 1999, p. 1)

Factors Affecting Schedule

A 1998 essay titled "Four Failures in Project Management" discusses what at the time were seen as some of the reasons for project management failure (Cooper, 1998). The essay describes the impact of a lack of systems thinking and a failure to appreciate the dynamics of a human-centered management process. Little has changed since 1998, and it is worthwhile not only to discuss the major points of that chapter, but to propose them as a framework to examine aspects of the F-35 development in the context of schedules.

The four failures are as follows:

- Failure to Know What to Expect
- Failure to Know What to Watch
- Failure to Know What to Do (and To Do It)
- Failure to Know What's What

Failure to know what to expect is about setting project targets including schedule:

> Setting and achieving an aggressive schedule is perhaps the most sacred of all sacred cows in the field of project management. It is also the source of the most destructive behavior and phenomena in projects. (Cooper 1998, p. 10)

The results of knowing what to expect are overlapped work stages, schedule pressure, resource inefficiencies, and worked morale (Cooper, 1998, p. 11).

Overlapped work stages occur when, in an effort to show progress, work is started that is scheduled later in the development in order to be able to show project progress, ultimately causing rework because of the out of sequence effects. Schedule pressure is just as it sounds: in an effort to demonstrate progress, the PM and management apply pressure on the workforce. The result of this pressure is a multiplication of the of out of sequence work and the resulting rework. Resource inefficiency occurs when the PM and management apply pressure, forcing overtime and other stress on the workforce.

Failure to know what to watch focuses on the idea of rework and the ultimate measure of quality. The "what to watch" aspect is about using perhaps the wrong tools to actually create schedules, and then not understanding what to do when rework happens. The basic challenge with the CPM/PERT scheduling method is that it does not account for what every PM knows occurs, which is rework. CPM/ PERT is a key problem because of its basic assumptions. The following are the basics of CPM/PERT:

- mean of activity duration = $(a + 4m + b)/6$
- standard deviation of activity duration = $(b - a)/6$

where a, m and b are the minimum, modal, and maximum of the activity duration

PERT uses four basic assumptions (Williams, 2002):

- there is a minimum, a maximum, and a median time provided by the estimator
- standard deviation is $(^1/^6)$ of the range $(b - a)$
- the distribution is $Beta$
- the activity durations are independent

The challenge with PERT is these assumptions. First, what is the max (a) and minimum (b)? What is the basis of these numbers? Second, given that a and b are estimates, how valid is the standard deviation? Third, why use a $beta$ distribution? Finally, although the network diagramming side of PERT is meant to disclose interdependencies and relationships, activity durations are rarely independent (Williams, 2002). The reality in today's complex projects is that the traditional methods of creating schedules are not robust enough or even complete enough for what will inevitably occur. Regardless of the causes of rework, the fact is it always occurs. We see it in the case of the F-35 weight problem discussed below, as well as any human endeavor.

The third failure is failing to know what to do. This failure points directly at the decisions a PM makes and is a result of the dynamics of the system. The fact is that a PM can influence but is hard-pressed to actually control the execution of a complex project. On the industry side, the PM is captured by his or her organization and the organizational process, as well the matrix-driven organizational structure of most defense companies. Knowing what to do is about the decisions PMs make to influence the project. Because Cooper is focused on rework, the focus of this "failure" is concerned with the decisions about how to apply resources when the project gets in trouble. A perfect example of this failure is captured by Brook's Law, to wit, "adding human resources to a late software project makes it later" (Brooks, 1995). The fact is, adding human resources to any project in progress has the unintended effect of slowing the overall project because the need to get the new people up to speed slows already slow progress, more workers end up getting in each other's way, and communication among the team members becomes challenging with the increase in numbers (keeping everyone aware of status and changes). The final failure, "What's What" relates to being able to learn. Otherwise known as lessons learned, this failure looks at an organization and its PMs' ability to actually learn from previous problems.

Complexity plus the failures provide an initial framework for analyzing existing development programs in general, and the F-35 in particular. Combining the ideas of complexity expressed as structural dynamics and decision dynamics emphasizes the issues of weapons system development complexity and the dynamics these forces create. The failures provide a means to look at development programs from a different perspective, and may also serve to help explain some of the challenges demonstrated in these programs.

Case Study: Two Episodes From the F-35 Program

Overview

The Joint Strike Fighter was originally intended to meet modest expectations: basically, a timely and affordable replacement for the F-16, F-18, and AV-8. Nonetheless it evolved and ended up a very tough task at the beginning of system development (SDD), as noted in a 2001 DoD independent cost estimate, which rated the F-35 program as high risk for both schedule and technical reasons (not an open source, but discussed in Blickstein et al., 2011, p. 37).

In particular, the original list of requirements turned out to be a highly effective way to reduce engineering "trade space." The F-35 requirements included being stealthy, supersonic, VSTOL capable (B model), and carrier capable (C model) (Blickstein et al., 2011, Table 4.6, p. 49).

The Narrow Path to Success

To accomplish a tough set of tasks in a timely manner, F-35 program management started with a number of highly optimistic fundamental premises (or "framing assumptions"). These, in turn, led to a program strategy that was success-oriented with little margin for error or surprises. Major assumptions included the following:[2]

JSF is readily available. Program management assumed the X-35 (a concept demonstrator) was a Y-35 (prototype for production, Blickstein, et al., 2011). This suggested that a development program (SDD) could proceed on an ambitious schedule and then transition quickly to full-rate production (~200 per year).

This time it's different. The program was structured (perhaps implicitly) on the promise of improved manufacturing methods and reformed acquisition practices, even though their value in practice had yet to be demonstrated. For example, an abbreviated test schedule was planned, enabled by improved computer simulation capability (unnamed source, 2018).[3] Also, new manufacturing methods, such as unitized wing, would save both time in development and money in procurement.

However, as the program progressed, system testing was generally in a catch-up mode as data from experimental airframes and computer simulations proved less useful than expected. And for example, the unitized wing was abandoned to save aircraft weight (discussed more below) but with a doubling of assembly time (Warwick, 2018).

This time it's the same. Cost estimates relied on experience gleaned from "legacy" aircraft, such as fourth-generation fighters, not accounting for, for example, increased complexity of the fifth generation. Also, the program started with a 6% weight growth allowance, in keeping with previous practice (Blickstein et al., 2011, p. 47).

Initial weight estimates used methods derived from experience with previous generations. But as one Lockheed Martin (LM) executive noted, "Legacy estimating techniques just don't work with this family of airplanes," which are highly complex, with densely-packed components in the airframe (Pappalardo, 2006).

---

[2] Franck et al. (2012) includes one discussion (esp. pp. 80–83).

[3] For which Chatham House Rules apply.

The real problem was perhaps less in the assumptions themselves, and more in the number. Even if each assumption was reasonable, it was also reasonable to expect that not all would work out. And if the road to success depends on all these bets coming in, the plan resembles a house of cards (cascading effects from small perturbations). In the event, the framing assumptions didn't all pan out, and the Joint Strike Fighter program got into trouble rather quickly.

The Weight Reduction and Redesign Episode of 2004

Because development of an operational platform was expected to be relatively quick and easy, initial design efforts could focus on cost ("affordability"), which included standard rather than custom parts. These measures added some weight. As one LM engineer put it, "The focus was very much on affordability at the time. People realized there was a penalty to be paid, and that was included in the weight estimates. It was higher than we thought" (Pappalardo, 2006). One likely reason for that situation is that LM's weight estimates were based on previous experience, as noted above.

The weight problems became obvious in 2003. The emerging F-35 design would be significantly over estimated weight, which would jeopardize meeting the program's KPPs (key performance parameters). Accordingly, weight was treated as an existential threat to the program, especially the STOVL model.

Weight Reduction Program Through Redesign (The Mother of All Rework Events)[4]

The weight problem brought the program to a "screeching halt" on April 7, 2004—with a "stand down" day. LM people were told that all work would stop until the weight problem was solved. This effort included substantial redesign work. LM's main focus shifted from affordability to "what's the lightest way to make it," according to another LM engineer.

The work was organized through a special project group called SWAT (Structural Weight Attack Team). SWAT was given very broad powers to waive LM's standard design change guidance and to offer incentives to employees who had weight reduction ideas. Supply chain firms were also involved and were credited with 586 pounds at the end.

Performance tradeoffs were likewise not off-limits. F-35B air-to-ground weapons carriage was reduced from two 2,000-lb bombs to 1,000 each. But a proposal to save structural weight through a reduction in maximum g-loads was disapproved by the DoD Joint Program Office (JPO).

In late 2004, LM declared victory. The exercise implemented more than 500 weight-loss recommendations. F-35B structural weight was reduced by 2,700 pounds; the A and C models 1,300 pounds each. Given the ingenuity of the engineering, some feelings of satisfaction were certainly warranted; according to one observer, "with SWAT, the program has a chance to come to fruition."

However, there were problems looming. One was cost. For example, "quick mate joints," which added 1,000 pounds to structural weight, were abandoned. To protect

[4] The main source for this section is Pappalardo's (2006) excellent article, "Weight Watchers."

commonality, the A and C models also lost their quick-mate joints. The result was an increase in manufacturing costs, due to "traditional, time-consuming" methods used instead.

Impacts known at the time were an increase in cost due to re-planning and an 18-month slip in the schedule, estimated at $6.2 billion and 18 months, respectively.

The Program Executive Officer at the time (PEO, Rear Admiral Steven Enewold) noted concerns going forward:

- increases in manufacturing costs (probably manageable);
- increased sustainability costs (unknown); and
- possible loss of durability-enhancing features ("good weight"), which was a matter of concern throughout the test program.

Continuing Concerns: That "Good Weight"

In some sense, the weight reduction exercise exchanged one set of problems for another. Among those problems was durability (operational life), especially for the B model.

Based on recent test data, the A and C models should last at least the planned operational life of 8,000 flight hours. However, estimates for expected B-model life vary considerably, from estimates of 2,100 (Trevithick, 2019) to 3,000 (DoD official), to well over 8,000 flight hours (LM, quoted in Trevithick, 2019). Part of this difference is due to characteristics of earlier vs. later production models (a result of program concurrency).

However, the F-35B encountered problems in durability testing that were significantly greater than the other models (e.g., DOT&E, 2010). At least some of this is due to the weight reduction exercise. For example, the 2010 DOT&E report on F-35 testing noted (p. 16), "The difference in bulkhead material is due to actions taken several years ago to reduce the weight of the STOVL aircraft. However, LM has recently stated that these problems are now solved: "The F-35B has completed full scale durability testing to 16,000 hours. Planned modifications and fleet management of the early contract F-35B aircraft will ensure that they meet the 8,000-hour service life requirement, and aircraft delivering today incorporate these design changes in the build process to ensure they'll meet 8,000 hours or more" (Trevithick, 2019).

However, DoD's Director of Operational Testing & Evaluation (DOT&E, 2019) had a less optimistic assessment for the B model. Early production units have expected operational lives significantly less than 8,000 hours, perhaps as little as 2,100 hours. This could mean B-model retirements as soon as 2026 or expensive retrofits. Moreover, the B-model was unable to complete its three-lifetime test profile, terminated due to numerous repairs on the test aircraft (p. 25).

Other issues have emerged. For example, a safety valve removal in 2008 (40+ pounds weight reduction) raised issues of aircraft vulnerability to combat damage (Copaccio, 2013).

An Engine Episode[5]

An interesting, and related, episode concerned the evolution of the F-119 engine (from the F-22) to the F-135.

In the early 1990s' programs, development efforts for a new strike fighter included with Advanced STOVL (ASTOVL) and Common Affordable Lightweight Fighter (CALF) programs. At this time, the strike fighter was viewed as being lightweight; one F-119 engine was deemed sufficient.[6] The problem emerged when specifications grew with JAST, and affordability was pursued, accepting increases in weight. The weight problem was not discovered quickly because of the parametric weight estimating models discussed above (Pappalardo, 2006; Warwick, 2018).

With increased requirements came an effort to increase F-119 thrust; at some point, the upgraded F-119 became the F-135. With the upgrade came a change in the JSF morphology, which necessitated a redesign, with a number of cascading effects, as reported in the RAND Root Cause Analysis (Blickstein et al., 2011). This RAND analysis reported and cascading major effects from this upgrade:

> Changes in the engine contributed to the weight growth of the JSF. Original plans called for the JSF to use the same engine as the F-22—the F-119 engine. However, the F-119 proved to be underpowered for the performance desired of the F-35, so the F-119 engine was altered to generate more thrust and became the F-135 engine. By enlarging the F-119 engine into the F-135 engine, engineering issues such as shaft length and efficiency had to be dealt with. However, the increase in thrust also lead to an increase in the engine size by a reported 1.5 inches in diameter.[7] *This small change in the engine generated a need to redesign the airframe, which in turn changed everything from aerodynamics to stealth signature, all of which needed to be re-baselined.* This engine issue also indicates lack of integration across the major contractors, which was Lockheed's responsibility as the prime contractor. (Blickstein et al., 2011; emphasis added)

However, the record also indicates that a need for a redesigned F-119 engine with increased thrust was recognized early in the program. That was a significant part of a 1997 contract with Pratt & Whitney in 1997 (Keijsper, 2007, p. 192). PW received a 10-year contract to develop the F-135 ("evolved" F-119) shortly after the F-35 source selection (over

---

[5] This is not THE engine episode. Another—in 2014—involved an engine fire traced to engine fan blades rubbing against their grooves.

[6] For example, the ASTOVL program was bound to an empty weight of 24,000 pounds. (Global Security, CALF). The F-119 was capable of supporting STOVL operations at that weight. However, the empty weight of the F-35A is about 29,000 lbs, an increase of 20+%. The F-135 max thrust is about 43,000 lbs and increase of 20+% above the F-119. So, using this back-of-the-envelope comparison, development of the F-135 makes good sense.

[7] There is some ambiguity in the open literature. Standard sources state that the F-119 and F-135 have the same diameter. However, the F-135 is longer: 220 inches vs. 203.

Boeing's F-32) in October 2001 (Global Security, Pratt & Whitney F-135 Engine). The first F-135 production unit was delivered in 2009 (Pratt & Whitney F-135, 2019).[8]

Engine Development: From F-119 To F-135

Although this paper focuses on the JSF program after Milestone B, events that preceded selection of the F-35 provide useful context. In May 1994, the Joint Advanced Strike Technology (JAST) program began. Early on the program focused on a single-engine, one-crewmember approach with affordability being a significant part of the rationale.

In July, the Advanced STOVL (ASTOVL) program chose GE, PW (with Allison) to conduct derivative engine studies, leading to demonstrations in FY97. Major issues at the time included single-engine reliability (Navy concern) and thrust.

The JAST and ASTOVL programs merged in October 1994 as JAST. In November, contracts were let for preliminary design of F-119 derivative. GE F-120 received less funding as an alternate engine.

In December 1994, Boeing, Lockheed Martin (LM) and McDonnell Douglas (with BAE) received 15-month conceptual design contracts. In the spring of 1995, all three JAST contractor teams choose the PW F-119 as the preferred engine for their development aircraft (JAST, n.d.).

In May 1996, the JAST program was renamed Joint Strike Fighter (JSF). In January 1997, PW received a contract to develop F-119 derivatives for the Boeing and LM test aircraft (Keijsper, 2007, p. 193). The DoD chose a Government Furnished Equipment (GFE) engine approach. That is, PW would supply engines to the government, which would then be delivered to Boeing and LM as GFE. There were various STOVL-variant problems. But most ground test objectives were met by the end of 2000.

After Milestone B

On October 26, 2001 (shortly after source selection), Pratt and Whitney received a 10-year contract for the design, development, fabrication, and test of the F-135 propulsion system and supporting equipment. It included system test and evaluation. PW was also to provide engines suitable for the F-35 flight testing program ("Pratt & Whitney F-135 Engine," n.d.).

PW assembled its first CTOL/CV test engine in September 2003 and conducted a successful test in October. The first F-135 STOVL propulsion system tests began on April 14, 2004.

In retrospect, however, the maturing engine and airframe designs were not proceeding as a coherent whole. What apparently happened was the F-135 was in development, with implications of the evolving new engine not yet fully known to the LM airframers (Blickstein et al., 2011). In retrospect, this was likely one factor in LM's overreliance on parametric weight estimations (Pappalardo, 2006). If so, it also means that LM not only had to rework the fuselage to save weight, but also to change the fuselage itself to deal with the F-135 engine.

---

[8] The Wikipedia article references a 2009 PW press release. That link is now broken.

Given the RAND findings above, it appears the F-135 (relative to F-119) was not jointly understood by PW and LM. The RAND Root Cause Analysis offers the hypothesis that LM failed to carry out this part of its prime contractor responsibilities (Blickstein et al., 2011).

Another interesting hypothesis is that the DoD decided to deal directly with Pratt & Whitney for the various engine variants associated with the Boeing (X-32) and Lockheed Martin (X-35) development efforts. The DoD would then deliver the engines to the airframers. In effect, DoD was the middleman in these transactions, which is unlikely to have improved information flow from PW to Boeing and LM.[9] That the F-135 (née F-119 variant) was in development at the same time as the F-35 airframe (and a DoD responsibility to boot) might well have been factor contributing to this outcome.

Another factor is that the F-35 airframe and the F-135 engine designs were progressing concurrently. PW assembled its first test CTOL engine in September 2003, and its first test STOVL engine in April 2004. In that regard, it's interesting that LM formed a special team (BRAT) over the 2002–2003 time period to address weight issues and brought F-35 development to a sudden halt, and commenced a redesign effort in April 2004, with a special team called SWAT (Pappalardo, 2006).[10] Engine-airframe program concurrency was a possible factor leading to the weight reduction and redesign episode of 2004.

Conclusion

We have argued that the act of weapon system development is, in and of itself, a system. Because it is a system, it has internal interrelationships and interdependencies that can fundamentally change the internal processes and outputs of that system. The F-35 activities described above are witness to that fact.

Further, we believe the F-35 discussions above serve as examples of the "F-35 Program System" and are thus susceptible to the complexity factors, as well as the four failures Cooper described. The complexity issues create an environment for the failures to occur.

Using the discussion on systems, complexity, and the Four Failures, Table 1 is a summary of the impact the dynamics of complexity and rework can have on a weapon systems development.

---

[9] Given there were two proposals in plan (Boeing and LM), the GFE approach for engines was likely reasonable at the time. However, it did have disadvantages that appeared later.

[10] However, it doesn't appear that the F-135 core engine weight was a problem. The F-119 "dry" weight is 3,900 lbs, while the F-135 weight is 3,750 lbs. Also, the F-119 and F-135 are described as having the same diameter (46 inches). However, the F-135 overall length (including tailpipe) is 17 inches longer (220 vs. 203).

Table 1. F-35 Dynamic Challenges

Events	Failure Modes			
	What to Expect *Excessive optimism in planning/ estimating the schedule*	**What to Watch** *Understanding and accepting the impact of rework*	**What to Do** *Understanding the impacts of complexity and feedback loops*	**What's What** *Lessons learned*
X-35 Prototype Assumption	D			D
Def Acquisition Reform Benefits	S			
Success-Based Development Strategy	S		S	
Cost-Reduction Exercise		D	S,D	D
Estimation Methods		S		S
Weight Reduction Exercise		S,D	S,D	
F-135 to F-119 Evolution		D	S,D	

Note. D = Decision Dynamics; S = Structural Dynamics

Measuring Performance in A Network-Centric-Combat Environment[11]

As indicated in previous reports (e.g., Franck et al., 2017, 2018), there are good reasons to consider the issue of performance measures in developing tools to analyze acquisition schedules. However, performance has become less a matter of platform attributes and more about what the new system adds to capabilities in an information-rich, networked, system-of-systems operational environment. "You look at an effect which you want to create with the overall force and you look at your mix of platforms and determine

[11] This section is abridged to conform with proceedings page limits. A more detailed discussion will appear in our final project paper.

which can lead the design change to achieve that effect” (John Blackburn, quoted in Laird, 2018, p. 4).

Also, program managers are (or should be) mindful of trade-offs being made. among the goals of cost, performance, and schedules (CJCS, 2015, p. A-9). With a better understanding of system performance in contemporary operational environments, such decisions could be improved.

Finally, useful measures of system performance can be a useful in estimating schedules—in schedule estimating relationships among other things.

There have been serious efforts in the past to formulate scalar performance measurements. However, previous efforts (e.g., Regan & Voigt, 1988) focused almost completely on platform characteristics and not on force characteristics. Operational capability is no longer a matter by adding up platform characteristics across the force, but by how a mix of different platform types operate together in the combat environment of the near future. As one observer put it, “the focus is less on what organically can be delivered by a new proposed new fighter than on its ability to interact with other platforms to deliver the desired combat effect” (Laird, 2019).

Accordingly, this section builds on previous reports (Franck et al., 2017, 2018) with a more general (but still simple) model of air combat in the near future. The essential features of our assumed scenario are as follows:

- two modern, high-technology air combat forces (Blue and Red);
- widely shared (but varying) operational situational awareness;[12]
- decentralized allocation of weapons to identified targets (like a “combat cloud,” Deptula, 2016, esp. p. 3);
- heterogeneous forces,[13] consisting of stealthy scouts (e.g., F-35), and less stealthy weapons carriers (e.g., F-15X);

Winning this engagement (as in all Lanchester-based models) requires inflicting losses on the opposing side. Accordingly, we examine the effects on air battle results of the following variables:

- Relative force sizes (R/B): even with high technology platforms and sophisticated networks, numbers probably still matter a great deal.
- Stealthy aircraft (Scouts) are survivable in a high-threat environment, while Weapons Carriers are not (Harrigian & Marosko, 2016, pp. 2–4, 7).
- Weapons Lethality, measured as a probability of success (kill).
- Battle management capabilities. It’s not “super simple” and “just battle management” (Miller, 2016) after targets have been identified. Moreover, it appears that contemporary combat air arms, such as the U.S. Air Force, do understand these difficulties (USAF, 2016, p. 6).

[12] We understand that fully shared situational awareness is still a work in progress (e.g., Laird, 2019).
[13] The U.S. Air Force Air Superiority Flight Plan for 2030 specifically calls for both “stand-in” (stealthy fighters) and “stand-off” (weapons carriers) airborne combat forces (USAF, 2016, esp. p. 7).

The Model

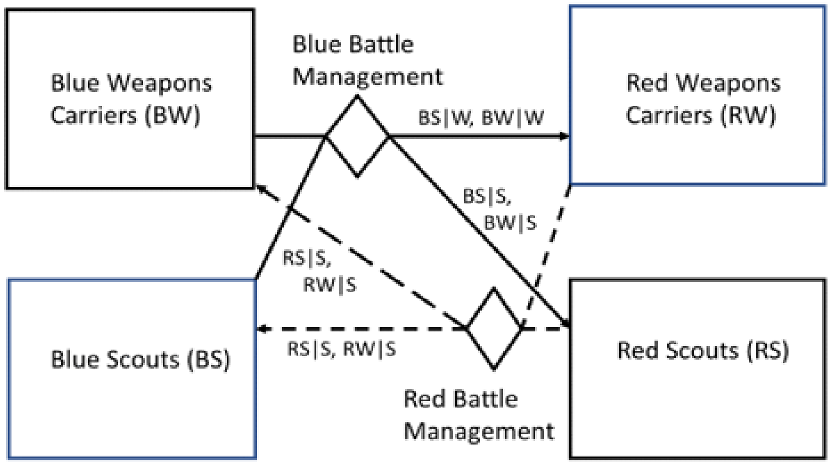


Figure 5. **Representation of a Generalized Lanchester Model of Air Combat**

Notation. XY|Z is side X (Blue or Red) units of type Y (Scout or Weapons Carrier) targeted against the opposing side's units of Type Z. For example, BS|W is number of Blue Scouts assigned against Red Weapons Carriers.

As noted above, our model involves an engagement of heterogeneous air combat forces: with stealthy scouts (with weapons) and non-stealthy weapons carriers.

Within that framework, we can consider effects of numbers, weapons lethality, stealth, and battle management effectiveness. A battle management decision process assigns Blue (Red) forces to Red (Blue) targets (that are detected and tracked). The air combat assets (both types) then attack their assigned targets.

By varying values for Blue (with Red characteristics held constant), what emerges is both interesting and suggestive. The various capabilities can be substitutes; that is, capability gaps in one characteristic can compensate for shortfalls in another characteristic. For example, Figure 6 depicts battle outcomes primarily as a function of Blue Stealth and Blue Lethality. To the upper right, Blue wins; at the lower left, Red wins. There are two curved corresponding to two levels of battle management denote a "tie."

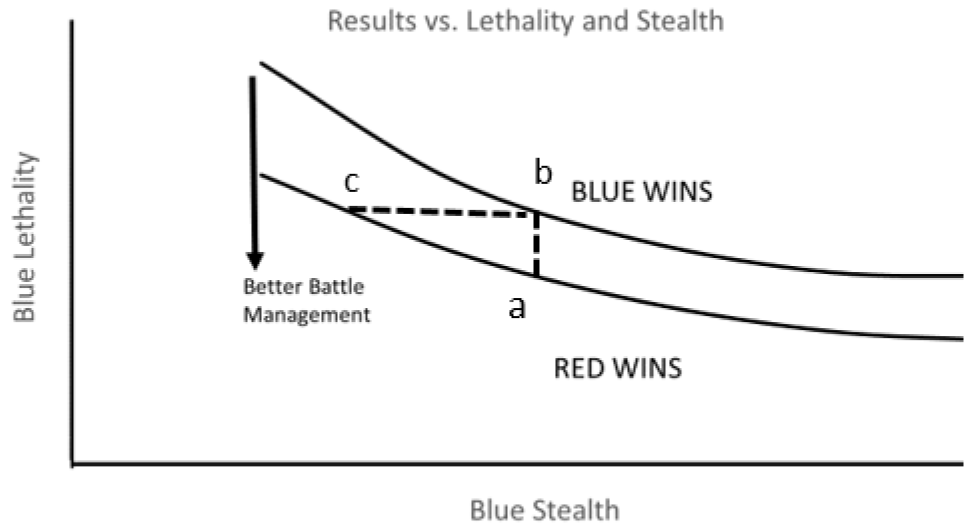


Figure 6. **Representation of a Generalized Lanchester Model of Air Combat**

Notation. XY|Z is side X (Blue or Red) units of type Y (Scout or Weapons Carrier) targeted against the opposing side's units of Type Z. For example, BS|W is number of Blue Scouts assigned against Red Weapons Carriers.

Also interesting is the relative percentage change in engagement outcome with changes in force ratio,[14] stealth, and battle management (against a Red with specified "baseline" capabilities).

These are given in Table 2.

Table 2: Responsiveness of Outcome to Changes in Force Characteristics

Variable	Force Ratio	Stealth	Lethality	Battle Management
Outcome/Variable Responsiveness	11.3	6.3	1.6	6.0

The magnitude of the numbers themselves should not be taken too seriously. The outcome variable is a measure of the margin of victory over the Red force (or defeat) rather than a raw measure of capability. In any situation of forces with about the same overall capability, any small change ("edge") can have a major effect on the margin of victory. However, the relative values are nonetheless interesting.

[14] This is an "elasticity," basically a ratio of percentage changes in outcome and force characteristic.

References

Baccarini, D. (1996). The concept of project complexity—A review. *International Journal of Project Management*, *14*(4), 201–204.

Blickstein, I. et al. (2011). Root cause analyses of Nunn-McCurdy breaches, Volume 1 Joint Strike Fighter. RAND Corporation. Retrieved from https://www.rand.org/pubs/monographs/MG1171z1.html

Brooks, F. P. (1995). *The mythical man-month: Essays on software engineering, anniversary edition, 2/E.* Addison-Wesley.

CALF. (n.d.). DARPA/Navy Common Affordable Lightweight Fighter (CALF) 1993–1994. Global Security. Retrieved February 22, 2019, from https://www.globalsecurity.org/military/systems/aircraft/calf.htm

Cooper, K. G., & Mullen, T. (1993). Swords and plowshares: The rework cycles of defense and commercial software development projects. *American Programmer*, 41–51.

Cooper, K. G. (1998). Four failures in project management. In *The Project Management Institute project management handbook (Jossey-Bass Business & Management Series)*. Jossey-Bass.

Copaccio, T. (2013). Pentagon mulls restoring F-35 safety gear to reduce risk. Retrieved from http://www.f-16.net/forum/viewtopic.php?t=22306&p=252263

Deptula, D. (2016, September). *Evolving technologies and warfare in the 21st century: Introducing the "Combat Cloud"* (Mitchell Institute Policy Papers, Vol. 4). Retrieved from http://docs.wixstatic.com/ugd/a2dd91_73faf7274e9c4e4ca605004dc6628a88.pdf

Dörner, D. (1990). The logic of failure. *Philosophical Transactions of the Royal Society*, *327*(1241), 463–473. http://doi.org/10.1098/rstb.1990.0089

Dörner, D. (1997). *The logic of failure*. Cambridge, MA: Perseus.

DOT&E (Director, Operational Test & Evaluation). (2010). Joint Strike Fighter. Retrieved from http://www.dote.osd.mil/pub/reports/FY2010/pdf/dod/2010f35jsf.pdf

DOT&E (Director of Operational Test & Evaluation). (2019). F-35 Joint Strike Fighter (JSF). Retrieved from http://www.dote.osd.mil/pub/reports/FY2018/pdf/dod/2018f35jsf.pdf

Drezner, J., & Smith, G. K. (1990). *An analysis of weapon system acquisition schedules*.

F-135 Engine. (n.d.). Pratt & Whitney F-135 engine. Global Security. Retrieved February 16, 2019, from https://www.globalsecurity.org/military/systems/aircraft/systems/f135.htm

Franck, R., Hildebrandt, G., & Udis, B. (2016). Toward realistic acquisition schedule estimates. Presented at the 13th Annual Acquisition Research Symposium, Monterey, CA.

Franck, R., Hildebrandt, G., Pickar, C., & Udis, B. (2017). Realistic acquisition schedule estimates: A follow-on inquiry. Presented at the 14th Annual Acquisition Research Symposium, Monterey, CA.

Franck, R., Hildebrandt, G., Pickar, C., & Udis, B. (2018). In the fullness of time: Towards realistic acquisition schedule estimates. Presented at the 15th Annual Acquisition Research Symposium, Monterey, CA.

Franck, R., Lewis, I., Simmonds, H., & Udis, B. (2012). *International defense acquisition management and the Fifth-Generation Fighter quandary*. Monterey, CA: Naval Postgraduate School. Retrieved from http://www.acquisitionresearch.net/files/FY2012/NPS-AM-12-214.pdf

Harrigian, J., & Marosko, M. (2016, July). *Fifth generation air combat: Maintaining the Joint Force advantage* (Michell Forum No. 6). Mitchell Institute. Retrieved from http://docs.wixstatic.com/ugd/a2dd91_bd906e69631146079c4d082d0eda1d68.pdf

Hennessy, G. (1996). From spreadsheets to system dynamic models. *The System Thinker*, 1–3.

JAST. (n.d.). Joint Advanced Strike Technology (JAST). Global Security. Retrieved February 22, 2019, from https://www.globalsecurity.org/military/systems/aircraft/jast.htm

Keijsper, G. (2007). *Joint Strike Fighter*. S. Yorkshire, England: Pen & Sword.

Marca, D. A., & McGowan, C. L. (2005). *IDEF0 and SADT: A modeler's guide*. Auburndale, MA: OpenProcess.

Laird, R. (2018, October 15). F-35 2.0 and changing the defense ecosystem. In *F-35 2.0, Second line of defense information*. Retrieved from https://sldinfo.com/wp-content/uploads/2018/09/F-35-2.0.pdf

Laird, R. (2019, February 6). Recrafting the fighter role. *Defense Information.* Retrieved from https://defense.info/air-power-dynamics/2019/02/recrafting-the-fighter-role/

Mawby, D. (1999). Managing the rework cycle not the schedule: A project management paradigm for the future? *INCOSE International Symposium*, *9*(1), 1175–1182. Retrieved from http://doi.org/10.1002/j.2334-5837.1999.tb00289.x

Pappalardo, J. (2006). Weight watchers: How a team of engineers and a crash diet saved the Joint Strike Fighter. *Air & Space Magazine*. Retrieved from https://www.airspacemag.com/military-aviation/weight-watchers-13117183/#AD2ai6Ui9K7c5tdv.99

Perrow, C. (1999). *Normal accidents: Living with high-risk technologies*. Princeton, NJ: Princeton University Press.

Pratt & Whitney F-135. (2018). In *Wikipedia*. Retrieved September 25, 2018, from https://en.wikipedia.org/wiki/Pratt_%26_Whitney_F135

Pugh, P. (1987). *Where does the time go? A review of development time scales for aircraft and guided weapons*. Presented at the Development Time Scales-Estimation and Control Royal Aeronautical Society.

Rothman, M. B. (1987). *Aerospace weapon system acquisition milestones: A data base*.

Senge, P. M. (2006). *The fifth discipline* (pp. 1–368). New York, NY: The Crown Publishing Group.

Stumpf, G. R. (2000). Schedule delay analysis. *Cost Engineering*, *42*.

Trevithick, J. (2019). USMC's older F-35Bs may only be able to fly around a quarter of their expected service life (Updated). *The Drive*. Retrieved from http://www.thedrive.com/the-war-zone/26269/usmcs-older-f-35bs-may-only-be-able-to-fly-around-a-quarter-of-their-expected-service-life

Regan, J. M., & Voigt, W. J. (1988, July 31). *The TASCFORM methodology: A technique for assessing comparative force modernization* (3rd ed.; Contract No. MDA903-85-C-908). Arlington, VA: The Analytic Sciences Corporation.

Tyson, K. W., Nelson, J. R., & Om, N. I. (1989). *Acquiring major systems: Cost and schedule trends and acquisition initiative effectiveness* (IDA Paper P-2201). Institute for Defense Analysis.

U.S. Air Force Enterprise Capability Collaboration Team. (2016, May). *Air superiority: 2030 flight plan*. Retrieved from https://www.af.mil/Portals/1/documents/airpower/Air%20Superiority%202030%20Flight%20Plan.pdf

Van Atta, R. H., Kneece, R., Patterson, C. M., & Hermes, A. C. (2015). *Assessing weapon system acquisition cycle times: Setting program schedules*. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/1014965.pdf

Van de Vonder, S., Demeulemeester, E., & Herroelen, W. (2007). A classification of predictive-reactive project scheduling procedures. *Journal of Scheduling*, *10*(3), 195–207. Retrieved from http://doi.org/10.1007/s10951-007-0011-2

Vandevoorde, S., & Vanhoucke, M. (2006). A comparison of different project duration forecasting methods using earned value metrics. *International Journal of Project Management*, *24*(4), 289–302. Retrieved from http://doi.org/10.1016/j.ijproman.2005.10.004

Warwick, G. (2018, September 17). How F-35 experience could reduce hurdles to developing fighters. *Aviation Week*, p. 44. Retrieved from http://aviationweek.com/air-dominance/how-f-35-experience-could-reduce-hurdles-developing-fighters

Williams, T. (2002). *Modelling complex projects*. West Sussex, England: John Wiley & Sons.

A Comparative Analysis of Advanced Methodologies to Improve the Acquisition of Information Technology for Optimal Risk Mitigation and Decision Support Systems to Avoid Cost and Schedule Overruns

Thomas J. Housel—PhD, is a Professor in the Department of Information Sciences at the Naval Postgraduate School, Monterey, CA. [tjhousel@nps.edu]

Johnathan Mun—PhD, is a Research Professor at NPS. [jcmun@nps.edu]

Ben Carlton—is an MS student at NPS.

Ray Jones—is a PhD Student at NPS.

Abstract

This study examines five advanced decision support methodologies, Lean Six Sigma (L6σ), Balanced Score Card (BSC), Integrated Risk Management (IRM), Knowledge Value Added (KVA), and Earned Value Management (EVM), in terms of how each can support the information technology (IT) acquisition process. In addition, the study provides guidance on when each methodology should be applied during the acquisition lifecycle of IT projects. This research includes an in-depth review of each methodology in the context of the acquisition lifecycle. All acquisition projects within the DoD must go through the acquisition lifecycle. While each acquisition project is unique, all must pass a series of common hurdles to succeed. Understanding how and when the methodologies can be applied to the acquisition of IT technologies is fundamental to the success of any IT acquisition. The study concludes with a set of recommendations for the use of each methodology in the acquisition lifecycle of IT projects.

Problem Statement

A recurring issue at the U.S. Department of Defense (DoD) is that acquisitions of information technology (IT) have been fraught with schedule and cost overruns. High profile programs such as the Joint Strike Fighter, Coast Guard Deepwater program, Army Comanche, and the Navy A-12 demonstrate the need for improvement within the acquisition process. The problem is the current suite of management tools do not seem to adequately provide sufficient early warning and fidelity into the root causes of fiscal overruns in order to provide the program manager time to adequately respond to program issues. This is a problem because the capabilities promised to the warfighter are not provided in a timely manner and the over-budgeted resources used to provide the capabilities could be more efficiently allocated to other programs.

There are a number of analytical and decision support methods that can be used to improve the acquisitions of IT. This study will provide an approach that will aid practitioners in selecting the best approach for a given phase of the acquisition lifecycle for IT systems. The methodologies that were reviewed for this study included Lean Six Sigma (L6σ), Balanced Score Card (BSC), Integrated Risk Management (IRM: Risk Simulation, Parametric Forecast Models, Portfolio Optimization, Strategic Flexibility, Economic Business Case Modeling), Knowledge Value Added (KVA), and Earned Value Management (EVM).

Research Questions and Objectives

The research questions are as follows:

1. When should the methodologies be used in the acquisition lifecycle to ensure successful acquisition of IT technologies?
2. How should the methodologies be used in the acquisition lifecycle to ensure successful acquisition of IT technologies?
3. What are the risks of using each of the methodologies for IT acquisitions?

The objective of the research was to provide a set of recommendations, based on comparison and contrast of the proposed methodologies, for when and how each method can be applied to improve the acquisitions lifecycle.

Overview

The authors have conducted numerous research studies on the effectiveness of IT acquisitions in, for example, the areas of signal intelligence, ship building, and ship maintenance, to name a few.[1] The prior studies focused on the return on investment of IT, valuation of IT real options, and IT investment portfolio optimization. For example, the shipbuilding and maintenance studies demonstrated the value added of acquiring additive manufacturing (AM), laser scanning technology (LST), and collaborative product lifecycle management tools (CPLM). This prior research revealed the need to understand how the IT acquisition lifecycle should optimally be managed within the context of the DoD existing acquisition lifecycle frameworks.

The need for these IT technologies to improve productivity has been addressed in these prior studies using the KVA and IRM approach. For example, the KVA analysis of the "as-is" ship maintenance processes identified opportunities for improvement in process efficiencies. L6σ has been used for similar purposes in other studies. These methodologies identify opportunities for productivity improvement using IT. The strategic planning for the possible insertion of these technologies was further addressed in the current study by use of the BSC methodology. The standard means for managing and monitoring the progress of an IT acquisition in the DoD is generally approached using the EVM methodology.

Each methodology has its place in ensuring a successful acquisition of IT technologies. In addition to these methodologies, past acquisition studies (e.g., signal intelligence, ship maintenance and building) have utilized the IRM methodology to forecast the future value of acquiring given IT technologies as well as the risks involved in those acquisitions. The challenge for the current study was to identify and justify the application of each of the five methodologies in the acquisition lifecycle. Each has its strengths and weaknesses as prior research has pointed out, and each was investigated in terms of how they could support the entire IT acquisition lifecycle as well as their inherent limitations in doing so.

---

[1] Most of these studies can be found on the Naval Postgraduate School Acquisitions Research Program website, https://my.nps.edu/web/acqnresearch.

This study will examine the potential use of the five methodologies to improve the chances for successful IT acquisitions. The methodologies were examined within the context of the routine (e.g., the 5000 series) acquisition lifecycle for IT. For the purposes of this study, the outputs from the Joint Capabilities Integration and Development System (JCIDS) and the Planning, Programming, and Budgeting Execution (PPBE) processes are presumed correct.

Literature Review

There are other numerous management tools that might be applied to IT acquisitions (e.g., activity-based costing and TQM, to name two). However, a review of the literature supported the focus on the five main analytical methodologies identified for this study. Expanding the potential scope of this research to include other methodologies was deemed to add minimal value given that these five approaches are in current use in acquisitions management and research. It was also assumed that beginning with these five methodologies would provide a platform for inclusion of other approaches in future research. A review of each of the methodologies is provided in what follows.

Lean Six Sigma[2]

Lean Six Sigma (L6σ) is a process performance-based methodology that focuses on improving efficiency, reducing costs, improving quality, and increasing process speed. It combines two kinds of process improvement techniques: lean and six sigma. Lean focuses on optimizing processes by eliminating waste while continuing to deliver valuable process outputs (GoLeanSixSigma, n.d.). Six Sigma is a method for reducing the number of defective process outputs with the goal of increasing efficiency as well as customer/user satisfaction. L6σ examines the details of the operations of processes in search of speed and cost improvements. L6σ consists of five phases, Define, Measure, Analyze, Improve, Control (DMAIC). First, the problem must be defined in order to limit the scope to an appropriate level. During measurement, analysts must quantify the problem to develop useful data for the following phase. Analyzing includes identifying the root cause of the problem in order to develop the necessary incremental steps to correct it. Improvement consists of implementing and verifying the incremental solution. Finally, control means maintaining the solution that was implemented during the improve phase.

L6σ requires certifications for its practitioners that range from Black Belt to White Belt. Each level or belt specifies the level of expertise a belt holder has in applying L6σ (MoreSteam, 2018). The belts, or experience levels, include the following:

- A Black Belt has expert knowledge and skills related to the DMAIC methodology, Lean methods, and team leadership.
- A Green Belt has strong knowledge and skills related to the DMAIC methodology and Lean methods, but typically does not have experience with advanced statistical tools such as design of experiments (DOE).
- A Yellow Belt is trained in the general Lean Six Sigma concepts and basic tools.

[2] This review of L6σ is taken from the following website: GoLeanSixSigma (2018). What is Lean Six Sigma? Retrieved from https://goleansixsigma.com/what-is-lean-six-sigma/.

- A project Champion is a high-ranking manager who will work with a Black Belt to ensure that barriers to project success are removed and the project team has the organizational support it needs to be effective.
- A White Belt has received a small amount (several hours) of awareness training.

This methodology has been used to incrementally improve the productivity of many DoD processes. It is also in current use as a means to help justify the future use of an IT system to improve process productivity within the DoD.

Balanced Score Card

The Balanced Score Card (BSC) is a strategic planning and management methodology developed by Kaplan and Norton (1996). The BSC includes financial metrics as well as nonfinancial performance measures, such as (1) leadership, (2) customer satisfaction, and (3) employee satisfaction, to achieve a balanced view of an organization's performance (Kaplan & Norton, 1996; also see Albert, 2002, and Niven, 2008). The BSC helps to strategically align an organization's actions to the vision and strategy of the organization, improve internal and external communications, and monitor organization performance against strategic goals.

The BSC typically uses four to five critical perspectives: (1) Organizational Capacity, (2) Customer/Stakeholder Satisfaction, (3) Financial Metrics, (4) Leadership Behavior, and (5) Internal Process Performance, to design a scorecard that reflects a company's vision and strategy. An organization can then develop strategic objectives, key performance indicators (KPIs), targets and initiatives relative to each of the perspectives, so that they can measure and monitor their progress based on the BSC (Balanced Scorecard Institute, n.d.).

Thereafter, the organization will need to convert the BSC into a strategy map so that it can be used to communicate and share with the rest of the organization. This strategy map will be a basic graphic that shows a logical, cause-and-effect connection among the critical perspectives. This is an important step that leads to high level vision and strategy statements that can be shared with the rest of the organization.

The organization should be able to measure the performance of its employees and management, based on the targets set in the BSC, as well as incentivize them with recognition and rewards. One of the roles of leadership is to ensure that the strategy map, based on the BSC, is clearly communicated and shared throughout the organization, so as to avoid strategic misalignments. The goal is to ensure accountability and ownership at the management level when the BSC has been executed, and employees should know what their performance targets are and what they need to do to achieve them. The organization should also conduct regular performance reviews to update and share the short-term results with its employees and management so that changes can be made based on a review of the progress toward a completed BSC. The goal of the BSC is to improve strategic alignment of all elements of the organization to ensure the BSC targets are the focus of the organization. A regular performance review also can help to motivate an underperforming area of the organization to improve its performance.

Integrated Risk Management

Integrated Risk Management (IRM) is a comprehensive methodology that is a forward-looking, risk-based decision support system incorporating various methods such as Monte Carlo Risk Simulation, Parametric Forecast Models, Portfolio Optimization, Strategic Flexibility, and Economic Business Case Modeling. Economic business cases using standard financial cash flows and cost estimates, as well as non-economic variables such as expected military value, strategic value, and other domain-specific SME metrics (e.g., Innovation Index, Conversion Capability, Ability to Meet Future Threats, Force Structure, Modernization and Technical Sophistication, Combat Readiness, Sustainability, Future Readiness to Meet Threats) can be incorporated. These metrics can be forecasted as well as risk-simulated to account for their uncertainties and modeled to determine their returns to acquisition cost (e.g., return on investment for innovation or return on sustainability). Capital investment and acquisition decisions within IT portfolios can then be tentatively made, subject to any budgetary, manpower, and schedule constraints.

In the U.S. military context, risk analysis, real options analysis, and portfolio optimization techniques are enablers of a new way of approaching the problems of estimating return on investment (ROI) and estimating the risk-value of various strategic real options. There are many new DoD requirements for using more advanced analytical techniques. For instance, the Clinger-Cohen Act of 1996 mandates the use of *portfolio management* for all federal agencies. The Government Accountability Office's (1997) *Assessing Risks and Returns: A Guide for Evaluating Federal Agencies' IT Investment Decision-Making* (Ver. 1) requires that IT investments apply ROI measures. DoD Directive 8115.01, issued October 2005, mandates the use of performance metrics based on outputs, with ROI analysis required for all current and planned IT investments. DoD Directive 8115.bb implements policy and assigns responsibilities for the management of DoD IT investments as portfolios within the DoD Enterprise, where they defined a portfolio to include outcome performance measures and an expected return on investment. The DoD (2017) Risk Management Guidance *Defense Acquisition Guidebook* requires that alternatives to the traditional cost estimation need to be considered because legacy cost models tend not to adequately address costs associated with information systems or the risks associated with them.

Projects can be broken down into their WBS and tasks, where these tasks can be combined in complex systems dynamic structures. The cost and schedule elements for each task can be modeled and risk simulated within the system to determine the total cost and schedule risk of a certain program. Program management (PM) is oftentimes integrated with IRM methods to provide a more holistic view in terms of acquisitions of IT programs.

Knowledge Value Added

As the U.S. military is not in the business of making money, referring to revenues throughout this paper may appear to be a misnomer. For nonprofit organizations, especially in the military, we require the KVA methodology to provide the required "benefits" or "revenue" proxy estimates to run a true ROI analysis. ROI is a basic productivity ratio with revenue in the numerator and cost to generate the revenue in the denominator (i.e., ROI is revenue-cost/cost). KVA generates ROI estimates by developing a market comparable price per common unit of output multiplied by the number of outputs to achieve a total revenue estimate.

KVA is a methodology whose primary purpose is to describe all organizational outputs in common units. This provides a means to compare the current and potential future outputs of all assets (human, machine, information technology) regardless of the aggregated outputs produced. For example, the purpose of a military process may be to gather signal intelligence or plan for a ship alternation. KVA would describe the outputs of both processes in common units, thus making the ROI performance of any of the processes comparable.

KVA measures the value provided by human capital assets and IT assets by analyzing an organization, process or function at the process-level. It provides insights into each dollar of IT investment by monetizing the outputs of all assets, including intangible assets (e.g., such as that produced by IT and humans). By capturing the value of knowledge embedded in an organization's core processes (i.e., employees and IT), KVA identifies the actual cost and revenue of a process, product, or service. Because KVA identifies every process required to produce an aggregated output in terms of the historical prices and costs per common unit of output of those processes, unit costs and unit prices can be calculated. The methodology has been applied in 45 areas within the DoD, from flight scheduling applications to ship maintenance and modernization processes.

As a performance tool, the KVA methodology

- Compares all processes in terms of relative productivity
- Allocates revenues and costs to common units of output
- Measures value added by IT by the outputs it produces
- Relates outputs to cost of producing those outputs in common units

Based on the tenets of complexity theory, KVA assumes that humans and technology in organizations add value by taking inputs and changing them (measured in units of complexity) into outputs through core processes. The amount of change an asset within a process produces can be a measure of value or benefit. The following are the additional assumptions in KVA:

- Describing all process outputs in common units (e.g., using a knowledge metaphor for the descriptive language in terms of the time it takes an average employee to learn how to produce the outputs) allows historical revenue and cost data to be assigned to those processes historically.
- All outputs can be described in terms of the time required to learn how to produce them.
- Learning Time, a surrogate for procedural knowledge required to produce process outputs, is measured in common units of time. Consequently, Units of Learning Time = Common Units of Output.
- Common unit of output makes it possible to compare all outputs in terms of cost per unit as well as price per unit because revenue can now be assigned at the sub-organizational level or at a DoD process level.
- Once cost and revenue streams have been assigned to sub-organizational or DoD process outputs, normal accounting and financial performance and profitability metrics can be applied (Rodgers & Housel, 2006; Pavlou et. al., 2005; Housel & Kanevsky, 1995).

KVA differs from other nonprofit ROI models because it allows for revenue estimates, enabling the use of traditional accounting, financial performance, and profitability measures at the sub-organizational level. KVA can rank processes by the degree to which they add value to the organization or its outputs. This assists decision-makers in identifying how much processes add value. Value is quantified in two key metrics: Return-on-Knowledge (ROK: revenue/cost) and ROI (revenue-investment cost/investment cost). The KVA method has been applied to numerous military core processes across the services. It was originally developed to estimate the ROI on IT acquisitions in the telecommunications industry at the sub-corporate level and has been used for the past 17 years in the DoD, with emphasis on the Navy, to assess the potential value added by IT acquisitions to core DoD processes.

Earned Value Management

Earned Value Management (EVM) provides cost and schedule metrics to track performance in accordance with an acquisition project plan during the developmental phase of the acquisition lifecycle after the Engineering Development contract is awarded. It uses a work breakdown structure (WBS) to try to measure the performance of a program based upon the amount of planned work that is done at any point in the program management baseline (PMB). EVM uses cost and schedule metrics that aid in performance trend analysis with a focus on identifying any budget and schedule deviations from the plan.

Given the propensity of IT acquisitions to be over budget and behind schedule, EVM metrics can help program managers identify and attempt to avoid overruns and schedule deviations. When variances in cost or schedule occur, EVM data can also be used to reforecast the budget and schedule with the focus of providing program managers with accurate performance information. It uses schedule and cost estimates to find the Planned Value (PV) of a given acquisition project. Cumulative PV provides the total value that should be achieved by a specified date (Reichel, 2006). The specific label for PV is Budget Cost for Work Scheduled (BCWS) within the DoD acquisitions community. Actual Cost (AC) is the accumulated accrued costs of labor and materials. The label for AC within the DoD acquisitions community is Actual Cost of Work Performed (ACWP). Earned Value (EV) measures the progress for a given plan. The DoD acquisitions label for EV is Budgeted Cost of Work Performed (BCWP; West, 2007). It may be possible to combine EVM with the IRM methodology to track IT acquisitions projects in a timelier manner, leading to fewer cost and schedule overruns.

Earned Value Management (EVM) is used by the DoD and industry for the planning and management of projects and programs. EVM is used in project management to locate emergent problems to allow the project team to take action as early as possible. EVM has been used for process improvements, but its strength is in providing a disciplined, structured, objective, and quantitative method to integrate performance, cost, and schedule objectives for tracking contract performance (DoD, 2015).

In order to utilize EVM, there are some variables and metrics that are calculated and compared to data contained in the project plan. The basis of EVM is an accurate plan with completion rates and budgeted costs (Reichel, 2006). This plan begins with the program work breakdown structure, in which the estimated cost and time are allocated to the PMB for every subordinate task as deemed appropriate by the management team and contract requirements. Once a solid plan is complete, the schedule and cost estimates will be used to find the *Planned Value (PV).* Cumulative PV will give the total value that should be achieved up to a specified date (Reichel, 2006). Period PV can be calculated for a specified period of time such as hour, day, week, and so forth, to get the amount of work that is planned over the duration selected. PV is called *Budget Cost for Work Scheduled (BCWS)* by the DoD when EVM is used to manage acquisitions. *Actual Cost (AC)* is the accumulated costs of labor and materials required to complete the project as they are accrued. The DoD refers to AC as *Actual Cost of Work Performed (ACWP)* for EVM in relation to acquisitions. *Earned Value (EV)* is a measure of progress against the plan. When referring to EVMs that are associated with DoD acquisitions, the EV term is *Budgeted Cost of Work Performed (BCWP)*.

There are several metrics and indexes that are computed using the previously computed variables discussed in the previous section. The primary metrics that are computed are variances that are related to cost and schedule.

1. *Cost Variance (CV)* is computed by subtracting AC from EV (West, 2007). A negative CV would represent a project that is over the planned budget.
2. *Cost Performance Index (CPI)* is computed by dividing EV by AC (West, 2007). As the metric moves towards zero, the higher the risk of going over the schedule. A CPI above one would indicate that the project is likely ahead of schedule.

3. *Schedule Variance (SV)* is computed by subtracting the PV from the EV (West, 2007). A negative SV would indicate a project that is behind the schedule.
4. *Schedule Performance Index (SPI)* is computed by dividing EV by AC (West, 2007). When the SPI moves lower than one towards zero, it indicates a potential cost or budget issue. An SPI above one would show that performance is likely exceeding the plan.

EVM is used to determine if a process is operating correctly with regards to costs and schedule. Using CV and SV, the process manager will be able to know if the process is not functioning as it was planned. As each step in the process is completed, the EV is computed and compared to the PV to determine the SV. A negative SV would indicate the process is running behind and action is needed to determine what is the problem. A positive number could indicate the process is over producing and creating pockets of inventory. Additionally, the CV can be computed to determine if a process is operating out of standard. A negative number could indicate over processing and exceeding standards, where a positive variance could indicate short cuts and possible defects. Regardless, any variance from the standard would only indicate that there may be an issue and the process owner needs to "go-look-go-see" and determine what is happening. The process manager would want to investigate the first step in a process that has a variance since that would be the strongest indicator of a possible issue. EVM would be most effective to monitor a stable process that has an established standard. It would not be very effective at determining if changing a process or automating a process would produce greater value. EVM is only focused on cost, and no metric or calculation is related to value, and it doesn't give insight into determining if a process's step should be automated.

Research Methodology

A review of each of the methodologies was conducted as well as a high-level review of the current phases of the acquisition lifecycle (i.e., DoDI 5000 series). The methodologies were evaluated in terms of each major phase of the acquisition lifecycle to suggest how they might be used to enhance the likelihood of successful completion of the phase. Analysis included a review of how the general overall acquisition lifecycle approach might be modified to incorporate the benefits from the methodologies, including the original motivations for the IT acquisition per the problems/challenges identified prior to the beginning of the acquisition process. It was presumed that it was possible that the acquisition lifecycle should include a formal review of the need for the IT in the first place. It also was presumed that it was possible that the acquisition lifecycle should not end when the IT is actually acquired. We examined how the methodologies might be used to monitor the ongoing return on the investments in the IT.[3]

---

[3] The future version of this research, to be completed by October 1, 2019, will include a review of prior case studies, conducted by the authors, as well as those reported in various journals and ARP publications and reports. These case studies will provide a rich source of information for the final research report. The case studies included in the review have used a number of the methodologies (e.g., KVA, IRM, EVM) and may have benefited from using the other methodologies (i.e., BSC, L6σ). The review will be augmented by interviews of acquisition SMEs at NPS to test the assumptions of the principal authors concerning their assumptions about how and when these methodologies should be used in the acquisition lifecycle.

A review of the generic IT acquisition lifecycle and the mapping of this generic lifecycle to the existing DoD acquisitions framework is provided in what follows. The section following this review will provide a review of the benefits and challenges of using each of the five methodologies with final recommendations about how to use each within the generic acquisition lifecycle. The final section will include the future research and limitations of this study.

Acquisition Lifecycle

This study developed a basic framework for placing the five methodologies within the generic IT acquisition lifecycle in Table 1, Five Approaches: When to Apply in the Tech Investment Lifecycle. Table 1 can be mapped to the standard DoD Acquisition framework. Doing so allows a comparison of where the two general frameworks match up and provides some preliminary guidance for how the five methodologies might be used in the standard 5000 series acquisition framework.

Table 1. Five Approaches: When to Apply in the Tech Investment Lifecycle

Pre-Investment	Strategic Goal Alignment	Implementation	Post Implementation
KVA (As-Is)	BSC (Align strategy with performance metrics)	EVM (Monitor cost and schedule, adjust as needed)	KVA (Monitor ROI, ROK)
L6(Id waste, value added)	IRM (Identify the strategic options for IT investments)	KVA (To-Be, ROI, ROK)	L6 (Assess and monitor cost, waste reduction)
Other	Other	IRM (Use the project management tools within the IRM suite)	Other

The Defense Acquisition lifecycle framework mirrors (i.e., Table 2, Aligning the Generic and 5000 Series Lifecycles) the generic technology investment acquisition lifecycle in that there exists a planning phase which includes activities consistent with pre-investment and strategic alignment, execution or implementation phase and an operations and sustainment phase, generally considered the post implementation phase of a program.

Table 2. Aligning the Generic and 5000 Series Lifecycles

Pre-Materiel Solutions Analysis	Materiel Solutions Analysis	Technology Maturation and Risk Reduction	Engineering and Manufacturing Development	Production and Development	Operations and Support
-Strategic goal alignment -Pre-investment	Pre-Investment	Pre-investment	Implementation	Implementation	Post-implementation

The DoD defines these phases as the Material Solution Analysis phase, Technology Maturation and Risk Reduction phase, Engineering and Manufacturing Development phase, Production and Deployment, and the Operations and Sustainment phase. Figure 1. The 5000 Series Acquisition Lifecycle is a visual representation of these phases as they are defined in DoDI 5000.02 (OUSD[AT&L], 2017).

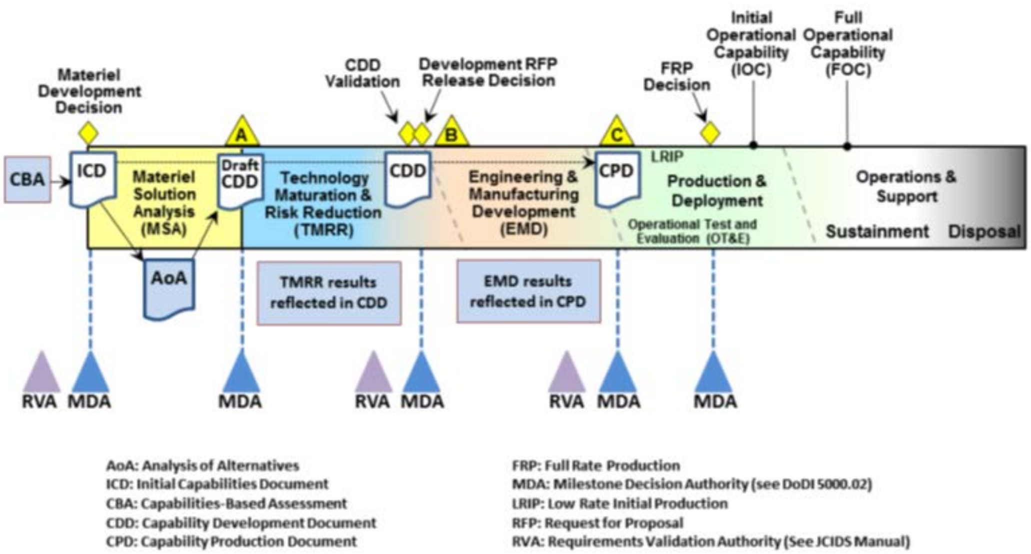


Figure 1. **The 5000 Series Acquisition Lifecycle**
(DoD, 2017)

Materiel Solution Analysis Phase

The Materiel Solution Analysis (MSA) Phase assesses potential solutions for a needed capability in an Initial Capabilities Document (ICD), which was developed during the defense requirements generation process known as the Joint Requirements Capability Determination System (JCIDS). The MSA phase is critical to program success and achieving materiel readiness because it is the first opportunity to influence systems supportability and affordability by balancing technology opportunities with operational and sustainment requirements. During this phase, various alternatives are analyzed to select the materiel solution and develop the Technology Development Strategy (TDS) which will be further assessed in the TMRR phase and eventually executed during EMD.

The MSA phase also includes identifying and evaluating affordable product support alternatives with their associated requirements to meet the operational requirements and associated risks. Consequently, in describing the desired performance to meet mission requirements, sustainment metrics are defined which will impact the overall system design strategy. One of the principle tasks that must be completed during this phase is the Analysis of Alternatives (AoA) suggesting that tools that offer robust tradeoff analysis might be better suited for this phase.

Significant events within the MSA and other phases of the acquisition lifecycle are listed in Table 3, Key Events Within the Phases of the 5000 Series. While this is not an all-inclusive list of events during each phase, important steps within a program's development are incorporated.

Technology Maturation and Risk Reduction Phase

The Technology Maturation & Risk Reduction (TMRR) Phase is designed to reduce technology risk, engineering integration, lifecycle cost risk and to determine the appropriate set of technologies to be integrated into a full system. The objective of the TMRR phase is to reduce technical risk and develop a sufficient understanding of a solution in order to make sound business decisions on initiating a formal acquisition program in the Engineering, Manufacturing and Development (EMD) Phase. This phase lends itself well to management tools that allow the Program Manager to conduct technical and business process tradeoff analysis studies relative to cost and schedule.

Table 3. Key Events Within the Phases of the 5000 Series

MSA	TMRR	EMD	P&D	O&S
Analysis of Alternatives	Preliminary Design Review	Complete detailed design	Low rate initial production	Lifecycle Sustainment Plan (LCSP)
Initial funding estimates	Capability Development Document	System-level Critical Design Review (CDR)	Initial Operational Test & Evaluation (IOT&E)	System Modifications
Technology Development Strategy	Competitive prototyping	Establish project baseline with Performance Measurement Baseline (PMB)	Full rate production decision	Sustainment
	Acquisition Program Baseline (APB) established		Initial and Full Operational Capability (IOC and FOC)	Disposal

Engineering and Manufacturing Development Phase

The Engineering & Manufacturing Development (EMD) Phase is where a system is developed and designed before going into production. The EMD Phase is considered the formal start of any program and the point at which a development contract is awarded based upon a specific statement of work (SOW). The goal of this phase is to complete the development of a system or increment of capability and evaluate the system for technical maturity before proceeding into the Production and Deployment (PD) Phase. This is the phase in which cost and schedule variance models that help the PM to better understand technical issues is best employed since requirements are fundamentally solidified and represented in the SOW. If requirements are shown to be less than optimal or there are other mitigating issues during this phase that impact cost and schedule, then decision support tools to facilitate tradeoffs may be used to help the PM maintain the program baseline and deliver user defined capability.

Production and Deployment Phase and Operations and Sustainment Phase

These phases are necessary for the program manager to ensure that the product being manufactured meets the operational effectiveness and suitability requirements for the user or customer. While the design is pretty well set at this point in the program, there may still be some trades that take place prior to the full rate production decision and fielding of the system. The program manager is less concerned with managing cost and schedule variance at this point since the contract types typically revert to a fixed price strategy. The biggest concern for the PM at this point is correcting any final deficiencies in the system and establishing a stable manufacturing and sustainment process.

The four generic phases listed in Table 1 align with the current DoD structure, as shown in Table 2. As the scope of this research is limited to the 5000 series, the pre-materiel solutions analysis column is for informational purposes only. The JCIDS process accomplishes strategic goal alignment, determining the necessary additions to the DoD's capabilities portfolio prior to the 5000 series. The Initial Capabilities Document (ICD) generated in the JCIDS process describes the high-level needs that the user requires, and these needs are assessed in the AoA process during the MSA phase. Within the scope of this paper, the DoD acquisition lifecycle and generic IT acquisition lifecycle begin with pre-investment during MSA.

If one discounts basic scheduling and cost management practices, the primary tools to monitor progress of an acquisition program during the MSA and EMD phases are EVM and the Risk Management Framework (RMF). Figure 2, Seven Steps to Risk Management Framework, shows the seven steps that comprise RMF, repeating in a cyclical pattern: prepare, categorize, select, implement, assess, authorize, and monitor.

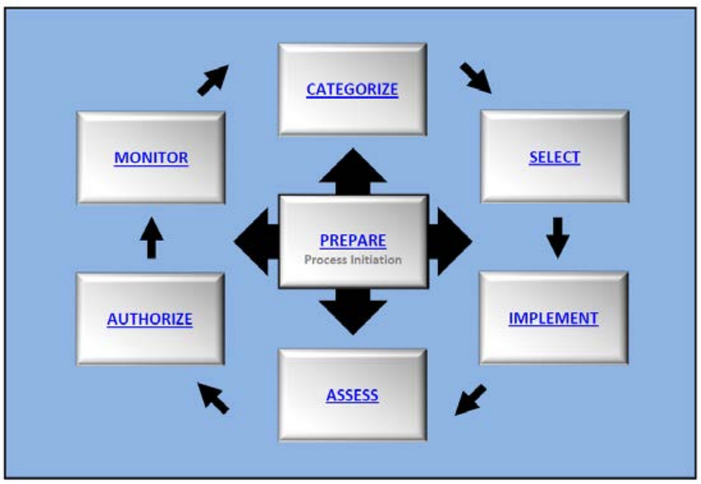


Figure 2. **Seven Steps to Risk Management Framework**
(Joint Task Force Transformation Initiative, 2018)

Preparation initiates the process, ensuring organizations are ready to execute RMF and giving context and priorities for managing risk (Joint Task Force Transformation Initiative, 2018). Categorization consists of organizing the system and the information used by the system based on an impact analysis (Joint Task Force Transformation Initiative, 2018). The risk manager then selects the appropriate security controls, tailoring them as necessary (Joint Task Force Transformation Initiative, 2018). The controls must then be implemented into the system and its operating environment before assessing the controls' effectiveness and authorizing the use of the information system (Joint Task Force Transformation Initiative, 2018). Finally, the manager must monitor the security controls on a continual basis, repeating the cycle as necessary when deficiencies are discovered (Joint Task Force Transformation Initiative, 2018). EMD is the first time program managers use EVM in an official capacity. The appropriate decision makers approved a schedule and budget for the program creating the Acquisition Program Baseline. Future progress is now measured against this benchmark. Even using these proven tools, cost and schedule overruns occur regularly, illustrating the need for a different approach.

The RMF is a broad analysis that covers multiple types of risk and is used throughout the entire lifecycle of a new development system. Implementing other tools into the process could help program managers better understand the risk involved at various decisions and points throughout the program. Within an acquisition there is an interdependence of risk. As the program progresses (and using the EVM methodology) and the ACWP increases, there are increasing levels of aggregation and abstraction of risk. For instance, to award an EMD contract, the technology involved must be at a Technology Readiness Level (TRL) of TRL 6, indicating the technology performed adequately in a relevant test environment (Assistant Secretary of Defense for Research and Engineering [ASD(R&E)], 2011). However, the technology is not yet completed and requires significant improvement before production. The current risk assessment program does not account for the possibility that this categorization is incorrect and may not lead to a fully operational system. As a result, program managers proceed with the assumption that the technology will continue development as planned. Any lack of progress will not become apparent until the ACWP begins to vary from the BCWP. It is often too late to make the appropriate corrections to the program in order to remain on budget by the time the discrepancy is discovered using EVM metrics.

Early risk management that focuses on the validity of the decision-making process using the RMF framework might introduce a higher level of understanding of the subordinate processes. For example, if at a particular milestone, the technology is not at the level of readiness it is being portrayed, then the consequences are x, y, and z. The results of each statement can be expressed in terms of time and money or, keeping with the already established EVM terminology, potential Cost Variance. A program manager can then assign a probability of success estimate to the state of the program which might drive a deeper understanding of the various interdependent program management processes.

Table 4, Methodologies Within the 5000 Series, shows when each methodology might be used in the 5000 series phases. What follows summarizes the alignment of the various program tools that can provide better insight into the lifecycle of a program. This table reflects that there are multiple tools for the various phases that should be used in concert and that certain tools are more appropriate for a particular phase than others. It is incumbent on the PM to use the tools appropriately in that they provide more information for a complex environment. The tools themselves do not provide the solutions to potential problems; they are simply indicators of underlying performance issues.

Table 4. Methodologies Within the 5000 Series

Materiel Solutions Analysis	Technology Maturation and Risk Reduction	Engineering and Manufacturing Development	Production and Development	Operations and Support
BSC	IRM	EVM	EVM	KVA
IRM	KVA	IRM	IRM	L6σ
KVA	L6σ	KVA	KVA	
L6σ				

Understanding the extent to which a particular tool might provide greater insight into program performance across the lifecycle, one should consider the level of analysis required and the viability of a particular tool to provide sufficient insight at that level of analysis. Three levels of analysis were considered for this initial survey: Organizational, Business Process, and Task Analysis.

It is clear from Table 5 that a variety of tools are required across the lifecycle in order for the PM to gain a more robust view of the program performance. The selection of the tool will depend on the particular focus and time horizon with which the tool is able to provide relevant information about the program. Simply relying on one tool will not allow the PM to adequately manage the program. Planning for the type and depth of the management tool is started early in the lifecycle and should be part of the overall acquisition strategy. Additionally, selecting contractors that are able to implement and manage these tools is critical in the decision-making process.

Table 5. Management Tool Selection Criteria Based Upon Level of Analysis, Focus of Analysis, and Acquisition Phase

Level of Analysis	Focus of Analysis	Acquisition Phase
Organization	-Strategic competitive advantages: BSC, IRM -Value=Revenue: BSC, IRM strategic options	MSA/TMRR/P&D/O&S
Business Process	-Cost savings: L6σ, EVM, BSC, IRM -Schedule: EVM Value: KVA outputs	MSA/TMRR/P&D/O&S
Task Analysis	-Cost savings: L6σ, IRM -Value=Cost+schedule cycle time: L6σ, BSC	TMRR/EMD/P&D

The BSC is an excellent tool when viewing a system holistically. It provides a way for managers to examine a project from a systems thinking approach. It may be most useful when strategizing about the potential use of an IT acquisition and how it might fit into the DoD's higher-level strategic goals prior to developing a requirements document. The statements derived from the BSC for general dissemination among all levels of the organizational structure must be translated into a simpler form presented in a set of objectives and targets that are clear for all levels within the organization. It is also important to understand that leadership is central to ensuring any IT acquisition will support the organization's overall strategy enumerated in the BSC. This is true in the DoD as well as in any organization's implementation of a BSC (Llach et al., 2017). Without leadership support and guidance, the BSC is unlikely to succeed, and the organization will not be able to generate acceptable returns on its IT investments.

Table 6. Benefits and Challenges of the Five Methodologies

	Extensible, quantitative value measurement	Time to Perform	Cost	Bottleneck Analysis
BSC	No, subjective measurement (revenue is exception)	3-6 months (depends on level of analysis)	Accounting based financial metrics only	None
EVM	No, cost measurement only	5+ months set up time (depends on requirements)	Cost of resources and time	No, linear tracking only
L6σ	No, nominal value only	3+ months (depends on level of process complexity)	Activity Based Costing approach	Direct bottleneck analysis
KVA	Yes	2 days – 1 month (depends on level of analysis)	Common units of cost	Elapsed time versus work time
IRM	Yes, KVA	3-6 months (Relatively quick once initial steps completed)	Cost accounting and KVA cost metrics	Monte Carlo simulation

The use of the BSC can result in a cursory review of key performance indicators (KPI) during the traditional acquisitions lifecycle management process. The BSC also avoids over relying on financial KPIs by viewing the effects of each of the KPIs on the other parts of the scorecard. While financial KPIs are reviewed with the BSC, the other segments are separated from a purely financial analysis, allowing managers to use their judgement in determining how the proposed solution will affect the scorecard as a whole. The problem is that without a quantifiable common units performance metric that allows the practitioner to determine the relative value between the different scorecards, it is difficult to determine which course of action would be optimal. There is no performance ratio that tells the manager that by performing a given action the financial KPIs will improve by a given amount, the stakeholder engagement will decrease by this amount, and the internal process will change by this amount. Instead, it is more of a conceptual thought exercise to ensure managers consider the effects of their decisions on the entire range of KPIs. Because of this, the BSC works best during the strategic goal alignment phase of the generic IT acquisition lifecycle and the pre-MSA portion of the DoD acquisition lifecycle. The MSA phase also includes aligning the stated requirements with the possible solutions to the capability gap during the AoA. An all-inclusive view of the effects the various IT solutions that are being considered will assist in the selection of the most appropriate option to continue towards acquisition. The BSC is recommended for implementation during the MSA phase.

EVM provides users with an easily understandable report of a project's advancement towards completion. Comparing the BCWP and the ACWP gives a clear view of how a system is progressing within the anticipated budget. The metrics used for cost and time are also clearly delineated. This delineation allows managers to compare the performance at different locations throughout the project, which can assist in determining where a project has changed trajectories. There are numerous challenges when using EVM as well. While cost is measured and tracked regularly, the value of the project is not monitored as closely. Despite the name, the amount of work performed does not tell a manager the actual quantifiable value (in a common units measurement) the project has incurred at a given point. There is no quantifiable measure of value within the methodology. The only quantitative measures of performance are measures of cost and time.

The ACWP assumes the outputs from all work were perfect upon completion. If there are issues with the results from earlier efforts, they must be reworked, changing the ACWP calculation. As in the earlier example, if the technology does not improve as expected because the TRL was not accurately portrayed, a program manager will believe the project is on schedule despite the ‘earned value’ lagging behind what the numbers are projecting. Additionally, and in some instances because of this assumption, EVM outputs are not timely. Conducting an accurate analysis of a program is time consuming and does not provide useful predictive information. By the time EVM alerts a program manager to a variance, the variance has already occurred. All corrections are reactive to bring the ACWP back to the baseline, which has proven to be a nearly impossible task in practice. EVM will only be effective when the baseline plan is well researched and accurate. Otherwise, the ACWP is compared against flawed data. EVM does provide valuable information to project managers during the EMD phase but should be supplemented with some of the other methodologies: L6σ, KVA, IRM throughout the project management cycle. EVM is recommend most for use during the EMD and P&D phases.

Successfully implementing L6σ into a process will lower the cost of the project by reducing the variation in a product run and the waste associated with its production. When additional steps or unnecessary waste is reduced, additional resources are now available for use in other processes. In identifying a bottleneck, L6σ can address multiple problems simultaneously depending on how the project is defined. By creating improvement in one area and freeing resources, other areas may benefit from an improved process work flow. L6σ can be costly to implement. The analysis requires a great deal of time and information to develop meaningful understanding of any problems. L6σ’s definition of value is at the nominal scale level: an item either adds value to a project or it does not. Reality is not often as black or white. There are required steps that must be conducted that do not necessarily add value to a product from the user’s standpoint. For instance, accounting departments do not attempt to directly add value to a final product, but any organization recognizes the need for accounting, suggesting the accounting department does add value. L6σ is time consuming when applied on a large scale, as would be the case in a DoD acquisition. Defining the problem and determining appropriate measurements in a step by step manner is a major undertaking. Contractors can gain much of the benefit from L6σ when performing their internal processes during the TMRR and EMD phases. However, acquisition professionals should use L6σ during the MSA and TMRR phases to ensure the project is defined and measured appropriately. L6σ should also be employed during the O&S phase to ensure the LCSP is carried out as efficiently as possible.

The greatest benefit from KVA is a quantifiable (common units) value metric which can be compared across various aspects of a project (Housel & Bell, 2001). If the value of an intermediate step is quantified, managers can compare the outputs of a component instead of simply the effort measured by time and cost that were inputs. KVA provides a value measurement for both tangible and intangible assets, making it especially well-suited for use with IT. A KVA analysis can be accomplished in a relatively short period of time in comparison with the other methodologies. A quick, rough-cut KVA analysis can provide rapid guidance for the project before sinking valuable time and resources into a more comprehensive examination. KVA is primarily a measurement tool that provides performance information to decision makers. It is not a system that will drive an acquisition project towards the goal on its own. As in the other methodologies mentioned thus far, KVA has limited value in making predictions for future value, focusing instead on the current value of systems in development. There must be another methodology employed with KVA to ensure a project’s success. Due to its ability to provide a quantitative, comparable value metric, KVA is recommended for use during all phases of the 5000 series.

IRM provides a foundation to incorporate the risk associated with a decision into a quantitative decision process. IRM's core premise maintains there is a probability for success and failure with every decision option during a project's lifecycle. Using statistical simulations, real options, and optimization will improve the quality of information a program manager has to determine the course of a project. Real options analysis can be used to frame strategies to mitigate risk, to value and find the optimal strategic pathway to pursue, and to generate options to enhance the value of the project while managing risks. IRM's drawback is that the analytical methods can sometimes be difficult to master. But with the requisite knowledge and training, coupled with the correct tools, the IRM methodology can provide a plethora of value-added information for making strategic and tactical decisions under uncertainty. The IRM methodology should be employed during the MSA, TMRR, EMD, and P&D phases.

Limitations and Future Research[4]

This research only examined the 5000 series acquisition lifecycle. It is probable that both the JCIDS and PPBE processes could benefit from the calculated implementation of some or all of the methodologies discussed. Improving one component of the Defense Acquisition Decision Support System will likely improve the outputs of the other two systems. Additional research into creating a quantifiable measure of risk will provide beneficial information that allows decision makers to understand the probability of success for subcomponents within a project.

Future research in how the five methodologies might be useful for other areas of investment in IT and DoD acquisitions of IT might prove useful in extending the current research study. The proposed five methodologies may be useful for researchers who are also interested in focusing on the following topics of acquisition research interest:

- Innovative Contracting Strategies—contracting at the speed of relevance (BSC, IRM)
- Breaking down silos, enterprise management (L6σ, KVA)
- Rapid Acquisition and Decision Support (IRM, KVA)
- Effects of Risk-Tolerant and Risk-Averse Behavior on Cost, Schedule, and Performance (IRM, EVM)
- The Role of Innovation in Improving Defense Acquisition Outcomes (BSC, IRM, EVM)
- Applying Model-Based Systems Engineering to Defense Acquisition (IRM, KVA)
- Augmenting the Acquisition Decision Processes with Data Analytics (IRM)

[4] Given that the case studies of IT technology acquisitions exist in various existing data sources and written case studies, there is very little risk associated compared to the normal generation of new data sets that were required in the prior studies performed by the authors for the ARP. Access to acquisition subject matter experts (SMEs) at NPS reduced the risk associated in seeking other SMEs to discuss IT acquisitions and the use of the methodologies within the IT acquisition lifecycle.

References

Advanced Performance Institute. (n.d.). Balanced Scorecard Part 1: A brief history. Retrieved

April 11, 2018, from https://bernardmarr.com/default.asp?contentID=1479

Albert, N. (2002). *Recommendations on the implementation of balanced scorecard metrics for the Department of Defense*. Retrieved from http://dbb.defense.gov/Portals/35/Documents/Reports/2002/FY02-2_Management_Information_2002-12.pdf

Assistant Secretary of Defense for Research and Engineering (ASD[R&E]). (2011, April). *Technology readiness assessment (TRA) guidance*. Washington, DC.

Balanced Scorecard Institute. (n.d.). What is the balanced scorecard? Retrieved April 11, 2018, from http://www.balancedscorecard.org/BSC-Basics/About-the-Balanced-Scorecard

De Lorenzo, R. (2005). The balanced scorecard in military medicine: What the clinical leader needs to know. Retrieved from http://www.biomedsearch.com/article/balanced-scorecard-in-military-medicine/147669360.html

DoD. (2017, February 26). *Defense acquisition guidebook*. Retrieved from https://www.dau.mil/tools/dag

GoLeanSixSigma. (2018). What is Lean Six Sigma? Retrieved from https://goleansixsigma.com/what-is-lean-six-sigma/

GAO. (1997). *Assessing risks and returns: A guide for evaluating federal agencies' IT investment decision-making* (GAO/AIMD-10.1.13). Washington, DC: Author.

Harvey, R. K. (2006, September 11). *Letterkenny Army Depot, the Army teaches business a lesson in Lean Six Sigma*. Retrieved from https://cle.nps.edu/access/content/group/f8731c51-dad6-4d56-ba30-b8cb7de32862/Course_Documents/Topic__1_-_Approaches_to_Measurement/Articles_and_Readings/Lean_Six_Sigma/L6S_-_Army_Case.pdf

Housel, T., & Bell, A. (2001). *Measuring and managing knowledge*. Boston, MA: McGraw Hill.

Housel, T., & Kanevsky, V. (1995). Reengineering business processes: A complexity theory approach to value added. *INFOR*, *33*, 248–262.

Joint Task Force Transformation Initiative. (2018). *Risk management framework for information systems and organizations: A system life cycle approach for security and privacy* (No. NIST SP 800-37r2). Gaithersburg, MD: National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-37r2

Kaplan, R. S., & Norton, D. P. (1996). Using the balanced scorecard as a strategic management system. *Harvard Business Review*.

Llach, J., Bagur, L., Perramon, J., & Marimon, F. (2017). Creating value through the balanced scorecard: How does it work? *Management Decision*, *55*(10), 2181–2199.

MCEN User Portal. (n.d.). United States Marine Corps. Retrieved April 18, 2018, from https://homeport.usmc.mil/eitsm/default.aspx

MoreSteam. (2018). New to Lean Six Sigma? Retrieved from https://www.moresteam.com/new-to-lean-six-sigma.cfm

Niven, P. R. (2008). *Balanced scorecard step-by-step for government and nonprofit agencies.*

Hoboken, NJ: John Wiley & Sons.

Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]). (2017, August 10). *Operation of the defense acquisition system* (DoD Instruction 5000.02). Washington, DC: Author.

Pavlou, P., Housel, T., Rodgers, W., & Jansen, E. (2005). Measuring the return on information technology: A knowledge-based approach for revenue allocation at the process and firm level. *Journal of Association of Information Systems*, *6*(7), 199–226.

Reichel, C. W. (2006). *Earned value management systems (EVMS): "You too can do earned value management."* Paper presented at the PMI® Global Congress 2006—North America. Newtown Square, PA: Project Management Institute. Retrieved from https://www.pmi.org/learning/library/earned-value-management-systems-analysis-8026

Rodgers, W., & Housel, T. J. (2006). Improvement of global performance measures related to intangible assets. In *Collected Papers of the 32nd European International Business Academy (EIBA) Conference* (Refereed proceeding).

United States Marine Corps. (n.d.). *Enterprise information technology service management identity and access management process guide* (IRM 2300-14). Washington, DC. Retrieved from https://homeport.usmc.mil/eitsm/eidam/Documents/IRM%202300-14%20EITSM%20Identity%20and%20Access%20Management%20Process%20Guide.pdf

West, C. (2007). Using earned value management for improving processes. Retrieved from https://www.stickyminds.com/article/using-earned-value-management-improving-processes

Zizlavsky, O. (2014). *The balanced scorecard: Innovation performance measurement and management control system*. Retrieved April 24, 2018, from https://scielo.conicyt.cl/pdf/ jotmi/v9n3/art16.pdf

System Maturity Estimation During Program Execution

John C. Kamp—is a retired naval officer and served as a program manager. He has over 30 years of direct experience in defense operations, research, engineering, program management, resource sponsorship, requirements generation, and acquisition spanning multiple domains and services. He received his Doctor of Engineering degree in Engineering Management from The George Washington University in 2019. Dr. Kamp is a Fellow in the Royal Institution of Naval Architects and has been a member of the ACM and IEEE since 2005. [jckamp2@gmail.com].

Ebrahim Malalla—is a Visiting Associate Professor of Engineering and Applied Science at The George Washington University. He worked previously as an Assistant Professor at Ahlia University, Bahrain, and at King Fahd University for Petroleum and Minerals, Saudi Arabia. While working at Ahlia University, he held the positions of the President Assistant for University Compliance (2015–2017), the Associate Dean of Graduate Studies and Research (2012–2014), the Chairperson of the IT Department (2008–2011), and the Chairperson of the M.Sc. Program in Engineering Management offered by The George Washington University at Ahlia University (2012–2017). Dr. Malalla received his PhD in Probabilistic Analysis of Algorithms from McGill University, Montreal, Canada, in 2004. His research interests include multidisciplinary problems in computer science, mathematics and engineering management and have led to many papers published in refereed international journals. [emalalla@gwu.edu].

Abstract

Defense acquisition programs integrate mature and immature new technologies into developing and in-service systems to offset future threats and needs. Mature technologies may be nearly ready-for-use; less mature technologies may mitigate anticipated threats or create new capabilities but may also take more time to develop and integrate into a system leading to schedule growth. The Department of Defense uses Technology Readiness Assessments to assess system technology maturity and to satisfy statutory requirements to evaluate system technical readiness prior to starting system development. The Government Accountability Office independently conducts annual assessments of selected weapon system programs. These are useful but require program offices to expend significant time and effort as part of program execution. This research examines different measures of technology and system maturity and identifies maturity-related factors. Regression analysis is used to identify statistically significant predictors of program technology and system maturity and schedule growth. The research results provide program offices insight into technology and system maturity and the sources of schedule growth based on resource, programmatic, operational testing, and schedule-related factors, allowing them to monitor and adjust acquisition program planning and execution.

Introduction

A recent unclassified summary of the National Defense Strategy described a changing acquisition strategy emphasizing speed of delivery as part of a response to capable, innovative adversaries (Mattis, 2018). One way that Department of Defense (DoD) Major Defense Acquisition Programs[1] (MDAPs) respond to an adversarial threat is by

[1] MDAPs are weapon system programs with research and development expenditures greater than $300 million or procurement expenditures greater than $1.80 billion indexed to fiscal year 1990 constant dollars (MDAP Defined, 2007).

integrating proven (*mature*) and emerging (*immature*) technologies into new and in-service systems. Katz et al. (2015) noted that DoD programs may select less mature technologies to hedge system performance against future threats or to create new capabilities but use more mature technologies to reduce the likelihood of schedule growth. The problem is that MDAP schedules can grow by over 25% when integrating immature systems. In 2013, the Government Accountability Office (GAO) reported an average schedule delay of 27 months, or 37%, for MDAPs to deliver an initial capability (Dodaro, 2013).

Maturity-Related Terms and Measures

System maturity is different than technology maturity. *System maturity* means the system satisfies the *design requirements*. The literature describes system maturity in terms of requirements validation (Tetlay & John, 2009) and includes validating functional requirements (Gove & Uzdzinski, 2013). *Technology maturity* describes how well a technology is understood. During system development, the DoD focuses on technology maturity instead of system maturity (Ramirez-Marquez & Sauser, 2009).

The GAO assesses MDAP *technology maturity, design maturity,* and *production maturity* as part of its annual independent assessments of selected DoD weapon system programs (Dodaro, 2007). Katz et al. (2015) showed that GAO technology maturity occurs for a system when the TRLs of all critical technologies were greater than or equal to 7. For the MDAPs considered in this research, most (about 54%) achieved GAO technology maturity, fewer (about 41%) achieved design maturity, and few (about 7%) achieved production maturity.

Product maturity reflects a product's market position. Day (1981) described product maturity in terms of customer understanding (learning), market share (potential) and competition (turbulence). Mature products respond more to customer and competitive demands (orientations) than to innovation (Wang, Wang, & Zhao, 2015). Nolte (2008) states that technology maturity is related to how *well* something is understood, while product maturity includes concepts of product obsolescence and competitive market share.

Product quality is a measure of how well a product meets customer requirements (Kandt et al., 2016). Azizian et al. (2011) identified the relationship between product (system) quality as measured by existing international standards and technology maturity, and found that system *development* and *operational tests*, system *prototyping,* and actual system *demonstrations* were statistically significant predictors of product quality.

Readiness describes context-specific system suitability for use (Tetlay & John, 2009), which is similar to maturity. Technology Readiness Assessments (TRAs) are used by the DoD to assess system technology maturity and to satisfy statutory requirements to evaluate system technical readiness prior to starting system development (Weapon System Acquisition Reform Act [WSARA], 2009). Bailey et al. (2014) noted that the TRA process is qualitative and subjective, and found the underlying system engineering activities, *not the TRA itself*, were statistically significant predictors of quality and program performance.

Technology Readiness Levels (TRLs) are used by the DoD to indicate where the maturity of either a technology or system lies within a qualitative nine-level ordinal scale (Mankins, 2009). Note that TRLs do not by themselves characterize risk or difficulty of progressing between levels (Conrow, 2011), nor do they describe integration readiness or risk (Ramirez-Marquez & Sauser, 2009). TRLs are characterized by completion of discrete events and activities, but are typically not reported in the public literature. There are tools such as TRL calculators (Nolte, 2008) to help consistently assign TRLs.

This paper summarizes recent research that explored how MDAP system maturity and performance are reflected in data routinely collected by program offices. Regression analysis was used to create *system maturity* models for selected MDAPs between Milestone B (approval to start Engineering and Manufacturing Development) and declaration of Initial Operational Capability (IOC). These system maturity models were used to test the hypothesis that *system maturity is correlated to program schedule growth*.

Data Sources and Dataset Creation

The original research dataset[2] (Kamp, 2019) was created from publicly-released annual reports to Congress issued by the GAO, the Director, Operational Test and Evaluation (DOT&E), and the DoD for MDAPs between 2007 through 2017. Program data (observations) were included in the dataset if a program was assessed in both the GAO and DOT&E annual reports in a given report year.[3] The dataset was filtered to eliminate cancelled programs and programs without published schedule estimates or with missing data elements, and recorded in comma-separated variable files. This resulted in 154 observations of 48 programs from the 2007 through 2017 reports. Three observations[4] were outside the research program window from Milestone B IOC and were eliminated, leaving 151 observations in the database. No programs had observations in all years, but three programs had more than six observations.[5] Tests for observation independence were performed on these observations and on the complete dataset, and no additional observations were deleted.

Response and Predictor Variables

There are two explicit technology maturity response variables in the dataset, GAO technology maturity and estimated TRL.[6] The GAO's technology maturity assessment is an independent check of technology maturity. By definition, a *technology may be mature* when demonstrated within a system in a relevant (TRL 6) or operational (TRL 7) environment, but a *system is mature* when tested in its production version under operational conditions (TRL 8) or when used in an actual mission (TRL 9; Assistant Secretary of Defense for Research and Engineering [ASD(R&E)], 2011). The predictor variables used in the research are clustered into four groups: resource-, programmatic-, operational testing- and schedule-related predictors. The following tables will summarize the predictors by group and summarize their significance to technology maturity response variables.

1. <u>*Resource-Related Predictors*</u>: Resources are planned or budgeted quantities. All resource-related variables in the dataset were continuous and were derived from GAO annual reports or Selected Acquisition Report (SAR) Summary reports. These are summarized in Table 1.

[2] Available upon request.
[3] The GAO and DOT&E do not issue publicly available reports on all MDAPs each year, resulting in relatively few programs reported by both agencies. This criterion eliminated about 90% of MDAPs in any given year, but ensured concurrent programmatic and operational testing information.
[4] The three program observations outside the research window were C-130J in 2008, JLTV in 2011, and MQ-9 in 2014.
[5] The Joint Strike Fighter (F-35), CVN-78 and WIN-T programs all had more than six observations.
[6] An estimated TRL was created as TRLs are not typically reported in public documents.

Table 1. Resource-Related Predictors

Descriptive Name	Symbol	Explanation/ Description	Source	Maturity significance
Cost change assigned to Engineering - current year	Eng	PM reported cost changes allocated to Engineering in the GAO year, $ millions	SAR Summary	
Cost change assigned to Engineering - one year prior	Eng.1	PM reported cost changes allocated to Engineering ONE YEAR PRIOR to the GAO year, $ millions	SAR Summary	
Cost change assigned to Engineering - two years prior	Eng.2	PM reported cost changes allocated to Engineering TWO YEARS PRIOR to the GAO year, $ millions	SAR Summary	Sys maturity Binary, p=0.071
procurement funding	P.M	GAO or Program Office reported procurement funding, $ Millions (GAO value) (natural log of P.M is LN.P.M)	GAO	
research and development funding	RD.M	GAO or PM reported total research and development funding, $ Millions (GAO value) (natural log of RD.M is LN.RD.M)	GAO	for LN.RD.M, Binary, p=0.000
procurement quantities	P_no	Planned procurement quantities	GAO	
External Program cycle, months	Cycle.Mo	GAO or Program Office reported Program Office cycle time estimate, months	GAO	TRL Ordinal, p=0.000

2. *Schedule Predictors*: Schedule-related predictors may be interpreted as mimicking the program office view of progressing between events. These are continuous valued predictors calculated as differences between key milestone dates (program start, Milestone B, OT events, Milestone C, and IOC) within the GAO reporting. The schedule-related predictors are summarized in Table 2.

Table 2. Schedule-Related Predictors

Descriptive Name	Symbol	Explanation/ Description	Source	Maturity significance
Time from program start to Milestone B	B.st	time between MILESTONE B and Program start date, years	Calculated	TRL Ordinal, p=0.000
Time from Milestone B to Milestone C	C.B	time between MILESTONE B and MILESTONE C, years	Calculated	TRL Ordinal, p=0.010
Time from Milestone C to EVENT	C.ev	time between MILESTONE C and EVENT, years	Calculated	System maturity, Binary, p=0.000
Time from program start to EVENT	ev.st	time between EVENT and program start date, years	Calculated	TRL Ordinal, p=0.000
Relative Schedule Change	RSC	Relative Schedule Change (RSC) - *H2 DEPENDENT VARIABLE*	Calculated	TRL Ordinal, p=0.001

3. *Programmatic-Related Predictors:* Programmatic predictors reflect both programmatic strategies and external factors. There were 11 categorical (TRUE or FALSE) predictors derived from GAO annual reports and cross checked against DOT&E reports and any available SAR reports. These predictors are summarized in Table 3.

Table 3. Programmatic-Related Categorical Predictors

Descriptive Name	Symbol	Explanation/ Description	Source	Maturity significance
Commercial basis	COML	Program procures Commercially available system design/product (i.e. a helicopter or ship)	GAO	Both models Binary, p=0.031, Ordinal p=0.038
Complex system	complex_sys	System is complex	GAO	
External program dependencies	DEPEND	SYSTEM function depends on other programs not controlled by Program Office	GAO	Sys maturity Binary, p=0.079
Unstable funding	Fin_Uns	Indications that program is financially unstable (i.e. funding change > 10% in a year)	GAO	
Joint program	Joint	Joint Program indicator	GAO	Sys maturity Binary, p=0.008
Nunn-McCurdy Breach	NM	Nunn-McCurdy Breach occurred	GAO	
External program issues	PM.oth	other Program management issues (political direction or sponsorship)	GAO	
System prototypes	Prototype	Program uses prototypes representative of objective system, capable of operating in realistic environments	GAO	Sys maturity Binary, p=0.006
Unstable requirements	Req_Uns	Program requirements are unstable (i.e. >10% change in procurement; identified requirement changes)	GAO	Sys maturity Binary, p=0.006
Restructured program	Restr	Program is restructured	GAO	
Part of a System of Systems	SoS_Part	System is required to operate as part of a system-of-systems	GAO	Sys maturity Binary, p=0.079

The Complex system predictor is identified by the program office controlling subsystem and system selection and integration, while the Part of a System of Systems predictor identifies a system requiring other systems to accomplish its design mission (such as an aircraft carrier needing aircraft; Stuckey, Sarkani, & Mazzuchi, 2017). Unstable requirements are primarily indicated by year-to-year procurement quantity changes of more than 10%. A nominal variable identifies the system type (symbol "Type")—such as an aircraft, ship, missile, or ground vehicle system. These have been used by other researchers (Tate, 2016). Some useful predictor variables such as system mission, prototyping, program funding, technology maturity (Monaco & White III, 2005), and Drezner and Smith's (1990) programmatic structural and external factors may be found in the GAO Annual Assessments. Other predictors found in the literature, such as Low Rate Initial Production quantities and contract type (Monaco & White III, 2005), are not within the research data sources.

4. *Operational Testing (OT) Predictors:* The OT categorical (binary) predictors represent system issues found during OT events and are described in Table 4.

Table 4. Operational Testing-Related Predictors

Descriptive Name	Symbol	Explanation/ Description	Source	Maturity significance
System command and control issues	C3I	Testing issues with command, control, communications, intelligence (i.e. communications range and data rate) (0=FALSE, 1=TRUE)	DOTE	System maturity, Binary, p=0.002
System control and controllability issues	CONTROL	Testing issues with system controllability (i.e. precision, maneuverability) (0=FALSE, 1=TRUE)	DOTE	
System integration issues	INTEG	Testing system integration issues (i.e. fit, quality, non-compliance with requirements) (0=FALSE, 1=TRUE)	DOTE	Both models Binary, p=0.011, Ordinal p=0.005
Interoperability issues with other systems	INTEROP	Testing issues with system interoperability (i.e. can't exchange data with other systems, crypto) (0=FALSE, 1=TRUE)	DOTE	
System reliablity, maintainability, availability issues	RMA	Testing issues with system reliability, maintainability, availability (i.e. mean time between failures) (0=FALSE, 1=TRUE)	DOTE	
System operator usability issues	OPER	Testing issues with operator usability (i.e. safety, tactics, doctrine, procedures, training, cybersecurity) (0=FALSE, 1=TRUE)	DOTE	
System propulsion issues	PROP.PW.EN	Testing issues with propulsion power or energy (i.e. underpowered, prime mover issues) (0=FALSE, 1=TRUE)	DOTE	
System payloads issues	SEN.W	Testing issues with system payloads (i.e. sensors or weapons) (0=FALSE, 1=TRUE)	DOTE	TRL Ordinal, p=0.028
System structural issues	STRUCT	Testing physical structural issues (i.e. cracking, vibration, loading) (0=FALSE, 1=TRUE)	DOTE	
System size, weight, or power issues	SWAP	Testing issues with size, weight, or power (i.e. overweight) (0=FALSE, 1=TRUE)	DOTE	System maturity, Binary, p=0.046
System software performance issues	SW	Testing issues with system software (i.e. logic errors, production, vulnerabilities) (0=FALSE, 1=TRUE)	DOTE	System maturity, Binary, p=0.000

Azizian et al. (2011) identified critical technologies, analyses of alternatives, operational tests, certification and accreditation, and system engineering plans (all processes supporting technology readiness assessments) as important predictors affecting technology maturity and program performance. The OT predictors were iteratively derived using word frequency counting software DOT&E annual reports, and by reading the reports to derive context and relevance to system effectiveness and suitability.

5. *Testing Events and Predictors*: In general, common developmental and operational testing events or milestones were used. Flight Test was generalized to include first operational test (First Flight or first underway sea test) to represent operation in a realistic environment. Additionally, this research mapped DOT&E reported test event completion to an estimated TRL between 5 and 9 inclusive.[7] The TRL mappings to events were based upon DoD TRL definitions (Mankins, 2009) and were reviewed by independent experts. These events and their TRL mappings are summarized in Table 5.

[7] In lieu of using Nolte's (2008) TRL calculator to estimate TRLs

Table 5. Testing Events and TRL Mapping

Descriptive Name	Symbol	Explanation/ Description	Source	Maturity significance
Crtical Design Review	CDR	Critical Design Review (TRL=6 after completion)	GAO/DOTE	Event dependent
Design Review	DR	Design review (unspecified) (TRL at least 5, value dependent on description)	GAO/DOTE	Event dependent
Development testing	DT	Development testing (unspecified) (TRL test dependent)	GAO/DOTE	Event dependent
Early Fielding	EFR	Early Fielding Report – following directed deployment (TRL =9)	GAO/DOTE	Mature system
Force Deployment Evaluation	FDE	Force Deployment Evaluation	GAO/DOTE	Event dependent
Follow-on Test	FOTE	Follow On Test and Evaluation – testing following IOC (TRL=9)	GAO/DOTE	Mature system
Flight Test	FT	Flight Test (or first Sea Test) (unspecified) (TRL at least 6, value dependent on test description)	GAO/DOTE	Event dependent
Initial Operational Test	IOTE	Initial Operational Test and Evaluation – uses a production representative system (TRL>7, value dependent on test description)	GAO/DOTE	Event dependent
Land Based Test	LBT	Land Based Test (of any type) (TRL at least 5, value dependent on test description)	GAO/DOTE	Event dependent
Live Fire Test	LFTE	Live Fire Test and Evaluation – survivability testing of components or system (TRL >6, value dependent on test description)	GAO/DOTE	Event dependent
Limited User Test	LUT	Limited User Test – a subset of OT – for example a subset of effectiveness testing (TRL> 6, value dependent on test description)	GAO/DOTE	Event dependent
Milestone B	MSB	Milestone B (not a test event) (statutory TRL=6 after 2008)	GAO/DOTE	Event dependent
Milestone C	MSC	Milestone C (not a test event) (TRL=8)	GAO/DOTE	Mature system
Operational Assessment	OA	Operational Assessment – a subset of operational test (specific objective) (TRL at least 6, value dependent on test description)	GAO/DOTE	Event dependent
Operational Test	OT	Operational Test (unspecified) (TRL at least 7, value dependent on test description)	GAO/DOTE	Event dependent
Quick Reaction Assessment	QRA	Quick Reaction Assessment – for a specific end use (TRL =8)	GAO/DOTE	Mature system

Methodology

We performed binary and ordinal logistic regression analyses on the dataset using Minitab 18 and SPSS. The response variables are GAO technology maturity for the binary logistic regression and estimated TRL for the ordinal logistic regression. The regressions were reduced using backwards elimination or manually (one variable at a time) until only significant predictors remained. A random 10% subset of the dataset was withheld for model validation. Residuals were inspected to assess if regression assumptions were satisfied; then model goodness of fit and accuracy were evaluated.

Results

GAO Technology Maturity Regression Model

Table 6 and Figure 1 summarize the GAO technology maturity binary logistic regression model. Significant terms are identified using the predictor symbols from Tables 1 through 5. The model is significant at the α = 0.05 level.

Table 6. Results of GAO Technology Maturity Binary Logistic Regression

Term	Coef	Contribution	P-Value	VIF	Odds Ratio
Regression		39.85%			
Constant	7.500		0.000		
LN.RD.M	-0.944	7.89%	0.000	1.60	0.389
C.ev	-0.381	8.12%	0.000	1.87	0.683
[Req_Uns=1]	-1.461	1.46%	0.009	1.57	0.232
[COML=1]	-1.303	1.29%	0.024	1.50	0.272
[Prototype=1]	1.404	2.46%	0.015	1.67	4.070
[SW=1]	2.536	7.11%	0.000	1.56	12.632
[C3I=1]	1.404	0.99%	0.010	1.29	4.071
[INTEG=1]	-1.217	3.15%	0.013	1.21	0.296
[DEPEND=1]	-1.271	3.71%	0.014	1.26	0.281
[Joint=1]	-1.814	3.67%	0.009	1.38	0.163

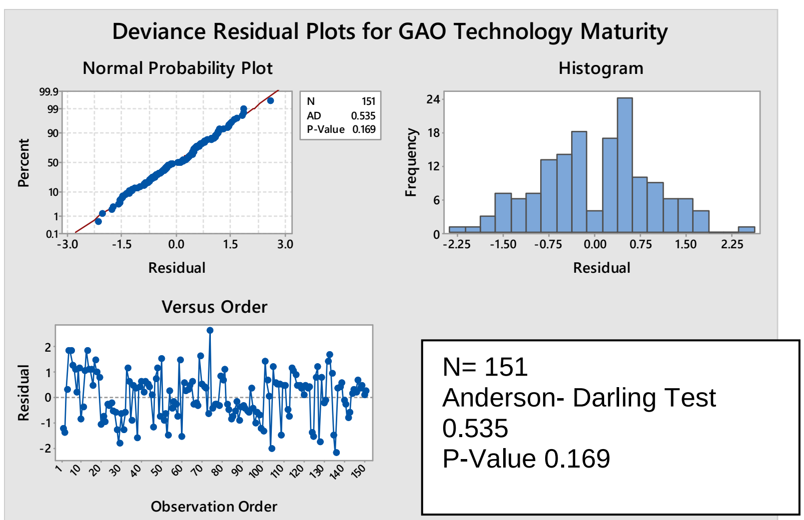


Figure 1. **Technology Maturity Regression Deviance Residual Plots**

The model satisfies all binary logistic regression assumptions. Model accuracy was assessed by withholding 14 random observations and re-performing the regression on the reduced dataset. This *accuracy test* model was used to predict the GAO technology maturity of these 14 withheld observations, and predictions and observations were compared. The regression predictors changed slightly between the two models. The accuracy test model correctly predicted GAO Technology Maturity 11 of 14 times (78.6%). Table 7 summarizes the goodness-of-fit differences between these two models.

Table 7. Binary Logistic Regression Goodness-of-Fit and Association Measures

Goodness-of-Fit Tests						Model Summary			Measures of Association
						Deviance	Deviance		
Hosmer-Lemeshow	Observations	Model α	DF	Chi-Square	P-Value	R-Sq	R-Sq(adj)	AIC	Kendall's Tau-a
Table 6 model	151	0.05	8	6.85	0.552	39.85%	35.06%	147.42	0.39
Accuracy test model	137	0.05	8	5.4	0.714	40.25%	34.98%	135.26	0.40

Association is between the response variable and predicted probabilities

Summary of GAO Technology Maturity Regression Results

The model correctly predicted GAO technology maturity over 75% of the time. Most of the 10 predictors contributed less than 4% of the variance. The following were the top three predictors in contribution order: time between Milestone C and defined events, the natural log of research and development funds, and software issues during Events. The model was most affected by programmatic (five predictors), then resource, operational testing (three predictors), and finally schedule predictors. Issues found during Testing Events indicate system immaturity, or a system without issues during an event is considered mature for that predictor. In particular, software issues discovered during operational testing had the largest odds ratio, as such issues were commonly discovered later in system development and testing relative to other issues.

TRL Ordinal Logistic Regression Model

An ordinal logistic regression of estimated technology readiness levels (TRLe) was performed in Minitab. Predictors were removed until only those with p-values less than or equal to 0.05 remained, and the regression is significant at α = 0.05. The model was re-run in SPSS using the Minitab predictors to test the proportional odds assumption using the test of parallel lines. The Minitab and SPSS logistic regression results are summarized in Table 8.

Table 8. Ordinal Logistic Regression of Technology Maturity (TRLs)

TRL Ordinal Logistic Regression Table

Minitab results					Odds	95% CI		SPSS results						95% CI	
Predictor	Coef	SE Coef	Z	P	Ratio	Lower	Upper		Estimate	Std. Error	Wald	df	Sig.	Lower	Upper
[TRL=5]	-6.609	0.955	-6.92	0.000				[TRLe = 5]	-7.219	1.002	51.883	1	0.000	-9.184	-5.255
[TRL=6]	-2.499	0.675	-3.70	0.000				[TRLe = 6]	-3.109	0.728	18.246	1	0.000	-4.536	-1.683
[TRL=7]	-1.341	0.656	-2.04	0.041				[TRLe = 7]	-1.951	0.701	7.745	1	0.005	-3.325	-0.577
[TRL=8]	1.475	0.662	2.23	0.026				[TRLe = 8]	0.864	0.692	1.559	1	0.212	-0.492	2.220
Cycle.Mo	0.030	0.007	4.47	0.000	1.03	1.02	1.04	Cycle.Mo	-0.030	0.007	19.950	1	0.000	-0.043	-0.017
ev.st	-0.561	0.075	-7.50	0.000	0.57	0.49	0.66	ev.st	0.561	0.075	56.224	1	0.000	0.414	0.708
B.st	0.523	0.131	4.00	0.000	1.69	1.31	2.18	B.st	-0.523	0.131	15.999	1	0.000	-0.779	-0.267
C.B	0.180	0.070	2.59	0.010	1.20	1.04	1.37	C.B	-0.180	0.070	6.683	1	0.010	-0.316	-0.044
RSC	2.194	0.673	3.26	0.001	8.97	2.40	33.55	RSC	-2.194	0.673	10.631	1	0.001	-3.513	-0.875
[COML=1]	-0.819	0.396	-2.07	0.038	0.44	0.20	0.96	[COML=0]	-0.819	0.396	4.286	1	0.038	-1.594	-0.044
[SEN.W=1]	-0.809	0.368	-2.20	0.028	0.45	0.22	0.92	[SEN.W=0]	-0.809	0.368	4.846	1	0.028	-1.530	-0.089
[INTEG=1]	1.018	0.363	2.80	0.005	2.77	1.36	5.63	[INTEG=0]	1.018	0.363	7.863	1	0.005	0.306	1.729

The model differences are due to software implementation differences. All ordinal logistic regression model assumptions were satisfied. The pseudo-R2 is 0.577. SPSS provides a simpler prediction performance summary, and full dataset results are shown in Table 9.

Table 9. SPSS Ordinal Logistic Regression TRL Prediction Results

SPSS model performance		Actual TRL estimate 5 Count	6 Count	7 Count	8 Count	9 Count
Predicted Response Category	5	0	0	0	0	0
	6	4	31	6	13	0
	7	0	0	0	0	0
	8	0	10	18	38	11
	9	0	0	0	4	16

		Actual TRL estimate [5,6,7]	[8,9]		
Predicted TRL estimate	[8,9]	28	69	exact	56.3%
	[5,6,7]	41	13	+/- 1	84.8%
		[5,6,7]	[8,9]	correct	72.8%
	[8,9]	18.5%	45.7%		
	[5,6,7]	27.2%	8.6%		

The predicted responses represent the highest probability for each observation; the actual TRL is the estimated TRL in the data set. The model predicts the exact TRL (matches the estimated TRL) 56.3% of the time and is within + 1 TRL level nearly 85% of the time. The response gap seen in Table 9 at a predicted TRL of 7 mimics the TRL and system maturity relationship (system mature if TRL is 8 or 9, system immature if TRL is 5, 6, or 7).

Summary of TRL Ordinal Regression Results

The TRL ordinal logistic regression model is dominated by schedule-related predictors. It predicted an *exact* TRL slightly over half the time; this is not useful if a program needs high confidence in the prediction precision. The SPSS model correctly predicted *system maturity* over 70% of the time and shows a significant relationship between TRL (maturity) and relative schedule change. *Program offices may use the inability to hold to schedule as an indicator of technology immaturity, and during Engineering and Manufacturing Development (EMD), as an indicator of system immaturity.*

System Maturity and Schedule Results

Finally, the dataset was partitioned into immature (estimated TRL = 5, 6, 7) and mature (estimated TRL = 8, 9) subsets. A linear regression of Relative Schedule Change (RSC) using the research predictors was performed on each partition. The immature model (estimated TRL = 5, 6, 7) did not satisfy residual normal distribution assumptions at α = 0.05, so comparisons are made when both models are significant at α = 0.1. These models are presented in Table 10 and Figure 2.

Table 10. RSC Regression Summaries by System Maturity

Systems with TRLe = [5,6,7]

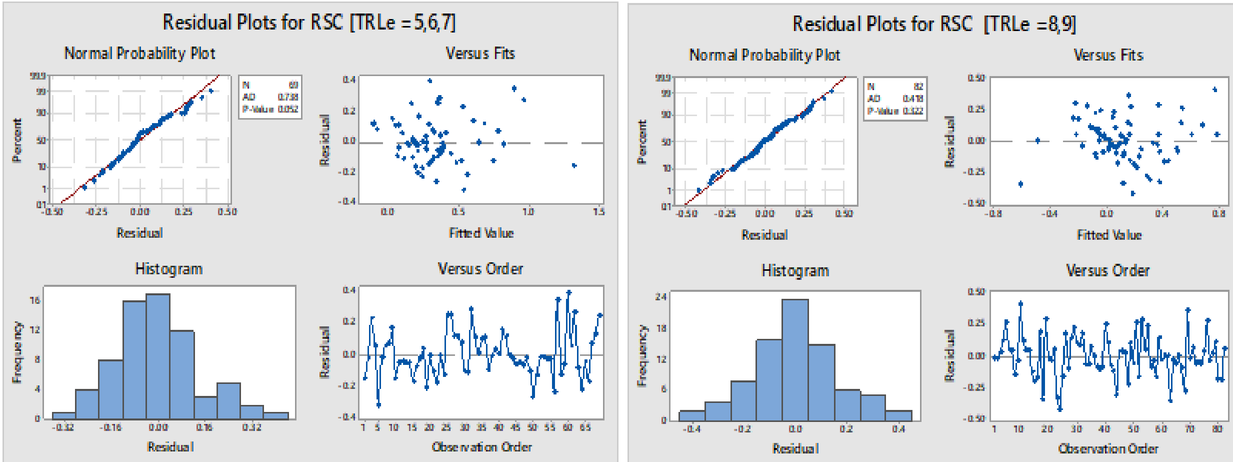
Source	Coef	Contribution	p	VIF
Regression	0.552	74.27%	0	
LN.RD.M	0.0635	0.25%	0.009	2.82
LN.P.M	-0.0807	16.51%	0	2.02
Cycle.Mo	-0.004734	27.51%	0	2.74
B.st	0.0633	9.63%	0	2.81
C.B	0.0207	0.48%	0.009	1.5
Restr	0.152	2.14%	0.002	1.49
PM.oth	-0.1146	1.49%	0.012	1.32
complex_sys	0.3115	0.31%	0	2.68
SoS_part	0.3782	14.22%	0	2.85
DEPEND	-0.1301	0.80%	0.014	1.58
OPER	-0.1003	0.92%	0.058	1.73

S	R-sq	R-sq(adj)	PRESS	R-sq(pred)
0.160606	74.27%	69.30%	2.3534	58.81%

Systems with TRLe = [8,9]

Source	Coef	Contribution	p	VIF
Regression	0.693	68.58%	0	
LN.P.M	-0.1058	10.59%	0	1.65
Eng.2	-0.000414	0.23%	0.008	3.26
P_no	-0.000001	3.43%	0.013	1.87
C.B	0.0278	0.01%	0.013	2.92
Restr	-0.3253	2.78%	0	3.05
PM.oth	0.3648	1.64%	0	2.03
COML	0.1854	1.75%	0.006	2.44
Prototype	-0.278	2.51%	0	2.23
NM	0.1334	0.54%	0.039	1.86
Type	* by factor	37.30%	0	* by factor
complex_sys	0.2167	1.73%	0.013	2.00
SoS_part	0.1271	1.68%	0.072	1.86
SWAP	0.2466	0.10%	0.003	*
RMA	0.1334	4.28%	0	1.96

S	R-sq	R-sq(adj)	PRESS	R-sq(pred)
0.188684	68.58%	59.60%	3.852	46.03%



Estimated TRL = [5,6,7]
N= 69
Anderson- Darling Test
0.738
P-Value 0.052

Estimated TRL = [8,9]
N= 82
Anderson- Darling Test
0.418
P-Value 0.322

Figure 2. **Residual Plots for System Maturity-Related Schedule Regressions**

Clustering the linear regression predictors into four broad groups—Resources, Programmatic, OT and Schedule—provides an efficient representation of the changing relative importance of factor groups as a system proceeds from immature to mature per Table 11.

Table 11. Schedule Variance Percentages by Factor Groups

	[TRLe=5,6,7]	[TRLe = 8,9]
Resources	44.27%	14.26%
Program	18.16%	49.93%
OT	1.72%	4.38%
Schedule	10.11%	0.01%
error	25.74%	31.42%

These schedule growth models showed that immature systems need adequate resources (including both time and money as resources), a good initial schedule plan (getting the schedule right), and a plan to manage system complexity driven by interactions with the larger system of systems. Mature system schedule growth is driven by commodity type (for example, aircraft or ship) integration issues.

Conclusions

This research examined different measures of technology and system maturity and identified maturity-related factors. We used logistic regression analysis to show relationships between system maturity and program schedule growth. This research is valid for MDAPs with reports issued by both the GAO and DOT&E in the same year from 2007 through 2017. Research findings may not be valid for MDAPs not in these reports, highly classified programs, defense business systems and smaller expenditure acquisition programs. The research provides program offices insight into technology and system maturity and the sources of schedule growth based on resource, programmatic, operational testing, and schedule-related factors, allowing them to monitor and adjust acquisition program planning and execution. Examples of useful results for program managers include the following:

- This research developed operational testing performance factors. These factors were shown to be related to system maturity and program schedule growth. Program managers may use such factors to help develop quantitative measures of system maturity related to factors seen during development and testing.
- Program managers may use the combination of a reported or estimated GAO technology maturity and a predicted or estimated TRL of 8 or 9 as an indicator of *system* maturity during system Engineering and Manufacturing Development.
- The research showed that resources (having enough money and time to develop a system) matter most before a system is mature and that program structure and execution matter later in program execution. However, program managers need both program resources and structure from the start to deliver and support their products.
- The research showed that an *inability* to adhere to planned schedule indicates system immaturity.

References

Assistant Secretary of Defense for Research and Engineering (ASD[R&E]). (2011, April). *Technology readiness assessment (TRA) guidance*. Washington, DC: DoD. Retrieved from https://www.acq.osd.mil/ecp/DOCS/DoDGuidance/TRA2011.pdf

Azizian, N., Mazzuchi, T., Sarkani, S., & Rico, D. F. (2011). A framework for evaluating technology readiness, system quality, and program performance of U.S. DoD acquisitions. *Systems Engineering*, *14*(4), 410–426.

Bailey, R. U., Mazzuchi, T. A., Sarkani, S., & Rico, D. F. (2014). A comparative analysis of the value of technology readiness assessments. *Defense Acquisition Research Journal*, *21*(4), 825–850.

Conrow, E. H. (2011). Estimating technology readiness level coefficients. *Journal of Spacecraft and Rockets*, *48*(1).

Day, G. (1981). The product life cycle: Analysis and applications issues. *Journal of Marketing*, *45*(4), 60. https://doi.org/10.2307/1251472

Dodaro, G. L. (2007). *Defense acquisitions: Assessments of selected weapon programs* (GAO-07-406SP; p. 178). Washington, DC: GAO.

Dodaro, G. L. (2013). *Defense acquisitions: Assessments of selected weapon programs* (GAO-13-294SP; p. 190). Washington, DC: GAO.

Drezner, J. A., & Smith, G. K. (1990). *An analysis of weapon system acquisition schedules* (Acquisition No. R3937; p. 212). Santa Monica, CA: RAND.

Gove, R., & Uzdzinski, J. (2013). A performance-based system maturity assessment framework. *2013 Conference on Systems Engineering Research*, *16*, 688–697. https://doi.org/10.1016/j.procs.2013.01.072

Kamp, J. (2019). *Integrating immature systems and program schedule growth*. Washington, DC: The George Washington University.

Kandt, A., Pickshaus, T., Fleischer, K., & Schmitt, R. (2016). A new model to ascertain product maturity in product development processes. *26th CIRP Design Conference*, *50*, 173–178. https://doi.org/10.1016/j.procir.2016.05.006

Katz, D. R., Sarkani, S., Mazzuchi, T., & Conrow, E. H. (2015). The relationship of technology and design maturity to DoD weapon system cost change and schedule change during engineering and manufacturing development: Relationship of maturity to DoD weapon system effectiveness. *Systems Engineering*, *18*(1), 1–15. https://doi.org/10.1111/sys.21281

Major Defense Acquisition Program Defined, 10 U.S.C. § 2430 (2007). Retrieved from https://www.gpo.gov/fdsys/pkg/USCODE-2006-title10/html/USCODE-2006-title10-subtitleA-partIV-chap144-sec2430.htm

Mankins, J. C. (2009). Technology readiness assessments: A retrospective. *Acta Astronautica*, *65*(9), 1216–1223. https://doi.org/10.1016/j.actaastro.2009.03.058

Mattis, J. (2018). *Summary of the 2018 national defense strategy: Sharpening the American military's competitive edge*. Washington, DC: DoD. Retrieved from https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf

Monaco, J. V., & White III, E. D. (2005). Investigating schedule slippage. Defense Technical Information Center.

Nolte, W. L. (2008). *Did I ever tell you about the whale? Or measuring technology maturity*. Charlotte, NC: Information Age.

Ramirez-Marquez, J. E., & Sauser, B. J. (2009). System development planning via system maturity optimization. *IEEE Transactions on Engineering Management*, *56*(3), 533–548. https://doi.org/10.1109/TEM.2009.2013830

Stuckey, R. M., Sarkani, S., & Mazzuchi, T. A. (2017). Complex acquisition requirements analysis: Using a systems engineering approach. *Defense Acquisition Research Journal: A Publication of the Defense Acquisition University*, *24*(2), 266–301.

Tate, D. M. (2016). *Acquisition cycle time: Defining the problem* (Revised).

Tetlay, A., & John, P. (2009). Determining the lines of system maturity, system readiness and capability. In *Readiness in the system development lifecycle*. England: Loughborough University. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.614.9724

Wang, Q., Wang, Z., & Zhao, X. (2015). Strategic orientations and mass customisation capability: The moderating effect of product life cycle. *International Journal of Production Research*, *53*(17), 1–18. https://doi.org/10.1080/00207543.2015.1027012

Weapon Systems Acquisition Reform Act of 2009, Pub. L. No. 111-23 (2009). Retrieved from https://www.gpo.gov/fdsys/pkg/PLAW-111publ23/content-detail.html

Panel 17. Setting Programs Up for Success: Effective Analysis and Alternatives

Thursday, May 9, 2019	
10:30 a.m. – 11:45 a.m.	**Chair: Phil Rodgers,** Director, Acquisition Management and Approaches, Office of the Under Secretary of Defense for Acquisition and Sustainment ***A Robust Framework for Analyzing Acquisition Alternatives (AoA)*** Brian Morgan and Moshe Kress, Naval Postgraduate School ***Rethinking Government Supplier Decisions: The Economic Evaluation of Alternatives (EEoA)*** James Fan and Francois Melese, Naval Postgraduate School ***A Framework for a Defense Systems Effectiveness Modeling and Analysis Capability: Systems Effectiveness Modeling for Acquisition*** John Green, Naval Postgraduate School

A Robust Framework for Analyzing Acquisition Alternatives (AoA)

Moshe Kress—Operations Research Department, Naval Postgraduate School. [mkress@nps.edu]

Brian Morgan—Operations Research Department, Naval Postgraduate School. [blmorgan@nps.edu]

Abstract

Analysis of Alternatives (AoA) is a complex multi-dimensional decision procedure used by the U.S. Department of Defense as part of the acquisition process. The four dimensions of the procedure are *alternatives*, *criteria*, *scenarios*, and *stakeholders*. Current AoA studies lack the structural rigor needed from such a complex procedure, which involves measurements, evaluations, analyses, and modeling, as well as social and group-decision aspects—all in a highly uncertain environment. We propose a structured paradigm for conducting AoA, rooted in well-established methods of multi-criteria decision analysis. The core of the methodology comprises the concepts of ratio-scale matrices and the Euclidean norm. The ratio-scale matrices are used to elicit evaluations, preferences, and opinions from individual stakeholders and analysts, and the Euclidean norm is utilized to mitigate possible preference inconsistencies and help form consensus.

Introduction

The U.S. Department of Defense (DoD) Acquisition System comprises three interconnected stages that start with specifying requirements—a procedure called Joint Capabilities integration and Development System (JCIDS). The second stage, called the Acquisition Process, focuses on determining appropriate materiel solutions for the requirements specified in the first stage. The third stage has to do with executing the decision made at the second stage. It includes funding and control activities contained in the Planning, Programming, and Budgeting Execution (PPBE) Process (DoD, 2017).

The first stage comprises tactical and operational analyses based on wargames, simulations combat models, and input from subject matter experts. It identifies gaps in current capabilities and projects future needs based on evolving threats and operational postures. The third stage is the Department of Defense's resource allocation process that includes an annual budget, for presentation to Congress, linking missions to the requested funding.

The second stage—the Acquisition Process—comprises two interrelated phases: a *creative* phase and an *analytic* phase. The outcome of the creative phase is a set of potential materiel solutions to the operational requirements specified in the Initial Capabilities Document (ICD) produced in the first stage. This set comprises acquisition alternatives to be analyzed in the analytic phase. Obviously, the set includes only those alternatives that are evidently reasonable and viable. In other words, no alternative in the set can be without a capability, or violate a clear requirement, specified in the ICD. The analytic phase focuses on evaluating the alternatives with respect to several criteria, while incorporating quantitative analysis with multiple stakeholders' opinions and preferences. The outcome of this phase is a recommendation on the most preferred alternative(s) to be considered for acquisition. This recommendation must be based on multi-criteria evaluations of the alternatives and reflect a consensus among stakeholders' opinions, goals, and preferences.

This study focuses on the analytic phase of the second stage—also called Analysis of Alternatives (AoA; see, for example, RAND Corporation, 2006). Our objective is to propose a comprehensive formal framework for executing AoA and introduce a unified analytic structure into it. The proposed framework is general enough to be easily tuned to any specific AoA study in any branch of the armed services.

The core of the analytic phase (AoA) is a Multi-Criteria Decision Analysis (MCDA) process. In this process, alternatives are evaluated according to a set of criteria, and the resulting evaluations are then aggregated into a rating or a score that represents the relative standing of each alternative. In a DoD acquisition context, the criteria typically include scenario-dependent operational effectiveness, technological feasibility and risk, supportability, compatibility (with existing systems), and cost. While the general spirit of MCDA is indeed present in typical current AoA projects, its actual manifestation varies significantly among studies (see, e.g., DoN, 2006; RAND Corporation, 2006, 2016; Souders et al., 2004; Smith et al., 2011; TRADOC, 2011). Crucially, most of these studies lack the structural or formal rigor that is desired in a critical decision process such as AoA. Typical weaknesses relate to in-context evaluations of alternatives with respect to criteria, determining the weights of criteria, treating uncertainty and risk, and adequately aggregating preferences among stakeholders. These issues are either not addressed in those studies at all or they are treated in inconsistent ways across studies. Moreover, as much as it is an analytic process, AoA is also a social process that involves several (sometimes many) stakeholders. Different stakeholders, representing various DoD branches and organizations, may have different opinions, points of view or preferences regarding the importance (i.e., weights) of criteria. They may also differ in their assessments about the likelihood of future scenarios and disagree about the values of alternatives with respect to qualitative (subjective) criteria, where measures of performance (MOPs) and/or measures of effectiveness (MOEs) are either difficult to compute or do not exist altogether. Even measurable (objective) criteria, such as detection range, velocity, and firing accuracy (say, probability of hit), may be scaled differently by different stakeholders. This important social aspect seems to be ignored in current AoA studies. In our proposed framework, we attempt to remedy these, and other, shortcomings.

The main contributions of this paper are (a) proposing a clear "standard" for conducting AoA in the U.S. Department of Defense, (b) explicitly addressing the role of scenarios and stakeholders in the AoA process, and (c) developing an all-inclusive distance-based model that addresses, simultaneously, all four dimensions of AoA—alternatives, criteria, scenarios, and stakeholders.

MCDA models considering alternatives and criteria are quite abundant (e.g., AHP, ELECTRE, and PROMETHEE; Behzadian et al., 2010; Figueira et al., 2005; Saaty, 1980). There are also MCDA models that consider scenarios (Montibeller, 2006; Stewart, 1997), and those which consider consensus formation among multiple stakeholders (Cook et al., 1996). But, to the best of our knowledge, the model presented here is the first attempt to tackle all four dimensions in a unified and robust fashion in the context of DoD AoA.

The paper is organized as follows. In the next section, we review the four main dimensions of an acquisition AoA: *alternatives*, *criteria*, *scenarios*, and *stakeholders.* Then we describe the set of criteria relevant to a typical DoD acquisition AoA, and their imbedded hierarchy. The next section is the heart of this paper; it formally describes the MCDA methodology underlying the AoA process. Then we address the uncertainty associated with future scenarios and the way it affects the choice of the winning alternative. Finally, we outline the six steps of a robust AoA study.

In the rest of the paper we refer to the subject of the AoA study as *item.* An item can be a Navy fighter, a radar system, a transport vehicle, a supply ship, a command and control system, etc. The objective of the AoA process is to select for acquisition the most appropriate item out of a set of alternative items.

The Four Dimensions of Acquisition AoA

Once the operational needs and/or capability gaps have been identified at the JCIDS stage, an initial set of items—potential materiel or non-materiel solutions to these needs—is generated at the onset of the next stage: the Acquisition Process. Generally speaking, the ultimate goal of the acquisition process is selecting, out of an initial set of possible items, an item that provides the best balance between (in-context) utility or value and potential cost and risk. The members of this set of items are called *alternatives*. The alternatives represent the first main component of the acquisition AoA. The other three main components are *criteria*—the touchstones according to which alternatives are evaluated, *scenarios* that provide the operational backdrop for the evaluation, and *stakeholders* who contribute analytic inputs, as well as preferences, opinions and judgements, into the acquisition decision process.

Alternatives

Generating the initial set of alternatives is the “creative” part in acquisition AoA. The generators of the alternatives are typically defense agencies, who may suggest existing materiel options or off-the-shelf items, and defense contractors who propose either existing products currently produced or items that are at various stages of maturity in the research-and-development stage. The items suggested may range from the mundane (e.g., the current “status quo” alternative) to the daring (e.g., an item based on revolutionary, and perhaps even immature, technological concept).

In some AoA studies, there exists a legacy item (ship, weapon, C2 system, etc.) that either is near its end of life or its capabilities are insufficient for emerging requirements. In such cases, it is important to clearly identify the characteristics of the legacy alternative, which can be considered as a baseline according to which potential upgrades are considered (MITRE, n.d.).

The set of alternatives should be carefully constructed. It must be non-trivial (e.g., just two alternatives where one clearly dominates the other), but also manageable in size. There is hardly an effective and meaningful way of handling the evaluation of dozens of alternatives. One way of reducing the size of the alternatives’ set is eliminating similar alternatives—alternatives that differ marginally or those that are evidently dominated by other alternatives.

The alternatives should also be realistic in the sense that they are technologically feasible and grounded in industry’s capabilities. The set of alternatives should not include idealized items that have no practical basis in industry or government. The set of acquisition alternatives may be divided into categories:

- Modified existing items currently in operation,
- As-is or modified off-the-shelf items available in the market but not yet in operational use,
- Repurposing and/or recombining existing items with new technologies, and
- Newly developed items (USAF, 2016).

The four categories differ in their potential effectiveness, cost, and the risk associated with their acquisition. To modify a legacy item would probably be cheaper and less risky than developing a completely new one, but a new item would most likely be more effective and more attuned to current requirements than the modified legacy item. Roughly and generally speaking, the main thrust of the AoA process is to tradeoff among these three contrasting aspects—effectiveness, cost, and risk.

Criteria

The merits and weaknesses of the alternatives are evaluated by criteria, which represent various aspects related to the operation, functionality and reliability of the alternatives, the risk associated with their selection, and the cost factors related to acquiring, handling, and maintaining them. In general, the set of criteria for evaluating defense (physical) acquisition items such as weapons, sensors, and platforms is divided into four subsets:

- Effectiveness
- Operationability, reliability, maintainability, and logistics (ORML)
- Cost
- Risk

While effectiveness is measured by specially constructed measures of effectiveness (MOE), and cost is typically measured in money spent (and/or to be spent), ORML criteria are measured by both MOEs and cost factors.

Effectiveness

An old adage states that "among all the alternative items that are completely useless for a certain requirement, the best one to be selected is the cheapest." In other words, the main driver for selecting an alternative is its usefulness or effectiveness with respect to the requirements that generated the acquisition process. The term *effectiveness* may be elusive and may mean many different things. Measuring effectiveness of an alternative is probably the most challenging part of an AoA study. To demonstrate the complexity of this challenge, consider the following simple (in fact, simplistic) example:

> The requirement is for an anti-air (AA) weapon, and the only two criteria are fire-rate and single-shot kill probability (SSKP). There are two alternative weapons for consideration. Weapon A has a higher fire rate than Weapon B, but smaller SSKP than Weapon B. Which weapon is more effective? Weapon A can deliver higher "quantity" of shots while Weapon B has a better "quality" per shot.

One way of measuring the (relative) effectiveness of the two AA weapons is to determine a tactical or operational objective (e.g., maximize number of targets killed), determine an appropriate MOE (e.g., expected number of killed targets within a certain time period) and construct a model (analytic or simulation) that calculates the values of the MOE for the two weapons. Another way to determine the relative effectiveness of the two weapons is to treat each attribute—fire-rate and SSKP—as separate criteria, give a score to each weapon with respect to each criterion, and then combine the scores of the two criteria, via, say, a weighted combination, into a single score—one for each weapon.

The first approach could be considered "objective" in the sense that there is a quantitative model that bridges between the data and assumptions, and the final evaluation of the two weapons. The second approach is "subjective" in the sense that stakeholders and/or subject-matter experts must provide their personal input in determining the scores of the alternatives with respect to the criteria and the weights of the criteria.

In reality, and unlike the above example, *effectiveness* has more than two aspects, and thus measuring effectiveness becomes more challenging. Ideally, there would be one measurable objective (e.g., maximize probability of winning the engagement/battle/campaign) that encompasses all relevant operational aspects of the item and the scenario in which it is to be implemented (see below). The measurable objective would be formalized as an MOE, which could be reliably computed in a comprehensive model. Unfortunately, this ideal setting seldom occurs. Either there are multiple objectives or the scenario and the role of the item in it are so complex that no model can reliably capture all the salient aspects.

The bottom line is that, in reality, effectiveness in an AoA is evaluated by a mixture of the two approaches—the analytic “objective” approach and the opinion- or experience-based, “subjective” approach. The goal is to enhance, as much as possible, the analytic side and thus minimize the possible biases and disagreements (see below) that may be generated from the subjective approach.

Operationability, Reliability, Maintainability, and Logistics (ORML)

During its course of operations, an item must be operated (or controlled) by qualified persons, professionally maintained, and regularly serviced and resupplied. These requirements result in operational, as well as economic, implications. Obviously, ceteris paribus, an item that is more reliable and requires less maintenance, less qualified operators and lighter logistic burden is preferred to an item that is rated worse on any of these aspects. The question is that of trade-offs; how much effectiveness one would be willing to sacrifice for a simpler, more robust, and lower-maintenance system?

Operationability is a criterion that reflects two salient aspects of a newly acquired item: (a) compatibility with existing systems, currently in use, with which the new item has to interact, and (b) human-system integration (HSI). A new radar must interact with existing sensors, command and control systems and weapons, and therefore must be *compatible* with them. However, measuring compatibility is challenging; there is no natural MOE that could be defined and objectively evaluated for measuring how well a certain alternative item interacts with current systems. This is a “subjective” criterion that must be evaluated qualitatively by subject-matter experts (SMEs). Similar restrictions also apply to the other part of Operationability—HSI. While, in principle, one could use the number of operators, classified by technical background, length of service and pay-grade, and estimated length of the training period, as surrogate MOEs for HSI requirements, in practice it would be difficult to do it. Here, once again, evaluating this criterion will most likely be done by qualitative input obtained from SMEs.

Reliability affects the readiness of the item. The more reliable an item is, the less frequently it is unexpectedly down. This criterion is quantitative and is typically measured by the mean time between failures (MTBF). While measuring MTBF of an existing system is a relatively straightforward statistical task, estimating the value of this criterion for items in a design or development phases is challenging because of lack of statistical data. Thus, reliability estimates must depend on engineering-based projections based on the item’s design and the technical specifications of its components, and perhaps some statistical data available about similar systems. In many cases, these estimates are provided by the vendors of the items, in which case the projected reliability values must be taken cautiously.

Maintainability is an attribute that describes the technical and physical burden associated with an item. Arguably, a modular item that requires a “plug-and-play”–type service is more maintainable than an item that comprises hard-wired components, which, when failed, result in the need for system-wide service. Similar to reliability, maintainability of an item could be measured by MOEs such as average mean service time over all components of the system, or the maximum mean service time, or other statistical measures of repair and maintenance services. The same challenges that apply to the reliability criterion, when an item is still in the design or development phases, apply here too. Maintainability could be considered a fully quantitative criterion only for existing items, which have accumulated enough maintenance experience and data. Otherwise, maintainability is evaluated by SMEs.

Logistics refers to the operations and cost aspects related to the transshipment of items, and the supply chain of consumables (e.g., ammunition, fuel) and repair parts needed for their operation and maintenance. There are typically two logistic aspects associated with an item: (a) the physical infrastructure needed for storage and maintenance of the item, and (b) transportation and handling equipment for transporting the item and its required supplies. For example, transporting fuel requires specially designated tankers. Certain items may also impose logistic constraints. For example, Vertical Launching Systems (VLS) missiles used by the U.S. Navy cannot currently be resupplied at sea. In order to replenish this type of ammunition, warships must return to port.

Cost

While end-users of a military item—commanders, combat developers, operations officers, etc.—are mostly concerned with effectiveness and operationability of the item, DoD program managers and budget officials may be mostly concerned with its overall cost (see the discussion of *stakeholders* below). Cost comprises several expenditures that vary in their nature (e.g., R&D, production, life-cycle), the time horizon during which they are to be realized, and the certainty regarding their monetary size. Arguably, costs related to future expenditures (e.g., maintenance) are more uncertain then the R&D cost for, say, an item in an advanced development stage, or purchasing price for an off-the-shelf item. The cost criterion can be broken down to sub-criteria representing its various components in order to reflect preferences of immediate versus future expenditures.

Risk

The most complex and elusive criterion in the AoA process is risk, and arguably, it mostly applies to items that are not readily available. The risks are: delays in development, disrupted production schedules, running costs over budget, and difficulties in assimilating the item in the force. Alternatives, which are already existing items or very close to maturity, have relatively little or no risk regarding their availability at the time when they are needed. But other risks, associated with re-production, costs, and assimilation may still exist. For less mature alternatives, the more technologically challenged the item, the higher the probability that something will go wrong during the research and development stages, as well as in the production phase. The problem is that it is extremely difficult to estimate this probability. Therefore, this criterion is essentially "subjective," where risk assessments are mostly based on inputs from subject matter experts or qualitative projections based on data from past similar experiences.

Scenarios

Scenarios may be considered as "Uber criteria." They form the settings in which the alternatives are evaluated with respect to the "regular" criteria described earlier. There are two types of scenarios to be considered in an AoA study. The first type refers to the operational setting in which the item is designated to operate. Military conflict scenarios—and in particular, combat scenarios—are used for in-context evaluation of the effectiveness, operationability, and logistics of items such as weapons, C2 systems, sensors, and other defense- and military-related items. For example, the importance of the *range* criterion of a sensor—an Effectiveness criterion—may depend on the typical detection ranges applicable in a certain operational scenario. The importance of the robustness of a vehicle to road conditions may depend on the typical terrain in a scenario. Thus, the designated operational setting of an item is important for evaluating the item's potential effectiveness. An alternative that performs well over a wide range of plausible scenarios may be preferred to an alternative that performs very well on limited operational settings but poorly on other likely settings.

The second type of scenarios applies to AoA of items that do not yet exist and are in various stages of the research and/or development phases. These scenarios describe economic, social, political, and technological factors that may affect the risk associated with selecting a certain alternative. For example, if a certain alternative requires a considerable R&D effort, the Risk, and perhaps the Cost, criteria associated with that alternative will be impacted by the availability of economic resources and technological capabilities.

Both types of scenarios incorporate a fair amount of uncertainty that must be factored in the AoA study. The way scenarios are incorporated in an AoA study is discussed in more detail in the modeling part described below.

Stakeholders

As much as a technical and analytical process, AoA is also a social phenomenon involving a plethora of stakeholders who may represent different interests, viewpoints, agendas, and goals. For example, combatants—the future users of the item—may focus on the effectiveness of the item and its compatibility with existing combat systems currently in use. Combat developers may look at a much wider picture and will be concerned with issues of force structure and other strategic considerations. Technical experts will focus on the scientific and engineering aspects, and in particular on potential technological challenges that may affect the Risk criterion. Finally, budget officials will naturally focus on the programmatic aspects associated with the developing, production, operation, and maintenance of the item. In other words, the Cost criterion plays a major part for these stakeholders.

Because the AoA process is complex and multidimensional, and some criteria (dimensions) may be conflicting, it is important to select a balanced mix of stakeholders for the study—representing all the aforementioned groups of decision makers and experts who represent different aspects of the decision problem.

The Set of Criteria

The criteria are the touchstones that determine the in-context value of an alternative. Obviously, the goal is to select the alternative with the highest overall value when all relevant criteria are considered. The set of criteria should adhere to some structural, as well as content, rules and properties, which are described in the following sections.

Criteria Tree

It is convenient to view the set of criteria as a hierarchical structure. This view is not new; it is manifested, for example, in the Analytic Hierarchy Process (AHP; Saaty, 1980), which is used by the DoD. The idea is to break down the main four criteria—(1) effectiveness, (2) operationability, reliability, maintainability, and logistics (ORML), (3) cost and (4) risk—into sub-criteria, sub-sub-criteria, and so on. This breakdown induces a tree structure whose leaves (lowest hierarchy) are criteria that can either be measured by Measures of Performance (MOPs) or Measures of Effectiveness (MOEs), or can meaningfully be evaluated qualitatively by subject matter experts (SMEs). The aforementioned four criteria constitute the first layer of the criteria tree, as shown in Figure 1.

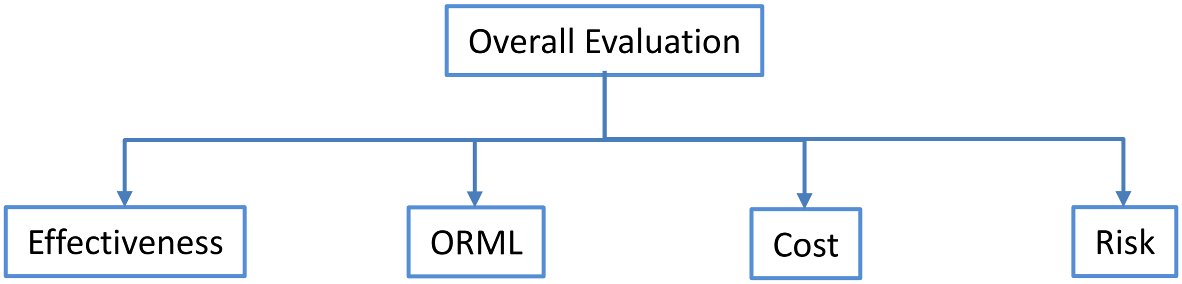


Figure 1. **First Layer of a Criteria Tree**

If the item to be selected is, for example, some kind of a ground fighting vehicle (e.g., a tank), then possible second and third layers of sub-criteria, which evolve from the Effectiveness criterion, are shown in Figure 2.

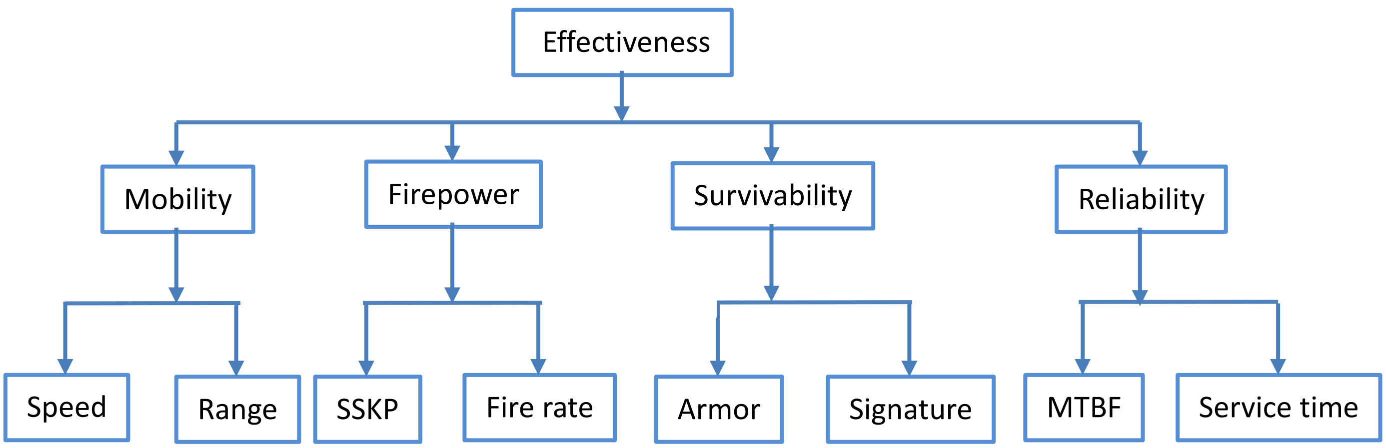


Figure 2. **Second and Third Layers of a Criteria Tree**

The third layer can be further broken down. For example, *speed* could be divided according to the type of terrain, *single-shot-kill-probability* (*SSKP*) may be separated into *Day-SSKP* and *Night-SSKP*, and so on.

Criteria Properties

A proper design of the criteria tree is crucial for the success of an AoA project. Specifically, the number of layers in the criteria tree and the granularity of each layer depend on the context and thrust of the analysis and on the complexity of the parent criterion in a higher layer. On the one hand, it is important to include all relevant sub-criteria that affect the parent criterion. On the other hand, we need to avoid over-cluttering the criteria tree such that it remains as manageable as possible. Keeney and Raiffa, in their seminal 1976 work, suggested some rules or properties that should guide the way criteria are selected for the analysis. In particular, the set of criteria must be *complete* in the sense that the "leaves" of the criteria tree—the end criteria at the lowest layer—cover all the aspects affecting the choice of the item. The criteria must also be *operational*—they must be relevant to the decision problem and meaningful to the decision-makers. Another important rule is to avoid *redundancy* that can lead to the undesirable effect of double counting. For example, the criterion “range of an aerial platform” may be redundant in the presence of the criterion “endurance of an aerial platform.” Finally, as mentioned above, the set of (end) criteria must be as small as possible, notwithstanding the other properties.

It is noted that in some cases, breaking down a criterion to more refined sub-criteria (in a lower layer of the criteria tree) may be counterproductive when the sub-criteria are interdependent. Two criteria are dependent if the importance or *weight* (see Weights of Criteria section below) of one criterion is affected by the evaluation of the alternative with respect to the other criterion. For example, the speed and maneuverability of a fighter aircraft might be dependent; if the aircraft is slow, the maneuverability may be more important than if the aircraft is fast. In that case the two sub-criteria may be combined into a single criterion such as *flight performance*.

Weights of Criteria

Different criteria may have different levels of importance, or different *weights*. An important fact to remember is that these weights are *subjective*. There is no scientific method that could measure the “true” weight of a criterion. Different stakeholders may have different opinions regarding the impact a certain criterion has on the overall value of an alternative. Moreover, the weight of a criterion may also depend on the scenario; a certain capability of an item may vary according to the scenario in which the item is to be employed. For example, the importance of the criterion “Electro-Optical Signature” of a platform depends on the detection capabilities of the adversary in a conflict scenario. Absent such capabilities, the weight of this criterion will most likely be quite low. Another example is the reliability of equipment. If the system has large redundancy with respect to the availability of this equipment, then the weight of the *reliability* criterion may be lower than in the case where the system relies on a single availability of that equipment. Also, the economic, political, social, and technological scenario may affect the weight to be assigned to the risk-related criteria. In the next section, we describe a method for eliciting weights that take into account the aforementioned factors: multiple stakeholders and multiple scenarios.

Methodology

The AoA process is about comparing the *values* of the alternatives. The best alternative—the one to be selected for acquisition—is the alternative that provides the highest overall value. But how can one combine multiple criteria and opinions into a single value? What is the scalar function that translates measurements and evaluations of the alternatives with respect to the various criteria, and evaluations of criteria weights, into a single value that can be compared among alternative items? The problem is exacerbated in the presence of multiple stakeholders who may provide a plethora of opinions and multiple scenarios that may result in different in-context evaluations.

We start with the basic construct, which is common in many decision settings—a linear value function (see e.g., Saaty, 1980). Simply stated, if w_j denotes the weight of criterion j, and v_{ij} is the value of alternative i with respect to criterion j, $i = 1,...,I,\ j = 1,...,J,$ then the overall value of alternative i, V_i^*, which is to be compared with the overall values of the other alternatives, is given by

$$V_i^* = \sum_{j=1}^{J} w_j v_{ij}, \ i = 1,...,I \qquad (1).$$

The alternative with the highest V_i^* value is the most likely alternative to be selected. Note that we state "Most likely" and not "definitely." This distinction is discussed further in the last two sections of this report.

As mentioned earlier, the challenges are to determine the values of $w_j, \quad j = 1,...,J,$ and $v_{ij}, \ i = 1,...,I, \ j = 1,...,J,$ taking into consideration the presence of multiple stakeholders and multiple scenarios. We will construct our value function step-by-step, starting with determining the values of the weights w_j.

Determining Criteria Weights for a Certain Scenario

Consider a certain reference scenario s. Assuming this scenario is realized, we wish to elicit from R stakeholders criteria weights $w_{js}, \ j = 1,...,J,$ that (a) reflect the relative importance of the various criteria if scenario s prevails, and (b) represent a consensus among the stakeholders regarding these weights. An efficient and effective way to elicit preferences from decision-makers is through ratio-scale matrices, similarly to the setup used in AHP (see Saaty, 1980). The idea is as follows: Each stakeholder $r, r = 1,...,R,$ is asked to compare two criteria weights, say w_{js} and w_{ks}, with respect to scenario s. In other words, the stakeholder provides an assessment regarding the extent one criterion is more (or less) important than the other. The comparison is in terms of the ratio between the two weights. That is, p_{jks}^r is the assessment of stakeholder r regarding the ratio $\frac{w_{js}}{w_{ks}}$. Different stakeholders may have different opinions regarding the very same ratio. In other words, for two stakeholders r and r', we may have $p_{jks}^r \neq p_{jks}^{r'}$. A natural way to mathematically resolve such discrepancies is using least squares. The same way least squares are used to fit a "consensus" line among a clutter of points in statistical linear regression, we can derive a consensus set of weights by minimizing the least-square or L_2 distance. The usefulness and effectiveness of the least-square (L_2) measure as consensus forming method in decision analysis is described in Golany and Kress (1993). Formally, we solve the following non-linear optimization problem:

$$Min \sum_{j<k} \sum_{r=1}^{R} \left(\frac{w_{js}}{w_{ks}} - p_{jks}^r \right)^2 \qquad (2)$$

$$st$$

$$\sum_{j=1}^{J} w_{js} = 1, \quad w_{js} \geq 0.$$

The objective function is separable and quadratic, and therefore the optimization problem is easily solvable for real-size problems by tools as simple as the MS Excel Solver. The constraint is just a normalization of the criteria weights, which facilitates simpler computations down the road.

The optimal solution of Problem 2 is a vector $W*_s = (w*_{1s},...,w*_{Js})$ of criteria weights that represent an L_2 consensus regarding the criteria importance in the presence of scenario s.

We solve Problem 2 S times—once for each possible scenario. For brevity and simplification, we drop the * sign from future notation.

The model in Problem 2 described above for criteria weights can obviously be applied, sequentially, to the different levels of the criteria tree (see above). For each master criterion at level l, we solve Problem 2 for the "child" criteria at level $l+1$. The weight of the end criterion at the bottom level is the product of the criteria weights leading to that criterion. For example, considering Figures 1 and 2, we first solve Problem 2 for *Effectiveness, ORML, Cost*, and *Risk.* Next, for the master criterion *effectiveness*, we solve 2 for *Mobility, Firepower, Survivability*, and *Mobility*. Similarly, Problem 2 is solved for the children (if any) of *ORML, Cost*, and *Risk.* Finally, we solve 2 for the lowest level (e.g., *Speed* and *Range* for *Mobility, SSKP* and *Fire-rate* for *Firepower,* etc.). The weight of the end-criterion in the value function 1, say *Speed,* is the product $w_{Effectiveness,s} \times w_{Mobility,s} \times w_{Speed,s}$.

Determining Alternatives' Values for a Certain Criterion and Scenario

Once again, we consider a certain scenario s. Let us also consider a certain criterion j. Similarly, to the way criteria's relative weights are elicited from stakeholders, the objective here is to obtain the ratio figure d^r_{iljs} that represents stakeholder's r opinion regarding the ratio between the value of alternative i and alternative l with respect to criterion j, in the presence of scenario s. Similarly to Problem 2, we solve now

$$Min \sum_{i<l} \sum_{r=1}^{R} \left(\frac{v_{ijs}}{v_{ljs}} - d^r_{iljs} \right)^2 \qquad (3)$$

$$st$$

$$\sum_{i=1}^{I} v_{ijs} = 1, \quad v_{ijs} \geq 0.$$

Problem 3 has an identical structure as Problem 2. Here, the optimal values $V*_{js} = (v*_{1js},...,v*_{Ijs})$ are the mathematical consensus values of the alternatives with respect to criterion j, under the assumption of scenario s. Problem 3 is solved $J \times S$ times, once for each criterion and each scenario. As before, for brevity and simplification, we drop the * sign from future notation.

Note that in both Problems 2 and 3 we assume a homogeneous or "democratic" set of stakeholders; no stakeholder's opinion is considered more influential, or with higher weight, than others. If this is not the case, and certain stakeholders' opinions weigh more than others, then the objective functions in 2 and 3 are weighted accordingly with stakeholders' *r*-indexed weights. The problems are still easy to solve.

The Alternative's Value Function in a Scenario

Following the solutions of Problems 2 and 3, we can compute the value of an alternative, with respect to a certain scenario. This value reflects the consensus weights of the criteria $W_s = (w_{1s},...,w_{Js})$, and the consensus (relative) values of the alternatives

$V_{js} = (v_{1js}, ..., v_{Ijs})$. Formally, the consensus overall value of alternative i in scenario s is given by

$$\bar{v}_{is} = \sum_{j=1}^{J} w_{js} v_{ijs} \qquad (4)$$

In other words, if an "Oracle" could tell us the scenario s to be unfold, then the alternative i to be considered for selection is the one for which $\bar{v}_{is}$ is maximized. Absent such Oracle, the probabilities of the various scenarios must be taken into consideration when trying to identify the best alternative.

Scenarios' Probabilities

The old adage claims that "it is very difficult to forecast, especially the future." Nobody knows for sure which of the possible scenarios will actually be realized. Different stakeholders may have different opinions about the likelihood of the various scenarios. The combined assessment of scenarios' probabilities is obtained using the same methodology as in Problems 2 and 3.

Let a_{st}^{r} denote the assessment of stakeholder r about the relative likelihood of scenarios s and t. That is, a_{st}^{r} is the subjective opinion of stakeholder r regarding the extent scenario s is more (or less) likely than scenario t. The consensus probabilities q_s, $s = 1, ..., S$, of the various scenarios is obtained as the solution of the quadratic optimization problem

$$Min \sum_{s<t} \sum_{r=1}^{R} \left(\frac{q_s}{q_t} - a_{st}^{r} \right)^2 \qquad (5)$$

$$st$$

$$\sum_{s=1}^{S} q_s = 1, \quad q_s \geq 0.$$

The optimal solution $Q^* = (q_1^*, ..., q_S^*)$ is the consensus probability distribution of the scenarios. As before, we drop the * sign from future notation.

Selecting the Winning Alternative

Following the operations described in the previous section, the AoA team has an initial set of parameters that reflect the stakeholders' L_2-consensus regarding (a) the weight of criteria in each scenario, (b) the relative value of each alternative with respect to each criterion in each scenario, and (c) the (subjective) probabilities of the scenarios.

Recall from the previous section that for each scenario s we have now a calculated value $\bar{v}_{is}$ for each alternative i. This value represents the L_2-consensus outcome of the stakeholders' group decision process with respect to the relative standing of alternative i, if scenario s is realized. The L_2-consensus about the likelihood of scenario s is q_S. Thus, we have now a (subjective) probability distribution of alternatives' values over scenarios where each value $\bar{v}_{is}$ is associated with a probability q_S.

There are several ways to proceed from this point and identify the alternative that is most likely to be the best among the set of I alternatives. The most natural measure is the expected value where the "winning" alternative is alternative i for which $\sum_{s=1}^{S} q_s \bar{v}_{is}$ is maximized. Here we choose the alternative that "on-average" over the possible scenarios produces the highest relative value. This linear measure is quite common and easy to explain to decision-makers, but it is not always the right yardstick for choosing an alternative, in particular when the specific likelihoods of scenarios are to be looked at in more detail.

Another possible measure is the mode of the distribution; we simply choose the alternative that performs the best with respect to the *most likely* scenario. That is, if $s' = \arg\max q_s$ then the selected alternative i is the one for which $\bar{v}_{is'}$ is maximal. This measure is appropriate if there is one scenario that stands out as very likely—much more than any other scenario. If the induced (subjective) entropy of the scenarios, as implied from the stakeholders' projections, is high, then obviously the mode measure will be inappropriate.

Lastly, and probably most appropriately, it would be better to select an alternative that is *good* over a large set of scenarios than an alternative that is *excellent* over a smaller set of scenarios. The goal here is to seek robustness in the choice of the winning alternative. The idea is as follows.

First, we set a probability threshold. This threshold represents the level of confidence, with respect to the realized scenario, which we wish to associate with the winning alternative. Suppose this probability level is α. Reasonable values of α are in the range 0.6–0.9. Next we generate all the minimal subsets of scenarios whose combined probabilities are at least α. For each such subset, we identify the alternative(s) for which the ***minimum*** value across the scenarios in the subset is ***maximal***. Formally, let $T_1,...T_M$ denote the set of all the α-valued subsets of scenarios. Each subset T_m comprises scenarios with combined probabilities of at least α, and any scenario removed from that set reduces the combined probabilities to less than α (hence, minimal subsets).

Consider an α-valued subset $T_m = \{s_1,...,s_{n_m}\},\ m = 1,...,M,$ where we have $\sum_{k=1}^{n_m} q_{s_k} \geq \alpha.$ Note that each α-valued subset of scenarios may contain different number of scenarios. Define $v_i(T_m) = Min_{s \in T_m} \bar{v}_{is}$ and $i_m = \arg\max\{v_1(T_m),...,v_I(T_m)\}$. Alternative i_m is the max-min alternative of the α-valued subset T_m. In other words, alternative i_m provides the highest ***guaranteed*** value among all alternatives if it is given that one of the scenarios in T_m is realized. Finally, $i^* = \arg\max\{v_{i_1},...v_{i_M}\}$ is the alternative that has the highest value with probability of at least α. Obviously, v_{i^*} is monotone non-increasing in α; the higher the required probability threshold, the smaller the assured alternative value. To demonstrate this procedure, consider the following example:

- Three scenarios, A, B, and C, with probabilities 0.3, 0.3, and 0.4, respectively
- There are 3 alternatives

- The overall values $\overline{v}_{is}$ of the (alternative x scenario) combinations are shown in Table 1.

Table 1: Values of Alternatives

Scenario / Alternative	A (0.3)	B (0.3)	C (0.4)
1	0.7	0.5	0.95
2	0.6	0.8	0.6
3	0.9	0.4	0.5

Suppose α =0.6, which implies that we have here three subsets of scenarios that satisfy the minimum probability threshold requirement: $T_1 = \{A, B\}$ with probability 0.3+0.3=0.6, $T_2 = \{A, C\}$ with probability 0.3+0.4=0.7, and $T_3 = \{B, C\}$ with probability 0.3+0.4=0.7.

For the first scenario set, we have: $v_1(T_1) = 0.5$, $v_2(T_1) = 0.6$, $v_3(T_1) = 0.4$, and therefore the max-min alternative i_1 is alternative 2 with value 0.6. For the second scenario set, we have: $v_1(T_2) = 0.7$, $v_2(T_2) = 0.6$, $v_3(T_2) = 0.5$, and therefore the max-min alternative i_2 is alternative 1 with value 0.7. For the third scenario set, we have: $v_1(T_3) = 0.5$, $v_2(T_3) = 0.6$, $v_3(T_3) = 0.4$, and therefore the max-min alternative i_3 is alternative 2 with value 0.6. Thus, alternatives 1 and 2 are candidates for selection. But the maximum value over the eligible α -valued scenario sets is 0.7 and is obtained by alternative 1. Therefore, at confidence level of at least 0.6, the highest valued alternative is alternative 1.

Notice how the likelihoods of the scenarios affect the choice of alternatives. If, instead of the probabilities values in Table 1, the scenario probabilities were 0.6, 0.2, and 0.2 for scenarios A, B, and C, respectively, then it is easily seen that alternative 3 becomes the most preferred one with value 0.9. Going back to the original probabilities, if the threshold α is now 0.8, then we only have one subset (the complete set of scenarios), and the max-min alternative is alternative 2 with min value of 0.6.

To summarize, this quantile-type approach is both flexible, in the sense that one could choose the confidence level for selecting the best alternative, and robust by adopting the max-min measure of alternatives' values. This approach selects an alternative that is good over a wide range of possible scenarios instead of an alternative that is excellent in only limited number of situations.

Implementation

In the last two sections, we described a formal decision process for conducting AoA, in the presence of several uncertain scenarios, by a group of stakeholders who may have different perspectives and opinions regarding the subject matter. Disagreements and inconsistencies in preferences and assessments may occur with respect to criteria weights, alternative valuations and scenario likelihoods. The proposed group-decision model produces a consensus rating of the alternatives based on minimizing disagreements in the L_2 metric sense.

While the model is transparent and relatively simple to implement in a spreadsheet, it **should not** be considered as a "black box" that automatically produces a winning alternative based on stakeholders' and analysts' inputs. The "winning" alternative that emerges from the model may not necessarily be the final choice in situations where the value(s) of the runner-up(s) is (are) not significantly different from the winner's value. The model is a technical tool that, following a properly designed sensitivity analysis, can help guide the AoA process towards a robust decision. A possible paradigm for conducting a well-structured analysis of acquisition alternatives is as follows:

Step 1: Establish an AoA team that is tailored in size and scope to the military problem being considered. The team must comprise a group of stakeholders (e.g., field commanders and end-users), decision makers (e.g., budget managers, defense officials) and analysts (e.g., engineers, cost-estimators, operations-research analysts, and other subject-matter experts).

Step 2: The AoA team reviews documents describing operational setting, requirements, and capability gaps. An initial set of possible acquisition alternatives is generated. The analysts in the team start gathering more detailed data and information about the operational setting and the possible alternatives.

Step 3: Non-starter alternatives are identified and removed from consideration. Such alternatives are items that are rejected up front because of reasons such as not meeting minimum capability thresholds, they are too costly, or they are based on immature technologies. The team defines the sets of alternatives, criteria, and scenarios. This step also includes open discussions that set the stage for the detailed analysis to follow.

Step 4: Each member r in the AoA team provides her/his estimates for p_{jks}^{r} (see Determining Criteria Weights for a Certain Scenario section), d_{iljs}^{r} (see Determining Alternatives' Values for a Certain Criterion and Scenario), and a_{st}^{r} (see Scenarios' Probabilities). This step includes also operations-research and cost-estimation analyses, which provide valuable inputs to the AoA team.

Step 5: Model implementation on data gathered in Step 4. Output: set of alternative ratings.

Step 6: Discussion on the model results (alternative ratings) and performing sensitivity analysis on all three factors: criteria weights, alternatives' values and scenarios' probabilities. Step 5 may be repeated several times based on the discussions in this step.

We can see that the model described in the Methodology section acts as a decision aid and facilitator for discussions among the team members rather than an "Oracle" that crunches numbers and provides a "solution."

References

Behzadian, M., Kazemzadeh, R. B., Albadvi, A., & Aghdasi, M. (2010). PROMETHEE: A comprehensive literature review on methodologies and applications. *European Journal of Operational Research, 200*(1), 198–215.

Cook, W. D., Kress, M., & Seiford, L. M. (1996). A general framework for distance-based consensus in ordinal ranking models. *European Journal of Operational Research, 96*, 392–397.

DoD. (2017, August 10). *Operation of the defense acquisition system* (DoD Instruction 5000.02). Washington, DC: Author.

DoN. (2006). *Seabase to Shore Connector analysis of alternatives final report*. Reston, VA: WBB Consulting

Figueira, J., Mousseau, V., & Roy, B. (2005). ELECTRE methods. In J. Figueira, S. Greco, & M. Ehrgott (Eds.), *Multiple criteria decision analysis: State of the art surveys* (pp. 133–162). New York, NY: Springer-Verlag.

Golany, B., & Kress, M. (1993). A multicriteria evaluation of methods for obtaining weights from ratio-scale matrices. *European Journal of Operational Research*, *69*, 210–220.

Keeney, R. L., & Raiffa H. (1976). *Decisions with multiple objectives*. J. Wiley & Sons.

MITRE. (n.d.). Systems engineering guide: Analyses of alternatives. Retrieved April 23, 2018, from https://www.mitre.org/publications/systems-engineering-guide/acquisition-systems-engineering/acquisition-program-planning/performing-analyses-of-alternatives

Montibeller, G., Gummer, H., & Tumide, D. (2006). Combining scenario planning and multi-criteria decision analysis in practice. *Journal of Multi-Criteria Decision Analysis, 14*, 5–20.

RAND Corporation. (2006). *RAND Project Air Force: Analysis of alternatives (AoA) for KC-135 Recapitalizatio*n (Executive summary). Santa Monica, CA: Author.

RAND Corporation. (2016). *Maritime Tactical Command and Control analysis of alternatives*. Santa Monica, CA: Author.

Saaty, T. L. (1980). *The analytic hierarchy process*. New York, NY: McGraw Hill.

Souders, R. M., Schulze, S., Ginburg, Y., & Goetke, J. (2004). *MPF(F) analysis of alternatives: Final summary report* (CNR D0009814.A2/Final). CNA.

Smith, M. W., Ritz, C., Smith, M., & Braun, C. (2011). *T-AO(X) analysis of alternatives final report* (CRM D0026011.A2/Final). CNA.

Stewart, T. J. (1997). Scenario analysis and multicriteria decision-making. In J. Climaco (Ed.), *Multicriteria analysis* (pp. 519–528). Berlin, Germany: Springer.

TRADOC. (2011). *Joint Light Tactical Vehicle analysis of alternatives final report* (TRAC-F-TR-11-025). TRADOC Analysis Center.

USAF, Office of Aerospace Studies. (2016, July 6). *Analysis of alternatives handbook*. Retrieved from http://afacpo.com/AQDocs/AoAHandbook.pdf

Rethinking Government Supplier Decisions: The Economic Evaluation of Alternatives (EEoA)

Francois Melese—Professor, Defense Resources Management Institute (DRMI), Graduate School of Business & Public Policy (GSBPP), Naval Postgraduate School, Monterey, CA [fmelese@nps.edu]

James Fan—Assistant Professor, Defense Resources Management Institute (DRMI), GSBPP, Naval Postgraduate School, Monterey, CA

"If we do not have a good economic model for supplier decisions, we are not on a level playing field. And we already spend [too] much … time on that uneven playing field."

—Colonel John T. Dillard, U.S. Army (Ret.),
Senior Lecturer, Naval Postgraduate School,
Past Program Manager for Advanced Acquisition Programs

Abstract

This paper offers an economic model to assist public procurement officials to rank competing vendors when benefits cannot be monetized. An important defense application is "source selection"—choosing the most cost-effective vendor to supply military equipment, facilities, services, or supplies. The problem of ranking public investment alternatives when benefits cannot be monetized has spawned an extensive literature that underpins widely applied decision tools. The bulk of the literature, and most government-mandated decision tools, focus on the demand side of a public procurement. The Economic Evaluation of Alternatives (EEoA) extends the analysis to the supply side. A unique feature of EEoA is to model vendor decisions in response to government funding projections. Given a parsimonious set of continuously differentiable evaluation criteria, EEoA provides a new tool to rank vendors. In other cases, it offers a valuable consistency check to guide government supplier decisions.

Introduction

Public procurement is big business. In 2017 the U.S. Department of Defense (DoD) spent over $300 billion on procurement, research, development, and test & evaluation, most of it sourced to the private sector (Schwartz et al., 2018). The Organization for Economic Cooperation and Development (OECD) reports member countries spend more than 12% of their cumulative GDP on public purchases (OECD, 2016). One of the biggest challenges for public procurement officials is to rank vendors when benefits cannot be monetized. Indeed, government benefits are often depicted as bundles of desirable characteristics, or attributes, that cannot easily be combined with costs into a single overall measure such as profitability. The problem of ranking public investment alternatives when benefits cannot be monetized has spawned an extensive literature generally referred to as "multi-criteria decision-making" (MCDM). A proliferation of applications of decision tools derived from this literature has appeared in management science, operations research, and decision sciences (prominent examples include Keeney & Raiffa, 1976; Dyer & Sarin, 1979; Kirkwood, 1995, 1997; Clemen, 1996; Che, 1993; Beil & Wein, 2003; and Parkes & Kalagnanam, 2005).

Today, widespread application of MCDM tools and techniques is mandated through various laws, rules, and regulations that govern public procurement. The main guide for federal procurement officials in the United States is the Federal Acquisition Regulation (FAR).[1]

> Evaluation criteria are the factors an agency uses to determine which of several competing proposals submitted in response to an RFP [Request for Proposal] would best meet the agency's needs. In establishing effective evaluation criteria, an agency must clearly identify the factors relevant to its selection of a vendor and then prioritize or weight the factors according to their importance in satisfying the agency's need in the procurement. … This allows the agency to rank the proposals received. (FAR, Proposal Development, Section M—Evaluation Factors for Award)

Similar source selection techniques are frequently applied in the United States at state and local levels, and in the private sector.

While demand side developments of MCDM models have been extensively studied in the academic literature, the literature is mostly silent about the supply side (vendor) problem. Vendor decisions (bid proposals) are generally treated as exogenous. In sharp contrast, the Economic Evaluation of Alternatives (EEoA) captures both the demand side (procurement official decisions) and the supply side (vendor optimization decisions).[2]

EEoA encourages public procurement officials to carefully consider the impact on vendor proposals of announced priorities—desired criteria, characteristics, or attributes for solicited quantities of products, services, or projects (e.g., computer systems, vehicles, weapon systems, logistics packages, and buildings). Officials should also consider the impact of anticipated future budgets. In response to government priorities—evaluation criteria, quantities, and funding—competing vendors, with different input costs and production functions, maximize their production offers, that is, bid proposals that consist of bundles of non-price characteristics or attributes.

EEoA models public procurement official decisions in two stages. In the first stage, along with the requirement (quantity demanded) and funding guidance, the procurement official reveals desired evaluation criteria (characteristics or attributes) of the product or service, but not the relative importance/weights. Given this information, competing vendors engage in constrained optimizations based on their respective production technologies and input costs to generate proposals that match anticipated future funding. Since input costs and production functions vary among vendors, they play a critical role in their bid proposals—interpreted as bundles of non-price characteristics or attributes embedded in each identical unit offered by a particular vendor. In the second stage, the procurement

---

[1] Note the exclusive focus on the demand side in the FAR, i.e., ranking exogenously-determined bids received from vendors (see https://www.acquisition.gov/browse/index/far). Also note the standard practice for U.S. military (and other procurement officials) is to (i) announce factors ("evaluation criteria") relevant to the selection, but only after receiving vendor proposals, and (ii) assign specific relative importance/weights to those factors to rank vendors. This practice is modeled in the Economic Evaluation of Alternatives (EEoA).

[2] The EEoA model is loosely based on Lancaster's "Characteristics Approach to Demand Theory" (Lancaster, 1966a, 1966b, 1971, 1979), as modified by Ratchford (1979). Applications of the model appear in Simon & Melese (2011), and in Chapter 4 of Melese, Richter, & Solomon (2015).

official ranks competing vendors according to the government's utility function over the evaluation criteria (see Figure 1).[3]

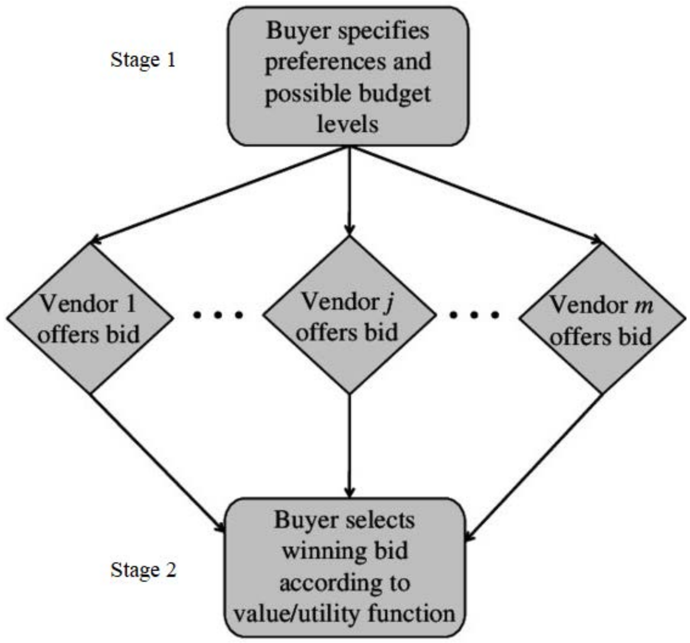


Figure 1. **The Two-Stage Procurement Process**

The dual objective of EEoA is to encourage governments (i) to consider the supply side by recognizing the importance of modeling vendor responses to information provided or inferred in public procurements and (ii) to offer an alternative to the standard MCDM approach when benefits cannot be monetized. An attractive feature of EEoA is that it offers a novel technique to measure benefits that serves as a valuable consistency check for MCDM preference trade-offs among key attributes.[4] We explore assumptions under which the two decision models (MCDM and EEoA) are isomorphic from a procurement official's perspective. In practice, however, we demonstrate how EEoA can yield significantly different solutions (rank orderings) than the standard MCDM approach.

The paper proceeds as follows. The first section develops the two-stage EEoA model. On the supply side, two cases are presented to illustrate the model (i) where vendors

[3] Note this is analogous to steps mandated in the FAR, except that, since funding is fixed in EEoA (i.e. the price is the same for each vendor), the second step involves the submission by vendors of sealed non-price bids for the announced level of funding, interpreted and evaluated by procurement officials as bundles of characteristics, attributes, etc. that respond to previously announced evaluation criteria (e.g., see FAR 14.5)
[4] Both Australian and Canadian Ministries of Defense are considering implementing this consistency check for the MCDM component of their portfolio decision models (Personal correspondence with fellow NATO SAS-134 Defence Official Panel Members studying Defence Portfolio Management for NATO; emails received 11/2018).

have identical attribute costs, but different production technologies and (ii) where vendors have different attribute costs, but identical production technologies. A simple example serves to integrate procurement official (demand) considerations, with vendor (supply) decisions, under varying (probabilistic) scenarios. The following section contrasts an application of EEoA, with a standard textbook application of MCDM. The final section concludes with recommendations for future research.

The Economic Evaluation of Alternatives (EEoA) Model

The challenge for our public procurement official is to select a competing vendor that delivers the best combination of desired non-price attributes for each identical unit of a requirement (e.g., 50 computers, or 20 drones, or 2 hospital ships, etc.) at affordable funding levels. The EEoA framework can be thought of as a multi-attribute sealed bid procurement auction that extends traditional price-only auctions to one in which competition among $j \in [1, m]$ vendors (bidders) takes place exclusively over bundles of $i \in [1, n]$ non-price characteristics or attributes (a_{ij}).

The EEoA model structures the problem as a two-stage optimization (see Figure 1). In the first stage, the public procurement official provides j competing vendors with the evaluation criteria, available funding, and the requirement (quantity demanded).[5] Given the anticipated budget, **B**, and their respective production technologies and input costs, competing vendors offer their best possible non-price attribute packages bundled into each identical unit required.[6] Note that the greater the funding available, the greater the available funding per unit, which allows vendors to bundle more of the desired attributes in each identical unit (e.g., better computers, drones, ships, etc.).[7]

The vendor (supply side) problem is formulated in the section titled First Stage EEoA: The Vendor's Problem. Competition takes place exclusively over non-price bid proposals from each vendor, evaluated by procurement officials as bundles of attributes offered by each vendor for a standard unit of the requirement. Whereas bundles of attributes for each unit of the requirement are identical for each vendor, they differ among vendors. Individual vendor bid proposals depend on their costs to generate each attribute, their production technology to combine those attributes, and anticipated future funding.

---

[5] Since there is a fixed requirement (quantity demanded), the budget, **B**, can be interpreted as the unit funding/budget available to vendors to produce a unit of the required product or service. For example, if we anticipate $25,000 of funding is available for 50 computers, the budget (**B**) used by competing vendors to build their proposals would be $500 per unit.

[6] For example, suppose we have $25,000 of funding for 50 computers, or a budget, **B**=$500/unit. Then, for example, each of 50 identical Apple laptop computers offered at $500/unit would satisfy the basic evaluation criteria (screen size, memory, battery life, software, etc.), but consist of a somewhat different bundle of those characteristics/attributes, than each of 50 identical Microsoft (or Dell, or HP, etc.) laptop computers.

[7] The greater the funding available, the greater the funding per unit, allowing vendors to offer more of the desired attributes for each identical unit demanded by the buyer. For example, suppose for our 50 computers, instead of $25,000 (**B**=$500) of funding, it turns out $50,000 (**B**=$1000) will be available. Then each of the 50 identical laptop computers offered by Apple will have more and/or better characteristics/attributes, and so will each of the 50 identical laptop computers offered by Microsoft (bigger screen size, more memory, longer battery life, etc.).

In the second stage, the procurement official's objective is to select the vendor j that maximizes the government's utility function, $U_j = U_j(a_{1j}, a_{2j,} \ldots, a_{nj})$, subject to projected funding (i.e., the per unit affordability or budget constraint), B. For analytic tractability we assume the utility function is quasi-concave, and that attributes are continuous, non-negative, monotonic increasing variables, that is, the domain of the buyer's utility function, and sellers' production functions and attribute cost functions, are the nonnegative real numbers. Non-satiation in the relevant range of attributes is also assumed, such that, $\partial U_j / \partial a_{ij}$>0, or the greater the score of the $i \in [1, n]$ desired attributes, a_{ij}, the more value (utility/benefit) for the buyer, but the more costly it is for sellers to produce.

Following the literature, we allow the buyer's utility function (scoring/ranking rule) to be linear, additive, and separable across attributes (see Keeney & Raiffa, 1976; Kirkword, 1997, etc.). The public procurement official's problem is to select a vendor $j \in [1, m]$ that maximizes the government's utility function:

$$(1)\ \boldsymbol{U_j} = \mathrm{U_j}\left(\boldsymbol{A}_j^T\right) = \boldsymbol{W}\boldsymbol{A}_j^{\mathrm{T}},$$

where desired attributes are known to sellers, and the bundle of attributes in vector $\boldsymbol{A_j}$ = $[a_{1j}\ a_{2j} \ldots a_{nj}]$ represents each vendor's offer (bid proposal) for each unit required. The relative weights for each attribute are the procurement official's private information, given by the vector:

$$\boldsymbol{W} = (w_1, w_2, w_3, \ldots, w_n \mid w_i \in \mathbb{R}^+, i \in [1, n]).$$

The procurement official maximizes (1) subject to a funding/affordability constraint:

$$(2)\ TC_j \leq \mathbf{B},$$

such that the total unit cost (price) of any vendor's bid proposal, TC_j, must fit within forecasted future funding, i.e., the per unit budget, **B.** Note that whereas the set of non-price attributes in the buyer's utility function are revealed to the $j \in [1, m]$ competing vendors, the ***relative*** (preference or "trade-off") ***weights***, w_i, are not.[8] This reflects practical application of the FAR:

> In government acquisition, procuring commands have their own best practices and priorities … but they all follow the [Federal Acquisition

[8] For example, consider the following summary of Federal Acquisitions Regulations (FAR) 15.1 and 15.3: "Evaluating proposals under the RFP [Request for Proposal] best value trade-off analysis criteria": In a negotiated bid there are factors [evaluation criteria] with varying weights assigned. The solicitation tells you the weight of each factor. However, government contracting agencies are not required to publicize the actual source selection plan [it is an internal document]. The agency has broad discretion on what it believes to be the best value. Note, however, the agency must be consistent in following their source selection plan in evaluating every vendor, or risk bid protests—e.g., see Melese (2018).

> Regulation]. And in their selection of suppliers, they assign weights to their parameter criteria in accord with their priorities. … These weights for scoring of proposals do not have to be specifically revealed as an algorithm, but are typically communicated to offerors in terms of [rank ordering of] importance. (Colonel John T. Dillard, U.S. Army (Ret.), Senior Lecturer, Naval Postgraduate School, Past Program Manager for Advanced Acquisition Programs)

In this formulation of the procurement problem, both buyer and seller suffer from imperfect and asymmetric information. While the seller does not know the specific relative importance/weights assigned to desired attributes (or "evaluation criteria"), the buyer (procurement official) does not know the vendors' costs of producing a particular attribute, nor the technology (production functions) that combines those attributes into vendor proposals.[9] The supply side vendor problem is examined in detail in the next section, followed by the demand side procurement problem.

First Stage EEoA: The Vendor's Problem (Supply Side)

The first stage of the two-stage EEoA optimization framework focuses on the vendor's problem. The economic approach assumes vendors are strategic players, so that the anticipated/forecasted (per unit) funding/budget, **B**, for the procurement, impacts vendors' formulation of their competing bid proposals (attribute bundles, $\boldsymbol{A_j}$).[10]

Given n desired attributes (a_{ij}) and anticipated future funding (the per unit budget, **B**), the m vendors each offer competing bid proposals (bundles of attributes), $\boldsymbol{A_j}$, based on their production technology, and their unit costs of producing each attribute, $c_{ij}(\boldsymbol{B})$.[11] For any fixed requirement (quantity demanded) and funding level (per unit budget, **B**), a representative vendor's problem is to maximize their attribute output function for each (identical) unit required, subject to the vendor's costs of producing each attribute. Competing vendors offer their best possible non-price attribute bundle for the projected per unit funding/budget, **B,** given their idiosyncratic technology. As Wise & Morrison (2000) observe, a multi-attribute auction allows competing vendors to differentiate themselves in the auction process and bid on their competitive advantages.

The vendor's problem can be expressed as selecting an attribute vector (bid proposal), $\boldsymbol{A_j}$ = $[a_{1j}, a_{2j}, \ldots, a_{nj}]$ that maximizes output:

$$(3)\ \boldsymbol{Q_j} = Q_j(\boldsymbol{A}_j^T),$$

subject to unit costs (TC) not exceeding anticipated per unit funding (**B**) for the project,

---

[9] "Seller costs can be expected to depend on [the] local manufacturing base, and sellers can be expected to be well informed about the cost of (upstream) raw materials" (Parkes & Kalagnanam, 2005, p. 437).

[10] Further implications are explored in Section 2c. Note the supply-side development in this section generalizes a special case of the multi-attribute auction found in Simon & Melese (2011).

[11] Each vendor's bundle is a technologically-determined combination of attributes. For instance, a computer is a combination of screen size, memory, battery life, etc., with unit costs associated with each attribute.

$$(4)\ \boldsymbol{TC_j} = \sum_{i=1}^{n} c_{ij}\,(B)\,a_{ij} \leq \mathbf{B}.$$

For ease of exposition, the remainder of the study focuses on two vendors and two (non-price) attributes.

Suppose each vendor has a different technology to combine the two attributes, and different attribute costs, then the Lagrangian function for the vendor's problem is given by:

$$(5)\ \mathcal{L}_j = Q_j(a_{1j}\,, a_{2j}, \boldsymbol{B}) + \lambda_j[\boldsymbol{B} - \textstyle\sum_{i=1}^{2} c_{ij}\,(\boldsymbol{B})\,a_{ij}], \text{ for } j=1,2.$$

If vendors compete on "quality," they are likely to use the maximum expected per unit funding, **B**, to develop their bid proposals, so (4) is an equality. So first order necessary conditions for an optimum are given by:

$$(5a)\ \partial\mathcal{L}_j/\partial a_{1j} = \partial Q_j/\partial a_{1j} - \lambda_j c_{1j}(\boldsymbol{B}) = 0,$$

$$(5b)\ \partial\mathcal{L}_j/\partial a_{2j} = \partial Q_j/\partial a_{2j} - \lambda_j c_{2j}(\boldsymbol{B}) = 0,$$

$$(5c)\ \partial\mathcal{L}_j/\partial \lambda_j = \boldsymbol{B} - \textstyle\sum_{i=1}^{2} c_{ij}\,(\boldsymbol{B})\,a_{ij} = 0.$$

Solving Equations 5a–5c yields optimal attribute bid proposals (outputs) for each vendor $j = 1{,}2$, for each identical unit required, for any given per unit budget, ***B***:

$$(6a)\ a_{1j}^{*} = a_{1j}^{*}(\alpha_{1j}(\boldsymbol{B}), \alpha_{2j}(\boldsymbol{B}), c_{1j}(\boldsymbol{B}), \boldsymbol{B}),$$

$$(6b)\ a_{2j}^{*} = a_{2j}^{*}(\alpha_{1j}(\boldsymbol{B}), \alpha_{2j}(\boldsymbol{B}), c_{2j}(\boldsymbol{B}), \boldsymbol{B}).$$

For tractability, we assume a standard Cobb-Douglas production function, with attributes $(a_{1j}\,, a_{2j})$ as inputs:

$$(6)\ Q_j\left(a_{1j}\,, a_{2j}\right) = a_{1j}^{\alpha_{1j}} a_{2j}^{\alpha_{2j}}.$$

Two special cases help illustrate the model: (i) where vendors share common attribute costs, but have different production technologies, and (ii) where vendors share the same production technology, but have different attribute costs.

Vendors With Different Production Technologies

In the first case (illustrated in Figure 2), vendors $j = 1{,}2$ have different, constant (i.e. independent of funding) technologies (i.e., in Equation 6: $\alpha_{1j}(\boldsymbol{B}) = \alpha_{1j}$ and $\alpha_{2j}(\boldsymbol{B}) = \alpha_{2j}$), but identical (constant) attribute costs (i.e., $c_{1j}(\boldsymbol{B}) = c_1$ and $c_{2j}(\boldsymbol{B}) = c_2$). From the first order necessary conditions for an optimum ((5a) – (5c)), and (6), competing vendors' optimal attribute bundle bid proposals, for the expected per unit funding/budget level **B**, are given by:

$$(6a')\ a_{1j}^{*} = [\alpha_{1j}/(\alpha_{1j} + \alpha_{2j})\ c_1]\ \mathbf{B}, \text{ and}$$

$$(6b')\ a_{2j}^{*} = [\alpha_{2j}/(\alpha_{1j} + \alpha_{2j})\ c_2]\ \mathbf{B}.$$

Figure 2 illustrates optimal attribute bundle bid proposals for each vendor for a specific unit funding/budget level, **B**: $A_1 = (a_{11}^*, a_{21}^*)$ and $A_2 = (a_{12}^*, a_{22}^*)$. The optimum for each vendor is determined graphically by the tangency of each vendor's isoquant (derived from their separate production functions), with the common budget constraint.

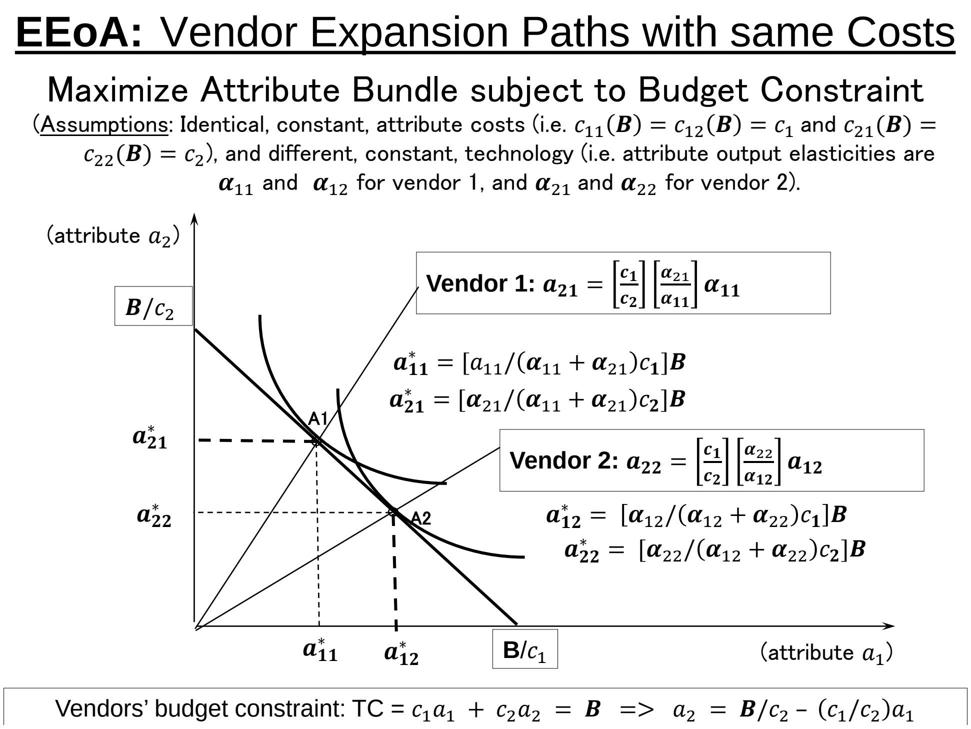


Figure 2. **Common Attribute Costs but Different Technologies**

Suppose instead of a single funding forecast, the buyer (procurement official) reveals a range of possible budget estimates for the procurement (say optimistic, pessimistic, and most likely).[12] Then Equations 6a' and 6b' can be combined to yield each vendor's expansion path, given by:

$$(7)\ a_{2j} = [(c_{1j}(B)/c_{2j}(B))\,(\alpha_{2j}(B)/\alpha_{1j}(B))]\ a_{1j}\ \text{, for } j = 1{,}2.$$

The two expansion paths defined by Equation 7 reveal optimal attribute bundles offered by each vendor at different possible funding levels, *B*. Each point on the expansion paths derived for each vendor reveals optimal attribute bundle offers (bid proposals) for each identical unit required, over different possible budgets.

Given this formulation, if attribute costs and technology parameters are constant (i.e., independent of funding levels), then the expansion paths are linear. Expansion paths for the first case, where vendors' share common costs but different technologies, are given by:

---

[12] For example, see Simon & Melese (2011).

(7a) $a_{21} = [c_1/c_2][\alpha_{21}/\alpha_{11}]\, a_{11}$, for vendor 1, and

(7b) $a_{22} = [c_1/c_2][\alpha_{22}/\alpha_{12}]\, a_{12}$, for vendor 2,

illustrated as two straight lines from the origin in Figure 2. For the specific per unit budget level, **B**, the two competing attribute bundle bid proposals offered by each vendor (from (6a') and (6b')) appear as points $A_1 = (a_{11}^*, a_{21}^*)$ and $A_2 = (a_{12}^*, a_{22}^*)$ on the competing vendors' expansion paths.

Vendors With Different Attribute Costs

Turning to the second example (illustrated in Figure 3), suppose vendors have different (constant) attribute costs, but identical (constant) production technologies (i.e., in Equation 6: $\alpha_{1j}(\mathbf{B}) = \alpha_1$ and $\alpha_{2j}(\boldsymbol{B}) = \alpha_2$ for j=1,2), together with constant returns to scale (such that: $\alpha_1 + \alpha_2 = 1;\ i.e. if\ \alpha_1 = \alpha\ then\ \alpha_2 = 1 - \alpha$). In this case the two vendor's optimal bid proposals for unit funding/budget level, **B**, are given by:

(6a'') a_{1j}^*= $[\alpha/c_{1j}]$ **B**, and

(6b'') a_{2j}^*= $[(1-\alpha)/c_{2j}]$ **B**, (j=1,2).

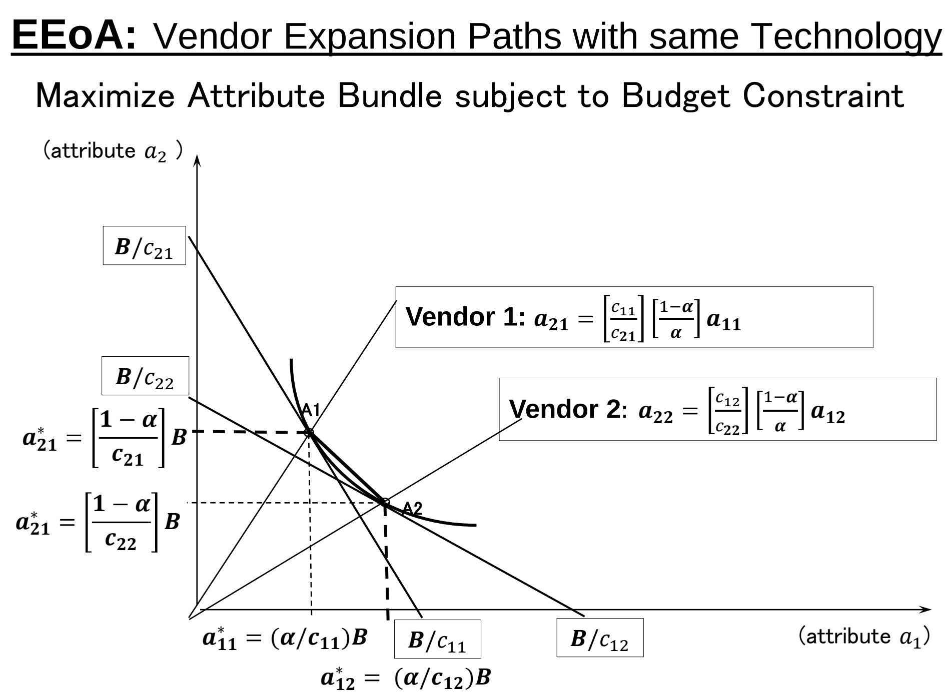


Figure 3. **Common Technology but Different Attribute Costs**

Similar to the first case, Figure 3 illustrates competing optimal attribute bundle bid proposals for each vendor, for the unit funding/budget level, **B**: $A_1 = (a_{11}^*, a_{21}^*)$ and $A_2 = (a_{12}^*, a_{22}^*)$. Now the optimum for each vendor occurs at the point where their respective budget constraints are tangent to their common isoquant. If vendors' technology and attribute cost parameters are constant (i.e., independent of funding levels), both expansion paths are again linear. Expansion paths for this second case (where vendors share a common technology but have different attribute costs) are illustrated as two straight lines from the origin in Figure 3, given by:

(7a') $a_{21} = [c_{11}/c_{21}][(1-\alpha)/\alpha]\, a_{11}$, for vendor 1, and

(7b') $a_{22} = [c_{12}/c_{22}][(1-\alpha)/\alpha]\, a_{12}$, for vendor 2.

Focusing on this second case (where vendors share a common technology but have different attribute costs) for any unit funding/budget level, **B**, connecting the two optimal vendor attribute production points (A_1 and A_2) creates an attribute "*production possibility frontier*" (PPF), illustrated in Figure 3. The slope of this PPF reflects attribute trade-offs possible in the marketplace by switching from one vendor to another. This technical (or engineering) trade-off is given by the slope: $\Delta a_2/\Delta a_1 = (a_{21}^* - a_{22}^*)/(a_{11}^* - a_{12}^*)$.

The first stage vendor optimization problem in the two-stage EEoA framework highlights the importance of modeling the supply side—considering vendor decisions in response to anticipated future funding. The second stage focuses on the demand side—the procurement official's source selection problem.[13]

Second Stage EEoA: Procurement Official's Problem (Demand Side)

For any given requirement (quantity demanded) and forecasted per unit funding/budget, **B**, the procurement official (decision-maker) must rank vendors' optimum bid proposals. For example, in Figure 3: Vendor 1=>(a_{11}^*, a_{21}^*) and Vendor 2=>(a_{12}^*, a_{22}^*). Recall the lens through which the government evaluates competing vendors is the utility function given by Equation 1.[14] In EEoA, the government supplier decision ("source selection") depends on the public procurement official's (decision-maker's) preferences revealed through explicit trade-offs for any pair of attributes that leave decision-makers indifferent in any given scenario. These explicit pair-wise comparisons elicited from a public procurement official (or expert decision-makers) generate relative weights assigned to the desired attributes.

The public procurement official's problem is to select a vendor $j \in [1, m]$ with bid proposal (per unit attribute bundle) $\boldsymbol{A}_j = [a_{1j}, a_{2j}, \ldots, a_{nj}]$) that maximizes the government's utility function given by Equation 1. Recall, following the standard assumption in the literature (see Keeney & Raiffa (1976), Kirkword (1997), etc.), the utility/benefit provided by any vendor j is given by the linear, separable utility function:

$$(1')\ \boldsymbol{U_j} = \mathrm{U_j}\left(\boldsymbol{A}_j^T\right) = \boldsymbol{W}\boldsymbol{A}_j^{\mathrm{T}} = \sum_{i=1}^{n} w_i\, a_{ij},$$

---

[13] Note that this second stage demand-side problem is the exclusive focus of most textbooks, the majority of the literature, and standard support tools and algorithms.
[14] An interesting extension of Equation 1 is developed later to address uncertainty when different possible scenarios (states of nature) impact the government's utility function (for example, due to possible future changes in the political, economic, or threat environment).

where the vector $\boldsymbol{A_j} = [a_{1j}\ a_{2j} \dots a_{nj}]$ represents the bundle of attributes for each unit, offered by each of the $j \in [1, m]$ competing vendors. As discussed earlier, specific relative trade-off weights for every attribute are the procurement official's private information, given by the vector:

$$\boldsymbol{W} = (w_1, w_2, w_3, \dots, w_n \mid w_i \in \mathbb{R}^+, i \in [1, n]).$$

The procurement official is also fiscally informed, with a forecasted funding/budget (affordability) constraint for the procurement given by Equation 2. So the per unit price (total unit costs) of any vendor proposal, TC_j, must fit within forecasted future funding (the anticipated per unit budget, $\mathbf{B}$), or $TC_j \leq \mathbf{B}$. The next step is to combine Demand and Supply (i.e., the procurement official's source selection problem), with vendors' (optimization-generated) bid proposals. The following simple source selection example demonstrates how EEoA integrates demand and supply.

Demand & Supply: A Two Scenario, Two Vendor, Two Attribute Example

For purposes of illustration, suppose a public procurement official responsible for UN peacekeeping missions is asked to select a vendor for a new fleet of Autonomous Electric Off-road Light Armored Transport Vehicle (AEOLATV). Assume the anticipated (per unit) budget, **B**, for the program allows two competing vendors to offer the required set of vehicles, and that there are only two evaluation criteria in the government's utility function: ***Top Speed*** of each vehicle measured in miles per hour (a_1), and ***Range*** measured in miles (a_2).[15] In Figure 3, this involves a choice between Vendor 1 that offers less speed but more range (a^*_{11}, a^*_{21}), and Vendor 2 that offers more speed but less range (a^*_{12}, a^*_{22}).

In EEoA, the source selection decision (vendor ranking) depends on the procurement official's (decision-maker's) preferences revealed through pair-wise comparisons, that is, explicit acceptable trade-offs between pairs of attributes within a particular scenario. This generates relative weights assigned to the desired attributes within a particular scenario.

A straightforward modification of (1') allows us to extend the analysis to address different possible scenarios (states of nature) that could impact the procurement official's pair-wise comparisons.[16] Equation 8 accounts for k possible scenarios (or "states of nature"), Ns, $\forall s \in$ [1,k], with corresponding probabilities, $P(Ns)$. This linear, separable ***expected*** utility function captures the differing relative weights, derived from explicit preference trade-offs among pairs of attributes that depend on specific scenarios (states of nature). Now the procurement official's problem is to select the vendor (e.g., bidder or investment alternative), $j \in$ [1,m], that maximizes the government's ***expected*** utility given by:

$$(8)\ \mathbf{E}(\boldsymbol{U_j}) = \sum_{s=1}^{k} P(Ns) \sum_{i=1}^{n} w_{is}\, a_{ij}.$$

[15] For example, we could assume all other characteristics (or attributes) of the vehicles offered by the vendors are the same, so top speed and range are the only differentiating factors.
[16] For example, different possible threat environments in which the UN might operate.

Consider a simple case with two possible states of nature N1 & N2, (e.g. Scenario s=1 a High Tech Threat environment, vs. Scenario s=2 a Low Tech Threat Environment), with corresponding probabilities, P(N1) and P(N2).[17] From Equation 8, the government's expected utility function (scoring rule) for the two scenario, two attribute case is:

$$(8')\ \mathbf{E}(\boldsymbol{U_j}) = P(N1)[w_{11}a_{1j}+w_{12}a_{2j}] + P(N2)[w_{21}a_{1j}+w_{22}a_{2j}].$$

Totally differentiating the procurement official's (government's) utility function (8') and setting the result equal to zero in each scenario (N1 & N2), generates two sets of relative weights (or indifference curves). In general, relative weights for any two pairs of attributes (a_1,a_2) in each of the k scenarios in (8) are given by:

$$(9)\ \partial a_2/\partial a_1 = -(w_{1s}/w_{2s}) = -X_s,\ \forall s \in [1,k].$$

The last term in Equation 9, X_s>0, represents the acceptable trade-off determined by a decision-maker (procurement official) between any pair of attributes (a_1,a_2) for a specific scenario: $w_{1s} = (w_{2s})$x(X_s). It reflects acceptable pair-wise trade-offs for the government over the relevant range of attributes in each scenario. These preference trade-offs define linear indifference curves between any two pairs of attributes in each scenario (or piecewise linear approximations over specific ranges of attributes). The slopes of these indifference curves are the relative weights for each pair of attributes, in each state of nature, over relevant ranges of each attribute.

Optimal vendor rankings in EEoA can be determined by comparing the slope of the government's (buyer's) revealed preferences (indifference curves), with the competing vendor-proposed bundles of attributes (production possibility frontiers). For example, Figure 4 illustrates two different sets of indifference curves (dashed lines) that reflect two different scenarios. In turn, these yield two different vendor rankings.

For a given per unit budget, **B**, if the slope of the indifference curve is steeper than the slope of the production possibility frontier (where the PPF reflects technical trade-offs available between competing vendors), or if from (9), -X = $-(w_1/w_2) < -(a^*_{21} - a^*_{22})/(a^*_{11} - a^*_{12})$, then vendor 2 is selected, since $U^*_2 > U_1$. If the reverse is true, then vendor 1 wins, since $U^*_1 > U_2$ (see Figure 4).

Suppose a government decision-maker is willing to trade off relatively more range (a_2) for the same incremental increase in top speed (a_1) in Scenario N1, than in Scenario N2. For example: 20 miles of range for an extra 10 mph top speed in N_1, versus only 10 miles for an extra 10 mph in N_2. In this case, $-X_1 = -2 < -X_2 = -1$, implies the slope of the indifference curve is steeper (more negative) in Scenario N_1 than in N_2.[18] From Figure 4, vendor 2 is ranked higher (offers greater utility) in scenario N1, and vendor 1 in scenario N2. This is consistent since the decision-maker revealed a stronger relative preference for top

[17] In the AEOLATV example, scenario N1 could represent the possibility of facing a fast adversary with limited range with probability P(N1), and scenario N2 a slower adversary with greater range with probability P(N2); where P(N1)+P(N2)=1.

[18] In this case, under scenario N1 vendor 2 ranks higher (offers greater utility) than vendor 1, and there is a rank reversal under scenario N2.

speed in scenario N1 (i.e., was willing to trade-off more range), and vendor 2 offers relatively higher top speed (a_{12}^*) than vendor 1 (a_{11}^*).

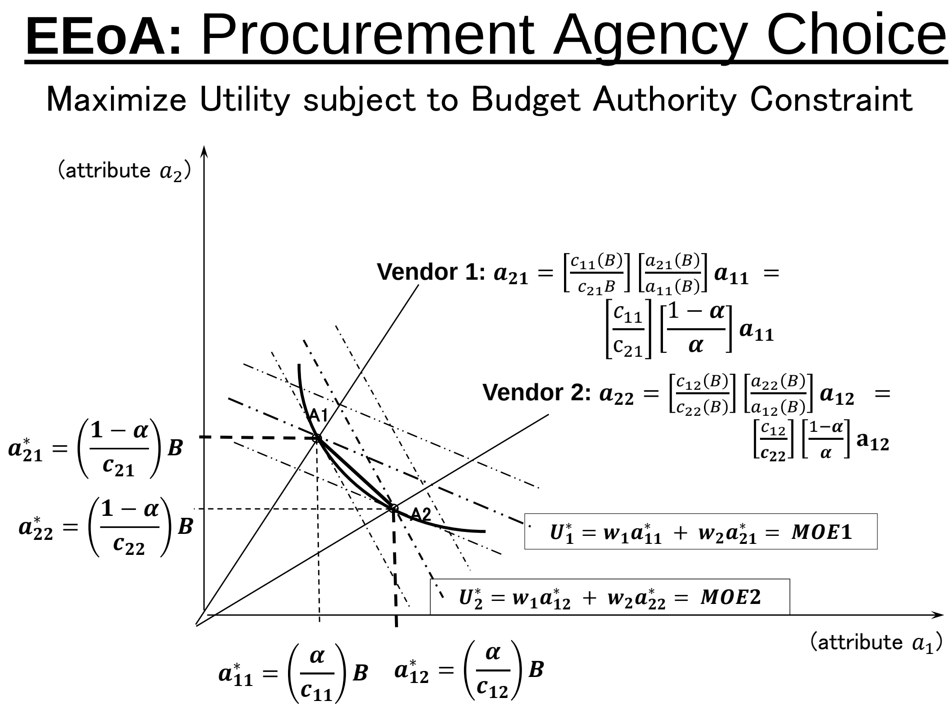


Figure 4. **Procurement Agency Vendor Selection**

In general, probabilities assigned to each scenario in Equations 8 or 8' generate an ***Expected Utility*** vendor ranking metric that consists of a probability-weighted average of pair-wise attribute trade-offs (-Xs) that define expected utility functions in each of the $s \in$ [1,k] scenarios. For example, in the two scenario, two vendors, two attribute case, this determines the slope of a new indifference curve that is a combination of the two indifference mappings illustrated in Figure 4. For any specified budget, the tangency (or corner point) of this new indifference curve with the PPF reveals the optimal Expected Utility ranking of the two vendors. The next section contrasts this Economic Evaluation of Alternatives (EEoA), with the standard textbook MCDM model commonly applied by public procurement officials to guide government supplier decisions.

Comparison of EEoA and MCDM Models

The topic of multi-criteria decision-making (MCDM) has spawned a rich literature with many variations to account for decision-making in complex scenarios. This section uses a standard textbook MCDM model frequently applied to guide government supplier decisions as a baseline (see Keeney & Raiffa, 1976; Kirkwood, 1997, etc.). We contrast this MCDM model with the EEoA approach within a single scenario. The MCDM additive value function typically used to rank vendors is given by:

$$(10)\ \boldsymbol{V_j} = \mathrm{V_j}\left(\boldsymbol{A}_j^T\right) = \boldsymbol{\lambda} v_i(a_{ij}) = \sum_{i=1}^{n} \lambda_i\, v_i(a_{ij}).$$

This value function is the sum of individual value functions, $v_i(a_{ij})$, defined over relevant ranges of each attribute $i \in [1, n]$, for any vendor j. The vector of preference weights is given by:

$$\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3 \ldots, \lambda_n \mid \lambda_i \in \mathbb{R}^+, i \in [1, n]).$$

The individual value functions $v_i(a_{ij})$ are typically monotonic and scaled (normalized), while the preference weights (λ_i) reflect the importance of each attribute. While these weights ($\boldsymbol{\lambda}$) are analogous to the relative weights ($\boldsymbol{W}$) in EEoA, they are only equivalent if raw attribute measures are used in MCDM instead of normalized values to determine pair-wise trade-offs (i.e., if $v_i(a_{ij}) = a_{ij}$). For purposes of comparison with EEoA, it is convenient to assume procurement officials (decision makers) are subject to the same funding/affordability constraint given by (2): $TC_j \leq \mathbf{B}$. Implications of this MCDM model are explored below under the usual assumption that attribute measures are normalized using individual value functions with preferential independence.

Implicit Trade-Offs in MCDM vs. Explicit Trade-Offs in EEoA

From Equation 10, the only theoretical difference between the procurement official's objective function (1) or (1') in EEoA, and MCDM is an additional step in Equation 10 that involves normalizing attribute measures through individual value functions. In fact, the demand side of EEoA can be thought of as a special case of MCDM, where $v_i(a_{ij}) = a_{ij}$.

In theory, any value function, v_i, in conjunction with the appropriate attribute weights λ_i, can recover the EEoA utility function for any given vector of attributes $\boldsymbol{A_j}$.This is clear when we consider a procurement official's value function with two attributes as before:

$$(10')\ \boldsymbol{V_j} = \sum_{i=1}^{n} \lambda_i v_i(a_{ij}) \Rightarrow [\lambda_1 v_1(a_{1j}) + \lambda_2 v_2(a_{2j})].$$

Totally differentiating (10) or (10') and setting the result equal to zero yields ***implicit*** trade-offs in the MCDM approach between any two pairs of attributes (a_1,a_2), that is, the first two terms in Equation 11 shown here. For sake of consistency given a particular decision-maker's preferences, this should precisely correspond to the explicit trade-offs (revealed preferences) obtained from that decision maker in EEoA, i.e., represented by the last two terms in Equation 9.

$$(11)\ \partial a_2/\partial a_1 = -[\lambda_1 v_1{}'(a_1)]/[\lambda_2 v_2{}'(a_2)] = -\frac{w_1}{w_2} = -X_s.$$

While the MCDM approach adds a degree of freedom for procurement officials and expands the decision space, it risks obscuring explicit trade-offs between attributes revealed in the EEoA approach. From Equation 11, we see that:

$$\lambda_1 / \lambda_2 = X_s[v_2{}'(a_2)/v_1{}'(a_1)], \text{ or}$$

$$Z = [v_2{}'(a_2)/v_1{}'(a_1)],$$

where the constant Z= $\lambda_1 / (\lambda_2 X_s)$. So in general, for any pair of attributes, and alternatives (i.e., vendors $j \in$ [1,m]),

$$(12)\ Z v_1'(a_{1j}) = v_2{}'(a_{2j}).$$

Integrating both sides of Equation 12 yields:

$$(13)\ v_2(a_{2j})/v_1(a_{ij}) = Z = \lambda_1/(\lambda_2 X_s).$$

That is to say, if the goal is to ensure EEoA and MCDM approaches generate the same rank ordering, procurement officials must set individual attribute value functions v_i's and attribute weights λ_i's in the precise ratio specified in Equation 13.

In practice, there is no reason to assume this happens, and reconciling the two approaches to generate the same rank ordering is non-trivial. While a procurement official may have a certain trade-off in mind between pairs of measurable attributes when developing the MCDM value function, normalizing each attribute with individual value functions, and selecting appropriate weights to assign to those value functions, can easily yield *implicit* pairwise trade-offs among attributes that generate different rank orderings than the *explicit* pairwise trade-offs determined in EEoA.[19] Which decision support model best elicits public procurement officials' (decision-makers') preferences remains an important empirical question and warrants further research.

From a practical standpoint, a limitation of the EEoA approach is that as the number of attributes (n) under consideration expands, it is increasingly burdensome to generate required pairwise comparisons. For example, assuming each alternative (vendor proposal) includes a set of n attributes, applying EEoA requires $\frac{n(n-1)}{2}$ pairwise comparisons. Interestingly however, EEoA could be applied in combination with MCDM as a consistency check for important attributes. That is to say, if $\partial a_2/\partial a_1 = -(w_{1s}/w_{2s}) = -X_s$ is the explicitly determined trade-off (indifference) that a public procurement official (decision-maker) is comfortable with in a particular scenario (for specific ranges of attribute measures) in EEoA, then weights developed in MCDM should reflect this relative preference (trade-off).[20] The test simply involves application of Equation 11 shown previously. We now turn to another important contribution of EEoA: the importance of modeling the supply side—specifically, accounting for vendor responses to anticipated future funding.

Accounting for Vendor Responses to Anticipated Future Funding

Traditionally, MCDM models focus on the demand side of a public procurement and treat supply side vendor decisions as exogenous. This section demonstrates the potential value of explicitly accounting for vendor responses to anticipated future funding (affordability or budget constraints).

Since each vendor's expansion path represents their optimal attribute bundle bid proposals for any given budget (see Figures 2, 3, and 4), these expansion paths can easily be converted, through the buyer's utility function (1'), into cost-effectiveness (or Budget-Utility) ***functions*** for each vendor. For example, substituting each vendor's optimal attribute bundle (6a'') & (6b'') into Equation 1' for any specific scenario yields two points in cost-effectiveness space that represent the utility of each vendor's bid proposal for the per unit

---

[19] Note: Linear normalization combined with careful swing weighting in MCDM could recover similar trade-offs to those explicitly revealed in EEOA (see Equation 9), resulting in an identical rank ordering of competing vendors. An example is available upon request.

[20] If the extra burden of normalization and swing weighting required in MCDM causes a decision-maker to "misevaluate" their trade-off preferences, then EEoA offers an alternative framework/perspective that can help to realign their weighting. Note that in theory a rational decision-maker with perfect information and infinite computational capability would never need to do this. Since in practice it is difficult to define a "correct" weighting, contrasting the development of weights in MCDM and EEoA may be an empirical question worth investigating.

funding/budget, **B**: $(\mathrm{U}_1^*, \mathbf{B})$ and $(\mathrm{U}_2^*, \mathbf{B})$. Different budgets represented along the expansion paths generate different utility. For example, the cost-effectiveness/utility relationships illustrated in Figure 6 reflect the value to the government of each vendor's offers at different funding levels.

There is an important contrast between endogenously derived EEoA cost-effectiveness functions for each vendor, and the exogenous cost-effectiveness points often illustrated in MCDM to represent vendor offers.[21] This becomes especially apparent when vendor costs depend on anticipated future funding. For instance, with bigger budgets, a vendor's costs to provide more of a particular attribute (say computer memory) might enjoy increasing returns to scale because of quantity discounts, the ability to employ just-in-time inventory techniques, or the possibility of adopting other process improvements that reduce a vendor's costs of incorporating/producing a desired attribute.

Consider the case illustrated in Figure 5, where vendor 1's costs of producing attribute 1 are assumed to depend on the funding level or anticipated per unit budget, **B** (i.e. $c_{11}(\boldsymbol{B})$). For ease of exposition, suppose both vendors $j = 1{,}2$ have identical, constant production technologies (i.e. $\alpha_{1j}(\mathbf{B}) = \alpha_1 \text{ and } \alpha_{2j}(\boldsymbol{B}) = \alpha_2$), and constant returns to scale $\alpha_1 + \alpha_2 = 1$. The difference between them is in their individual attribute costs. As before, let $c_{12}(\boldsymbol{B}) = c_{12}$; $c_{22}(\boldsymbol{B}) = c_{22}$; and $c_{21}(\boldsymbol{B}) = c_{21}$, but now suppose vendor 1's costs for attribute 1 depends on the budget. For example, assume the following relationship: $c_{11}(B) = c_{11} - k\boldsymbol{B}$ >0. Also let $\boldsymbol{B}$<c_{11}/k, $c_{11} > c_{12}$, and $k \in [0,1)$.[22] In this case (from (6a'') and (6b'')) each vendor's optimal attribute bundle proposals for a unit funding/budget level $\boldsymbol{B}$ is given by:

(14a) $a_{11}^*= [\alpha/c_{11}(\boldsymbol{B})]\ \boldsymbol{B} = [\alpha/(c_{11} - k\boldsymbol{B})]\boldsymbol{B}$,

(14b) $a_{21}^*= [(1-\alpha)/c_{21}]\boldsymbol{B}$, and

(15a) $a_{12}^*= [\alpha/c_{12}]\boldsymbol{B}$,

(15b) $a_{22}^*= [(1-\alpha)/c_{22}]\boldsymbol{B}$.

---

[21] For an example of the latter, see the Defense Acquisition Guidebook, which states, "Cost-effectiveness comparisons in theory would be best if the analysis structured the alternatives so that all the alternatives have equal effectiveness (the best alternative is the one with lowest cost) or equal cost (the best alternative is the one with the greatest effectiveness). Either case would be preferred; however, in actual practice, in many cases the ideal of equal effectiveness or equal cost alternatives is difficult or impossible to achieve due to the complexity of AoA [Analysis of Alternatives] issues. ***A common method for dealing with such situations is to provide a scatter plot of [competing vendor proposals'] effectiveness versus cost***" [emphasis added] (CH 2–2.3.2.7 AoA Study Plan-Cost-Effectiveness Comparisons, https://www.dau.mil/tools/dag/Pages/DAG-Page-Viewer.aspx?source=https://www.dau.mil/guidebooks).
[22] These simple assumptions help illustrate our point. A model with quadratic costs could add another dimension (a "knee of the curve," i.e., monotonic increasing with a single inflection point) to the cost-effectiveness function, which could offer an interesting extension of the model.

Figure 5 illustrates each vendor's optimal attribute bundle bid proposals (given by (14a,b) and (15a,b)) for a specific budget, **B**, i.e. points $A_1\colon(a_{11}^*, a_{21}^*)$ and $A_2\colon(a_{12}^*, a_{22}^*)$.

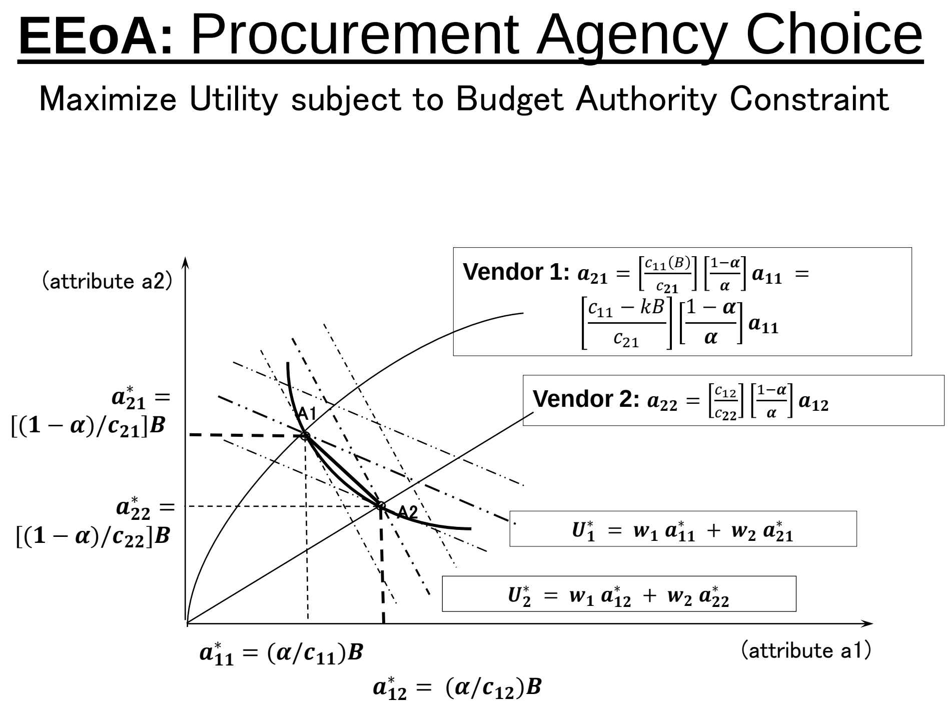


Figure 5. **Vendor Selection When Vendor 1's Attribute Costs Depend on Budget**

The expansion path for vendor 2 is again linear, with the same positive, constant slope for any budget (i.e., identical to (7b')). However, since vendor 1's attribute costs now depend on the anticipated per unit funding/budget, B, vendor 1's expansion path is nonlinear, increasing at a decreasing rate as illustrated in Figure 5 and given by:[23]

(16) $a_{21} = [c_{11}(B)/c_{21}][(1-\alpha)/\alpha]\, a_{11} = [(c_{11} - kB)/c_{21}][(1-\alpha)/\alpha]\, a_{11}$,

where the slope (first derivative) is given by:

(16') $\partial a_{21}/\partial a_{11} = [c_{11}(B)/c_{21}][(1-\alpha)/\alpha] = [(c_{11} - kB)/c_{21}][(1-\alpha)/\alpha] > 0$,

and change in slope with a change in the budget (second derivative) given by:

(16'') $\partial(\partial a_{21}/\partial a_{11})/\partial B = [c_{11}'(B)/c_{21}][(1-\alpha)/\alpha] < 0$.

---

[23] The illustration of the two expansion paths assumes that throughout the relevant range of budgets (funding levels), $(c_{11}(B)/c_{21}) > (c_{12}/c_{22})$.

Substituting vendor 1 and 2's optimal attribute bundle offers ((14a,b) and (15a,b)) into the procurement official's (buyer's) utility function for any given scenario in Equation 8' yields:[24]

$$(17)\ \mathrm{U}_1^* = w_1 a_{11}^* + w_1 a_{21}^* = w_1\,[\alpha/c_{11}(B)]\,B + w_2\,[(1-\alpha)/c_{21}]\,B$$

$$(18)\ \mathrm{U}_2^* = w_1 a_{12}^* + w_1 a_{22}^* = w_1\,[\alpha/c_{12}]\,B + w_2\,[(1-\alpha)/c_{22}]B.$$

Equations 17 & 18 represent functions that can be plotted in cost-effectiveness (Budget-Utility) space over a relevant range of funding scenarios (see Figure 6). In this case, assuming identical, constant costs for attribute 2 (i.e. $c_{21} = c_{22} = c_2$), from (17) and (18),

$$(19)\ \mathrm{U}_1^* \gtreqless \mathrm{U}_2^* \text{ as } c_{12} \gtreqless c_{11}(\boldsymbol{B}) = c_{11} - k\boldsymbol{B} \text{ or as } \boldsymbol{B} \gtreqless (c_{11} - c_{12})/k = B'.$$

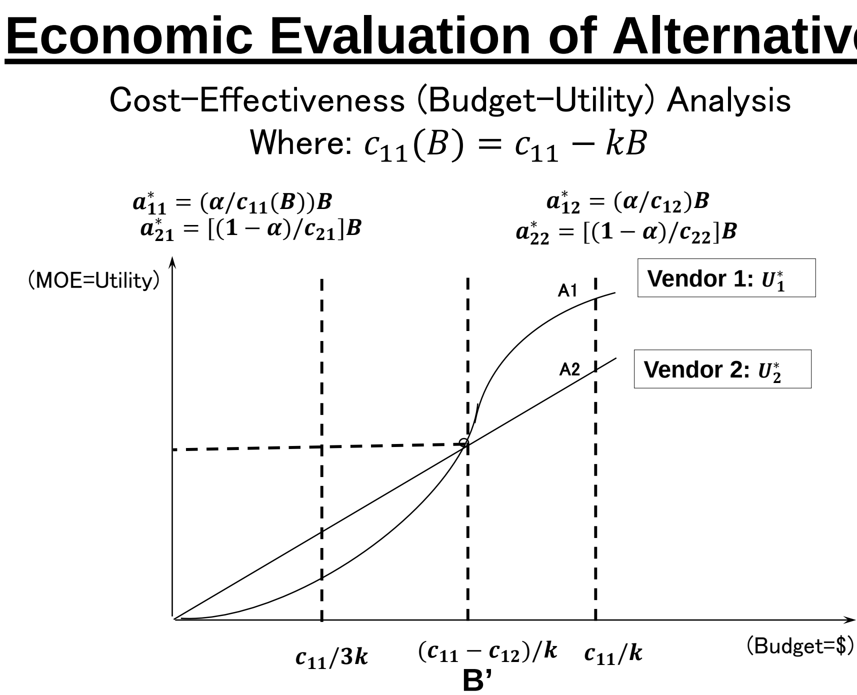


Figure 6. **Vendor Selection in Cost-Effectiveness Space**

What is revealed in Figure 6, is that the relation given by Equation 19 indicates it is optimal for the buyer to switch vendors at B' (i.e., an *optimal rank reversal*). For any unit funding/budgets **B**>B', vendor 1 is ranked higher than vendor 2. The two are ranked the same for the budget, **B**=B', and for budgets **B**<B', vendor 2 is ranked higher than 1. As

[24] For a specific funding level **B**, this represents two optima that can be compared that represent the maximum utility a buyer can obtain from each vendor. This is illustrated in Figure 4 as the highest indifference curve attainable given the corresponding point on the attribute production possibility frontier.

expected, evaluating the slopes of the two vendors' cost-effectiveness functions at the switch point, B'=$(c_{11} - c_{12})/k$, yields:

(20) $\partial \mathrm{U}_1^*/\partial \boldsymbol{B} > \partial \mathrm{U}_2^*/\partial \boldsymbol{B}$ or $(c_{11}(\boldsymbol{B}) - c_{11}'(\boldsymbol{B})\boldsymbol{B})/c_{11}(\boldsymbol{B})^2 > 1/c_{12}$ since $c_{11} > c_{12}$.

This highlights the importance of modeling the supply side. Specifically, this example emphasizes the importance for public procurement officials to obtain realistic budget forecasts for government programs, and to offer those as guidance to vendors. As two pioneers in defense economics Hitch & McKean (1967) wisely counseled,

> As a starter ... several budget sizes can be assumed. If the same [vendor] is preferred for all … budgets, that system is dominant. If the same [vendor] is not dominant, use of several … budgets is nevertheless an essential step, because it provides vital information to the decision maker.

Instead of plotting procurement alternatives (vendor bid proposals) as single points in cost-effectiveness (budget-value) space, EEoA encourages procurement officials in fiscally constrained environments to solicit bids over a range of possible budget scenarios.[25]

Conclusion and Avenues for Future Research

This paper offers an economic model to assist public procurement officials to rank competing vendors when benefits cannot be monetized. The problem of ranking public investment alternatives when benefits cannot be monetized has spawned an extensive literature that underpins widely applied decision tools. The bulk of the literature, and most government-mandated decision tools, focuses on the demand side of a public procurement. The Economic Evaluation of Alternatives (EEoA) extends the analysis to the supply side.

Introducing the supply side offers multiple avenues for further research. Notably, it provides fertile ground to apply both auction and game theory literatures. An interesting extension would be to leverage auction theory and introduce strategic shading of bids by vendors. Another is to consider the risk of collusion among vendors, or allow some vendors to enjoy economies of scale (i.e., to make production technology parameters a function of the budget). Whereas EEoA models vendors as proposing bundles of characteristics to win a budget "prize," alternative optimization assumptions and strategic behaviors could be assumed.

A rich opportunity also exists for both experimental and qualitative research to significantly improve public procurement. An important empirical question is whether procurement officials and managers would have an easier time using EEoA or MCDM (or some combination). Consistency tests could be conducted in experimental settings to explore when the two techniques converge (offer identical vendor rankings), and when (and why) they diverge.

In conclusion, the Economic Evaluation of Alternatives (EEoA) captures both the demand side (government procurement official decisions) and the supply side (vendor

---

[25] In this case, the standard technique of eliminating "dominated alternatives" could lead to sub-optimal decisions. For example, see Chapter 4 in Melese, Richter, & Solomon (2015), or the specific example of the EEoA model developed in Simon & Melese (2011).

optimization decisions). A unique feature of EEoA is to model vendor decisions in response to government funding projections. Given a parsimonious set of continuously differentiable evaluation criteria, EEoA provides a new tool to rank vendors. In other cases, it offers a valuable consistency check to guide government supplier decisions.

References

Beil, D., & Wein, L. (2003). An inverse-optimization-based auction mechanism to support a multiattribute RFQ process. *Management Science, 49*(11), 1529–1545.

Blondal. (2006). International experience using outsourcing, public-private partnerships, and vouchers. In J. M. Kamensky & A. Morales (Eds.)*, Competition, choice, and incentives in government programs* (pp. 121–159). Lanham, MD: Rowman & Littlefield.

Buede, D. M., & Bresnick, T. A. (1999). Applications of decision analysis to the military systems acquisition process. *Interfaces, 22*(6), 110–125.

Che, Y. (1993). Design competition through multidimensional auctions. *RAND Journal of Economics, 24*(4), 668–680.

Darji, V. P., & Rao, R. V. (2014). Intelligent multi criteria decision making methods for material selection in sugar industry. *Procedia Materials Science*, *5*, 2585–2594. Retrieved from https://doi.org/10.1016/j.mspro.2014.07.519

Defense Acquisition University (DAU). (n.d.) *Defense acquisition guidebook*. Retrieved from https://www.dau.mil/tools/dag

Dyer, J. S., & Sarin, R. K. (1979). Measurable multiattribute value functions. *Operations Research, 27*(4), 810–822.

Dyer, J. S., & Sarin, R. K. (1982). Relative risk aversion. *Management Science, 28*(8), 875–886.

Ewing P. L., Jr., Tarantino, W., & Parnell, G. S. (2006). Use of decision analysis in the army base realignment and closure (BRAC) 2005 military value analysis. *Decision Analysis, 3*(1), 33–49.

Federal Acquisition Regulation (FAR), 48 C.F.R. ch. 1 (2018). Retrieved from https://www.acquisition.gov/sites/default/files/current/far/pdf/FAR.pdf

Gal, T., Stewart, T., & Hanne, T. (Eds.). (2013). *Multicriteria decision making: Advances in MCDM models, algorithms, theory, and applications* (Vol. 21). Springer Science & Business Media.

Hitch, C., & McKean, R. (1967). *The economics of defense in the nuclear age*. Cambridge, MA: Harvard University Press.

Jamil, N., Besar, R., & Sim, H. K. (2013, January). A study of multicriteria decision making for supplier selection in automotive industry. *Journal of Industrial Engineering*, 23. Retrieved from https://doi.org/10.1155/2013/841584

Jordan, K. (2007). The NMCI experience and lessons learned. The consolidation of networks by outsourcing. *Case Studies in National Security Transformation, 12.*

Keeney, R. L. (1992). *Value-focused thinking: A path to creative decision-making*. Cambridge, MA: Harvard University Press.

Keeney, R. L. (2004). Making better decision makers. *Decision Analysis, 1*(4), 193–204.

Keeney, R. L., & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value trade-offs*. New York, NY: John Wiley and Sons.

Kennedy, B. R. (1999). Historical realities of c-17 program pose challenge for future acquisitions. *Program Manager Magazine, 28*(6), 70–78.

Kirkwood, C. (1997). *Strategic decision making*. New York, NY: Duxbury Press.

Kirkwood, C. W., & Sarin, R. K. (1980). Preference conditions for multi-attribute value functions. *Operations Research, 28*(1), 225–232.

Lancaster, K. J. (1966a). A new approach to consumer theory. *Journal of Political Economy, 74*(2), 132–157.

Lancaster, K. J. (1966b). Change and innovation in the technology of consumption. *American Economic Review, 56*(2), 14–23.

Lancaster, K. J. (1971). *Consumer demand: A new approach*. New York, NY: Columbia University Press.

Lancaster, K. J. (1979). *Variety, equity, and efficiency*. New York, NY: Columbia University Press.

Larsen, R. F. (2007). The evolution of the Pentagon's strategic warfighting resources and risk process. *Land Warfare Papers*, No. 64.

Loerch, A. G., Koury, R. R., & Maxwell, D. T. (1999). Value added analysis for army equipment modernization. *Naval Research Logistics, 46(*3), 233–253.

Matheson, J. E., & Abbas, A. E. (2005). Utility transversality: A value-based approach. *Journal of Multi-Criteria Decision Analysis, 13*(5–6), 229–238.

Melese, F. (2015). The economic evaluation of alternatives. In F. Melese, A. Richter, & B. Solomon, *Military cost-benefit analysis: Theory & practice* (Chapter 4). Routledge.

Melese, F. (2018). Cost-benefit analysis of bid protests: A representative bidder model. *Defence & Peace Economics.*

Michael, R. T., & Becker, G. S. (1973). On the new theory of consumer behavior. *The Swedish Journal of Economics, 75*(4), 378–396.

OECD Public Procurement Dataset. (2016). Retrieved from https://qdd.oecd.org/subject.aspx?Subject=GOV_PUBPRO_2016

Parkes, D., & Kalagnanam, J. (2005). Models for iterative multi-attribute procurement auctions. *Management Science, 51*(3), 435–451.

Parnell, G. S. (2007). Value-focused thinking using multiple objective decision analysis. In A. Loerch & L. Rainey (Eds.), *Methods for conducting military operational analysis: Best practices in use throughout the Department of Defense* (pp. 619–656). Alexandria, VA: Military Operations Research Society.

Pratt, J. W., Raiffa, H., & Schlaifer, R. (1964). The foundations of decision under uncertainty: An elementary exposition. *Journal of the American Statistical Association, 59*(306), 353–375.

Quade, E. S. (1989). *Analysis for public decisions* (Rev. ed. by G. M. Carter). New York, NY: North-Holland.

Ratchford, B. (1979). Operationalizing economic models of demand for product characteristics. *Journal of Consumer Research, 6*(1), 76–85.

Schwartz, M., Sargent, J., & Mann, C. (2018). Defense acquisitions: How and where DOD spends its contracting dollars. Washington, DC: Congressional Research Service. Retrieved from https://fas.org/sgp/crs/natsec/R44010.pdf

Simon, J., & Melese, F. (2011). A multiattribute sealed-bid procurement auction with multiple budgets for government vendor selection. *Decision Analysis, 8*(3), 170–179.

Sproles, N. (2000). Coming to grips with measures of effectiveness. *Systems Engineering, 3*(1), 50–58.

Wilson, L. E. (2006). *NMCI: The silver lining* (EWS Contemporary Issue Paper).

Wise, R., & Morrison, D. (2000). Beyond the exchange—The future of B2B. *Harvard Business Review, 78*(6), 86–96.

Acknowledgements & Disclaimer

We are grateful to Colonel (Ret.) John Dillard, Dr. Bill Gates, Lt. Colonel (Ret.) Jeff House, Dr. Eva Regnier, Dr. Anke Richter, and other NPS colleagues for valuable comments and suggestions. The views expressed belong exclusively to the authors and do not necessarily reflect those of our colleagues, the Naval Postgraduate School, the U.S. Navy, or the Department of Defense.

A Framework for a Defense Systems Effectiveness Modeling and Analysis Capability: Systems Effectiveness Modeling for Acquisition

John M. Green—is currently a Senior Lecturer in the Systems Engineering Department at the Naval Postgraduate School. He holds an MBA and MS in computer science from University of New Haven, an MA in international relations from Salve Regina College, and a BS in physics from Saginaw Valley State University. He is also a graduate of the Naval War College, College of Command and Staff. Green is a Senior Member of IEEE and AIAA. He is also a member of MORS, ASNE, INFORMS, the Association of Old Crows, and INCOSE. [jmgreen@nps.edu]

Jerrell Stracener—PhD, is a Senior Research Associate in the Southern Methodist University (SMU) AT&T Center for Virtualization with a research focus on U.S. defense applications and systems effectiveness. He was Founding Director of the SMU Systems Engineering Program, and as Professor of Practice, taught graduate-level courses in engineering probability and statistics, systems reliability and availability analysis, integrated logistics support (ILS), and performed/directed systems engineering research and supervised PhD student research. He served in the U.S. Navy and earned both PhD and MS degrees in statistics from SMU and a BS in math from University of Texas at Arlington. [jerrell@smu.edu]

Abstract

The purpose of this paper is to present a response to two current Department of Defense (DoD) initiatives. The first is the DoD National Defense Strategy of 2018, which encourages the adoption of new practices to improve system performance and affordability to meet current and future threats. The second initiative is the DoD Digital Engineering Strategy, which outlines five strategic goals in support of the first initiative. The first strategic goal—"Formalize the development, integration, and use of models to inform enterprise and program decision making"—is the specific subject of this paper. The response is a conceptual methodology that addresses an analytic deficiency identified by a 2017 congressional commission that examined the capabilities of the DoD civilian staff in their determination of force and weapons systems requirements. Specifically, this paper presents a framework for a "Defense Systems Effectiveness Modeling and Analysis Capability" whose metric is the probability of mission success. The objective is the application of modeling and analysis to guide decisions leading to fielding systems having optimum effectiveness constrained by affordability and reduced development time. While the current U.S. focus is on systems readiness, it is an integral element of the more robust systems effectiveness.

Introduction

The 2018 National Defense Strategy (NDS; DoD, 2018) makes readiness and warfighter needs a priority, with lethality and warfighting the primary objective. The strategy emphasizes affordability with sustained and predictable investment to achieve greater performance through modernizing the military and restoring readiness. Within this context, improvement of readiness involves developing the right systems or systems of systems with alacrity.

To support the goals of the NDS, the DoD's Under Secretary of Defense for Research and Engineering has initiated the Digital Engineering Strategy (DES), which has five goals intended to drive the acquisition of future systems (Office of the Deputy Assistant Secretary of Defense for Systems Engineering, 2018). The five goals promote a model-based, systems engineering (MBSE) wherein systems are digitally rendered. The resulting digital artifacts become the means of communications between stakeholders. The goals are as follows:

1. Formalize the development, integration, and use of models to inform enterprise and program decision making;
2. Provide an enduring, authoritative source of truth;
3. Incorporate technological innovation to improve the engineering practice;
4. Establish a supporting infrastructure and environments to perform activities, collaborate, and communicate across stakeholders; and
5. Transform the culture and workforce to adopt and support digital engineering across the lifecycle.

Purpose

An approach to the first goal of the DES is the purpose of this paper. A crucial element of the formalization process is the development of an effectiveness modeling and analysis framework. The advent of DES is important because recent criticism by a bipartisan congressional commission noted that civilian analytical capabilities for force and weapons development within the DoD have severely degraded since their original establishment in the 1960s by Robert McNamara (Gordon & Lubold, 2018). The truth of this statement is borne out by the lack of an established methodology within the DoD for acquiring systems of systems. There is current work underway addressing systems of systems, mission engineering, and capability portfolio analysis but not at the level of the Weapon System Effectiveness Industry Advisory Committee (WSEIAC) study to be discussed shortly.

Specific Contribution of This Paper

The contribution of this paper is twofold. First, it provides clarity of purpose for readiness, an oft used and abused term. Why not readiness? A focus on readiness may lead to sub-optimum system solution because it ignores three other factors important to systems effectiveness and mission success. Mission success is the applicable measure because it drives force projection and war-fighting capability. Second, the paper presents a framework that addresses the role of readiness within the context of mission success. This framework applies to both systems and systems of systems acquisition, providing the stakeholders with quantified results.[1]

Organization of Paper

The paper provides a brief discussion of relevant past work that is foundational to the development of the Defense Systems Effectiveness Modeling and Analysis Capability (DSEMAC). Key terms are defined mathematically, followed by a brief discussion of the requirements for a framework that provides the needed structure for the DSEMAC, which in turn is followed by a description of the proposed framework. A summary and a description of future work conclude the paper.

Past Work

A focus on readiness ignores the larger context of systems effectiveness and the additional attributes of mission reliability, mission survivability, and mission capability. It is

[1] *System* will be used throughout this paper.

the premise of this paper that system effectiveness and mission success are the same and the overarching goal. Readiness is a subset of the larger picture that includes mission reliability, mission survivability, and mission capability as shown in Figure 1. This view is not a new concept. The relationships have a long history that started in the 1950s and was extensively documented in a report published by the Weapon System Effectiveness Industry Advisory Committee (WSEIAC) in the 1960s (WSEIAC, 1965). Figure 1 is based on the WSEIAC report and illustrates the relationship between overall mission effectiveness and its constituent components of mission readiness, mission reliability, and mission capability. Note that mission survivability is not included in the report and thus omitted from Figure 1. Survivability is included in this paper for completeness, as shown in Figure 2.

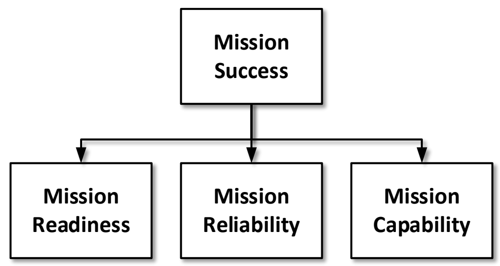


Figure 1. **The WSEIAC Systems Effectiveness Hierarchy**
(WSEIAC, 1965)

As defined by the WSEIAC report, mission readiness (often known as operation availability [A_o] or operational readiness [OR]) quantifies the percentage of time that the system is ready at the start of the mission. Mission reliability (or dependability) quantifies the likelihood that the system will perform its mission essential functions throughout the mission. Both these terms are well represented in the literature. Mission capability quantifies the adequacy of the system to meet the mission goals. Capability is about ways and means. It matters not if the system is available and reliable throughout the mission if it cannot achieve the desired results because the said ways and means were insufficient or incorrect.

Figure 2 presents a complete view of the relationships with the addition of mission survivability. The probability of mission success is a function of the four terms. Therefore, the graphic is a top-level objective hierarchy. As an objective tree, the goal is to maximize the probability of mission success. The lower-level objectives each describe a specific aspect of mission success and are, therefore, inherently important. The lower-level objectives can be expanded by including another level of detail. For example, mission survivability can be expanded to susceptibility and vulnerability. In this case, the goal is to reduce both to increase survivability.

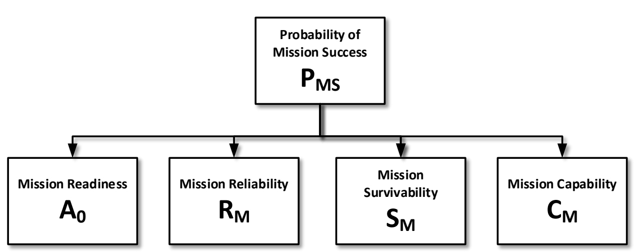


Figure 2. **The Revised WSEIAC Systems Effectiveness Hierarchy**

The systems effectiveness hierarchy and the following equation for P_{MS} provides a quantitative basis for the acquisition of weapons systems and systems of systems. The WSEIAC report provided a general mathematical relationship for mission success as follows:

$$P_{MS}=(P_{Ao})(P_{RM})(P_{SM})(P_{CM})$$

where,

P_{MS} ≡ the probability of mission success for a specified mission

P_{Ao} ≡ the probability that the system is available at the start of the mission

P_{RM} ≡ the probability that the system will successfully perform specified mission essential functions by mission phase

P_{SM} ≡ the probability that the system will survive the mission

P_{CM} ≡ the probability that the system meets the capability objectives

Note the probabilistic formulation of mission success. There are several valid reasons for this approach. First, military operations are characterized by random variables, for example, probability of detection or probability of kill. Second, probabilities are dimensionless, making them easier to work with across diverse system elements such as sensors and weapons.

Figure 3 is a summary of the relationships that contribute to a framework for system effectiveness.

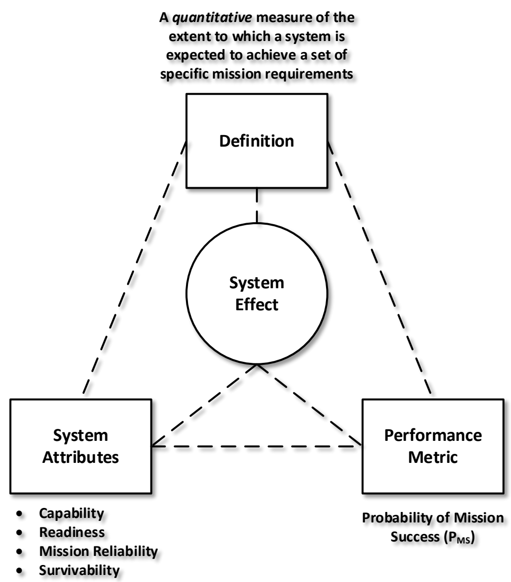


Figure 3. **Systems Effectiveness Relationships**

In systems terminology, Figure 3 is a context diagram that becomes a starting point for the framework requirements discussed in the following section.

Framework Requirements

A framework is a structured way of relating objects of interest and their resulting interactions. The importance of a framework in the acquisition of systems cannot be understated. First, a framework organizes theory and practice and provides a structure for methods. Second, complex systems and systems of systems are typically not developed as a single architecture. Thus, there are time-phasing and contractual issues. Individual systems are usually single function, and system couplings are interdependent (Luman, 2000).

Third, there currently is no systematic method of measuring systems effectiveness. The literature is devoid of theory and standards. Most approaches center on qualitative methods, which are subjective at best.

Basic Requirements

There are four major requirements for the framework: The supporting methods must be quantitative, the supporting methods must present results probabilistically, the supporting methods must be reliability based, and, finally, the framework must support hierarchy and abstraction. The end goal is a framework that supports evaluation of mission success versus cost, where the emphasis is on the likelihood of mission success.

Quantitative

One of the first steps in an analysis is to describe the processes involved. Mathematics is precise and explanatory, facilitating analysis and explanation of more complex problems than possible using qualitative methods. The model for the probability of mission success must be based on proven methodology. The challenge is developing and maintaining a model for each mission which will be large and complex for complex systems.

Probabilistic

Military operations are about achieving success and the estimation of event probabilities, typically described as measures of effectiveness (MOE) or measures of performance (MOP). Often parametric values are used incorrectly as measures. For example, detection of a threat is expressed as a probability of detection and is a function of several parameters including range. The outcome is the probability of detection as a function of range.

Reliability-Based

Reliability theory is based on the premise of system success and failure ($P_{success} = 1 - P_{failure}$). Many of its concepts are foundational precepts to quantifying system effectiveness. Further, most of the system variables of interest are reliability related. Figure 3 identifies them as key system attributes.

Hierarchy and Abstraction

Systems are hierarchical by nature with increasing detail at each level of expansion. The framework must support models that describe each level of expansion. This paper suggests a black box approach at each layer.

A Notional Effectiveness Model

Systems concepts are based on a need to meet an operational requirement. The effectiveness of how well this need is met (mission success) is a measure of its tactical utility and its value to the force structure. Figure 4 is a notional model adapted from Figure 2-1 found in the *Reliability Engineering Handbook* (Bureau of Naval Weapons, 1964). It summarizes the first three figures and is intended to convey several points: how well the system will perform, how long the system will perform, and how often the system can perform.

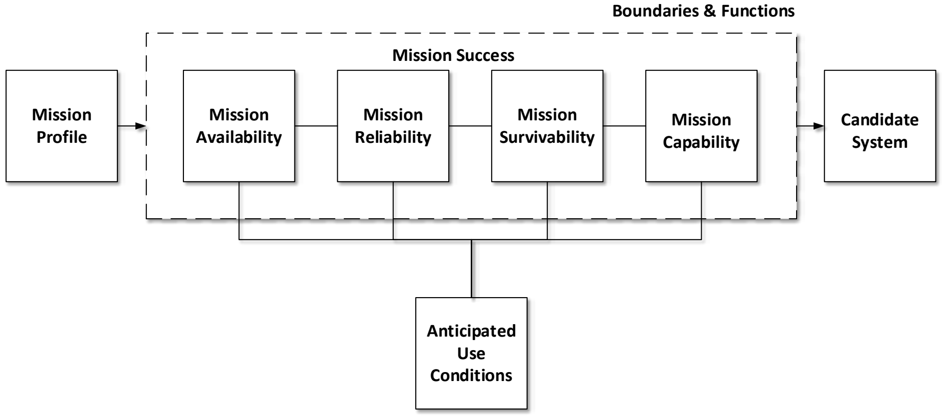


Figure 4. **A Systems Effectiveness Model**
(Adapted from Bureau of Naval Weapons, 1964)

This model, when combined with a decision process, becomes the basis for the overall framework model.

Proposed Framework

Figure 5 is a generic decision process. It serves as a guide to understanding how to incorporate Figure 4 into a larger context. Figure 6 is the resulting proposed framework.

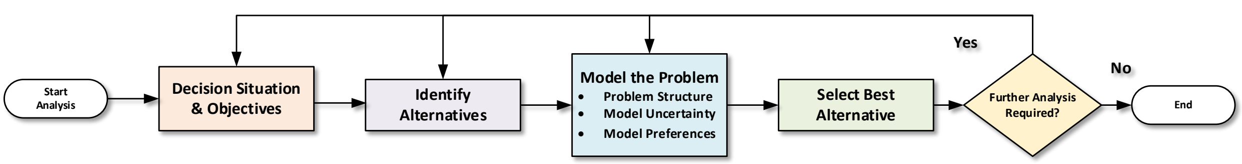


Figure 5. **Generic Decision Process**

Problem Formulation

With the framework in place, it is appropriate to return to the purpose of the framework to wit: to make decisions about system selection. There are three basic steps to the decision process. First, understand the set of system variables and how they interact quantitatively and accurately. Knowledge of the system is imperative. In the framework, this is represented by the upper five boxes (orange and purple). Second, select a single MOE expressible in terms of the variables represented by the blue boxes. A premise of this paper is that mission success is that MOE. The final step is to select the method by which the best system is selected represented by the green boxes.

The decisions involve making choices from a set of candidate solutions in order to find the most desirable solution. Once the decision is made, it becomes an irrevocable allocation of resources. Given the set of candidate solutions, the task becomes one of defining a system such that:

$$\text{Maximize } P_{MS} = (P_{Ao})(P_{RM})(P_{SM})(P_{CM}),$$

subject to the following constraints:

- Specified mission
- Required performance
- Budget

This is a basic optimization problem. It is decisive because the result is one system—the best one.

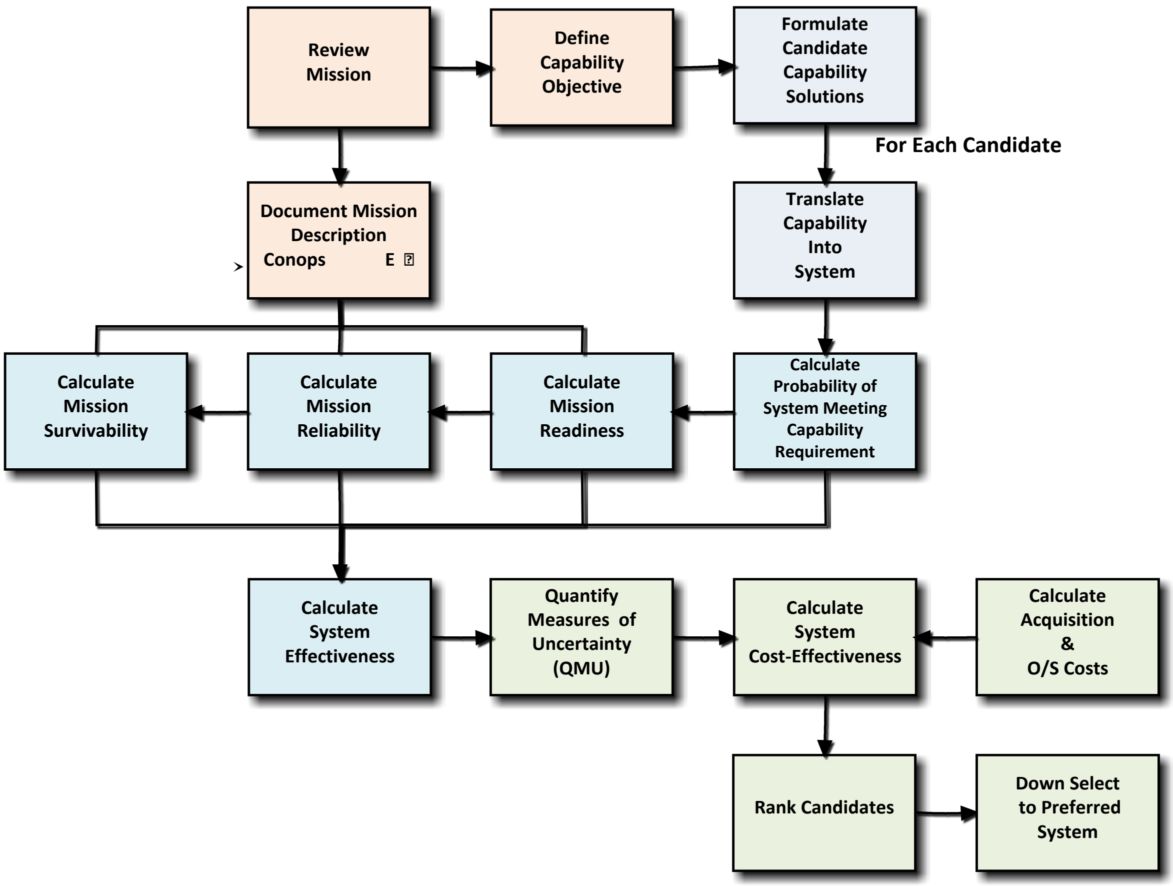


Figure 6. **A Framework for a Defense Systems Effectiveness Modeling and Analysis Capability**

Comments on Cost-Effectiveness

In the model described above, cost-effectiveness has been chosen as the criterion for the model because it is best used for ranking alternatives that are relatively similar especially when there is a single dominant objective whose attainment can be assessed directly or for which a good proxy value exists (Quade, 1982). It is axiomatic in the world of quantitative analysis that, in general, the possibility of selecting between two alternatives based on cost and effectiveness data alone is not possible. It is a choice between specifying performance or cost. If the former, then cost is minimized; if the latter, then effectiveness is maximized.

Summary

This paper presents the rationale for a framework for a Defense Systems Effectiveness Modeling and Analysis Capability. It describes why the key decision criterion is the probability of mission success and shows the approach to the derivation of the framework. This framework is inclusive of capability, readiness, mission reliability, and survivability (which is typically omitted in system effectiveness evaluations).

Future Research

As noted, survivability is not usually included. While availability and readiness have a large literature base, there is very little material on survivability.

A second research topic is Candidate Capability Architecture solution development. There is no literature on performance-based architecture development.

References

Bureau of Naval Weapons. (1964). *Handbook: Reliability engineering*. Defense Technical Information Center. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/a286606.pdf

DoD. (2018). *National defense strategy.* Washington, DC: DoD.

Gordon, M., & Lubold, G. (2018). Study cites weak civilian control of military. *Wall Street Journal*.

Luman, R. R. (2000). Integrating cost and performance models to determine requirements allocation for complex systems. *Johns Hopkins APL Technical Digest.* Retrieved from http://www.jhuapl.edu/techdigest/TD/td2103/luman.pdf

Office of the Deputy Assistant Secretary of Defense for Systems Engineering. (2018). *DoD digital engineering strategy.* Washington, DC: Author.

Quade, E. S. (1982). *Analysis for public decisions.* New York, NY: Elsevier Science.

Weapons Systems Effectiveness Industry Advisory Committee (WSEIAC). (1965). *Final report of Task Group 1* (AFSC-TR-65-2).

Panel 18. Insights Into Defense Acquisition Program Oversight

Thursday, May 9, 2019	
12:45 p.m. – 2:00 p.m.	**Chair: Travis Masters,** Assistant Director, Contracting and National Security Acquisitions, U.S. Government Accountability Office ***Foundational Assumptions and Associated Observables*** Gregory Davis, John Bailey, and Matthew Goldberg, Institute for Defense Analyses ***Do We Need a Different Approach to Statistical Analysis of Research, Development, Test, and Evaluation Cost Growth?*** David McNicol, Institute for Defense Analyses ***Quantifying the Year-by-Year Cost Variance of Major Defense Programs*** David Tate and Michael Guggisberg, Institute for Defense Analyses

Foundational Assumptions and Associated Observables

Gregory A. Davis—PhD, has worked at the Institute for Defense Analyses (IDA) since 2006, conducting research on as broad a range of topics as he can find. Before coming to IDA, he was an AAAS Science and Technology Policy Fellow in OSD(PA&E). Most of his recent work has focused on major defense acquisition programs that have experienced cost growth. Davis also won the Kubla Kahn 2004 Diplomacy championship and is now in another Diplomacy game, where he isn't doing nearly as well. [gdavis@ida.org]

John W. Bailey—PhD, has worked for both the former Computer Science Division and the Cost Analysis Research Division of IDA since the 1980s. Prior to that, he conducted software programming research for General Electric and co-founded Software Metrics, Inc. While completing his PhD in software reuse methodology, he helped Rational establish its Ada development environment, which was later acquired by IBM. Bailey is now applying computer science best practices to a 70-acre cattle farm in northern Virginia. [jbailey@ida.org]

Matthew S. Goldberg—is an economist with over 30 years of experience in defense analysis. He is on his second tour at IDA, having also worked at CNA, RAND, and as a Deputy Division Director at the Congressional Budget Office. He has published extensively in the fields of economics, cost analysis, operations research, and applied statistics. He is also a published jazz composer and arranger. [mgoldber@ida.org]

Abstract

The Institute for Defense Analyses worked with the Office of the Secretary of Defense (OSD) to invent Foundational Assumptions and Associated Observables (FAAOs). FAAOs are a tool to help oversight organizations monitor progress in acquisition programs between milestones. FAAOs are similar to the Framing Assumptions required in the 2015 version of the DoD's acquisition regulations, but they are created by and for oversight organizations, not those executing the programs. In addition to inventing the process, we also delivered five sets of FAAOs to the OSD for use in oversight, which was an essential step in creating the process.

Introduction

It may be apocryphal, but some say that if you drop a frog into a pot full of hot water, it will jump out without injury. But if you drop it into a pot of water at a comfortable temperature and heat it up gradually, you can boil it.[1] Acquisition programs seem to be similar. In this metaphor, when a program goes through a milestone review, its temperature is measured and actions are taken to make sure all is well. Between milestones, programs can quietly morph into something quite different, and the metaphorical temperature can rise as the program experiences unanticipated challenges. The purpose of Foundational Assumptions and Associated Observables, or FAAOs, is to note the temperature at the milestone review and to create a process for the oversight community to measure it regularly.

[1] While frogs may not behave this way, we do not want to discourage you from catching a frog and testing it out, if you are so inclined.

History of Framing Assumptions

The Weapon Systems Acquisition Reform Act of 2009 (WSARA) mandated the appointment of a Director of Performance Assessment and Root Cause Analysis (PARCA) within the Office of the Secretary of Defense (OSD), who, under certain circumstances, must perform root cause analyses (RCAs) on major defense acquisition programs (MDAPs), most often when too much cost growth has triggered a Nunn-McCurdy (NM) breach.[2] Immediately upon the director's appointment, he and his first staff members conducted five RCAs simultaneously: the Advanced Threat Infrared Countermeasure and Common Missile Warning System (ATIRCM/CMWS), the Apache Longbow Block III, the DDG-1000, the F-35, and the Remote Minehunting System (RMS). This was a period of unusually intense effort. In the years since, they have conducted 17 more RCAs.

Early on, PARCA noticed a pattern in these results. MDAPs do not generally suffer cost growth leading to NM breaches because of small errors. Rather, it was often the case that programs suffered from invalid major assumptions starting early in each MDAP's life cycle. Often the problems grew out of management errors or unrealistic baselines. These incorrect assumptions later caused major difficulties for the program.

PARCA's staff termed these incorrect assumptions in the troubled programs *framing assumptions*, or FAs. They realized that all acquisition programs must depend on such assumptions, most of which are accurate and therefore cause no problems. They also figured that if these assumptions were made explicit early on in an MDAP's life cycle and then monitored, programs that experienced problems because of failed FAs would be identified sooner and dealt with more easily.

PARCA disseminated information about FAs. In January 2015, the Under Secretary of Defense for Acquisition, Technology and Logistics (USD[AT&L]) signed a new instruction, DoD Instruction (DoDI) 5000.02, *Operation of the Defense Acquisition System*, which mandated that program managers report their FAs at the Milestone A review, again at the Request for Proposal (RFP) release decision point, and again at Milestone B Defense Acquisition Board (DAB) meetings.

DAB briefings given after DoDI 5000.02 was signed included discussions of FAs, but the FAs presented varied significantly in both quality and follow-through. Some MDAPs built elaborate checklists, while others seem to have had no more than a single PowerPoint slide with a few bullet points. Some MDAP program offices worked with PARCA staff to help develop their FAs while others did not. Some MDAPs presented inconsistent lists of FAs in their DAB charts and other documents, such as their acquisition strategies (ASs).

In December 2014, just before the instruction was signed, the Armored Multipurpose Vehicle (AMPV) program came up for a Milestone B review. AMPV's DAB charts had these FAs:

---

[2] NM breaches are triggered by growth in average unit cost, which is the quotient of total dollars in the program divided by the number of units. Both the numerator and denominator include the past and the projected future. The details of the triggers are too complex for a footnote and not directly relevant to this paper. A succinct description can be found in Appendix A of Arnold et al. (2010).

- Vendors will offer military derivative solutions similar to what the Army used for cost estimating purposes.
- Vendors offer mature designs and deliver prototypes 24 months after contract award
- Mission Equipment Package (MEP) Configuration locked at Preliminary Design Review (PDR) for Engineering and Manufacturing Development (EMD), Critical Design Review (CDR) for Low Rate Initial Production (LRIP)

The AMPV AS contains the following paragraph:

> Framing assumptions for the development of this AS include: A Bradley-based solution is an adequate analogy for representing the cost and schedule estimates for the offeror selected for the EMD program (the Request for Proposal [RFP] was built to allow any military vehicle or derivative that meets performance requirements); vendors will offer mature designs and are able to deliver prototypes 24 months after contract award; AMPV will remain in the fleet for 50+ years, similar to the M113; and six months is sufficient time to execute source selection.

None of these reported FAs look like big bets upon which AMPV's success is highly dependent. If the first prototype arrives in 30 months instead of 24 months or source selection takes a year, neither implies that the program is fundamentally different. The procurement is expected to run for 18 years. These delays might indicate deeper problems with other assumptions, but they are not overly important by themselves. In fact, AMPV has missed many deadlines, but there is no discussion of cancellation or of an NM breach. The assumption about remaining in the fleet for 50+ years is noteworthy. Such an assumption should have an influence on the vehicle's designers, but a 50-year life span cannot be verified, so it is unclear how stating it as an FA matters for monitoring the health of the program. Our foundational assumptions for this program are presented in a later section, Final Deliverable.

While each MDAP generates FAs for itself, resulting in considerable variation, the assumptions in the AMPV program were not atypical. PARCA wanted to use FAs to monitor program health, but few of them were suitable. They had tried to train program managers to do a better job, but instead developed an alternative internal approach that became FAAOs.

The Origin of FAAOs

FAAOs were created by the Institute for Defense Analyses (IDA) as tools to help PARCA conduct its regular work assessing the performance of MDAPs. Philosophically, foundational assumptions are the same as FAs; both are attempts to identify the big bets in an acquisition program and to track how well that the program is doing between the milestones of the program's life cycle relative to those assumptions. The associated observables enable PARCA to ascertain whether or not the foundational assumptions remain true. However, instead of being the responsibility of the program manager, FAAOs reside at PARCA and may have been written in-house or by external contractors.

If foundational assumptions are philosophically the same as FAs, why do both exist? The motivations of a program office and an oversight organization do not always align. The FAs are generated and owned by the program office, whereas FAAOs belong to PARCA. However, PARCA's FAAOs have no legal or regulatory power. PARCA may seek comments or help from anybody they like, but there is no obligation for the program office to respond. There is no statutory requirement that FAAOs exist.

Because PARCA has no authority, no coordination with external offices is necessary, although PARCA can offer recommendations. FAAOs are brand new and have not yet proven their utility, but we expect them to be useful because PARCA will monitor them, and if they find something alarming, they can take their finding to a senior official in OSD and recommend some extra investigation. The senior official would then decide whether it is worth making a deep dive on the program. It is worth noting that in 2016, the official to notify and convince was the USD(AT&L). Today, that office no longer exists, and it is unclear who in OSD would be the most appropriate official to notify. PARCA could notify the Secretary of Defense or his deputy; if either of those officials decided to act, they could investigate the program and mandate changes, even if there is no relevant lower-level official in OSD.

This Project

This paper is part of PARCA's first endeavor into FAAOs. PARCA contracted with IDA to generate several sets of FAAOs and report on the process, which we did, making us the *writer* of the first FAAOs. We expect PARCA to use these FAAOs by reading them quarterly and comparing what they say to the current status of the programs. The action officer (AO) who performs that role is the *reader*.

We generated FAAOs for five programs. The first was on the AMPV program, for which the FAAOs were in the form of a memo that was revised twice during coordination between our team and PARCA before it achieved consensus. Upon reflection, the disagreements between our team and the sponsor over AMPV were partly about style but also about content. The 2013 definition of an FA, which we adhered to in the AMPV case, is not sufficient. We will discuss this matter further in the following section, Framing Assumptions. The problem with the definition of an FA may have been clearest in the AMPV case not just because it was the first program we examined, but also because—of all the programs we looked at—it was by far the furthest along.

After AMPV, we delivered FAAOs on four more programs as briefing charts:

- Three-Dimensional Expeditionary Long-Range Radar (3DELRR)
- Columbia Class submarine
- Global Positioning System (GPS) IIIF satellites
- DDG-51 flight III ships with their new Air and Missile Defense Radar (AMDR)

The final set of FAAOs should be thought of as one "program," although it spans two MDAPs: the DDG-51 MDAP, which has produced 64 currently active ships with more on the way, and the AMDR, which achieved Milestone B in 2013 but has yet to have its hardware taken to sea for testing.

Framing Assumptions

FAAOs grew out of FAs. Relatively rapidly, the DoD moved to instantiate the idea of requiring certain acquisition programs to identify potential assumptions that, if they were violated, could significantly affect cost, schedule, or performance outcomes. The short time between the report on the theorized benefits of FAs by PARCA analysts in 2013 (Husband, 2013) and the promulgation of FAs in DoDI 5000.02[3] just 16 months later is remarkable.

FAs are defined broadly as "any explicit or implicit assumption that is central to shaping the cost, schedule, and/or performance expectations of a program." The PARCA office in 2013, and, later, Arena and Mayer (2014, p. 2), attempted to more precisely define the attributes of an FA:[4]

- *Critical: Significantly affects program expectations*. This criterion means that FAs, when they fail or are incorrect, will have significant cost, schedule, and/or performance effects on the program. One possible consequence is a formal program breach. Another—arguably appropriate—possible consequence is that the program is cancelled. The criterion is meant to exclude the many smaller assumptions made for a program that do not result in significant consequences.
- *No workarounds: Consequences cannot be easily mitigated*. This criterion implies that valid FAs have no obvious workarounds or potential fixes if they are wrong. When an FA is wrong, there is a very high probability of significant cost and/or schedule implications.
- *Foundational: Not derivative of other assumptions*. This criterion is, perhaps, the hardest to understand and define. An FA is foundational if it is a high-level and encompassing assumption. An FA might have derivative assumptions associated with it, but a proper FA will not be derivative or subordinate to other major assumptions.
- *Program-specific: Not generally applicable to all programs.* This criterion implies that FAs should reflect some unique aspects of the program. For example, an FA is not, "The contractor will perform well." However, an FA might be, "The key technologies are sufficiently mature such that no component development or prototyping is necessary."

This last constraint on the definition of FAs solves one problem, that of bounding the set of applicable FAs, but leads to others by omitting many relevant questions for the health of a program. Bailey and Frazier (2018) discuss the fact that many of the problems in acquisition programs are about general best practices, not program-specific issues. Another problem with this definition is that problems in a program can be at a level where they will not cause an NM breach or cancellation, but still rise to the interest of oversight because of short-term issues.

The FAAOs for AMPV suffered both kinds of issues. Individuals within the Pentagon expressed concerns that the first AMPVs would not be delivered in time to satisfy operational demands because there was insufficient manufacturing capacity for the vehicles. Some felt that for the FAAOs to be useful, they would need to touch on this point, especially because the Under Secretary was worried about them at the time; however, others resisted

[3]No definition is provided, however, which is why research to help operationalize the selection of FAs is needed.

[4]Italicized text is from the PARCA original. RAND also published a slightly earlier treatment (Arena et al., 2013).

because the FAAOs did not satisfy the described criteria in two ways. First, the need for sufficient manufacturing capability was not program-specific, and second, the funding required to fix the manufacturing deficit was very small compared to the total cost of the program. It is possible that we should rethink whether these are the best criteria for FAAOs, but at this point, they have been adopted.

Writing FAAOs

In most instances, a set of FAAOs ought to be associated with a baseline for schedule and cost. The fundamental question the FAAO writer is answering is, “What are the big bets associated with this baseline?”

Data Sources

To answer this question, we started by reading every program-related document we could find. The following list should not be thought of as either necessary or sufficient, but it is suggestive, and for each set of FAAOs, we at least sought out these documents:

1. Requirements Documents (usually a Capability Development Document, or CDD, but not necessarily)
2. Selected Acquisition Reports (SARs)
3. Defense Acquisition Executive Summaries (DAES) reports
4. Approved Program Baseline (APB)
5. Acquisition Strategy (AS)
6. Defense Acquisition Board (DAB) briefing charts
7. Test and Evaluation Master Plan (TEMP)
8. Budget displays in the lead service’s Procurement and RDT&E budget justification books
9. Reports from congressional agencies, typically the Congressional Budget Office (CBO), the Congressional Research Service (CRS), and the Government Accountability Office (GAO).[5]

Availability of these documents varies considerably. They may not all yet exist, depending on the phase of the acquisition program in question. IDA has contacts within the office of the Director of Operational Test and Evaluation (DOT&E), which allows us to access many of them. IDA’s testing experts have also provided numerous briefings from either the program office or the lead service’s testing community. Those briefings often highlight technical or programmatic issues that have been identified, along with proposed mitigation strategies. The progression from one briefing to another is often informative as well—for example, revealing slips in scheduled testing events, operational dates, or the like.

In addition to reading documents, the IDA research team interviewed our testing experts on the programs, PARCA’s AOs who are following the programs and any other

[5] The GAO produces a useful annual summary, most recently GAO (2017).

experts we could find: some from government oversight organizations and others from within IDA. We interviewed at least two experts for each program.

It is critical to capture Key Performance Parameters (KPPs) and Key System Attributes (KSAs), which can be numerous but quite revealing. For example, the 3DELRR program had requirements for interoperability with radars and command systems from all of the other services, but the latter systems were themselves evolving into new versions, so in effect 3DELRR was chasing a moving target. In turn, monitoring the progress of an acquisition program might require PARCA AOs to query the program offices for related programs, perhaps residing in other services.

Thinking

After collecting information, we applied two approaches to developing FAAOs: a *direct approach* and a *data-driven approach*.

The direct approach evaluates the assumptions and their implications directly. Can we design a system that meets these requirements? Is the threat environment stable enough to warrant this investment? What would go wrong if the assumptions aren't met, and how might OSD monitor the status of these assumptions? The direct approach has been written about in one way or another in all of our references on FAs.

While the foundational assumption comes first in a direct approach, in the data-driven approach, the associated observable is the starting point of the analysis. In this approach, we think about what data are available on the program, what issues those data are revealing (or perhaps concealing), and what data the writer would like to have to obtain clear resolution. The AMPV program provides a good example. Like all program offices, AMPV reports regularly on how many units they plan to build each year. The annual totals are reported in quarterly DAES reports and annual SARs, as well as in the APB and AS. However, although the AMPV has five variants, only the AS said how many of each variant of those planned vehicles would be built each year.[6] The DAES reports and the SARs, which are continually updated, track the total number of vehicles, but not separately by variant. Having continuous data on variants would tell analysts a great deal about whether the program is sticking to plan or some variants are being delayed. If, for example, the mortar carriers were delayed, the program could report that it is in good health when, in fact, there is a serious unresolved problem the program office could be choosing not to reveal.

Final Deliverable

For the five programs, one of our deliverables was a memo and the other four were sets of briefing charts.[7] Each document contains a table like Table 1 from AMPV with simple instructions for the AO who would follow the program quarterly, along with general information to back up the chart and provide context on the program.

---

[6] The year-by-year totals for each variant didn't sum to the full program totals for all variants in the AS's table. Therefore, the only document we found that touched on this question beyond what is currently under contract had two contradictory answers for what the Army plans to buy.

[7] As some of the FAAOs were marked For Official Use Only (FOUO), we did not attach them to this paper to allow it to circulate more easily. All five can be found by contacting the authors of this paper.

Table 1. FAAO Table for the AMPV

Foundational Assumptions	Associated Observables
The AMPV program takes currently used mission equipment and mounts it onto a proven chassis that is larger and more capable than the original. The design process is low-risk and easily understood.	[The observable in this box was For Official Use Only and has been removed.]
	Monitoring technical measurements can show trends, and three important ones for any ground vehicle are weight, horsepower, and electrical power. For each variant, find the current value and target for each variable and plot them as a function of time. (Note: 5 variants × 3 technical parameters × (actual + goal) = 30 numbers each quarter.)
	The number of each variant produced and projected by year should be tracked. If the plan shifts some variants earlier and others later, this suggests a problem in the variant with the delayed production. The mix has presumably been set so each brigade can replace all M113s with AMPVs at once, preventing the need to support both simultaneously; a delay in any variant would change this.
AMPVs are one-for-one replacements of existing M113s currently in the ABCTs.	The total number of each type of vehicle in the program's plan should be monitored. Changes here indicate this is no longer a one-for-one replacement program.

Each quarter, the PARCA AO assessing a program should read the FAAOs for that program (which should reflect the status of the program the last time OSD examined it) and follow up their reading in two ways. First, they should attempt to collect the data that the FAAOs call for and add them to the data that have been collected in past quarters to see how they are trending. Second, they should reflect on the program described in the FAAOs and consider if the essence of the program has changed. If either the data or the reflection suggest that major changes in the program have occurred, PARCA should notify management and encourage a deep dive into the state of the program.

The reflection step is important. It is likely that the last time OSD leadership thought about this program was at a milestone that could have been several years in the past.[8] The last review may have corresponded to a requirements document, an AS, a TEMP, a set of briefing charts, a set of FAs, or more. To know if the program has changed, the obvious method would be to read all of those documents and see if they are still correct, but this is too much effort to perform quarterly. The author of the FAAO memo should capture the understanding of the program at the time of this review and report it all in a short document. For example, if everything in the FAAO memo is about aerodynamic challenges that have remained under control while the AO is now hearing about major challenges in software development, it is time for senior managers to investigate the program again, as the challenges they are facing today were not anticipated when it was last reviewed. The reflection stage calls for the AO's judgment, and the writer of the FAAOs must make sure they have given the AO enough information to allow them to exercise that judgment.

Lessons Learned From Our Five Sets of FAAOs

Our five sets of FAAOs were about programs that varied in many ways, including technical difficulty, phase of development, interoperability requirements, level of

[8] It is possible that no political appointees in OSD were in their jobs at the last milestone and that none have ever thought about this program, for which they are now responsible.

classification, and service. All of these differences matter, and the purpose of this section is to discuss them.

Table 2 contains data on each of our programs. Research, Development, Test and Evaluation (RDT&E) Funding Fraction is a variable we designed. Using the data source in the right-hand column, it takes the RDT&E appropriations in base year dollars and reports what fraction of those dollars were appropriated before fiscal year (FY) 2018.[9]

Table 2. FAAO Program List

Program	Service	Milestone B Date	Milestone C Date	RDT&E Funding Fraction before FY 2018	Source
AMPV	Army	December 2014	February 2019	53%	Dec 2017 SAR
3DELRR	Air Force	September 2014	June 2022	66%	Apr 2018 APB
Columbia	Navy	November 2016	NA	64%	Dec 2017 SAR
GPS IIIF	Air Force	NA	March 2020	0%	PB 2019 AF RDT&E *J Book* Volume 2 PE 1203269F (page 889)
DDG Flight III & AMDR	Navy	September 2013	April 2017	87%	Dec 2017 SAR (AMDR Only)

In the rest of this section, we discuss our learning process and sum up the lessons at the end.

Program Stage

Conceptually, FAAOs are connected to a baseline because the question we are asking is what assumptions must be made to meet that baseline. The different programs, with their different stages, made that question more or less complicated.

AMPV

Our first set of FAAOs was completed in December 2017. AMPV has not yet been fielded, but it received its first procurement funding in 2018 to begin LRIP. Our research showed two different things, one that is comforting to oversight and another that is not.

Overall, everything seems to be in order. These vehicles are mostly derivative designs that are relatively simple. The total RDT&E funding is about $1 billion, but that funding will design five different vehicles.

In the short term, however, things are not comfortable at all because of production difficulties. Because these vehicles are supposed to be simple, the Army believes they can deploy them to Eastern Europe quickly, but the facility where BAE is planning to manufacture them is not currently capable of the notional production rate. This delay and the

9 Normally a budget justification book wouldn't be sufficient for calculating RDT&E budget fraction because it reports only then-year dollars and doesn't break out year-by-year to allow conversion to a base year. However, since our source showed no funds were appropriated to GPS IIIF before FY2019, the fraction is exactly zero.

expenditure required to eliminate it are not significant on the scale of the program as a whole, but it is a serious issue for the program today. Whether or not that should be considered in the FAAOs is still an open question. To date, all guidance on FAs has been at the level of NM breaches, and they have been required to be program-specific. The assumption here is that there is sufficient capacity to build the required hardware, which is an essential condition for success for every program that produces hardware.

3DELRR

Our second set of FAAOs was for 3DELRR; it was presented to PARCA in February 2018. Table 2 shows that 3DELLR had its Milestone B review in 2014, but this is misleading. Then-USD(AT&L) Frank Kendall approved entry into EMD with a Milestone B decision in September 2014, apparently before an APB was finished. However, that initial development contract award to Raytheon was held up for 2.5 years because of bid protests by two competitors (Lockheed Martin and Northrop Grumman) and a lawsuit filed by Raytheon. The EMD phase began in earnest in May 2017, when the contract was again awarded to Raytheon.

This program was effectively initiated too early for FAAO development. A typical program would have released at least four SARs by the time we conducted this analysis, dated December 2014 through December 2017; 3DELRR had produced zero. 3DELRR was selected for our study because it was re-emerging. When we worked on our FAAOs, the program had no baseline, although one would later be approved by the Air Force's acquisition executive in April 2018. Still, the data source and thinking processes enabled us to produce a set of briefing charts that could help inform future AOs by highlighting some potential difficulties in the program.

Columbia

The FAAOs on Columbia were completed in April 2018. Milestone B took place in November 2016, and lead ship construction is scheduled to begin in October 2020. Long-lead items are already being built. Like AMPV, Columbia was well positioned for this analysis. The program's mission and requirements are clear, as is the budget.

GPS IIIF

We delivered FAAOs on GPS IIIF in July 2018, but this program is premature. We did have a requirements document, but all the documents we read, with one exception, suggested that this program is not a major change from the GPS III program that precedes it. The sole exception was the set of cost numbers we found in the Defense Acquisition Management Information Retrieval (DAMIR) system's "PB" section, which is not an official report. However, these numbers were high enough to give us pause because they predicted that each GPS IIIF satellite would cost about 1.5 times the cost of a GPS III satellite, and the total RDT&E costs were projected to be similar to the costs on GPS III. At this stage in the program, it was difficult for us to understand the reason for all of those extra costs. We were able to identify the assumptions that were being discussed, but there is a mystery in this program that we could not uncover at this stage.

DDG-51 Flight III and AMDR

This program consists of two MDAPs: DDG-51, an established system that has produced many of the Navy's current ships, and AMDR, a new radar that has not been fielded yet. The DDG-51 flight III ships are variants of the older ships in the *Arleigh Burke* class.[10] The AMDR program will produce the new SPY-6 radar system, which should make these new ships more capable than their predecessors. We delivered our FAAOs in July 2018. While this combination of MDAPs is not technically a program, we adopted the word *program* to describe it because that seems appropriate; the Navy is planning to buy 22 ships that are unlike any previous ships.

While AMDR has already passed Milestone C, it is worth noting that the radar has not yet been tested at sea, nor has it been tested with multiple arrays, even though the operational configuration is to consist of four arrays working together. With all of the requirements and costs laid out but the system not yet in production, this was a good time to identify the program's FAAOs.

Interoperability Requirements

The programs we studied varied considerably in how interoperable they need to be. At one extreme is the Columbia class submarine, and at the other is 3DELRR. Interoperability is difficult and needs to be considered when identifying FAAOs. A system that must be interoperable may perform the same way in two separate instances and be useful the first time but not the second, because of how other systems interact.

Columbia class boats are expected to remain hidden while waiting for an order to strike, orders that can only come via a limited number of channels. From an interoperability standpoint, this is about as isolated as a system can be.

At the other extreme is 3DELRR, which is envisioned to provide data that contribute to a picture of everything in the air over a theater. Other data will come from other services, and possibly also civilian agencies and international partners. The picture generated will include aircraft from all of our services as well as allies and rivals. Similarly, that picture may be used by joint headquarters and operators from every service and our allies. To understand whether this system will be useful requires looking at how it interacts with all those other systems, and our FAAOs for 3DELRR call for monitoring the progress of its interfaces with two other systems.

Classification

Our FAAOs on the Columbia class submarines were made more difficult because of classification issues, and we have some concern that because of the classification, we may have missed something that is relevant. There was a lot written about the "coordinated stern," but in the unclassified world, it was rarely more than concern. We found a document classified at the SECRET level that illuminated this conversation considerably, although it was still difficult to incorporate that information into the unclassified final product PARCA requested.

We also began the process of working on FAAOs for the Air Force's Long Range Stand-off Weapon (LRSO), which is a new nuclear-armed cruise missile. While the existence of the program is not only public but has been advertised, once we dug below the surface we found that everything was classified beyond SECRET. We agreed along with PARCA that even if we could get read-in to learn about the program, PARCA's analysts

[10] Previous ships in the class are already subdivided into three flights called I, II, and IIa.

would not be able to take advantage of a document that was classified beyond SECRET, so we went no further.

None of the other programs we studied seemed to have significant problems with classification issues. For every one of them, we reviewed some documents that were classified at the SECRET level, but we were able to produce unclassified final products.

Deliverable Format

As stated earlier, IDA delivered the first set of FAAOs in a memo and the others in briefing charts. The sponsor was happier with the briefing chart of our second set than the memo of the first. We suspect there were more differences than the format, but we decided to continue with the format that was well received. If we are asked to deliver additional sets of FAAOs, we may revisit this decision, as we expect that a written document is more useful than a set of charts. We will also consult with the AOs and see how they were used.

Resources Required to Write FAAOs

The first two sets of FAAOs, for AMPV and 3DELRR, each required about 200 hours of researchers' time to put together. Once we had a better feel for the process, the number of hours dropped to between 80 and 100 per set.

There were disagreements among the PARCA and IDA staff members about which sets of FAAOs were best. Some of the disagreements stemmed from the requirement that FAs be program-specific, some from the format of the IDA deliverables, and others over program-specific issues.

Summary of Lessons Learned

It is never too late in a program's life cycle to attempt to identify FAAOs, as long as the government is planning to spend more money on the program; however, the reverse is not true. Writing FAAOs only makes sense once there is a record of what the program is supposed to deliver. 3DELRR had a Milestone B review, but the program stalled so soon after that it never even had an APB. An FAAO writer is not going to do a more complete job than a DAB, so the FAAOs should wait until after that review. GPS IIIF seemed to be much too early.

It only makes sense to generate FAAOs at a level of classification that is high enough to know what is going on in the program. There is no reason that highly classified FAAOs could not exist, but they would also require both an audience and storage containers (i.e., safes) that are cleared at that level, and that is not how PARCA has operated.

Conclusion

Time will tell if any of these sets of FAAOs turn out to be useful. If AOs read the FAAOs quarterly and track the observables we recommended, that is one form of success. If the data are consistent with program health, then we will have done better still. Our first hope is that the five programs we studied will match their baselines in cost, schedule, and performance. If any one of them slips, we hope that the FAAOs will allow the OSD to detect those problems early.

References

Acquisition program baseline (APB): Three-dimensional expeditionary long-range radar (3DELRR). (2018, April). Washington, DC: Defense Acquisition Management Information Retrieval (DAMIR).

Acquisition strategy for armored multi-purpose vehicle, ACAT 1D, supporting Milestone B and engineering and manufacturing development phase, version 1.6. (2014, December). Washington, DC: Program Manager, Armored Brigade Combat Team. For Official Use Only.

Arena, M. V., Blickstein, I., Doll, A., Drezner, J. A., Kallimani, J. G., Kavanagh, J., … Wong, C. (2013). Framing assumptions. In *Management perspectives pertaining to root cause analyses of Nunn-McCurdy Breaches (Vol. 4). Program manager tenure, oversight of Acquisition Category II programs, and framing assumptions* (pp. 83–104). Retrieved from http://www.jstor.org/stable/10.7249/j.ctt5hhttw.12

Arena, M. V., & Mayer, L. A. (2014). *Identifying acquisition framing assumptions through structured deliberation* (TL-153-OSD). Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/pubs/tools/TL153.html

Arnold, S. A., Byun, J. S., Cloud, H. A., Gallo, A. O., Gonwa, M. W., Harmon, B. R., ... Bronson, P. F. (2010). *WSARA 2009: Joint Strike Fighter root cause analysis* (IDA P-4612). Alexandria, VA: Institute for Defense Analyses.

Bailey, J. W., & Frazier, T. P. (2018). *Operationalizing framing assumptions using case studies* (IDA Paper P-9178). Alexandria, VA: Institute for Defense Analyses.

Defense Acquisition Board, Milestone B. (2014, December). *Armored multi-purpose vehicle program (AMPV)* (Presentation). For Official Use Only.

DoD. (2015, January). *Operation of the defense acquisition system* (DoD Instruction 5000.02). Retrieved from https://www.acq.osd.mil/fo/docs/500002p.pdf

Department of the Air Force. (2018, February). *DoD fiscal year (FY) 2019 budget estimates: Justification book* (Vol. 2). Washington, DC: Department of the Air Force.

GAO. (2017). *Defense acquisitions: Assessments of selected weapon programs* (GAO-17-333SP). Retrieved from https://www.gao.gov/products/GAO-17-333SP

Husband, D. M. (2013). *Information paper on framing assumptions*. Retrieved from https://www.acq.osd.mil/parca/docs/2013-09-13-information-paper-framing-assumptions.pdf

PARCA. (n.d.). Root cause analyses. Retrieved from https://www.acq.osd.mil/parca /references.shtml

Selected acquisition report (SAR): Air and missile defense radar (AMDR). (2017, December). Washington, DC: DAMIR.

Selected acquisition report (SAR): Armored multi-purpose vehicle (AMPV). (2017, December). Washington, DC: DAMIR.

Selected acquisition report (SAR): SSBN 826 Columbia class submarine (SSBN 826). (2017, December). Washington, DC: DAMIR. For Official Use Only.

Weapon Systems Acquisition Reform Act of 2009 (WSARA), Pub. L. No. 111-23.

Acknowledgments

The first person to thank is David Cadman, formerly of PARCA and now Acquisition Enablers, who sponsored this work. We also thank our technical reviewers, Patricia Bronson and Nancy Huff, for their technical reviews. Dr. Bronson's high praise confirmed that some readers will find this work worthwhile. Dr. Huff's review improved the paper.

Do We Need a Different Approach to Statistical Analysis of Research, Development, Test, and Evaluation Cost Growth?

David L. McNicol—joined the Department of Defense (DoD) in 1982. From 1988 until 2002, he was a Deputy Director of Program Analysis and Evaluation (PA&E). Earlier, McNicol taught at the University of Pennsylvania and the California Institute of Technology. He holds a BA in economics from Harvard and an MS in management and a PhD in economics and finance from MIT. Employed at the Institute for Defense Analyses (IDA) since his retirement from the DoD, he became director of the Cost Analysis and Research Division in 2006. Still at IDA, McNicol stepped down in 2012 to return to his previous role as a research staff member. [dmcnicol@ida.org]

Abstract

McNicol (2018; hereafter *Acquisition Policy*) obtained remarkably strong statistical results for a simple model of cost growth in Major Defense Acquisition Programs (MDAPs). Following previous studies, *Acquisition Policy* used Program Acquisition Unit Cost (PAUC), the numerator of which is the sum of Research, Development, Test, and Evaluation (RDT&E) cost and procurement cost. This paper asks whether the model used by *Acquisition Policy* characterizes RDT&E cost growth and growth in procurement cost individually as well as it does PAUC. It does not. As would be expected, the results for Average Procurement Unit Cost (APUC) are very similar to those for PAUC and are marginally stronger statistically. The results for RDT&E also are quantitatively similar to those for PAUC, but the explanatory power of the model is far lower, suggesting either much greater variability in RDT&E cost estimates or flaws in the model as applied to RDT&E. The paper concludes with suggestions for improving models of RDT&E cost growth.

Introduction

McNicol (2018; hereafter *Acquisition Policy*) obtained remarkably strong statistical results for a simple model of root causes of cost growth in Major Defense Acquisition Programs (MDAPs). That study considered Program Acquisition Unit Cost (PAUC). Program acquisition cost is the sum of Research, Development, Test, and Evaluation (RDT&E) cost and procurement cost (that is, the cost of buying a system once it has been developed). PAUC is acquisition cost divided by the number of fully configured units acquired. Procurement typically is four to five times as large as RDT&E. Consequently, PAUC is dominated by Average Procurement Unit Cost (APUC), which is the program’s procurement dollars divided by the number of units purchased with them. As would be expected, the model of *Acquisition Policy* works well for APUC. This paper first asks whether it also provides a solid account of RDT&E cost growth. After finding that it does not, the paper examines ways to improve the basic model.

The next section provides the minimal background needed to follow this paper. The Results for APUC and RDT&E Cost Growth section presents estimates of the model for APUC and RDT&E cost growth. The Extensions of the Funding Climate-Acquisition Policy section considers expansion of the basic model to include other variables that may help explain the competition for RDT&E funds at Milestone (MS) B.

Background

The model of *Acquisition Policy* was mainly directed to what have come to be called Errors of Inception[1]—that is, cost growth attributable to unrealistic assumptions embedded in the program's MS B baseline. The proximate causes of Errors of Inception are, by definition, characteristics of the program, e.g., the maturity of critical technologies, concurrency between development and production, and the amount of computer code to be taken from legacy systems, among others. *Acquisition Policy* argues that the root causes of these proximate causes lie in the intensity of competition for funds at MS B, marked by funding climate, as modified by acquisition policy and process. The model will be referred to here as the Funding Climate-Acquisition Policy Model.[2]

The Funding Climate-Acquisition Policy Model focuses on the competition for funds at the Service level during the Program/Budget cycle before an MDAP undergoes MS B review at the Office of the Secretary of Defense (OSD) level. The model of *Acquisition Policy* tacitly assumes that that process will consider the funding decision in terms of the acquisition cost of the program—that is, the sum of RDT&E funding post-MS B and the cost of procuring the system once it has been fully developed. This is a reasonable position, but also one that is subject to a reasonable challenge.

The challenge rests on a combination of two sets of facts. First, the program/budget process develops the Future Years Defense Program (FYDP), which, for most of the period covered by *Acquisition Policy*, included the upcoming budget year and the four succeeding years. The Services must build their FYDPs subject to hard ceilings imposed on the authority of the Secretary of Defense. Second, most MDAPs typically spend at least three years in Engineering and Manufacturing Development (EMD), which begins at MS B, and only then move into the first part of the procurement phase. Consequently, for most MDAPs at MS B, the procurement phase starts late in the FYDP or beyond it. The Services track, and on a case-by-case basis limit, planned funding for MDAPs beyond the FYDP, but these limits are softer than the controls on the FYDP period during the program/budget process. The implication of these comments is that MDAPs coming to an MS B review perhaps compete for RDT&E and procurement funds under somewhat different conditions.

Results for APUC and RDT&E Cost Growth

The model of Acquisition Policy, applied to APUC growth, is

$$Ch_{APUCi} = a_0 + a_1 Climate_i + a_2 DSARC_i + a_3 PCDSARC_i + a_4 DAB_i + a_5 AR_i + a_6 T_{boomi} + a_7 T_{busti} + e_i$$

ChAPUCi is computed by comparing the MS B baseline value of APUC—which can be thought of as a goal or a prediction—to the actual APUC reported in the final Selected

[1] This term was introduced by the Office of Program Assessments and Root Cause Analyses (PARCA) in connection with its root cause analyses. The Weapon Systems Acquisition Reform Act of 2009 established PARCA, which in 2018 was renamed the Office of Acquisition, Analytics and Policy (AAP).

[2] The Funding Climate-Acquisition Policy Model is an extension of the "speeding" model of cost growth offered by McNicol (2004, pp. 37–49). *Acquisition Policy* (Chapter 2, pp. 9–24) elaborates on these ideas.

Acquisition Report (SAR) for the program. Both the MS B baseline and the final value[3] of APUC are stated in program base year dollars. The actual value is adjusted on the basis of the MS B baseline quantity by moving up or down the cost progress curve as appropriate. The ratio of the MS B baseline value of APUC to the quantity-adjusted actual value is an estimate of what APUC growth would have been had the MS B baseline quantity been acquired.

Table 1 defines the categorical variables used in the study. The study period (Fiscal Year [FY] 1965–FY 2009) includes two complete bust-boom cycles in Department of Defense (DoD) funding. The first of the acquisition policy bins (McNamara-Clifford) does not appear explicitly in the model because it is used as the reference category. Acquisition Policy identifies the factors used to establish the break points between the acquisition policy bins and between bust and boom climates (McNicol, 2018, Chapter 2, pp. 11–13; Chapter 3, pp. 13–16).

Table 1. Categorical Variables of the Funding Climate-Acquisition Policy Model

Variable	Short Name	Period (Fiscal Years)
Climate	bust climates	1965–1982, 1987–2002
	boom climates	1983–1987, 2003–2008
McNamara-Clifford	McNamara-Clifford	1965–1969
Defense System Acquisition Review Council	DSARC	1970–1982
Post-Carlucci DSARC	PC DSARC	1983–1989
Defense Acquisition Board	DAB	1990–1993 2001–2009
Acquisition Reform	AR	1994–2000

Finally, Tboomi and Tbusti are the numbers of years the ith program spent in boom and bust years, respectively. These two variables effectively are measures of program duration. They are included in the Acquisition Policy model of PAUC growth as a rough and ready way of accounting for the cost growth due to Errors of Execution and Program Changes. Errors of Execution are errors that arise post-MS B, typically errors by government or contractor management. Program changes are unforced changes made post-MS B to increase or, in a few cases, decrease, the capabilities of the system acquired. Tboomi and Tbusti are retained in the model for RDT&E growth because long duration programs may incur RDT&E costs to develop improvements or even new variants long after the original EMD work has been completed. Finally, the term ei is a random variable that is assumed to have a constant mean and variance.

Table 2 presents the estimated parameter values and their associated p-values using growth in APUC (adjusted to the MS B quantity) as the dependent variable.[4] Given the underlying model, the intercept term is the expected average APUC growth for MDAPs

[3] For a program that is still underway, the most recent estimate (as reported in the SAR) of the final value was used.
[4] Estimates of the model's parameters for PAUC are in the appendix.

that passed MS B during McNamara-Clifford.[5] The actual average for this bin for the sample used to compute the estimates in Table is 88.7%. We expect the estimated coefficient of Climate to be negative, which it is, and the magnitude of the estimate also is reasonable. The estimated coefficient for each of the acquisition policy bins should be negative, which they are. The estimated coefficients of Tboom and Tbust should be positive (they are) and Tboom should have the larger coefficient (it does). The estimated coefficients, then, satisfy prior expectations and each except that for Tbust is significant at the 5% level or less. The estimated equation explains about 22% of the variation in APUC, which is quite high for a pooled time series-cross section dataset and a model that does not include the lagged dependent variable. In short, the statistical results cast the explanation underlying the model estimated in a favorable light.

Table 2. Estimate of the Funding Climate-Acquisition Policy Model for APUC Growth

	Coefficients	**p-value**
Intercept	74.8%***	< 0.001
Errors of Inception–Intensity of Competition for Funds		
Climate	-26.7%**	0.02
Error of Inception–Acquisition Policy		
DSARC	-58.8%***	< 0.001
PC DSARC	-46.4%**	0.004
DAB	-60.8%***	< 0.001
AR	-81.0%***	< 0.001
Errors of Execution and Program Changes		
T_{boom}	3.8%/yr**	0.03
T_{bust}	0.5%/yr	0.61

***Statistically significant at less than the 1% level.

** Statistically significant at less than the 5% level.

Note: R-Squared = 0.22, F = 5.46 (P < 0.001), N= 145. Estimated using ordinary least squares (OLS). The regression was computed using the 145 MDAPs in the database for which both APUC growth and RDT&E growth are available.

The results for RDT&E cost growth are reported In Table 3. Only the estimate of the intercept is statistically significant; the estimated equation explains only 4% of the variation in RDT&E cost growth; and the equation as a whole is not statistically significant. Low explanatory power is understandable, as estimates of RDT&E cost at MS B are generally thought to be more uncertain than estimates of procurement cost (and therefore APUC). Nonetheless, the p-values and other test statistics have nothing good to say about the Funding Climate-Acquisition Policy model as applied to RDT&E cost growth.

[5] The intercept term also will pick up the effects of non-linearities and other specification errors, omitted variables, and errors in measurement of variables that are included.

Table 3. Estimate of the Funding Climate-Acquisition Policy Model for RDT&E Cost Growth

	Coefficients	p-value
Intercept	75.4%**	0.018
Errors of Inception–Intensity of Competition for Funds		
Climate	-13.1%	0.602
Error of Inception–Acquisition Policy		
DSARC	-50.2%*	0.101
PC DSARC	-34.9%	0.309
DAB	-53.8%	0.122
AR	-33.2%	0.397
Errors of Execution and Program Changes		
T_{boom}	2.2%/yr	0.573
T_{bust}	0.8%/yr	0.662

** Statistically significant at less than the 5% level.

* Marginally statistically significant at the 10% level.

Note. R-Squared = 0.04, F = 0.763 (P = 0.619), N= 145. Estimated using ordinary least squares (OLS). The regression was computed using the 145 MDAPs in the database for which both APUC growth and RDT&E growth are available.

The coefficient estimates suggest, however, that the sensible course may not be to scrap the Funding Climate-Acquisition Policy model as applied to RDT&E cost growth but to incorporate within it additional variables. This suggestion is conveyed by the fact that the estimated coefficients all have the expected sign and reasonable magnitudes. This combination—coefficient estimates that are reasonable but not significant—could arise if there is one or more important variables missing from the model estimated and not highly correlated with variables that are included. Given this possibility, the relevant question is: What are these omitted variables? The discussion of this question that follows is exploratory in character. The underlying objective is simply to gauge whether the Funding Climate-Acquisition Policy model as applied to RDT&E cost growth shows substantial signs of promise.

Extensions of the Funding Climate-Acquisition Policy

Within the logic of the Funding Climate-Acquisition Policy model, a relevant "missing variable" would be one that influences competition for acquisition funding during the POM cycle or a change in acquisition policy not captured in the policy bins used. These are discussed in turn.

One obvious consideration in competition for funds is the priority that the sponsoring Service places on a program. Only very rarely does a Service's ranking of its investment priorities become public, however. Consequently, a proxy for program priority must be found if it is to be included in the model. One useful point that can be made in this connection is that each of the Services affords high priority to platforms that have a central role in its main warfighting missions. The F-22, then, was a very high priority program for the Air Force, and, similarly, the M-1 Abrams tank, the DDG-51 Arleigh Burke destroyer, and the V-22 were very high priority programs for the Army, Navy, and Marine Corps, respectively. The data set used in this paper contains 31 MDAPs with both APUC and RDT&E cost growth estimates that acquired a platform central to one of the Services' warfighting mission. These were assigned a value of 1 in the categorical variable called High Priority; all other programs in the dataset were assigned a value of 0 for this variable.

Another obvious consideration is program size—that is, in the present context, the amount of RDT&E funding requested at MS B, which will be treated as distinct from priority. For a given priority level, large programs presumably face stiffer competition if for no other reason than that they attract opposition from programs they would displace. Consequently, again at a given priority level, we would expect larger programs to have higher RDT&E cost growth than smaller programs.

There are two problems with including program size in the model, one statistical and the other a limitation of the database available for this paper. The statistical problem is that program size is correlated with priority. Size is not always a reliable guide to priority, however; there are some large programs (for example, Family of Medium Tactical Vehicles) that have a relatively low priority for funding purposes and some relatively small programs (for example, Javelin) that had a high priority. Accordingly, it is reasonable to include both priority and size in the model, although it may prove to be impossible to capture their separate effects.

The second problem is that the database included the RDT&E cost projected at MS B for only about one-third of the programs, and the resources required to collect the data for the other two-thirds were not available.[6] One way to ameliorate this problem is to include in the model the number of MDAPs that passed MS B each year (for each Service and joint programs). This variable (#Competing) should provide a rough measure of the extent of competition for RDT&E funds in the given year. Another is to include categorical variables for satellites, which have large RDT&E funding requirements, and ships, which at MS B require relatively little RDT&E funding. We would expect the estimated coefficient for satellites to be positive and that for ships to be negative.[7] Of course, categorical variables for satellites and ships pick up several differences, so even if these expectations are met, we cannot confidently attribute the effects to the size of RDT&E funding.

The discussion now turns to additional acquisition policy and process variables that might be incorporated in the model. The policy variables of the model of Acquisition Policy mark time periods. Within the first one to three years of each of these periods, several major changes in acquisition policy were made, most of which persisted to (and past) the end of this study (FY 2009). It is arguable that no major acquisition policy changes were implemented after the first few years of each period.[8] The policy variables currently in the model could be replaced by categorical variables marking the major policy changes. This

---

[6] MS B funding for MDAPs that passed MS B in FY 1997 and later years is readily available on the Defense Acquisition Management Information Retrieval (DAMIR) system. Funding data for MDAPs that passed MS B before FY 1997 are available in the Selected Acquisition Reports (SARs), but their extraction for programs that began EMD can be difficult and require searching information sources other than the relevant SARs.

[7] There is nothing novel about using categorical variables for commodity types in a statistical analysis of cost growth. The novelty here is in the suggestion that differences in cost growth across various commodity types reflect the amounts of RDT&E they require at MS B.

[8] The main challenge to this proposition is the changes adopted by the DoD in 1986 and 1987 as a result of presidential direction and legislation that implemented some of the recommendations of the Packard Commission report and the Goldwater-Nichols Act. While adopted in 1986 and 1987, these changes were not effectively implemented until about 1990. See *Acquisition Policy* (Appendix B, pp. B-10–B-11).

would be a considerable amount of work (there would be two to three dozen such variables) with little prospect of gain, because the changes cluster in distinct sets and none is directed especially to RDT&E cost growth. Acquisition Policy assumed that the policy variables defined in terms of distinct periods was a reasonable if imperfect way to represent changes in acquisition policy over the study period.

The results obtained when the four additional variables are included in the model are presented in Table 4 (Recall Table 3) that the estimated coefficient of only one of the seven variables of the basic model is even marginally statistically significant. In such a circumstance, when additional variables are introduced, it is often found that the signs of the estimated coefficients of the original model change and estimated magnitudes can change dramatically. Such an outcome would have ended discussion of the Funding Climate-Acquisition Policy construct as a useful model of RDT&E cost growth.

Table 4. Estimate of an Extended Funding Climate-Acquisition Policy Model for RDT&E Cost Growth

	Coefficients	p-value
Intercept	77.0%	0.021**
Errors of Inception–Intensity of Competition for Funds		
Climate	-12.8%	0.613
High Priority	-14.5%	0.529
#Competing	1.0%	0.741
Satellites	63.0%	0.113
Ships	-37.1%	0.216
Error of Inception–Acquisition Policy		
DSARC	-41.2%	0.179
PC DSARC	-31.6%	0.363
DAB	-52.8%	0.128
AR	-24.3%	0.547
Errors of Execution and Program Changes		
T_{boom}	0.2%	0.957
T_{bust}	1.0%	0.615

** Statistically significant at less than the 5% level.

Note. R-Squared =0.08, F = 1.060 (P = 0.399), N= 145. Estimated using ordinary least squares (OLS). The regression was computed using the 145 MDAPs in the database for which both APUC growth and RDT&E growth are available.

Those are not the results obtained, however. First, the estimated coefficients for the basic Funding Climate-Acquisition Policy model all have the expected sign, and (except for T_{boom}) their magnitudes do not change drastically. Second, the coefficient of each of the additional variables introduced has the expected sign and a reasonable magnitude, although only that of satellites approaches statistical significance. That is a modest amount of evidence, but enough to suggest that there may be merit in trying to understand more fully the competition for RDT&E funding at MS B and to obtain good measures of the key variables involved.

Concluding Comment

There currently is no consensus model of RDT&E cost growth, and the only contender in the lists seems to be the Funding Climate-Acquisition Policy model.[9] So, in view of the results provided above, the answer to the question asked in the title of this paper is part "no" since the Funding Climate-Acquisition Policy model provides a reasonable basis for further work, and "yes" in that much remains to be done for that model to provide a solid statistical account of RDT&E cost growth of MDAPs.

References

McNicol, D. L. (2004). *Cost growth in major weapon procurement programs*. Alexandria, VA: Institute for Defense Analyses.

McNicol, D. L. (2018). *Acquisition policy, cost growth, and cancellations of major defense acquisition programs* (IDA Report R-8396). Alexandria, VA: Institute for Defense Analyses.

Younossi, O., Arena, M. V., Leonard, R. S., Roll, C. R., Jr., Jain, A., & Sallinger, J. M. (2007). *Is weapon system cost growth increasing?* (MG588). Retrieved from https://www.rand.org/pubs/monographs/MG588.html

Acknowledgments

This paper was reviewed by Dr. Gregory A. Davis of the Institute for Defense Analyses. The author is grateful for his comments on this paper and his continuing support of the project of which this paper is a part.

Appendix. Estimate of the Funding Climate-Acquisition Policy Model for PAUC Growth

Table 5. Estimate of the Funding Climate-Acquisition Policy Model for PAUC Growth

	Coefficients	p-value
Intercept	100.2%***	< 0.001
Errors of Inception–Intensity of Competition for Funds		
Climate	-30.4%**	0.046
Error of Inception–Acquisition Policy		
DSARC	-81.5%***	< 0.001
PC DSARC	-67.7%***	0.001
DAB	-84.7%***	< 0.001
AR	-101.1%***	< 0.001
Errors of Execution and Program Changes		
T_{boom}	4,4%/yr	0.061
T_{bust}	-0.07 %/yr	0.952

** Statistically significant at less than the 5% level.
Note. R-Squared = 0.20, F = 5.047 (P < 0.001), N= 145. Estimated using ordinary least squares (OLS). The regression was computed using the 145 MDAPs in the database for which both APUC growth and RDT&E growth are available.

[9] Younossi et al. (2007) is a statistical study of RDT&E cost growth. It attempts to answer the question of whether cost growth, particularly RDT&E cost growth, has been increasing since the 1970s. It does not attempt to account for either the proximate or root causes of RDT&E cost growth.

Quantifying the Year-by-Year Cost Variance of Major Defense Programs

David M. Tate—joined the research staff of the Institute for Defense Analyses' (IDA) Cost Analysis and Research Division in 2000. Prior to that, he was an Assistant Professor of Industrial Engineering at the University of Pittsburgh, and the Senior Operations Research Analyst (Telecom) for Decision-Science Applications, Inc. At IDA, he has worked on a wide variety of resource analysis and quantitative modeling projects related to national security. These include an independent cost estimate of Future Combat Systems development costs, investigation of apparent inequities in Veterans' Disability Benefit adjudications, and modeling and optimization of resource-constrained acquisition portfolios. Dr. Tate holds bachelor's degrees in philosophy and mathematical sciences from the Johns Hopkins University, and MS and PhD degrees in operations research from Cornell University. [dtate@ida.org]

Michael R. Guggisberg—is a Research Staff Member in the Strategy, Forces, and Resources Division at the Institute for Defense Analyses. He earned his PhD in economics at the University of California, Irvine in 2017. Prior to that, he was a squad leader in the United States Marine Corps Reserve, with tours in Iraq and Afghanistan. [mguggisb@ida.org]

Abstract

To a first approximation, acquisition programs never spend what they originally said they would spend when they began. In fact, the uncertainty in initial funding profile estimates is much larger than is generally understood; the possibility of program cancellations, restructurings, truncations, and block upgrades are often not accounted for. Worse yet, all of this uncertainty arises in a context in which programs must fit within annual budgets—it is not enough to only spend as much as you said you would; you must also spend it when you said you would, or problems ensue.

In 2018, we presented a methodology that uses historical program outcomes to characterize the year-by-year development and procurement cost risk associated with a major acquisition program. That work used functional regression to characterize changes in development profiles, modeled as Weibull curves. This paper improves and extends that work, using a novel application of Functional Principal Component Analysis (FPCA) to characterize the distributions of future RDT&E and Procurement profiles of both new and continuing acquisition programs.

Introduction—The Research Program

Recap of Prior Work

To a first approximation, acquisition programs never spend what they said they would when they began. In fact, the error bars around an initial cost estimate are much larger than is generally understood once program cancellations, restructurings, truncations, and block upgrades have been accounted for. Worse yet, all of this uncertainty arises in a context where programs must fit within annual budgets—it is not enough to only spend as much as you said you would; you must also spend it when you said you would, or problems ensue.

We have developed a methodology to characterize the year-by-year budget risk associated with a major acquisition program. This methodology can be applied to both development costs (Research, Development, Test, and Evaluation, or RDT&E) and procurement costs, and can be extended to understand the aggregate affordability risk of portfolios of programs. The method allows resource managers to estimate annual budget risk levels, required contingency amounts to achieve a specified probability of staying within a given budget, and a host of other relevant risk metrics for programs. It also allows policy makers to predict the impact on program affordability of proposed changes in how contingency funds are managed.

Many researchers have studied cost growth in major defense programs. The vast majority of this work has looked at either the ratio of eventual total cost to the originally estimated total cost, or the increase in some unit cost measure. Neither of these approaches addresses the problem that funds are authorized year-by-year, and that the affordability of a program or portfolio requires having enough obligation authority in each year to do the work needed over the next few years.

In Tate, Coonce, and Guggisberg (2018), we introduced an analytical approach for quantifying how likely a given set of programs is to fit within a projected budget over a planning horizon. This paper improves and extends that previous work.

To recap the approach: using historical Selected Acquisition Report (SAR) data, we look at how the profile of annual funding changed from initial estimates to actual authorized amounts, looking only at programs that are no longer spending. We do this separately for RDT&E costs and procurement costs. Our approach is agnostic about causes of these changes—the possibility that a program might be cancelled, or that the buyer might decide to triple the quantity or modify the design, is treated as part of the uncertainty to be accounted for in forecasting future budget demands. Posterior estimates of the distribution of possible cost profile outcomes are generated as a function of initial budget estimates, attributes of the program (e.g., that it is a joint program, or an aircraft program), and environmental conditions (e.g., that overall defense budgets are relatively tight at the moment).

Reminder: Desirable Outputs of a Model

Given a planned program (or set of programs—we'll get to that later) and a budget, resource managers would very much like to answer questions such as the following:

- What is the distribution of funding the program will receive in year t = 1, 2, ?
- What is the probability that the program will receive more funding in year t than is currently budgeted, for t = 1, 2, ?
- How many total contingency dollars would be enough to achieve a given probability that the current budget plus the contingency is enough to fund the program over the Future Years Defense Plan (FYDP)?
- What is the probability that the program will receive at least $X less than planned over the FYDP, for various values of X?

The goal of our research is to develop empirical models, based on historical program attributes, environments, and outcomes, that will allow us to answer questions like these. To do that, we need a few specific tools:

- A way to describe funding profiles mathematically;
- A list of program attributes and environmental factors that help predict program outcomes;

- A statistical model to estimate the probability distribution of final funding profile shapes, given the initial or midlife funding profile, environmental factors, and other program attributes;
- A mathematical characterization of how well the shape tends to fit actual data; and
- Historical data on program initial plans, midlife plans/outcomes and final outcomes.

Tate et al. (2018) illustrated this approach using Weibull curves to model RDT&E cost profiles of major defense acquisition programs (MDAPs). That work used *functional regression*, in which the shape of the realized cost profiles is assumed to have a particular functional form. Specifically, we assumed a Weibull distribution for the (scaled) initial and final profiles. That approach proved to be unsatisfactory in a couple of ways.

For one thing, many historical programs have realized RDT&E spending profiles that do not look like a single Weibull profile. For example, the Advanced Medium-Range Air-to-Air Missile (AMRAAM) and DDG-51 Destroyer programs each consists of a sequence of block upgrades (or new developments) of the product. They behave, in essence, like multiple sequential acquisition programs under a single funding line.

Alternatively, some RDT&E programs function more like services contracts than like product development contracts, consisting of a level of effort to improve capabilities over time, rather than one or more development and production projects with a discrete beginning and end. Ballistic Missile Defense System (BMDS) is perhaps the best exemplar of this approach, but there are others. The Evolved Expendable Launch Vehicle program was originally designed as a program to procure a set number of launch vehicles for satellites. Now called the National Security Space Launch program, it represents ongoing modernization and improvement of space launch capabilities. Figure 1 shows annual RDT&E funding for BMDS.

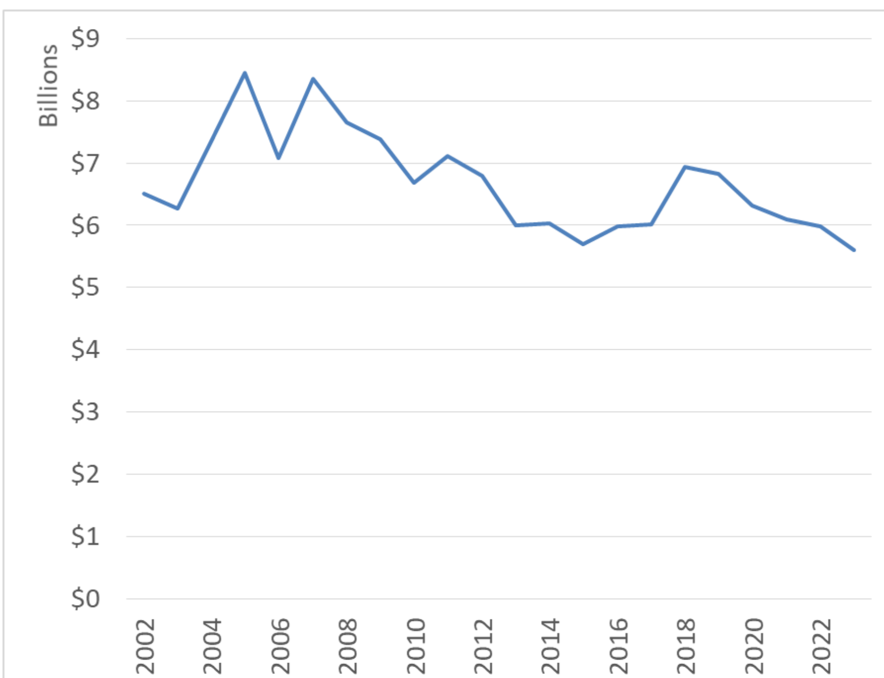


Figure 1. **Annual RDT&E Funding for BMDS**

Improved Modeling Approach

In the previous section, we noted that parametric functional families (and Weibull curves in particular) lack the flexibility to capture the variety of shapes shown by historical funding profiles. Our examination of historical spending patterns suggests that this is true not only of RDT&E profiles, but also of procurement profiles. As a result, we have adopted a nonparametric approach to characterizing profile shapes.

Instead of treating the year-by-year outcomes as having some complicated joint distribution, we will instead use nonparametric techniques from functional data analysis to treat the individual year-by-year outcomes as having been generated by some (noisy) underlying set of basic profile shapes, and then think about probability distributions over the parameters of those generating functions. Brown et al. (2015) provide a good summary of past approaches. Our first attempts (reported in Tate et al., 2018) attempted to fit the cost profiles to a pre-specified parametric functional family such as Weibull curves. Our revised approach uses a more flexible methodology based on Functional Principal Component Analysis (FPCA), described next.

Functional Principal Component Analysis

Define the set of programs to be $\{1,2,\ldots,I\}$. Let $C_{ij}(t)$ represent the planned spending for program i in fiscal year t as estimated in fiscal year j. Elements of $C_{ij}(t)$ reflect predictions if $t > j$ or actual spending for year t if $t \leq j$. This definition of a program can capture all stages of a program's lifecycle. A program is defined as initial if $C_{ij}(t) = 0$ for all $t < j$. A program is defined as completed (possibly cancelled) if $C_{ij}(t) = 0$ for all $t \geq j$. All other programs are considered "midlife"—their cost profiles are partly realized, but not yet completed.

A set of functional observations is notoriously difficult to summarize, since they are elements of an infinite dimensional space. One tool for summarizing such collections of functions is FPCA. FPCA is the infinite dimensional generalization of Principal Component Analysis, which is a methodology that provides an orthonormal basis to represent vectors in a finite dimensional Euclidean space. Its principal use in statistics is to find transformations of the predictive variables that are approximately independent in their effects on the outcomes of interest.

The FPCA process identifies a mean function $\mu(t)$ and a set of K eigenfunctions, $\xi_k(t)$, for $k \in \{1,2,\ldots,K\}$, that represent recurring patterns of deviation from the mean function. The eigenfunctions form an orthonormal basis in the L^2 Hilbert space (Yao, Müller, & Wang, 2009). The FPCA basis explains more variation than any alternative basis expansion when using a fixed K number of eigenfunctions.

Given an observed historical cost profile $C_{ij}(t)$, FPCA represents the profile as a weighted sum of the eigenfunctions, plus the mean function:

$$\log\left(C_{ij}(t)\right) = \mu(t) + \sum_{k=1}^{K} \omega_{ijk}\xi_k(t) + \epsilon_{ij}(t).$$

That is, ω_{ijk} (usually called "FPCA scores") function as weights on the eigenfunctions for generating log cost profiles in the new basis. The mean function and eigenfunctions are common for all programs in all stages of their life. The kth FPCA score is specific to program i in fiscal year j. The discrepancy from using a fixed finite K number of eigenfunctions is represented by $\epsilon_{ij}(t)$.

We apply FPCA separately to RDT&E and procurement cost profiles, for several reasons:

- RDT&E spending profiles and procurement spending profiles tend to have different shapes;
- RDT&E profiles and procurement profiles are offset in time, with procurement spending beginning later; and
- RDT&E and procurement fall under different "colors of money," and must therefore be separately evaluated against their respective budgets.

We use $K = 3$ for RDT&E and $K = 2$ for procurement. The value K was chosen such that the cumulative fraction of explained variation was over 90%. Applying FPCA generates mean profiles and principal eigenfunctions for both RDT&E and procurement. The shape fits are done using profiles that have been scaled in duration such that they begin at time 0 and end at time 1. The mean and principal eigenfunctions are shown in Figure 2.. The mean function is in the upper two subplots and the first K principal eigenfunctions are in the bottom two subplots. Note that while the mean RDT&E profile does have a roughly Weibull shape, the FPCA method can also account for more complex shapes using different weights on the various eigenfunctions. This is an improvement over the previous method, which would force a Weibull shape even where not appropriate.

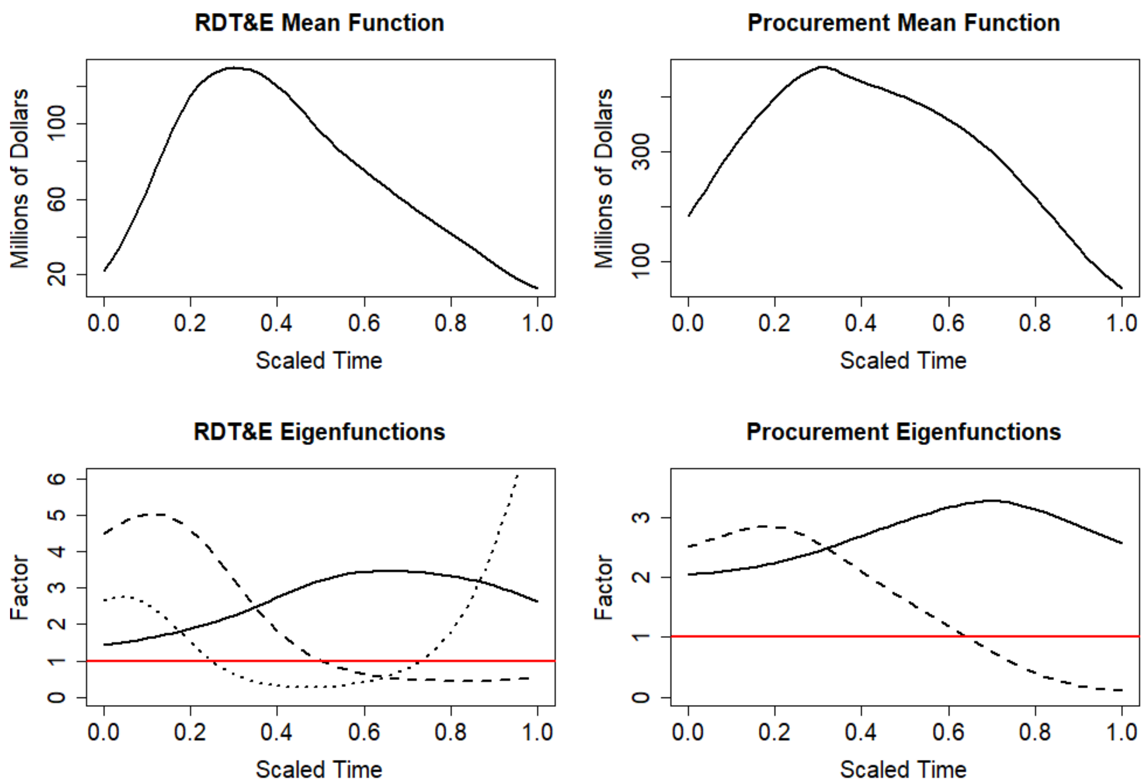


Figure 2. **Mean Shape and Eigenfunctions for RDT&E and Procurement**

The FPCA process was fit to logged spending profiles; these curves have been transformed back into dollar units. The first principal eigenfunction is represented by the solid line, the second principal eigenfunction is represented by the dashed line, and the third principal eigenfunction is represented by the dotted line. Since eigenfunctions have been exponentiated, these represent multiplicative deviations from the mean. If the eigenfunction is greater than 1, it induces a positive deviation from the mean; if it is less than 1, it is a negative deviation from the mean. The solid red line at 1 represents no deviation from the mean response. The scale of the deviation is determined by the FPCA scores. The FPCA scores are real-valued; thus, if a score is below zero, the eigenfunction flips over the red line (but not symmetrically due to the non-linear transformation).

Identifying Potential Predictor Variables

Given choices for functional forms, the next challenge is to somehow characterize how the distribution of possible actual outcome profiles could be derived for a given initial plan. It seems obvious that different kinds of programs involve different levels of cost risk. There is a substantial literature attempting to identify specific factors that are correlated with program cost and schedule growth. Some factors that have been found by past researchers to be correlated with (unit) cost growth and/or total program cost growth risk include:

- Commodity type (e.g., helicopter, satellite, MAIS, missile, or submarine; Arena et al., 2006; Drezner et al., 1993; Tyson, Harmon, & Utech, 1994)
- Acquiring Service (Army, Navy, Air Force, Joint, Department of Energy) (Drezner & Smith, 1990; Jessup & Williams, 2015; Light et al., 2017; McNicol, 2004)
- New design vs. modification of existing design (Arena et al., 2006; Coonce et al., 2010; Drezner et al., 1993; Jimenez et al., 2016; Marshall & Meckling, 1959)
- New build vs. remanufacture of existing units (Tyson et al., 1989)
- Budget climate at Milestone B (Asher & Maggelet, 1984; McNicol, 2017)
- Number of years of spending prior to Milestone B (Jimenez et al., 2016; Light et al., 2017)
- Schedule optimism (Arena et al., 2006; Asher & Maggelet, 1984; Glennan et al., 1993; Tate, 2016)
- Technology maturity of the program (Adoko, Mazzuchi, & Sarkani, 2015; GAO, 2006)
- Investment size (Bliss, 1991; Creedy, Skitmore, & Wong, 2010)

Because we are not attempting to diagnose causes of cost growth, but are instead only trying to understand and characterize risk (on the assumption that the past is a reasonable guide to the future), we do not distinguish here among risks arising from discretionary choices, environmental factors, or intrinsic program features.

Describing Changes in Cost and Schedule as Changes in Profile Functions

We can model the change in an initial or midlife profile to a final profile by modeling the change in FPCA scores, total spending, and total duration. We saw above that we can model development costs or production costs as being generated from a basis of eigenfunctions. Define ω_{ijk}^{0} to be the FPCA scores of the planned or midlife programs and ω_{ik}^1 to be the FPCA scores of the completed program. Define T_{ij}^0 to be the planned total duration until program completion and T_i^1 to be the total duration when program i is actually completed. Define C_{ij}^0 to be the total cost of an initial or midlife program and C_i^1 to be the total cost of a completed program. Then $\theta_{ij}^0 = (\omega_{ij1}^0, \dots, \omega_{ijK}^0, T_{ij}^0, C_{ij}^0)$ fully characterizes the cost profile of an initial or midlife program and $\theta_{ij}^1 = (\omega_{i1}^1, \dots, \omega_{iK}^1, T_i^1, C_i^1)$ fully characterizes a final cost profile. We estimate the conditional (joint) distribution of θ_{ij}^1 given the appropriate program and environmental attributes and the fact that the program's previous estimate was best fit by the FPCA expansion.

There are several possible approaches to this and many choices of how to parametrize the family of curves being fit, but the general method will be the same in all cases. We estimate the distribution of θ_i^1 as a function of parameters θ_{ij}^0 and the historical program characteristics X_{ij}:

$$\theta_i^1 = \left(X_{ij}, \theta_{ij}^0\right)\beta + \eta_{ij},$$

where X_{ij} includes factors such as current estimated cost, Service, budget climate, and so forth, and η_{ij} are independent and identically distributed draws from a multivariate normal distribution centered at the vector 0 with covariance Σ. The vector X_{ij} gives the values of the predictors for historical program i in year j. The vector θ_{ij}^0 contains the elements that describe cost profile for program i in year j. The vector (X_{ij}, θ_{ij}^0) denotes the component-wise concatenation of θ_{ij}^0 onto X_{ij}.

This linear regression model implies a functional fit and distribution over the annual cost profile function $C_{ij}(t)$. Rather than attempting to predict eventual actual cost as a function of initial estimated cost and other predictors, we instead attempt to predict the distribution of the parameters of a function that *generates* eventual cost, given program-specific attributes and the parameters that generate the initial estimate. Note that this is a multiple output regression—we are simultaneously estimating all of the best-fit parameters θ_i^1 and the covariance matrix that describes how those parameters are correlated.

We use a Bayesian estimation framework, starting with a weakly informative prior distribution $F_{prior}\left(\theta_i^1\right)$ and using Markov Chain Monte Carlo (MCMC) estimation to derive a posterior distribution $F_{posterior}\left(\theta_i^1\right)$, including the covariance matrix (Chib & Greenberg, 1995). We do this separately for RDT&E costs and procurement costs, using different families of profile-generating functions, treating their changes in shape and size as independent. Treating development and procurement jointly is a potential area for future research.

Regression Methodology

We compiled original and midlife estimates and actual outcomes for $I = 1{,}278$ historical profiles for RDT&E and $I = 828$ for procurement. For each historical program i at year t, we fit scaled, FPCA scores to the full current profiles (including completed programs):

$$C_{ij}(t) = \mu(t) + \sum_{k=1}^{K} \omega_{ijk}\xi_k(t) + \epsilon_{ij}(t).$$

Let $\theta_{ij} = (\omega_{ij1}, \ldots, \omega_{ijK}, , C_{ij}, T_{ij}^{p})$ be the parameters of those best-fit curves. Then θ_{ij}^{0} are the best fit parameters to the initial and midlife profiles and θ_i^1 are the best fit parameters to the actual completed profiles. We further decompose C_{ij}^{0} into $(C_{ij}^{0a}, C_{ij}^{0p})$ to be the actual spending that has already occurred (if any) and the planned spending yet to occur. We model the distribution of θ_i^1 as a function of θ_{ij}^{0} and a set of predictor variables X_{ij} simultaneously over all programs, where X_{ij} includes the program-specific and environmental factors previously listed. Parametric linear models are simultaneously fit to obtain a predictive model for the final profile parameters θ_i^1. The models are as follows:

$$\omega_{\mathrm{i1}}^{1} = (X_{ij}; \theta_{ij}^{0})\beta_{\omega_1} + \eta_{\omega_1},$$

$$\vdots$$

$$\omega_{\mathrm{iK}}^{1} = (X_{ij}; \theta_{ij}^{0})\beta_{\omega_{\mathrm{K}}} + \eta_{\omega_{\mathrm{K}}},$$

$$\sqrt{\mathrm{C}_{\mathrm{i}}^{1}} = \left(X_{ij}; \theta_{ij}^{0}\right)\beta_C + \eta_C ,$$

$$\log(\mathrm{T}_{\mathrm{i}}^{1}) = (X_{ij}; \theta_{ij}^{0})\beta_T + \eta_T,$$

where the error terms $(\eta_{\omega_1}, \ldots, \eta_{\omega_K}, \eta_C, \eta_T)$ are assumed to be jointly normally distributed. The covariates X_{ij} include information about previously finished programs that had initial planned spending profiles and actual final profiles. Using these historical data, the model is fit to predict final actual profiles using only information available from a program's Milestone B date. The parameters $\beta = (\beta_{\omega_1}, , \beta_{\omega_K}, \beta_C, \beta_T)$ are jointly estimated using a Bayesian Seemingly Unrelated Regressions model with prior distributions on the parameters $E\left[\theta_i^1 | X_{ij}\right] \equiv \beta$ and $Var\left[\theta_i^1 | X_{ij}\right] \equiv \Sigma$.

The prior for β has a multivariate normal distribution, calibrated such that prior belief is that there is no change in the profile from the current estimate to final actual profile and no other traits of the initial profile are predictive of the final actual profile. This prior belief is fairly strong in order to induce regularization. This prior choice balances the bias-vs.-variance tradeoff to produce better out-of-sample predictions.

The prior for Σ has an inverse Wishart distribution, chosen such that the equations are uncorrelated and the prior variance is 1.

The joint posterior distribution of β and Σ incorporates the prior beliefs and the historical data to arrive at an updated posterior belief. The Bayesian machinery is especially useful for our purposes because it allows us to obtain random draws from the posterior distribution of β and Σ, which in turn allows us to generate random draws of a final profile distribution $\hat{\theta}_i^1$ for any program with known initial profile characterized by covariates X_{ij} and θ_{ij}^{0}. This lets us estimate the complete (posterior) distribution of final profiles, rather than just a point estimate and variance measure. We sample from the posterior with an MCMC Gibbs algorithm from Rossi, Allenby, and McCulloch (2005). We draw 400,000 MCMC samples, keep every fourth draw, and discard the first 1,000, leaving us with a Monte Carlo sample of 99,000 draws. Keeping every fourth draw is called "thinning" and reduces the MCMC autocorrelation; discarding the first 1,000 draws is called "burn-in" and ensures we utilize draws after the MCMC algorithm has converged to the posterior distribution.

Regression Data

The data for the regression are the initial estimate and final actual cost profiles for completed historical MDAPs. The earliest program in the data set passed Milestone B in 1982. The data are taken from SARs, together with compiled attributes and environmental factors (as enumerated above) for each program. We apply this method to both development (RDT&E) cost risk and procurement cost risk models, which differ only in which predictor values are used, the number of eigenvectors fitted, and the eigenvector shapes resulting from the estimation.

The following are the specific predictor variables used in this paper:

- ω_{ij1}—the first FPCA score
- ω_{ij2}—the second FPCA score
- ω_{ij3}—the third FPCA score
- $\sqrt{C_{ij}^{a}} = \sqrt{\sum_{t=0}^{j} C_{ij}(t)}$—square root of actual spending
- $\sqrt{C_{ij}^{p}} = \sqrt{\sum_{t=j+1}^{\infty} C_{ij}(t)}$—square root of planned spending
- $\log(T_{ij}^{0})$—natural log of the planned number of future spending years
- The Service overseeing the program (Navy, DoD, Air Force, Army, Department of Energy)
- A commodity type (Air; Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance [C4ISR]; Ground; Ordnance; Sea; Space; other)[1]
- A measure of relative Service budget tightness compared to two years ago
- A measure of relative Service budget tightness over the last 10 years
- A measure of budget optimism—planned spending divided by the mean historical actual spending for this commodity type
- A measure of schedule optimism—planned duration divided by the mean historical actual duration for this commodity type
- Whether the program is based on a modification of a preexisting design (binary)

The measures of relative budget tightness were based on the year the program passed Milestone II/B. The measures of budget and schedule optimism reflect the

[1] More precise commodity categories—e.g., distinguishing helicopters from fixed-wing aircraft—might be useful, given enough data. We found that increasing the sample size in each category led to better results than increasing the precision of the categories.

empirical observation that the average behavior of programs in a given commodity class is a better predictor of cost and schedule than early cost and schedule estimates of individual programs in that class.

Monte Carlo Risk Analysis

General Approach

Suppose that we have budgeted a program at some level, possibly different from its predicted cost profile. Let $B(t)$ be the budgeted funds in year t, and let $C_0(t)$ be the predicted cost that will be incurred in year t. There are many questions we might wish to ask about the program's affordability risk:

- In how many years will the program exceed the planned budget?
- How many total dollars over budget will the program spend?
- What is the probability of exceeding the budget at least once over the FYDP?
- How much contingency funding would be needed to achieve 90% confidence of staying within budget, depending on whether unspent contingency carries over to the next year?

These are all questions of potential interest to both program managers and resource managers. Using the posterior final profile distribution derived from the original profile C_0, we can perform many counterfactual Monte Carlo analyses to answer these kinds of questions. The general pattern for these analyses is as follows:

1. Given the initial development estimate for a program …
2. Define a yearly budget level $B(t)$, and a contingency fund size (if any).
3. Use the regression described above to determine the posterior distribution on the parameters of the best fit to the final actual development profile for the program.
4. Define outcomes or events of interest—e.g., exceeding the budget in some year, or staying within the budget through the entire FYDP, or having planned funds at least as large as spent funds in year 7.
5. For s = 1,…,S (indexing over iterations of the Monte Carlo algorithm):
   a. "Draw" random parameter vector $\theta^{1(s)}$from the posterior distribution.
   b. Compute the corresponding yearly values by evaluating the best fit curve at t = 1, …, $T^{1(s)}$ and computing $\exp(C^{1(s)})\frac{\exp\left(\mu(t)+\sum_{k=1}^{3}\omega_{ik}^{1(s)}\xi(t)\right)}{\sum_{t=1}^{T^{1(s)}}\exp\left(\mu(t)+\sum_{k=1}^{3}\omega_{ik}^{1(s)}\xi(t)\right)}$.
   c. Evaluate and store any events or outcomes of interest.

Note that the value of $T^{1(s)}$ used in step 5b is determined as part of $\theta^{1(s)}$ in step 5a.

After S iterations, calculate the statistics of interest over the stored events or outcomes. For example, count the number of times N that $C^{1(s)}(t) < B(t)$ for t = 1…5, and compute $\frac{N}{S}$. This is the estimated probability of staying within budget for the first five years. The Monte Carlo framework can also allow comparison of different management policies. For example, one could compare the effect of pre-allocating contingency to specific program years, versus maintaining a contingency fund to be spent down over time as needed.

In general, we would do this not only for development profiles, but also for procurement spending. In that case, policy makers might be interested in how much difference it would make to be able to manage both RDT&E and procurement using a single combined budget and/or a single program contingency fund, rather than having to manage separate budgets and contingency amounts due to “color of money” prescriptions.

Figure 3 shows an example of applying this method to a new program, using actual RDT&E cost estimates for a current MDAP. The vertical red line marks the date of the analysis, which is the boundary between past actual funding and projected future funding. The green line shows actual past annual funding; the dashed blue line shows the program estimate of annual future funding. The solid black line is the mean projected funding derived from the FPCA Monte Carlo methodology; the dashed lines mark upper and lower 10% prediction interval bounds[2] year by year as determined using the weighted Monte Carlo.

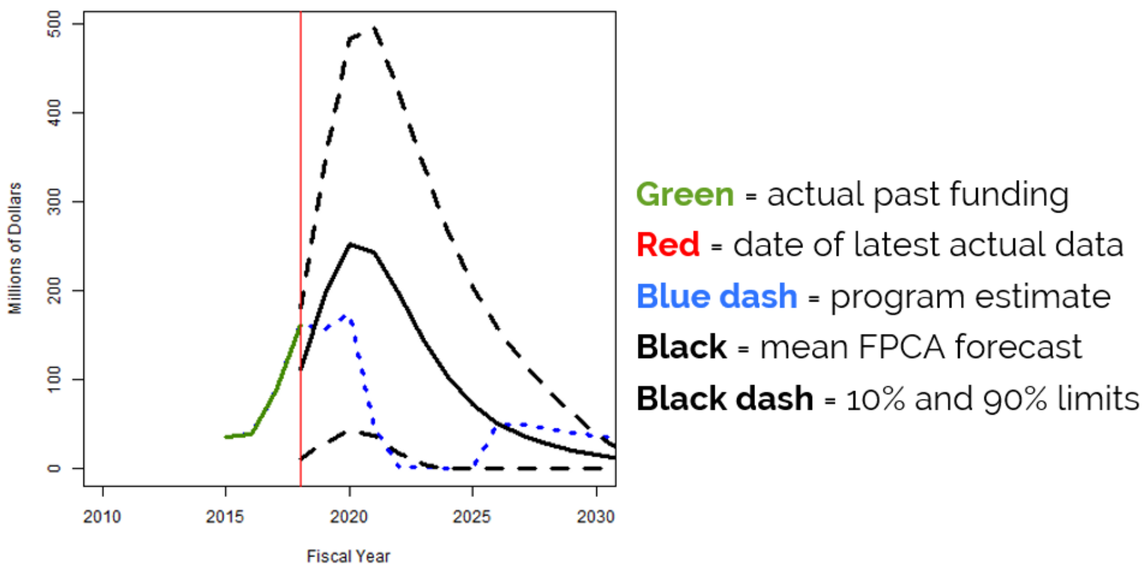


Figure 3. **Predicted Future RDT&E Spending for a New Program**

Midlife Programs

The method we described in Tate et al. (2018) applies specifically to programs that are just beginning, using an estimate of their future spending profiles. In practice, however, most acquisition programs in any given year are already partially complete, and part of their realized spending profiles is known. We need a method that accounts

[2] That is, the 10% and 90% quantiles of the estimated distribution of possible funding in each year.

for those actual costs to date, and generates profile distributions for the future that are conditioned on that history.

Statistically, it would be difficult to perform conditional FPCA regressions to predict the remaining profiles for RDT&E and procurement, taking the actual costs to date as input factors. Not only would the power of the regressions be greatly reduced (due to paucity of historical programs with a specific cost history), but also the characterization of shapes of actual spending would be at least as complicated as for overall profiles. In the worst case, we would need a separate set of FPCA eigenfunctions for programs with one year of actuals, programs with two years of actuals, and so forth. Fortunately, the Monte Carlo framework for generating posterior empirical distributions provides an alternative that is both computationally efficient and effective.

We implement this method as follows. The original Monte Carlo method as described in Tate et al. (2018) weighs all random draws from the posterior distribution equally, in order to produce year-by-year empirical spending distributions for the program. We modify the method for midlife programs using unequal weighting of these random draws. Instead of weighing all draws equally when estimating the distribution of future profiles, we instead give higher weight to those draws that more closely match the observed history of the program to date. This is comparable to a Nadaraya-Watson estimator (Nadaraya, 1964), and has the effect of conditioning the Monte Carlo–based future estimates on the observed history. The exact weighting scheme can be adjusted to balance between computational efficiency and strict enforcement of the conditioning on the past.

Figure 4 shows an example of this method, applied to a different actual MDAP. Note how the model predicts mean funding levels below the planned level, but with significant uncertainty (including some chance of program cancellation).

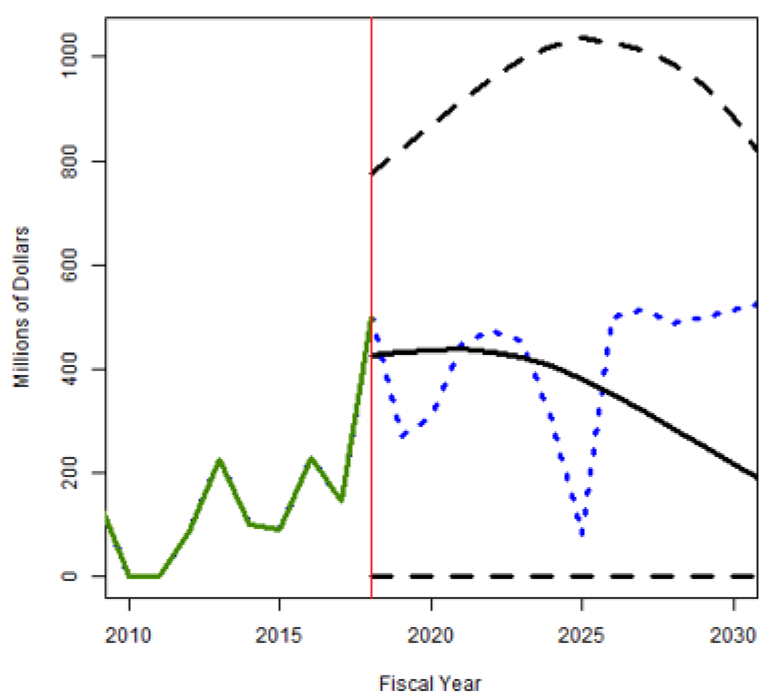


Figure 4. **Prediction Intervals for Future Funding of an Ongoing Program**

Portfolios: More Than One Program at a Time

We have shown how our model can characterize the affordability risk of a single program's development budget. We noted in Tate et al. (2018) that it would be even more useful to be able to characterize the affordability risk of a group of projects or programs being managed with a common contingency pool. If the conditional outcomes of these programs were approximately independent, this would not be much more complicated than the single-program case. In practice, we know that funding levels among programs within a portfolio are negatively correlated; this is a potential area for future research. For the moment, we treat programs as if they were independent, and incorporate current funding tightness as a predictor of outcomes. This does not affect the mean outcome for each program but does increase the variance.

If we have estimated the $F_{posterior}(\theta^1)$ distributions for each of a set of programs, we can apply the same kind of Monte Carlo analysis to the sum of their annual costs, compared against a collective portfolio budget and contingency fund. This could be done separately for RDT&E and procurement, each with its own budget, or it could be done using a combined investment budget. This would enable true affordability analysis of portfolios as envisioned by the Better Buying Power initiatives,[3] but with considerably more realism than current affordability analyses that are based on point-estimate cost profiles assuming fixed program content and quantities.

One potential use of such a model would be to quantify the benefits of portfolio-level contingency funding versus program-level contingency funding. It is well known in the project management world that allocating reserve funds to specific cost areas before you actually know where the cost growth is going to occur leads to less efficient use of those reserve funds. However, it has historically been difficult to protect funds that are not part of the base budget for some cost element. In the DoD, apart from a highly limited ability to reprogram funds from one program element or line item to another, there is currently no ability to reserve funds for contingency use outside of a specific program's budget. The recent report of the Section 809 Panel specifically recommended expanding the ability of the DoD to reprogram funds across programs and manage contingency at the portfolio level. This research provides some analytical support for those recommendations.

Potential Criticisms of the Method

We noted in Tate et al. (2018) that the utility of these methods assumes, among other things, that historical patterns of cost growth and schedule stretch will persist into the future. This is a conservative assumption, given that observed patterns of cost and schedule growth in major programs have persisted across multiple acquisition systems and regulatory regimes over the past decades. A more nuanced concern is that if resource managers were to actually use these methods to manage portfolios of programs more efficiently, the resulting changes in program outcomes ought to invalidate the models, at least until a new collection of historical outcomes under the new regime could be assembled.

---

[3] Department of Defense, *Better Buying Power*, http://bbp.dau.mil/index.html.

We also noted that these methods offer no insights into *why* costs and schedules deviate from their original estimates (or how this could be "fixed"), and that these methods explicitly model how much funding a program *will receive* in a given year—not how much it needs, or ought to receive, or would receive if there were more money to go around. As such, the model data incorporate the history of negotiations between the Services, the Office of the Secretary of Defense, and the Congress regarding how much to fund programs year by year, and when to cancel them. If there were to be a fundamental change in the dynamic of how those decisions are made, then that, too, might invalidate the link between historical outcomes and future program outcomes, at least until enough new data could be collected.

Finally, we note that the current portfolio modeling approach treats individual program funding levels as independent draws from their respective posterior distributions. This is known to be a weak assumption and is a potential area for future research.

Conclusions

Quantifying Annual Resource Risks for a Program or Portfolio

We have developed a methodology to characterize the year-by-year budget risk associated with a major acquisition program. This methodology can be applied to both development costs and procurement costs, and can be extended to understand the aggregate affordability risk of portfolios of programs. The method allows resource managers to estimate annual budget risk levels, required contingency amounts to achieve a specified probability of staying within a given budget, and a host of other relevant risk metrics for programs. It also allows policy makers to predict the impact on program affordability of proposed changes in how contingency funds are managed.

Research Program Status

The switch from functional regression using parametric curve families to FPCA using nonparametric eigenfunction kernels has significantly improved both the fidelity of the curve fits and the flexibility of the predictive aspects of our approach. We have established that FPCA methods can accurately reproduce historical funding profile shapes for both RDT&E and procurement profiles, and that it is possible to characterize the uncertainty in future spending profiles using the outputs of FPCA and weighted Monte Carlo techniques to sample from the distribution of overall funding profiles while accounting for actual program history to date. This represents a significant improvement in the state of the art; we are not aware of any other technique that has been proposed that can predict time-phased cost and schedule growth distributions for any kind of defense acquisition program, much less a general approach that potentially can be applied to all programs.

Future Research

This technique is currently in the prototype stage and is based on a relatively sparse set of historical program outcome data. There is still much work to be done on establishing the ideal number of eigenfunctions to use in fitting initial and final RDT&E and procurement profiles (respectively), characterizing the distribution of residuals around the best-fit functional curve and utilizing a more flexible mean function in the SUR regression. Additionally, we would like to assess the model's predictive power with an out-of-sample prediction exercise. However, measures for out-of-sample predictive accuracy of functional distributions is an unexplored topic in statistical methodology. There is also a great deal to be learned about how managers could best use the information provided by this method to manage actual programs and portfolios, and what the implications might be for recommending policy changes to acquisition law and regulations. As noted in the Portfolios: More Than One Program at a Time and the Potential Criticisms of the Method sections, the current portfolio modeling approach treats program outcomes as independent. It would be useful to extend this approach to account for correlations among funding levels within a portfolio, or to explicitly model priorities among programs.

References

Adoko, M. T., Mazzuchi, T. A., & Sarkani, S. (2015). Developing a cost overrun predictive model for complex systems development projects. *Project Management Journal*, *46*(6), 111–125. doi:10.1002/pmj21545

Anderson, T. P., & Cherwonik, J. (1997, Summer). Cost estimating risk and cost estimating uncertainty guidelines. *Acquisition Review Quarterly*, 339–348. Retrieved from www.dtic.mil/get-tr-doc/pdf?AD=ADA487532

Arena, M. V., Leonard, R. S., Murray, S. E., & Younossi, O. (2006). *Historical cost growth of completed weapon system programs* (TR-343-AF). Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/content/dam/rand/pubs/technical_reports/2006/RAND_TR343.pdf

Asher, N. J., & Maggelet, T. F. (1984). *On estimating the cost growth of weapon systems* (IDA Paper P-1494). Alexandria, VA: Institute for Defense Analyses.

Bitten, R. E., & Hunt, C. (2017). *Assessing the impact of confidence levels in funding and budgeting NASA science missions*. Paper presented at the ICEAA Professional Development & Training Workshop, Portland, OR.

Bliss, G. (1991, June). *The accuracy of weapon systems cost estimates*. Presented at the 59th Military Operations Research Symposium, U.S. Military Academy, West Point, NY.

Brown, G. E., White, E. D., Ritschel, J. D., & Seibel, M. J. (2015). Time phasing aircraft R&D using the Weibull and Beta distributions. *Journal of Cost Analysis and Parametrics*, *8*(3), 150–164. doi: 10.1080/1941658X.2015.1096219

Chib, S., & Greenberg, E. (1995). Hierarchical analysis of SUR models extensions to correlated serial errors time-varying parameter models. *Journal of Econometrics*, *68*, 339–360. doi:10.1016/0304-4076(94)01653-H

Congressional Budget Office. (2017). *An analysis of the Obama administration's final future defense program* (CBO Report No. 52450). Retrieved from https://www.cbo.gov/system/files/115th-congress-2017-2018/reports/52450-fydp.pdf

Coonce, T., Bitten, R., Hamaker, J., & Hertzfeld, H. (2010). NASA productivity. *Journal of Cost Analysis and Parametrics*, *3*(1), 59–78. doi:10.1080/1941658X.2010.10462228

Creedy, G. D., Skitmore, M., & Wong, J. K. W. (2010). Evaluation of risk factors leading to cost overrun in delivery of highway construction projects. *Journal of Construction Engineering and Management*, *136*(5), 528–537. doi:10.1061/(ASCE)CO.1943-7862.0000160

Drezner, J. A., Jarvaise, J. M., Hess, R. W., Hough, P. G., & Norton, D. (1993). *An analysis of weapon system cost growth*. Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/content/dam/rand/pubs/monograph_reports/2006/MR291.pdf

Drezner, J. A., & Smith, G. K. (1990). *An analysis of weapon system acquisition schedules* (R-3937-ACQ). Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/pubs/reports/R3937.html

Glennan, T. K., Bodilly, S. J., Camm, F., Mayer, K. R., & Webb, T. (1993). *Barriers to managing risk in large scale weapons system development programs* (MR-248-AF). Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/pubs/monograph_reports/MR248.html

GAO. (2006). *Space acquisitions: DOD needs to take action to address unrealistic initial cost estimates of space systems* (GAO-07-96). Washington, DC: Author. Retrieved from https://www.gao.gov/products/GAO-07-96

Jessup, R., & Williams, J. (2015). *The cost of commonality: Assessing value in joint programs* (MBA professional report). Monterey, CA: Naval Postgraduate School.

Jimenez, C., White, E. D., Brown, G. E., Ritschel, J. D., Lucas, B. M., & Seibel, M. J. (2016). Using pre-Milestone B data to predict schedule duration for defense acquisition programs. *Journal of Cost Analysis and Parametrics*, *9*(2), 112–126. doi:10.1080/1941658X.2016.1201024

Light, T., Leonard, R. S., Pollak, J., Smith, M., & Wallace, A. (2017). *Quantifying cost and schedule uncertainty for major defense acquisition programs (MDAPs)*. Santa Monica, CA: RAND Corporation. doi:10.7249/RR1723

Marshall, A. W., & Meckling, W. H. (1959). *Predictability of the costs, time, and success of development* (P-1821). Santa Monica, CA: RAND Corporation. Retrieved from https://www.rand.org/pubs/papers/P1821.html

McNicol, D. L. (2004). *Cost growth in major weapon procurement programs*. Alexandria, VA: Institute for Defense Analyses.

McNicol, D. L. (2017). *Influences on the timing and frequency of cancellations and truncations of major defense acquisition programs* (IDA Paper P-8280). Alexandria, VA: Institute for Defense Analyses.

McNicol, D. L., Gould, B. R., Wu, L., Fuchs, E., & Bronson, P. (2013). *Evidence on the effectiveness of the Defense Acquisition Board process* (IDA Paper P-5011). Alexandria, VA: Institute for Defense Analyses.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, *9*(1), 141–142. doi:10.1137/1109020

Rossi, P. E., Allenby, G. M, & McCulloch, R. E. (2005). *Bayesian statistics and marketing*. Hoboken, NJ: Wiley.

Tate, D. M. (2016). *Acquisition cycle time: Defining the problem* (IDA Document NS D-5762). Alexandria, VA: Institute for Defense Analyses.

Tate, D. M., Coonce, T. J., & Guggisberg, M. R. (2018) Quantifying annual affordability risk of major defense programs (IDA Document NS D-9019). In *Proceedings of* *the 15th Annual Acquisition Research Symposium*. Monterey, CA: Naval Postgraduate School.

Tyson, K. W., Harmon, B. R., & Utech, D. M. (1994). *Understanding cost and schedule growth in acquisition programs* (IDA Paper P-2967). Alexandria, VA: Institute for Defense Analyses.

Tyson, K. W., Nelson, J. R., Om, N. I., & Palmer, P. R. (1989). *Acquiring major systems: Cost and schedule trends and acquisition initiative effectiveness* (IDA Paper P-2201). Alexandria, VA: Institute for Defense Analyses.

Tyson, K. W., Nelson, J. R., & Utech, D. M. (1992). *A perspective on acquisition of NASA space systems* (IDA Document D-1224). Alexandria, VA: Institute for Defense Analyses.

Yao, F., Müller, H., & Wang, J. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, *100*(470), 577–590. doi:10.1198/016214504000001745

Acknowledgments

This work was originally conceived and funded by the congressionally established Section 809 Panel (https://section809panel.org/), and in particular by Commissioners Dr. Allan Burman and Mr. David Ahern. Additional funding and data were provided by the Office of the Secretary of Defense. Dan Cuda and James Bishop of the Institute for Defense Analyses were technical reviewers, and their observations and suggestions improved the paper significantly.

Panel 19. Improving the Effectiveness of Ship Acquisition

Thursday, May 9, 2019	
12:45 p.m. – 2:00 p.m.	**Chair: Rear Admiral Michael J. Johnston, USCG,** Assistant Commandant for Acquisition & Chief Acquisition Officer ***Flexible Ships*** Johnathan Mun, Naval Postgraduate School ***Naval Combat System Product Line Architecture Economics*** Ray Madachy and John Green, Naval Postgraduate School ***Defining a Model-Based Systems Engineering Approach for Systems Engineering Technical Reviews*** Warren Vaneman and Ronald Carlson, Naval Postgraduate School

Rear Admiral Michael J. Johnston, USCG— Rear Admiral Johnston currently serves as the U.S. Coast Guard's Director of Acquisition Programs and Program Executive Officer (PEO). His duties include management oversight of all Coast Guard acquisition programs and projects for the modernization and recapitalization of surface, air, command and control, and logistics assets in support of the Coast Guard's multiple maritime missions.

Rear Admiral Johnston graduated from the United States Coast Guard Academy in 1990 with a Bachelors of Science in Electrical Engineering. After serving two consecutive tours afloat, he attended the Georgia Institute of Technology to complete a Master of Science in Electrical Engineering. He then served as a project manager and senior communications network engineer for the Vessel Traffic Systems and the Differential Global Positioning System.

He was selected to serve as Aide to the Commandant of the Coast Guard from 2000-2002. During this tumultuous timeframe, his work helped restore the Nation to normalcy following the 9/11 terrorists attacks. He also helped to transition the Coast Guard into the newly-formed Department of Homeland Security.

Rear Admiral Johnston went back to operations as the Deputy Commander for Group St. Petersburg, FL coordinating operations with partner agencies and restructuring the unit into a Sector Command. Following this assignment, he reported as Commanding Officer of the Electronics Systems Support Unit in New Orleans, LA; just two months before Hurricane Katrina. He coordinated response, recovery, and repair efforts for all command and control systems across the Gulf Coast from Mexico to Florida.

In 2008, Rear Admiral Johnston graduated from the Massachusetts Institute of Technology with a Master of Science in Management and Systems Engineering. A DHS Level III Acquisitions Program Manager, he went on to serve as the Deputy Project Manager within the Coast Guard Acquisitions program, delivering significant capabilities to the newest cutters and aircraft for the Service.

From 2011 to 2014, Rear Admiral Johnston served as the Commanding Officer of the Coast Guard's Command, Control, and Communications Engineering Center. He was responsible for the development, deployment, sustainment, logistics support, training, and configuration management for all C3 systems in the Coast Guard.

He served in Officer Personnel Management; then, in June 2015, he was selected to serve as Executive Director for the Deputy Commandant for Mission Support (DCMS), responsible for staffing, training and equipping the Coast Guard for its missions.

Flexible Ships

Johnathan Mun—PhD, is a Research Professor at the Naval Postgraduate School, Monterey, CA. [jcmun@nps.edu]

Abstract

The current global security environment is changing at a faster pace than ever before with higher levels of complexity and competitiveness, with a complex dynamic of possibilities. The U.S. Navy not only needs more platforms or ships, but it needs them with the ability to adapt to changes with new technologies and operational concepts. One such concept is that of flexibility in our fleet of ships. To successfully implement the Surface Navy's Flexible Ships concept, PEO-SHIPS requires a new methodology that assesses the total future value of various combinations of Flexible Ships' design features and how they will enable affordable warfighting relevance over the ship's full-service life. Examples of Flexible Ships design features include decoupling payloads from platforms, standardizing platform-to-payload interfaces, implementing allowance for rapid reconfiguration of onboard electronics and weapons systems, preplanning access routes for mission bays and mission decks, and allowing for sufficient growth margins for various distributed systems. This research analyzes the application of strategic Real Options Valuation methodology within the Integrated Risk Management process to assess the total future value of Flexible Ships design features and for use in the Future Surface Combatant Analysis of Alternatives. The current research has the explicit goal of proposing a reusable, extensible, adaptable, and comprehensive advanced analytical modeling process to help the U.S. Navy in quantifying, modeling, valuing, and optimizing a set of ship design options to create and value a business case for making strategic decisions under uncertainty.

Introduction

The current global security environment is changing at a faster pace than ever before with higher levels of complexity and competitiveness, with a complex dynamic of possibilities. Not only does the U.S. Navy need more platforms or ships, but it needs them with the ability to adapt to changes with new technologies and operational concepts. One such concept is that of flexibility in our fleet of ships.

In the Flexible Ships IPT Charter signed out by the nine Flag Officers/SESs last year, the IPT is tasked to "make recommendations to leadership for policy/process changes that will foster incorporation of flexible ships features into current and future ships."

The U.S. Navy is tasked with fulfilling its missions globally in environments with rapidly changing threats using an equally rapidly evolving technological base of platform, mission, electronic, and weapon systems. The challenge the U.S. Navy faces is to retain and maintain sufficient military relevance during wartime as well as peacetime, with the added goal of minimizing highly intrusive and costly modernization throughout a ship's service life by incorporating Modular Adaptable Ships (MAS) and Flexible and Adaptable Ship Options (FASO) in the ship design. Accomplishing this goal has the added benefit of allowing the Navy to affordably and quickly transform a ship's mission systems over its service life to maintain its required military capabilities (Doerry, 2012).

The operative term in FASO is *flexibility*. To have flexibility means to have the options to make midcourse corrections and changes as required, when uncertainties become resolved over the passage of time, actions, and events. In other words, the uncertainties of what future technologies may look like and the ever-changing complexity of global threats balanced with the need for the U.S. Navy to be responsive and persistent, and to maintain a fleet that provides leaders and decision-makers with credible and exercisable options, requires flexibility. When uncertainties become known, the correct course of action can be implemented, but only if the flexibilities exist. Therefore, by definition, having flexibility or options in place and ready to execute provides value as well as a lower overall cost of making major ship changes and alterations when the occasion calls for them.

Having a 355-ship navy is by itself insufficient regarding effectiveness. It is becoming more critical to consider what these platforms are capable of in terms of creating an effect and affecting desired outcomes, having flexibility to adapt to ever-changing threats, and possessing the ability to incorporate evolutionary technology upgrades as they become available. Current levels of technology are insufficient for maintaining maritime superiority. New operational concepts and technologies need to be consistently updated in our fleet. In most cases, timing is critical. We cannot wait until our adversaries have a new technology before testing and implementation begins on our end.

Flexibility and modularity provide the U.S. Navy with the options to execute various capabilities quickly. Future flexible ships can be designed such that they are modular with rapidly swappable components. New technologies (sensors, weapon systems, system upgrades) can be implemented rapidly at lower costs if the ship was designed with future flexibility and upgrades in mind. Flexibility is synonymous with *options*. That is, with flexible ships, leadership has the option, but not the obligation, to execute any of the available trigger points when conditions make it optimal to do so.

To understand the Flexible Ships concept, one has to first look at its basic principles and tenets. Sturtevant (2015) lists the five main tenets of flexible ships:

1. Decoupled payloads (capabilities) from platforms (ships)
2. Standard platform-to-payload interfaces—well-defined, common interfaces for distributed ship services that are prescribed and managed by the U.S. Navy
3. Rapid reconfiguration—specific C5I compartments that can be easily reconfigured with upgraded equipment or new systems
4. Preplanned access routes—used for the easy removal and replacement of interior equipment or systems
5. Sufficient service life allowance growth margins (space and weight for future capabilities, and provision for projected demand for distributed systems such as electric power, cooling, and network bandwidth)

This current research focuses on applying a series of analytical methodologies, such as Real Options Valuation (ROV), to support development of a business model or business case analysis that supports strategic decision-making in the context of uncertainty. This analysis identifies, models, values, and optimizes the various strategic real options identified for flexible ship designs. Currently, there is only a limited set of real-life applications of FASO/MAS in ship design, and they are classified; therefore, actual empirical data is not used in this research. In addition, because the objective of this research is to illustrate in detail the business case modeling process and analytical methodologies such that the method and process can be replicated and used in all future FASO/MAS design decisions, subject matter expert (SME) inputs, publicly available information, and a set of basic assumptions or rough order magnitude (ROM) estimates are used. The use of the ROM or SME inputs, while subjective in nature, in no way detracts from the analytical power, efficacy, or applicability of these methods, because the values they supply to the model parameters can be replaced with more objective values as they become available.

Literature Review

The concept of FASO is not new to the Navy. In fact, benefits of FASO/MAS concepts were detailed in the mid-1970s by Jolliff (1974), Simmons (1975), Drewry and Jons (1975), and others. Even as recently as 2015, the Naval Sea Systems Command's (NAVSEA's) Program Executive Office, Ships (PEO-SHIPS) put out a presentation on Flexible Ships, detailing its "Affordable Relevance over the Ship's Life Cycle" (Sturtevant, 2015). In it, the director of science and technology, Glen Sturtevant, noted that the main current and future challenges confronting the surface Navy include facing unknown but evolving global threats while managing an accelerated pace of technological changes, coupled with handling rising costs and declining budgets. The analysis found that ships currently cost too much to build and sustain, the ships (platforms) are too tightly coupled with their capabilities (payloads), and inflexible and fixed architectures of legacy ships limit growth and capability upgrades or result in lengthy and costly upgrades. The effects of these issues, of course, are compounded by ever-evolving, unknown global threats.

When the Freedom and Independence classes of American littoral combat ships (LCSs) were planned in the late 1990s, designers focused on swappable rapid reconfiguration of combat capabilities through interchangeable mission modules. Anti-submarine and surface warfare mission modules could be interchanged within hours in the presence of an evolving threat. Beyond the LCS's standard littoral combat and protection missions, these "plug-and-fight" mission modules provide significant combat flexibility within a single hull and cost savings in terms of having to maintain a smaller number of ships. In contrast to the traditional shipbuilding approach of cramming a wide-ranging set of bolted-in equipment into fixed installations, flexible ships can radically change the ships' capabilities, by swapping in a full breadth of equipment focused on a particular need (Berkok, Penney, & Kivinen, 2013).

Some examples of MAS and FASO that had been espoused in Navy research literature include decoupling of payloads from platforms, standardizing platform-to-payload interfaces, rapid reconfiguration, preplanned access routes, and sufficient service life allowance for growth (such as in Sturtevant, 2015; Doerry, 2012; Koenig, 2009; Koenig, Czapiewski, & Hootman, 2008; and others). These FASO approaches can be applied to a whole host of systems such as weapons, sensors, aircraft, unmanned vehicles, combat systems, C4I, flexible infrastructure, flexible mission bays and mission decks, vertical launch systems (VLSs) for various multiple missile types, future high-powered surface weapons (laser weapon systems and electromagnetic railguns), and modular payloads (e.g., anti-submarine warfare, special operations, mine warfare, intelligence gathering, close-in weapon systems, harpoon launchers, rigid hull inflatable boats, and gun systems).

More recently, Real Options theory has been recommended for evaluating the value of MAS technologies. Real Options theory proposes to apply financial options and analysis techniques to nonfinancial applications. Ship acquisition programs are characteristic of projects that benefit from investment options. MAS technologies provide those options. Real Options theory projects the value of being able to make decisions in the future when improved information is available to make a better decision. Gregor (2003), Koenig (2009), and Page (2011) provide good insights into the benefits and limitations of applying Real Options theory to naval ship acquisitions.

Flexible and Adaptable Ship Design

Seventy percent of the world is covered by water. To ensure freedom of navigation, economic independence, and national sovereignty, countries must maintain a highly efficient and technologically advanced fleet. With shrinking defense budgets, the current trend is to build fewer warships but maintain the same operational tempo. To continually meet the demands of a larger operational fleet, these new smaller fleets must be built on flexible and adaptable platforms with decoupled payloads that allow the vessel to accomplish a multitude of mission sets. This type of modular design and build "offers an opportunity for a ship to affordably transform its mission systems over its service life to maintain military relevance" (Doerry, 2012). The design characteristics that allow these fleets to flourish are Modular Adaptable Ships (MAS) and Flexible and Adaptable Ship Options (FASO; Mun & Housel, 2016). MAS- and FASO-incorporated designs provide an economical platform for a seagoing navy to build highly effective warships capable of performing various missions in a multitude of environments.

Flexible and adaptable ship designs are centered around a standard hull with modular mission payloads that offer a wide mission set, affordable scalability, reduced operational downtime, increased availability of the ship, and a reduced total number of mission modules for the fleet (Thorsteinson, 2013). For navies with limited budgets, having a flexible and modular platform allows a vessel to perform at times like a frigate and at other times like a corvette (Paris, Brussels, & Fiorenza, 2013). These new multi-mission vessels are already operational in blue-water fleets around the world operated by countries including Denmark, Germany, France, Italy, Australia, and the United States.

Modular build and design have been in use since the mid-20th century. During World War II, Henry Kaiser's shipyards were able to produce Liberty ships in minimal time due in part to the heavy use of modular construction, and the Germans constructed their Type 21 submarines with modular build principles (Abbott, Levine, & Vasilakos, 2008). Starting in 1979, the German shipyard Blohm & Voss began building modular corvettes and frigates for third-world navies using a modular concept known as MEKO. The MEKO concept has continually evolved with time producing the more mature MEKO A-100, A-200, and now A-400. In 1986, the Royal Danish Navy (RDN) began implementation of a modular concept called STANFLEX for a new class of patrol craft (Abbott et al., 2008) known as the Flyvefisken (SF 300) class. The specific use of modular mission payload within the SF 300s directly translated into the future design and development of the RDN Absalon support ships and Iver Huitfeldt class frigates. The French and Italians have worked together to design a flexible multi-mission frigate known as the FREMM class, while the Australian Royal Navy has the modular Anzac class of frigates and Hobart class of air-warfare destroyers (AWDs).

Cost reduction may come in the form of increases in timely switches between capabilities that may also affect procurement, maintenance, and operating costs. For instance, mission modules stored at forward hubs or onboard, while increasing ship costs, may reduce the number of trips to the home port significantly. Moreover, since modularity severs the jointness of various capabilities onboard the platform, a given module can be upgraded relatively independently of others (Berkok et al., 2013).

For many years, ships have been constructed in a modular fashion. That is, significant portions of the ships are built as modules, and the modules are then put together as a final assembly. But "modular capability focuses not on the overall construction of the ship but, rather, on the rapid plug-and-play installation of capabilities such as guns, missiles, unmanned vehicles, SONARs, special forces accommodations, etc." (MacKenzie & Tuteja, 2006). In addition, according to the authors, there are three key modular design types within the naval ship context:

- STANFLEX concept of the Royal Denmark Navy
- MEKO concept of Blohm + Voss GmbH
- Modular Platform Concept (MOPCO) of Abeking & Rasmussen (A&R)

Of the three design types, MEKO is most popular internationally. MEKO vessels are employed by Australia, Turkey, Greece, Germany, South Africa, and other countries. It is interesting to note that although MEKO naval ships are modular in design, there is little evidence that modularity is actually being used in the operation of these vessels. The main benefits of modularity and flexible ships include the following:

- Operational flexibility (i.e., the ability to reconfigure ship for various missions)
- Increased availability of the ship (i.e., reduced operational downtime)
- Reduced total number of mission modules for the fleet, resulting in cost savings

MEKO platforms are designed specifically for the varied deployment of standardized modules (weapons, electronics, and the ship's technical equipment) which are also connected with the power supply, the air-conditioning and ventilation system, and the data network, for example, via standardized interfaces. All the components needed to run a specific system are accommodated in a single module (MacKenzie &Tuteja, 2006).

Royal Danish Navy

The Royal Danish Navy (RDN) has been at the forefront of modular ship design since 1987 when the first of 14 Flyvefisken class or STANFLEX 300 (SF 300) multi-role vessels (MRVs) were commissioned to replace its fleet of 24 mission-specific ships (eight Fast Attack Craft [FAC], eight patrol boats, and eight mine countermeasure vessels) with a smaller number of multi-role vessels (MRVs; Pike, 2011). The design was based on a standard hull that used modular bays, four interchangeable mission containers, to change mission type through use of the Standard Flex (STANFLEX) concept. The ability to quickly and efficiently swap payload allowed these MRVs to serve the following mission sets: anti-air defense (AAW); anti-surface warfare (ASuW); anti-submarine warfare (ASW); electronic warfare (EW); mine countermeasures (MCM); patrol and surveillance; and pollution control (Pike, 2011). STANFLEX and modular payload allowed for containers to be pre-staged for mission flex while simultaneously reducing downtime for technological upgrades, which could be applied to the appropriate container as opposed to the ship itself. The Flyvefisken class was ultimately decommissioned in October 2010 ("Flyvefisken Class (SF 300), Denmark," n.d.), but the use of the STANFLEX concept played a fundamental role in the design and development of the larger follow-on modular designs seen in the Absalon class littoral support ships and Iver Huitfeldt class frigates.

To meet the rising need for a blue-water navy, the RDN commissioned two flexible support ships, HDMS Absalon in 2004 and HDMS Exbern Snare in 2005. Capitalizing on the success of the STANFLEX design and the use of payload modularity, the Absalon class features five STANFELX container wells located amidships on the weapons deck (Lundquist, 2012; Pike, 2016). Under the various configurations, the Absalon class could be "equipped for naval warfare, land attack, strategic sealift missions, emergency disaster relief or as a hospital ship" ("Absalon," n.d.). The Absalon class demonstrated that modularity could be applied to larger combatant ships and was not localized to smaller littoral ships. Continuing to capitalize on the growing success of its modular techniques, the RDN moved forward with designing and building a flexible and adaptable frigate fleet.

German Navy

At the forefront of modular design for the German Navy is Blohm + Voss. The design concept known as Mehrzweck-Kombination (MEKO), which translates as "multi-purpose combination," has been utilized in ship construction and design since the 1970s. MEKO designs rely heavily on modularity that increases the speed at which the ship can be built and facilitates faster upgrades and refits. The success of the MEKO class can be seen in 13 navies worldwide in various corvettes and frigates (Kamerman, 2015). The modular mission payloads in 20-foot standardized ISO containers create adaptability and flexibility and allow navies to rapidly reconfigure mission type based on operational needs. Modules can be rotated for upgrades and maintenance or passed between ships, which reduces the number of overall payloads required for the fleet. This simple reduction results in significant cost savings in procurement and maintenance over the life cycle of the ship ("ThyssenKrupp," n.d.). The MEKO class is the backbone for the new German frigate class, the Baden-Württemberg (F125).

The German Navy will acquire four Baden-Württemberg class frigates to replace the eight frigates in the Bremen class (F122) commissioned in the 1980s. The Baden-Württemberg frigate design incorporates enhanced survivability capabilities that include floating, moving, and fighting after sustaining damage; embarking and deploying special forces; and maintaining prolonged periods at sea with little maintenance. It also incorporates modular mission capabilities (Kamerman, 2015). The design flexibility of the F125 will double the availability of the current German frigate fleet (Kamerman, 2015) while simultaneously reducing overhead.

American Navy

Littoral Combat Ship

As the U.S. Navy began to phase out its fleet of 51 Oliver Hazard Perry class frigates, its leadership began to look for a high-tech platform that could be used as a replacement (Osborn, 2015). The end result was the Littoral Combat Ship (LCS) in two variants: the trimaran-hull Independence class and the mono-hull Freedom class. The concept of the LCS was a highly flexible and adaptable ship that would allow the U.S. Navy to operate in littoral areas with a focus on maritime security and anti-piracy (Stashwick, 2016). "The ships were designed to be high-speed (over 40 knots) and highly maneuverable, with the ability to swap out modules to provide mission-specific capabilities like anti-submarine, anti-surface, and mine-clearing" (Stashwick, 2016). Both variants of the LCS included a mission bay in the design to house elements of mission packages. Within the LCS class, "mission packages are composed of mission modules, aircraft, and crew detachments to support the mission modules and aircraft" (Doerry, 2012).

The Freedom class is built on a steel monohull and is capable of incorporating three mission packages: anti-submarine warfare (ASW), anti-surface warfare (ASuW), and mine countermeasures (MCM). The flexible design "incorporates a large reconfigurable seaframe to allow rapidly interchangeable mission modules, a flight deck with integrated helicopter launch, recovery, and handling systems and the capability to launch and recover maritime vehicles (manned and unmanned) from the stern side" ("Freedom," 2016). The low total gross weight allows the LCS class to obtain speeds greater than 40 knots. The design trade-off for speed was sustained battle damage capability. Where other surface combatants could withstand and potentially recover from sustained-high intensity conflict, the Freedom class would likely result in abandonment of the vessel in the same type of conflict (Stashwick, 2016).

The Independence class of littoral combat ships was designed on an aluminum trimaran hull form based on a commercial ferry design used by the Norwegian company Fred Olsen ("Independence," 2016). As with the Freedom class, the Independence is capable of performing three mission packages: anti-submarine warfare (ASW), anti-surface warfare (ASuW), and mine countermeasures (MCM). The flexible design incorporates a "reconfigurable seaframe to allow rapidly interchangeable mission modules" that include three modular weapon stations ("Independence," 2016).

Small Surface Combatant

The U.S. Navy originally contracted 52 littoral combat ships, but in January 2014, Secretary of Defense Chuck Hagel "instructed the Navy that there would be no new contracts awarded for LCS production beyond 32 ships" (Osborn, 2015). In place of the remaining 20 littoral combat ships, the Navy was to build a Small Surface Combatant ship. On January 15, 2015, Secretary of the Navy Mabus stated that a new class of ship was required to have frigate-like capabilities and thus would change the designation of the last 20 ships from LCS to FF (Osborn, 2015).

The new frigate will capitalize on the two existing hull variants used in the Freedom and Independence classes (Eckstein, 2015). Speaking at an American Society for Naval Engineers event, CAPT Dan Brintzinghoffer stated that the new frigate would take the basic LCS design but differ in that it "will be more lethal, more survivable, and will be able to conduct surface warfare and anti-submarine warfare simultaneously, whereas the LCS had to choose only one mission package to work with at any given time" (Eckstein, 2015). The new class of frigates will trade the high-speed capability of the LCS class in order to accommodate the additional weight created by the heavier armor for increased survivability (Eckstein, 2015). To make the new frigate class cost-efficient, CAPT Brintzinghoffer stated that commonality will be required across both variants, and it will likely need to share some modular aspects with the LCS class or some commonalities with other classes of surface combatants (Eckstein, 2015). This new class of frigates represents an opportunity for the U.S. Navy to build on an existing hull, capitalizing on cost savings in the early design process, and incorporate more advanced flexible and adaptable modules in the payload design. A proven example of a flexible and adaptable frigate with modular payload with similar tonnage is the MEKO A-200 class frigate. The MEKO class family is designed for sustained battle damage and could provide guidance for enhanced survivability options for this new class of American frigates.

FASO/MAS at PEO-SHIPS: Guided Missile Destroyers

DDG 51 FLIGHT III

The Arleigh Burke class of Guided Missile Destroyers (DDG) is the U.S. Navy's first class of destroyer built around the Aegis Combat System and the SPY-1D multi-function passive electronically scanned array radar. The class is named for Admiral Arleigh Burke, the most famous American destroyer officer of World War II and later chief of naval operations. The class leader, USS *Arleigh Burke*, was commissioned during Admiral Burke's lifetime (Office of the Director, Operational Test and Evaluation [DOT&E], 2013).

The DDG class ships were designed as multi-mission destroyers to fit the AAW role with their powerful Aegis radar and surface-to-air missiles; the anti-submarine warfare (ASW) role with their towed sonar array, anti-submarine rockets, and ASW helicopter; the anti-surface warfare (ASuW) role with their Harpoon missile launcher; and the strategic land strike role with their Tomahawk missiles. With upgrades to AN/SPY-1 phased radar systems and their associated missile payloads, as part of the Aegis Ballistic Missile Defense System, members of this class have also begun to demonstrate some promise as mobile anti-ballistic missile and anti-satellite weaponry platforms. Some versions of the class no longer have the towed sonar or Harpoon missile launcher (DOT&E, 2013).

The DDG 51 class destroyers have been designed to support carrier strike groups, surface action groups, amphibious groups, and replenishment groups. They perform primarily AAW with secondary land attack, ASW, and ASuW capabilities. The MK41 vertical launch system has expanded the role of the destroyers in strike warfare, as well as their overall performance. The U.S. Navy will use the DDG 51 Flight III Destroyer equipped with the Aegis modernization program and AMDR to provide joint battlespace threat awareness and defense capability to counter current and future threats in support of joint forces ashore and afloat.

Step 1: Identification of FASO/MAS Options

The following provides two high-level examples of identifying and framing strategic flexibility options in the DDG 51 and DDG1000 environments. These are only notional examples with rough order magnitude values to illustrate the options framing approach.

This real options example illustrates the implications of the standard LM2500 GE Marine Gas Turbines for DDG 51 FLT III ships versus the Rolls-Royce MT30 Marine Gas Turbine Engines for the Zumwalt DDG 1000, where the latter can satisfy large power requirements in warships. The LM2500 provides 105,000 shaft hp for a four-engine plant. In comparison, the MT30 can generate upwards of 35.4 MW, and its auxiliary RR4500 Rolls-Royce turbine generators can produce an added 3.8 MW, and each DDG1000 carries two MT30s and two RR4500s. This means that the combined energy output from the Zumwalt can fulfill the electricity demands in a small- to medium-sized city. In contrast, two LM2500 gas turbines can only produce a total of 95.2 kW, which is approximately 0.12% or 1/825 of the power the Zumwalt can produce. Manufacturer specifications indicate that the LM2500 has an associated Cost/kW of energy of $0.34 and the MT30 Cost/kW is $0.37. In addition, the MT30 prevents warships from running off balance when an engine cannot be restarted until it has cooled down, as is the case in the LM2500.

Figure 1 illustrates a real options strategy tree with four mutually exclusive paths. Additional strategies and pathways can be similarly created, but these initial strategies are sufficient to illustrate the options framing approach. Path 1 shows the As-Is strategy, where no additional higher capacity power plant is used; that is, only two standard LM2500 units are deployed, maintain zero design margins for growth, and only the requirements for the current ship configuration are designed and built. Medium and large upgrades will require major ship alterations, with high cost and delayed schedule. Path 2 implements the two required LM2500 units with additional and sufficient growth margins for one MT30 power plant but currently only with a smaller power plant incorporated into the design. Sufficient area or modularity is available where parts of the machinery can be removed and replaced with the higher energy production unit if needed. Upfront cost is reduced, while future cost and schedule delays are also reduced. Path 3 is to have two prebuilt MT30s and RR4500s initially. While providing the fastest implementation pathway, the cost is higher in the beginning, but the total cost is lower if, indeed, higher energy weapons will be implemented. Path 4 is an option to switch whereby one LM2500 is built with one MT30 unit. Depending on conditions, either the LM2500 or MT30 will be used (switched between units). When higher-powered future weapons are required, such as electromagnetic railguns (EM Rail Guns) or high-intensity lasers (HI Lasers) as well as other similarly futuristic weapons and systems, the MT30 can be turned on.

Having a warship flexibility with two LM2500s (As-Is base case) allows the Navy a savings of $31.76 million by deferring the option of the other two additional LM2500s. Therefore, by having a flexible ship, the Navy can invest later in one LM2500 and attach another MT30 (preventing any engine off-balance effects when the engines cannot be restarted due to excessive heat) and can save $34.58 million. The usage of options to defer/invest that combine gas turbine specifications allows the Navy to prevent high sunk costs, properly adjusting the true kW requirements, and allows different combinations of propulsion and energy plants. This analysis can be further extended into any direction as needed based on ship designs and Navy requirements.

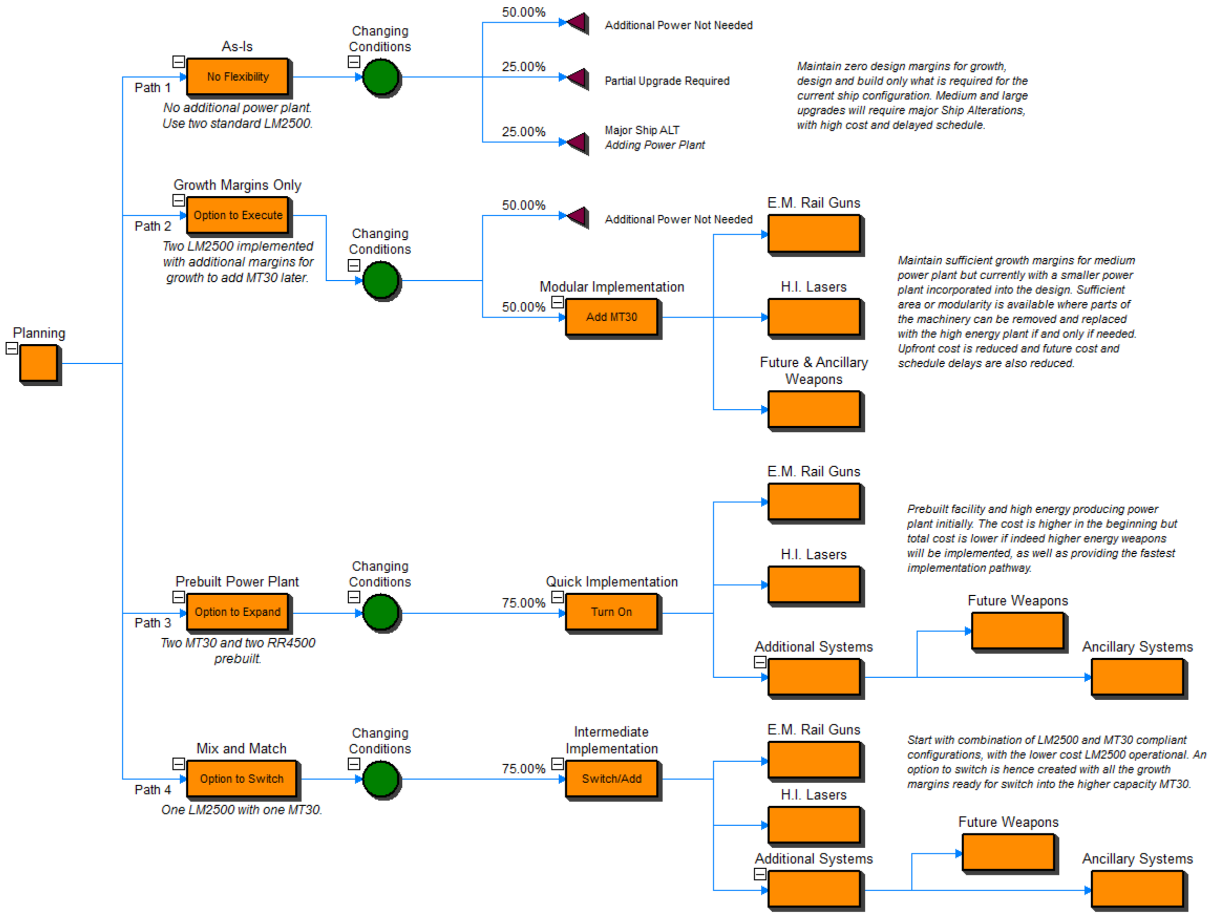


Figure 1. **Options Framing on Power Generation**

Vertical Launch Systems

Another concern of the DoD is the large capital investment required in vertical launch systems (VLSs) in U.S. Navy ships. VLSs need to be developed and integrated per Navy requirements, which are constrained by rapid technological change and high uncertainties in costs. The usage of strategic real options aims to assess whether the Navy can *keep the option open* to defer the large investments to help avoid high sunk costs and quick technological obsolescence or should pre-invest in a new VLS. Consequently, flexibility and uncertainty create the right environment to model VLSs using a real options framework. According to DDG 51 (Flight II and Flight III) specifications, the estimated cost of a single VLS is approximately $228 million. The most expensive subarea is the MK41 subsystem (DDG 51 contains two MK41s). This current example is developed based on the assumptions of a rapid technological obsolescence, high integration costs, time delays, and reduced capability, which can all jeopardize Navy investments in the VLS.

In addition, using a real options framework to possibly defer the implementation of MK41 would allow ship designers and engineers to incorporate modernization and upgrade margins in the VLS within the ship design early on, and to defer the exact configuration of the VLS until a future date when uncertainties on capability requirements are resolved over the passage of time, action, and events. These capability requirements include integration, upgrades, changes, new technology, new requirements, and updated military warfighter needs. Also, we can evaluate the option to invest in the second or third MK41 as situational needs arise. Figure 2 shows two simple option paths, in which the first path indicates immediate execution, where two MK1s are implemented immediately, not knowing if both are actually needed, as opposed to the second strategic path, where the VLS is designed so that either two MK1s can be implemented or only one. Therefore, one MK1 can be first inserted and the second added on later only when required, where the VLS has design growth margins to adapt to slightly different technological configurations. The question, of course, is which strategic pathway makes most sense, as computed using strategic real options value.

When the flexibility value is added into the mix, the expected total cost is reduced from $110.10 million to $98.51 million. Finally, wartime scenarios can be incorporated into the analysis whereby if there is a higher probability of conflict where the VLS is required, the value to keep open the option to defer is reduced and the Navy is better off executing the option immediately and having the required VLS in place.

The project with flexibility is $118.22 million (flexible VLS warship open to integrate another MK41 in the future as and when needed) against $228.34 million (base case DDG 51 with no flexibility options, where the VLS is already built in). The Navy can save or delay the usage of $110.10 million by holding on to the option of deferring the second MK41. In addition, in the near future, the cost to implement the second MK41 can be reduced due to a flatter learning curve, economies of scale, and the specific technology becoming more readily available, less complex, and easier to implement, or it can be more expensive because the technology experiences new updates, higher performance, and greater efficiency. If cost volatility is the main variable for the Navy, we contrast deferring the second MK41 against the base case. It means that we compare the VLS system with no flexibility ($228.32 million) against the cost changes in the second MK41 (assuming Navy engineers develop a plug-and-play structure to integrate the next MK41 quickly).

This assumption can be relaxed using cost and schedule modeling and Monte Carlo simulation methods. In terms of the options valuation, the option to defer for the Navy follows cost comparisons. In other words, it reduces the cost exposure for the second MK41 from $110.10 million to an expected cost of $69.89 million. In addition, decision-makers observe in the options strategy tree and decision tree where they can keep the option to defer *open* and under what conditions the Navy should execute and invest in the second MK41. One likely extension is where the decision-maker can introduce probabilities or expectations of Navy actions (new missions and new requirements) or events (wartime, peacetime). This affects the flexibility of the second MK41 by constraining the option's flexibility to defer. For instance, if the Navy has strong expectations of requiring the second MK1 (wartime probability is higher than 30%), it reduces the value of the option to defer and accelerates the availability and execution of the second MK41 option earlier. In peacetime, the Navy has more flexibility in terms of how it implements or assesses its real options to wait and defer.

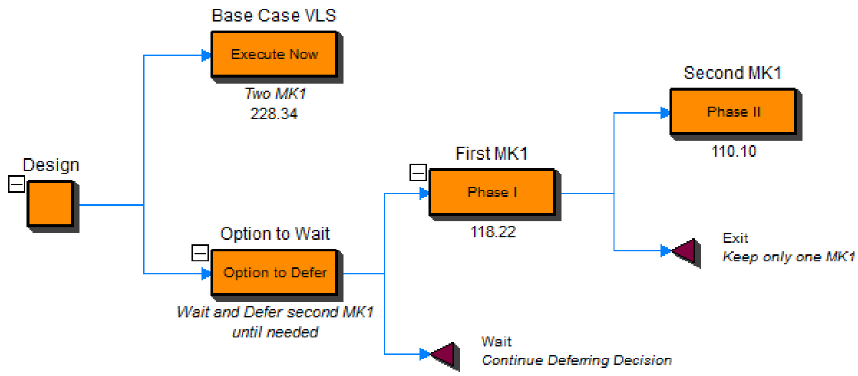


Figure 2. **Options Framing on Vertical Launch Systems**

Step 2: Cost Analysis and Data Gathering

Once the various FASO/MASO options are framed and modeled, as shown in the previous step, the modeling process continues with additional data gathering activities. Figure 3 shows some examples of shadow revenues (i.e., cost savings from lowered cost of future upgrades and technology insertions, costs mitigated by reducing the need for alternative equipment and lower spare parts, and other costs deferred by reducing the need for maintenance and operating costs) or cost savings, additional direct and indirect costs of implementing the new option, and capital requirements.

Step 3: Financial Modeling

The *Discounted Cash Flow* section, shown in Figure 3, is at the heart of the input assumptions for the analysis. Analysts would enter their input assumptions—such as starting and ending years of the analysis, the discount rate to use, and the marginal tax rate—and set up the project economics model (adding or deleting rows in each subcategory of the financial model). Additional time-series inputs are entered in the data grid as required, while some elements of this grid are intermediate computed values.

Analysts can also identify and create the various options and compute the economic and financial results, such as net present value (NPV), internal rate of return (IRR), modified internal rate of return (MIRR), profitability index (PI), return on investment (ROI), payback period (PP), and discounted payback period (DPP). This is shown in Figures 3 and 4, complete with various charts, cash flow ratios and models, intermediate calculations, and comparisons of the options within a portfolio view, as illustrated in the figure. As a side note, the term *Option* is used to represent a generic analysis option, where each project can be a different asset, project, acquisition, investment, research and development, or simply variations of the same investment (e.g., different financing methods when acquiring the same firm, different market conditions and outcomes, or different scenarios or implementation paths). Therefore, the more flexible terminology of *Project* is adopted instead.

Figure 5 illustrates the *Economic Results* of each project. This figure shows the results from the chosen project and returns the NPV, IRR, MIRR, PI, ROI, PP, and DPP. These computed results are based on the analyst's selection of the discounting convention, if there is a constant terminal growth rate, and the cash flow to use (e.g., net cash flow versus net income or operating cash flow). An *NPV Profile* table and chart are also provided in the figure, where different discount rates and their respective NPV results are shown and charted. The *Economic Results* shown are for each individual project, whereas the *Portfolio Analysis* (see Figure 5) compares the economic results of all projects at once. The *Terminal Value Annualized Growth Rate* is applied to the last year's cash flow to account for a perpetual constant growth rate cash flow model, and these future cash flows, depending on the cash flow type chosen, are discounted back to the base year and added to the NPV to arrive at the perpetual valuation.

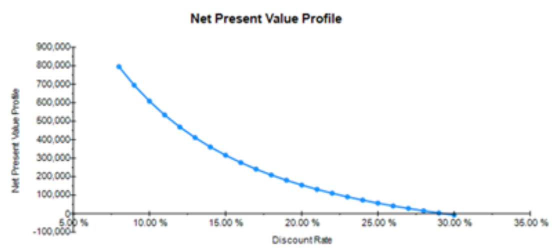
Static Portfolio Analysis and Comparisons of Multiple Projects

Figure 5 illustrates the *Portfolio Analysis* of multiple projects. This Portfolio Analysis returns the computed economic and financial indicators such as NPV, IRR, MIRR, PI, ROI, PP, and DPP for all the projects combined into a portfolio view (these results can be stand-alone with no base case or computed as incremental values above and beyond the chosen base case). The *Economic Results* show the individual project's economic and financial indicators, whereas this Level 2 *Portfolio Analysis* view shows the results of all projects' indicators and compares them side by side. There are also two charts available for comparing these individual projects' results. The *Portfolio Analysis* is used to obtain a side-by-side comparison of all the main economic and financial indicators of all the projects at once. For instance, analysts can compare all the NPVs from each project in a single results grid. The bubble chart on the left provides a visual representation of up to three chosen variables at once (e.g., the *y*-axis shows the IRR, the *x*-axis represents the NPV, and the size of the bubble may represent the capital investment; in such a situation, one would prefer a smaller bubble that is in the top right quadrant of the chart). These charts have associated icons that can be used to modify their settings (chart type, color, legend, etc.).

Year	2016	2017	2018	2019	2020	2021	2022	2023	...	2041	2042	2043
Revenues	$1,742.51	$11,737.14	$225,850.13	$225,850.13	$225,850.13	$225,850.13	$225,850.13	$225,850.13	...	$235,437.44	$235,437.44	$235,437.44
Cost Savings (Future Upgrades and Insertion)	$1,132.63	$7,629.14	$146,802.58	$146,802.58	$146,802.58	$146,802.58	$146,802.58	$146,802.58	...	$153,034.34	$153,034.34	$153,034.34
Cost Mitigated (Alternative Equipment)	$522.75	$3,521.14	$67,755.04	$67,755.04	$67,755.04	$67,755.04	$67,755.04	$67,755.04	...	$70,631.23	$70,631.23	$70,631.23
Cost Deferred (Maintenance and Operations)	$87.13	$586.86	$11,292.51	$11,292.51	$11,292.51	$11,292.51	$11,292.51	$11,292.51	...	$11,771.87	$11,771.87	$11,771.87
Direct Costs	$1,141.09	$1,141.09	$26,392.75	$26,392.75	$26,392.75	$26,456.81	$27,888.82	$27,888.82	...	$32,021.41	$32,021.41	$32,021.41
Direct Expenses	$1,110.26	$1,110.26	$24,896.68	$24,896.68	$24,896.68	$24,896.68	$24,896.68	$24,896.68	...	$25,961.75	$25,961.75	$25,961.75
Operational Costs	$18.50	$18.50	$414.95	$414.95	$414.95	$453.38	$829.89	$829.89	...	$1,730.79	$1,730.79	$1,730.79
Maintenance	$12.33	$12.33	$25.62	$25.62	$25.62	$51.25	$51.25	$51.25	...	$106.87	$106.87	$106.87
Direct Expenses	$0.00	$0.00	$1,055.50	$1,055.50	$1,055.50	$1,055.50	$2,111.00	$2,111.00	...	$4,222.00	$4,222.00	$4,222.00
Gross Profit (Operating Income)	$601.42	$10,596.05	$199,457.38	$199,457.38	$199,457.38	$199,393.32	$197,961.31	$197,961.31	...	$203,416.03	$203,416.03	$203,416.03
Indirect Expenses (General & Administrative)	$799.42	$3,073.28	$9,212.61	$9,212.61	$9,212.61	$9,212.61	$9,212.61	$10,877.49	...	$12,259.92	$9,465.41	$9,465.41
Training and Administrative	$0.00	$31.00	$703.00	$703.00	$703.00	$703.00	$703.00	$703.00	...	$733.00	$733.00	$733.00
Contracts and Bidding	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	...	$0.00	$0.00	$0.00
Operations	$0.00	$0.00	$1,248.07	$1,248.07	$1,248.07	$1,248.07	$1,248.07	$1,248.07	...	$1,248.07	$1,248.07	$1,248.07
Maintenance	$799.42	$2,997.82	$4,758.48	$4,758.48	$4,758.48	$4,758.48	$4,758.48	$6,423.36	...	$7,733.14	$4,938.63	$4,938.63
Parts and Service	$0.00	$0.00	$1,506.00	$1,506.00	$1,506.00	$1,506.00	$1,506.00	$1,506.00	...	$1,506.00	$1,506.00	$1,506.00
Miscellaneous	$0.00	$44.46	$997.06	$997.06	$997.06	$997.06	$997.06	$997.06	...	$1,039.71	$1,039.71	$1,039.71
EBITDA: Earnings Before Interest, Taxes, Depreciation, and Amortization	($198.00)	$7,522.77	$190,244.77	$190,244.77	$190,244.77	$190,180.71	$188,748.70	$187,083.82	...	$191,156.11	$193,950.62	$193,950.62
Depreciation	$0.00	$9,874.00	$39,827.00	$39,074.00	$38,161.00	$37,206.00	$36,172.00	$35,223.00	...	$24,502.00	$23,977.00	$23,444.00
Amortization	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	...	$0.00	$0.00	$0.00
EBIT: Earnings Before Interest and Taxes	($198.00)	($2,351.23)	$150,417.77	$151,170.77	$152,083.77	$152,974.71	$152,576.70	$151,860.82	...	$166,654.11	$169,973.62	$170,506.62
Interest	$0.00	$6,779.32	$25,892.66	$22,767.15	$19,224.35	$15,842.53	$13,062.00	$12,303.79	...	$653.99	$666.90	$667.48
EBT: Earnings Before Taxes	($198.00)	($9,130.55)	$124,525.11	$128,403.62	$132,859.42	$137,132.18	$139,514.70	$139,557.03	...	$166,000.12	$169,306.72	$169,839.14
Corporate Taxes	($56.43)	($2,602.21)	$35,489.66	$36,595.03	$37,864.93	$39,082.67	$39,761.69	$39,773.75	...	$47,310.03	$48,252.42	$48,404.15
NET INCOME	($141.57)	($6,528.34)	$89,035.45	$91,808.59	$94,994.49	$98,049.51	$99,753.01	$99,783.28	...	$118,690.09	$121,054.30	$121,434.99
Total Noncash Expense Items	$0.00	$9,874.00	$39,827.00	$39,074.00	$38,161.00	$37,206.00	$36,172.00	$35,223.00	...	$24,502.00	$23,977.00	$23,444.00
Change in Net Working Capital	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	...	$0.00	$0.00	$0.00
Capital Expenditures	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	...	$0.00	$0.00	$0.00
Other Noncash Expenses	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	...	$0.00	$0.00	$0.00
Total Gross Invested Operating Capital	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	...	$0.00	$0.00	$0.00
CAPITAL INVESTMENTS	$250,000.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	...	$0.00	$0.00	$0.00
NET OPERATING PROFIT AFTER TAXES (NOPAT)	($141.57)	($1,681.13)	$107,548.71	$108,087.10	$108,739.90	$109,376.92	$109,092.34	$108,580.49	...	$119,157.69	$121,531.14	$121,912.23
NET CASH FLOW (NCF)	($141.57)	$3,345.66	$128,862.45	$130,882.59	$133,155.49	$135,255.51	$135,925.01	$135,006.28	...	$143,192.09	$145,031.30	$144,878.99
OPERATING CASH FLOW (OCF)	($141.57)	$8,192.87	$147,375.71	$147,161.10	$146,900.90	$146,582.92	$145,264.34	$143,803.49	...	$143,659.69	$145,508.14	$145,356.23
FREE CASH FLOW (FCF)	($141.57)	$8,192.87	$147,375.71	$147,161.10	$146,900.90	$146,582.92	$145,264.34	$143,803.49	...	$143,659.69	$145,508.14	$145,356.23

Figure 3. **Financial Economic Cost Savings and Aversion Cash Flow Model**

ECONOMIC INDICATORS	RESULTS
Net Present Value (NPV)	$608,388.34
Net Present Value (NPV) with Terminal Value	$726,488.77
Internal Rate of Return (IRR)	29.31%
Modified Internal Rate of Return (MIRR)	15.07%
Profitability Index (PI)	3.43
Return on Investment (ROI)	243.36%
Payback Period (PP)	3.80
Discounted Payback Period (DPP)	4.80



NPV PROFILE	
DISCOUNT %	NPV
8.00%	$794,590.13
9.00%	$694,674.49
10.00%	$608,388.34
11.00%	$533,487.10
12.00%	$468,141.66
13.00%	$410,854.85
14.00%	$360,395.93
15.00%	$315,748.74
16.00%	$276,070.74
17.00%	$240,660.35
18.00%	$208,930.96
19.00%	$180,389.99
20.00%	$154,622.20
21.00%	$131,276.09
22.00%	$110,052.97
23.00%	$90,698.01
24.00%	$72,993.02
25.00%	$56,750.42
26.00%	$41,808.40
27.00%	$28,026.89
28.00%	$15,284.18
29.00%	$3,474.16
30.00%	($7,495.93)

Figure 4. **Financial and Economic Results**

Economic Results	Option 1	Option 2	Option 3	Option 4	Option 5	Option 6	Option 7	Option 8	Option 9	Option 10
Net Present Value (NPV)	$608,388.34	$427,132.76	$40,765.22	$41,613.74	($10,610.44)	($23,774.85)	$728,339.38	$554,258.99	$31,837.41	$46,377.25
Net Present Value (NPV) with Terminal Value	72648877.00%	54046710.00%	7033594.00%	6615455.00%	1525722.00%	2718633.00%	112457959.00%	74402419.00%	16876439.00%	12973950.00%
Internal Rate of Return (IRR)	29.31%	11.17%	16.21%	19.21%	8.55%	6.76%	11.20%	13.74%	9.29%	14.77%
Modified Internal Rate of Return (MIRR)	0.15	0.10	0.12	0.13	0.09	0.08	0.10	0.11	0.07	0.10
Profitability Index (PI)	343.00%	114.00%	137.00%	154.00%	90.00%	86.00%	109.00%	129.00%	134.00%	176.00%
Return on Investment (ROI)	2.43	0.14	0.37	0.54	-0.10	-0.14	0.09	0.29	0.34	0.76
Payback Period (PP)	3.80	10.82	6.38	5.45	10.32	9.78	9.98	7.87	9.19	7.30
Discounted Payback Period (DPP)	$4.80	$22.81	$9.80	$7.84			$22.35	$13.79	$12.18	$9.21
Show on Charts										

Figure 5. **Static Portfolio Analysis**

Step 4: Tornado and Sensitivity Analytics

Figure 6 illustrates the *Applied Analytics* results, which allows analysts to run *Tornado Analysis* and *Scenario Analysis* on any one of the projects previously modeled—the analytics cover all the various projects and options. We can, therefore, run tornado or scenario analyses on any one of the projects or options. Tornado analysis is a static sensitivity analysis of the selected model's output to each input assumption, performed one at a time, and ranked from most impactful to least impactful. We can start the analysis by first choosing the output variable to test.

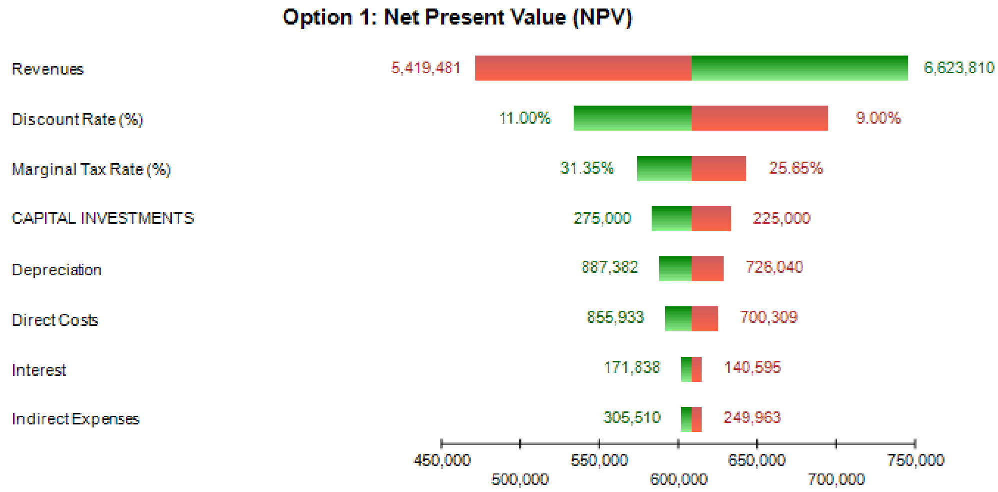


Figure 6. **Applied Analytics—Tornado Analysis**

We used the default sensitivity settings of ±10% on each input assumption to test and decide how many input variables to chart (large models with many inputs may generate unsightly and less useful charts, whereas showing just the top variables reveals more information through a more elegant chart). Analysts can also choose to run the input assumptions as unique inputs, group them as a line item (all individual inputs on a single line item are assumed to be one variable), or run as variable groups (e.g., all line items under Revenue will be assumed to be a single variable). Analysts will need to remember to click Update to run the analysis if they make any changes to any of the settings. The sensitivity run was based on the input assumptions as unique inputs, but the inputs can also be grouped as a line item (all individual inputs on a single line item are assumed to be one variable), or the analysis can be run as variable groups (e.g., all line items under Revenue will be assumed to be a single variable). The following summarizes the tornado analysis chart's main characteristics:

- Each horizontal bar indicates a unique input assumption that constitutes a precedent to the selected output variable.
- The *x*-axis represents the values of the selected output variable. The wider the bar chart, the greater the impact/swing the input assumption has on the output.
- A green bar on the right indicates that the input assumption has a positive effect on the selected output (conversely, a red bar on the right indicates a negative effect).
- Each of the precedent or input assumptions that directly affect the NPV with Terminal Value is tested ±10% by default (this setting can be changed); the top 10 variables are shown on the chart by default (this setting can be changed), with a 2-decimal precision setting; and each unique input is tested individually.
- The default sensitivity is globally ±10% of each input variable, but each of these inputs can be individually modified in the data grid. Note that a larger percentage variation will test for nonlinear effects as well.

- The model's granularity can be set (e.g., Variable Groups look at an entire variable group, such as all revenues, and will be modified at once; Line Items change the entire row for multiple years at once; and Individual Unique Inputs look at modifying each input cell).

Figure 7 illustrates the *Scenario Analysis* results, where the scenario analysis can be easily performed through a two-step process: Identify the model input settings and run the model to obtain scenario output tables. In the *Scenario Input settings*, analysts start by selecting the output variable they wish to test from the droplist. Then, based on the selection, the precedents of the output will be listed under two categories (*Line Item*, which will change all input assumptions in the entire line item in the model simultaneously, and *Single Item*, which will change individual input assumption items). Analysts select one or two checkboxes at a time, along with the inputs they wish to run scenarios on, and enter the plus/minus percentage and the number of steps between these two values to test. Analysts can also add color coding of sweetspots or hotspots in the scenario analysis (values falling within different ranges have unique colors). Analysts can create multiple scenarios and *Save As* each one (enter a name and model notes for each saved scenario).

Scenario analysis results can sometimes be used as heat maps to identify the combinations of input parameter conditions whereby the calculated outputs will be above or below certain thresholds. A visual heat map can be created by adding color thresholds in the scenario results table:

- Create and run scenario analysis on either one or two input variables at once.
- Scenario settings can be saved for future retrieval, which means analysts can modify any input assumptions in the options models and come back to rerun the saved scenarios.
- Increase/decrease decimals in the scenario results tables, as well as change colors in the tables for easier visual interpretation (especially when trying to identify scenario combinations, or so-called sweetspots and hotspots).
- Additional input variables are available by scrolling down the form.
- Line items can be changed using ±X%, where all inputs in the line are changed multiple times within this specific range all at once. Individual items can be changed ±Y units, where each input is changed multiple times within this specific range.
- Sweetspots and hotspots refer to specific combinations of two input variables that will drive the output up or down. For instance, suppose investments are below a certain threshold and revenues are above a certain barrier. The NPV will then be in excess of the expected budget (the sweetspots). Or if investments are above a certain value, NPV will turn negative if revenues fall below a certain threshold (the hotspots).

	20.00%	21.00%	22.00%	23.00%	24.00%	25.00%	26.00%	27.00%	28.00%	29.00%	30.00%	31.00%	32.00%	33.00%	34.00%	35.00%	36.00%	37.00%	38.00%
3,718,366.00	132,576.00	110,413.00	90,269.00	71,903.00	55,105.00	39,697.00	25,526.00	12,458.00	376.90	-10,818.00	-21,215.00	-30,891.00	-39,916.00	-48,348.00	-56,241.00	-63,642.00	-70,595.00	-77,135.00	-83,297.00
3,737,936.00	134,780.00	112,499.00	92,248.00	73,782.00	56,894.00	41,403.00	27,154.00	14,015.00	1,867.60	-9,388.80	-19,843.00	-29,573.00	-38,647.00	-47,126.00	-55,063.00	-62,507.00	-69,498.00	-76,075.00	-82,272.00
3,757,507.00	136,985.00	114,585.00	94,226.00	75,662.00	58,683.00	43,108.00	28,783.00	15,572.00	3,358.40	-7,959.50	-18,471.00	-28,255.00	-37,379.00	-45,905.00	-53,886.00	-61,371.00	-68,401.00	-75,015.00	-81,246.00
3,777,077.00	139,190.00	116,672.00	96,205.00	77,541.00	60,471.00	44,813.00	30,411.00	17,129.00	4,849.10	-6,530.30	-17,099.00	-26,936.00	-36,111.00	-44,683.00	-52,709.00	-60,235.00	-67,304.00	-73,954.00	-80,221.00
3,796,647.00	141,394.00	118,758.00	98,183.00	79,421.00	62,260.00	46,519.00	32,039.00	18,686.00	6,339.80	-5,101.10	-15,727.00	-25,618.00	-34,842.00	-43,462.00	-51,531.00	-59,099.00	-66,207.00	-72,894.00	-79,195.00
3,816,218.00	143,599.00	120,844.00	100,161.00	81,300.00	64,049.00	48,224.00	33,667.00	20,242.00	7,830.50	-3,671.90	-14,355.00	-24,299.00	-33,574.00	-42,241.00	-50,354.00	-57,963.00	-65,110.00	-71,834.00	-78,170.00
3,835,788.00	145,804.00	122,931.00	102,140.00	83,180.00	65,838.00	49,929.00	35,296.00	21,799.00	9,321.30	-2,242.70	-12,984.00	-22,981.00	-32,306.00	-41,019.00	-49,177.00	-56,827.00	-64,013.00	-70,774.00	-77,145.00
3,855,358.00	148,008.00	125,017.00	104,118.00	85,059.00	67,627.00	51,635.00	36,924.00	23,356.00	10,812.00	-813.48	-11,612.00	-21,663.00	-31,037.00	-39,798.00	-47,999.00	-55,691.00	-62,916.00	-69,714.00	-76,119.00
3,874,929.00	150,213.00	127,103.00	106,096.00	86,939.00	69,415.00	53,340.00	38,552.00	24,913.00	12,303.00	615.74	-10,240.00	-20,344.00	-29,769.00	-38,576.00	-46,822.00	-54,555.00	-61,819.00	-68,654.00	-75,094.00
3,894,499.00	152,418.00	129,190.00	108,075.00	88,818.00	71,204.00	55,045.00	40,180.00	26,470.00	13,793.00	2,045.00	-8,867.80	-19,026.00	-28,501.00	-37,355.00	-45,644.00	-53,419.00	-60,722.00	-67,594.00	-74,068.00
3,914,069.00	154,622.00	131,276.00	110,053.00	90,698.00	72,993.00	56,750.00	41,808.00	28,027.00	15,284.00	3,474.20	-7,495.90	-17,708.00	-27,232.00	-36,133.00	-44,467.00	-52,283.00	-59,625.00	-66,533.00	-73,043.00
3,933,640.00	156,827.00	133,362.00	112,031.00	92,578.00	74,782.00	58,456.00	43,437.00	29,584.00	16,775.00	4,903.40	-6,124.00	-16,389.00	-25,964.00	-34,912.00	-43,290.00	-51,147.00	-58,528.00	-65,473.00	-72,018.00
3,953,210.00	159,032.00	135,449.00	114,010.00	94,457.00	76,571.00	60,161.00	45,065.00	31,141.00	18,266.00	6,332.60	-4,752.10	-15,071.00	-24,696.00	-33,691.00	-42,112.00	-50,011.00	-57,431.00	-64,413.00	-70,992.00
3,972,780.00	161,236.00	137,535.00	115,988.00	96,337.00	78,359.00	61,866.00	46,693.00	32,698.00	19,756.00	7,761.80	-3,380.20	-13,752.00	-23,427.00	-32,469.00	-40,935.00	-48,875.00	-56,334.00	-63,353.00	-69,967.00
3,992,351.00	163,441.00	139,621.00	117,966.00	98,216.00	80,148.00	63,572.00	48,321.00	34,254.00	21,247.00	9,191.00	-2,008.30	-12,434.00	-22,159.00	-31,248.00	-39,758.00	-47,739.00	-55,237.00	-62,293.00	-68,942.00
4,011,921.00	165,646.00	141,708.00	119,945.00	100,096.00	81,937.00	65,277.00	49,950.00	35,811.00	22,738.00	10,620.00	-636.44	-11,116.00	-20,891.00	-30,026.00	-38,580.00	-46,603.00	-54,140.00	-61,233.00	-67,916.00
4,031,491.00	167,850.00	143,794.00	121,923.00	101,975.00	83,726.00	66,982.00	51,578.00	37,368.00	24,229.00	12,049.00	735.46	-9,797.20	-19,622.00	-28,805.00	-37,403.00	-45,467.00	-53,043.00	-60,173.00	-66,891.00
4,051,062.00	170,055.00	145,880.00	123,901.00	103,855.00	85,515.00	68,688.00	53,206.00	38,925.00	25,719.00	13,479.00	2,107.40	-8,478.80	-18,354.00	-27,584.00	-36,225.00	-44,331.00	-51,947.00	-59,112.00	-65,865.00
4,070,632.00	172,260.00	147,967.00	125,880.00	105,734.00	87,304.00	70,393.00	54,834.00	40,482.00	27,210.00	14,908.00	3,479.30	-7,160.40	-17,086.00	-26,362.00	-35,048.00	-43,195.00	-50,850.00	-58,052.00	-64,840.00
4,090,203.00	174,464.00	150,053.00	127,858.00	107,614.00	89,092.00	72,098.00	56,462.00	42,039.00	28,701.00	16,337.00	4,851.20	-5,842.00	-15,817.00	-25,141.00	-33,871.00	-42,059.00	-49,753.00	-56,992.00	-63,815.00
4,109,773.00	176,669.00	152,139.00	129,836.00	109,493.00	90,881.00	73,803.00	58,091.00	43,596.00	30,191.00	17,766.00	6,223.10	-4,523.60	-14,549.00	-23,919.00	-32,693.00	-40,923.00	-48,656.00	-55,932.00	-62,789.00

Figure 7. **Applied Analytics—Scenario Tables**

Step 5: Monte Carlo Risk Simulation

Figure 8 illustrates the *Risk Simulation* analysis, where Monte Carlo risk simulations can be set up and run. Analysts can set up probability distribution assumptions on any combination of inputs, run a risk simulation tens to hundreds of thousands of trials, and retrieve the simulated forecast outputs as charts, statistics, probabilities, and confidence intervals to develop comprehensive risk profiles of the projects.

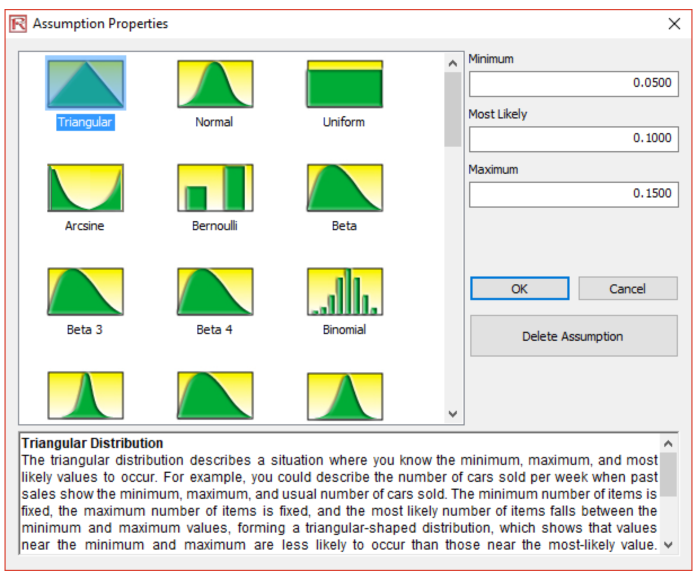


Figure 8. **Monte Carlo Risk Simulation Input Assumptions**

Simulation Results, Confidence Intervals, and Probabilities

Figure 9 illustrates the *Risk Simulation* results. The simulation forecast chart is shown on the right, while percentiles and simulation statistics are presented on the left.

Statistics/Percentiles	Value
Trials	10,000.00
Mean	1,264,569.20
Median	1,223,025.65
Stdev	323,440.89
CV	0.26
Skew	0.59
Kurtosis	0.07
Minimum	528,515.81
Maximum	2,690,456.43
Range	2,161,940.62
0.00%	528,515.81
5.00%	806,158.76
10.00%	873,477.43
20%	$975,191
30%	1,066,998.73
40.00%	1,146,536
50.00%	1,223,025.65
60.00%	1,310,131.64
70.00%	1,408,675.46
80.00%	152992680.50%
90.00%	1,711,752.31
95.00%	1,871,099.08
100.00%	2,690,456.43

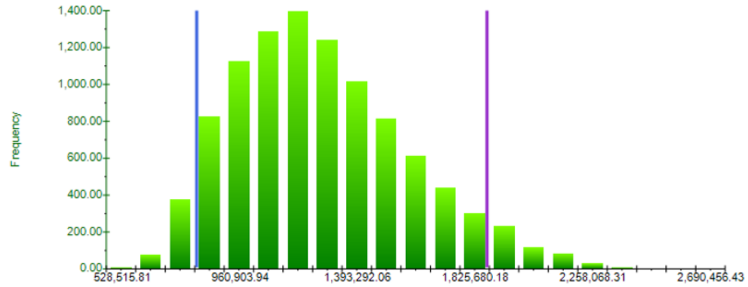


Figure 9. **Monte Carlo Risk Simulation Results**

Probability Distribution Overlay Charts

Figure 10 illustrates the *Overlay Results*. Multiple simulation output variables can be compared at once using the overlay charts. Analysts simply check/uncheck the simulated outputs they wish to compare and select the chart type to show (e.g., S-Curves, CDF, PDF). Analysts can also add percentile or certainty lines by first selecting the output chart, then entering the relevant values, and finally clicking the *Update* button. The generated charts are highly flexible in that analysts can modify them using the included chart icons (as well as whether to show or hide gridlines), and the chart can be copied into the Microsoft Windows clipboard for pasting into another software application. Typically, S-curves or CDF curves are used in overlay analysis when comparing the risk profile of multiple simulated forecast results.

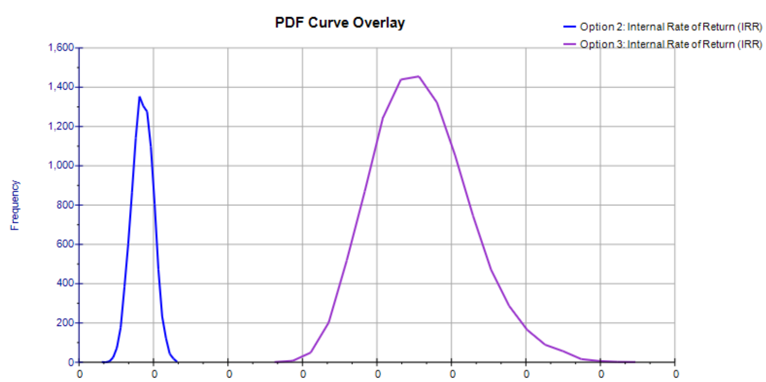


Figure 10. **Simulated Overlay Results**

Analysis of Alternatives and Dynamic Sensitivity Analysis

Figure 11 illustrates the *Analysis of Alternatives* results. Whereas the *Overlay Results* shows the simulated results as charts (PDF/CDF), the *Analysis of Alternatives* shows the results of the simulation statistics in a table format as well as a chart of the statistics so that one project can be compared against another. The standard approach is to run an analysis of alternatives to compare one project to another, but analysts can also choose to analyze the results on an incremental analysis basis.

OPTIONS	Option 1	Option 2	Option 3	Option 4	Option 5	Option 6	Option 7	Option 8	Option 9	Option 10
Acronym										
Mean	1,264,569.20	-129,548.71	115,517.52	104,069.16	186,548.19	50,886.00	982,026.93	2,222,255.83	120,139.43	115,842.87
Median	1,223,025.65	-269,130.82	111,946.33	101,647.69	182,533.06	48,903.31	809,951.34	2,125,984.32	118,821.08	113,873.98
Stdev	323,440.89	756,980.40	33,246.01	27,242.55	42,751.66	24,260.40	1,841,083.18	774,387.88	22,748.49	26,884.24
CV	25.58%	584.32%	28.78%	26.18%	22.92%	47.68%	187.48%	34.85%	18.94%	23.21%
Skew	0.59	0.83	0.50	0.46	0.46	0.41	0.48	0.64	0.31	0.38
Kurtosis	0.07	0.42	-0.06	-0.02	-0.05	-0.06	0.18	0.29	-0.06	-0.02
Minimum	528,515.81	-1,612,389.19	33,915.58	34,781.88	72,360.45	-13,544.30	-4,062,019.59	409,940.95	53,294.46	45,687.22
Maximum	2,690,456.43	3,186,736.54	238,613.77	208,832.14	361,534.33	150,216.66	8,105,902.62	5,492,144.74	213,754.98	234,987.45
Range	2,161,940.62	4,799,125.72	204,698.19	174,050.27	289,173.87	163,760.96	12,167,922.22	5,082,203.79	160,460.52	189,300.23
0% Percentile	528,515.81	-1,612,389.19	33,915.58	34,781.88	72,360.45	-13,544.30	-4,062,019.59	409,940.95	53,294.46	45,687.22
5% Percentile	806,158.76	-1,126,810.37	66,833.21	63,341.55	123,647.00	14,407.25	-1,721,563.12	1,118,422.25	85,021.02	74,880.17
10% Percentile	873,477.43	-985,321.30	75,292.44	70,536.07	134,523.19	21,061.66	-1,234,281.88	1,308,170.97	91,480.18	82,478.56
20% Percentile	975,191.16	-787,408.29	86,592.52	80,551.23	149,044.20	29,596.95	-575,471.45	1,553,575.46	100,319.83	92,352.05
30% Percentile	1,066,998.73	-607,802.26	95,258.85	87,689.17	160,430.73	36,309.09	-88,078.13	1,756,505.72	107,072.89	100,314.71
40% Percentile	1,146,536.48	-435,048.39	103,478.90	94,871.67	171,896.55	42,716.18	366,415.18	1,937,629.66	113,197.08	106,958.95
50% Percentile	1,223,025.65	-269,130.82	111,946.33	101,647.69	182,533.06	48,903.31	809,951.34	2,125,984.32	118,821.08	113,873.98
60% Percentile	1,310,131.64	-70,640.29	120,924.23	108,733.13	193,556.04	55,280.95	1,260,274.23	2,321,587.17	124,955.05	121,112.91
70% Percentile	1,408,675.46	168,159.10	130,549.63	116,448.61	206,334.42	62,431.52	1,795,294.68	2,549,885.15	131,348.03	128,934.06
80% Percentile	1,529,926.80	474,006.49	143,040.93	127,054.19	222,193.22	71,595.89	2,464,689.40	2,848,749.31	139,239.82	138,456.24
90% Percentile	1,711,752.31	956,928.86	161,722.53	141,005.97	245,083.83	83,760.91	3,475,890.85	3,275,508.70	150,031.84	151,426.79
95% Percentile	1,871,099.08	1,333,426.70	177,165.94	153,294.40	263,658.89	93,827.83	4,368,679.78	3,668,268.40	159,781.91	163,466.89
100% Percentile	2,690,456.43	3,186,736.54	238,613.77	208,832.14	361,534.33	150,216.66	8,105,902.62	5,492,144.74	213,754.98	234,987.45

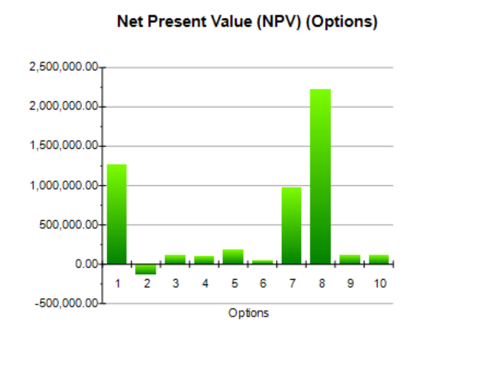


Figure 11. **Simulated Analysis of Alternatives**

Step 6: Strategic Real Options Valuation Modeling

Figure 12 illustrates the *Options Strategies* tab. *Options Strategies* is where analysts can draw their own custom strategic maps, and each map can have multiple strategic real options paths. This analysis allows analysts to draw and visualize these strategic pathways and does not perform any computations. The examples in Figures 1 and 2 can be easily incorporated into the strategy tree seen in Figure 12.

Real Options Valuation Modeling

Figure 13 illustrates the *Options Valuation* and the *Strategy View*. This part of the analysis performs the calculations of real options valuation models. Analysts must understand the basic concepts of real options before proceeding. This analysis internalizes the more sophisticated Real Options SLS software (see Chapter 13 of Mun's *Modeling Risk* book [Mun, 2015]). Instead of requiring more advanced knowledge of real options analysis and modeling, analysts can simply choose the real option types, and the required inputs will be displayed for entry. Analysts can compute and obtain the real options value quickly and efficiently, as well as run the subsequent tornado, sensitivity, and scenario analyses.

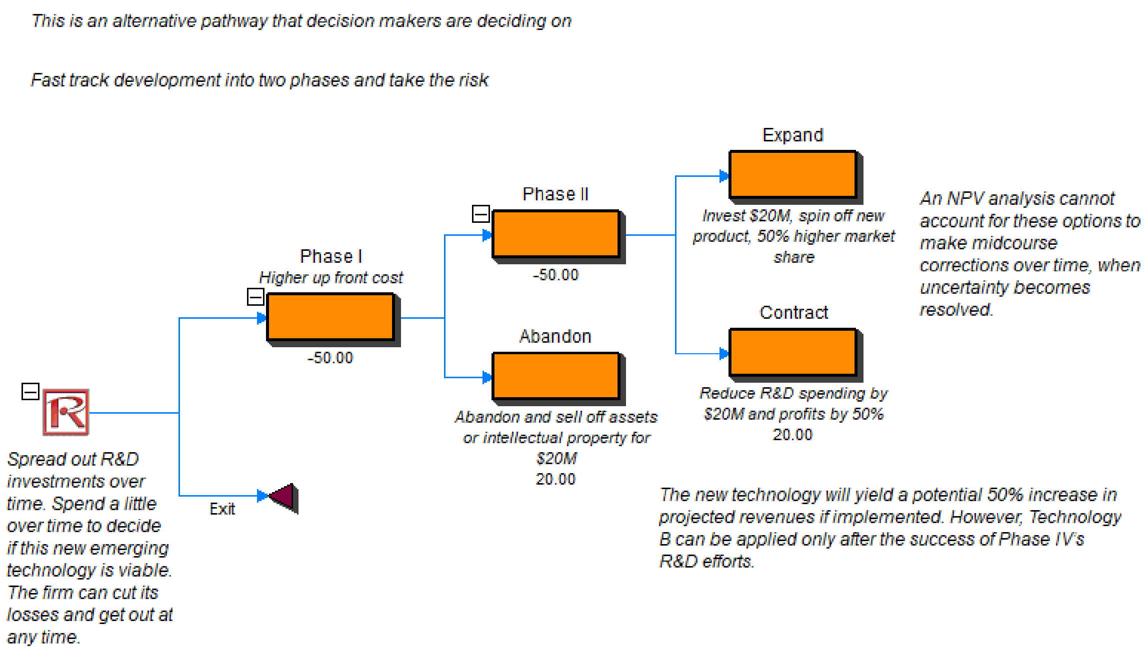


Figure 12. **Framing Flexibility and Options Strategies**

American: Option to Abandon	450,355.44		Result:
Asset Value (Present Value of Net Benefits):	445,625.18		450,355.44
Volatility (Annualized Risk %):	0.22		
Maturity (Total Years to Option Expiration):	5.00		
Risk-Free Rate (Riskless Discount Rate %):	0.04		
Dividend Rate (Opportunity Cost %):	0.00		
Lattice Steps (Typically 100 to 1000):	100.00		
Salvage:	250,000.00		

Figure 13. **Value of Flexibility Options**

The strategic real options analysis is solved by employing various methodologies, including the use of binomial lattices with a market-replicating portfolios approach, and backed up using modified closed-form sequential compound option models. The value of a compound option is based on the value of another option. That is, the underlying variable for the compound option is another option, and the compound option can be either sequential in nature or simultaneous. Solving such a model requires programming capabilities. This subsection is meant as a quick peek into the math underlying a very basic closed-form compound option. It is only a preview of the detailed modeling techniques used in the current analysis and should not be assumed to be the final word. For instance, as suggested in Mun (2016), we first start by solving for the critical value of *I,* an iterative component in the model, using the following equation:

$$X_2 = Ie^{-q(T_2-t_1)}\Phi\left(\frac{\ln(I/X_1)+(r-q+\sigma^2/2)(T_2-t_1)}{\sigma\sqrt{(T_2-t_1)}}\right) - X_1e^{-r(T_2-t_1)}\Phi\left(\frac{\ln(I/X_1)+(r-q-\sigma^2/2)(T_2-t_1)}{\sigma\sqrt{(T_2-t_1)}}\right)$$

Then, solve recursively for the value from the previous equation and input it into the model:

$$Compound\ Option = Se^{-qT_2}\Omega\left[\begin{array}{l}\dfrac{\ln(S/X_1)+(r-q+\sigma^2/2)T_2}{\sigma\sqrt{T_2}};\\ \dfrac{\ln(S/I)+(r-q+\sigma^2/2)t_1}{\sigma\sqrt{t_1}};\sqrt{t_1/T_2}\end{array}\right]$$

$$-X_1e^{-rT_2}\Omega\left[\begin{array}{l}\dfrac{\ln(S/X_1)+(r-q+\sigma^2/2)T_2}{\sigma\sqrt{T_2}}-\sigma\sqrt{T_2};\\ \dfrac{\ln(S/I)+(r-q+\sigma^2/2)t_1}{\sigma\sqrt{t_1}}-\sigma\sqrt{t_1};\sqrt{t_1/T_2}\end{array}\right]$$

$$-X_2e^{-rt_1}\Phi\left[\dfrac{\ln(S/I)+(r-q+\sigma^2/2)t_1}{\sigma\sqrt{t_1}}-\sigma\sqrt{t_1}\right]$$

Additional methods using closed-form solutions, binomial and trinomial lattices, and simulation approaches, as well as dynamic simulated decision trees are used in computing the relevant option values of each strategic pathway as previously indicated.

Step 7: Portfolio Optimization

In today’s competitive global economy, companies are faced with many difficult decisions. These decisions include allocating financial resources, building or expanding facilities, managing inventories, and determining product-mix strategies. Such decisions might involve thousands or millions of potential alternatives. Considering and evaluating each of them would be impractical and maybe even impossible.

A model can provide valuable assistance in incorporating relevant variables when analyzing decisions and in finding the best solutions for making decisions. Models capture the most important features of a problem and present them in a form that is easy to interpret. Models often provide insights that intuition alone cannot. An optimization model has three major elements: decision variables, constraints, and an objective. In short, the optimization methodology finds the best combination or permutation of decision variables (e.g., which products to sell or which projects to execute) in every conceivable way such that the objective is maximized (e.g., revenues and net income) or minimized (e.g., risk and costs) while still satisfying the constraints (e.g., budget and resources).

The projects can be modeled within the ROV software as a portfolio and optimized to determine the best combination of projects for the portfolio in the *Optimization Settings* subtab. Analysts start by selecting the optimization method (Static or Dynamic Optimization). Then they select the decision variable type of *Discrete Binary* (choose which Project or Options to execute with a Go/No-Go Binary 1/0 decision) or *Continuous Budget Allocation* (returns % of budget to allocate to each *option* or *project* as long as the total portfolio is 100%); select the *Objective* (Max NPV, Min Risk, etc.); set up any *Constraints* (e.g., budget restrictions, number of projects restrictions, or create customized restrictions); select the options or projects to optimize/allocate/choose (default selection is *all options*); and when completed, run the Optimization.

Figure 15 illustrates the *Optimization Results*, which returns the results from the portfolio optimization analysis. The main results are provided in the data grid, showing the final *Objective Function* results, final *Optimized Constraints*, and the allocation, selection, or optimization across all individual options or projects within this optimized portfolio. The top portion of the figure shows the textual details and results of the optimization algorithms applied, and the chart illustrates the final objective function. The chart will only show a single point for regular optimizations, whereas it will return an investment efficient frontier curve if the optional *Efficient Frontier* settings are set (min, max, step size). Figures 15 and 16 provide examples of the critical results for decision-makers as they allow flexibility in designing their own portfolio of options. For instance, Figure 15 shows an efficient frontier of portfolios, where each of the points along the curve represents an optimized portfolio subject to a certain set of constraints. In this example, the constraints were the number of options that can be selected in a ship, and the total cost of obtaining these options is subject to a budget constraint. The colored columns on the right in Figure 15 show the various combinations of budget limits and maximum number of options allowed. For instance, if a program office in the Navy allocates only $2.5 million (see the Frontier Variable located on the second row) and no more than four options per ship, then only options 3, 7, 9, and 10 are feasible, and this portfolio combination would generate the biggest bang for the buck while simultaneously satisfying the budgetary and number of options constraints. If the constraints were relaxed to, say, five options and a $3.5 million budget, then option 5 is added to the mix. Finally, at $4.5 million and no more than seven options per ship, options 1 and 2 should be added to the mix. Interestingly, even with a higher budget of $5.5 million, the same portfolio of options is selected. In fact, the Optimized Constraint 2 shows that only $4.1 million is used. Therefore, as a decision-making tool for the budget-setting officials, the maximum budget that should be set for this portfolio of options should be $4.1 million. Similarly, the decision-maker can move backwards, where, say, if the original budget of $4.5 million were slashed by the U.S. Congress to $3.5 million, then the options that should be eliminated are options 1 and 2.

While Figure 15 shows the efficient frontier where the constraints such as number of options allowed and budget were varied to determine the efficient portfolio selection, Figure 16 shows multiple portfolios with different objectives. For instance, the five models shown were to maximize the financial bang for the buck (minimizing cost and maximizing value while simultaneously minimizing risk), maximizing OPNAV value, maximizing KVA value, maximizing Command value, and maximizing a Weighted Average of all objectives. This capability is important because analysts’ objectives and decisions will differ based on different perspectives. Using a multiple criteria optimization approach allows us to see the scoring from all perspectives. Options with the highest count (e.g., 5) would receive the highest priority in the final portfolio, as it satisfies all stakeholders’ perspectives, and would, hence, be considered first, followed by options with counts of 4, 3, 2, and 1.

Conclusions and Recommendations

Strategic real options valuation (ROV) provides the decision-maker the right, but not the obligation, to hold off on executing a certain decision until a later time when uncertainties are resolved and when better information is available. The option implies that flexibility to execute a certain path exists and was predetermined or predesigned in advance. Based on the research performed thus far, we conclude that the methodology has significant merits and is worthy of more detailed follow-on analysis. It is therefore recommended that the ROV methodology be applied on a real case facing the Navy with actual data, and the project’s outcomes tracked over time.

Objective Function	6.1286	6.7465	6.9478	6.9478	6.9478
Frontier Variable	2000000.0000	2500000.0000	3000000.0000	3500000.0000	4000000.0000
Optimized Constraint	1978818.0000	2487042.0000	2718646.0000	2718646.0000	2718646.0000
Option1	1.00	1.00	1.00	1.00	1.00
Option2	0.00	1.00	1.00	1.00	1.00
Option3	1.00	1.00	1.00	1.00	1.00
Option4	1.00	1.00	1.00	1.00	1.00
Option5	1.00	0.00	1.00	1.00	1.00
Option6	0.00	0.00	1.00	1.00	1.00
Option7	0.00	0.00	0.00	0.00	0.00
Option8	1.00	1.00	1.00	1.00	1.00
Option9	0.00	0.00	1.00	1.00	1.00
Option10	0.00	1.00	1.00	1.00	1.00

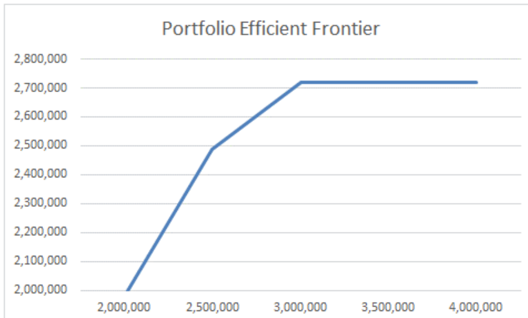


Figure 14. **Portfolio Optimization Results**

Model	Model 1	Model 2	Model 3	Model 4	Model 5	Count
Objective	1,408,735.73	51.16	53.56	48.10	53.56	
Budget Constraint	3,800,000	4,000,000	4,000,000	3,750,000	4,000,000	
Program Constraint	6	7	7	6	7	
Option 1	1.00	1.00	1.00	0.00	1.00	4
Option 2	0.00	0.00	0.00	0.00	0.00	0
Option 3	1.00	1.00	1.00	1.00	1.00	5
Option 4	0.00	1.00	1.00	0.00	1.00	3
Option 5	1.00	1.00	1.00	1.00	1.00	5
Option 6	0.00	1.00	1.00	1.00	1.00	4
Option 7	1.00	0.00	0.00	0.00	0.00	1
Option 8	0.00	1.00	1.00	1.00	1.00	4
Option 9	1.00	0.00	0.00	1.00	0.00	2
Option 10	1.00	1.00	1.00	1.00	1.00	5

Figure 15. **Multi-Criteria Portfolio Optimization Results**

References

Abbott, J. W., Levine, A., & Vasilakos, J. (2008). Modular/open systems to support ship acquisition strategies. In *Proceedings of ASNE Day* (pp. 23–25). Retrieved from http://navalengineers.net/Proceedings/AD08/documents/paper32.pdf

Berkok, U., Penney J., & Kivinen, S. (2013, December). *Factor and organizational substitutions to minimize costs in the Navy*. Ottawa, ON, Canada: Defense R&D Canada, Centre for Operational Research and Analysis.

Cavas, C., & Tran, P. (2016, October 18). France unveils new FTI frigate designed for the French Navy and export. *Defense News.* Retrieved from http://www.defensenews.com/articles/france-unveils-new-fti-frigate-ship-is-designed-for-the-french-navy-and-for-exportFlyvefisken

Doerry, N. H. (2012, August). Institutionalizing modular adaptable ship technologies. *Journal of Ship Production and Design, 30*(3), 126–141. Retrieved from http://www.sname.org/pubs/viewtechnicalpaper?DocumentKey=f20b0085-7010-4f09-873b-546862111d08; also available at http://doerry.org/norbert/papers/20120628MAS-doerry.pdf

Drewry, J. T., & Jons, O. P. (1975, April). Modularity: Maximizing the return on the Navy's investment. *Naval Engineer's Journal, 87*(2), 198–214.

Eckstein, M. (2015, October 15). Navy's future frigate will be optimized for lethality, survivability; will not retain LCS's speed. Retrieved from https://news.usni.org/2015/10/15/navys-future-frigate-will-be-optimized-for-lethality-survivability-will-not-retain-lcss-speed

Flyvefisken Class (SF 300), Denmark. (n.d.) Retrieved from http://www.naval-technology.com/projects/fly/

Freedom Class Littoral Combat Ship Flight 0. (2016, November 16). Retrieved from http://janes.ihs.com.libproxy.nps.edu/FightingShips/Display/1357098

Italian Navy receives fifth FREMM frigate Alpino. (2016, September 30). Retrieved from http://navaltoday.com/2016/09/30/italian-navy-receives-fifth-fremm-frigate-alpino/

Jolliff, J. V. (1974, October). Modular ship design concepts. *Naval Engineers Journal*, 11–30.

Kamerman, J. (2015). *Meeting the future: The German experience* [PowerPoint slides]. Retrieved from https://www.aspi.org.au/__data/assets/pdf_file/0016/26503/Kamerman-The-German-experience-slides.pdf

Koenig, P. C. (2009, April). *Real options in ship and force structure analysis: A research agenda.* Paper presented at American Society of Naval Engineers Day. Retrieved from http://navalengineers.net/Proceedings/AD09/Papers/ASNEoptionspaper.pdf

Koenig, P. C., Czapiewski, P. M., & Hootman, J. C. (2008). Synthesis and analysis of future naval fleets. *Ships and Offshore Structures*, *3*(2), 81–89.

Lundquist, E. (2012, October 22). Absalon class littoral support ships. Retrieved from http://www.defensemedianetwork.com/stories/absalon-class-littoral-support-ships-lcs-on-steroids/

MacKenzie, S. C., & Tuteja, R. (2006, February). *Modular capabilities for the Canadian Navy's single class surface combatant*. Defense R&D Canada.

Mun, J. (2015). *Modeling risk: Applying Monte Carlo simulation, real options analysis, forecasting, and optimization techniques* (3rd ed.). Dexter, MI: Thomson-Shore.

Mun, J. (2016). *Real options analysis: Tools and techniques* (3rd ed.). Dexter, MI: Thomson-Shore.

Mun, J. C., & Housel, T. J. (2016). *Flexible and adaptable ship options: Assessing the future value of incorporating flexible ships design features into new Navy ship concepts*. Monterey, CA: Naval Postgraduate School.

Office of the Director, Operational Test and Evaluation (DOT&E). (2013). *DDG 51 Flight III Destroyer/Air and Missile Defense Radar (AMDR)/Aegis modernization*. Retrieved from http://www.dote.osd.mil/pub/reports/FY2014/pdf/navy/2014ddg51.pdf

Osborn, K. (2015, January 1). Navy changes new LCS name to frigate. Retrieved from http://www.military.com/daily-news/2015/01/15/navy-changes-new-lcs-name-to-fast-frigate.html

Page, J. (2011). *Flexibility in early stage design of U.S. Navy ships: An analysis of options* (Master's thesis). Cambridge, MA: Massachusetts Institute of Technology.

Paris, C. M., Brussels, N. F., & Fiorenza, N. (2013, March 25). Complex tradeoffs between specialized and modular combat ships. Retrieved from http://aviationweek.com/awin/complex-tradeoffs-between-specialized-and-modular-combat-ships

Peruzzi, L., Scott, R., & Pape, A. (2016, October 20). Euronaval 2016: French Navy's new frigate design unveiled. Retrieved from http://www.janes.com/article/64763/euronaval-2016-french-navy-s-new-frigate-design-unveiled

Pike, J. (2011, November 11). Flyvefisken-class STANFLEX 300 ships. Retrieved from http://www.globalsecurity.org/military/world/europe/hdms-flyvefisken.htm

Pike, J. (2016, November 29). Absalon Class Flexible Support/Command Support Ship. Retrieved from http://www.globalsecurity.org/military/world/europe/hdms-absalon.htm

Simmons, J. L. (1975, April). Design for change: The impact of changing threats and missions on system design philosophy. *Naval Engineers Journal*, 120–125.

Stashwick, S. (2016, December 12). Littoral combat ship: The U.S. Navy's "alleged warship." Retrieved from http://thediplomat.com/2016/12/littoral-combat-ship-the-us-navys-alleged-warship/

Sturtevant, G. (2015, January 21). Flexible ships: Affordable relevance over the ship's life cycle. Retrieved from http://www.asne-tw.org/asne/events/presentation/2015-01Sturtevant.pdf

Thorsteinson, J. (2013). Modular warships. *Canadian Naval Review, 8*(4), 29–30. http://www.navalreview.ca/wp-content/uploads/public/vol8num4/vol8num4art7.pdf

Naval Combat System Product Line Architecture Economics

Ray Madachy—is a Professor at the Naval Postgraduate School (NPS) in the Department of Systems Engineering. His research interests include systems and software cost modeling, system affordability and tradespace analysis, total ownership cost, modeling and simulation of systems and software engineering processes, and integrating systems engineering and software engineering disciplines. He wrote the book *Software Process Dynamics* and co-authored *Software Cost Estimation With COCOMO II*, *Software Cost Estimation Metrics Manual for Defense Systems*, and *What Every Engineer Should Know About Modeling and Simulation.* [rjmadach@nps.edu]

John (Mike) Green—is a senior lecturer at the Naval Postgraduate School (NPS), Department of Systems Engineering. Green holds multiple degrees in physics, computer science, international relations, and business. His research interests include sensor fusion, open architecture, human systems integration (HSI), the application of fuzzy logic to combat systems, and layered weapons systems concepts. [jmgreen@nps.edu]

Abstract

Navy combat systems are currently ship class dependent and acquired as stovepipes, yet there are many commonalities among them. This disaggregated nature leads to suboptimal designs and exorbitant costs throughout the system's life cycle. Product line approaches may reduce acquisition costs, increase mission effectiveness, enable more rapid deployment, and provide other benefits across the DoD.

A method for economic tradeoff analysis of system product lines is presented as a model-based systems engineering (MBSE) approach that integrates parametric cost modeling with architecture modeling. The modeling framework includes both a reference architecture and cost model for a general combat system product line.

The economic value of investing in product line flexibility is assessed with the System Constructive Product Line Investment Model (COPLIMO). Empirical DoD cost data is allocated to system functions in the architecture models to calibrate the cost model and populate it for specific system configurations. It is then used to assess the costs and benefits of product line architecting versus traditional one-off designs.

Results of case studies to-date indicate a strong ROI when using a product line approach. Further case studies are ongoing, and the framework will be generalized for other DoD domains to assess product line practices and economics.

Introduction

This ongoing research is assessing economic consequences of product line architecture approaches and refining a framework for others to use similarly. It is being conducted in the Department of Systems Engineering at the Naval Postgraduate School with student involvement.

The technical approach employs parametric cost modeling, empirical data collection of DoD programs for model calibration, application of model-based systems engineering methods to product line architectures, and integration of the modeling methods. The product line options are assessed with economic measures of return-on-investment.

A primary contribution is the integration of parametric cost modeling within MBSE for economic tradeoff analysis of system product lines. Product line costs and benefits are assessed across all life cycle phases to address total ownership cost (TOC).

The research problem being addressed is how to best architect Naval combat systems to be most economical while meeting mission needs. The research is relevant to public procurement policy and management in terms of how combat systems and associated acquisition processes can improve by focusing on product line efficiencies. The goals of improving acquisition processes, increasing combat system mission effectiveness, meeting cost and schedule budgets drive the research questions. The answers to the questions will inform whether the goals are achieved. The questions can be answered by quantitative indicators provided by the cost models and empirical data. The following elements facilitate better-informed acquisition decisions.

Goals:

- Improve combat system acquisition processes
- Meet cost budgets
- Provide rapid capability within schedule constraints
- Improve cost and schedule prediction of system product lines

Questions:

- What are the economic returns of combat system product line architectures versus one-off system designs?
- What is the optimal design approach for product line system development for naval combat systems?
- What system modeling concepts can be implemented for product line architectures that support analyses of both mission effectiveness and cost?
- What are relevant cost factors for product line development?
- Can the results be generalized and/or models used for other Naval and DoD domains?
- What are the limitations and refinements needed to apply the models across domains?

Metrics:

- Product line architecture return-on-investment
- System development and change costs
- Architectural variance points

To address the above goals, combat systems architectures are being formally modeled to identify common functions and variations for different case studies. Empirical cost data from Naval weapons systems programs collected from DoD databases are then allocated to the same system functions in the architecture models. The data is being used to calibrate the parametric product-line investment model and populate it for specific system configurations. It can then be used more generally to assess costs and benefits of product line architecting approaches versus traditional one-off designs for specific systems and their constituent elements.

When TOC is considered for development and maintenance, product lines can have a considerably larger payoff, as there is a smaller base to undergo corrective, adaptive, and perfective maintenance. The value of investing in product-line flexibility using return-on-investment (ROI) and TOC is assessed with System COPLIMO.

We are first assessing the economics of Navy combat system product line architecture approaches with domain case studies and associated economic analyses. The case studies and analyses are at a system-level, sub-system or component level. Systems and all their constituent elements including software, hardware, facilities, or personnel are modeled.

An overall economic business case analysis for product line practices in DoD acquisition will be performed as a synthesis of the case studies covering combat system elements including hardware and software at various system levels. Insights gained from the cost model will provide for more informed acquisition decision-making, and recommendations will be discussed.

Cross Domain Applicability

The method for coupling cost modeling and architectural modeling has wide application across DoD domains. The concept and execution of product line architectures extends across all system application domains where related systems share features. Similarly, many DoD domains and industries can benefit with the capability to analyze the economic consequences of their product line architecture options. It is valid for all the services, the intelligence community, other government operations, and commercial industry across numerous domains (though some already leverage product line architectures).

The systems engineering modeling methods for product architecture and cost modeling are transferable in several ways. The modeled generic system architecture containing the detect, control, engage paradigm as a central premise of combat systems is the same across many DoD application domains beyond the Navy. The architecture model can thus be used as a template for many DoD system product lines. The general method can also be used for different non-combat system types with relevant architecture models.

The modeling framework includes a reference architecture and cost model for a general combat system product line that is extensible to other DoD and government domains. A cross-domain analysis is first being performed within the Navy and the lessons extrapolated. Tools and guidance will be provided for others to adapt and use the framework for investment analysis of product line architecting in different environments.

Background and Previous Work

Product line investment returns accrue from reusing common pieces in different systems/products that share features. Furthermore, systems can be fielded faster leading to increased overall mission effectiveness. Flexibility is enhanced increasing the option space. These benefits occur because previously built components reduce the effort and enable more rapid development. Employing a product line engineering approach to future combat system design is beneficial for all stakeholders.

There are other significant product line benefits besides life cycle cost savings, such as rapid development time and adaptability to mission changes. Cost models provide an easy-to-use framework for performing these broader “ility” and affordability analyses.

The models also demonstrate that not all attempts at product line reuse will generate large savings. A good deal of domain engineering needs to be done well to identify product line portions of the most likely to be product specific, fully reusable, or reusable with adaptation. Product line architecting needs to be done well to effectively encapsulate the sources of product line variation. Cost models help evaluate the tradeoffs of different architectural options and determine when product line approaches are justified.

System Architecting for Change

Composable systems allow for selecting and assembling components in different ways to meet user requirements. In order for a system to be composable, its components must also be reusable, interoperable, extensible, and modular. A reusable artifact is one that provides a capability that can be used in multiple contexts. Reuse is not confined to a software or hardware component but any life cycle artifact.

Efficient product line architecting requires modularization of the system's architecture around its most frequent sources of change (Parnas, 1979) as a key principle for affordability. When changes are needed, their side effects are contained in a single systems element, rather than rippling across the entire system. For modularization, it is desirable to identify the commonalities and variability across the families of products or products and develop architectures for creating and evolving the common elements with plug-compatible interfaces to insert the variable elements.

The methods of MBSE have been demonstrated for implementing these product line best practices. Our integrated method extracts cost elements from the architecture models.

Parametric Cost Modeling for Product Line Economics

Product line models for TOC provide strong capabilities for analyzing economic consequences of alternative system acquisition approaches. They show that if total life cycle costs are considered for development and maintenance, product lines can have a considerably larger payoff, as there is a smaller base to undergo corrective, adaptive, and perfective maintenance.

The initial basic version of COPLIMO was designed to assess the costs, savings, and return-on-investment associated with developing and reusing software product line assets across families of similar applications (Hall, 2018). Several extended parametric models adapted from COPLIMO have been employed since then.

Most software product line cost estimation models are calibrated only to local product line data rather than to a broad range of product lines. They also underestimate the return-on-investment for product lines by focusing only on development versus life cycle savings, and by applying writing-for-reuse surcharges to the entire product rather than to the portions of the product being reused.

COPLIMO addresses these shortfalls and consists of two components: a product line development cost model and an annualized post-development life cycle extension. It models the portions of software that involve product-specific, newly-built software; fully reused black-box product line components; and product line components that are reused with adaptation. It is an extension built upon the well-calibrated and most widely used software cost model Constructive Cost Model (COCOMO) II, tailored for strategic software product line decision issues with available supporting industry data (Boehm et al., 2000).

Product line investment models must address two sources of cost investment or savings:

- The relative cost of developing for product lines: The added effort of developing flexible product line architectures to be most cost-effectively reused across a product line family of applications, relative to the cost of developing a single system.
- The relative cost of reuse: The cost of reusing system architecture in a new product line family application relative to developing new systems.

The original COPLIMO was developed as a detailed model for software product lines and was also extended for software quality. The software model was later modified for systems-level product lines on the DoD System Engineering Research Center's (SERC's) *Valuing Flexibility* research project (SERC, 2012; Boehm, Lane, & Madachy, 2011). It was demonstrated for representative DoD system types using empirical system maintenance data.

The System COPLIMO framework is a model extension at the systems level, used to assess flexibility and ROI tradeoffs (SERC, 2012; Boehm, Lane, & Madachy, 2011). The same concepts and phenomena of software product lines also apply at the system level. It models up-front investment in creating reusable system architectures for product lines composed of software and hardware. It performs a TOC analysis for a family of systems. The TOC covers the full system lifespan of and normalized to net present value at specified interest rates. Figure 1 shows the model inputs and outputs as a black box.

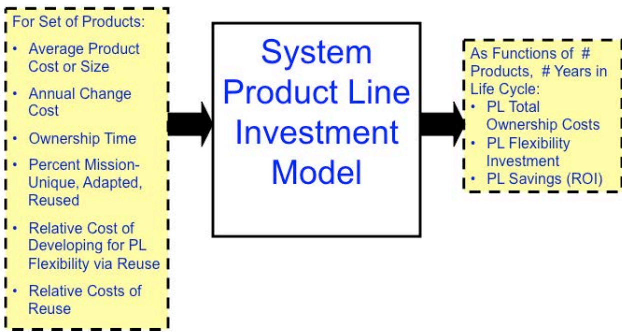


Figure 1. **System COPLIMO Inputs and Outputs**

The general model was enhanced to handle specific DoD application domains with Monte Carlo simulation capabilities. We incorporated the life cycle cost ratios for Operations and Support (O&S) for hardware and software system types derived from Redman, Crepea, and Stratton (2008) and Koskinen (2010). Choosing system type impacts the general model inputs for Ownership Time and Annual Change Cost based on the O&S cost ratios. The user chooses a system type and ownership time, which invokes a calculated annual change costs for the relevant domain.

The software product line model was then enhanced and adopted for NAVAIR avionics software. The product line research at NAVAIR involved cost and ROI modeling of avionics software development on the Future Airborne Capability Environment (FACE). COPLIMO helped validate product line costing efforts across different airborne platforms.

Subsequently we devised an integrated method for representing architectural variants to enumerate as parametric inputs for the System COPLIMO cost model described next.

Method

The technical approach integrates parametric cost modeling with MBSE product modeling methods to enable economic tradeoff analysis of system product lines. Product line architectures of common system designs for future Navy combat systems are modeled including hardware and software architectural options. A functional decomposition of current Navy combat system suites provides the framework for product lines incorporating the commonalities needed for effective combat capabilities regardless of platform or ship class.

Navy combat systems have a variety of configurations that include sensors, weapons, and hardware/software integrations to accomplish similar goals. These common elements and their interfaces are modeled as flexible product lines. Our method assumes each system utilizes the generic detect, control, engage paradigm as the central premise of the combat system architecture, both functional and physical. This is our modeling starting point.

The modeling sequence below is used for a given system product line and undertaken in the case studies:

1. Describe a general domain model of the given system with common elements.
2. Develop a reference product architecture with variation points.
3. Map existing systems to the reference architecture.
4. Collect empirical costs and map them to system elements from above. Develop new cost models for each application, as necessary.
5. Tailor the COPLIMO framework model for the reference architecture.
6. Assess product line economics for the given system.

Product Line Architecture Modeling

The system architecture modeling uses the Hatley-Pirbhai structured methodology and an associated architecture template. See Figure 2 for the Hatley-Pirbhai architecture template that is instantiated for each system.

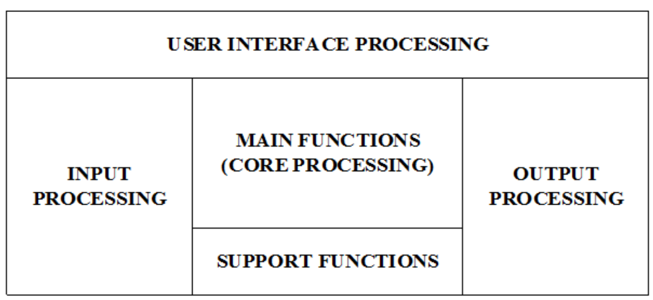


Figure 2. **Hatley-Pirbhai Architecture Template**

An Enhanced Data Flow Diagram (EDFD) in Figure 3 and related Architectural Flow Diagram (AFD) in Figure 4 describe the functional and physical behavior of the combat system. Each system architecture diagram utilizes the detect, control, engage paradigm as the central premise of the combat system architecture, both functional and physical, in the EDFDs and AFDs.

The AFD provides a structure for variation point identification necessary for orthogonal variability modeling (OVM) in a product line construct. Variations points are identified for sensors, HSI/consoles, weapons, and data links with alternative choices for a combat system product line.

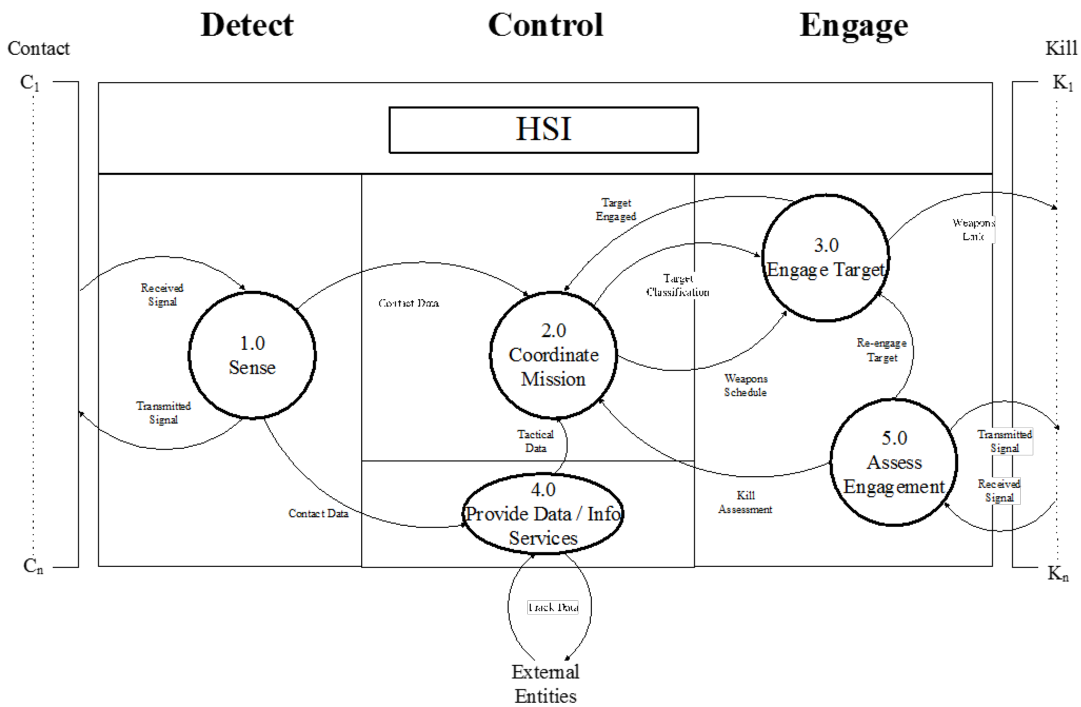


Figure 3. **Enhanced Data Flow Diagram (EDFD)**

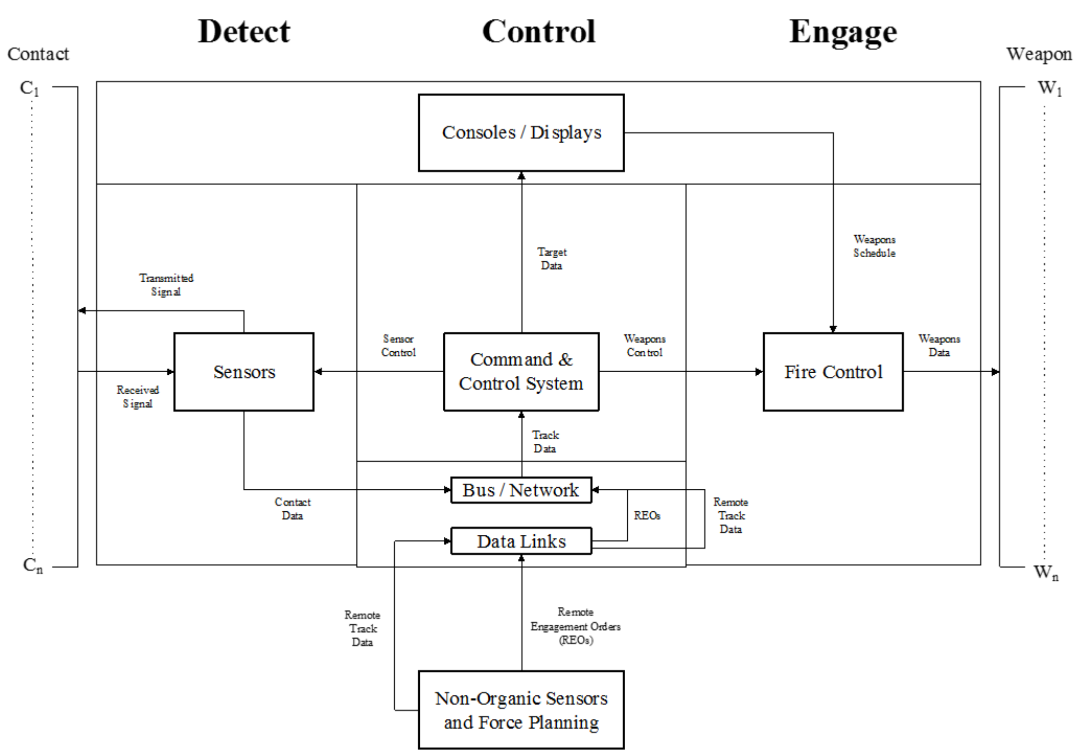


Figure 4. **Architectural Flow Diagram (AFD)**

The AFD provides the structure for variation point identification necessary for orthogonal variability modeling (OVM) in the product line construct. Variations points are identified for sensors, HSI/consoles, weapons, and data links.

The variation points and associated variants are presented as OVMs, showing alternative choices for each variation point. The variation point OVMs are consolidated into a product line OVM with packaged variants and constraint dependencies. The constraint dependencies demonstrate feasible combinations of packaged variants, variation points, and variants for the combat system product line. The notation for an OVM is shown in Figure 5. See the case study section for an applied OVM example.

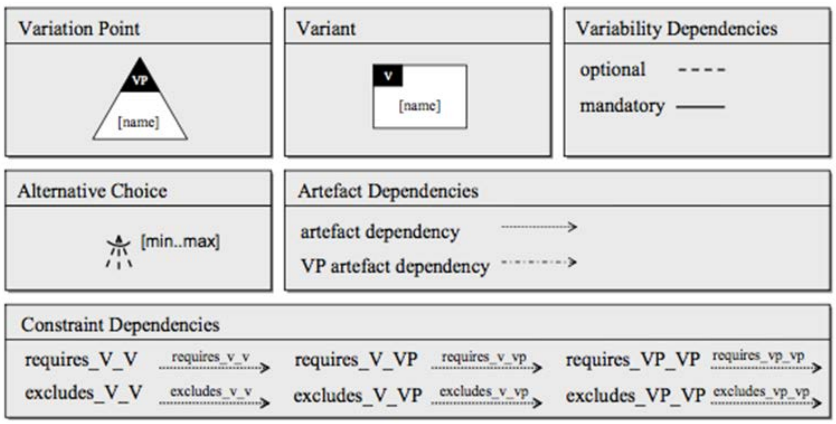


Figure 5. **OVM: Halmans and Pohl Notation**

An OVM uses graphic notation (Halmans and Pohl notation) to display the variability within a product line. The two classes within the OVM are the variation point and variant. Variability dependencies show the association between the variation point and variant classes.

Variation points and variants must follow the following associative conditions:

1. Each variation point must be associated with at least one variant.
2. Each variant must be associated with at least one variation point.
3. A variation point can offer more than one variant.
4. A variant can be associated with different variation points.

DoD Empirical Cost Data Collection

To collect relevant data on systems development costs, the Defense Acquisition Management Information Retrieval (DAMIR) repository has been a primary source. All the weapons cost data required for three tiers of a cruise missile defense system in Hall (2018) were obtained in President's Budget Submission reports (DoD, 2016) and DOD selected acquisition reports (DoD, 2015) for chosen programs. The DOD Selected Acquisition Reports also provide data on the system ownership times.

Data required for the investment model on inflation rates come from the Bureau of the Fiscal Service, U.S. Department of the Treasury. The Navy Visibility and Management of Operating and Support Costs (VAMOSC) management information system has also been used by students to obtain actual costs. It has data for different levels of system elements useful for the product line variation modeling and WBS cost mapping.

Software development cost data is analyzed from the DoD Cost Assessment Data Enterprise (CADE) Software Resources Data Report (SRDR) records (DoD, 2011). This repository provides actual software development costs that can be tied to contractor product line components and practices. Additionally, it is a rich database containing essential data on software reuse and modification parameters that can be directly used to set defaults and tailor the COPLIMO model. The relative costs of reuse, adapted and developing for product line flexibility can be inferred for given programs and application domains (Clark & Madachy, 2015). Software maintenance SRDRs can provide insight into annual system change costs and percentages.

Tiered Combat System Case Study

The concept for the integrated method of representing architectural variants to enumerate as parametric inputs for the System COPLIMO cost model was first proven in a student master's thesis. In Hall (2018), it was applied to successive tiers of a cruise missile **combat system product line** using rigorously collected actual system costs. The tiers were modeled as product line architectures suitable for further system development activities and automatic cost estimation.

The modeling sequence undertaken for the case study is detailed in Figure 6 and as follows:

1. Conduct an architectural analysis of current combat systems (scoped to surface combatant applications).
2. Determine necessary architectural functions and commonalities.
3. Model a case study 3 Tier Product Line with increasing capability in each tier while still utilizing architectural component commonalities.
4. Use identified commonalities to determine percentage of unique, reused, and adapted components.
5. Apply percentages to System COPLIMO to determine return on investment of a product line approach.

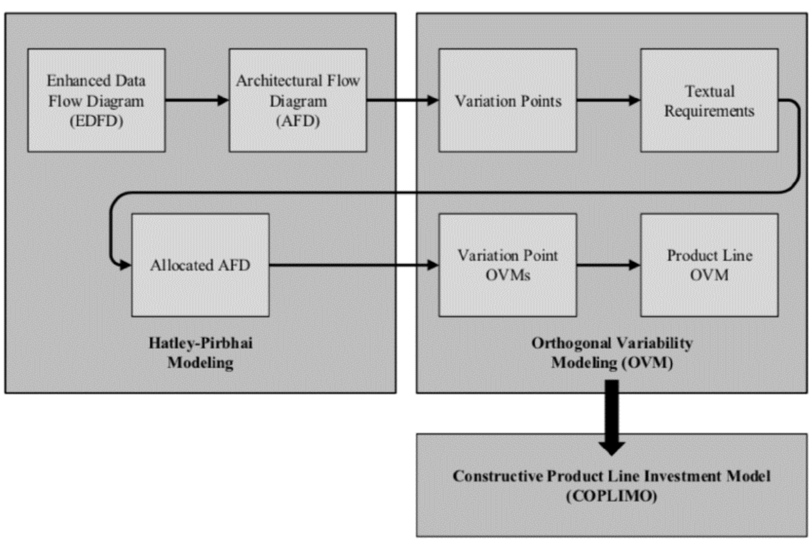


Figure 6. **Modeling Sequence for Tiered Combat System Product Line Analysis**

The System COPLIMO tool used in Hall (2018) was an adaption of the system-level product line flexibility tool described in Boehm et al. (2000). The pre-sets for domain-specific defaults were replaced with provisions for actual system costs and maintenance parameters. This was done by accessing and consolidating empirical weapons cost data from DoD repositories to populate the model.

First tier includes a surface warfare (SUW) capability designed for a small surface combatant. The second tier is designed around a cruise missile defense capability that could be employed on a future frigate (FFGX), amphibious assault ship, and aircraft carrier (CVN) platforms. The third tier includes theater ballistic missile defense (TBMD) and cruise missile defense capabilities, designed to facilitate the needs of a future guided missile destroyer (DDGX) and guided missile cruiser (CGX). See Figure 7.

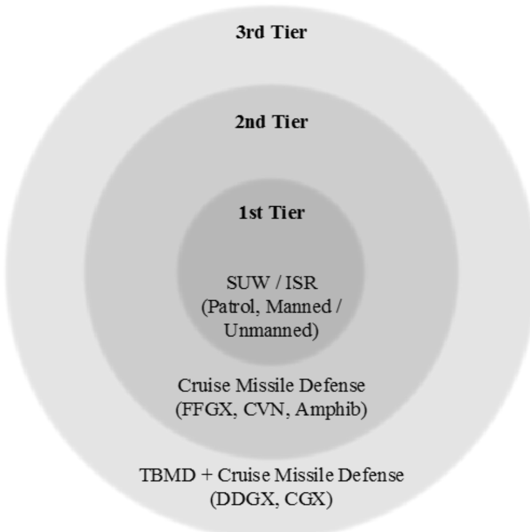


Figure 7. **Combat System Product Line Tiers**

The combat system functional and physical architectures provided the construct for identifying variability subjects within the combat system. For orthogonal variability modeling after analyzing the functional and physical constructs of the EDFD and AFD, four variation points were identified for further decomposition and component allocation:

1. Sensors
2. HSI/Console
3. Weapons
4. Data Links

Each variant textual requirement is associated a variation point. Textual requirements do not specify what the variant is. Textual requirements were generated for all variation points based on review of current combat system mission capabilities. An example is shown in Figure 8.

Variation Point	The sensors shall have the ability to...
Variant	...conduct volume air search and tracking...
Variant	...and conduct surface search and tracking...
Variant	...and search / track in the electro-optical (EO) / infrared (IR) spectrum...
Variant	...and provide high resolution imagery for identification and targeting...
Variant	...and query manned / unmanned aerial systems...
Variant	...and provide passive electromagnetic (EM) wave detection.

Figure 8. **Example Textual Requirements for Sensors Variation Point**

Physical components identified from textual requirements were then assigned to the AFD. Components are variants which will be used for orthogonal variability modeling. These components are general, for example, without specifying specific types of sensors. Figure 9 shows the allocated AFD.

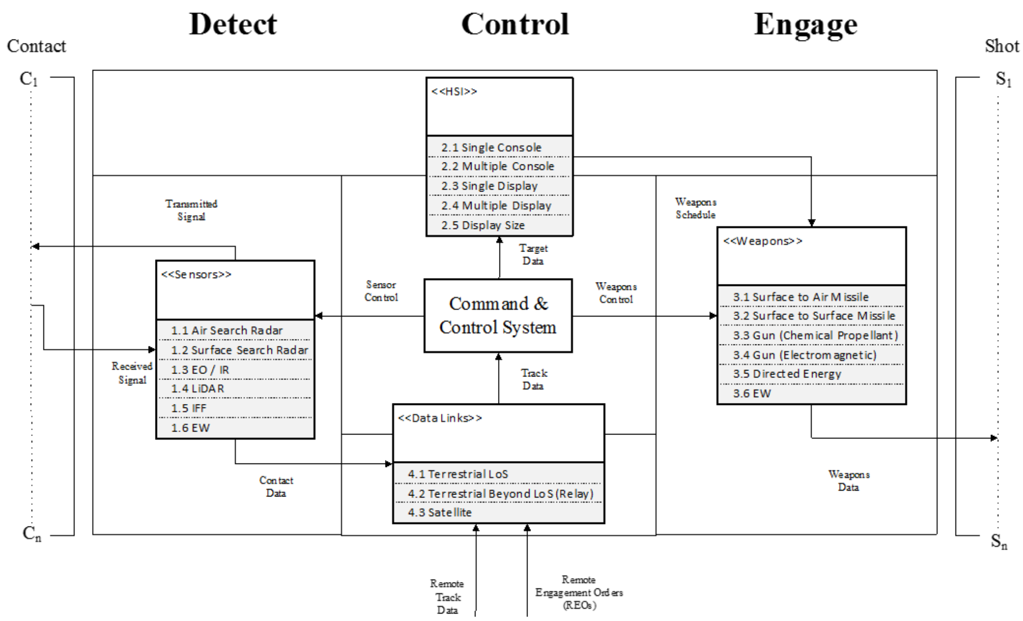


Figure 9. **Allocated Architectural Flow Diagram**

OVMs were then generated for the variation points. See Figure 10 for the sensors OVM. The product line OVM in Figure 11 shows constraint dependencies between variation points and variants at a product-line level. The packaged variants require or exclude different variants depending on the capabilities of the combat system tier. These variant requirements and exclusions parallel the detect, control, engage paradigm.

The Product Line OVM helps identify reused, adapted, and mission unique components within the product line, necessary for COPLIMO. The OVM used to quantify variation points for COPLIMO product line percentage inputs for the tiers is in Figure 11.

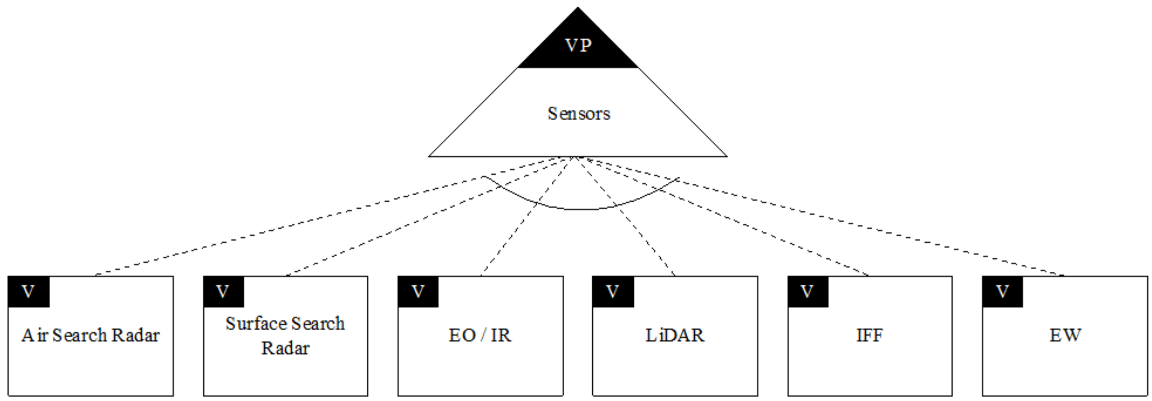


Figure 10. **Sensor Variation Point Orthogonal Variability Model**

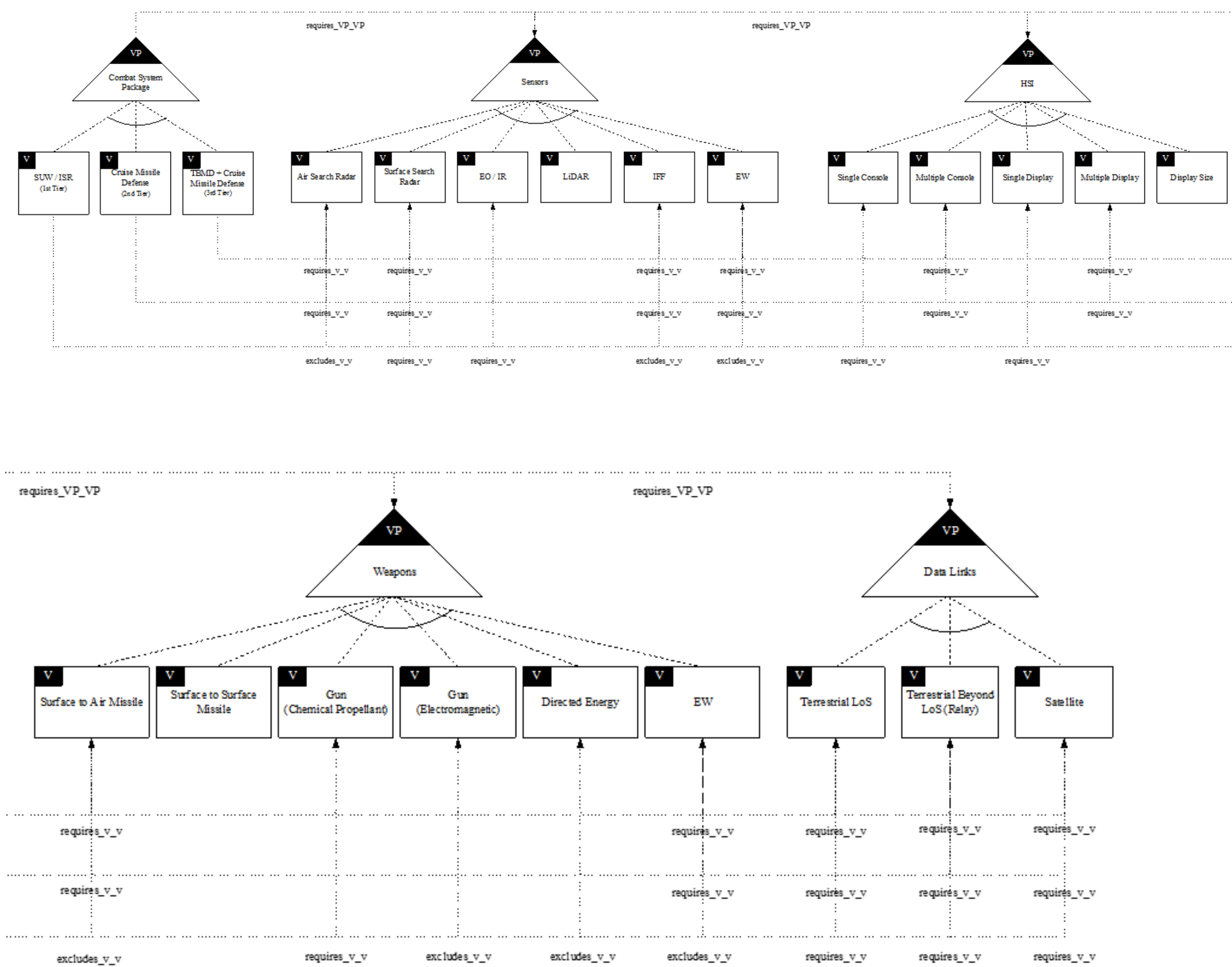


Figure 11. **Combat System Product Line Orthogonal Variability Model (Portion)**

The product line orthogonal variability model describes the three tiers of combat systems that are proposed for the product line. This OVM introduces the concept of packaged variants to reduce complexity of the model when representing each of the tiers. The variation point of “Combat System Package” includes three variants: SUW (1st tier), cruise missile defense (2nd tier), and TBMD + cruise missile defense (3rd tier). These variants are all optional, packaged variants that can be chosen based on the customer’s needs. Such variation points are shown textually in Figure 12.

Variation Point	The console / HSI shall be equipped with...
Variant	...either single...
Variant	...or multiple consoles...
Variant	...and single...
Variant	...or multiple displays...
Variant	...and allow for various display sizes.

Variation Point	The weapons shall have the ability to...
Variant	...target and engage air targets at long range...
Variant	...and target and engage surface targets at long range...
Variant	...and target and engage air / surface targets a short range...
Variant	...and provide long range naval surface fire support...
Variant	...and provide supportability for future weapons technology...
Variant	...and provide offensive capability in the EM spectrum.

Variation Point	The data links shall have the ability to...
Variant	...transfer data with assets within line of sight (LoS)...
Variant	...and transfer data with assets beyond LoS...
Variant	...and transfer data via satellite...

Figure 12. **Variation Points**

The product line components are enumerated in Figure 13. They are classified as adapted, reused, or mission-unique to specify for COPLIMO. The COPLIMO model inputs and their rationales are shown in Figure 14. These inputs model the Tier 3 Capability for Theater Ballistic Missile Defense and Cruise Missile Defense Capable.

Variation Point: Sensors		
Product Line Classification	**Variant**	**Justification**
Adapted	Air Search Radar	Power, beam forming, and search / track functions different for 2nd and 3rd tier packaged variants.
Adapted	EW	Power and physical size requirements may be different for 2nd and 3rd tier packaged variants.
Reused	Surface Search Radar	Physical size and capabilities of sensor can be used for 1st, 2nd, and 3rd tier packaged variants.
Reused	EO / IR Sensor	See Surface Search Radar justification.
Reused	LiDAR	See Surface Search Radar justification.
Reused	IFF	Hardware and interfaces are the same for 2nd and 3rd tier packaged variants.
Variation Point: HSI		
Product Line Classification	**Variant**	**Justification**
Reused	Single Console	Consoles common across 1st, 2nd, and 3rd tier packaged variants.
Reused	Multiple Console	See Single Console justification.
Reused	Single Display	Displays common across 1st, 2nd, and 3rd tier packaged variants.
Reused	Multiple Display	See Single Display justification.
Adapted	Display Size	Displays are common but size can be specified by customer.
Variation Point: Data Links		
Product Line Classification	**Variant**	**Justification**
Reused	Terrestrial LoS	Data links standardized across US and NATO platforms, therefore they will also be common across 1st, 2nd, and 3rd tier packaged variants.
Reused	Terrestrial Beyond LoS	See Terrestrial LoS justification.
Reused	Satellite	See Terrestrial LoS justification.
Variation Point: Weapons		
Product Line Classification	**Variant**	**Justification**
Mission Unique	Surface to Air Missile	Ranges and kill mechanisms are different for 2nd and 3rd tiers.
Mission Unique	Surface to Surface Missile	Ranges and size of missile different for 1st, 2nd and 3rd tiers based on mission and ship size.
Mission Unique	Gun Electro-Magnetic	Power and size constraints dependent on ship size and cost for 2nd and 3rd tiers.

Figure 13. **Product Line Components**

System COPLIMO Input Summary (3rd Tier Packaged Variant)		
Input	**Value**	**Rationale**
System Costs		
Average Product Development Cost	$322M	Department of Defense Fiscal Year (FY) 2017 President's Budget Submission 2016, 127-138
Annual Change Cost	10 %	Estimate
Ownership Time	40 years	DoD Selected Acquisition Report 2015, 48
Interest Rate	2.625 %	Bureau of the Fiscal Service, U.S. Department of the Treasury 2018
Product Line Percentages		
Mission Unique	20 %	From system architecture analysis
Adapted	25 %	From system architecture analysis
Reused	55 %	From system architecture analysis
Relative Cost of Reuse		
Relative Cost of Reuse for Adapted	40 %	COPLIMO default
Relative Cost of Reuse for Reused	5 %	COPLIMO default
Investment Cost		
Relative Cost of Developing for PL Flexibility via Reuse	1.7	COPLIMO default

Figure 14. **Model Input for Tier 3 Combat System Product Line**

An example product line investment analysis for the tiered product line using System COPLIMO is shown in Figure 15. Inputs were based on rigorous data collection for cruise missile programs from the DoD databases.

The return on investment (ROI) output provides a metric for determining the cost benefit of a product line engineering approach. ROI is defined as the net effort savings (PL Effort Savings), divided by the product line (PL) flexibility investment. The results suggest a very strong ROI as the number of cruise missile in the product line increases. For simplification in this case, each successive product was modeled with the same change percentage parameters. With these assumptions, the results indicate an ROI greater than 20 after the seventh built system.

System Costs

Average Product Development Cost (Burdened $M) 322 — Ownership Time (Years) 30

Annual Change Cost (% of Development Cost) 10 — Interest Rate (Annual %) 2.6

Product Line Percentages

Unique % 20

Adapted % 25

Reused % 55

Relative Costs of Reuse (%)

Relative Cost of Reuse for Adapted 40

Relative Cost of Reuse for Reused 5

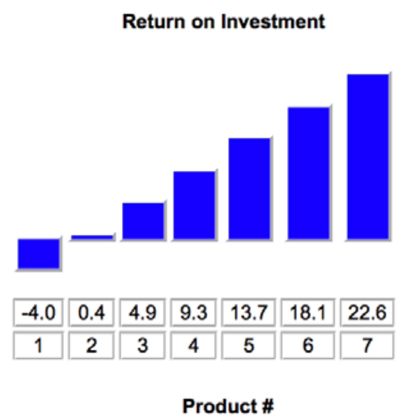
Investment Cost

Relative Cost of Developing for PL Flexibility via Reuse 1.7

Calculate — Monte Carlo Off

Results

# of Products	1	2	3	4	5	6	7
Development Cost ($M)	$457.2	$172.3	$172.3	$172.3	$172.3	$172.3	$172.3
Ownership Cost ($M)	$1,371.7	$516.8	$516.8	$516.8	$516.8	$516.8	$516.8
Cum. PL Cost ($M)	$1,829.0	$2,518.0	$3,207.1	$3,896.2	$4,585.3	$5,274.4	$5,963.4
PL Flexibility Investment ($M)	$135.2	$0	$0	$0	$0	$0	$0
PL Effort Savings	($541.0)	$58.0	$656.9	$1,255.8	$1,854.7	$2,453.6	$3,052.6
Return on Investment	-4.00	0.43	4.86	9.29	13.71	18.14	22.57



Created by Ray Madachy at the Naval Postgraduate School. For more information contact him at rjmadach@nps.edu

Figure 15. **System COPLIMO Results for Tier 3 Cruise Missile Defense Product Line Investment**

Current Case Studies

Coordinated case studies are currently being performed by student capstone teams and on individual theses. The research is divided into a set of sub-problems driven by the level of student involvement for each thesis or group capstone project. They cover different combat systems at varying levels within the system architectures.

The current case studies in-process involve the following:

- Aegis ship class software product line economics
- Ship bridge system product line architecting
- ASW product lines for air, surface, and subsurface applications

A capstone based in Newport, RI is addressing cross-domain applicability. They are investigating the product line potential for ASW systems to include air, surface, and subsurface applications (SH-60, Trident, Virginia, SQQ-89). Currently they are developing the reference architectures for the ASW systems to capture the variability for each of the platform applications for the cost model.

The ship bridge product line case study has extensively researched surface ship control to investigate the cause of the collisions involving the McCain and the Fitzgerald. An overarching process common to all ships and a notional reference architecture for a common ship control for all ships is being developed.

For the Aegis software product lines, substantial data has been collected from the contractor and government program office. Preliminary results indicate substantial savings which are being analyzed and documented. SRDR data is also being sought for more thorough and crosschecking analysis of software size and cost. A revision of COPLIMO will be done for the case study specifics.

We will synthesize the results of the case studies covering different system elements including hardware, software, etc. at various system levels. Specific product line practices and economics are expected to vary by subsystem-type.

Conclusions and Future Work

Results of the case studies to-date indicate a strong ROI when using a product line approach for Naval combat systems. We have found that high-level system architecture design for future U.S. Navy combat systems should focus on the product line, instead of platform specific combat systems. They should plan for the reuse of system components over time.

Applying the engineering product line methodology to combat system architecture design and development needs to happen at the earliest stage of design. System COPLIMO provides a trade space for determining initial investment and future return on investment (ROI) with respect to product line systems versus non-product line systems. Integrated modeling as this should be done to support early architectural decisions.

Further case studies are ongoing, and the framework will be generalized for other DoD domains to assess product line practices and economics. Future work includes developing engineering product line models for additional warfare areas such as anti-submarine warfare (ASW), electronic warfare (EW), cyber warfare, and others. Functional and physical architectural hierarchy can also be further decomposed into third and fourth levels to provide greater level of detail at the subsystem level.

Thus far our product models have been static. However, even greater insight is possible with dynamics models. For example, we can test executable EDFDs and AFDs in simulation software, following the detect, control, engage paradigm for different mission scenarios.

We will collect more empirical data to further validate COPLIMO at a system level, instead of using software engineering default calibrations. To further improve cost estimation fidelity, we will account for individual component complexities in the effort model. We will also model with product-specific inputs for individual products in a line versus homogeneity of change percentages.

References

Boehm, B., Abts, C., Brown, A. W., Chulani, S., Clark, B., Horowitz, E., … Steece, B. (2000). *Software cost estimation with COCOMO II*. Upper Saddle River, NJ: Prentice Hall.

Boehm, B., Brown, A. W., Madachy, R., & Yang, Y. (2004). A software product line life cycle cost estimation model. In *Proceedings of the 2004 International Symposium on Empirical Software Engineering*.

Boehm, B., Lane, J., & Madachy, R. (2011, April). Total ownership cost models for valuing system flexibility. Conference on Systems Engineering Research, Los Angeles, CA.

Clark, B., & Madachy, R. (Eds.). (2015). *Software cost estimation metrics manual for defense systems*. Haymarket, VA: Software Metrics Inc.

DoD. (2011, November). *Software resources data report.* Retrieved from http://dcarc.cape.osd.mil/Files/Policy/2011-SRDRFinal.pdf

DoD. (2015). *Selected acquisition report (SAR): DDG 51 Arleigh Burke Class Guided Missile Destroyer (DDG 51).* Retrieved from http://www.esd.whs.mil/Portals/54/Documents/FOID/Reading%20Room/Selected_Acquisition_Reports/15-F-0540_DDG_51_SAR_Dec_2014.PDF

DoD. (2016). *Fiscal Year (FY) 2017 President's Budget submission: Navy justification book, Volume 1 of 1: Shipbuilding and conversion, Navy.* Retrieved from http://www.secnav.navy.mil/fmc/fmb/Documents/17pres/SCN_Book.pdf

Hall, R. (2018, June). *Utilizing a model based systems engineering approach to develop a combat system product line* (Master's thesis). Naval Postgraduate School.

Koskinen, J. (2010). *Software maintenance costs.* Jyväskylä, Finland: University of Jyväskylä. Retrieved from http://users.jyu.fi/~koskinen/smcosts.htm

Parnas, D. L. (1979). Designing software for ease of extension and contraction. *IEEE Transactions on Software Engineering, 5*(2), 128–138.

Redman, Q., Crepea, A. T., & Stratton, G. (2008). *Weapon design trade off using life cycle costs*. Retrieved from http://www.galorath.com/images/uploads/3_-_Quentin_Redman_-_Design_trades_using_Life_Cycle_Costs_short_version.pdf

Systems Engineering Research Center (SERC). (2012). *Valuing flexibility: Phase II* (SERC-2012-TR-010-2). Retrieved from https://web.sercuarc.org/documents/technical_reports/1525448307-SERC-2012-TR-010-2-Valuing-Flexibility-Phase-2-RT-18a.pdf

Defining a Model-Based Systems Engineering Approach for Systems Engineering Technical Reviews

Warren Vaneman—has more than 32 years of leadership and systems engineering experience from various positions within the intelligence community, including as Chief Architect of the Enterprise Ground Architecture at the National Reconnaissance Office. He is also a retired Navy Reserve Captain. He has a BS from the State University of New York Maritime College, an MS in systems engineering and a PhD in industrial and systems engineering from Virginia Tech, and a Joint Professional Military Education Phase 1 Certificate from the Naval War College. The International Council on System Engineering (INCOSE) certifies him as a Certified Systems Engineering Professional (CSEP). [wvaneman@nps.edu]

Ron Carlson—served 26 years in naval aviation as a pilot, seven years of which were at NAVAIR where he led NAVAIR Systems Engineers through several years of systems engineering revitalization. He joined the NPS SE department 10 years ago. He has a Master of Philosophy in Systems Engineering from Stevens Institute of Technology, master's degrees in strategic studies and national policy from the Naval War College and business administration–aviation from Embry Riddle Aeronautical University, and his Bachelor of Science in nuclear engineering from the University of Michigan. [rrcarlso@nps.edu]

Abstract

The DoD Digital Engineering Strategy calls for formalized planning, development, integration, management, and use of models to support systems engineering activities and decision-making across the lifecycle. As DoD organizations migrate to a Model-Based Systems Engineering (MBSE) environment, efficiencies will be gained by making the model the focus of engineering development activities throughout the engineering and acquisition lifecycle. Technical reviews will be key benefactors of this environment because model-based reviews allow for complexity to be managed more efficiently, and data, in lieu of "systems engineering products," will be the commodity used to evaluate the technical review criteria. Current technical reviews are based around lengthy reviews of static, contractually obligated documents that are used to demonstrate successful completion of the review criteria. MBSE technical reviews will provide greater insight with faster comprehension for the details across a program's lifecycle. This will not only provide efficiencies for the review, but will also improve the program's cost and schedule efficiency. This paper presents preliminary findings from our ongoing research by discussing the systems engineering activities that are performed during the system acquisition lifecycle and technical reviews from an MBSE perspective. These activities will then be evaluated to see how MBSE will complement technical reviews.

Introduction

> "Advancements in computing, modeling, data management, and analytical capabilities offer great opportunities for the engineering practice. Applying these tools and methods, we are shifting toward a dynamic digital engineering ecosystem. This digital engineering transformation is necessary to meet new threats, maintain overmatch, and leverage technology advancements."
>
> —Kristin Baldwin, Acting Deputy Assistant Secretary of Defense for Systems Engineering (DASD[SE], 2018)

Model-based processes are one of the most widely-discussed issues within the Department of Defense (DoD) today. The DoD Digital Engineering Strategy (2018) provides a vision on how the DoD will modernize, develop, deliver, operate, and sustain systems.

This strategy is important because advances in technology have led to larger and more complex systems. This implies a need for a clear, concise way to express the system design (clear, logically consistent semantics), and a need to represent systems differently to account for emergent behavior within the system due to the increased complexity.

The Digital Engineering Strategy provides five goals (DASD[SE], 2018).[1] This paper is the first step in defining a Model-Based Systems Engineering[2] (MBSE) approach for Naval Systems Engineering Technical Reviews. While our research will likely address each of the five goals, the most significant goal for this paper is as follows (DASD[SE], 2018):

Goal 1: Formalize the Development, Integration, and Use of Models to Inform Enterprise and Program Decision-making.

1.1 Formalize the planning for models to support engineering activities and decision-making across the lifecycle.
1.2 Formally develop, integrate, and curate models.
1.3 Use models to support engineering activities and decision-making across the lifecycle.

There is a strong need to ensure that the systems engineers and stakeholders understand the different model types and what information can be gleaned from them. When developed properly, models can provide a precise virtual representation of the functional, physical, parametric, and program entities of the systems. Increased emphasis is on the model itself, specifically the objects and relationships it contains, rather than the diagram to encourage better model development, usage, and decision-making. To enable this, new policies must be established to defined model-based processes, and governance of the authoritative source of truth—often known as the single source of technical truth.

Our ongoing research is defining how DoD organizations can conduct milestone reviews in a MBSE-environment. This effort requires an examination of current technical review processes; a derivation of new MBSE processes that will provide the requisite system and programmatic information to satisfy the review criteria; and a demonstrated model-based technical review environment. This paper takes the first step. The next section discusses the essence of MBSE. Then we provide a framework that establishes the relationships between key elements that are used for system definition and development, and establishes the framework from which technical reviews in a MBSE environment can be addressed. The next section provides a background of technical reviews. The last section provides our initial conclusion, and the direction for our research.

---

[1] GOAL 1: Formalize the development, integration, and use of models to inform enterprise and program decision-making.
GOAL 2: Provide an enduring, authoritative source of truth.
GOAL 3: Incorporate technological innovation to improve the engineering practice.
GOAL 4: Establish a supporting infrastructure and environments to perform activities, collaborate, and communicate across stakeholders.
GOAL 5: Transform the culture and workforce to adopt and support digital engineering across the lifecycle.

[2] For the purpose of this paper, the terms “Model-Based Systems Engineering” and “Digital Engineering” will be considered synonymous. Model-Based Systems Engineering is defined in the second section.

The Essence of Model-Based Systems Engineering

The objective of systems engineering is to facilitate a process that consistently leads to the development of successful systems (Long & Scott, 2011). Model-Based Systems Engineering (MBSE) was envisioned to transform the reliance of traditional document-based work products to an engineering environment based on models. Model-Based Systems Engineering is the formalized application of modeling (static and dynamic) to support system design and analysis, throughout all phases of the system lifecycle, through the collection of modeling languages,[3] structures,[4] model-based processes,[5] and presentation frameworks[6] used to support the discipline of systems engineering in a model-based or model-driven context (Vaneman, 2016).

One can argue that systems engineering has always used models (i.e., diagrams, documents, matrices, tables, etc.) to the represent systems. In these traditional document-based models, the system's entities were represented multiple times, making it difficult, if not impossible, to view the system holistically. The transformation to MBSE means more than using model-based tools and processes to create document-based models, but shifts the focus to a virtual system model of the system, where there exists a singular definition for any system element.

To illustrate the concept of a virtual model of system, consider the dimensions of a systems engineering project (Figure 1; Larson et al., 2013; Vaneman & Vaneman 2018), where the cube represents a system. The system has height, width, and depth. System height provides a decomposition from the highest system level down to components and parts. System width defines the lifecycle of the system, and provides insight across the entire system lifecycle from concept definition to disposal. System depth provides the complex relationships between systems, functions, requirements, and so forth. The system

- satisfies capabilities;
- performs functions and has behavior;
- is defined by requirements;
- is testable;
- has risks; and

---

[3] Modeling Languages—Serve as the basis of tools and enable the development of system models. Modeling languages are based on a logical construct (visual representation) and/or an ontology. An ontology is a collection of standardized, defined terms and concepts and the relationships among the terms and concepts.

[4] Structure—Defines the relationships between the system's entities. It is these structures that allow for the emergence of system behaviors and performance characterizations within the model.

[5] Model-Based Processes—Provides the analytical framework to conduct the analysis of the system virtually defined in the model. The model-based processes may be traditional systems engineering processes such as requirements management, risk management, or analytical methods such as discrete event simulation, systems dynamics modeling, and dynamic programming.

[6] Presentation Frameworks—Provides the framework for the logical constructs of the system data in visualization models that are appropriate for the given stakeholders. These visualization models take the form of traditional systems engineering models. These individual models are often grouped into frameworks that provide the standard views and descriptions of the models, and the standard data structure of architecture models. The Department of Defense Architecture Framework (DoDAF) is an example.

- incurs costs.

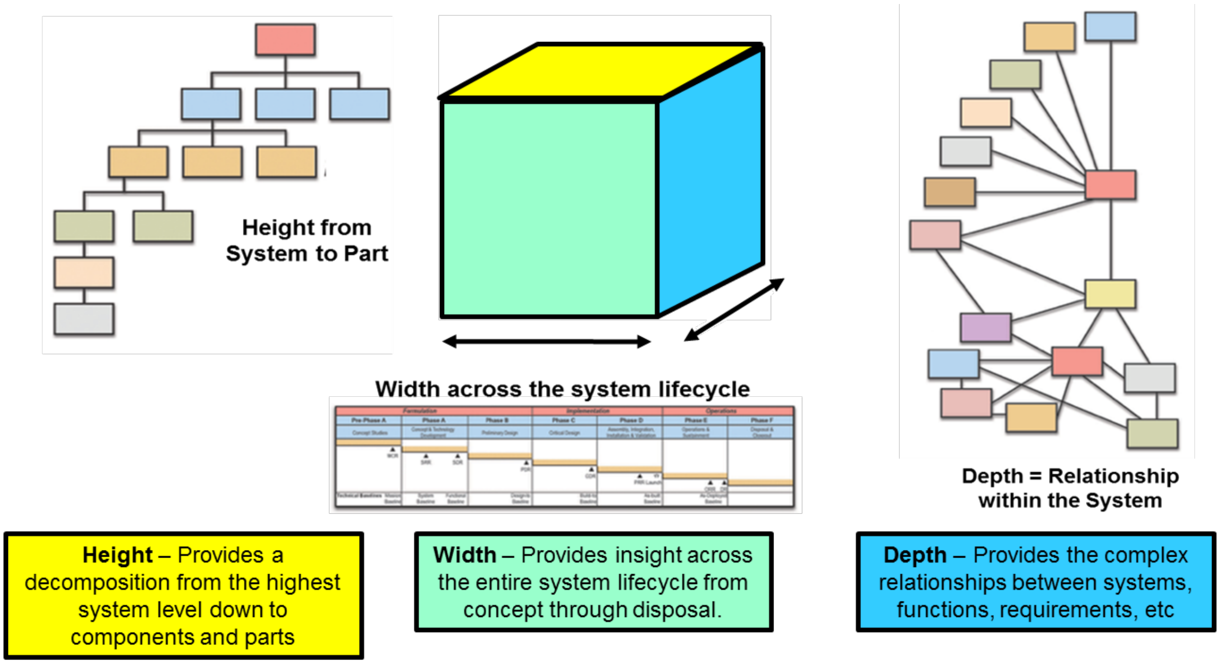


Figure 1. **Dimensions of a Systems Engineering Project**
(Larson et al., 2013; Vaneman & Vaneman, 2018)

In this virtual system model, each entity is represented as data, ideally only once, with all necessary attributes and relationships of that entity being portrayed. The key to defining this virtual system is model structure. Model structure defines the relationships between the system's entities, establishes concordance[7] within the model, and allows for the emergence of system behaviors and performance characterizations within the model (Vaneman, 2016).

To use the system entities to make programmatic decisions, the area of system focus must be isolated and portrayed in a manner so that decision-makers can arrive at an answer and make decisions. In MBSE, this is accomplished through the presentation framework, which provides the logical constructs of the system data in visualization models that are appropriate for the given stakeholders. These visualization models take the form of traditional systems engineering models, and are often grouped into standard viewpoints[8] and views[9]. The standard framework within the DoD is the Department of Defense Architecture Framework (DoDAF; Dam, 2014).

---

[7] Concordance (or referential integrity) is the ability to represent entity data so that it is consistent across views and abstraction levels (Vaneman, 2016).
[8] A viewpoint describes data drawn from one or more perspectives and organized in a particular way useful to decision-making.
[9] A view is a related set of information using models for the representation of data in any understandable format.

DoDAF defines eight viewpoints (Figure 2; Dam, 2014) and 52 views. The framework provides the flexibility for other “fit for purpose” views to be defined as needed to address a problem, provided that the spirit of the viewpoint is maintained.

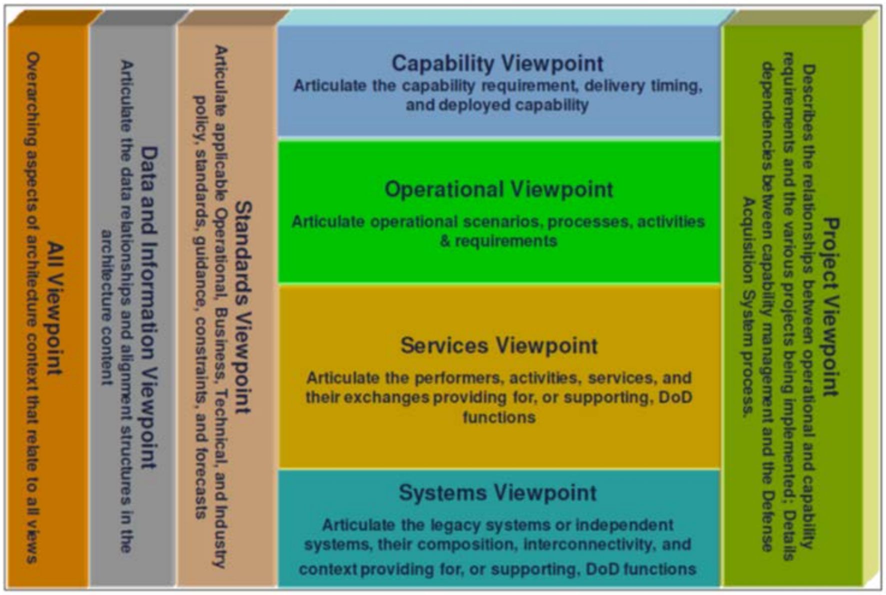


Figure 2. **DoD Architecture Framework**
(Dam, 2014)

This is an important feature of DoDAF since the framework only covers the architectural perspectives of the system, and does not include other system perspectives encountered throughout the lifecycle such as behavior, requirements, risks, verification and validation, and costs.

The 52 different DoDAF views can be represented in a document-based systems engineering environment. In such an environment, the diagram, not the system entities, becomes the “atomic” level and do not contain structure, and therefore lack concordance. In a MBSE environment, the system entities are at the “atomic” level, are related by structure, have concordance, and are represented in the 52 views.

These MBSE concepts represent a fundamental change in the systems engineering discipline, practices, and processes because they allow for the precise representation of the system’s entities and attribute, and through model structure, provide concordance. Complexity in the model-based environment is significantly reduced by separating and characterizing systems issues into various entity-based viewpoints and views. As such, MBSE requires a mindset change, a change in systems engineering processes, and a change in expectations of the artifacts required during the systems engineering process.

MBSE Development Throughout the System Acquisition Lifecycle

The DoD Digital Engineering Strategy Goal 1 calls for formalized planning, development, integration, management, and use of models to support engineering activities and decision-making across the lifecycle (DASD[SE], 2018). The realization of these goals will satisfy the transformation from the traditional document-based, to a model-based, systems engineering environment. This requires a fundamental shift in the development and use of engineering data to support system and programmatic decisions. In this environment, the model becomes central to the engineering of systems, and ultimately the way that decisions are made.

The System Acquisition Lifecycle Model identifies five primary phases, which take the system from concept develop and materiel solution analysis through operations and support. These phases, with their associated technical reviews, are briefly described in Table 1 (derived from Manning, 2019). The first three phases of the system acquisition lifecycle, through Engineering and Manufacturing Development culminating with Acquisition Milestone C, is where the most significant systems engineering activities occur. Implementing MBSE during later phases of the system acquisition lifecycle is possible, but programs should consider model adoption carefully. Beaufait (2018) demonstrated that MBSE can benefit programs post-Milestone C; however, introducing MBSE that far into the lifecycle of the program will face challenges related to cost, schedule, and a lack of understanding of MBSE. At this stage of the program, the implementation of MBSE has an additional cost that is likely not planned in the budget, and skeptical program managers are reluctant to make that investment in exchange for the promised benefits of MBSE (Beaufait, 2018).

The following discussion addresses model-development across the system acquisition lifecycle through Engineering and Manufacturing Development. Figure 3[10] is a relationship diagram that will be used to depict and explain model development and use throughout the lifecycle. While various DoDAF views and other systems engineering artifacts are shown in the diagram, the instantiation of these views only represents how the system data will be displayed within the presentation framework. Again, in an MBSE environment, the system is represented virtually; therefore, the data and relationships, not the views, are the "atomic" level of detail.

The System Lifecycle Model During the Materiel Solution Analysis Phase

The Materiel Solution Analysis (MSA) Phase assesses potential solutions for a needed capabilities identified by the stakeholder and formally documents in the Initial Capabilities Document (ICD). During this phase, various alternatives are analyzed to select the materiel solution and develop the strategy to fill any technology gaps. This phase describes the desired performance to meet mission requirements, defines metrics, identifies the operational requirements needed to satisfy the capabilities, and provides an initial analysis of risks (Manning, 2019).

[10] Figure 3 is meant to be viewed digitally so that it can be expanded.

Table 1. Summary of the DoD System Acquisition Lifecycle Phases

Lifecycle Phase	Description of the Lifecycle	Technical Reviews within Lifecycle
Materiel Solution Analysis (MSA)	MSA assesses potential solutions for a needed capability in an Initial Capabilities Document (ICD) The MSA phase is critical to program success and achieving materiel readiness because it's the first opportunity to influence systems supportability and affordability by balancing technology opportunities with operational and sustainment requirements.	• Initial Technical Review (ITR) • Analysis of Alternatives (AoA) • Alternative System Review (ASR) ♦ **Milestone A**
Technology Maturation and Risk Reduction (TMRR)	The purpose of TMRR is to reduce technology risk, engineering integration, lifecycle cost risk and to determine the appropriate set of technologies to be integrated into a full system. The TMRR phase conducts competitive prototyping of system elements, refines requirements, and develops the functional and allocated baselines of the end-item system configuration.	• System Requirement Review (SRR) • System Functional Review (SFR) • Preliminary Design Review (PDR) ♦ **Milestone B**
Engineering and Manufacturing Development (EMD)	EMD is where a system is developed and designed before going into production. The phase starts after a successful Milestone B - the formal start of any program. The goal of this phase is to complete the development of a system or increment of capability, complete full system integration, develop affordable and executable manufacturing processes, complete system fabrication, and test and evaluate the system before proceeding into the Production and Deployment (PD) Phase.	• Critical Design Review (CDR) • Test Readiness Review (TRR) ♦ **Milestone C**
Production and Development (PD)	PD is where a system that satisfies an operational capability is produced and deployed to an end user. The phase has two major efforts; (1) Low-Rate Initial Production (LRIP) and (2) Full-Rate Production and Deployment (FRP&D). The phase begins after a successful Milestone C review.	• Full Rate Production (FRP) • Initial Operational Capability (IOC) ♦ **Full Operational Capability (FOC)**
Operation and Support (OS)	OS is where a system that satisfies an operational capability is produced and deployed to an end user. The phase has two major efforts: (1) Low-Rate Initial Production (LRIP) and (2) Full-Rate Production and Deployment (FRP&D). The phase begins after a successful Milestone C review	• Sustainment ♦ **Disposal**

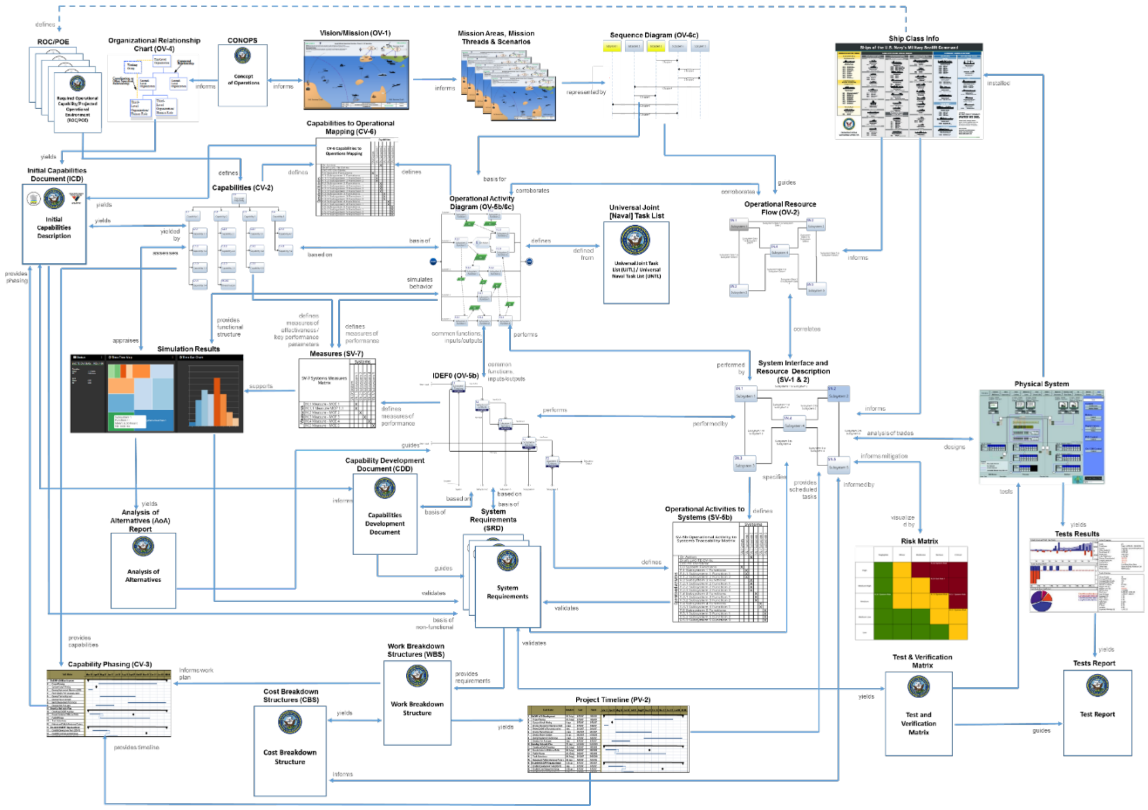


Figure 3. **MBSE Development Process**

The MSA process (see Figure 3) begins with the identification of stakeholder-needed capabilities, often contained in the Required Operational Capability/Projected Operational Environment (ROC/POE) documents. Often a system will be governed by multiple ROC/POE documents due the breadth of the future system deployment. The ROC/POE serves as the basis for the Capability Taxonomy (CV-2), the beginning of the modeling effort. Many ROC/POEs capture the majority of the capabilities to be satisfied, however, usually does not contain all of them.

The Concept of Operations (CONOPS), often provided by the stakeholders, provides additional insights into the capabilities required. The CONOPS often includes an overarching High-Level Operational Concept Graphic (OV-1), which shows an overview of the operational concept, as well as the vision and mission of the system. The CONOPS usually identifies mission areas that contain mission threads and scenarios. The mission threads can be represented in scenario-focused OV-1. These OV-1s offer sufficient detail to visualize the steps of the operations. These mission threads can be further represented by sequence diagrams (OV-6c). The OV-6c serves as the basis for the Operational Activity Model (OV-5b-6c).

The OV-5b/6c is a fit for purpose view that represents the sequence of functions as well as the inputs and outputs for each function. The functions in the OV-5b/6c are the same functions contained in the OV-6c, viewed from a different perspective. The functions can be grouped by the subsystem that they are assigned to. The OV-5b/6c can be further developed by information from the Universal Joint Task List (UJTL).

The operational entities depicted in the OV-5b/6c are based on the capability entities depicted in the CV-2. Thus, function *x* is based on capability *y.* These relationships are shown in the Capabilities to Operational Mapping (CV-6). Using the Organizational

Relationship Chart (OV-4), the initial Capability Phasing (CV-3), the CV-2, and the CV-6, the ICD can be defined. In a MBSE environment, the ICD is an integral part of the model and thus has concordance with the views used to portray it.

The functions contained in the OV-5b/6c can be viewed differently by using the IDEF0 (OV-5b). The functional entities in the OV-5b are the same functional entities in the OV-5b/6c. These entities are only represented once on the model, but can be viewed in several different ways, thus the model exhibits concordance. The OV-5b also contains the inputs and outputs included in the OV-5b/6c. The OV-5b goes further in capturing system data by identifying the policies, guidelines, rules, and regulations that govern the functions. This view also initially identifies the system elements and relates them to the functions that they satisfy.

With the data captured thus far, two additional complimentary views—the Operational Resource Flow (OV-2) and the System Interface Description (SV-1)—can be developed. Both of these views have a common structure that depicts the system elements that were first identified in the OV-5b. The connections in the OV-2, influenced by the functions in the OV-5b/6c, represent the data, and data characteristics (i.e., direction of flow, type, size, frequency, and duration), that flow between two system elements. The connections in the SV-1 represent the physical means (e.g., pipes, data links) by which data is transferred. The OV-2 defines the “what” that needs to be transferred, and is correlated to SV-1, which shows “how” the data is transferred.

With the data developed to this point, system measures can be defined in the Systems Measures Matrix (SV-7). The Measures of Effectiveness (MOE) and the Key Performance Parameters (KPP) are defined by the capabilities depicted in the CV-2. The Measures of Performance (MOP) are derived by the operational entities depicted in the OV-5b and OV-5b/6c.

At this point, the data captured can be used to perform the analysis of alternatives (AoA). An AoA typically consist of the initial assessment of three areas—cost, risk, and performance. The system entities are related to operational entities via the OV-5b, and risk, and initial costs, in the SV-1. System performance is represented mathematically within the operational entities. Many MBSE tools allow for these entities to be defined by several statistical distributions, thereby allowing for discrete event and Monte Carlo Simulation.

The last activity engineered in the MSA is development of the draft Capabilities Development Document (CDD). The CDD specifies the operational requirements for the system that will deliver the capabilities, that meet the operational performance requirements, specified in the ICD, and depicted by the entities developed thus far (Manning, 2019). The primary views used to develop the CDD are the CV-2 and OV-5b.

Milestone A marks the end of the MSA Phase. The purpose of Milestone A is to make recommendations and seek permission to enter the Technology Maturation and Risk Reduction (TMRR) Phase (Manning, 2019).

The System Lifecycle Model During Technology Maturation and Risk Reduction

The purpose of the Technology Maturation & Risk Reduction Phase is to reduce risks associated with technology, integration, and lifecycle cost, determine the appropriate set of technologies to be integrated into a full system, validate designs and costs, and evaluate manufacturing processes for the system build. TMRR refines requirements, conducts competitive prototyping of system elements, and develops the functional and allocated baselines of the final system configuration (Manning, 2019).

The modeling process (see Figure 3) continues with the further development of the CDD. The CDD guides the development of the system requirements document (SRD). The SRD defines system level functional and performance requirements for a system (Manning, 2019). While the SRD is guided by the CDD in a document-based engineering environment, in a MBSE environment it is primarily derived from the OV-5b, SV-1, and the Operational Activities to Systems Matrix (SV-5b). As the system engineering effort progresses, these views are iteratively refined, with more detailed data being developed with each iteration, thereby allowing for a natural progression of the requirements hierarchy from ICD to the CDD, to the SRD, and ultimately to sub-system requirements documents.

In a MBSE environment, requirements are derived from the system-entity data, and corresponding relationships, in the model. The primary view to visualize the relationships used to derive functional requirements is the OV-5b. This view contains all of the data required (system elements, functions, inputs, outputs, controls) to generate requirements. The initial system structure also influences the system requirements.

The interfaces are defined via the SV-1. As previously stated, the flow interfaces between system elements in the OV-2 need to be correlated with the physical interfaces in the SV-1 to identify the proper interface requirements. The SV-5b is used to validate the system requirements by ensuring that each operation is satisfied by a system element, and each system element is assigned to to an operation. The draft CV-3, which was developed in MSA, is matured here.

A correllary to the SRD is the Test and Verification Matrix, which shows how the system will be tested. Developing a Test and Verification Matrix in conjunction with the SRD is a good practice that validates that the requirements can be tested as written.

Once a detailed set of requirments is defined, the Work Breakdown Structure (WBS) can be developed. A WBS is a tool used to define a project in discrete work elements. It relates the elements of work to be accomplished to each other and to the end product. It's used for planning, development of the Cost Breakdown Sructure (CBS), and the execution and control of the system development (Manning, 2019). The CBS allocates costs to the various levels of the WBS.

The WBS informs the development of the final Capability Phasing (CV-3). A Project Timeline (PV-2) is derived from the WBS. This view depicts the detailed schedule for system development.

During TMRR, the system in interatively developed, and a comprehensive risk assessment is conducted. The purpose of the risk assessment is to identify the root cause of cost, schedule, and performance issues within the systems. In a MBSE environment, the risks are related to system elements portrayted in the SV-1 and SV-2.

Towards the end of TMRR, system devlopment has sufficiently matured where three-dimensional models and prototypes are developed. TMRR ends with Milestone B, where the program office seeks approval to enter the Engineering and Manufacturing Development (EMD) Phase. Milestone B is considered the official start of the program (Manning 2019).

The System Lifeycle Model During Engineering and Manufacturing Development

Systems design and development continue with the Engineering and Manufacturing Development (EMD) Phase, where the system is developed and designed before going into production. The goal of EMD is to complete the development of a system or increment of capability, complete full system integration, develop affordable and executable manufacturing processes, complete system fabrication, and test and evaluate the system before proceeding into the Production and Deployment (PD) Phase (Manning, 2019).

EMD consists of two major efforts: integrated system design and system capability; and manufacturing process demonstration. These two major efforts integrate the end item components and subsystems into a fully operational and supportable system. They also complete the detailed design to meet performance requirements with a producible and sustainable design and reduce system level risk. EMD typically includes the demonstration of a production prototype (Manning, 2019).

During EMD, MBSE is used for further iterative developed. As the system models are refined and further developed, other models within the framework must be changed to represent the new system baseline. Different system components lead to different operations. As the system and operations are changed, the capabilities must be re-evaluated to ensure that they are still being satisfied. Changes in the system baseline also impact risks—maybe new risks emerge, or current risks are mitigated. The change in the system baseline will likely have an impact on both cost and schedule. Given that the MBSE environment exhibits concordance, when a change is made in a system element it is captured in the model and then the changed element is portrayed throughout the model and all of the different viewpoints.

The MBSE environment can also be used to support the testing and verification of the system. During the development of the SRD, a Test and Verification Matrix was developed. This Test and Verification Matrix can be used to develop a test plan, which can be executed throughout the test and verification process.

Milestone C marks the end of the EMD Phase. The purpose of Milestone C is to make a recommendation or seek approval to enter the Production and Deployment (PD) Phase (Manning, 2019).

The system model discussed in this section provides the data required to make programmatic decisions. The system model will be used in Section IV to address the criteria during the system milestones reviews.

Technical Reviews in an MBSE Environment

The DoD Digital Engineering Stategy Goal 1 specifically states that the model of the system should be used for decision-making. A series of decision-making events within the system acquisition lifecycle that could benefit from the MBSE approach are the system acquisition technical reviews.

System acquisition technical reviews are discrete points in time, within a system's lifecycle, where the system is evaluated against a set of program-specific accomplishments (criteria). These criteria are used to track the technical progress, schedule, and program risks. The technical reviews serve as gates, that when successfully evaluated, demonstrate that the program is on track to achieve its final program goals, and should be allowed to proceed to the next acquisition phase. Figure 4 shows the technical reviews superimposed on the Systems Acquisition Lifecycle Model (derived from Defense Acquisition University, 2018).

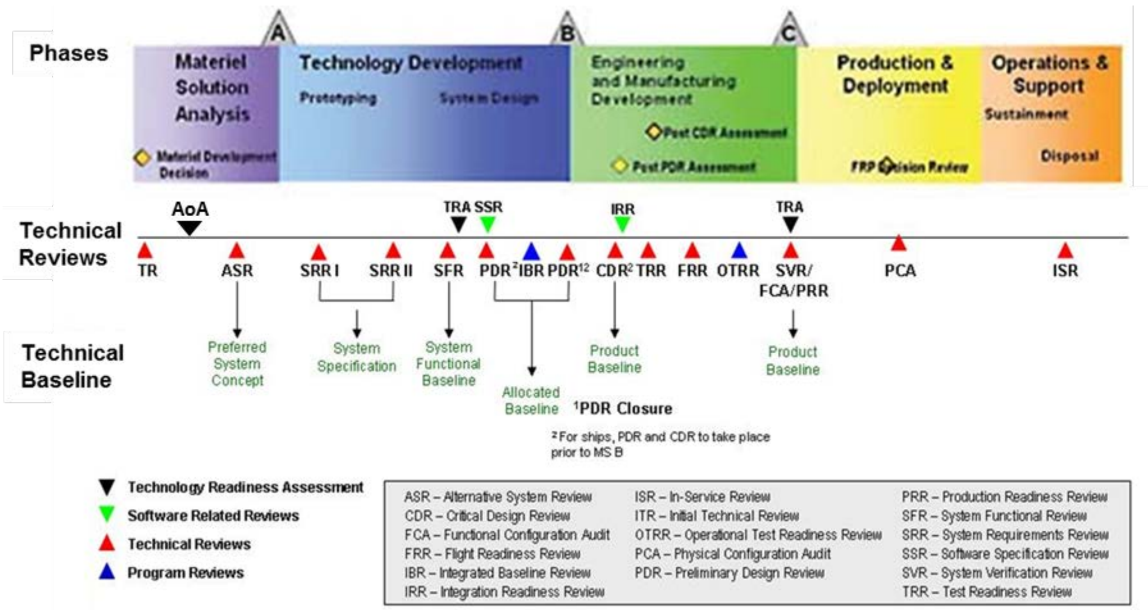


Figure 4. **System Acquisition Lifecycle Model**
(Adapted from Defense Acquisition University, 2018)

Currently, milestone reviews are based around lengthy reviews of static, contractually obligated documents that are used to demonstrate successful completion of the exit criteria. System documents and artifacts are baselined to represent the system and traditionally serve as evidence of programmatic progress. Typically, these documents are not synchronized, thus lack concordance. As discussed in the MBSE approach above, the "virtual" model of the system is created where each entity is ideally modeled once but represented several times. For technical reviews, the model-based data is depicted by views within a presentation framework, similar to a document-based review.

The difference in concordance is maintained, allowing decision-makers access to insights that have been heretofore unavailable. This includes emerging system behavior, and the assurance that a common system baseline is used to report on various aspects of the systems. Using the model as the source for decision-making throughout the system acquisition lifecycle is a significant departure since programs often generate unique artifacts for the sole purpose of the reviews.

Table 2 shows the applicability of model-based systems engineering views to the system acquisition lifecycle. The relationships in the matrix were made by correlating the generic criteria for each review, or content of the major documents, to the data in each system engineering view. The existing review criteria is designed to be addressed by document-based processes. These criteria need to be revised to account for the new insights that can be gleaned through a model-based approach.

As an example, consider the Alternative Systems Review (ASR). The ASR assesses the preliminary materiel solutions that have been developed during the Materiel Solution Analysis (MSA) Phase. The technical review ensures that one or more proposed materiel solution(s) have the best potential to be cost-effective, affordable, and operationally effective and suitable, and can be developed to provide a timely solution to at an acceptable level of risk to satisfy the capabilities listed in an Initial Capabilities Document (ICD; Manning, 2019).

Table 2. Applicability of Systems Engineering Views With the Systems Acquisition Lifecycle

	Materiel Solution Analysis			Technology Delevoment				Engineering and Manufacturing Development			Documents			
Systems Engineering Views	Analysis of Alternatives (AoA)	Alternative Systems Review (ASR)	Milestone A	System Requirements Review (SRR)	System Functional Baseline (SFB)	Preliminary Design Review (PDR)	Milestone B	Critical Design Review (CDR)	Test Readiness Review	Milestone C	Initial Capabilities Document	Capability Development Document (CDD)	System Requirements Specifications	Test Report
CV-2	x	x		x							x	x		
CV-3	x	x		x							x	x		
CV-6		x		x							x	x		
OV-1	x	x		x							x	x		
OV-2	x	x		x								x	x	
OV-4		x		x							x	x		
OV-5b	x	x		x	x	x						x	x	
OV-5b/6c	x	x		x	x	x					x	x	x	
OV-6c	x	x		x	x	x					x	x	x	
PV-2					x	x		x						
SV-1	x	x		x	x	x		x	x			x	x	x
SV-2					x	x		x	x				x	x
SV-5b				x	x	x		x	x			x	x	x
SV-7	x	x		x	x	x		x	x			x		x
Cost Esttimate	x			x				x						
Risk Matrix	x	x			x	x		x						
Similation Results	x			x		x		x	x				x	x
Test and Verification Matrix						x		x	x					x
Test Results								x	x					x
Work Breakdown Structure					x	x		x						

The system engineering process typcially has progressed to the point where the following information is available for the ASR (TTCP, 2014):

- Description of how the users will conduct operations, and how they expect to use the new system in this context of major mission areas and scenarios;
- Statement of need, and capabilities, in terms oriented to the system users, the stakeholders, and independent of specific technology solutions;
- The required system characteristics and context of use of services and operational concepts are specified;
- Major stakeholder capabilities are identified and documented, but detailed system requirements analysis has yet to be completed;
- The constraints on a system solution are defined;
- Results of an analysis of alternatives with a recommended preferred solution;
- Initial plans for systems engineering (e.g., Overview and Summary information [AV-1], Systems Engineering Plan [SEP], Systems Engineering Management Plan [SEMP]) providing the notion of “how” this system can be realized, including the level of process and process maturity needed to generate a system of the required complexity;
- Initial definition of the environment and the characteristics of the threat;

- Initial test & evaluation strategy including test cases derived from user operational vignettes, concept of operations, and capability description;
- An understanding of where the greatest risks and challenges may reside.

An analysis of the ASR generic criteria (DAU, 2018) is shown in Table 3. First the criteria is reviewed in the context of traditional reviews. Many of the criteria were assessed to be partially satisfied. These results do not suggest that ASRs have not been performed properly in the past. Rather, given the absence of concordance in document-based reviews, the criteria requiring different types of data, using different artifacts is extrmely difficult to achieve efficiently and effectively. All of the criteria were assessed to be satisfied in MBSE environment because of the concordance. The model-based systems engineering views needed to address the criteria are also shown in the table.

Conclusions and Future Work

As DoD organizations migrate to the MBSE environment, efficiencies will be gained by transitioning from the document-based reviews to model-based reviews. Model-based reviews allow for complexity to be managed more efficiently because data, in lieu of "systems engineering products," is the commodity that will be used to evaluate the exit criteria. The MBSE milestone reviews will provide greater insight with faster comprehension of the details across a program's lifecycle. This will not only provide efficiencies for the review, but will also improve the program's cost and schedule efficiency.

This paper provided some additional concepts developed during the initial phase of our research. These concepts are in the spirit of the DoD Digital Engineering Strategy Goal 1: "Formalize the development, integration, and use of models to inform enterprise and program decision-making" (DASD[SE], 2018).

While Goal 1 became the natural focus, other goals need to be considered when developing processes to implement a true MBSE environment. The most signifiant goal is one that is often overlooked, Goal 5: Transform the culture and workforce to adopt and support digital engineering across the lifecycle.

The systems engineering culture must change to focus on the virtual model of the system, and away from technical documentation. This is critical when considering conducting technical review in an MBSE environment.

Table 3. ASR Criteria and Related Views

Criteria	Satisfied by Traditional Review?	Satisfied by MBSE?	Views
Is the initial CONOPS updated to reflect current user position about capability gap(s), supported missions, interfacing/enabling systems in the operational architecture?	Partial	Yes	CV-2, CV-6, OV-1, OV-6c, OV-5b/6c
Are the required related solutions and supporting references (ICD and CDDs) identified?	Partial	Yes	CV-2, CV-3, CV-6, OV-4, OV-5b, OV-5b/6c
Are the thresholds and objectives initially stated as broad measures of effectiveness and suitability (e.g., KPPs)?	Yes	Yes	CV-2, OV-5b, OV-5b/6c, SV-7
Is there a clear understanding of the system requirements consistent with the ICD?	Yes	Yes	CV-2, CV-3, CV-6, OV-4
Are high-level descriptions of the preferred materiel solution(s) available and sufficiently detailed and understood to enable further technical analysis in preparation for Milestone A?	Partial	Yes	OV-2, OV-5b, SV-1
Are interfaces and external dependencies adequately defined for this stage in lifecycle?	Partial	Yes	OV-2, SV-1
Are system requirements sufficiently understood to enable functional definition?	Partial	Yes	OV-5b, OV-5b/6c
Is a comprehensive rationale available for the preferred materiel solution(s), based on the AoA?	Partial	Yes	CV-2, CV-3, CV-6, OV-2, OV-4, OV-5b, OV-5b/6c.
Can the proposed material solution(s) satisfy the user needs?	Partial	Yes	CV-2, CV-3, CV-6, OV-2, OV-5b, OV-5b/6c.
Have cost estimates been developed, and were the cost comparisons across alternatives balanced and validated?	Partial	Yes	OV-2, OV-5b, SV-1
Have key assumptions and constraints associated with preferred materiel solution(s) been identified?	Partial	Yes	OV-2, OV-5b, SV-1

This paper considers systems engineering throughout the acquisition lifecycle using a model-based approach. While MBSE was discussed, and the underlying principles of capturing system elements only once and using model structure to establish concordance were briefly discussed, this research focused heavily on the information portrayed in the various views within the presentation framework. In a true MBSE environment, systems engineering will be conducted at the entity level, thus making the model the focus and the views secondary.

The systems engineering community has not widely considered the effects on making the model the focus. One area that is being explored by our ongoing research is how will the technical review criteria need to be changed to gain the full benefit of model-based insights.

References

Beaufait, J. (2018). *Decision tools to implement MBSE Post Milestone C* (Master's thesis).

Dam, S. (2014). *DoD Architecture Framework 2.02*. Manassas, VA: SPEC Innovations.

Defense Acquisition University. (2019). *Defense acquisition guidebook.* Retrieved from https://www.dau.mil/tools/dag

Deputy Secretary of Defense for Systems Engineering (DASD[SE]). (2018). *Department of Defense digital engineering strategy.* Washington, DC: Author.

Larson, W., Kirkpatrick, D., Verma, D., Date, T., & Sellers, J. J. (2013). *Space systems engineering.* Boston, MA: McGraw Hill.

Long, D., & Scott, Z. (2011). *A primer for model-based systems engineering* (2nd ed.). Blacksburg, VA: Vitech Corporation.

Manning. (2019). *AcqNotes: Defense acquisition made easy.* Retrieved from http://acqnotes.com

The Technical Cooperation Program (TTCP). (2014). *Recommended practices: System of systems considerations in the engineering of systems* (TTCP Technical Report TR-JSA/TP4-1-2014).

Vaneman, W. K. (2016). Enhancing model-based systems engineering with the life-cycle modeling language. In *Proceedings of the 10th Annual IEEE Systems Conference*. IEEE.

Vaneman, W. K. (2019). *Model-based systems engineering de-mystified*. Unpublished document.

Vaneman, W. K., & Vaneman, N. (2018). *De-mystifying model-based systems engineering.* 21st Annual Systems Engineering Conference.

Panel 20. Contract Management Challenges

Thursday, May 9, 2019	
12:45 p.m. – 2:00 p.m.	**Chair: Stuart Hazlett,** Deputy Assistant Secretary of the Army (Procurement) ***Measuring Service Contract Performance: Preliminary Findings on Effects of Service Complexity, Managerial Capacity, and Prior Relationship*** Gregory Sanders, Andrew Hunter, Robert Karlen, and Zachary Huitink, Center for Strategic and International Studies ***Predicting Bid Protests: What Should Acquisition Teams (Not) Do?*** Tim Hawkins, Western Kentucky University ***Commercial When Convenient—A Gray Area in the Federal Acquisition Regulations That Makes Determining Fair and Reasonable Pricing Difficult*** Brian Gladstone and Mark Kaye, Institute for Defense Analyses

Stuart Hazlett—Mr. Hazlett assumed the position of Director of Contracting, United States Army Corps of Engineers (USACE) in January 2012. He serves as the delegated Head of the Contracting Activity by the Chief of Engineers and directly exercises authority over three Principal Assistants Responsible for Contracting, ten

Regional Contracting Chiefs, and four Center Contracting Chiefs. Mr. Hazlett has overall responsibility for managing the contracting activity in support of USACE that has presence in more than 30 countries supporting Military Programs, Civil Works, Real Estate, and Research and Development. Under his cognizance, Mr. Hazlett executes approximately 62,000 contracting actions with obligations annually of $25 billion.

Before taking his current position, Mr Hazlett served as the Deputy Director of Program Acquisition & Strategic Sourcing (PASS) for Defense Procurement and Acquisition Policy (DPAP). In this role he served as a senior advisor to the Director of DPAP, Director of Defense Pricing, and the Under Secretary of Defense for Acquisitions, Technology, and Logistics in the Office of the Secretary of Defense. Mr. Hazlett supported Defense Acquisition Boards, Defense Acquisition Executive Summaries, and various Overarching Integrated Product Teams for major defense acquisition and special interest programs. He promoted the development of sound program acquisition strategies and fostered continuous and effective communication within the acquisition community to ensure effective application of associated statute and policy.

Prior to the aforementioned position he served as the Deputy Director of Strategic Sourcing for the Director of DPAP. In collaboration with the DoD components, his directorate conducted spend analyses and facilitated business solutions that permitted the department to achieve best value in strategically sourcing goods and services within a $370B spend across the department. He served as the Chair of the Strategic Sourcing Directors Board (SSDB) and as the department's representative to the Federal Chief Acquisition Officer Council Strategic Sourcing Working Group.

Measuring Service Contract Performance: Preliminary Findings on Effects of Service Complexity, Managerial Capacity, and Prior Relationship

Andrew Hunter—is a senior fellow in the International Security Program and director of the Defense-Industrial Initiatives Group at CSIS. From 2011 to 2014, he served as a senior executive in the Department of Defense, serving first as Chief of Staff to Under Secretaries of Defense (AT&L) Ashton B. Carter and Frank Kendall, before directing the Joint Rapid Acquisition Cell. From 2005 to 2011, Hunter served as a professional staff member of the House Armed Services Committee. Hunter holds an MA degree in applied economics from the Johns Hopkins University and a BA in social studies from Harvard University.

Greg Sanders—is a fellow in the International Security Program and deputy director of the Defense-Industrial Initiatives Group at CSIS, where he manages a research team that analyzes data on U.S. government contract spending and other budget and acquisition issues. In support of these goals, he employs SQL Server, as well as the statistical programming language R. Sanders holds an MA in international studies from the University of Denver and a BA in government and politics, as well as a BS in computer science, from the University of Maryland. [gsanders@csis.org]

Jonathan Roberts—is currently an Army ROTC cadet pursuing an MA in Applied Economics at Georgetown University. He is also a graduate of Louisiana Tech University with a BS in Mechanical Engineering and a BA in Political Science. His research interests include economic statecraft in the era of great power competition and the economics of technological innovation. [jroberts1187@gmail.com]

Contributing Authors:

Zach Huitink—is a postdoctoral fellow at Syracuse University and an externally affiliated researcher working with the Defense-Industrial Initiatives Group at CSIS. His research and analysis focuses on national security, government contracting, and public-private partnerships. Huitink holds a PhD from the Department of Public Administration and International Affairs at Syracuse University's Maxwell School.

Xin Yuan—is an intern with the Defense-Industrial Initiatives Group (DIIG) at CSIS. She mainly works on service contract inventory data consolidation, processing, and visualization. She holds a BS in economics with a concentration on public finance & taxation from Tianjin University of Finance and Economics, and an MS in business analytics from George Washington University.

Robert Karlén—graduated from the University of Washington with a BA in political science and international security. At CSIS, Karlén helped with research and writing for various projects, including artificial intelligence, contracting performance, and Future Vertical Lift. Outside of CSIS, his work focuses on defense innovation and contracting.

Xinyi Wang—was a research intern with the Defense-Industrial Initiatives Group (DIIG) at CSIS. Her work here includes exploratory statistics and econometrics modeling for multiple quantitative research topics, like monopolies in defense industry. She holds a BA in mathematical economics from Shanghai University of Finance and Economics, and an MS in business analytics from George Washington University.

Abstract

Services contracts have a distinct set of challenges relating to the uncertainty and the challenges of measuring performance. Past researchers identified three overarching characteristics of interest: service complexity, management capacity, and the relationship between the buyer and the contractor. Researchers have often turned to surveys of government contracting personnel to take on the challenge of measuring service contract performance. This report takes a large-dataset, quantitative approach to looking at service

contract outcomes derived from information in the publicly available Federal Procurement Data System. This iteration of the report focuses on a newly developed outcome: the extent to which the government exercised available options as an indication of positive performance outcomes.

Introduction

Services contracts have long been understood to be distinct in key ways from their product counterparts in ways that add ambiguity and their own sets of challenges. Products are countable or otherwise objectively measurable, and while testing to see whether they meet requirements can prove complicated and controversial, there is at least a common item being argued over and measured. Service contracts inherently put more attention on the qualitative aspects of labor. Simple service contracts, like transportation or custodial services, have straightforward results to evaluate but can nonetheless introduce a host of concerns if, for example, taking place in contingency environments such as Afghanistan. Even familiar services like construction often must be evaluated not just on the quality of the final product but also the creation process which is often not contained in an easily measurable outcome and can bring a host of disruptions. The most challenging services can be those that do something new or ill-defined, where trying to put all the details in the contract at the outset might not only be an exercise in futility, but actively counterproductive. In such situations, the buyer and contractor have to solve problems together that were not fully anticipated when the contract was initiated.

For the U.S. federal government, in particular the Department of Defense (DoD), services constitute a significant portion of contract spending but are often a lower priority from a regulatory and policy perspective. This is even more so the case when R&D is classified as a service, as it is for the purposes of this iteration of the paper. This observation comes not just from critics in Congress, which has a range of concerns about services contracts, but also from the DoD itself where improving services acquisition tradecraft was a prominent part of the Better Buying Power initiatives. Some of the history of these acquisition reform efforts can be seen in McCormick et al. (2015), *Measuring the Outcomes of Acquisition Reform by Major DoD Components*, but suffice to say the problems of services contracting have long been a known issue.

While the prioritization of major defense acquisition programs over services acquisition is specific to the DoD, the challenges of services contracting are universally acknowledged—in the private sector, by sectors and levels of government, and in nonprofit organizations. This study takes a new quantitative look at services contract performance by employing the Federal Procurement Data System, an open source transaction database with records of more than a million service contracts within the past decade. This large dataset approach builds on past research regarding the public and private sector that often relied on surveys with smaller sample or case studies.

Scope

The research project seeks to answer the following questions:

- Under what circumstances are services contracts likely to encounter challenges, as measured by terminations and cost ceiling breaches, or prosper, as measured by the exercising of options?
- What services contracting policy choices influence these outcomes, for better or worse?

This iteration of the study focuses specifically on DoD contracts within a 2008 to 2015 study period. Past CSIS work with the contract dataset have covered both products and services contracts, but to better focus on the challenges of services contracting, this effort has focused on service contract complexity, contract management capacity, and the relationship between the contractor and the buyer. Past CSIS studies have looked at the performance outcomes of terminations and ceiling breaches, and the final technical deliverable for this project will as well. However, for reasons of novelty and brevity, this paper focuses on a performance outcome the study team newly explored in this project: the choice of whether to exercise contract options and the implication of positive performance when the acquirer chooses to do so.

Literature Review

This literature review will clearly delineate the different aspects of services contract management in several sections. In general, this includes contract complexity, management capacity, and trust/relationships. Additionally, there are some service specific considerations that will be of use to this study. Finally, this review includes only the academic theory and evidence. It does not include GAO reports, but including these reports in the final product will provide a level of context for our findings and conclusions that will be invaluable. Useful tables have been included from the literature because they have value in articulating some of the more ambiguous concepts.

There are incredibly few comprehensive definitions for contract management. For a broad definition, contract management may be defined as

> all activities performed by the government … that are relevant to contracts with private or nonprofit organizations … such as writing or creating the Invitation to Bid or Request for Proposal, devising a rating system for bid responses, rating the bid responses, awarding the contract, additional negotiations leading to a signed contract, and contract administration. (Lawther, 2002)

Contract Complexity

It stands to reason that the relative complexity of a given contract is a determinant of the level of cost (in labor, funds, or both) required by the government to effectively manage it, and in this vein, the literature delineates between low-, mid-, and high-complexity. For low-complexity contracts, “specificity and monitoring are simple and undemanding” (Joaquin & Greitens, 2012, p. 809). “Under mid-complexity scenarios, requests for proposals are more detailed and specific, and managers need to possess more technical expertise” (Joaquin & Greitens, 2012, 809). For highly complex services, and when understanding of the service delivery means are not clear, the agency and the contractor should enter into a true public-private partnership and recognize that the service to be provided will evolve in a dynamic manner, echoing studies by Fernandez (2007, 2009; Joaquin & Greitens, 2012, p. 809).

High levels of task complexity and uncertainty at the federal level can be considered as those tasks where the government requires a definitively new service or capability. From the defense perspective, this could be new software architecture, an emerging hardware technology, or specified research and development. Such high-complexity contracts complicate the ability of contract managers to write contractual requirements that are comprehensive and highly detailed, which most literature has presumed was a necessity in successful contracting. The results are mixed on the need for specificity. Brown and Potoski (2003, 2006) find support for this in overall contracting, while Fernandez (2007, 2009) determines that asset specificity is unrelated to service contracting success. Moreover, even moderate levels of complexity and uncertainty increase the likelihood that some of the contract requirements will be erroneous. A successful contract, then, may require that the principal and agent interact frequently to make “necessary adjustments in work processes, performance standards, quantities, and prices” and fill in the gaps in the contract (Fernandez, 2007, p. 1127). One additional consideration is that, contracting for management tasks can carry a large amount of risk, specifically that the government will enter into a monopoly relationship with the vendor (Brown & Potoski, 2006, p. 327).

Management Capacity

While there are various definitions of management capacity in the literature, many of them do not create a complete picture of the actual scope of managing contracts. The all-encompassing definition, as is required here, is provided by Brown and Potoski (2003):

> Contracting is not a one-size-fits-all proposition. The success or failure of any alternative service-delivery arrangement likely depends on how well governments can manage the entire contract process, from assessing the feasibility of contracting through implementation to monitoring and evaluation-activities that require strong government contracting capacity. Governments investing in contract-management capacity may be better positioned to harness the promise of effective contracting while avoiding its pitfalls. (p. 153).

Governments invest in contract-management capacity because contracting is a complex process, fraught with potential problems and pitfalls. In fact, “governments can respond to poor conditions by investing in the managerial capacity to identify suitable situations for contracting, negotiate strong contracts, and monitor vendor performance” (Brown & Potoski, 2003, p. 162) Romzek and Johnston (2002) determine seven factors that positively influence service-contracting effectiveness: competition among providers, resource adequacy, planning for performance measurement, training for state contract managers, evaluation of contractor staffing capacity, evaluation of contractor financial management capacity, and theoretical rationale for reform. These and other responsibilities of the government as contract managers have been discussed, but they often fall into general bins. The overall literature expresses a range of opinions on the bins that explain management capacity. This is largely due to their different definitions and measures. However, the most popular systems come from Brown and Potoski (2003) and Yang, Hsieh, and Li (2009). Brown and Potoski (2003) determine three subfields of management capacity: assessment, implementation, and evaluation capacities. Yang et al. (2009) build on this model by adding another measure to Brown and Potoski’s system and renaming it. Therefore, formulation capacity for them is the same as implementation capacity for Brown and Potoski while Yang et al. determines implementation capacity to be the ability of the government to create and sustain a public-private partnership for contracts. This is an important delineation because many findings detail the effects of public-private partnership ability which is the capacity of the government to manage relationships and, as such, falls under management capacity.

Diving into the separate bins, contract assessment is first in the process. Yang et al. (2009) find that increased assessment capacity is positively associated with cost reduction, but it is not associated with efficiency increase or quality improvement. Additionally, Fernandez (2007), using substantively weighted least squares to statistically identify the top performers, finds that ex ante evaluation (an assessment responsibility) is a significant determinant of the most successful contracts. Moreover, Yang et al. find a time component to management capacities. For agenda setting, the “process during which the values and preferences of stakeholders are manifested and compromised," the impact on cost reduction decreases as time passes for assessment (Yang et al., 2009, p. 687). Another aspect of assessment is the determination of asset specificity from the outset. Planning asset-specific requirements for either end product or production tasks requires significant expertise and insight. Brown and Potoski (2006) state that "managers need to master the complex production process in order to ensure that production tasks integrate with other processes required to deliver the service." However, Fernandez (2007, 2009) finds under many different statistical methods (OLS, SWLS, 2SLS) that although it is common for government managers to develop design specifications and hand it to industry to build, this is a retrograde approach, which "makes it impossible to hold contractors responsible for delivering solutions that work, because as long as what's delivered meets the specifications, it's the government's fault if the products don't work." Interestingly enough, Fernandez (2007) finds that task uncertainty is a very significant factor in diminishing contract performance. This differentiation bears the distinction that defining the asset as specifically as possible does not necessarily define what the government wants contractors to accomplish in the contract. Additionally, this suggests another side where contract specification capacity is not the problem, but rather a cultural shift away from strictly measurable design specification into more of a capability-based contract could produce more efficient and higher quality products.

As for implementation (formulation) capacities, the research is fairly anemic. These responsibilities vary but generally fall under "setting a fair bidding process, identifying the best-fit contractor, and reaching an excellent contract" (Yang et al., 2009, p. 683). Yang et al. find that an increase in formulation capacity does not affect contract performance. This is speculation, but it could be because much of this is process dependent, and much of where the human capital of capacity comes into effect is in the agenda setting stage, where a high-level of skill and expertise is needed, whereas implementation capacity, the stage for creating the actual final contract, could be most affected by personnel numbers. One requires a few highly capable people for negotiation and technical requirements creation, while the act of creating the actual contract could a require a larger number of less skilled workers, but both could have the same overall management capacity. One illuminating point by Fernandez (2009) regarding the system rather than the workers indicates that "ongoing competition between contractors during the implementation phase, rather than ex ante competition during the bidding phase, appears to be the form of competition that improves overall contracting performance" (p. 86).

As for the public-private governance capability, the literature is pretty clear-cut. There is widespread support among the evidence supporting the contract enhancing capacities of government and the private sector working together during the contracting period to increase the quality of the service. Speaking to the ability of the government in this respect, Yang et al. provide the most evidence. They use the term "to capture government agencies' active, ongoing involvement in or support for the contractor's operation. The core question involved here is, how can we help the contractor succeed?" It seems that the relationship is n-shaped, supporting the argument that, after reaching a point, extensive implementation activities may engender too much cost, red tape, and X inefficiencies. However, the function also shows that time has a magnifying effect. The impact of implementation capacity on cost reduction turns positive after the interaction efforts gain steam, and then, the impact of implementation activities on efficiency increases accelerates in that the benefits multiply as time passes, indicating that efforts to develop collaboration and mutual support will have long-term advantages (Yang et al., 2009, p. 692). Their results suggest that the government's ability to influence mechanisms such as trust, parallel expectations, and joint action hold significant potential to improve contracting performance. Additionally, Fernandez

(2009) finds evidence that joint problem-solving efforts are positively correlated with overall contracting performance.

In the last bin we have evaluation or monitoring capacity, which is the ability of the government to monitor a contractor's performance and/or enforce the contract. Yang et al. (2009) suggest that the impact of the evaluation capacity depends on its strength: a strong evaluation system is beneficial, but a weak evaluation system does no good. They also show that a "strong evaluation capacity may promote cost reduction and efficiency increases but may not help improve quality" (Yang et al., 2009, p. 691). However, they also find that the benefits of evaluation activities decrease over time. This suggests that a contract needs more evaluation in the beginning, but that once the expectations are clearly established, things run much more smoothly. On the other hand, Fernandez (2007, 2009) do not show any significance of the impact of monitoring activities in either scope or intensity while the enforcement mechanism is mixed. Fernandez's (2007) findings indicate that the most successful contractual relationships perform at higher levels when public managers make periodic use of the "stick" to enforce the contract.

> In fact, among the high performing cases, tactics such as imposing financial penalties and threatening to terminate the contract seem to enhance contracting performance more than alternative means for resolving disputes, such as negotiation and mediation, since the coefficient for reliance on alternative means for resolving disputes is not statistically significant. (Fernandez, 2007, p. 1135)

Then, Fernandez (2009) finds the complete opposite for services for the exact same dataset. This would seem completely contradictory, but Fernandez (2007) utilized SWLS to identify the top performers and then conducted an OLS analysis of the whole sample and a WLS analysis with the weights going to the high-performers. Negotiation and mediation seem to work for the overall sample, but when compared to the most successful contracts, legal enforcement and threats win the day. This heavily indicates the need to differentiate between the different types of contracts.

As for services specifically, much of the earlier literature evaluates service contracting as augmenting management capacity. Service delivery contracting includes producing the service but can also include delegating to vendors management responsibilities, such as monitoring outcomes.

> All service delivery management need not occur within government, though effective contracting clearly requires that governments maintain some management capacity. For example, even though governments can transfer some monitoring responsibilities to vendors, they likely still need to monitor their vendors' performance to some degree. (Brown & Potoski, 2006, p. 324)

Alternatively, by contracting for management responsibilities contracts and introducing third party verifiers, governments may accumulate more monitoring than they would have been able to conduct on their own (Brown & Potoski, 2006) In the case of easy-to-measure services, contract managers can focus more on outcome monitoring and less on the actual production of the service. In such cases, external monitoring becomes an attractive option, contract managers can quickly check the vendors' intensive reports against their own outcome observations. Unfortunately, these cases are only available when services are easily monitorable with specific requirements. Otherwise, delegating complex monitoring to the vendor is obviously easily susceptible to the agent opportunism problem (Brown & Potoski, 2006).

Lastly, it is important to make the distinction between in-house management capacity and overall government capacity. While some cases of governance “may indeed see the abdication of management responsibilities, cutting management staff and activities does not necessarily translate into reducing management capacity” (Brown & Potoski, 2006, p. 325). Earlier literature indicated that government outsourcing the capacity to deliver the service diminished their direct capacity to manage the service. Yet, governments can, in fact “buy” management services to adequately address their own deficits in capacity (Brown & Potoski, 2006, p. 324). This is corroborated in GAO reports of the defense acquisition workforce. For example, at the national level, many federal agencies now employ third-party evaluators to assess the quality of production activities for which they have contracted (e.g., information technology), a practice often referred to as “independent verification and validation.” Therefore, while the government may have reduced their in-house capacity, the overall capacity remains the same or even increases at lower costs. The tradeoff is that contracting and other forms of alternative service delivery do not eliminate the need for management capacity, but instead create an imperative for new types of management capacities. These problems may be more likely to occur in cases such as:

- Limited or no competition among potential vendors
- Contracted products and services that are difficult to specify and describe in written contracts
- Vendors that have special knowledge or skills about the product that is unavailable to public managers
- Public managers that have a hard time monitoring vendor performance once the contract has been let. (Brown & Potoski, 2003, p. 154)

Trust/Relationships

As has been mentioned above in the capacity responsibilities, trust, joint problem-solving, and public-private partnerships have a huge impact on contracting performance. The earlier literature showed theoretical divides on the principal-agent problem and how government and the private sector should behave. Johnston and Romzek (1999) conclude that the agent’s (contractor’s) response to the principal’s monitoring system depends on many factors such as the reliability and credibility of the system as well as the principal’s willingness to enforce punishment. This game theory understanding of the principal-agent problem is complicated in government contracting as elected officials and networks of contractors add multiple layers of accountability. Additionally, “scholars have found that the overreliance on legal means of conflict resolution may evoke conflict, opportunism, and defensive behavior” (Yang et al., 2009, p. 686). Brown and Potoski (2006) provide evidence that longer contracts may also begin to mirror monopoly relationships, exposing governments to the risk that vendors will shirk their responsibilities. “Governments that entered into longer contracts spot checked vendor performance only 68 times a year on average, compared to 95 times a year, on average, for governments operating with short term contracts” (Brown & Potoski, 2006, p. 336)

As the literature matured and developed ways to measure the relationship of the government and contractors, the tone adapted. Fernandez (2007) found that the effect of joint problem-solving on contracting performance is greater among the most successful contractual relationships than in the average case. Since contract managers work more closely with the contractor’s staff to solve issues that arise, the level of contracting performance tends to increase. In a later study, Fernandez found that trust has a positive independent effect on overall contracting performance and the largest coefficient in his model (Fernandez, 2009). However, some of the literature on trust suggests the possibility of an endogenous relationship between trust and performance (Fernandez, 2009, p. 86). Fernandez then conducted a 2SLS regression to account for endogeneity. He determined that contract duration does not appear to interact with trust. That is, “the effect of trust on contracting performance does not increase as the relationship evolves over time” (Fernandez, 2009, p. 87). Additionally, he discovered that monitoring activities and trust do not serve as substitutes. Going deeper into the model as it is of great interest to this study,

> Factor analysis was used to develop multi-dimensional measures of communication, and joint problem-solving efforts after contract award. Since contracts of longer duration can facilitate learning and allow the parties more time to iron out the kinks in service delivery, the model also includes a measure of the duration of the contract, measured in months. (Fernandez, 2007, p. 1127)

Finally, there are some scattered findings throughout the literature on the effect of management responsibilities that can affect relationships. When contracts specify in great detail how a service should be delivered, the contractor may have less incentive to innovate. Additionally, “lengthy negotiations can damage the relationship between partners and inhibit their adaptation to unanticipated situations” (Yang et al., 2009, p. 686). Yang et al. argue that “information searching, contract negotiation, and contract writing” give rise to transaction costs that can offset their cost-saving benefits, and that overuse of contracts for enforcement can curtail the development of trust and collaboration (Yang et al., 2009, p. 690). As time goes on, what is more important is to develop authentic partnerships between the government and the contractor so that information can be shared and collaboration achieved (Yang et al., 2009, p. 693) As for efficiency, the overuse of contracts for enforcement may decrease efficiency, but, developing authentic partnerships during the implementation phase counteracts this, and the effect increases over time. Yang et al. go so far as to suggest that “the best contracting strategy for government is to depart from pure contracting and shift to a collaborative model such as public-private partnerships” (Yang et al., 2009, p. 692). Mentioning sub-relationships, Fernandez (2007) also examined the use of subcontractors because “arrangements involving multiple subcontractors imposes additional burdens on the prime contractor, including higher coordination costs, the likelihood of delays, and sometimes even conflict over the choice of goals and means, all of which ultimately weaken performance” (Fernandez, 2007, p. 1129). The use of multiple subcontractors was significant in the overall OLS sample but was not significant in the case of high performers. This indicates once again that it is paramount to find the distinctions between types of contracts as different types of contracts may have different mechanisms to develop trust.

Conceptual Framework and Hypothesis

This paper posits and tests a conceptual argument linking three categories of characteristics with services contract performance: first, service contract complexity; second, management capacity on the part of the buyer; and third, the strength of the relationship between the buyer and the contractor. By specifying all three characteristics, the argument captures the inherit challenges of services contracting, those most under the control of the buyer, and those most of interest to individual vendors. FPDS does not contain direct measures of these variables, and so the paper introduces proxies for each under the relevant hypothesis.

Service Complexity

The complexity of the underlying service can introduce challenges in two broad ways. First, it may raise the technical expertise required from acquisition officials. A simple service, such as lawn mowing, can be easily specified and overseen while a more complicated service, like maintaining aircraft, requires a higher level of understanding and assurance, as important problems might not be immediately visible. The second aspect of complexity is the challenge of specifying the service in clear and comprehensive terms. When acquiring new services or ones that otherwise involve significant uncertainty, acquisition officials and contractors cannot simply rely on the initial performance work statement to deliver a successful outcome but will have to flexibly incorporate changing conditions or new information. This greater requirement for partnership asks more of both buyer and vendor and leaves much room for disagreement and conflicting interest. In both cases, this complexity makes the work more demanding and thus, all else equal, raises the risks of negative contracting outcomes.

> H_1: As service complexity increases (decreases), the likelihood of cost ceiling breaches and terminations increases (decreases) and the likelihood of exercised options decreases (increases).

The paper employs two labor-based measures to attempt to capture service complexity. Service contracting inherently emphasizes labor and measures of pay, and number of employees is a metric that can be relevant across disparate forms of services contracting.

The first measure is the average salary for the North American Industrial Classification System (NAICS) detailed industry that the contract is classified under. Higher salaries may have multiple sources, but one of them is the difficulty of the work and the experience and education required.

> H_{1A}: As average salary increases (decreases), the likelihood of cost ceiling breaches and terminations increases (decreases) and the likelihood of exercised options decreases (increases).

The second measure is more services contracting specific: average cost per employee. At this stage of the research, the average cost is calculated based on averages for the given product or service code, though the study team hopes to incorporate direct contract-level measures where available in future iterations. It employs an existing government metric, called the invoice rate, that approximates how much the government is charged annually for each comparable full-time employee supporting the service contracts. A services contract with a large number of lower-paid staff would have a lower invoice rate, while one that employed a small number of experts or that had extensive capital costs would have a higher invoice rate. Similarly, a service contract that was just making contracting personnel directly available to the buyer in government facilities and using government equipment would, all else equal, have a lower invoice rate than a than one that also promised a full package of services and charged overhead for the infrastructure in place to help deliver them. As with average salary, this hypothesis assumes that scarcer labor or labor acquired at a greater premium, all else equal, indicates a more complex service.

> H_{1B}: As average cost increase (decreases), the likelihood of cost ceiling breaches and terminations decrease (increases) and the likelihood of exercised options increase (decreases).

Contract Management Capacity

Contract management capacity can manifest in a variety of forms, including assessment, contract formulation capacity, evaluation, and ability to sustain a public-private partnership. The literature affirmed the importance of this capacity, in particular for the more complex services discussed for H_1.

> H_2: As a contracting office's contract management capacity increases (decreases), the likelihood of cost ceiling breaches and terminations decreases (increases).

The first measure considered is the only one where FPDS reports on one of the capabilities discussed in the literature review: performance-based services acquisition (PBSA). Defined in FAR 37.601, PBSA tracks multiple measures relevant to public-private partnership governance including the foundation of how the contract is defined. A performance-based services acquisition "describes the requirements in terms of results required rather than the methods of performance of the work" (GSA Federal Procurement Data System, 2017, p. 52). Other characters included measurable performances standards, plans for monitoring, and the potential for monetary adjustments depending on the quality of the output.

> H_{2A}: As contract office usage of performance-based services acquisition increases (decreases), the likelihood of cost ceiling breaches and terminations decreases (increases) and the likelihood of exercised options increases (decreases).

For the other forms of management capacity, specific measures employed by prior surveys and case studies are not available within FPDS, and headcount data for contracting officers is not publicly available at the contracting office level. To capture this important but elusive variable, this paper employs a measure that scales based on the contracting office's history. This approach assumes that the throughput with a given type of product or service code correlates with the development of technical expertise. As the prior section covered, complexity and expertise requirements can vary greatly from one category to another, and a contracting office may have high capacity in one area that would not translate to a new area.

> H_{2B}: As the share of contracting office obligations for a given service code increases (decreases), the likelihood of cost ceiling breaches and terminations decreases (increases) and the exercised options decrease (increase) for that service.

Extent of Prior Relationship

The importance of partnership, trust, and handling difficult problems and uncertainty together naturally lead into the last characteristic: the relationship between the contractor and buyer. The literature suggests that a perfectly written contract is no guarantee of, nor substitute to, effective collaboration. In the absence of data directly on trust, this hypothesis focuses on the level of interaction that provides the opportunity to build a deeper relationship.

> H_3: As the extent of the government's prior relationship with its vendor increases (decreases), the likelihood of cost ceiling breaches and terminations for that partnership decreases (increases).

The first measure is the number of past years of the relationship between the contracting office and the contractors with a single transaction in a given fiscal year enough to qualify. The second measure is the number of actions on the vendors contracts with that office in the prior year. Contract action counts vary wildly from contract to contract, but even if the obligated amount per action is small, they still represent more opportunities for interaction for the office and contractor.

H_{3A}: As the number of past years of a vendor has contracted with an office increases (decreases), the likelihood of cost ceiling breaches and terminations for that partnership decreases (increases).

H_{3B}: As the number of contract actions a vendor has performed for an office in the past year increases (decreases), the likelihood of cost ceiling breaches and terminations for that partnership decreases (increases).

Data and Methods

Data Sources and Structure

The primary source of this paper is FPDS, which is the transaction database for U.S. government contracts, including military and civilian as well as products and services. With some exclusions, such as classified contracts, the U.S. postal services, and the Defense Commissary Agency, U.S. federal government contracts above a $3,500 threshold are reported into FPDS. Services contracts are delineated using the product or service codes including in FPDS, and include R&D contracting for the purposes of this report. The study team maintains their own copy of the FPDS, which has been supplemented by the ad hoc search tool and information from various data dictionaries. This and past contract datasets are freely available for download for other researchers.

FPDS data has been supplemented using the Services Contract Inventory mandated by the 2010 Consolidated Appropriations Act (GSA, n.d.). The study team continues working on importing and matching contracts from both the civilian agency data held by the GSA and the separate DoD dataset. At this stage in the research, the analysis relies not on the contract inventory itself, which is only available for larger contracts in the first place, but on the invoice rates derived for Product and Service Codes through the work of the U.S. Army. Those invoice rates are used on an annual basis to estimate the number of comparative full-time employees for contracts in the inventory that lack more detailed data. They are broken out for both Overseas Contingency Operations, which are of special interest because they imply coverage of contractors supporting military operations overseas including those directly present in Iraq and Afghanistan.

This report uses a unit of analysis of individual service contracts and task orders. These are identified in FPDS through the unique combination of a procurement identifier and, for task orders, a parent procurement identifier. The dataset is made up of completed contracts and task orders for services contracts for the DoD, completed between fiscal years 2008 and 2015.[1] Many of the variables in the dataset have been built up and tuned over three CSIS reports on Fixed-Price contracts, industrial consolidation and competition, and crisis-funded contracts (Hunter & Sanders, 2019; Hunter et al., 2019; Sanders & Hunter, 2017). Services contracts are less numerically prevalent than their products counterparts but still constitute 1.3 million contracts and task orders. At this stage of the research, 24.2% by count and 21.0% by values obligated are eliminated from the sample because of missing data. The study team believes that recent upgrades to USAspending.gov may enable a reduction in this missing data rate.

[1] Completion is measured by having surpassed the current completion date of the contract or task order by at least one year or by contract close out or a partial or complete contract termination.

The exercised options outcome variable focuses on a narrower subset of contracts and task orders, namely those with unexercised options as of their initial transaction. This reduces the count tenfold, only 103,000 contracts and task orders qualify. However, excluding these contracts from the options exercised sample is important because the choice of whether or not to include options in a contract is a contract formulation decision and not a direct reflection of performance on a given contract. The importance of contracts and task orders with options is affirmed by their value, they account for 23.7% of the total services dataset. Their missing data rate is similar to the overall dataset with a reversal between the metrics, data is missing for 20.1% of contracts and task orders by count and 23.7% by value. Henceforth in this study, for simplicity, both contract awards and task orders will be referred to simply as contracts, except in those cases where the distinction matters.

Measures of Dependent and Independent Variables

This paper focuses on the new dependent variable, options exercised, though the final report of this study will include all three variables.

Dependent Variables

Terminations evaluates whether a contract has experienced a partial or complete termination at any point in its lifespan. This includes terminations for default and convenience (partial or complete) as well as terminations for cause and legal contract cancellations. Perhaps unintuitively, this can include both a traditional cancellation of a major weapon system and the cancellation and reassignment of a contract due to a bid protest.

Ceiling Breaches is a measure that attempts to track the risk cost increases. It tracks whether a contract's cost ceiling has increased as part of a change order or definitize change order. This measure focuses on change orders, rather than modifications for additional work, because the combination of a change order and an increase in ceiling suggests an unanticipated development that will cost the acquirer more money. As shown in Table 1 both ceiling breaches and terminations are rare, though contracts experiencing ceiling breaches account for a bit under a fifth of all contract obligations. Perhaps surprisingly, overlap between these variables is small.

Table 1. Dependent Variables

Variable	Value	%of records	% of $s
Ceiling Breach	0 (None)	99.0%	81.4%
	1 (Ceiling Breach)	1.0%	18.6%
Terminations	0 (Unterminated)	99.1%	97.1%
	1 (Partial or Complete Termination)	0.9%	2.9%

Exercised Options, in contrast to the other two metrics, is a positive measure of contract performance. They reflect that the buyer has chosen to acquire additional services within the scope of the original contract and is willing to pay a higher price as a result. One common source of options is multiple year contracts where the original "base" contract only covers the first year. Both government and contractor may assume that this extension will take place with a high degree of confidence, but in strictly legalistic terms the buyer is under no obligation to continue and may unilaterally allow the contract to end without the liability that may be incurred in a termination.

A transaction only qualifies as an exercised option if it meets all of the three following criteria:

1. The reason for modification is an exercised option, a supplemental agreement for work within scope, or a funding only action.
2. The base and *exercised* options value of the contract increases as part of the transaction.
3. The base and *all* options value of the contract does not increase as part of the transaction.

The study team used this conservative definition in order to ensure that exercised options were clearly differentiated from cost overruns. The metric is calculated in obligated dollars as follows.

$$\boldsymbol{Exercised\ Options\ (Base = 1)} = 1 + \frac{\text{Change in Base and Exercised Options Value due to qualifying modifications}}{\text{Base and Exercised Options Value of the unmodified contract}}$$

Taking a simple example, imagine a five-year contract with a ceiling of $50,000 that starts with a base of one year and $10,000. If no options were ever exercised the value of metric would be steady at 1 (1+$0/$10,000). If a $10,000 second year option was exercised, then then the metric would rise to 2 (1+$10,000/$10,000). If all four options were exercised, then the metric would rise to 5 (1+$40,000/$10,000). The variable is logged, but not centered because it is an outcome variable.

Variable Name	Min	Max	Median	Geometric Mean	1 Unit Below	1 Unit Above	% of Records NA	% of Obligation to NA Records
Exercised Options	1	58,837,341	1	1.047	0.642*	1.706	0.287%	0.418%

* 1 Unit below value below minimum value.

Study Independent Variables

Service Contract Complexity

Average Salary: Each contract in FPDS is labeled by its NAICS Detailed Industry category, the most granular level available. The U.S. Economic Census provides enough data to calculate average wage, although it is only available every five years and thus has a variable lag based of one to five years based on the time since the last census.

Invoice Rate: What is the average annual charge rate for comparable full-time employees. The invoice rate is available through the Service Contract Inventory and is dependent on U.S. Army calculations at the individual Product or Service Code level or for the broad service category.[2] When the invoice rate for a specific product or service code is available for the prior fiscal year, that factor is used. When the invoice rate is available for a code but not for the prior year, the average across all years is imputed. For those codes with no reported invoice rates, the broad service code is used instead for that year if available, and an average of the invoice rate for all available years is used otherwise.

Table 2 shows the descriptive statistics for these variables, which are logged and rescaled in the model.

Table 2. Average Salary and Invoice Rate

Variable Name	Min	Max	Median	Geometric Mean	1 Unit Below	1 Unit Above	% of Records NA	% of Obligation to NA Records
Average Salary	$8,690	$24,7576	$54,192	$50,890	$22,358	$115,834	2.83%	0.899%
Invoice Rate	$9,710	$1,762,137	$170,918	$167,767	$63,862.	$440,726.2	7.53%	13.0%

Contract Management Capacity

Partnership (Lagged): What share of office obligations for a given office were for Performance Based Services Contracting in the prior year.

Service Experience: For any given contract, what percentage of obligations for the office went to contracts with the same product or service code over the past seven years. Table 3 shows the descriptive statistics for these variables, which are rescaled in the model.

Table 3. Partnership and Service Experience

Variable Name	Min	Max	Median	Arithmetic Mean	1 Unit Below	1 Unit Above	% of Records NA	% of Obligation to NA Records
Partnership	0%	100%	27.8%	33.9%	-26.2%*	93.9%	0.03%	0%
Service Experience	0%	100%	1.9%	14.0%	-37.9% *	62.7%	0.03%	0%
Paired Years	0	7	4	3.49	-1.35*	8.33*	0.147%	0.344%

* 1 unit below values are less than minimal value for variable.

Extent of Prior Relationship

Paired Years: For any given contract's vendor and office pairing, how many of the past seven years involved interaction between the vendor and the office. For a new relationship, this value would be zero. Table 43 shows the descriptive statics for this variable, which is rescaled in the model.

[2] Product or Service Codes have four characters. Services codes start with a letter, while product codes start with a number. The broad services category (e.g., the letter Y for construction or the letter D for automated data processing) refers to the first letter of the services code.

Paired Actions (Lagged): For any given contract's vendor and office pairing, how many contracting actions did the vendor perform for that office across all contracts in the prior year. Table 43 shows the descriptive statistics for this variable, which is incremented by 1 to make zeros eligible for logarithmic transformation and is then logged and rescaled.

Table 43. Paired Actions

Variable Name	Min	Max	Median	Geometric Mean	1 Unit Below	1 Unit Above	% of Records NA	% of Obligation to NA Records
Paired Actions	1**	7,806,579	26	32.751	0.5 *	2,249	0.032%	0.013%

* 1 unit below values are less than minimal value for variable. ** True minimum value is 0.

Empirical Approach

At this stage of the research, the study team has created the exercise options models for each of the independent variables. The study used ordinary least squares regression to analyze the logged proportion of growth in exercised options compared to the base. The additional models for the other two dependent variables, ceiling breaches and terminations, will be added at a subsequent stage of the study but will use a broadly similar structure. Those additional models are presently intended to use maximum likelihood logit analysis, as they are presently structured as binary variables.

For all of these models, the study team captures the residual differences between the contracting office and agencies, the detailed industries and sub-sectors, and the countries of performances through the use of multilevel modeling techniques. This approach adopts techniques employed by Gelman and Hill (2017) and Sommet and Morselli (2017) that allow for a different intercept for each of the hierarchical industrial sectors, customers, and places of performance. These are referred to as level 2 and 3 variables, with the level 2 variables, office, and detailed NAICS6 nested under the level 3 variables, agency, and NAICS3. The five multilevel groupings employed in this model are shown in Table 5.

Table 5.. Level 2 and Level 3 Variables Included in the Model

Name	Level	Type	Description
NAICS3	3	Categorical	Subsector Code with 108 groups for services contracts within the study period.
NAICS6	2	Categorical	Detailed Industry Code with 1m069 groups for services contracts within the study period.
Agency	3	Categorical	Contracting Agency Code with 24 groups for services contracts within the study period.
Office	2	Categorical	Contracting Office Code with 1,185 groups for services contracts within the study period.
Place	2	Categorical	Country in which the contract was performed with 198 groups for services contracts within the study period.

The more traditional level 1 inputs, in addition to the study variables discussed in the prior section, included three varieties of inputs as controls. The first category focus on the pairing of the contract's vendor and office and are new to this paper.

- Office and vendor-office pair variables:
  - Office Volume: Total office obligations in the prior seven fiscal years.
  - Office Count: Number of distinct contracts and task orders the office managed in the prior fiscal year.
  - Market Share: What percentage of an office's obligations are accounted for by this vendor. This can be driven by a multitude of factors, including

general vendor success and size regardless of the relationship with a given office. A high value in this variable may reflect vendor lock.

- Subsector-level and detailed industry variables:
    - Both levels included the Herfindahl-Hirschman Index measures that consolidation measures.
    - Both levels also include the ratio of total defense obligations to U.S.-wide revenues.
    - In addition, the defense obligations for each detailed industry are added to the model.
- Contract-level variables:
    - Scope as measured by initial contract ceiling and duration.
    - Competition, which has a baseline of no competition and three alternatives: competition, available for competition but receiving only 1 offer, 2–4 offers, 5+ offers.
    - Vehicle, which has a baseline of definitive contracts and purchase orders, but also includes four types of indefinite delivery vehicles: Single-Award IDCs (S-IDC); Multi-award IDCs (M-IDCs); Federal Supply Schedule or Government-Wide Acquisition Contract (FSS-GWAC); and Blank Purchase Agreement or Basic Ordering Agreement (BPA-BOA).
    - Pricing, which uses firm-fixed price as a baseline with six alternatives handled by dummy variables: incentive fee contracts (whether fixed price or cost-based), combination; combination or other contracts which include multiple types, time and materials, labor hours, or fixed price: level of effort (T&M/LH/FP:LoE); other fixed price (other FP) including all types of fixed price not covered by earlier categories; whether the contract began as an undefinitized contract award (UCA); and other cost-based (other CB) covering all types of cost-based contracts not covered by earlier categories.
    - Crisis Funding: Baseline of drawing from non-emergency accounts with three alternatives for OCO, disaster response, and the Recovery Act (ARRA).

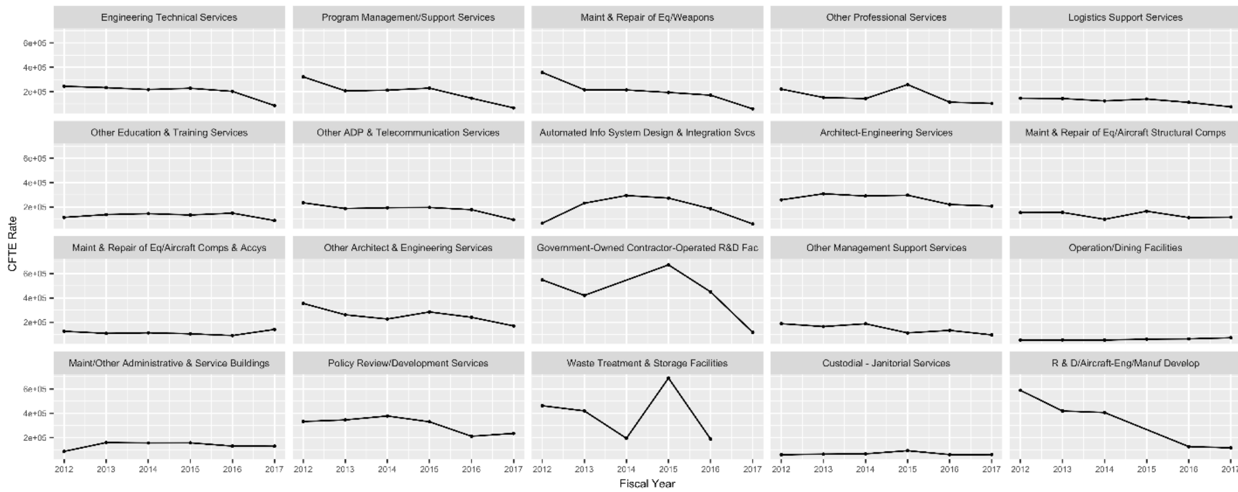
Preliminary Results and Discussion

Inventory of Contracts Services

The Inventory of Contracted Services (ICS) is mandated across the federal government and has an obvious value to this project above and beyond the inclusion of the invoice rate variable. By statute, the DoD has a separate inventory process from the GSA process that includes extracting contract data from the Federal Procurement Data System (FPDS) and the System for Award Management (SAM). The study team analyzed ICS data from both the DoD and the GSA to better understand service contract complexity and found that each source has its own set of challenges. Generally, since FY 2012, DoD ICS data includes comparable contractor full-time equivalents (CFTE) related information with clear ICS guidance and information available. GSA ICS data by comparison is easier to import because it is not spread across many, somewhat inconsistent, Excel tabs. However, GSA ICS relies on supplemental documents for explanation and has not published these supporting documents at all for some years since the start of the ICS, posing difficulties in cross-checking and reference.

According to CSIS data processing methods, three-stage standardization was applied to the raw data from government websites. The main challenge along the procedure was the inconsistent format in which the data was reported and published, which potentially complicated the consolidation process before import to CSIS database system, especially for validating data type, generating unique identifiers, etc. Additionally, inadequate explanation of the certain amount of missing values hinders the utility for further analysis.

Figure 5. **Comparable Full-Time Employee Invoice Rate of Top 20 Prime Service Code (Ordered Horizontally by Invoice Amount)**



Nonetheless, the study team was successful in importing and using key DoD guidance documents. One key piece of the results is shown Figure 5 where the invoice rates for top 20 service codes are shown, listed by volume. The broad patterns align with expectations: “Operation/Dining Facilities” and “Custodial–Janitorial Services” have the lowest invoice rates. By comparison, more complex services like “Government-Owned Contractor Operated R&D Facilities” and “Architect-Engineering Services” are fairly high. However, this investigation did reveal oddities in the later years, with multiple categories suddenly declining in 2016 and with some categories collapsing their rate in 2017. Sometimes the end of a single large contract can do a great deal to explain fluctuations, as the study team found with Waste Treatment & Storage Facilities, but that explanation did not hold in other cases. Because this variable is defined to use lagged data, the 2016 and 2017 invoice rates are not included in any of the statistics; that said, the study team intends a closer examination of this issue as one of the next steps of this project.

Bivariate Analysis of Study Variables

The first step in model building is to look at the relationship of each of the study variables to the output variables. Table 6 shows the minimal logit model, not yet including controls or multiple levels, for the six variables covered in the hypotheses. Each variable is examined one-on-one followed by Model 7 which includes all six variables. When analyzing these coefficients, the greater the magnitude of the coefficient, the greater the estimated risk of ceiling breaches.

Table 6. Logit Bivariate Look at Study Variables and Ceiling Breaches

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
(Intercept)	**-2.83** (0.01)***	**-2.83** (0.01)***	**-2.83** (0.01)***	**-2.88** (0.01)***	**-2.83** (0.01)***	**-2.82** (0.01)***	**-2.89** (0.01)***
Services Complexity							
Log(Det. Ind. Salary)	**0.12** (0.02)***						-0.03 (0.02)
Log(Service Invoice Rate)		**0.15** (0.02)***					**0.16** (0.02)***
Office Capacity							
Office Perf.-Based %			**0.14** (0.02)***				**-0.16** (0.02)***
Office Service Exp. %				**0.66** (0.01)***			**0.72** (0.02)***
Past Relationship							
Paired Years					**-0.10** (0.02)***		**-0.20** (0.02)***
Log(Paired Actions)						**0.29** (0.02)***	0.03 (0.02)
AIC	107637.39	107605.76	107621.00	105650.20	107646.22	107430.81	105389.27
BIC	107658.25	107626.62	107641.86	105671.06	107667.08	107451.67	105462.28
Log Likelihood	-53816.69	-53800.88	-53808.50	-52823.10	-53821.11	-53713.41	-52687.64
Deviance	107633.39	107601.76	107617.00	105646.20	107642.22	107426.81	105375.27
Num. obs.	250000	250000	250000	250000	250000	250000	250000

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$. Numerical inputs are rescaled.

When considered alone, both measures of service complexity match H_1 as higher salaries and invoice rates estimate a higher risk of ceiling breaches. However, when both variables are included in the same model, only the Service Invoice Rate remains significant and the direction of the relationship for average salary flips. For Office capacity, the individual results do not support $H_{2,}$ as both the percent of performance-based services an office performs and the office share of experience with a given service predict a higher rate of ceiling breaches. When the variables are combined, H_{2A} is supported but H_{2B} is not. A look at summary statistics for Performance-Based experience did find that as the percent of performance-based service went from 0% to 75%, the ceiling breach rate declined. Above 75%, it rose dramatically, suggesting an additional variable may influence that relationship. The results were also mixed for $H_{3,}$ as more paired years of history was associated with fewer breaches while more contracting actions between the office and the vendor in the prior year was associated with more. When all the study variables were included, the results for H_{3A} remained significant and the results for H_{3B} lost significance.

The next output variable is terminations and the bivariate results are shown in Table 7.

Table 7. Logit Bivariate Look at Study Variables and Terminations

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
(Intercept)	**-4.04** (0.02)***	**-4.04** (0.02)***	**-4.04** (0.02)***	**-4.06** (0.02)***	**-4.04** (0.02)***	**-4.04** (0.02)***	**-4.06** (0.02)***
Services Complexity							
Log(Det. Ind. Salary)	**-0.12** (0.03)***						**-0.25** (0.04)***
Log(Service Invoice Rate)		-0.03 (0.03)					0.06 (0.03)$^{\cdot}$
Office Capacity							
Office Perf.-Based %			**0.17** (0.03)***				0.00 (0.03)
Office Service Exp. %				**0.41** (0.03)***			**0.28** (0.03)***
Past Relationship							
Paired Years					-0.03 (0.03)		**-0.25** (0.04)***
Log(Paired Actions)						**0.50** (0.03)***	**0.51** (0.04)***
AIC	43737.75	43751.99	43723.28	43523.03	43751.85	43508.97	43317.42
BIC	43758.60	43772.85	43744.14	43543.88	43772.71	43529.83	43390.43
Log Likelihood	-21866.87	-21873.99	-21859.64	-21759.51	-21873.92	-21752.49	-21651.71
Deviance	43733.75	43747.99	43719.28	43519.03	43747.85	43504.97	43303.42
Num. obs.	250000	250000	250000	250000	250000	250000	250000

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$. Numerical inputs are rescaled.

The termination results, with a single exception, do not support the hypotheses or are not significant. For H_1, higher average salaries predict a lower risk of terminations. The relationship with invoice rate does predict a lower risk of terminations once all six variables are included, but is only significant with a p-value <0.1. For H_2 both office capacity variables estimate a higher risk of termination in isolation and the office use of performance-based services contracting has no estimated influence on outcomes once all six variables are included. The one place the hypotheses finds support is H_{3A}, paired years, which estimates a lower risk of termination, but only once all six variables are included. Paired actions is significant in both cases, but estimates a greater, not a lesser, risk of terminations.

The final output variable is options exercised, which involves two important changes in interpretation. First, this model uses regression because it is estimated the amount of growth attributable to options and not simply whether or not any options were exercised. Second, for the prior to variables, negative coefficients are associated with better outcomes, that is, fewer ceiling breaches and terminations. For Table 8, the direction is reversed, as positive values in the coefficient indicate that proportionally larger options are likely to be awarded. Finally, note that the number of contracts in the sample drops dramatically, as contracts with available options are less common, though higher in value on average.

Table 8. Regression Bivariate Look at Study Variables and Log(Options Growth)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
(Intercept)	**0.65** (0.00)***	**0.63** (0.00)***	**0.62** (0.00)***	**0.62** (0.00)***	**0.62** (0.00)***	**0.63** (0.00)***	**0.66** (0.00)***
Services Complexity							
Log(Det. Ind. Salary)	**-0.19** (0.00)***						**-0.14** (0.01)***
Log(Service Invoice Rate)		**-0.21** (0.00)***					**-0.14** (0.01)***
Office Capacity							
Office Perf.-Based %			**-0.06** (0.01)***				**-0.06** (0.01)***
Office Service Exp. %				**0.10** (0.01)***			**0.14** (0.01)***
Past Relationship							
Paired Years					**-0.01** (0.00)**		**-0.04** (0.01)***
Log(Paired Actions)						**0.06** (0.00)***	**0.07** (0.01)***
AIC	165175.62	165387.96	167067.06	166831.95	167195.69	167039.85	163626.85
BIC	165203.59	165415.93	167095.03	166859.92	167223.66	167067.82	163701.43
Log Likelihood	-82584.81	-82690.98	-83530.53	-83412.97	-83594.85	-83516.93	-81805.42
Deviance	35689.57	35781.35	36515.49	36411.79	36572.34	36503.47	35022.97
Num. obs.	82675	82675	82675	82675	82675	82675	82675

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, ·$p < 0.1$. Numerical inputs are rescaled.

The options variable presently offers the strongest support for the hypothesis. As per H_1, both individually and together, higher salaries and invoice rates estimate lower value of exercised options. For H_2, office capacity, the result is split. Greater experience with performance-based services is associated with a lower value of exercised options while office experience with a given service is associated with a greater value in exercised options. The support for H_3 is also split, where paired years are associated with lesser growth in options exercised while paired contracts actions in the prior year are associated with more options.

Next Steps

The study team is presently incorporating a range of controls into all three models. Of particular interest are the controls for contracting office and the office-vendor relationship, which might help clarify the initially contradictory results for H_2 and H_3. The study team is also looking at whether converting to constant dollars, perhaps using the contract start year for deflation purposes, is an option, as temporal effects could be particularly important for invoice rates and salaries, as both tend to rise over time. Additional attention to quality issues in the services contract inventory is also planned, as the information captured by the inventory is of great interest, but the initial analysis and importing process revealed a variety of data quality challenges. The study team is looking with interest whether, after introduction of controls, some hypotheses retain split decisions. For example, more years of paired experience estimates a lower risk of ceiling breaches and terminations, but also a lesser growth in exercised options.

References

Brown, T. L., & Potoski, M. (2003, March 4). Managing contract performance: A transaction costs approach. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/pam.10117

Brown, T., & Potoski, M. (2006, March 2). Contracting for management: Assessing management capacity under alternative service delivery arrangements. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/pam.20175

GSA. (n.d.). *Service contract inventory*. Retrieved from https://www.acquisition.gov/service-contract-inventory

GSA Federal Procurement Data System–Next Generation (FPDS-NG) Data Element Dictionary. (2017, November 10).

Hunter, A. P., & Sanders, G. (2019, April 2). *Avoiding terminations, single-offer competition, and costly changes with fixed-price contracts*. Retrieved from https://www.csis.org/analysis/avoiding-terminations-single-offer-competition-and-costly-changes-fixed-price-contracts

Hunter, A. P., Sanders, G., McCormick, R., Meitiv, A., Nzeribe, G., & McQuade, M. R. (2019, April 2). *Evaluating consolidation and the threat of monopolies within industrial sectors.* Retrieved from https://www.csis.org/analysis/evaluating-consolidation-and-threat-monopolies-within-industrial-sectors

Joaquin, M. E., & Greitens, T. J. (2012, August 23). Contract management capacity breakdown? An analysis of U.S. local governments. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6210.2012.02587.x

McCormick, R., Hunter, A. P., Sanders, G., Cohen, S., & McQuade, M. R. (2015, September 30). *Measuring the outcomes of acquisition reform by major DoD components*. Retrieved from https://www.csis.org/analysis/measuring-outcomes-acquisition-reform-major-dod-components

Romzek, B. S., & Johnston, J. M. (2002, July 1). Effective contract implementation and management: A preliminary model. Retrieved from https://academic.oup.com/jpart/article-abstract/12/3/423/909448?redirectedFrom=fulltext

Sanders, G., & Hunter, A. P. (2017). *Overseas contingency operations contracts after Iraq: Enabling financial management research and transparency through contract labeling*. Retrieved from https://www.researchsymposium.com/conf/app/researchsymposium/unsecured/file/145/SYM-AM-17-051-005_Sanders.pdf

Witesman, E. M., & Fernandez, S. (2013, August). Government contracts with private organizations. Retrieved from https://journals.sagepub.com/doi/abs/10.1177/0899764012442592

Disclaimer

The Center for Strategic and International Studies (CSIS) does not take specific policy positions; accordingly, all views expressed in this presentation should be understood to be solely those of the author(s).

Predicting Bid Protests: What Should Acquisition Teams (Not) Do?

Timothy G. Hawkins, Lt Col, USAF (Ret.)—PhD, is an associate professor in the Department of Marketing, Western Kentucky University, and a National Contract Management Association Fellow. He researches and teaches in the realms of supply chain management, marketing, government contracting, and strategic sourcing. He has 20 years of sourcing experience in industry and government. Dr. Hawkins has published articles on opportunism in buyer–supplier relationships, source selection, services procurement, performance-based logistics, collaborative pricing, and electronic reverse auctions in scholarly publications such as the *Journal of Supply Chain Management*, *Journal of Business Logistics*, *Journal of Purchasing and Supply Management*, *Journal of Business Research*, *International Journal of Logistics Management*, *Journal of Defense Analytics and Logistics*, *Defense Acquisition Research Journal*, *Industrial Marketing Management*, *Journal of Business Ethics*, *Supply Chain Management: An International Journal*, *Journal of Marketing Channels*, *Air Force Journal of Logistics*, *Journal of Contract Management*, *International Journal of Procurement Management*, *Journal of Product and Brand Management*, and *Journal of Public Procurement*. His current research interests include procurement ethics, buyer–supplier relationships, strategic sourcing, services procurement, and supplier performance management. [timothy.hawkins@wku.edu]

Abstract

Bid protests are increasing, and the effectiveness for protestors is relatively high. Bid protests delay receipt of needed goods and services. They are costly to prevent and to adjudicate. The purpose of this research is to better understand why bid protests are lodged by interested parties. This research concentrates on meso-level factors controlled by the acquisition team that affect the receipt of a bid protest, namely, the characteristics of the procurement, acquisition strategy decisions, and human factors. Using an existing data set of 240 government source selections resulting from a survey of U.S. Navy contracting officials, 19 antecedent factors will be explored.

Introduction

A central tenet of a public contracting system is to maintain the public's trust via instilled integrity, fairness, and openness (Hawkins et al., 2016). A bid protest is a corrective mechanism to ensure integrity and fairness by providing an interested party with a process to air complaints and obtain relief (Manuel & Schwartz, 2011). It is a written objection that can occur at any stage of the contract award process. Often, protests result from alleged errors or mistakes committed by the buying agency. The most common errors cited in protests are poorly written or vague contract requirements, failure to follow the process or evaluation criteria laid out in the request for proposals, unequal treatment of offerors, and failure to adequately document the record (GAO, 2014). Said errors can result in unfair discrimination against an offeror, and thus, lost business. Nevertheless, offerors also protest for opportunistic reasons such as to increase revenue, harm competitors, obtain competitive intelligence, prospect for protest viability, and negotiate a subcontract award (Maser & Thompson, 2010).

Bid protests have become a substantial aspect of government procurement (Cibinic et al., 2011). In 2016, 2,621 protests were received by the Government Accountability Office (GAO; 2016), double the number received in 2008 (Arena et al., 2018). This number trended steeply upward from 2007–2011, then levelled. "From FY2008–FY2014 total government spending, adjusted for inflation, decreased 25% while total protests increased 45%" (Schwartz & Manuel, 2015, p. 8). Thus, protests as a percentage of protest opportunities (i.e., awarded contract actions) increased from 0.16% in 2008 to 0.26% in 2016 (Arena et al., 2018). Of those protest cases that made it to a decision from 2009–2014 (i.e., the few that were not dismissed, settled, or withdrawn), only 17% were sustained, but an average of 42% of all protest cases were effective (either sustained or resulted in corrective action taken by the buying agency prior to a decision). The effectiveness rate for 2017 grew to 47% (GAO, 2017).

Acquisition officials and end users loathe the receipt of a bid protest (Hawkins et al, 2016). The potential to receive a bid protest drives agencies to incur transaction costs to (1) prevent a protest by thoroughly documenting and substantiating proposal evaluations and trade-off decisions (Hawkins et al., 2016), (2) defend against an actual protest lodged (NASPO, 2013), and (3) take corrective actions. Responding to a protest requires the agency to generate a statement of facts and a memorandum of law, and to gather all of the pertinent supporting documents such as the solicitation, evaluations, proposals, and so forth, for distribution to the GAO and, in some cases, the protestor's legal counsel (Rumbaugh, 2010). The GAO resolves 70% of cases within 60 days, but consumes 90–100 days resolving the remaining 30%, which are complex cases (Arena et al., 2018). At best, an agency's voluntary corrective action means the competition is reopened, and proposals are allowed to be revised, necessitating further evaluations and delaying the contract award. At worst, an authority such as the GAO or Court of Federal Claims (COFC) sustains the protest, meaning that the procurement process must often start anew. This adds even more time and delays the receipt of needed goods and services, resulting in significant rework. The end users bear costs as well since their requirements are delayed or go unfulfilled. Bid protests are such a persistent concern that the U.S. federal government recently proposed legislation to impose a $350 filing fee to dissuade frivolous protests (Poling, 2016), and the GAO, for the first time ever, temporarily banned a frequent protestor, Latvian Connection, from federal contract awards (Mlinarchik, 2016). Congress took a step further in its Conference Report for the fiscal year 2018 National Defense Authorization Act (NDAA), which included a pilot program to test the effects of an unsuccessful protestor paying the government's protest processing costs. Additionally, federal government agencies (Camm et al., 2012) and Congress (Arena et al., 2018) continue to commission studies to understand and mitigate problems. Furthermore, state governments are not immune to the public's concern for fair tendering; thus, they commissioned research of their own (Molenaar & Tran, 2015).

While some research downplays the impact of protests by emphasizing their relatively rare occurrence (Arena et al., 2018; Gordon, 2013), the buyer's reaction to the bid protest system is to apply extraordinary effort to defend acquisitions against a protest. Measures taken to avoid protests include (1) added layers of reviewers and legal counsel to scrutinize every document (and revision thereto) of the source selection record, (2) added procurement lead time, (3) conducting additional rounds of discussions to allow offerors an opportunity to rectify weaknesses and deficiencies rather than eliminating them from the competitive range, (4) unnecessarily retaining offerors in the competitive range, (5) awarding more contracts than intended, (6) modifying existing contracts rather than conducting full-and-open source selections, (7) shopping requirements to existing contracts for task order awards rather than conducting a full-and-open source selection, (8) utilizing a more objective, price-based source selection method such as LPTA rather than a full trade-off, (9) increasing the size of the acquisition team, and (10) offering more extensive debriefings. Furthermore, practitioners continue to devise procedures to mitigate protests (Curry, 2018; Finkenstadt & Hawkins, 2016). Together, efforts during source selections amount to an average $235,000 of transaction costs each, or 7.7% of the contract value (Hawkins et al., 2016). These burdens and costs are not trivial, which suggests that the bid protest system will continue to be controversial.

Periodically, the GAO publishes a list of common infractions leading to sustained protests. Such micro-level factors include a failure to follow the solicitation evaluation criteria; inadequate documentation of the record; unequal treatment of offerors; and unreasonable price or cost evaluation (GAO, 2014). Certain meso-level systemic characteristics could facilitate these micro-level mistakes. Surprisingly, however, few studies have examined the meso-level factors pertaining to the structure of an acquisition, the context of the procurement, and human factors. One study by Maser and Thompson (2010) found that protests are more likely in cases of (1) more bidders, (2) smaller bidders, (3) a high value of the protested contract as a percentage of the protestor's revenue, (4) contracts with long delivery times (i.e., extended lock-outs), (5) buying services, and (6) international winners. But what other strategy decisions are being made by the acquisition team that contribute to an offeror's decision to protest? Other factors could include the source selection method applied, whether oral presentations are conducted, whether sufficient procurement lead time is allotted, whether discussions were conducted, the size of the acquisition team, and the experience level of personnel involved. Further, do characteristics of the procurement affect an offeror's decision to protest?

In addition to the very practical utility of unveiling factors that may reduce or increase bid protests, perhaps greater value from investigating this line of logic is the extension of inter-organizational justice theory to pre-award supplier selection (i.e., not just pertaining to managing established post-award supplier relations). After all, a bid protest is purportedly a manifestation of a supplier's perceived injustice. Heretofore, justice theory applied to inter-organizational contexts is scant (Liu et al., 2012) and has ignored a challenging stage of supplier relationships— relationship initiation (Dwyer et al., 1987). However, the intersection of justice expectations and a competitive supplier selection presents a "sticky" situation in need of clarity.

This research, backed by quantitative data, seeks to bridge this gap. In doing so, all business-to-business/business-to-government (B2B/B2G) relationships stand to benefit by a better understanding of the specific phenomena leading to more efficient and effective supplier relationship formation (i.e., less perceived injustice and conflict).

Research Questions and Objectives

The purpose of this research is to better understand why bid protests are lodged by interested parties. An objective is to identify various meso-level decisions and actions of buy-side acquisition teams that affect the receipt of a bid protest. Another objective is to seek extensions to inter-organizational justice theory based on the findings. The following research questions (RQ) will be explored:

- RQ1: What characteristics of a procurement affect whether a bid protest is received?
- RQ2: What acquisition strategy variables/decisions affect whether a bid protest is received?
- RQ3: What human factors contribute to receipt of a bid protest?
- RQ4: Are the pertinent theories surrounding inter-organizational exchange complete, and if not, what extensions should be considered?

Research Scope

This research examines only sources of bid protests attributable to buying organizations. The scope excludes examining non-buyer sources of bid protests such as those lodged for reasons other than buyer action or inaction. Allegedly, it is common for businesses to protest a contract award due to business strategy reasons such as to buy more time (i.e., revenue) on a service contract as an incumbent, to gain another chance to secure an otherwise lost business opportunity, or to disadvantage a competitor in some way.

The remainder of this research is organized as follows. First, the relevant literature is surveyed raising a conceptual framework and proposed hypotheses. Next, the the research design and methodology are explained. Then, the study provides an analysis of the proposed models and reports the findings. Lastly, the study offers a summary discussion, offers implications for theory and practice, and concludes with study limitations and logical and useful vectors for future research.

Literature Review

Bid Protest Evolution and Diffusion

The U.S. government's bid protest system evolved as a means to ensure fairness to taxpayers, whose resources should not be wasted, and to suppliers that relied upon the government for business. Its origin traces to the Tucker Act of 1887; thereby, the government waived its sovereign immunity, allowing it to be sued in certain contractual matters (Arena et al., 2018). The U.S. Government Accounting Office was created in 1921 (Arena et al., 2018) and began hearing bid protests shortly thereafter, with the first recorded decision in 1925 (Gordon, 2013). Eventually, the courts also gained jurisdiction to hear protests, currently the Court of Federal Claims. An underlying theory of the bid protest system is equity; private firms should have an equivalent chance to secure government contracts (Arena et al., 2018). For protests filed at the GAO, relief is restricted to *interested parties*—those firms deemed to have direct economic interest (Cibinic et al., 2011) by being in a position for contract award given a sustained protest decision (Edwards, 2006).

Bid protest systems for the deterence and relief of injustice are not unique to the U.S. federal government. Their effectiveness in fostering integrity and fairness is so recognized that protests became part of international trade through forums such as the North American Free Trade Agreement, World Trade Organization, the United Nations Commission on International Trade Law, and the European Union (Gordon, 2013). Nothwithstanding, most U.S. state governments allow for administrative bid protests without having to resort directly to a lawsuit (NASPO, 2013).

Justice Theory

Because of its importance, justice is receiving increased academic attention (Kaynak et al., 2015). Perceived (in)justice affects key outcomes such as trust, satisfaction, commitment, and unethical behaviors (Greenberg, 1990) and has been positively associated with alliance profitability (Beugre & Acar, 2008). Of the three dimensions of justice, distributive justice dominated early work. Distributive justice represents an individual's assessment of the distribution of outcomes (Gilliland, 1993). Interested parties often seek to ensure that outcomes are distributed among the parties fairly. Commonly, the basis of those assessments is equity—a comparison of an individual's own *get* versus *give* ratio versus that of a referent. When this investment-to-outcome ratio is approximately equal among parties, justice is perceived, and vice versa. An inequity results in decreased satisfaction and often a search for alternatives.

Similar to findings in organizational theory (Gilliland, 1993; Leventhal, 1980; Thibaut & Walker, 1975), channel members expect to be treated fairly, a dimension referred to as procedural justice. “Procedural fairness is the glue that holds the relationship together” (Kumar, 1996, p. 104). It has been found to be more important than distributive justice (Gilliland, 1993). Procedural justice increases knowledge sharing, continuous commitment, and relationship investment, which, in turn, increase buyer–supplier relationship performance (Liu et al., 2012).

Procedures are seen as just when they include the following six principles: (1) bilateral communication, (2) impartiality (equal opportunity), (3) refutability, (4) explanation, (5) familiarity, and (6) respect (Kumar, 1996). Other important aspects of procedural justice include the following: consistent decisions based on accurate information, consideration of the ethical values of affected individuals, and outcomes that could be modified (Leventhal,1980). A nuance of procedural justice concerns the treatment of affected individuals while enacting a decision—a phenomenon termed *interactional justice* (Bies & Moag, 1986). Not only is the content of a decision important, but so is the way in which it is communicated. Affected people’s justice perceptions are affected by whether they receive an explanation for a decision (i.e., justification), and whether they are respected and not treated rudely (i.e., treated well).

Inter-organizational justice has been defined as “boundary spanners’ perceptions of the fairness of each other’s actions in interorganizational relationships” (Beugre & Acar, 2008, p. 452). Inter-organizational justice during sourcing processes is important due to its effect on relationship continuity (Kaynak et al., 2015). In procurement, justice or fairness has been examined in relation to many essential processes such as supplier selection (Plank et al., 1994), inspection and acceptance (Plank et al., 1994), dispute resolution (Lu et al., 2017), post-award negotiations of changes (Lu et al., 2017), forecast information sharing (Blancero & Ellram, 1997), and supplier performance evaluation (Blancero & Ellram, 1997; Hawkins & Gravier, 2016), to name a few.

A common thread across inter-organizational justice theory and social exchange theory is communication. Most of the aforementioned principles pertain in some way to communication. The theory of channel communication might be instructive (Blancero & Ellram, 1997; Mohr & Sohi, 1995), but pertains to ex post versus ex ante relationship formation. Very little research addresses the essential elements of communication during relationship formation, and particularly the interplay of these communication elements with perceptions of justice. Therefore, the focus here entails supplier selection prior to relationship formation.

Many of the meso-level factors predicting bid protests should focus on the seminal effects of buyer–supplier communication. As such, this research addresses how the structural design of the acquisition process either hinders or facilitates the communication of expectations, explanations of decisions, respect, disagreement, and opportunity. Pertinent factors can be organized as characteristics of the procurement, acquisition strategy components, and human factors.

Characteristics of the Procurement

It has been suggested that when revenue is at stake, incumbents who are unsuccessful offerors on the follow-on contract source selection are likely to protest (Arena et al., 2018). We also know that protests increase as the contract value as a proportion of the offeror’s total revenue increases (Maser & Thompson, 2010). Similarly, requirement criticality represents the level of contribution an acquired good or service makes to the requiring activity’s mission (Kraljic, 1983). When goods and services are critically important, the requiring activity is likely to have a persistent need. This means that not only is the revenue and profit of the current requirement at stake, but so is that of future, repeat procurements. Offerors may protest so as to not lose out on the promise of persistent income. Thus, it is expected that:

> **H1: There will be a positive relationship between dollar value of the proposed contract and receipt of a bid protest.**

> **H2: There will be a positive relationship between criticality of the requirement and receipt of a bid protest.**

Maser and Thompson (2010) found that protests are more likely in cases of procured services versus goods. The more difficult the definition of requirements (i.e., the communication of all expectations and performance levels), the more likely the buyer's evaluation team will misunderstand the proposed value offering. Hence, an overly strict evaluation criterion rating, a weakness, or a deficiency could be undeservingly assigned to the offeror's proposal. Similarly, the more intangible the service or its outcome, the more likely the buyer will omit a specification or inadequately define it for offerors. Thus, offerors may not adequately address a true underlying, yet undescribed, need. The mis-evaluation of poorly or under-specified needs may raise perceptions of procedural injustice. Therefore, it is posited that:

> **H3: The type of value procured will be associated with receipt of a bid protest.**

Protest risk has been found to be a significant predictor of fear of protest (Hawkins et al., 2016). Protest risk represents the product of the probability of receiving a bid protest and the magnitude of the consequences of receiving a protest. As previously discussed, negative consequences could include delayed receipt of needed goods and services, added effort of a source selection team increasing transaction costs, litigation costs such as bid and proposal costs, contract termination for convenience costs, potential shame and embarrassment to the acquisition team, and even adverse personnel action to those committing errors.

Not all acquisitions are equally susceptible to protest. For instance, a 10-year, multi-billion-dollar, unique service contract (e.g., cloud computing or cybersecurity) will have higher odds of being protested than a similar single-year contract due to its dollar amount, duration, and associated compounding reputational effects. Neither are the consequences of a protest the same for each acquisition. For example, redoing an evaluation of three proposals entails less transaction costs than that of 14. Similarly, redoing evaluations involving four evaluation criteria entails less transaction costs than that involving 20. Further, delaying the award of a $5 billion acquisition would likely cost the buyer more than that of a $2,000 acquisition. In terms of justice theory, where the distribution of negative consequences is unbalanced between buyer and seller or between competing offerors, protest risk should increase. Where the product of protest probability and magnitude of consequences is large, a protest is more likely. Thus, it is posited that:

> **H4: There will be a positive relationship between protest risk and receipt of a bid protest.**

Acquisition Strategy Variables/Decisions

Government source selections take time. But, agencies, in their acquisition processes, should not consume too much time, thereby dissuading the best firms from participating in the government market (Edwards, 2006). Sometimes, the allotted procurement lead time is limited in order to receive the goods and services when needed, and sometimes proper advance planning does not occur necessitating expedited sourcing. It is logical that when the myriad of tasks associated with source selection are rushed, mistakes may occur. Likewise, the insufficiency of planned procurement lead time has been found to increase the fear of protest (Hawkins et al., 2016). Ill-suited procurement lead time may signal to offerors that their proposals have not been thoroughly or fairly evaluated or that reasonable and legitimate trade-off decisions have been made and documented; thus, perceptions of procedural justice may suffer. It is thus posited that:

H5: There will be a negative relationship between sufficiency of planned procurement lead time and receipt of a bid protest.

Various methods are available to source selection teams to evaluate offers and choose between them. The three best value methods mentioned in FAR Part 15 include a full trade-off (FT), a price-past performance trade-off (PPT), and the low-price, technically-acceptable (LPTA) method. The FT method allows for trade-offs between price and non-price factors. Hence, using a FT method, a buyer is permitted to pay more for higher performance. In contrast, under a LPTA method, non-price factors are evaluated as acceptable or unacceptable. Once proposals are deemed acceptable on each non-price evaluation criterion, the award decision defaults to the low-price offer. Therefore, a binary rating of acceptable or unacceptable under an LPTA method is, in general, easier to defend than is an ordered-categorical-scale rating (e.g., outstanding, good, acceptable, marginal, unacceptable). Further, making and justifying trade-offs between such categorical ratings and price poses challenges in order to withstand scrutiny. For example, how outstanding does an offer need to be to warrant paying a 5% higher price?

Qualitative evidence suggests that contracting officers believe that their choice of source selection method can affect the receipt of a bid protest, and that this impacted their decision (Arena et al., 2018; Gordon, 2013). The LPTA method, due to its lower subjectivity, is more easily defendable and is less prone to errors than is the FT method. Under an FT method, multiple criteria and multiple evaluators could invite dissonance in evaluations among team members of the meaning of criteria, and could invite the subliminal use of unstated evaluation criteria that, arguably, needed to have been in the solicitation. Evaluations conducted contrary to the process prescribed in the solicitation can raise perceptions of procedural injustice by offerors. In several cases, an LPTA source selection has been used or suggested explicitly as a means of avoiding a bid protest (Pocock, 2009; Schwartz & Manuel, 2015). As such, it is posited that:

H6: There will be a negative relationship between source selection method appropriateness and receipt of a bid protest.

H7: The LPTA source selection method will be negatively associated with receipt of a bid protest.

Bid protests have been associated with socio-economic status (Maser & Thompson, 2010). Small businesses account for most protests at the GAO (53%) and at the COFC (58%) (Arena et al., 2018). Maser and Thompson (2010) posited that small businesses are more likely to protest than are large businesses, and further, that small businesses commonly protest other small businesses' contract awards. Given that protests are related to the procurement's proportion of the offeror's revenue (Maser & Thompson, 2010), this proportion will be higher for small businesses. Thus, it is posited that:

H8: There will be a positive relationship between a small business set-aside and receipt of a bid protest.

The acquisition strategy encompasses the source selection method but is more broad. It also entails such components as the contract type, milestones, team members, team size, evaluation criteria, contract duration, incentives, options, number of contracts, contract line item structure, price and cost analysis method, contract clauses and solicitation provisions, and payments method—to name a few. The extent to which these components of strategy do not fit the procurement could invite errors in the evaluation of proposals. As such, it is posited that:

H9: There will be a negative relationship between acquisition strategy appropriateness and receipt of a bid protest.

In source selection, often all technical evaluators are not involved in the determination of evaluation criteria or in the definition of their meaning. Furthermore, often, technical evaluators are not versed in the nuances of the rules of proposal evaluation and bid protests (Molenaar & Tran, 2015). Criteria that should have been in the solicitation but were omitted, for whatever reason, can by mistake or otherwise, inappropriately creep into the evaluation. A failure to follow the stated evaluation criteria is cited as a leading cause of sustained protests (GAO, 2014). Evaluation comments and proposal critiques that are useful in discriminating between offers can, therefore, be discouraged by review committees and legal counsel (Arena et al., 2018). The extent to which technical evaluator's evaluations are sanitized by reviewers should mitigate procedural injustices, and therefore, protests. Thus,

H10: There will be a negative relationship between compromised technical evaluations and receipt of a bid protest.

Often, source selection teams are rushed by aggressive milestones for contract award. One way to reduce procurement lead time is to bypass discussions (i.e., negotiations or, more often, the resolution of weaknesses and deficiencies in proposals). In order to award a contract without discussions, the contracting officer must notify offerors in the solicitation of the intent to award without discussions, making it a deliberate acquisition strategy decision. Rushing the process and forgoing an opportunity to fully understand each aspect of each proposal might invite errors to the evaluations. Additionally, one aspect of procedural justice is to afford individuals an opportunity to impact the decision process (e.g., proposal evaluations) or offer input (Thibaut & Walker, 1975). Forgoing discussions denies such input. Thus, it is posited that:

H11: There will be a positive relationship between intent to award without discussions and receipt of a bid protest.

On the other hand, discussions entail strict procedural rules ripe for errors. For example, discussing one aspect of a proposal with one offeror and failing to check the same with each other offeror (e.g., past performance reference relevance in terms of type of work, location, or weather) could be a protestable offense (Wallace, 2018). The unequal treatment of offerors was cited as a leading cause of sustained protests (GAO, 2014). If discussions are opened, the procurement becomes substantially more error-prone due to the strict procedures and documentation required. Inadequate documentation is cited as a leading cause of sustained bid protests (GAO, 2014; Wallace, 2018). For this reason, discussions are sometimes avoided by contracting officers (Gordon, 2013). As such,

H12: There will be a positive relationship between conducting discussions and receipt of a bid protest.

Oral presentations constitute the submission of proposal information orally (Edwards, 2006). Oral presentations were codified in the FAR in concert with the rewrite of Part 15 in 1997 as a tool to streamline the source selection process and to improve pre-award communications between offerors and the government (Hannaway, 2000). Oral presentations facilitate communication from the offeror of its understanding of the work, its capabilities (Edwards, 2006), its past performance, and its technical approach (Rumbaugh, 2010). This explanation should enhance evaluators' understanding of the proposals resulting in more accurate evaluations and ratings (e.g., proposal risk). Indeed, explanation and bilateral communication are among the six principles of the theory of justice (Kumar, 1996).

On the other hand, oral presentations add one more step to a complicated evaluation process (i.e., more opportunity to make a mistake). Specifically, entertaining oral presentations without opening discussions means that source selection team members, in their communications, must be careful not to allow an offeror to revise its proposal—even orally (Cibinic et al., 2011; Edwards, 2006). Of course, this requires a perfect knowledge of each element of an offeror's written proposal in order to recognize whether any statement made during an oral presentation constitutes a change to any prior written or oral proposal submission. Obviously, prospective contracts with expansive or complicated scopes of work can render such perfect knowledge untenable. Proposal revisions may inadvertently be made. Consider also that salespeople naturally want to satisfy evaluators (i.e., avoid negative ratings or perceptions of weaknesses); thus, changes to proposals can be difficult to avoid as salespeople can sense evaluators' concerns by either non-verbal cues or by the ensuing line of questioning. Given the aforementioned conflicting arguments to the benefit or harm of an oral presentation, no directional claim is made.

H13: There will be a relationship between the use of oral presentations and receipt of a bid protest.

The GAO (2014) repeatedly cites inadequate documentation of the record as a chief culprit of sustained bid protests. Poor documentation could include contradictions in the records and omissions of details needed to justify ratings and trade-off decisions. Documents relied upon during proposal evaluations include the source selection decision document, comparative analysis of proposals, evaluation notices to offerors, source selection plan, debriefing scripts, technical evaluations, past performance evaluations, cost or price analyses, rating charts, and evaluation briefing charts. Additionally, protest probability has been qualitatively associated with source selection document scrutiny (Arena et al., 2018). The purpose of the scrutiny is to avoid a protest. Thus, logic holds that more revisions reduce errors and thereby lower the chances of receiving a bid protest. Added scrutiny entails often multiple acquisition team members poring over all of the documents to prevent errors such as those cited by the GAO— unequal treatment of offerors and following the evaluation process and criteria per the RFP. As such, it is posited that:

H14: There will be a negative relationship between the number of source selection document revisions and receipt of a bid protest.

In order to appease otherwise unsuccessful offerors and thwart a protest, contracting officers will sometimes award more contracts than planned. In essence, the work gets split among two or more contractors so that there are no losers. For example, building, fielding, and sustaining two varieties of Littoral Combat Ship platforms substantially increased costs relative to doing so for a single platform (O'Rourke, 2014) but mitigated the threat of a protest. Thus,

H15: There will be a negative relationship between increased actual number of contracts awarded versus that intended and receipt of a bid protest.

Qualitative evidence suggests that contracting officers adjust the chosen type of contract to the probability of a protest (Arena et al., 2018). More complicated contract types (e.g., cost reimbursement) entail more complicated cost analyses that are prone to controversy and error (e.g., should-cost analysis). Prior research found that cost plus-type contracts are more likely to be protested (Maser & Thompson, 2010). Thus, it is posited that:

H16: Contract type will be associated with receipt of a bid protest.

Acquisition officials exercise judgment in assigning an appropriate amount of resources to conduct a source selection. They must consider evaluators' availabilities, expertise, and location. Potential resources are balanced with the task demands such as the award milestones, required travel, quantity of expected proposals, and quantity of evaluation factors and sub-factors (Edwards, 2006). For source selections with higher protest risk, acquisition officials may assign more evaluators and other team members and for a larger portion of their time. Logically, more people and more effort should mitigate protest-worthy mistakes. More resources can be indicated by transaction costs, determined by the number of full-time equivalent personnel working on the source selection. Therefore,

H17: There will be a negative relationship between transaction costs and receipt of a bid protest.

Human Factors

Fear of protest describes the level of apprehension a contracting professional has about receiving a bid protest (Hawkins et al., 2016). It follows that in cases in which contracting officers are worried about a protest, the acquisition team will take added measures to prevent a protest. Thus, it is posited that:

H18: There will be a negative relationship between fear of protest and receipt of a bid protest.

The RAND Corporation's study of bid protests revealed that industry representatives question the competency of the acquisition workforce, citing a need for additional training (Arena et al., 2018). Additionally, source selection experience has been found to reduce fear of protest (Hawkins et al., 2016). Experience appears to yield confidence in the compliance of the procurement process. Training and education may also provide the necessary awareness of the myriad of laws, regulations, and case law—any of the peculiarities of which could jeopardize a procurement. Therefore, it is reasonable to expect that:

H19: There will be a negative relationship between experience and receipt of a bid protest.

Combined, this set of hypotheses should help predict bid protests. The conceptual mode (Figure 1) is sufficiently comprehensive to enable practitioners to determine needed definitive action to improve the effectiveness of their source selections.

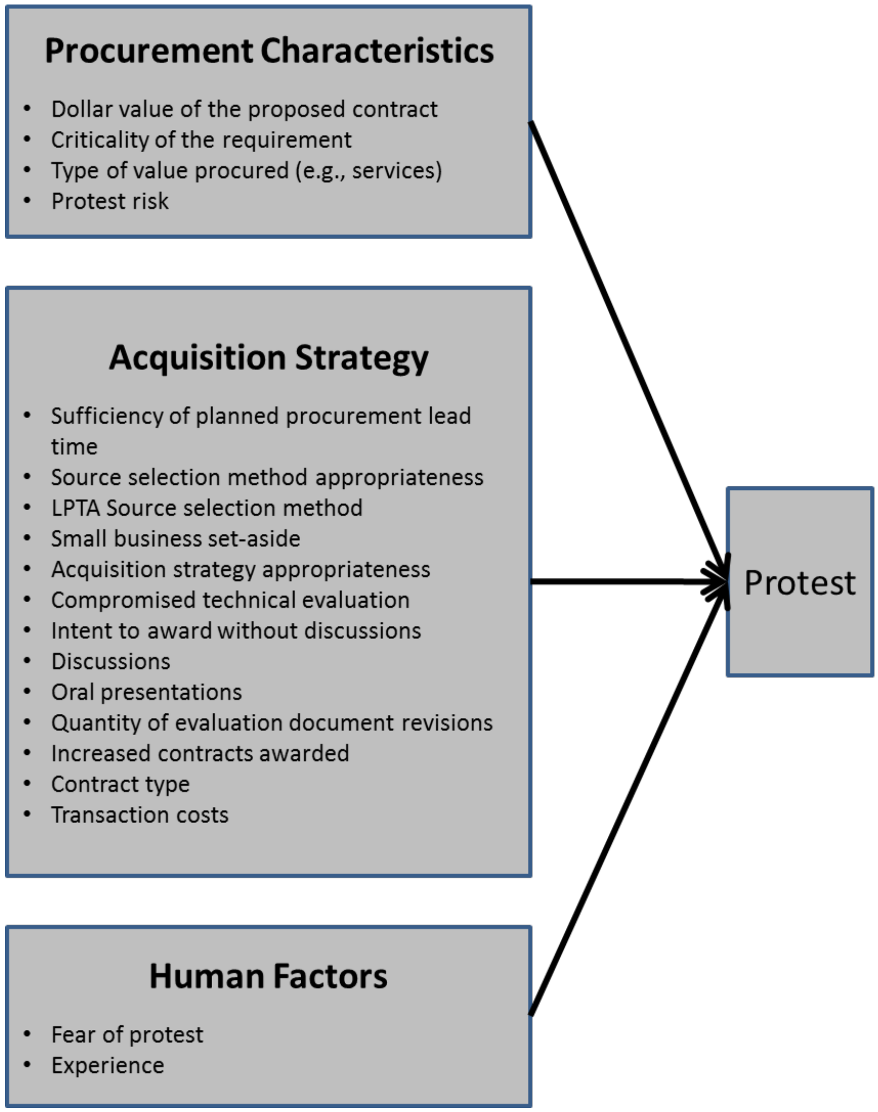


Figure 1. **Conceptual Model**

Methodology

The purpose of this research was to identify factors associated with the receipt of bid protests. This study examines a unique, rich data set of 350 government source selections resulting from a survey of U.S. Navy contracting officials. The data went beyond that of typical protest research that relies on summary-level contract award data from FPDS-NG and GAO's Electronic Protest Docketing System. Rather, the data set includes unique insights from those involved in the source selection, including perceptions (e.g., source selection method appropriateness) and objective data elements not captured elsewhere (e.g., intent to award without discussions). Given the exploratory nature of the research, a backward stepwise logistic regression model will be applied to the data.

Unit of Analysis

The unit of analysis for this data is a U.S. federal government source selection. Since many bid protests stem from a protestable action associated with a source selection (e.g., a proposal rating, rating justification, or basis of a trade-off analysis), this is the proper unit of analysis for the study. The data pertained to source selections conducted pursuant to FAR Part 15; those conducted using simplified acquisition procedures and task order competitions will be excluded.

Data

The data set included 350 records of source selections. Many records were omitted from this analysis due to missing data and conflicting data. Five records reported zero PALT, which is not possible. Another 32 records reported PALT less than 45 days. While the original survey instructed respondents to complete the survey pertaining to a FAR Part 15 source selection, some respondents may have reported on task order competitions. Due to advertising requirements (15 days), proposal preparation time (30 days), and time for evaluations, FAR Part 15 source selections should consume at least 45 days from receipt of a complete requirements package. Also, 15 records either included no dollar value or a value that was less than the simplified acquisition threshold ($150,000—meaning FAR Part 13 procedures or task order procedures were more likely). Finally, 66 records did not include sufficient transaction cost data to determine full-time equivalents. Together, for the sake of complete data and consistency of source selection rules, these 110 records were removed, leaving a data set of 240 records for analysis.

Summary

This research offers a first step toward quantitative, transaction-level investigation into reasons for bid protests. While no one can prevent an interested party from filing a bid protest (Rumbaugh, 2010), the factors identified herein can help acquisition managers hedge against the likelihood. This research will explore 19 meso-level antecedent factors that can be categorized as characteristics of the procurement, acquisition strategy decisions, and human factors. Based on the exploratory findings, the research will draw implications for theory and practice, and chart promising directions for future research into this important stream of acquisition research.

References

Arena, M.V., Persons, B., Blickstein, I., Chenoweth, M. E., Lee, G. T., Luckey, D., & Schendt, A. (2018). *Assessing bid protests of U.S. Department of Defense procurements: Identifying issues, trends, and drivers*. Santa Monica, CA: RAND Corporation.

Beugre, C. D., & Acar, W. (2008). Offshoring and cross-border interorganizational relationships: A justice model. *Decision Sciences*, *39*(3), 445–468.

Bies, R. J., & Moag, S. (1986). Interactional justice: Communication criteria of fairness. *Research on Negotiation in Organizations*, *1*, 43–55.

Blancero, D., & Ellram, I. (1997). Strategic supplier partnering: A psychological contract perspective. *International Journal of Physical Distribution & Logistics Management*, *27*(9), 616–629.

Camm, F., Chenoweth, M. E., Graser, J. C., Light, T., Lorell, M. A., Rudavsky, R., & Lewis, P.A. (2012). Government Accountability Office bid protests in Air Force source selections: Evidence and options. Santa Monica, CA: RAND Corporation.

Cibinic, J., Jr., Nash, R. C., Jr., & O-Brien-DeBakey, K. (2011). *Competitive negotiation: The source selection process* (3rd ed.) Washington, DC: The George Washington University Press.

Curry, W. S. (2018, August). Reversing the surge in protests: The "total weighed score" proposal scoring method. *Contract Management*, 48–59.

Dwyer, F. R., Schurr, P. H., & Oh, S. (1987, April). Developing buyer-seller relationships. *Journal of Marketing*, *51*, 11–27.

Edwards, V. (2006). *Source selection answer book* (2nd ed.). Vienna, VA: Management Concepts.

Finkenstadt, D., & Hawkins, T. (2016). #eVALUate: Monetizing service acquisition trade-offs using a Quality-Infused Price© methodology. *Defense Acquisition Research Journal*, *23*(2), 202–230.

Gilliland, S. (1993). The perceived fairness of selection systems: An organisational justice perspective. *Academy of Management Review*, *18*(4), 694–734.

Gordon, D. I. (2013). Bid protests: The costs are real, but the benefits outweigh them. *Public Contract Law Journal*, *42*(3), 489–516.

Government Accountability Office. (2014). *GAO Bid Protest Annual Report to Congress for Fiscal Year 2013* (GAO-14-276SP). Retrieved from www.gao.gov/assets/660/659993.pdf

Government Accountability Office. (2016). *GAO Bid Protest Annual Report to Congress for Fiscal Year 2016* (GAO-17-314SP). Retrieved from http://www.gao.gov/assets/690/681662.pdf

Government Accountability Office. (2017). *GAO Bid Protest Annual Report to Congress for Fiscal Year 2017* (GAO-18-237SP). Retrieved from https://www.gao.gov/assets/690/688362.pdf

Greenberg, J. (1990). Organizational justice: Yesterday, today, and tomorrow. *Journal of Management, 16*(2), 399.

Hair, J. F., Jr., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Hannaway, S. (2000). Oral presentations in negotiated procurements: Panacea or Pandora's box. *Public Contract Law Journal*, *29*(3), 455–509.

Hawkins, T., & Gravier, M. (2016). Antecedents and consequences of supplier performance evaluation efficacy. In *Proceedings of the 13th Annual Acquisition Research Symposium* (pp. 178–193). Monterey, CA: Naval Postgraduate School, Acquisition Research Program.

Hawkins, T., Gravier, M., & Yoder, E. C. (2016). Federal bid protests: Is the tail wagging the dog? *Journal of Public Procurement*, *16*(2), 152–190.

Kaynak, R., Sert, T., Sert, G., & Akyuz, B. (2015). Supply chain unethical behaviors and continuity of relationship: Using the PLS approach for testing moderation effects of inter-organizational justice. *International Journal of Production Economics*, *162*(4), 83–91.

Kraljic, P. (1983). Purchasing must become supply management. *Harvard Business Review*, *61*(5), 109–117.

Kumar, N. (1996). The power of trust in manufacturer-retailer relationships. *Harvard Business Review*, *74*(6), 92–106.

Leventhal, G. S. (1980). What should be done with equity theory? In K. J. Gergen, M. S. Greenberg, & R. H. Willis (Eds.), *Social exchange: Advances in theory and research*. New York, NY: Plenum Press.

Liu, Y., Huang, Y., Luo, Y., & Zhao, Y. (2012). How does justice matter in achieving buyer–supplier relationship performance? *Journal of Operations Management*, *30*(5), 355–367.

Manuel, K. M., & Schwartz, M. (2011). *GAO bid protests: An overview of time frames and procedures* (R40228). Retrieved from www.fas.org/sgp/crs/misc/R40228.pdf

Maser, S. M., & Thompson, G. F. (2010). *Understanding and mitigating protests of Department of Defense acquisition contracts* (WIL-CM-10-164). Salem, OR: Willamette University. Retrieved from http://www.acquisitionresearch.org/files/FY2010/WIL-CM-10-164.pdf

Mlinarchik, C. (2016). How many bid protests is too many? *Contract Management*, *56*(11), 38–47.

Mohr, J. J., & Sohi, R. S. (1995). Communication flows in distribution channels: Impact on assessments of communication quality and satisfaction. *Journal of Retailing*, *71*(4), 393–416.

Molenaar, K. R., & Tran, D. (2015). *Practices for developing transparent best value selection procedures*. Washington, DC: Transporation Research Board.

National Association of State Procurement Officials (NASPO). (2013). *Research brief: State bid protests*. Retrieved from http://www.naspo.org/dnn/portals/16/documents/.FINAL_NASPO_BidProtests_Research_Brief_042413.pdf

O'Rourke, R. (2014). *Navy littoral combat ship (LCS) program: Background and issues for Congress* (RL33741). Retrieved from http://www.fas.org/sgp/crs/weapons/RL33741.pdf

Plank, R. E., Landeros, R., & Plank, L. F. (1994). Values driving decisions in questionable purchasing situations. *International Journal of Purchasing and Materials Management*, *30*(4), 45–53.

Pocock, C. (2009, June 17). Tanker protagonists square off for round three of U.S. bid. AIN Online. Retrieved from http://www.ainonline.com/aviation-news/paris-air-show/2009-06-17/tanker-protagonists-square-round-three-us-bid

Poling, S. A. (2016, April 15). Proposed rules: Government Accountability Office. *Federal Register*, *81*(73), 22197–22203.

Rumbaugh, M. G. (2010). *Understanding government contract source selection*. Vienna, VA: Management Concepts.

Schwartz, M., & Manuel, K. M. (2015). *GAO bid protests: Trends and analysis* (R40227). Washington, DC: Congressional Research Service. Retrieved from https://www.fas.org/sgp/crs/misc/R40227.pdf

Thibaut, J., & Walker, L. (1975). *Procedural justice: A psychological analysis*. Hillsdale, NJ: Erlbaum.

Wallace, J. C. (2018, January–February). GAO rulings in contract protests. *Defense AT&L*, 39–43.

Panel 21. GAO Panel Discussion on Weapon System Outcomes and the Changing Oversight Environment

Thursday, May 9, 2019	
2:15 p.m. – 3:30 p.m.	**Chair: Shelby Oakley,** Director, Contracting and National Security Acquisition, Government Accountability Office ***How Are Recent Acquisition Reform Initiatives Changing the Oversight and Management of Major Defense Acquisition Programs?*** Cheryl Andrew, Government Accountability Office ***GAO Annual Assessment of Weapon Acquisition Programs*** Chris Durbin and Cheryl Andrew, Government Accountability Office ***Navy Shipbuilding: Past Performance Provides Valuable Lessons for Future Investments*** Diana Moldafsky and Cheryl Andrew, Government Accountability Office

Shelby Oakley—Ms. Oakley is a Director in the Government Accountability Office's (GAO) Contracting and National Security Acquisitions team. In her role, she oversees GAO's portfolio of work examining the most complex and expensive acquisitions within the federal government. Her portfolio includes Navy Shipbuilding and Nuclear Triad modernization programs as well as GAO's annual work to assess the cost, schedule, and performance of DOD's entire portfolio of major weapon system development programs – almost a two trillion dollar investment. In addition, her portfolio also includes oversight of contracting activities of the Veterans Affairs Department and whistleblower protections. Shelby previously served as Director in GAO's Natural Resources and Environment team where she led teams reviewing a range of nuclear security, policy, and nonproliferation related issues. In her role as Director, she has testified before Congress multiple times.

From 2004 to 2015, Shelby led teams reviewing the activities of the National Aeronautics and Space Administration (NASA) with a focus on helping NASA improve its acquisition management practices. Her reviews covered key aspects of NASA's operations, such as Space Shuttle workforce transition and sustainment of the International Space Station, as well as reviews of all major NASA systems including in depth reviews of NASA's human spaceflight programs and the James Webb Space Telescope.

Shelby joined GAO in 2001 after earning a Master's Degree in Public Administration from the University of Pittsburgh's Graduate School of Public and International Affairs in 2001 and her Bachelor of Arts Degree in both Psychology and Sociology from Washington and Jefferson College.

Panel 22. Planning, Programming, and Budgeting Execution Reform Within Defense Acquisition

Thursday, May 9, 2019	
2:45 p.m. – 3:30 p.m.	**Chair: John Terence Blake, VADM, USN (Ret.),** Professor of Practice and Conrad Chair, Naval Postgraduate School ***Capital Budgeting and Portfolio Optimization in the U.S. Navy and Department of Defense*** Johnathan Mun, Naval Postgraduate School ***An Approximate Dynamic Programming Approach for Weapon System Financial Execution Management*** Erich Morman and Jefferson Huang, Naval Postgraduate School ***Actual Obligation Versus Comptroller Projected Obligation Rates*** Todd Harrison and Seamus Daniels, Center for Strategic and International Studies

John Terence Blake, VADM, USN (Ret.)—Vice Admiral John Terence Blake was appointed to the United States Naval Academy from the state of New York, he graduated in 1975. His sea duty assignments include: USS New (DD 818), USS Sarfield (DD 837), USS Joseph Strauss (DDG 16), USS John Young (DD 973), USS Chandler (DDG 996), USS Leahy (CG 16), and USS Blue Ridge (LCC 19).

Blake commanded the destroyer USS O'Brien (DD 975), served on the 7th Fleet Staff as current operations and assistant chief of staff for Operations, commanded the guided-missile cruiser USS Normandy (CG 60) and served as commander, Carrier Strike Group 11.

His shore duty assignments include: flag lieutenant to commander, Navy Recruiting Command; Naval Post Graduate School where he earned a masters degree in Finance; Navy Staff (N80) head, Sea Control Section and program manager for the Navy Shipbuilding account; National War College where he earned a masters degree in National Security; Joint Staff (J8) division chief and head of the Combat Identification Joint Warfare Capability Assessment Team; director, Programming Division (N80); director, Operations Division, Office of Budget in the Office of the Assistant Secretary of the Navy (Financial Management/Comptroller); director, Operations Division, Fiscal Management Division in the Office of the Chief of Naval Operations; deputy director for Resources and Acquisition on the Joint Chiefs of Staff (J8) and deputy assistant secretary of the Navy for Budget.

He is authorized to wear the Navy Distinguished Service Medal, Defense Superior Service Medal with oak leaf cluster, the Legion of Merit with four gold stars, the Meritorious Service Medal with two gold stars, the Navy and Marine Corps Commendation Medal with two gold stars and various service and campaign medals.

Capital Budgeting and Portfolio Optimization in the U.S. Navy and Department of Defense

Johnathan Mun—PhD, is a Research Professor at the Naval Postgraduate School, Monterey, CA. [jcmun@nps.edu]

Abstract

This research has the explicit goal of proposing a reusable, extensible, adaptable, and comprehensive advanced analytical modeling process to help the U.S. Department of Defense (DoD) with risk-based capital budgeting and optimizing acquisitions and programs portfolios with multiple competing stakeholders while subject to budgetary, risk, schedule, and strategic constraints. The research covers topics of traditional capital budgeting methodologies used in industry, including the market, cost, and income approaches, and explains how some of these traditional methods can be applied in the DoD by using DoD-centric non-economic, logistic, readiness, capabilities, and requirements variables. Portfolio optimization for the purposes of selecting the best combination of programs and capabilities is also addressed, as are other alternative methods such as average ranking, risk metrics, lexicographic methods, PROMETHEE, ELECTRE, and others. Finally, an illustration at Program Executive Office Integrated Warfare Systems (PEO IWS) and Naval Sea Systems Command (NAVSEA) is presented to showcase how the methodologies can be applied to develop a comprehensive and analytically robust case study that senior leadership at the DOD may utilize to make optimal decisions.

Introduction

The United States Department of Defense (DoD) is always looking for better theoretically justifiable and quantitatively rigorous analytical methods for capital budgeting and portfolio optimization. Specific interest lies in how to identify and quantify the value of each program to the military and optimally select the correct mix of programs, systems, and capabilities that maximizes some military “value” (strategic, operational, economic) while subject to budgetary, cost, schedule, and risk constraints.

This research applies some private-sector and industry best practices coupled with advanced analytical methods and models to help create these methodologies. However, the uniqueness of the DoD requires that additional work be done to determine the concept of value to the military while considering competing stakeholders’ needs. We still need a defensible, quantitatively robust concept of military value to use in the modeling.

The purpose of this research is to illustrate and recommend approaches of modeling methodology and development of military value metrics, and how to combine them into a defensible, reusable, extensible, and practical approach within portfolios of programs.

This research specifically showcases how capital budgeting and portfolio optimization methods can be applied in the U.S. Navy as well as across the DoD in general, where multiple stakeholders (e.g., Office of the Secretary of Defense, Office of the Chief of Naval Operations, Congress) have their own specific objectives (e.g., capability, efficiency, cost effectiveness, competitiveness, lethality) as well as constraints (e.g., time, budget, schedule, manpower) and domain requirements (e.g., balancing the needs of anti-submarine warfare, anti-aircraft warfare, missile defense). This first-step research project provides an overview of the methodology employing nominal data variables to illustrate the analytics; it will be followed up by future research with more case-specific examples using actual subject matter expert (SME) data from the Office of the Chief of Naval Operations.

Capital Budgeting

The concept of capital budgeting and portfolio optimization has far-reaching consequences beyond the DoD. Private industry can greatly benefit from the concepts and methodologies developed in this research to apply portfolio optimization to its respective capital investment portfolios. These optimized portfolios are, by definition, the best and most efficient usage of a firm's capital to generate the greatest amount of value to the entire economy while mitigating risks for the organization and keeping limited budgetary and human resource constraints in check. More technically savvy individuals can apply the same methodologies in their retirement and investment portfolios, and portfolio managers can also leverage the knowledge and insights from the research to apply efficient frontier analyses for their clients' invested portfolios.

Portfolio Optimization

A portfolio, by definition, is any combination of two or more assets, projects, capabilities, or options. The whole portfolio is usually assumed to be greater than the sum of its parts, based on outcome performance measures, expected return on investment (ROI), capabilities, and other metrics (Mun, 2015). This assumption is due to the potential risk reduction, leverage, and synergy in terms of lower cost, interoperability, and flatter learning curve when multiple programs or capabilities are combined into a more cohesive portfolio (Mun, 2015, 2016).

In today's competitive global economy, companies in the private sector are faced with many difficult decisions. These decisions include allocating financial resources, building or expanding facilities, managing inventories, and determining product-mix strategies. The U.S. military is no different. The DoD, as a whole, has oftentimes struggled with trying to find the best force mix, or optimal programs that maximize military capabilities within set budgetary, scheduling, and human resource constraints.

Such decisions might involve thousands or millions of potential alternatives. Considering and evaluating each of them would be impractical or even impossible. An optimization model can provide valuable assistance in incorporating relevant variables when analyzing decisions and finding the best solutions for making decisions. These models capture the most important features of a problem and present them in a form that is easy to interpret. Models often provide insights that intuition alone cannot. An optimization model has three major elements: decision variables, constraints, and an objective. In short, the optimization methodology finds the best combination or permutation of decision variables (e.g., which programs or capabilities the DoD should acquire and which projects to eliminate) in every conceivable way such that the objective is maximized (e.g., maximum capabilities, highest expected military value, maximum military utility) or minimized (e.g., cost risk and schedule risk) while still satisfying the constraints (e.g., budget, political, human resources, and other non-economic resources).

Obtaining optimal values generally requires that you search in an iterative or ad hoc fashion. This search involves running one iteration for an initial set of values, analyzing the results, changing one or more values, rerunning the model, and repeating the process until you find a satisfactory solution. This process can be very tedious and time-consuming even for small models, and often it is not clear how to adjust the values from one iteration to the next. Using the proposed modeling process can eliminate the negatives of searching in an iterative or ad hoc fashion.

Research Questions and Objectives

The proposed research attempts to answer the following research questions:

- Can the DoD perform credible and defensible portfolio optimization on capabilities and programs?
- How are military-based definitions of value created and used in developing optimal portfolios?
- What are the best approaches and algorithms that are most amenable to defense acquisition portfolios?

The proposed modeling methodology and process to be developed has the following objectives:

- Create and model multiple-objective optimization models based with competing stakeholders.
- Develop models based on the integrated risk management (IRM) methodology where Monte Carlo risk simulation methods will be employed to analyze risks and uncertainties in the portfolio's inputs.
- Optimize the portfolio of options (i.e., given a set of projects, programs, acquisition, or capability options with different costs, benefits, capabilities, and uncertainties, helps identify which programs or capabilities should be chosen given constraints in budget, schedule, and capability requirements, all the while considering various viewpoints from different stakeholders, including Navy leadership, field commanders, and technical engineering, and economic and strategic points of view).

Consider that, to maintain a high level of competitiveness, corporations in the private sector need to continually invest in technology, research and development (R&D), and other capital investment projects. But resource constraints require organizations to strategically allocate resources to a subset of possible projects. A variety of tools and methods can be used to select the optimal set of technology projects. However, these methods are only applicable when projects are independent and are evaluated in a common funding cycle. When projects are interdependent, the complexity of optimizing even a moderate number of projects over a small number of objectives and constraints can become overwhelming. Dickinson, Thornton, and Graves (2001) presented a model developed for the Boeing Company in Seattle to optimize a portfolio of product development improvement projects. The authors illustrate how a dependency matrix (modeling of interdependencies among projects) is applied in a nonlinear integer programming methodology to optimize project selection. The model also balances risk, overall objectives, and the cost and benefit of the entire portfolio. Once the optimum strategy is identified, the model enables the team to quickly quantify and evaluate small changes to the portfolio.

In the U.S. military context, risk analysis, real options analysis, and portfolio optimization techniques enable a new way of approaching the problems of estimating return on investment (ROI) and the risk value of various strategic real options. There are many DoD requirements for using more advanced analytical techniques. For instance, the Clinger-Cohen Act of 1996 mandates the use of portfolio management for all federal agencies. The GAO's 1997 report entitled *Assessing Risks and Returns: A Guide for Evaluating Federal Agencies' IT Investment Decision-Making* requires that IT investments apply ROI measures. DoD Directive (DoDD) 8115.01 (DoD, 2005) mandates the use of performance metrics based on outputs, with ROI analysis required for all current and planned IT investments.

DoDD 8115.bb (2006) implements policy and assigns responsibilities for the management of DoD IT investments as portfolios within the DoD enterprise where it defines a portfolio to include outcome performance measures and an expected return on investment. The DoD's *Risk Management Guidance Defense Acquisition Guidebook* requires that alternatives to the traditional cost estimation need to be considered because legacy cost models tend not to adequately address costs associated with information systems or the risks associated with them (see Mun, Ford, & Housel, 2012).

Literature Review

Portfolio Modeling in Military Applications

Optimization is a rich and storied discipline designed to use data and information to guide decision making in order to produce an optimal, or very close to optimal, outcome. However, "government agencies have been much slower to use these approaches to increase efficiency and mission effectiveness, even though they collect more data than ever before" (Bennett, 2017). For these government agencies, optimization solutions can utilize the large amounts of data from different sources to provide decision makers with alternative choices that optimally meet agency objectives.

Greiner, McNutt, Shunk, and Fowler (2001) correctly stated that standard economic measures such as internal rate of return (IRR), net present value (NPV), and return on investment (ROI) are commonly used in evaluating commercial-based R&D projects to help identify optimal choices. However, such economic measures in their commercial form are of little use in evaluating weapon systems development efforts. Therefore, this paper examines the challenges faced by the DoD in determining the value of weapon systems during the R&D portfolio selection processes.

Burk and Parnell (2011) reviewed the use of portfolio decision analysis in military applications, such as weapon systems, types of forces, installations, and military R&D projects. They began with comparing military and commercial portfolio problems in general and discussing the distinguishing characteristics of the military decision environment: hostile and adaptive adversaries, a public decision process with multiple stakeholders, and high system complexity. Based on their work, the authors observed that the "most widespread prominent feature of these applications is the careful modeling of value from multiple objectives" (Burk & Parnell, 2011). What they found surprising was that "quantitative methods of measuring and valuing risk are surprisingly rare, considering the high level of uncertainty in the military environment" (Burk & Parnell, 2011). Their analysis examined portfolio applications in more detail, looking at how military analysts model portfolio values, weight assessments, constraints and dependencies, and uncertainty and risk.

Davendralingam and DeLaurentis (2015) looked at analyzing military capabilities as a system of systems (SoS) approach. According to the authors, this approach creates significant development challenges in terms of technical, operational, and programmatic dimensions. Tools for deciding how to form and evolve SoS that consider performance and risk are lacking. Their research leveraged tools from financial engineering and operations research perspectives in portfolio optimization to assist decision making within SoS. The authors recommended the use of more robust portfolio algorithms to address inherent real-world issues of data uncertainty, inter-nodal performance, and developmental risk. A naval warfare situation was developed in the paper to model scenario applications to find portfolios of systems from a candidate list of available systems. Their results show how the optimization framework effectively reduces the combinatorial complexity of trade-space exploration by allowing the optimization problem to handle the mathematically intensive aspects of the decision-making process. As a result, the authors concluded that human decision makers can be tasked to focus on choosing the appropriate weights for risk aversion in making final decisions rather than on the mathematical constructs of the portfolio.

Sidiropoulos, Sidiropoulou, and Lalagas (2014) ran a portfolio management analysis with a focus on identifying and assessing current commercial off-the-shelf (COTS) Portfolio Analysis (PA) software products and solutions. *Risk Simulator* was used to develop portfolio models. These models were populated with relevant data and then run through an appropriate number of simulation iterations to assess candidate projects with respect to risk and Expected Military Value (EMV). The examples and models used in this paper discuss Portfolio Management Analysis (PMA) during various stages of project management and systems engineering. The goal for PMA is realized after the entire project design infrastructure is implemented and the end users' instruments are provided for implementation. The authors' intent was to identify "approaches and tools to incorporate PMA net-centric strategies to meet war fighter and business operations requirements, while continuing to maintain current levels of service, ensuring conservation of manpower and meeting infrastructure resource requirements" (Sidiropoulos, Sidiropoulou, & Lalagas, 2014).

Flynn and Field (2006) looked at quantitative measures that were under development to assess the Department of the Navy's (DoN's) portfolio of acquisitions to improve business practices through better analytical tools and models. The authors found that the DoN's time would be better served by shifting its attention from analyzing individual acquisition programs (now studied exhaustively) to analyzing a portfolio of systems as a whole. This approach is similar to the methodology employed as a best practice in the private sector. According to the research, this high-level view provides senior military leaders valuable metrics for measuring risks and uncertainties of costs, capabilities, and requirements. Armed with these metrics, senior leaders can make better choices, among a set of plausible portfolios, to satisfy the Navy's national security objectives. To support their analysis, a subset of the then-current DoN portfolio was selected by financial management and acquisition staff with which to test a methodology of portfolio analysis in the area of Mine Countermeasures, a diverse, representative system of programs. This pilot model was a multi-phase process that included gathering life-cycle cost data for the various systems to be analyzed, establishing a scoring system using subject matter experts to determine how effectively current and future systems match capabilities to requirements, and developing a means to display results by which decision makers can examine risk-reward analysis and conduct trade-offs. The researchers' ultimate goal was to assess military investments using portfolio analysis methodology.

The GAO (1997, 2007) emphasized the approach of optimizing a portfolio mix to manage risk and maximize the rate of return. Although the DoD produces superior weapons, the GAO reported that the department has failed to deliver weapon systems on time, within budget, and with desired capabilities. While recent changes to the DoD's acquisition policy held the potential to improve outcomes, programs continue to experience significant cost and schedule overruns. The GAO was asked to examine how the DoD's processes for determining needs and allocating resources can better support weapon system program stability. To do this, according to the report, the GAO compared the DoD's processes for investing in weapon systems to the best practices that successful commercial companies use to achieve a balanced mix of new products, including companies such as Caterpillar, Eli Lilly, IBM, Motorola, and Procter and Gamble. Based on the reports, the GAO found that to achieve a balanced mix of executable development programs and ensure a good return on their investments, the successful commercial companies the GAO reviewed take an integrated, portfolio management approach to product development. Through this approach, companies assess product investments collectively from an enterprise level, rather than as independent and unrelated initiatives. These commercial entities weigh the relative costs, benefits, and risks of proposed products using established criteria and methods and select those products that can exploit promising market opportunities within resource constraints and move the company toward meeting its strategic goals and objectives. In these firms, investment decisions are frequently revisited, and if a product falls short of expectations, companies make tough go/no-go decisions over time.

Wismeth (2012) noted that the Army has implemented the Army Portfolio Management Solution (APMS) to facilitate collection and analysis of information necessary to prioritize the thousands of IT investments within its portfolio. IT investments are grouped according to the mission capabilities they support: Warfighter, Business, and Enterprise Information Environment Mission Areas, each of which is led by a three- or four-star-level general officer or senior executive.

Janiga and Modigliani (2014) recommended that the DoD foster dynamic and innovative solutions for tomorrow's warfighter by designing acquisition portfolios that deliver an integrated suite of capabilities. Program executive officers (PEOs) today often focus on executing a dozen similar but independent programs. In contrast, large commercial businesses manage integrated product lines for items ranging from automobiles and electronics to software and health services. The DoD could leverage this model as a basis for constructing portfolios of similar programs that deliver enhanced capabilities in shorter timeframes.

The Institute for Defense Analyses (IDA) prepared a document for the Office of the Director, Acquisition Resources and Analysis, under a task titled "Portfolio Optimization Feasibility Study" (Weber et al., 2003). The objective was to study the feasibility of using optimization technology to improve long-term planning of defense acquisition. The model described in this document is an example of optimization technology that can estimate and optimize production schedules of Acquisition Category I programs over a period of 18 years.

Vascik, Ross, and Rhodes (2015) found that the modern warfighter operates in an environment that has dramatically evolved in sophistication and interconnectedness over the past half century. With each passing year, the infusion of ever more complex technologies and integrated systems places increasing burdens on acquisition officers to make decisions regarding potential programs with respect to the joint capability portfolio. Furthermore, significant cost overruns in recent acquisition programs reveal that, despite efforts since 2010 to ensure the affordability of systems, additional work is needed to develop enhanced approaches and methods. Vascik et al.'s paper discussed research that builds on prior work that explored system design trade-spaces for affordability under uncertainty, extending it to the program and portfolio level. Time-varying exogenous factors, such as resource availability, stakeholder needs, or production delays, may influence the potential for value contribution by constituent systems over the life cycle of a portfolio and make an initially attractive design less attractive over time. Vascik et al. (2015) introduced a method to conduct portfolio design for affordability by augmenting Epoch-Era Analysis with aspects of Modern Portfolio Theory. The method is demonstrated through the design of a carrier strike group portfolio involving the integration of multiple legacy systems with the acquisition of new vessels.

According to DoDD 5100.96 (DoD, 2017), the DoD Space Assessment (PDSA) monitors and oversees the performance of the entire DoD space portfolio. The PDSA, in assessing space-related threats, requirements, architectures, programs, and their synchronization, advises senior DoD leadership and recommends NSS enterprise-level adjustments. It conducts an annual strategic assessment, or Space Strategic Portfolio Review (SPR) when directed, assisted by the DSC and DCAPE, to address space posture and enterprise-level issues and provides the DMAG and the secretary and deputy secretary of defense with results of the analysis, which may include prioritized programmatic choices for space capabilities.

Capital Budgeting and the Value Concept

The Traditional Views

Value is defined as the single time-value discounted number that is representative of all future net profitability. In contrast, the market price of an asset may or may not be identical to its value ("assets," "projects," and "strategies" are used interchangeably). For instance, when an asset is sold at a significant bargain, its price may be somewhat lower than its value, and one would surmise that the purchaser has obtained a significant amount of value. The idea of valuation in creating a fair market value is to determine the price that closely resembles the true value of an asset. This true value comes from the physical aspects of the asset as well as its nonphysical, intrinsic, or intangible aspects. Both aspects have the capability to generate extrinsic monetary value or intrinsic strategic value. Traditionally, there are three mainstream methodologies to valuation, namely, the market approach, the income approach, and the cost approach (see Mun, Hernandez, & Rocco, 2016, for more details). Other approaches used in valuation, more appropriately applied to the valuation of intangibles, rely on quantifying the economic viability and economic gains the asset brings to the firm. There are several well-known methodologies for intangible-asset valuation, particularly in valuing trademarks and brand names. These methodologies apply the combination of the market, income, and cost approaches just described. Although the financial theories underlying these approaches are sound in the more traditional deterministic view, they cannot be reasonably used in isolation when analyzing the true strategic flexibility value of a firm, project, or asset.

Portfolio Optimization

In today's competitive global conditions, the DoD is faced with many difficult decisions. These decisions include allocating financial resources, building or expanding facilities, managing inventories for maintenance, and determining force-mix strategies. Such decisions might involve thousands or millions of potential alternatives. Considering and evaluating each of them would be impractical or even impossible. A model can provide valuable assistance in incorporating relevant variables when analyzing decisions and in finding the best solutions for making decisions. Models capture the most important features of a problem and present them in a form that is easy to interpret. Models often provide insights that intuition alone cannot. An optimization model has three major elements: decision variables, constraints, and an objective. In short, the optimization methodology finds the best combination or permutation of decision variables (e.g., which products to sell and which projects to execute) such that the objective is maximized (e.g., in revenues and net income) or minimized (e.g., in risk and costs) while still satisfying the constraints (e.g., budget and resources).

Obtaining optimal values generally requires that you search in an iterative or ad hoc fashion. This search involves running one iteration for an initial set of values, analyzing the results, changing one or more values, rerunning the model, and repeating the process until you find a satisfactory solution. This process can be very tedious and time consuming even for small models, and it is often not clear how to adjust the values from one iteration to the next.

A more rigorous method systematically enumerates all possible alternatives. This approach guarantees optimal solutions if the model is correctly specified. Suppose that an optimization model depends on only two decision variables. If each variable has 10 possible values, trying each combination requires 100 iterations (102 alternatives). If each iteration is very short (e.g., two seconds), then the entire process could be done in approximately three minutes of computer time.

However, instead of two decision variables, consider six, then consider that trying all combinations requires 1,000,000 iterations (106 alternatives). It is easily possible for complete enumeration to take weeks, months, or even years to carry out (Mun, 2015). To run the analysis, we use the *Portfolio Optimization* tool in the ROV PEAT software application (courtesy of www.realoptionsvaluation.com). In the Portfolio Optimization section of this tool, the individual projects can be modeled as a portfolio and optimized to determine the best combination of projects for the portfolio.

The projects can be modeled as a portfolio and optimized to determine the best combination of projects for the portfolio in the *Optimization Settings* subtab. Analysts start by selecting the optimization method (Static or Dynamic Optimization). Then they select the decision variable type *Discrete Binary* (choose which Project or Options to execute with a go/no-go binary 1/0 decision) or *Continuous Budget Allocation* (returns percentage of budget to allocate to each *option* or *project* as long as the total portfolio is 100%); select the *Objective* (Max NPV, Min Risk, etc.); set up any *Constraints* (e.g., budget restrictions, number of projects restrictions, or create customized restrictions); select the options or projects to optimize/allocate/choose (default selection is *all options*); and when completed, click *Run Optimization*.

Figure 1 illustrates the *Optimization Results*, which returns the results from the portfolio optimization analysis. The main results are provided in the data grid, showing the final *Objective Function* results, final *Optimized Constraints*, and the allocation, selection, or optimization across all individual options or projects within this optimized portfolio. The top left portion of the screen shows the textual details and results of the optimization algorithms applied, and the chart illustrates the final objective function. The chart will only show a single point for regular optimizations, whereas it will return an investment efficient frontier curve if the optional *Efficient Frontier* settings are set (min, max, step size).

Figures 1 and 2 are critical results for decision makers as they allow decision makers flexibility in designing their own portfolio of options. For instance, Figure 1 shows an efficient frontier of portfolios, where each of the points along the curve are optimized portfolios subject to a certain set of constraints. In this example, the constraints were the number of options that can be selected in a ship and the total cost of obtaining these options, which is subject to a budget constraint. The colored columns on the right in Figure 1 show the various combinations of budget limits and maximum number of options allowed. For instance, if a program office in the Navy only allocates $2.5 million (see the Frontier Variable located on the second row) and no more than four options per ship, then only options 3, 7, 9, and 10 are feasible, and this portfolio combination would generate the biggest bang for the buck while simultaneously satisfying the budgetary and number of options constraints. If the constraints were relaxed to, say, five options and a $3.5 million budget, then option 5 is added to the mix. Finally, at $4.5 million and no more than seven options per ship, options 1 and 2 should be added to the mix. Interestingly, even with a higher budget of $5.5 million, the same portfolio of options is selected. In fact, the Optimized Constraint 2 shows that only

$4.1 million is used. Therefore, as a decision-making tool for the budget-setting officials, the maximum budget that should be set for this portfolio of options should be $4.1 million. Similarly, the decision maker can move backwards, where, say, if the original budget of $4.5 million was slashed by Congress to $3.5 million, then the options that should be eliminated would be options 1 and 2. While Figure 1 shows the efficient frontier where the constraints such as number of options allowed and budget were varied to determine the efficient portfolio selection, Figure 2 shows multiple portfolios with different objectives. For instance, the five models shown were to maximize the financial bang for the buck (minimizing cost and maximizing value while simultaneously minimizing risk), maximizing Naval Operations (OPNAV) value, maximizing KVA value, maximizing Command value, and maximizing a Weighted Average of all objectives. This capability is important because depending on who is doing the analysis, their objectives and decisions will differ based on different perspectives. Using a multiple criteria optimization approach allows one to see the scoring from all perspectives. The option with the highest count (e.g., option 5) would receive the highest priority in the final portfolio, as it satisfies all stakeholders' perspectives and would hence be considered first, followed by options with counts of 4, 3, 2, and 1.

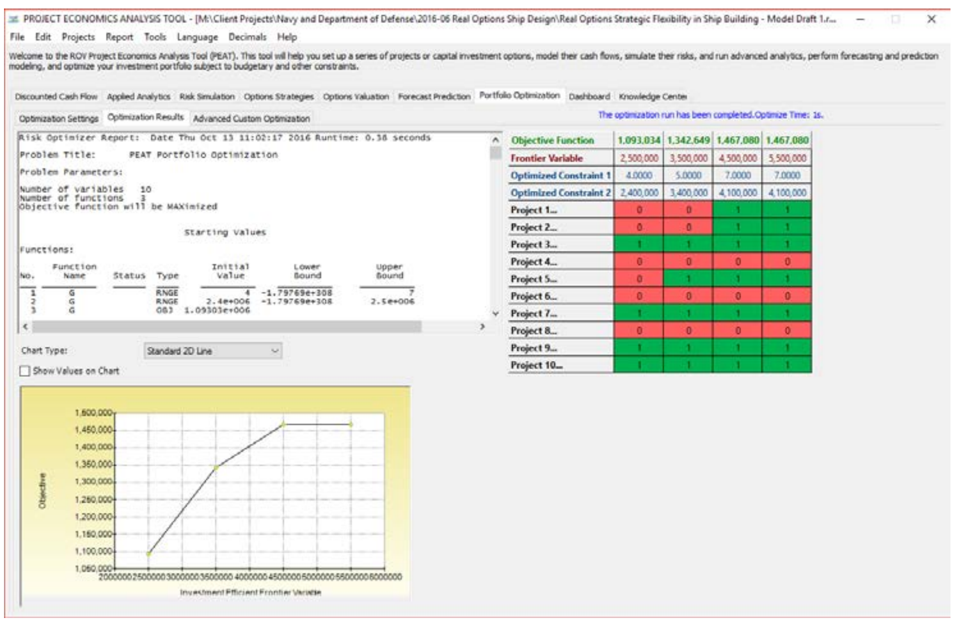


Figure 1. **Portfolio Optimization Results**

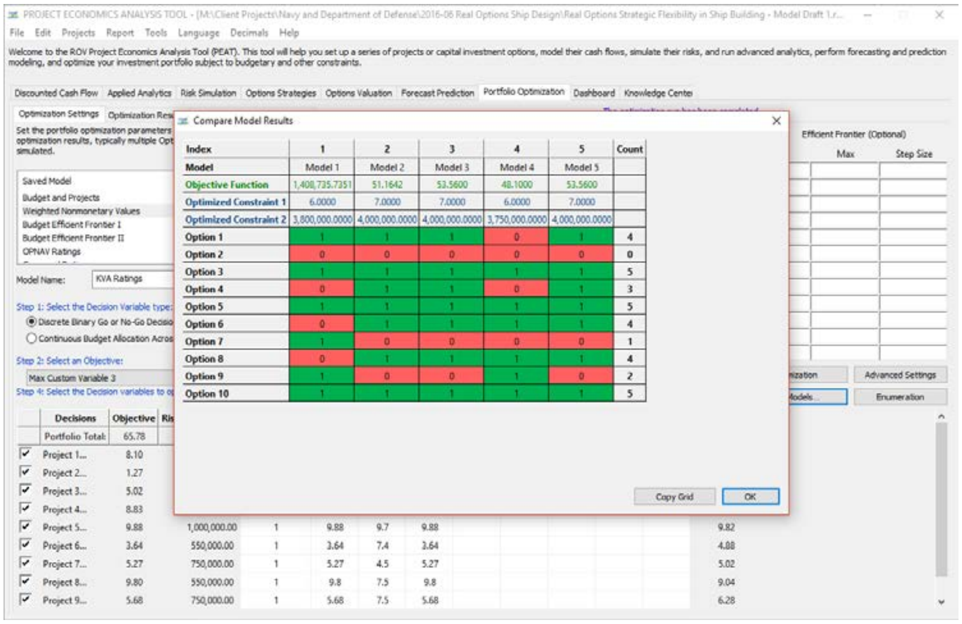


Figure 2. **Multi-Criteria Portfolio Optimization Results**

Alternative Analytical Approaches

Lexicographic Average Rank for Evaluating Uncertain Multi-Indicator Matrices With Risk Metrics

In many situations, projects are characterized by several criteria or attributes that can be assessed from multiple perspectives (financial, economic, etc.). Each criterion is quantified via performance values (PV), which can either be numerical or categorical. This information is typically structured in a multi-indicator matrix Q. A typical problem faced by a decision maker is to define an aggregate quality (AQ) able to synthesize the global characteristics of each project and then derive the rankings from the best to the worst base-case ranking (Mun et al., 2016). Ranking techniques can be classified as parametric and nonparametric. A parametric technique requires information about decision-maker preferences (e.g., criterion weights). According to Dorini, Kapelan, and Azapagic (2011), some examples of parametric techniques include the ELECTRE methods (Roy, 1968) and PROMETHEE—Preference Ranking Organization Methods for Enrichment Evaluations (Brans & Vincke, 1985). Nonparametric techniques, such as Partial Order Ranking (Bruggemann et al., 1999) and Copeland Scores (Al-Sharrah, 2010), do not require information from the decision maker. In general, all of these techniques are able to produce a ranking of the alternatives from the best to the worst.

Therefore, given a matrix **Q**, the selected procedure generates a ranking, defined as the base-case rank (BCR). As a result of this assessment, for each alternative, a specific rank R_i that considers the multiple perspectives defined by the decision maker is obtained. The set of R_i corresponds to the global evaluation under the first synthetic attribute, defined and named as *base ranking*, and capable of characterizing the alternatives in the base case.

However, in real-life situations, each performance value could be affected by uncertain factors. Several approaches have been presented for analyzing how the uncertainty in the performance values (the input) affects the ranking of the objects (the output; Rocco & Tarantola, 2014; Corrente, Figueira, & Greco, 2014; Hyde, Maier, & Colby, 2004; Hyde & Maier, 2006; Yu et al., 2012). The approaches, based on Monte Carlo simulation, consider each uncertain factor as a random variable with known probability density functions. As a result, the AQ of each alternative and, therefore, its ranking also become random variables, with approximated probability distributions. In such situations, the decision maker could perform probability distribution evaluations. For example, the decision maker could be interested in determining not only what the worst rank of a specific alternative is, but also its probability and volatility (risk evaluation).

In the standard approach, the probability of an alternative being ranked as in the BCR is selected as the synthetic attribute *probability* able to characterize the alternatives under uncertainty.

The stochastic nature of the AQ of each alternative could be further assessed in order to reflect the risk evaluation induced by uncertainty. In this case, it is required to compare several random variables synthesized through their percentiles and statistical moments. Several approaches have been proposed to this end, such as a simple comparison of the expected value, the expected utility (Von Neumann & Morgenstern, 1947), the use of low order moments (Markowitz, 1952), risk measures (Jorion, 2007; Mansini, Ogryczak, & Speranza, 2007; Rockafellar & Uryasev, 2000), the Partitioned Multiobjective Risk Method (PMRM; Asbeck & Haimes, 1984; Haimes, 2009), and the stochastic dominance theory (Levy, 2006), among others.

To consider the risk evaluation induced by uncertainty, each alternative is represented by the third synthetic attribute: *compliance*. This new attribute is based on a simultaneous assessment of several risk measures and some moments of each AQ distribution (Mun et al., 2016).

At this point, each alternative is assessed from three different angles:

1. Multiple decision-making perspectives that include several aspects such as economic, financial, technical, and social (*base ranking*)
2. Uncertainty propagation on performance values (*probability*)
3. A risk evaluation based on the generated probability distribution (*compliance*)

These perspectives are then used for defining a new multi-indicator matrix $\mathbf{Q}_1$ correlated to projects and synthesized using a ranking technique. However, in some situations, decision makers need to select projects following their most-preferred criteria successively. For this reason, an aggregation ranking technique that allows compensation is useless.

Therefore, the final assessment is derived using a combined approach based on a *nonparametric aggregation rule* (using the concept of average rank) for attributes 1 and 2; a simple procedure for score assignment for attribute 3; and a *lexicographic rule*. In addition, a preliminary analysis of the alternatives is performed by using a Hasse diagram (Bruggemann & Patil, 2011). To the best of the researcher’s knowledge, this type of combined assessment has not been reported in the literature.

Average Rank Approach

Let P define a set of n objects (e.g., alternatives) to be analyzed and let the descriptors q_1, q_2..., q_m define m different attributes or criteria selected to assess the objects in P (e.g., cost, availability, environmental impact). It is important that attributes are defined to reflect, for example, that a low value indicates low rankings (best positions), while a high value indicates high ranking (worst positions; Restrepo et al., 2008). However, for a given problem or case study, this convention could be reversed.

If only one descriptor is used to rank the objects, then it is possible to define a total order in *P*. In general, given $x, y \in P$, if $q_i(x) \leq q_i(y)\ \forall i$, then *x* and *y* are said to be comparable. However, if two descriptors are used simultaneously, the following could happen: $q_1(x) \leq q_1(y)$ and $q_2(x) > q_2(y)$. In such a case, *x* and *y* are said to be incomparable (denoted by $x \| y$). If several objects are mutually incomparable, set *P* is called a partially ordered set or *poset*. Note that since comparisons are made for each criterion, no normalization is required.

The objects in a poset can be represented by a directed acyclic graph whose vertices are the objects $\in P$, and there is an edge between two objects only if they are comparable and one covers the other, that is, when no other element is in between the two. Such a chart is termed a Hasse diagram (Bruggemann, Schwaiger, & Negele, 1995).

A Hasse diagram is, then, a nonparametric ranking technique and can perform ranking decisions from the available information without using any aggregation criterion. However, while it cannot always provide a total order of objects, it does provide an interesting overall picture of the relationships among objects.

A useful approach to produce a ranking is based on the concept of the average rank of each object in the set of linear extensions of a poset (De Loof, De Baets, & De Meyer, 2011). Since the algorithms suggested for calculating such average ranks are exponential in nature (De Loof et al., 2011), special approximations have been developed, such as the Local Partial Order Model (LPOM; Bruggemann et al., 2004), the extended LPOM (LPOMext; Bruggemann & Carlsen, 2011), or the approximation suggested by De Loof et al. (2011).

From the Hasse diagram, several sets can be derived (Bruggemann & Carlsen, 2011). If $x \in P$,

1. $U(x)$, the set of objects incomparable with x: $U(x) := \{y \in P : x \| y\}$
2. $O(x)$, the *down* set: $O(x) := \{y \in P : y \leq x\}$
3. $S(x)$, the successor set: $S(x) := O(x) - \{x\}$
4. $F(x)$, the *up* set: $F(x) := \{y \in P : x \leq y\}$

Then, the following average rank indexes are defined:

a) $LPOM(x) = (|S(x)| + 1) \times (n + 1) \div (n + 1 - |U(x)|)$

b) $LPOMext(x) = |O(x)| + \sum_{y \in U(x)} \frac{p_y^<}{p_y^< + p_y^>}$

where *n* is the number of objects,

$|V|$ defines the cardinality of the set *V,*

$p_y^< = |O(x) \cap U(y)|$, $p_y^> = |F(x) \cap U(y)|$, and $y \in U(x)$

Lexicographic Approach

A lexicographic approach allows decision makers to introduce decision rules in which they select more objects impacting on their most-preferred criteria. According to Saban and Sethuraman (2014), when two objects have the same impact on the most-preferred criteria, decision makers prefer the one with the highest impact on the second most-preferred criteria, and so forth. This lexicographic representation models the problems where decision makers strictly prefer one criterion over another or they are managing noncompensatory aggregation (Yaman et al., 2011; Pulido, Mandow, & de la Cruz, 2014).

Finally, decision makers can model their strong preferences over the criteria selected mainly because, after further analysis of the problem, they are not indifferent or only weakly sure about their preferences on the criteria taken into consideration. In other words, they will always prefer one criterion to another without considering criterion weights explicitly.

Risk Metrics and Compliance

Risk metrics are statistical indicators or measurements that allow decision makers to analyze the dispersion (volatility) of certain events or outcomes. Hence, a random variable can be evaluated using statistical moments (e.g., mean, variance, skewness, kurtosis), or risk measurements can be used to analyze extreme values, such as Value at Risk (VaR) and Conditional VaR (Bodie, Kane, & Marcus, 2009; Fabozzi, 2010; Matos, 2007; Mun, 2015).

In decision problems, risk metrics play an important role in analyzing the volatility or stability of a set of options or a portfolio of alternatives, for example, in financial risk management (Chong, 2004), portfolio risk management (Bodie, Kane, & Marcus, 2009), and enterprise risk management (Scarlat, Chirita, & Bradea, 2012), as well as a variety of other areas (Fabozzi, 2010; Szolgayová et al., 2011).

In order to determine how risky an object is and its relationship with other objects, a compliance approach is followed, that is, the definition of a set of rules to guide decision makers (Hopkins, 2011). Several approaches have been proposed for assessing the compliance. For example, Barrett and Donald (2003) propose a stochastic dominance analysis to compare probability distributions before establishing a hierarchy; Boucher et al. (2014) rely on risk metrics and forecasting to adjust models by historical performance; and Zanoli et al. (2014) analyze impacts of risk factors on noncompliance in UK farming.

The compliance approach is more user-friendly for decision making because it allows evaluating whether an object performs according to decision-makers' preferences over defined risk metrics. The basic idea is to dichotomize the risk continuum (Hopkins, 2011). Therefore, the higher the compliance with a defined risk metric, the higher the alignment with the decision-makers' preferences. Similar approaches are considered by Scarlat, Chirita, and Bradea (2012) and Tarantino (2008) relying on key risk indicators.

Multicriteria Analysis

In addition to uncertainty and flexibility, another complexity appears when decision makers need to introduce potentially conflicting decision criteria (quantitative or qualitative, monetary and nonmonetary) into project management, such as legal (taxes, compliance, social responsibility, etc.), environmental (level of pollution, noise, watershed issues, etc.), economic (level of economic growth, national income, inflation, unemployment, etc.), and social (number of employees, value to society, safety and security, community development). Furthermore, those criteria might have different relative importance (RI) or weights.

To address this concern, multicriteria analysis (MCA) has become a powerful mechanism to handle multidimensional problems and to obtain an Aggregate Quality (AQ) supporting the final decision (Bouyssou et al., 2006; Brito, de Almeida, & Mota, 2010). MCA refers to a set of methods, techniques, and tools that help people with their decision problems (description, clustering, ranking, and selection) by simultaneously considering more than one objective or criterion (Roy, 1996; Ghafghazi et al., 2010; Kaya & Kahraman, 2011; Afsordegan et al., 2016).

The Preference Ranking Organization Method for Enrichment Evaluations (PROMETHEE; Goumas & Lygerou, 2000; Brans & Mareschal, 2005; Behzadian et al., 2010; Tavana et al., 2013) has been proposed as a proper MCA technique. PROMETHEE methods are based on outranking the relationship *S*. This concept does not determine if the relationship among two alternatives *a* and *b* is a strong preference (*a P b*), weak preference (*a Q b*), or indifference (*a I b*), but instead it establishes if “the alternative *a* is at least as good as the alternative *b*” (Brans & Mareschal, 2005).

PROMETHEE methods are suitable because of their theoretical and practical advantages. For instance, they can associate to each project an AQ index that maximizes the available information in terms of decision-makers’ preferences over the criteria selected, as well as the preferences’ intensities among alternatives and the nature of each criteria (Bouyssou et al., 2006).

Other methods could also be allowed to handle this multicriteria approach, for example, the ELECTRE methods (Bouyssou et al., 2006), AHP—Analytical Hierarchy Process (Desai, Bidanda, & Lovell, 2012; Saaty, 2013), MACBETH (Cliville, Berrah, & Mauris, 2007; Costa, De Corte, & Vansnick, 2012), and TOPSIS (Kaya & Kahraman, 2011; Sakthivel et al., 2013), to name some. However, these other methods do not clearly state the advantages aforementioned, and the AQ is difficult to interpret.

Capital Budgeting and Portfolio Optimization in the DoD

Operational and Logistics

- **Inherent Availability** (IA). Measures operational percentage in an ideal support environment per design specifications.

$$IA = \frac{MTBF}{MTBF+MTTR}$$

- **Effective Availability** (EA). Probability a ship’s system is available at any instant during the maximum operational period, accounting for all critical failures, reparable and nonrepairable at sea, and preventive maintenance.

$$EA = 1 - \frac{MTTR}{MTBF+MTTR} - \frac{MDT}{MT} - 0.5\frac{MT}{MTTF}$$

- **Mission Reliability** (MR). Operational Ready Rate (ORR) at the start of a mission compared to its Inherent Reliability (IR).

$$MR = ORR * IR$$

- **Operational Dependability** (OD). Probability a system can be used to perform a specified mission when desired.

$$OD = \frac{MTTF}{MTBF}$$

- **Mean Down Time** (MDT), **Mean Maintenance Time** (MMT), **Logistics Delay Time** (LDT), and their combinations.
- **Achieved Availability** (AA), **Operational Availability** (OA), **Mission Availability** (MA)

Financial and Economic

Cost Deterrence and Avoidance. Soft or shadow-revenue (cost savings) over the economic and operational life of the program or system. Milestone A, B, C.

Traditional Financial Metrics. Net Present Value (NPV), **Internal Rate of Return** (IRR), **Return on Investment** (ROI), and other metrics, as long as there are financial and monetary values.

Budget Constraint. FY Budget limitations and probabilities of budgetary overruns.

Total Ownership Cost (TOC) and **Total Lifecycle Cost** (TLC). Accounting for the cost of developing, producing, deploying, maintaining, operating, and disposing of a system over its entire lifespan. Uses **Work Breakout Structures** (WBS), **Cost Estimating Categories** (CEC), and **Cost Element Structures** (CES).

Knowledge Value Added (KVA). **Monetizing Learning Time**, **Number of Times Executed**, **Automation**, **Training Time**, and **Knowledge Content**.

Strategic and Capability

Multiple value metrics can be determined from Subject Matter Experts (SME): Expected Military Value and Strategic Value

Future Weapon Strategy

Capability Measures (CM). Difficult to quantify and needs SME judgment: Innovation Index, Conversion Capability, Ability to Meet Future Threats; Force Structure (size/units), Modernization (technical sophistication), Combat Readiness, Sustainability; Future Readiness (ability to meet evolving threats, ability to integrate future weapons systems)

Domain Capabilities (DC). Portfolios are divided into different domains, and each domain is optimized separately and then combined into the enterprise level and re-optimized; example domains include Coastal Defense, Anti-Air Surface Warfare, Anti-Surface Warfare, Anti-Submarine Warfare, Naval Strike, Multi-Mission Air Control, Sea Control, Deep Strike, Missile Defense, and so on. Constraints can be added whereby each domain needs to have a minimum amount of capability or systems, and within each domain, different “value” parameters can be utilized.

Optimization Application at PEO-IWS and NAVSEA

The following is a case illustration of portfolio optimization. The values and variables shown are nominal and used for illustration only; they should not be, and have not been, used for making any actual decisions. Nonetheless, all that has to be done in any future real-life applications is to change the names of these options and the values. The analytical process and portfolio methodology remain the same.

The Program Executive Office—Integrated Warfare Systems (PEO-IWS) at the DoD engaged a graduate student team from the Naval Postgraduate School (NPS) to conduct a study to apply the Integrated Risk Management (IRM) method to estimate the value stream and cost savings in its Advanced Concept Build (ACB) for Navy ships, and to provide a set of solid recommendations to its multiple stakeholders going forward. Every few years, Navy destroyers will receive ACB updates to the Aegis ship defense system. These updates include basic hardware enhancement but are mostly software patches and updates for their various capabilities (e.g., ballistic missile defense systems, or BMD 5.X; carry-on cryptologic programs, or CCOPS; weather sensor algorithm updates, or Weather NOW; and many others). The issue is that there are more ACB capabilities than there is budget available for them. The cost to implement new ACB updates can be rather high, and sometimes there are several implementation paths or strategic options to consider in each ACB capability. The task is to model each of these approaches and provide an assessment and recommendation of the best path forward, model each capability, and recommend the best combinatorial portfolio that maximizes the utility to the Navy, both monetary (cost savings, KVA analysis, benefits) and nonmonetary (OPNAV leadership requirements, force readiness, systems integration, obsolescence, etc.).

One of the modeling problems is that the DoD is not in the business of selling its products and services, and, consequently, obtaining a solid set of revenues would prove to be difficult. In such situations, one can resort to using KVA analysis or cost savings approaches. KVA allows us to generate market comparables as proxy variables to determine a shadow price and provide comparable *revenues*. Alternatively, cost savings, or the amount of money that would not have to be spent, can similarly be used as proxy for benefits or revenues in a discounted cash flow model. In addition, there might be competing stakeholders and requirements. For instance, BMD 5.X is very expensive, provides low cost savings (monetary benefits), and is not used often (sometimes not used at all between ACB cycles), but OPNAV and the office of the CNO may want this update to maintain readiness for the fleet and see this upgrade as critical. These considerations need to be modeled.

To summarize, this case illustration requires the following assumptions:

- Each of these ACB capabilities was modeled and compared as a portfolio of static NPV, IRR, ROI, and so forth.
- Using the ROV PEAT software, Monte Carlo risk simulations were run on the main inputs based on the *Air Force Cost Analysis Agency Handbook* (*AFCAA Handbook*) and used to interpret the dynamic results.
- Portfolio optimization algorithms were run using budgetary and project constraints, and efficient frontier analyses based on changing budgets were then executed. Finally, OPNAV requirements, KVA valuation, and other non-economic military values were used to run multi-criteria portfolio optimizations.

The following are the parameters of the ACB program under consideration:

- For all models, we assumed a 10-year time horizon for the cost savings (all future savings past Year 10 after discounting will be assumed to be negligible). The discounting base year is 2017 (Year 0 and Capital Investment is required in 2017), whereas immediate savings and short-term benefits and maintenance savings start in Year 1 (2018). This means Year 10 is 2027.
- Table 1 shows the remaining relevant information needed to run the models. All monetary values are in thousands of dollars.

Table 1. Information Needed to Run the Models

Capability Acronym	Savings Now	Short-Term Benefits	Mainten-ance Savings	Capital Cost	Fixed Cost	Operating Cost	OPNAV Value	Command Value	KVA Value
MH60R	$550	$30	$60	$400	$3	$2	8.1	1.2	9.11
CCOPS	$650	$5	$10	$300	$3	$2	1.27	2.5	1.43
Weather	$700	$35	$10	$350	$3	$2	5.02	7.5	5.65
SSDS	$1,000	$50	$20	$600	$3	$2	8.83	4.5	9.93
BMD	$2,000	$100	$20	$1,000	$3	$2	9.88	9.7	11.11
NIFC-CA	$1,000	$10	$20	$550	$3	$2	3.64	7.4	4.09
SPQ-9B	$2,000	$100	$20	$750	$3	$2	5.27	4.5	5.93
CIWS-CEC	$850	$75	$20	$550	$3	$2	9.8	7.5	11.02
RDDL	$1,500	$125	$20	$750	$3	$2	5.68	7.5	6.39
SM-2 BLK	$1,000	$125	$20	$550	$3	$2	8.29	8.5	9.33

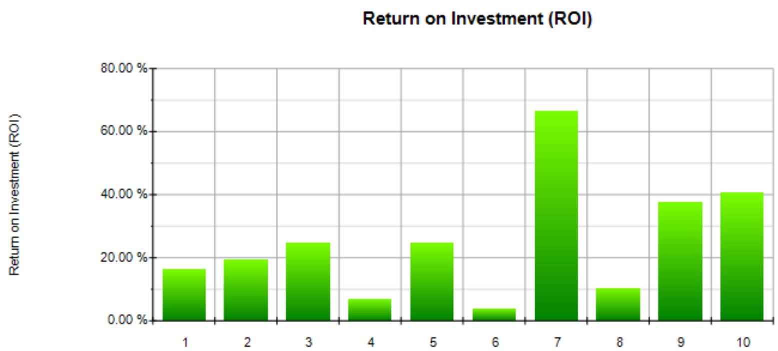
- "Savings Now" is the immediate monetary cost savings benefits obtained by implementing the new upgraded system (e.g., lower overhead requirements, reduced parts and labor requirements). This amount is applied in the first year of the cash flow stream only (Year 1 or 2018) as its effects are deemed to be immediate.
- "Short-Term Benefits" is the savings per year for the first 5 years, stemming from reduction in staffing requirements, but these savings are deemed to be reabsorbed later on. Savings apply from 2018 to 2022.
- "Maintenance Savings" is the savings each year for all 10 years starting in 2018 where system maintenance cost is reduced and saved.
- "Capital Cost" is applied in Year 0 or 2017 as a one-time capital expenditure.
- Assume a "Fixed [Direct] Cost" and constant "[Indirect] Operating Cost" per year for all 10 years starting in 2018. The new equipment upgrades will require some fixed overhead cost and operating expenses to maintain. The idea is that these will be less than the total sum of benefits obtained by implementing the capability.
- Value metrics on Innovation, Capability, Time to Intercept, Warfighting Impact, Health, and Execution were compiled with the help of subject matter experts, and these values are weighted and summarized as "OPNAV" (Innovation, Capability, and Execution Health) and "Command" (Time to Intercept and Warfighting Impact) variables. These are weighted average values of multiple subject matter experts' estimates of the criticality (1–10, with 10 being the highest) of each capability. "KVA" is unit equivalence (this can be multiplied by any market price comparable such

as $1 million per unit or used as-is in the optimization model). These will be used later in the optimization section that follows.

- Tornado analysis was run using ROV PEAT.
- The *AFCAA Handbook* recommendations for uncertainty and risk distributions were used, with the following parameters for simulation:
    - Savings Now and Capital Investment inputs were set using Triangular distributions based on the risk and uncertainty levels perceived by the subject matter experts, or they can be based on a fitting of historical data.
    - Run 10,000 to 1,000,000 simulation trials.
    - The multiple simulated distributions' results were compared using Overlay Charts and Analysis of Alternatives.
- Finally, multiple portfolio optimization models were run in this case illustration using the following parameters:
    - Constraints for the portfolio optimization were a $4,000,000 budget and less than or equal to 7 Opportunities. The portfolio's NPV was maximized.
    - Investment Efficient Frontier was run between $2,500,000 and $5,500,000 with a step of $1,000,000 and no more than 7 Opportunities. The portfolio's NPV was maximized.
    - Another Investment Efficient Frontier was run between $2,500,000 and $5,000,000 with a step of $500,000 and no more than 7 Opportunities. The portfolio's NPV was maximized.
    - Finally, a series of portfolios using the nonmonetary, non-economic military OPNAV, COMMAND, and KVA estimates were applied in the portfolio model but using budgetary constraints. The relevant custom military values and their weighted average values for the portfolio were maximized.

Figure 3 shows the results of a capital budgeting analysis. The 10 programs under consideration were evaluated based on their financial and economic viability. The standard economic metrics such as NPV, IRR, MIRR, ROI, and others are shown. The bar chart provides a visual representation of one of the metrics, whereas the bubble chart shows multiple result metrics at once (e.g., the NPV on the x-axis and the IRR on the y-axis, and size represents NPV with Terminal Value). In this chart, the large-ball programs on the top far right of the chart would be better ranked than smaller-ball projects on the bottom left.

	Economic Results	MH60R	CCOPS	Weather	SSDS	BMD	NIFC-CA	SPQ-9B	CIWS-CEC	RDDL	SM-2 BLK
☑	Net Present Value (NPV)	66,086.45	58,344.30	86,785.26	42,214.01	249,615.61	22,292.73	499,615.61	57,914.81	283,316.41	223,316.41
☑	Net Present Value (NPV) with Terminal Value	83,109.93	59,891.88	88,332.84	46,856.77	254,258.37	26,935.49	504,258.37	62,557.57	287,959.17	227,959.17
☑	Internal Rate of Return (IRR)	36.02%	47.04%	49.72%	31.53%	49.84%	29.20%	93.31%	33.94%	59.65%	58.85%
☑	Modified Internal Rate of Return (MIRR)	26.93%	27.24%	27.80%	25.85%	27.82%	25.50%	31.55%	26.26%	29.07%	29.33%
☑	Profitability Index (PI)	1.17	1.19	1.25	1.07	1.25	1.04	1.67	1.11	1.38	1.41
☑	Return on Investment (ROI)	16.52%	19.45%	24.80%	7.04%	24.96%	4.05%	66.62%	10.53%	37.78%	40.60%
☑	Payback Period (PP)	0.9691	0.6993	0.7277	0.8667	0.7274	0.8255	0.5456	0.9002	0.7036	0.7422
☑	Discounted Payback Period (DPP)	3.2718	0.8741	0.9096	2.8857	0.9093	2.7933	0.6819	2.7933	0.8795	0.9278



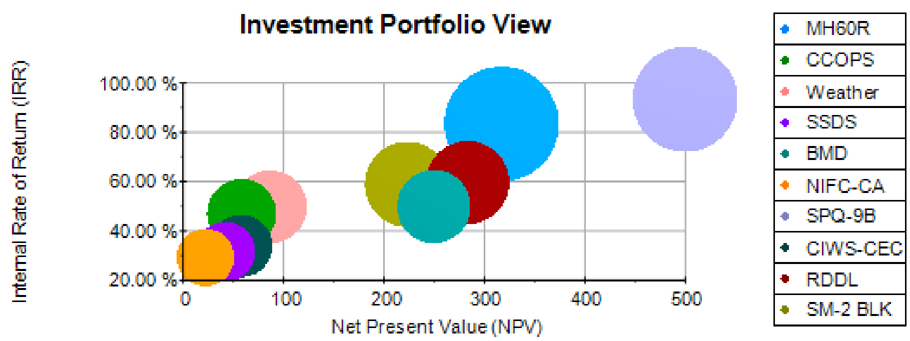


Figure 3. **Capital Budgeting Results Comparison**

NPV	
Rank	Project
1	SPQ-9B
2	RDDL
3	BMD
4	SM-2 BLK
5	Weather
6	MH60R
7	CCOPS
8	CIWS-CEC
9	SSDS
10	NIFC-CA

ROI	
Rank	Project
1	SPQ-9B
2	SM-2 BLK
3	RDDL
4	BMD
5	Weather
6	CCOPS
7	MH60R
8	CIWS-CEC
9	SSDS
10	NIFC-CA

PP	
Rank	Project
1	SPQ-9B
2	CCOPS
3	RDDL
4	BMD
5	Weather
6	SM-2 BLK
7	NIFC-CA
8	SSDS
9	CIWS-CEC
10	MH60R

Figure 4. **Program Rankings**

According to the analysis, the top five recommended ACB capabilities based on Static Portfolio Analysis are SPQ-9B, SM-2 BLK, MH60R, BMD, and RDDL. Figure 4 shows a summary of the ranking. Three main distinctions include the following:

- The highest NPV belongs to SPQ-9B.
- Middle range NPVs belong to BMD, RDDL, and SM-2 BLK.
- The lowest range of NPVs belong to MH-60R, CCOPS, Weather, SSDS, NIFC-CA, and CIWS-CEC.

This distinction is generally true for all other metrics. Data from all metrics are compared to create a numerical ranking from key figures. Although not black and white, this linear ranking helps in decision-making comparative analysis. Figure 5 shows the PDF Curve Overlay where all the programs' simulation results are overlaid on top of each other. Only the SPQ-9B has a positive NPV across all trials. This finding is consistent with the results of the ACB Capability Comparison.

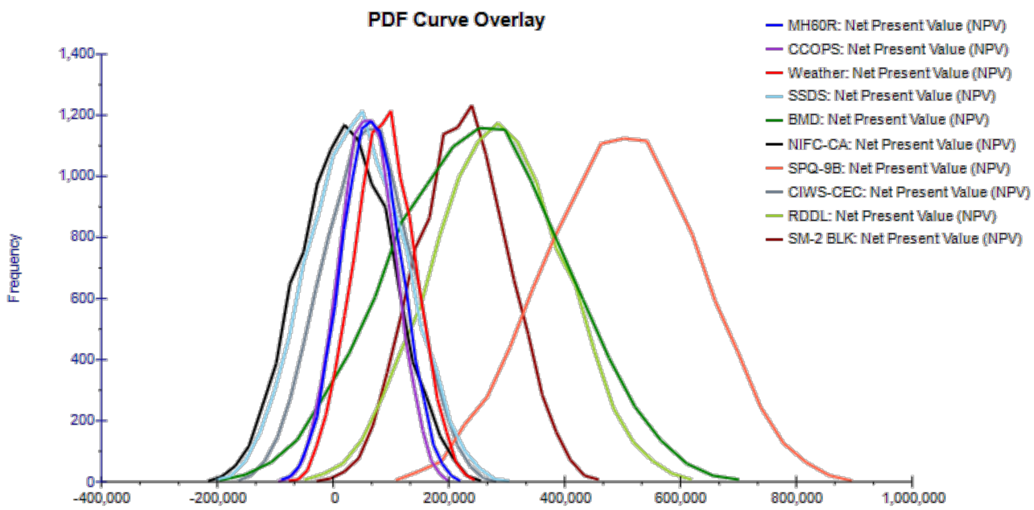


Figure 5. **Comparison of Simulated NPV Probability Distributions**

Figure 6 shows the probability of success of each program. These are currently based on using NPV but can be applied to any non-economic variable. The definition used here is the probability (PROB) of NPV > 0. Based on the values below, (1 – PROB)%, is the probability of failure.

PEAT NPV Probabilities	
100.00%	SPQ-9B
99.94%	SM-2 BLK
99.62%	RDDL
97.61%	Weather
95.41%	BMD
89.90%	MH60R
89.37%	CCOPS
77.58%	CIWS-CEC
70.11%	SSDS
61.34%	NIFC-CA

Figure 6. **Economic Probability of Success**

Figure 7 shows the results of Portfolio Optimization 1, which assumes a budget of $4 million, Portfolio Size: ≤7, and the goal of Maximizing Portfolio NPV. In this simple optimization, the model recommends excluding CCOPS, SSDS, NIFC-CA, and CIWS-CEC from the portfolio. Figure 8 shows Portfolio Optimization 2, which runs an Investment Efficient Frontier. It assumes a budgetary range of $2.5–$5 million with a step size of $500,000. It also assumes a Portfolio Size ≤7 and the explicit goal of Maximizing Portfolio NPV. Weather, SPQ-9B, RDDL, and SM-2 BLK were consistently in the optimal portfolio. Based on budget, other capabilities were recommended. Above $4.5 million, there is no change to the portfolio.

Objective Function	1,408,736
Optimized Constraint 1	7.0000
Optimized Constraint 2	3,800,000
MH60R	1.00
CCOPS	0.00
Weather	1.00
SSDS	0.00
BMD	1.00
NIFC-CA	0.00
SPQ-9B	1.00
CIWS-CEC	0.00
RDDL	1.00
SM-2BLK	1.00

Figure 7. **Portfolio Optimization 1**

Objective Function	1,093,034	1,159,120	1,342,649	1,408,736	1,467,080	1,467,080
Frontier Variable	2,500,000	3,000,000	3,500,000	4,000,000	4,500,000	5,000,000
Optimized Constraint	2,400,000	2,800,000	3,400,000	3,800,000	4,100,000	4,100,000
MH60R	0.00	1.00	0.00	1.00	1.00	1.00
CCOPS	0.00	0.00	0.00	0.00	1.00	1.00
Weather	1.00	1.00	1.00	1.00	1.00	1.00
SSDS	0.00	0.00	0.00	0.00	0.00	0.00
BMD	0.00	0.00	1.00	1.00	1.00	1.00
NIFC-CA	0.00	0.00	0.00	0.00	0.00	0.00
SPQ-9B	1.00	1.00	1.00	1.00	1.00	1.00
CIWS-CEC	0.00	0.00	0.00	0.00	0.00	0.00
RDDL	1.00	1.00	1.00	1.00	1.00	1.00
SM-2BLK	1.00	1.00	1.00	1.00	1.00	1.00

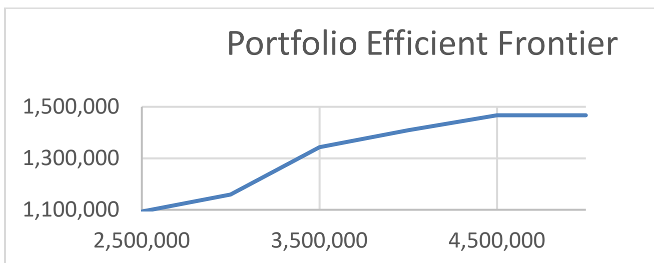


Figure 8. **Portfolio Optimization 2**

Figure 9 shows the results for OPNAV. Similar results were run on COMMAND and KVA objectives. OPNAV Value is a combination of subject matter experts' assessments of Innovation, Capability, and Execution Health metrics. Command Value is the subject matter experts' assessments of Time to Intercept and Warfighting Impact.

Objective Function	40.04	43.68	49.92	53.56	56.87	60.87	64.51
Frontier Variable	2,500,000	3,000,000	3,500,000	4,000,000	4,500,000	5,000,000	5,500,000
Optimized Constraint	2,450,000	3,000,000	3,450,000	4,000,000	4,500,000	4,950,000	5,500,000
MH60R	1.00	1.00	1.00	1.00	1.00	1.00	1.00
CCOPS	0.00	0.00	0.00	0.00	1.00	0.00	0.00
Weather	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SSDS	1.00	1.00	1.00	1.00	1.00	1.00	1.00
BMD	0.00	0.00	1.00	1.00	1.00	1.00	1.00
NIFC-CA	0.00	1.00	0.00	1.00	0.00	0.00	1.00
SPQ-9B	0.00	0.00	0.00	0.00	0.00	1.00	1.00
CIWS-CEC	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RDDL	0.00	0.00	0.00	0.00	1.00	1.00	1.00
SM-2BLK	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Figure 9. **Portfolio Optimization 3 (OPNAV)**

Figure 10 shows a combined view where multiple optimizations were run and compared against one another. Additional constraints can be added as needed, but the case illustration applies a $4 million budget, and no more than seven programs can be chosen at a time. In other words, the following monetary and nonmonetary portfolios were optimized:

- Model 1—Maximize Monetary Values (NPV)
- Model 2—Maximize OPNAV Value (i.e., subject matter experts' assessments of Innovation, Capability, and Execution Health)
- Model 3—Maximize All Weighted Average Nonmonetary Values (this is a percentage weighted average of all nonmonetary military values that are part of the OPNAV and COMMAND variables, as well as any other variables of interest to senior leadership)
- Model 4—Maximize Military Command Value (i.e., subject matter experts' assessments of Time to Intercept and Warfighting Impact)
- Model 5—Maximize KVA Value

As seen in Figure 10, these five portfolios are combined into a matrix that shows the count of GO decisions. Clearly, for a decision maker, the lowest-hanging fruits would be to execute the programs starting with the highest count. For instance, Weather, BMD, and SM-2BLK would be considered the highest priority, as, regardless of the point of view and stakeholder under consideration, these programs have always been chosen.

Model	1. NPV	2. OPNAV	3. W/AVG	4. COMMAND	5. KVA	Count
Objective	1,408,735.73	51.16	53.56	48.10	53.56	
Budget Constraint	3,800,000	4,000,000	4,000,000	3,750,000	4,000,000	
Program Constraint	6	7	7	6	7	
MH60R	1.00	1.00	1.00	0.00	1.00	4
CCOPS	0.00	0.00	0.00	0.00	0.00	0
Weather	1.00	1.00	1.00	1.00	1.00	5
SSDS	0.00	1.00	1.00	0.00	1.00	3
BMD	1.00	1.00	1.00	1.00	1.00	5
NIFC-CA	0.00	1.00	1.00	1.00	1.00	4
SPQ-9B	1.00	0.00	0.00	0.00	0.00	1
CIWS-CEC	0.00	1.00	1.00	1.00	1.00	4
RDDL	1.00	0.00	0.00	1.00	0.00	2
SM-2BLK	1.00	1.00	1.00	1.00	1.00	5

Figure 10. **Portfolio Optimization 7 (Combined View)**

Conclusions and Recommendations

The analytical methods illustrated in the case study apply stochastic risk-based Monte Carlo simulations to generate tens of thousands to millions of scenarios and algorithmic portfolio optimization by applying economic and non-economic military values. The methods are objective, verifiable, replicable, and extensible and can be easily modified to incorporate additional constraints and limitations (e.g., manpower, force mix, minimum capability requirements, domain-specific requirements, cross-domain needs, etc.).

It is recommended that any follow-on research incorporate the following items:

- Apply the methods to actual programs with real-life data and assumptions, with SME estimates.
  - Create new or evaluate existing concepts of military value. These will incorporate
    - Data validity tests using applied statistical tests (from basic linear and nonlinear correlations to econometric models and nonparametric hypothesis tests). These are applied over time to identify if the collected data are valid and actually describe what the researcher wants or expects the data to describe. In other words, are the data collected valid, accurate, and precise?
    - Big data analysis—trying to find patterns and analytical relationships in large data sets.
    - Historical data to perform backcasting (back testing historical data to known historical events).
    - Tweaking and creating lighthouse events and programs in the past, assigning critical value metrics to these events and programs, and using these as guideposts for generating future SME estimates.
    - Creating more exact definitions and methods for SME assumptions that allow for collecting a more objective and defensible data set.
- Utilize multi-objective optimization. Interdependencies and competing stakeholder needs (e.g., Congress versus Office of the Secretary of Defense [OSD] and other external stakeholders) need to be considered. These competing objectives need to be reconciled to determine a Pareto optimal portfolio.
- Evaluate analytical hierarchical processes, multi-objective optimization, and other algorithms and compare the results.
- Within the portfolio, model and account for risks of cost and budget overruns as well as delivery delays using risk-based simulations.

To summarize, based on the research performed thus far, the researcher concludes that the methodology has significant merits and is worthy of more detailed follow-on analysis. It is therefore recommended that the portfolio optimization methodology outlined in this research be applied on a real case study facing the U.S. Navy, using actual data and tracking the project’s outcomes over time. The approach described does not necessarily have to be performed in lieu of existing methods, but in conjunction with them. After all, if the Navy and the DoD are spending hundreds of billions of dollars on capability upgrades, the least that can be done is to have another point of view, an analytically robust and verifiable way of looking at the decision portfolios. The more information decision makers have, the better informed they will be and the better their decision outcomes will be.

References

Afsordegan, A., Sánchez, M., Agell, N., Zahedi, S., & Cremades, L. (2016). Decision making under uncertainty using a qualitative TOPSIS method for selecting sustainable energy alternatives. *International Journal of Environmental Science and Technology, 13*(6), 1419–1432.

Al-Sharrah, G. (2010). Ranking using the Copeland Score: A comparison with the Hasse Diagram. *Journal of Chemical Information and Modeling, 50*, 785–791.

Asbeck, E., & Haimes, Y. Y. (1984). The partitioned multiobjective risk method. *Large Scale Systems, 6*(1), 13–38.

Barrett, G. F., & Donald, S. G. (2003). Consistent tests for stochastic dominance. *Econometrica, 71*(1), 71–104.

Behzadian, M., Kazemzadeh, R. B., Albadvi, A., & Aghdasi, M. (2010). PROMETHEE: A comprehensive literature review on methodologies and applications. *European Journal of Operational Research, 200*(1), 198–215.

Bennett, S. (2017, July). *Optimization aids in quest for government efficiency.* Retrieved from https://federalnewsradio.com/commentary-analysis/2017/07/optimization-aids-in-the-quest-for-government-efficiency/

Bodie, Z., Kane, A., & Marcus, A. J. (2009). *Investments*. New York, NY: McGraw-Hill.

Botkin, B. (2007, June). *Applying financial portfolio analysis to government program portfolios* (Master's thesis, Naval Postgraduate School). Retrieved from https://calhoun.nps.edu/handle/10945/3481

Boucher, C. M., Danielsson, J., Kouontchou, P. S., & Maillet, B. B. (2014). Risk models-at-risk. *Journal of Banking and Finance, 44,* 72–92.

Bouyssou, D., Marchant, T., Pirlot, M., Tsoukias, A., & Vincke, P. (2006). *Evaluation and decision models with multiple criteria* (Operational Research and Management Sciences). New York, NY: Springer's International Series.

Brans, J. P., & Vincke, P. H. (1985). A preference ranking organization method: The PROMETHEE method for multiple criteria decision making. *Management Science, 31*(6), 647–656.

Brans, J.-P., & Mareschal, B. (2005). Multicriteria decision aid. The PROMETHEE-GAIA solution in multiple criteria decision analysis: State of the art surveys. *International Series in Operations Research & Management Science, 78*, 163–186.

Brito, A. J., de Almeida, A. T., & Mota, C. M. M. (2010). A multicriteria model for risk sorting of natural gas pipelines based on ELECTRE TRI integrating utility theory. *European Journal of Operational Research, 200*(3), 812–821.

Bruggemann, R., & Carlsen, L. (2011). An improved estimation of averaged ranks of partial orders. *Match Communications in Mathematical and in Computer Chemistry, 65,* 383–414.

Bruggemann, R., Schwaiger, J., & Negele, R. D. (1995). Applying Hasse diagram technique for the evaluation of toxicological fish tests. *Chemosphere, 30,* 1767–1780.

Bruggemann, R., Bücherl, C., Pudenz, S., & Steinberg, C. (1999). Application of the concept of partial order on comparative evaluation of environmental chemicals. *Acta hydrochimica et hydrobiologica, 27,* 170–178.

Burk, R. C., & Parnell, G. S. (2011). Portfolio decision analysis: Lessons from military applications. In A. Salo, J. Keisler, & A. Morton (Eds.), *Portfolio decision analysis* (pp. 333–357). New York, NY: Springer.

Chong, Y. Y. (2004). *Investment risk management.* Oxford, UK: John Wiley & Sons.

Cliville, V., Berrah, L., & Mauris, G. (2007). Quantitative expression and aggregation of performance measurements based on the MACBETH multi-criteria method. *International Journal of Production Economics, 105*(1), 171–189.

Corrente, S., Figueira, J. R., & Greco, S. (2014). The SMAA-PROMETHEE method. *European Journal of Operational Research, 239,* 514–522.

Costa, C., De Corte, J. M., & Vansnick, J. C. (2012). MACBETH. *International Journal of Information Technology & Decision Making, 11*(2), 359–387.

Davendralingam, N., & DeLaurentis, D. (2015, May). A robust portfolio optimization approach to system of system architectures. *Systems Engineering, 18*(3), 269–283.

De Loof, K., De Baets, B., & De Meyer, H. (2011). Approximation of average ranks in posets. *Match Communications in Mathematical and in Computer Chemistry, 66,* 219–229.

Department of Defense (DoD). (2005, October 10). *Information technology portfolio management* (DoD Directive 8115.01). Retrieved from http://www.dtic.mil/whs/directives/corres/pdf/811501p.pdf

Department of Defense (DoD). (2006). (DoD Directive 8115.bb).

Department of Defense (DoD). (2017, June 9). *DoD space enterprise governance and principal DoD space advisor (PDSA)* (DoD Directive 5100.96). Retrieved from https://fas.org/irp/doddir/dod/d5100_96.pdf

Desai, S., Bidanda, B., & Lovell, M. R. (2012). Material and process selection in product design using decision-making technique (AHP). *European Journal of Industrial Engineering, 6*(3), 322–346.

Dickinson, M. W., Thornton, A. C., & Graves, S. (2001, November). Technology portfolio management: Optimizing interdependent projects over multiple time periods. *IEEE Transactions on Engineering Management, 48*(4), 518–527.

Dorini, G., Kapelan, Z., & Azapagic, A. (2011). Managing uncertainty in multiple-criteria decision making related to sustainability assessment. *Clean Technologies and Environmental Policy, 13,* 133–139.

Fabozzi, F. J. (2010). *An introduction to financial markets, business finance, and portfolio management.* Hoboken, NJ: John Wiley & Sons.

Flynn, B., & Field, J. (2006, January 1). Transformation of analytical tools: Using portfolio analysis techniques in defense applications. *Armed Forces Comptroller.* Retrieved from https://www.thefreelibrary.com/Transformation+of+analytical+tools%3A+using+portfolio+analysis+...-a0145158636

General Accounting Office (GAO). (1997, February). *Assessing risk and returns: A guide for evaluating federal agencies' IT investment decision-making* (GAO/AIMD-10.1.13). Retrieved from http://www.gao.gov/special.pubs/ai10113.pdf

Ghafghazi, S., Sowlati, T., Sokhansanj, S., & Melin, S. (2010). A multicriteria approach to evaluate district heating system options. *Applied Energy, 87*(4), 1134–1140.

Goumas, M., & Lygerou, V. (2000). An extension of the PROMETHEE method for decision making in fuzzy environment: Ranking of alternative energy exploitation projects. *European Journal of Operational Research, 123*(3), 606–613.

Government Accountability Office (GAO). (2007). *Best practices: An integrated portfolio management approach to weapon system investments could improve DOD's acquisition outcomes* (GAO-07-388). Washington, DC: Author.

Greiner, M. A., McNutt, R. T., Shunk, D. L., & Fowler, J. W. (2001). Selecting military weapon systems development portfolios: Challenges in value measurement. In *Proceedings of the Portland International Conference on Management of Engineering and Technology, 2001 (PICMET '01)* (pp. 403–410). doi:10.1109/PICMET.2001.952153

Haimes, Y. Y. (2009). *Risk modeling, assessment, and management*. Hoboken, NJ: John Wiley & Sons.

Hopkins, A. (2011). Risk-management and rule-compliance: Decision-making in hazardous industries. *Safety Science, 49,* 110–120.

Hyde, K. M., & Maier, H. R. (2006). Distance-based and stochastic uncertainty analysis for multi-criteria decision analysis in Excel using Visual Basic for applications. *Environmental Modelling & Software, 21,* 1695–1710.

Hyde, K. M., Maier, H. R., & Colby, C. B. (2004). Reliability-based approach to multicriteria decision analysis for water resources. *Journal of Water Resource Planning Management, 130*(6), 429–438.

Janiga, M., & Modigliani, P. (2014, November–December). Think portfolios, not programs. *Defense AT&L Magazine,* 12–16. Retrieved from www.dtic.mil/get-tr-doc/pdf?AD=ADA612303

Jorion, P. (2007). *Value at risk: The new benchmark for managing financial risk*. New York, NY: McGraw-Hill.

Kaya, T., & Kahraman, C. (2011). Multicriteria decision making in energy planning using a modified fuzzy TOPSIS methodology. *Expert Systems with Applications, 38*(6), 6577–6585.

Levy, H. (2006). *Stochastic dominance: Investment decision making under uncertainty.* Berlin, Germany: Springer Science+Business Media.

Mansini, R., Ogryczak, W., & Speranza, M. G. (2007). Conditional value at risk and related linear programming models for portfolio optimization. *Annals of Operations Research, 152*(1), 227–256.

Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance,* 77–91.

Matos, M. A. (2007). Decision under risk as a multicriteria problem. *European Journal of Operational Research, 181*(3), 1516–1529.

Mun, J. (2015). *Modeling risk: Applying Monte Carlo risk simulation, strategic real options, stochastic forecasting, portfolio optimization, data analytics, business intelligence, and decision modeling*. CA: Thomson-Shore and ROV Press.

Mun, J. (2016). *Real options analysis: Tools and techniques for valuing strategic investments and decisions with integrated risk management and advanced quantitative decision analytics*. CA: Thomson-Shore and ROV Press.

Mun, J., Ford, D., & Housel, T. (2012, October 10). *Naval ship maintenance: An analysis of the Dutch shipbuilding industry using the knowledge value added, systems dynamics, and integrated risk management methodologies* (NPS-AM-12-204). Retrieved from https://calhoun.nps.edu/bitstream/handle/10945/33851/NPS-AM-12-204.pdf?sequence=1

Mun, J., Hernandez, E., & Rocco, C. (2016). A combined lexicographic average rank approach for evaluating uncertain multi-indicator matrices with risk metrics. In M. Fattore & R. Brüggemann (Eds.), *Partial order concepts in applied sciences*. Retrieved from http://www.springer.com/la/book/9783319454191

Pulido, F. J., Mandow, L., & de la Cruz, J. L. P. (2014). Multiobjective shortest path problems with lexicographic goal-based preferences. *European Journal of Operational Research, 239*(1), 89–101.

Restrepo, G., Brüggemann, R., Weckert, M., Gerstmann, S., & Frank, H. (2008). Ranking patterns, an application to refrigerants. *Match Communications in Mathematical and in Computer Chemistry, 59,* 555–584.

Rocco, C. M., & Tarantola, S. (2014). Evaluating ranking robustness in multi-indicator uncertain matrices: An application based on simulation and global sensitivity analysis. In R. Brüggemann, L. Carlsen, & J. Wittmann (Eds.), *Multi-indicator systems and modelling in partial order* (pp. 275–292). New York, NY: Springer.

Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk, 2,* 21–41.

Roy, B. (1968). Classement et choix en presence de points devue multiples (la methode ELECTRE). *Revue d'Informatique et de recherché opérationelle, 6*(8), 57–75.

Roy, B. (1996). *Multicriteria methodology for decision aiding.* Berlin, Germany: Springer Science+Business Media.

Saaty, T. (2013). The modern science of multicriteria decision making and its practical applications: The AHP/ANP approach. *Operations Research, 61*(5), 1101–1118.

Saban, D., & Sethuraman, J. (2014). A note on object allocation under lexicographic preferences. *Journal of Mathematical Economics, 50,* 283–289.

Sakthivel, G., Ilangkumaran, M., Nagarajan, G., & Shanmugam, P. (2013). Selection of best biodiesel blend for IC engines: An integrated approach with FAHP-TOPSIS and FAHP-VIKOR. *International Journal of Oil Gas and Coal Technology, 6*(5), 581–612.

Scarlat, E., Chirita, N., & Bradea, I. A. (2012). Indicators and metrics used in the enterprise risk management (ERM). *Economic Computation and Economic Cybernetics Studies and Research, 46*(4), 5–18.

Sidiropoulos, L., Sidiropoulou, A., & Lalagas, S. (2014). Defense portfolio analysis. *Journal of Computations & Modelling, 4*(1), 327–347.

Szolgayová, J., Fuss, S., Khabarov, N., & Obersteiner, M. (2011). A dynamic CVaR-portfolio approach using real options: An application to energy investments. *European Transactions on Electrical Power, 21*(6), 1825–1841.

Tarantino, A. (2008). *Governance, risk, and compliance handbook: Technology, finance, environmental, and international guidance and best practices.* Hoboken, NJ: John Wiley & Sons.

Tavana, M., Behzadian, M., Pirdashti, M., & Pirdashti, H. (2013). A PROMETHEE-GDSS for oil and gas pipeline planning in the Caspian Sea basin. *Energy Economics, 36,* 716–728.

Vascik, P., Ross, A., & Rhodes, D. (2015). *A method for exploring program and portfolio affordability tradeoffs under uncertainty using Epoch-Era Analysis: A case application to carrier strike group design*. Paper presented at 12th Annual Acquisition Research Symposium, Naval Postgraduate School, Monterey, CA.

Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior.* Princeton, NJ: Princeton University Press.

Weber, C. A., Balut, S. J., Cloos, J. J., Frazier, T. P., Hiller, J. R., Hunter, D. E., . . . Tran, D. (2003, October). *The acquisition portfolio schedule costing/optimization model: A tool for analyzing the RDT&E and production schedules of DoD ACAT I systems* (IDA Document D-2835). Alexandria, VA: Institute for Defense Analyses. Retrieved from www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA421123

Yaman, F., Walsh, T. J., Littman, M. L., & Desjardins, M. (2011). Democratic approximation of lexicographic preference models. *Artificial Intelligence, 175*(78), 1290–1307.

An Approximate Dynamic Programming Approach for Weapon System Financial Execution Management

Erich Morman—is an assistant professor at the Naval Postgraduate School in the Defense Resources Management Institute (DRMI). He received a BA in mathematics and economics (1996) from the University of Dayton, and an MS in systems engineering and operations research (2003) and a PhD in systems engineering and operations research (2013) from George Mason University. Prior to joining DRMI, Dr. Morman spent 17 years with the Missile Defense Agency (MDA) supporting various weapon system program offices in the area of Planning, Programming, Budgeting, and Execution (PPBE). While working for the MDA, he completed his dissertation, which focused on applications of reinforcement learning algorithms to improve fiscal year cash allocation decisions. Additionally, Dr. Morman spent nearly two years as a support contractor for the DoD's Suicide Prevention Office (DSPO) conducting measures of effectiveness and data analysis. [edmorman@nps.edu]

Jefferson Huang—is an assistant professor in the Operations Research Department at the Naval Postgraduate School. Prior to joining NPS in 2018, Dr. Huang was a postdoctoral associate with the School of Operations Research and Information Engineering at Cornell University. He received a PhD in applied mathematics in 2016 from Stony Brook University, with a dissertation on the computational complexity of algorithms for Markov decision processes. Huang also holds an MS in transportation from the Massachusetts Institute of Technology and a BS in civil and environmental engineering from the University of California at Berkeley. His research interests include applied probability, optimization under uncertainty, and operations management. [jefferson.huang@nps.edu]

Abstract

A fiscal year (FY) starts on 1 October and runs to 30 September of the following year. During this time frame, a Department of Defense (DoD) weapon system program office allocates financial resources to vendors working various projects. As the calendar gets closer to 30 September, program offices undergo a FY closeout review. During this time, considerable energy is invested in assessing cash utilization levels (disbursements) and taking corrective actions related to projects that are not sufficiently spending their allocated funds. Since the DoD operates under a use-or-lose budgetary environment, projects that are behind in meeting their spending goals are at risk of losing a portion, if not all, of their unutilized allocated funding. This financial closeout process is an annual tradition that involves considerable time and resources. The purpose of this research is to assess the viability of using approximate dynamic programming (ADP) to create and manage financial execution plans throughout the FY. The research examines the difficulties of adopting ADP as an execution management tool as well as the potential this methodology has for reducing the total amount of unspent money a program office has on hand during the FY closeout period.

Introduction

As with most public sector organizations, Department of Defense (DoD) money that is managed by weapon system program offices contains an expiration point. Dollars not spent or utilized within a defined timeframe are taken away and are no longer available as a resource to pay for support projects or activities. Organizations that manage money with this type of constraint are operating with what is informally referred to as a use-or-lose budget. Functioning under this framework, weapon system program managers and their financial officers must consider how to strategically allocate funding over an annual time horizon that balances between the immediate day-to-day cash allocation decisions and the aggregate long-term impact these decisions will have on the program office's fiscal year (FY) financial closeout position.

Although a weapon system program manager (PM) is ultimately tasked with efficiently and effectively delivering a weapon system platform or capability, it is the responsibility of their business financial manager (BFM) to ensure that the flow of financial resources is conducted in a manner that complements the PM's mission. While analyzing FY2012 DoD budget data, Conley et al. (2014) point out that the rate of spending as measured by expenditure rates across the DoD was declining for several years prior. The report highlights how spending benchmarks issued by the Office of the Secretary of Defense (OSD) are based on 30 years of financial execution history. Theoretically, this means that DoD spending benchmarks are correlated to the work schedules and associated spending patterns that are emblematic of the acquisition efforts within a typical DoD weapon system program office. However, the actual acquisition experience for each weapon system program is unique and always evolving, compounding the difficulties faced by PMs, BFMs, and their staff.

Serving as additional evidence that there are cash flow problems within the DoD, a 2013 Defense Acquisition University (DAU) study provides a summary of survey results from 229 DoD personnel that responded to questions regarding the top challenges they see as factors impeding cash flow and hindering the ability of a program office to meet OSD's spending benchmarks (Tremaine & Seligman, 2013). In their report, they provide a summary of key factors that program offices indicate are barriers to improving spending efficiency. The report highlights a myriad of growing challenges and endogenous issues that DoD personnel working in a weapon system program office contend with on a routine basis. The following is a short list of standard problems that are impediments and bottlenecks to efficiently allocating and spending money in a timely manner:[1]

- The more routine use of continuing resolution authorities (CRAs) by Congress to issue yearly budgets through multiple installments
- Congressional marks or program cuts
- Delays in contract negotiations and awards
- A high volume of contract modifications related to warfighter requirement changes
- Constant rotation or shortages of key program office personnel
- Complications with getting funding documents issued and approved in a timely manner
- An inability to obtain timely data on contractor outlays or expenditure positions

[1] The list includes items from the survey results of Termaine & Seligman (2013) as well as factors mentioned in Cooley & Ruhm (2014). Some of the additional items contained in the list are from the author's first-hand knowledge of working directly for DoD weapon system program offices for 15 years.

A question to consider is whether or not DoD financial execution performance has improved at all over the past five to six years since the publication of the Conley et al. (2014) and Tremaine & Seligman (2013) reports. However, it is difficult to find open-source data or information that suggests that weapon system expenditure performance is improving. Rather, popular press headlines that currently occur during the traditional annual FY closeout period suggest that efficient cash flow remains a problem and is becoming worse (Mehta, 2018; Moritz-Rabson, 2018).

In this study, we look to the use of ADP as a solution approach to the financial execution problem for weapon system program offices. Fundamentally, the financial execution problem confronted by program offices is a dynamic sequential resource allocation problem, where the resource variable in question is the amount of cash that is committed to projects on a daily basis. Although use-or-lose budget resource problems are not explicitly addressed, there are a number of publications that highlight ADP's applicability to solving other types of resource allocation problems. ADP contains a number of features that make it an attractive tool for the financial execution challenges of weapon system program offices that are operating with use-or-lose budgets. First, ADP is a well-established prescriptive analytical tool. It is designed to create a sequential decision-making policy. In the case of the financial execution problem, a program office must consider a cash allocation policy over a fiscal year that provides an appropriate level or installments of funding to projects that minimize the amount of vulnerable end-of-year money. Second, ADP "learns" a financial execution policy by iteratively interacting with the decision environment. Lastly, the ADP methodology can be adjusted to incorporate the uncertainty and stochastic information of separate program offices. In this manner, ADP can be specialized for individual program offices to more readily account for their unique financial challenges and circumstances.

The remainder of the paper is organized as follows. In the following section, we provide a short literature survey that includes background and context information on ADP. After that is an overview of the DoD financial execution process. We provide a dynamic programming formulation of the use-or-lose program office budget problem in the following section, and then a numerical example. The final section includes conclusions and directions for future research.

Literature Review

Dynamic programming has a history as a mathematical tool for modeling and solving sequential decision-making problems that traces back to the 1950s and early 1960s. A number of the seminal works at this time that set the foundations for dynamic programming include publications by Bellman (1954), Bellman (1957), Howard (1960), and Bellman and Dreyfus

(1962). Since then, the dynamic programming field has grown to include newer techniques such as ADP that address the inherent difficulties with using traditional dynamic programming solution approaches and the complexities of real-world problem structures. Unfortunately, as pointed out by Powell (2009), the various sub-communities working to advance dynamic programming concepts use different vernacular and notional symbols to essentially express the same fundamental ideas. For further discussion on relationships between ADP and artificial intelligence, see, for example, Powell (2010), Tsitsiklis (2010), and Gosavi (2009).

Overview of DoD Financial Execution

A program office acquisition environment is interwoven with a number of important schedules and critical timelines. The more prominent time-oriented processes that a PM must adhere to include (1) a schedule for budget preparation, review, submission, and approval; (2) the timeline for prime contract awards or modifications which can include periods for request for proposals (RFPs), time for proposal preparations and responses to proposal questions, review and assessment of submitted proposals, and time for resolving a possible bid protest after a contract award is announced; (3) the fiscal year calendar that involves mid-year financial reviews, end-of-year closeout reviews, and even possible monthly spending benchmark reviews; and (4) programmatic schedules with well-defined milestone review thresholds. Unfortunately, these separate process schedules do not always complement one another or align cohesively in a streamlined method that facilitates the delivery of a weapon system platform.

> It's tough to manage an event-driven program in a schedule-driven budget.
> —William T. Cooley (Cooley & Ruhm, 2014)

The FY calendar includes important start dates (1 October) and stop dates (30 September) that are necessary for comptrollers and budgetary personnel to track and manage funding that supports the acquisition of a weapon system. However, the fact that the fiscal year calendar starts on 1 October and ends on 30 September has little to do with timing for parts, materials, test events, or other programmatic activities necessary for fielding a weapon system. Nonetheless, the reality is that these dates have considerable influence on when funding is available and the timing of financial commitment actions or cash allocation decisions a program office is likely to take. In the remainder of this section, we take a closer look at different aspects of the DoD financial execution environment: stages of a transaction, appropriation categories, and spending timelines and benchmarks.

Stages of a Transaction

Once a cash determination is made to allocate money for a particular project, the transaction moves through formal DoD financial execution stages. The flow chart in Figure 1 from the Army's financial management operations field manual provides the order of execution stages (Department of the Army, 2014). This financial execution process is the standard used throughout the DoD. The first step is the authorization of a funding transaction. After the appropriate authorization documentation is completed and signed, the funding is said to be committed. Committing dollars is an important first step in the execution process that occurs prior to the actual movement of money to a recipient. This initial stage serves as a cross-check that helps to avoid anti-deficiency violations that result when funding is issued to a contractor or service provider in excess of what is available. Committed dollars are then used to prepare formal and legal contractual obligations between the weapon system program office and a hired vendor. The obligation creates a legal reservation of funds and represents the allocated funds that are available for paying for a project. As work is performed on the project, expenses are accrued. A vendor then provides invoices to the program office for which payment is issued. Once payment is received by the vendor or contractor the funding is considered disbursed. The terms outlays and expenditures can also be used to refer to disbursed funding. Throughout the course of a fiscal year, the financial execution status of a weapon system program office is routinely tracked and assessed. The basis of measurement used to evaluate fiscal year execution is the amount of overall budget that currently resides in each of these respective stages. However, significant attention is paid particularly to the obligation and expenditure positions of a weapon system program. To highlight the magnitude of the amount of funding that moves through this process each year, the Defense Finance and Accounting Service (DFAS) reported that it paid out $554 billion in disbursements for FY2017 and $558 billion in disbursements for FY2018 (DFAS, n.d.).

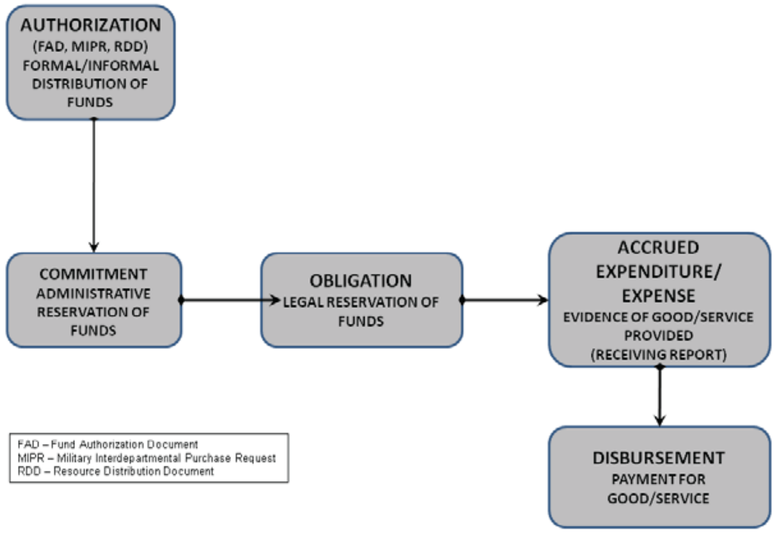


Figure 1. **Stages of a Transaction**

Appropriation Categories

An additional factor that contributes to the complexity of financial execution at the DoD is the agency's use of different appropriation categories. When creating a budget for a weapon system program office, similar types of projects or work are categorized together in the same appropriation category. Furthermore, the activities of the separate appropriation categories are funded with unique types of money or with what is more commonly referred to as different "colors"-of-money. These categorizations of activities and funding allow regulators, comptrollers, and other oversight officials to have better insight on how money is spent and on what activities constitute most of the defense budget. However, weapon system program managers and their financial staff are now encumbered with the additional responsibility of managing their programs to correct appropriation categories and must account for these delineations when making decisions related to budget preparations, funding requests, and cash allocations. The following is a short summary of the more common appropriations:[2]

- Military Personnel (MILPERS): Funds salary and benefits of military personnel to include active duty, reserve, as well as DoD government civilian employees.
- Research, Development, Test, and Evaluation (RDT&E): Funds projects and initiatives that support program research, technology development, engineering development, manufacturing development, and programmatic test events.

[2] More extensive details regarding what each appropriation category funds can be found in the DoD Financial Management Regulation (FMR) 7000.14-R, Volumes 2a and 2b.

- Procurement: Funds the purchase of military equipment and weapon systems to include the production and fielding costs associated with the assets.
- Operation and Maintenance (O&M): Funds activities directly related to the operations, servicing, and upkeep of fielding military systems and platforms.
- Military Construction (MILCON): Funds construction projects related to buildings, facilities, and property improvement efforts that directly support the operations and maintenance of a fielded weapon system.

Spending Timelines and Benchmarks

Each of the DoD's appropriation categories are subject to guidance regarding the amount of time allowable for moving money through the different stages of a transaction, as described previously. Particular attention is paid to the rate at which funding is obligated and disbursed. Within DoD financial execution, regardless of the appropriation category, money exists in two possible periods: (1) the current period and (2) the expired period. Weapon system program offices must ensure all new obligation actions occur during the current period. The length of the current period is different for each "colors"-of-money or appropriation category. O&M and MILPERS have the shortest current period at one year, RDT&E funding has a two-year current period time frame, the current period for procurement funding can range between three to five years, and military construction has the longest current period at five years. Once the current period for an appropriation has lapsed, the funding moves into an expired period. Irrespective of the appropriation, the expired period lasts for five years once the current period is over. During the expired period, no new obligations are allowed. However, funds that were already obligated during the current period can be expensed and recorded as an outlay. Once the expired period has lapsed, the funding is considered canceled and can no longer be used for obligations or expenditures.

The current period and expired period set strict cash flow stopping points; however, the cash flow performance of a weapon system program office is judged on a continual basis. If for any reason it appears that a program office is falling too far behind in its ability to effectively issue and spend money, it runs the risk of being perceived as having too large of a budget for its mission. Comptroller officials and leadership at a more senior level to the program office have the authority to reallocate funding from underperforming program offices to other program offices or activities. Thus, there is an imperative for program offices to maintain constant vigilance of their financial execution position and to make quality cash allocations to contracts and vendors that will expeditiously accrue and expense their funding allotments.

From the perspective of purely protecting funds in a use-or-lose environment, the sooner money moves through the complete stages of a transaction, the better it is for the program office. Unfortunately, programmatic activities and acquisition initiatives that require funding are not always conveniently timed or necessarily ready to receive funds in a manner that allows program offices to keep pace with the spending benchmarks in Table 1. Furthermore, if a program office expends funding too quickly, it runs the risk of running over its budget before the fiscal year is over. Much like underutilizing funds, overrunning a budget is another financial execution position that a program office needs to avoid and must take into consideration when making cash allocation determinations.

Table 1. DoD Spending Guidance by Appropriation

		RDT&E		Procurement		O&M		MILCON	
	Month	**Obl.**	**Exp.**	**Obl.**	**Exp.**	**Obl.**	**Exp.**	**Obl.**	**Exp.**
First Year of Availability	Oct	7.5%	4.6%	6.7%	N/A	8.3%	6.3%	5.4%	1.2%
	Nov	15.0%	9.2%	13.3%	N/A	16.7%	12.5%	10.8%	2.3%
	Dec	22.5%	13.8%	20.0%	N/A	25.0%	18.8%	16.3%	3.5%
	Jan	30.0%	18.3%	26.7%	N/A	33.3%	25.0%	21.7%	4.7%
	Feb	37.5%	22.9%	33.3%	N/A	41.7%	31.3%	27.1%	5.8%
	Mar	45.0%	27.5%	40.0%	N/A	50.0%	37.5%	32.5%	7.0%
	Apr	52.5%	32.1%	46.7%	N/A	58.3%	43.8%	37.9%	8.2%
	May	60.0%	36.7%	53.3%	N/A	66.7%	50.0%	43.3%	9.3%
	Jun	67.5%	41.3%	60.0%	N/A	75.0%	56.3%	48.8%	10.5%
	Jul	75.0%	45.8%	66.7%	N/A	83.3%	62.5%	54.2%	11.7%
	Aug	82.5%	50.4%	73.3%	N/A	91.7%	68.8%	59.6%	12.8%
	Sep	90.0%	55.0%	80.0%	N/A	100.0%	75.0%	65.0%	14.0%
Second Year of Availability	Oct	90.8%	57.9%	80.8%	N/A	100.0%	77.1%	67.1%	18.1%
	Nov	91.7%	60.8%	81.7%	N/A	100.0%	79.2%	69.2%	22.2%
	Dec	92.5%	63.8%	82.5%	N/A	100.0%	81.3%	71.3%	26.3%
	Jan	93.3%	66.7%	83.3%	N/A	100.0%	83.3%	73.3%	30.3%
	Feb	94.2%	69.6%	84.2%	N/A	100.0%	85.4%	75.4%	34.4%
	Mar	95.0%	72.5%	85.0%	N/A	100.0%	87.5%	77.5%	38.5%
	Apr	95.8%	75.4%	85.8%	N/A	100.0%	89.6%	79.6%	42.6%
	May	96.7%	78.3%	86.7%	N/A	100.0%	91.7%	81.7%	46.7%
	Jun	97.5%	81.3%	87.5%	N/A	100.0%	93.8%	83.8%	50.8%
	Jul	98.3%	84.2%	88.3%	N/A	100.0%	95.8%	85.8%	54.8%
	Aug	99.2%	87.1%	89.2%	N/A	100.0%	97.9%	87.9%	58.9%
	Sep	100.0%	90.0%	90.0%	N/A	100.0%	100.0%	90.0%	63.0%
Third Year of Availability	Oct	100.0%	90.8%	90.8%	N/A	100.0%	100.0%	90.4%	65.5%
	Nov	100.0%	91.7%	91.7%	N/A	100.0%	100.0%	90.8%	68.1%
	Dec	100.0%	92.5%	92.5%	N/A	100.0%	100.0%	91.3%	70.6%
	Jan	100.0%	93.3%	93.3%	N/A	100.0%	100.0%	91.7%	73.2%
	Feb	100.0%	94.2%	94.2%	N/A	100.0%	100.0%	92.1%	75.7%
	Mar	100.0%	95.0%	95.0%	N/A	100.0%	100.0%	92.5%	78.3%
	Apr	100.0%	95.8%	95.8%	N/A	100.0%	100.0%	92.9%	80.8%
	May	100.0%	96.7%	96.7%	N/A	100.0%	100.0%	93.3%	83.3%
	Jun	100.0%	97.5%	97.5%	N/A	100.0%	100.0%	93.8%	85.9%
	Jul	100.0%	98.3%	98.3%	N/A	100.0%	100.0%	94.2%	88.4%
	Aug	100.0%	99.2%	99.2%	N/A	100.0%	100.0%	94.6%	91.0%
	Sep	100.0%	100.0%	100.0%	N/A	100.0%	100.0%	95.0%	93.5%

Table 1 provides DoD spending guidance that serves to assist program offices with determining if their cash flow performance is maintaining an adequate pace. A close examination of the information in Table 1 reinforces the concept that there are different benchmark spending expectations for the different “colors”-of-money. Not shown on the chart is MILPER. Since this appropriation is primarily for salaries, its expenditure cycle occurs at a relatively predictable and standard pace. Also, procurement funding does not show a monthly expenditure rate. Since procurement is used to buy and support the purchase of large weapon systems and platform end items, its expenditures often occur in single large sums, as opposed to small monthly incremental allotments. However, the remaining three appropriations—RDT&E, O&M, MILCON—represent initiatives that a program office could fund and receive outlays against in relatively smaller installment amounts to projects. Table 1 reveals that after the first year of availability, the expectation is that RDT&E funds will be 55% expended, O&M funds will be 75% expended, and MILCON funding will be 14% expended. It is these appropriations that are of interest for use in an

ADP approach for financial execution management. ADP is ideal for either appropriation categories or specific projects where a program office would consider issuing staggered multiple allotments of cash or commitment actions to pay for the activity. This cash allocation approach is one where the program office is attempting to determine if the contractor or vendor will spend the current funds allotted to it before another installment of money is provided.

A Financial Execution Management Model

The following section provides a mathematical formulation for the financial execution problem of weapon system program offices. We define critical variables of the financial execution system and adopt them to a dynamic programming formulation.

At the start of the fiscal year, a budget of bud_i is allocated to each of a finite number $I$of projects $i \epsilon \{1, \dots, I\}$. During each of a finite number of time periods $t = 1, \dots, T$, each project i has a (random) disbursement need $\widehat{D}_{i,t}$, which must be satisfied from the current "inventory" of funds that have been committed and have become available to project i by period t.

The agency's objective is to allocate funds in a way that tracks the actual disbursements as closely as possible. This is reflected in the model as follows. For $t = 1, \dots, T$, let $b_{i,t}^c$ denote the total amount committed to project i by the end of period t. In particular,

$$b_{i,t}^c = \sum_{s=1}^{t} x_{i,s}$$

where $b_{i,s}^c = 0$ for $s \leq 0$. Moreover, we assume that at the start of each period, the agency has a cumulative disbursement schedule $\bar{b}_{i,t}^d = [\bar{b}_{i,t}^d(1), \dots, \bar{b}_{i,t}^d(T)]$ for each project i, where $\bar{b}_{i,t}^d(n)$ denotes the current (i.e., at the end of period *t*) projected amount of money that project i will need during time n. Once the actual disbursement requirement $\widehat{D}_{i,t}$ for project i during period t is revealed, the disbursements for each project i are updated according to a given function F^d, so that

$$\left(\bar{b}_{1,t+1}^d, \dots, \bar{b}_{I,t+1}^d\right) = F^d\left[\left(\bar{b}_{1,t}^d, \dots, \bar{b}_{I,t}^d\right), \left(\widehat{D}_{1,t}, \dots, \widehat{D}_{I,t}\right)\right]. \quad (1)$$

At the start of each period $t = 1, \dots T$, and for each project i, the agency must decide on a total amount x_t to commit. This amount is allocated to the I projects based on fixed allocation rules and is subject to constraints that depend on the cumulative commitments $b_{i,t}^c$ and current disbursement schedule $\bar{b}_{i,t}^d$ for each project i. Given $b_{i,t}^c, \dots, b_{I,t}^c$ and $\bar{b}_{1,t}^d, \dots, \bar{b}_{I,t}^d$, let

$$\chi(b_{i,t}^c, \dots, b_{I,t}^c, \bar{b}_{1,t}^d, \dots, \bar{b}_{I,t}^d).$$

Denote the corresponding set of feasible total commitment amounts x_t. If the agency elects to commit x_t, the cumulative commitments for each project i are updated according to a given function F^c(describing a given allocation rule), so that

$$\left(b_{1,t+1}^c, \dots, b_{I,t+1}^c\right) = F^c\left[\left(b_{1,t}^c, \dots, b_{I,t}^c\right), x_t\right]. \tag{2}$$

If the agency commits x_t at time t, its associated "cost" for that time period is the absolute difference between the cumulative amount committed by the end of time t, and the cumulative projected disbursement by the end of time $t + \alpha_i$ (which is when x_t first becomes available for disbursement), that is,

$$\left|\sum_{i=1}^{I} b_{i,t-1}^c + x_t - \sum_{i=I}^{I} \bar{b}_{i,t}^d(t + \alpha_i)\right|.$$

The term α_i is a project specific sensitivity parameter. The choice α_i reflects the number of time periods beyond the current time period t that a program office wants to provide an incremental amount of funding that will sufficiently cover project i costs occurring between time periods t and $t + \alpha_i$.

Formulation as a Dynamic Program

To formulate the agency's sequential decision problem as a dynamic program, we need to specify the *state variables*, the *decision variables,* the *exogenous information processes, transition function*, and the *objective function.*

State Variables: For $t = 1, \dots, T$, the *state* S_t at the start of period t is a pair that includes, for each project $i \in \{1, \dots, I\}$, the values $b_{i,t-1}^c$ (i.e., the cumulative commitment to project i by the end of time t-1) and $\bar{b}_{i,t-1}^d$ (i.e., the projected disbursement schedule for project i as of the end of period $t - 1$), that is,

$$S_t = \left[\left(b_{1,t-1}^c, \dots, b_{I,t-1}^c\right), \left(\bar{b}_{1,t-1}^d, \dots, \bar{b}_{I,t-1}^d\right)\right].$$

Decision Variables: For $t = 1, \dots, T$ and $i = 1, \dots, I$ the *decision variable* x_t denotes the amount that the agency commits at the start of time t. If the start at the start of period t is S_t, then x_t is constrained to satisfy

$$x_t \in A(S_t) \coloneqq \chi\left(b_{1,t}^c, \dots, b_{I,t}^c, \bar{b}_{1,t}^d, \dots, \bar{b}_{I,t-1}^d\right).$$

Exogenous Information Process: There is a single exogenous information process $\left\{\widehat{D}_{i,t}\right\}_{t=1}^T$ associated with each project i, where $\widehat{D}_{i,t}$ are simulated actual disbursement requirements for each project i during period t.

Transition Function: Suppose that at the start of period t, the state is S_t. If the decision $x_t = (x_{1,t}, \dots, x_{I,t})$ is made, and the exogenous information for that period is $\widehat{D}_t = (\widehat{D}_{1,t}, \dots, \widehat{D}_{I,t})$, then the state at the start of period $t + 1$ is

$$\begin{aligned} S_{t+1} &= S^M(S_t, x_t, \widehat{D}_t) \\ &= \left[\left(b_{1,t}^c, \dots, b_{I,t}^c\right)\left(\bar{b}_{1,t}^d, \dots, \bar{b}_{I,t-1}^d\right)\right] \end{aligned}$$

$$= \left[F^c\left(\left(b^c_{1,t-1}, \ldots, b^c_{I,t-1}\right), x_t\right), F^d\left(\left(\bar{b}^d_{1,t-1}, \ldots, \bar{b}^d_{I,t-1}\right), \left(\widehat{D}_{1,t}, \ldots, \widehat{D}_{I,t}\right)\right)\right],$$

where F^c and F^d come from (2) and (1), respectively. Figure 2 depicts the relationship that exists between the *state variables* S_t, *decision variables* x_t, and *exogenous information process* $\widehat{D}_t$. At the beginning of a time period *t*, the financial execution status of a program office is captured by S_t which includes the cumulative commitment amounts and project disbursement schedules for each project i. At this point, exogenous information $\widehat{D}_t$ regarding the previous time period's disbursements is revealed. The decision process utilizes information from the state position S_t and exogenous information $\widehat{D}_t$ to select a commitment action x_t regarding the amount of additional incremental funding to allocate to each project i. This commitment action x_t, along with our knowledge regarding the current actual project disbursement amounts $\widehat{D}_t$, allows our decision system to step forward one time period and into the next state position S_{t+1}, which contains updated information regarding our program office's cumulative commitment amounts and project disbursement schedules. The process continues for a pre-defined limited number T of time periods or decision periods t.

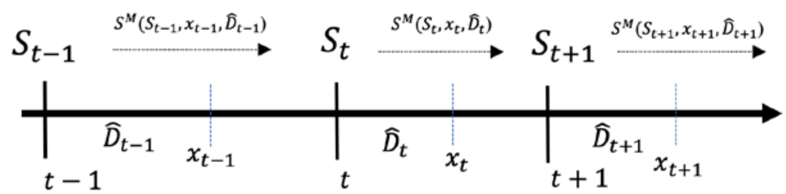


Figure 2. State-to-State Transitions

Objective Function: Suppose that at the start of period t, the state is S_t and the decision $x_t$is made. Then the corresponding *contribution* of period t is

$$\hat{C}(S_t, x_t) := -\left|\sum_{i=1}^{I} b^c_{i,t-1} + x_t - \sum_{i=1}^{I} \bar{b}^d_{i,t-1}(t + \alpha_i)\right|. \tag{3}$$

The objective is to find a policy that maximizes the expected total contribution over the T periods, that is, a policy that maximizes

$$\mathbb{E}\left\{\sum_{t=1}^{T} \hat{C}(S_t, x_t)|S_0\right\}.$$

Cash Allocation Example

We now consider the simple case of allocating funding for a single project with a total project budget $bud_1 = 27$. We define the time period t as a month and consider the cash allocation process for this single project over a fiscal year horizon $T = 12$ months. The choice of t reflects the frequency of how often a program office wants to assess their financial execution status and make an allotment of funding decision x_t across all the projects within their budget. Additionally, we'll select $\alpha_1 = 2$, to indicate that the program office wants to consider funding allotments in amounts that cover three-month time frames. An initial cumulative disbursement schedule $b^d_{i,t}$ is created from either a direct vendor quote, similar work completed in the past, or from any other viable technique available to the program office that can be used to create an initial spend plan forecast. For our single project, we'll assume the following cumulative disbursement schedule in millions of dollars ($M):

$$\bar{b}^d_{1,1} = [0,0,0,3,6,9,12,15,18,21,24,27].$$

This disbursement profile represents a project that starts work in the fourth month of the fiscal year, January, and requires $3M per month for the remainder of the fiscal year.

Let's consider a case where the decision system arrives at time period $t = 4$, January, with $S_t = (6, [0,0,0,3,6,9,12,15,18,21,24,27])$. At this point, $6M are committed to the project and $0M are disbursed. The decision system makes a commitment action according to (3). Given that $\alpha_1 = 2$, the next allocation of funding will attempt to bring the current total committed funding level $b^c_{1,4}$ up to a level that matches as close as possible the estimated cumulative disbursement amount for March (time period $t + 2$). In our example, we'll assume that the choice for the next allotment of funding is $3M. The decision system moves into the next time period, $t = 5$, February. At this point, exogenous information is revealed regarding actual disbursements that occurred in time period $t = 4$. This information is then used to create an updated cumulative disbursement schedule. For example, if the actual disbursement amount in January was only $1M as opposed to the anticipated $3M that was expected, an updated disbursement schedule might look like the following

$$\bar{b}^d_{1,5} = [0,0,0,1,3,6,9,12,16,20,24,27].$$

The implication is that the contractor supporting the work fell behind schedule during the month of January; however, the updated cumulative disbursement schedule indicates a belief that the contractor will be able to make up the additional work prior to the end of the fiscal year and will still require a full $27M to pay for the project prior to the end of the 12-month period.

Curse of Dimensionality

One drawback of using the dynamic programming formulation for solving the financial execution problem is that it suffers from the "curse of dimensionality," which is a common issue for many optimization modeling approaches. Using the single project scenario described in the previous section, we can consider the computational demands of our decision system based on the size of the action space x_t and state-space S_t. In order to determine these dimensions, we will first need to make an assumption about the discretized amount with which our project receives and disburses dollars. For simplicity, we assume money is received and spent to the nearest $1M increment. Additionally, we need to make another assumption about the range of variability that can occur with our simulated exogenous data $\widehat{D}_{1,t}$. In this case, we'll assume that disbursements can occur with variability of +$2M to -$2M, above and below the forecasted amount for a given time period t. Given these parameters, we can now calculate both the sizes of both the action-space and state-space.

Given that the project receives money to the nearest $1M increments, this means that for each time period t, there are 28 possible commitment or de-commitment actions to our $27M project. De-commitment actions are allowed as long as sufficient funding remains committed to the project to cover all expenses (disbursements) that have occurred to date. The state-space is defined as the combination of our cumulative commitment amount $b^c_{1,t}$ and disbursement schedule $\bar{b}^d_{1,t}$. For the $27M project, there are 28 possible values for the scalar $b^c_{1,t}$. Furthermore, since we are anticipating disbursements to occur in nine out of our 12-month time frame, there are 5^9 possible vectors combinations for $\bar{b}^d_{1,t}$, and when combined with the 28 possible values of $b^c_{1,t}$ means that there are over 54 million state-space possibilities. Even for this single project situation, to model all possible outcomes for all the possible state-action pairings is computationally intractable. This difficulty is further exacerbated when we consider budget scenarios that examine multiple projects simultaneously.

As an alternative, we consider using an approximate dynamic programming (ADP) modeling approach to the financial execution problem. ADP allows us to estimate a "good" decision-making solution without having to explicitly enumerate and calculate the values of all possible action-outcome pairings. Rather, it provides a means of approximating state-space values through the use of Bellman's formula:

$$V_t(S_t) = \max_{x_t}\left(\hat{C}(S_t, x_t) + \gamma\mathbb{E}\{V_{t+1}(S_{t+1})|S_0\}\right).$$

Bellman's formulation contains two components. It retains the *contribution* from the previously stated objective function, $\hat{C}(S_t, x_t)$, and combines with it a discounted expected value of the state the decision system arrives at as a result of the action x_t taken at time period t. Through the use of simulation, the ADP approach allows us to approximate or "learn" the values of state-spaces in our decision system. As a result, the ADP algorithm can generate a cash allocation policy that directs a program office to allocate funding during each time period t to successively move the decision-maker from one high valued state-space (financial execution position) to another high valued-state space position. Therefore, the cash allocation policy generated by the ADP algorithm will balance between allocation decisions taken earlier in the FY with those generated later, creating a sequential cash allocation policy that limits that amount of over-committed funding without shortchanging funding for projects.

Conclusion

This paper presents a framework for integrating ADP as a solution approach to DoD financial execution management. At the end of each FY, millions of unspent dollars are returned by weapon system program offices to DoD comptrollers as a result of use-or-lose budget environments. Currently, traditional FY cash allocation strategies implemented by program offices are myopic and risk projects receiving more funding than what can be spent within the FY calendar. ADP offers an alternative analytical tool that creates a sequential cash allocation plan balancing between the current allotment of funding to a project and the final end of year financial position of a project.

The next steps of this research involve testing the ADP algorithm in a theoretical DoD financial execution construct. ADP is a solution approach that contains flexibility that allows its structure to be modified to accommodate different parameters and facets that are unique to separate program offices. Further work will focus on experimenting with three of our ADP problem variables and determining how they can be used to customize our ADP algorithms. First, we will consider how different definitions of the epoch period t will impact the effectiveness of our model. In the example provided, t represented making a cash allocation decision, x_t, every month. Other options for t can include weekly or daily epochs. One rationale for changing the definition of t is to be able to better align it to the actual decision periods used by program offices. Another reason would be to evaluate to what extent making more cash allocation or fewer cash allocation decisions over a FY has on the objective of reducing the total amount of vulnerable end of year overcommitted funding. Another feature to closely examine is the sensitivity variable α_i. The value α_i is a parameter that establishes how many time periods, t, into the future the current allotment of cash will be able to pay for project disbursements. In the above example, we defined $\alpha_i = 2$, meaning that our objective function formulation would pick cash allocation amounts that funded projects for the next three months. Realistically, this value would be dynamic and not static; its value would be dependent on the point in time in the fiscal year in which a cash allocation decision is being made. If it is early in the FY, program office may be comfortable with setting α_i at a larger value given that the contractor has a longer time period before the end of the FY to utilize the money, and then slowly reducing the parameter α_i as the FY calendar starts to approach the end of the year. Another strategy to use if the program office is operating under a CRA is to set α_i to the length of time of the CRA. Under this scenario, program offices are aligning a project's cash allocation with the CRA timeframe. Lastly, we look to consider different ways of defining the exogenous data $\widehat{D}_t$. At the start of each time period t, the ADP model simulates a sample of exogenous data $\widehat{D}_t$ and uses the information to define the current period's state-space S_t. The variable $\widehat{D}_t$ represents both the expenses (i.e., disbursement information) that occurred for a project in the previous time period along with the strategy for how this information is used to update the cumulative disbursement schedule $\bar{b}^d_{i,t}$. To provide more fidelity to the ADP model, $\widehat{D}_t$ can be uniquely defined for each project. For example, $\widehat{D}_t$ would take into consideration any available historical spending data on the project as well as subject matter expert input specifically related to the execution management of the project.

Reference List

Bellman, R. (1954). The theory of dynamic programming. *Bulletin of the American Mathematical Society, 60*(6), 503–515.

Bellman, R. (1957). *Dynamic programming.* Princeton, NJ: Princeton University Press.

Bellman, E., & Dreyfus, S. (1962). *Applied dynamic programming.* Princeton, NJ: Princeton University Press.

Cooley, W. T., & Ruhm, B. C. (2014). *A guide for DoD program managers*. Fort Belvoir, VA: Defense Acquisition University Press.

Conley, K. M., Dominy, J. R., Kneece, R. R., Jr., Mandelbaum, J., & Whitehead, S. K. (2014). *Implications of DoD funds execution policy for acquisition program management*. Alexandria, VA: Institute for Defense Analysis.

Defense Finance Accounting Service (DFAS). (n.d.). Agency overview. Retrieved from https://www.dfas.mil/pressroom/aboutdfas.html

Department of the Army. (2014). *Financial Management Operations* (FM 1-06). Retrieved from https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/fm1_06.pdf

Gosavi, A. (2003). Reinforcement learning: A tutorial survey and recent advances. *INFORMS Journal on Computing, 21*(2), 178–192.

Howard, R. (1960). *Dynamic programming and Markov processes.* Cambridge, MA: MIT Press.

Mehta, A. (2018, October 1). DoD hands out $7.54 billion in year-end contract bonanza. *DefenseNews.*

Moritz-Rabson, D. (2018, September 28). Pentagon rushes to spend billions before end of month. *Newsweek.*

Powell, W. B. (2009). What you should know about approximate dynamic programming. *Naval Research Logistics, 56,* 239–249.

Powell, W. B. (2010). Merging AI and OR to solve high-dimensional stochastic optimization problems using approximate dynamic programming. *INFORMS Journal on Computing, 22*(1), 2–17.

Tremaine, R. L., & Seligman, D. J. (2013). OSD's obligation and expenditure rate goals: An examination of the factors contributing to the interference. Fort Belvoir, VA: Defense Acquisition University.

Tskitsiklis, J. N. (2010). Perspectives on stochastic optimization over time. *INFORMS Journal on Computing, 22*(1), 18–19.

Actual Obligation Versus Comptroller Projected Obligation Rates

Seamus P. Daniels—is a research associate and program manager for Defense Budget Analysis in the International Security Program at the Center for Strategic and International Studies (CSIS), where he contributes to the Center's research on issues related to defense funding and military readiness. He has authored publications on trends in the overall defense budget, the legislative process surrounding defense appropriations, Navy readiness funding, and NATO burden sharing. He holds a BA with honors from Princeton University's Woodrow Wilson School of Public and International Affairs with minors in Near Eastern studies and Arabic language and culture. [sdaniels@csis.org]

Todd Harrison—is the director of Defense Budget Analysis and the director of the Aerospace Security Project at CSIS. As a senior fellow in the International Security Program, he leads the Center's efforts to provide in-depth, nonpartisan research and analysis of space security, air power, and defense funding issues. He has authored publications on trends in the overall defense budget, military space systems, defense acquisitions, military compensation, military readiness, nuclear forces, and the cost of overseas military operations. He graduated from the Massachusetts Institute of Technology with both a BS and an MS in aeronautics and astronautics. [tharrison@csis.org]

Abstract

This paper seeks to enhance understanding of the formulation and accuracy of Department of Defense (DoD) Comptroller projected obligation rates in the defense acquisition sector. These projections are published annually for each appropriation account in the Financial Summary Tables released by the office of the DoD Comptroller. To understand the implications of these forecasts on the contracting acquisition area, this paper compares the Comptroller projected obligation rates for procurement accounts with actual obligation rates as well as budget execution benchmarks also compiled by the Comptroller's office. It assesses the reliability of the projections and their consistency with other DoD targets, identifies trends in the accuracy of obligations rates across different accounts, and attempts to isolate factors that may influence the formulation and accuracy of the projections.

Introduction

Obligation rates are considered one of the "key financial metrics" for the Department of Defense (DoD) in monitoring how programs allocate their funding and whether they remain on schedule (*Unobligated Balances*, 2006). While budget authority and total obligational authority track how much money is appropriated by Congress each year, obligations track how funding is committed by signing contracts, employing personnel, or otherwise making commitments to spend money (Schwartz et al., 2018). When determining the amount of funding that may be made available for an appropriation account in an upcoming fiscal year, DoD offices and the authorization and appropriations committees in Congress take previous years' actual obligation rates into consideration (Defense Security Cooperation Agency [DSCA], 2012). Programs that have not been able to adequately obligate prior year funding are less likely to receive the funding they are requesting for future years and, in more extreme cases, may have prior year unobligated funding rescinded by Congress.

The Office of the Under Secretary of Defense (Comptroller; OUSDC) publishes a baseline standard for cumulative obligation and expenditure rates by title of funding (procurement, RDT&E, O&M, etc.). This table of benchmarks, derived from 30 years of execution history, is intended to serve as a rule-of-thumb for the military services when planning their program expenditures (Conley et al., 2014, pp. vi–vii). For procurement accounts, the benchmarks state that a cumulative 80% of the funding should be obligated by the end of the first year, 90% by the end of the second year, and 100% by the end of the third year (when the funding would otherwise expire; OUSDC, 2017). Expenditure rates are higher for RDT&E accounts given their two-year period of availability. It is expected that 90% of RDT&E funding should be obligated by the end of the first year and 100% by the end of the second year (OUSDC, 2017). Congressional staffs use this table as a baseline reference for judging whether particular funding lines and programs are obligating money on track or are falling behind.

The Comptroller's office also publishes Financial Summary Tables annually with the president's budget request that include a more granular projection of obligation rates by individual appropriation account. These projected rates, presented as a percentage of "total operating authority," represent the percent of a particular budget year of funding that the DoD expects to obligate over the course of the fiscal years that follow. Importantly, the Comptroller projected obligation rates are not cumulative, whereas the Comptroller benchmarks are cumulative obligation rates. For example, in the Army's Aircraft Procurement FY 2017 appropriations, DoD projected that 64.44% will be obligated in FY 2017, 25.00% in FY 2018, and 10.56% in FY 2019 (OUSDC, 2016, p. 004). Since this is a procurement account, the money is only available for three years and any leftover funding after that time would expire. In comparison, Research, Development, Test, and Evaluation (RDT&E) funding must be fully obligated within two years, while Operation and Maintenance (O&M) and Military Personnel (MILPERS) accounts must be obligated within one year (Schwartz, 2017).

Given the obligation rate benchmarks set by the Department, one would expect most if not all of the projections for the first year of obligations to meet or exceed the 80% threshold. However, the projected obligation rates fail to meet the execution benchmarks for many accounts. This lack of alignment poses questions over the usefulness of the projections as well as their accuracy in anticipating the rate of actual obligations. Similarly, the lack of change in the projections from year to year (the Army Aircraft Procurement account has had identical projections from FY 2013 through FY 2019), even as the funding and status of programs within the accounts changed considerably, calls into question the DoD's model for deriving projected obligation rates.

Obligation rates can be important for industry and investors as a measure of government contracting for current and future fiscal years. Private sector partners rely on projections for their own strategic planning, forecasting the overall potential for sales and revenue for the defense industry based in part on the expected obligation rates. Consequently, the obligation and outlay rates can impact the stock valuations of companies via their revenue forecasts. These projections are also important to defense companies themselves, particularly smaller ones, because the timing of programs can mean the difference between smooth cash flow and challenges to solvency. Given the significance these forecasts play in the acquisition sector, inaccurate projections could contribute to poor decision-making in the private sector that could lead to inefficiencies in the market and sub-optimum management decisions within companies.

In an effort to assess the reliability and accuracy of the Comptroller projected obligation rates, this paper provides an analysis of the projections for procurement accounts. It includes a survey of the projected obligation rates for a variety of procurement accounts from FY 2012 to FY 2019 and compares those projections against both the obligation rate benchmarks and actual obligation rates (from FY 2012 to FY 2015). The paper analyzes that data by military department and certain types of account to identify trends and draw conclusions.

Literature Review

Previous studies have assessed the execution of DoD programs against the Department's obligation rate goals. A 2013 Defense Acquisition University (DAU) report examined potential causal factors preventing acquisition programs from meeting the execution benchmarks. The study surveyed 229 DoD personnel who ranked the impact of 64 factors on the performance of acquisition programs. According to the results, the late release of full obligation/budget authority due to continuing resolution authority, contract negotiations' delays, and contract award delays had the highest adverse impact on the achievement of execution goals (Tremaine & Kinnear-Seligman, 2013).

A 2014 study from the Institute for Defense Analyses (IDA) similarly assessed the underperformance of acquisition programs against the Comptroller execution benchmarks and investigated factors related to program execution. The report found that the rates for procurement obligations and RDT&E disbursements have been decreasing since 2006 and 2009, respectively. While the research team found that the benchmarks—though "potentially arbitrary to some extent"—are "a reasonable means of identifying funds for possible reallocation to higher priority needs," it concluded that "management attention unduly focuses on meeting benchmarks" and offered recommendations to improve program execution (Conley et al., 2014, pp. vi–vii).

Both of the aforementioned studies focused on the execution of acquisition programs against the benchmarks for obligation and expenditure rates. This paper builds on the existing research by comparing the actual obligation rates against the Comptroller projected obligation rates found in the Financial Summary Tables and the Comptroller benchmarks. Given the focus on the Comptroller projected obligation rates, this study is also conducted at the broader appropriation account level rather than the budget line level of detail used in the IDA analysis.

Methodology

Collecting Comptroller Projected Obligation Rates

The analysis in this report was conducted in three phases. The first phase entailed the collection of the Comptroller projected obligation rates from Section F of the Financial Summary Tables. Projections were captured for procurement appropriation accounts from FY 2012 to FY 2019. Data collection posed a challenge given the Financial Summary Tables' lack of a machine-readable format, forcing the research team to manually input the projections.

The research team then measured how often the projected obligation rates change from fiscal year to fiscal year before comparing them to the cumulative program execution benchmarks. The projections for procurement accounts were assessed to determine how often they met the 80% threshold for the first year of obligations and 90% for the second year. Accounts' alignment with the benchmarks were measured as a percentage of the total number of budget years in which the projected obligation rates met or exceeded the benchmarks. For the purposes of this analysis, the term "budget year" is used to refer to the year in which funding is originally appropriated for an account. Funding can then be obligated in that fiscal year and in the fiscal years that follow.

The research team studied 18 procurement accounts. Several procurement accounts were excluded from the analysis as exceptions because they do not follow the standard obligation practices for procurement accounts. They include Shipbuilding & Conversion, Navy; Coastal Defense Augmentation; Defense Production Act Purchases; Chemical Agents & Munitions Destruction; and the MRAP Vehicle Fund.

Two procurement accounts contain less data than the other accounts. The Space Procurement, Air Force account was only created in FY 2016 so there are only four budget years' worth of projections and no budget years' worth of actual obligation rates. The National Guard and Reserve Equipment account also lacked projections for FY 2019.

Calculating Actual Obligation Rates

The second phase of this analysis entailed the calculation of the actual obligation rates of the selected procurement accounts. To calculate the actual obligation rate of funds, the total obligations in a given fiscal year from a particular budget year's funding is divided by the total available for obligation for that budget year including any adjustments that may occur in subsequent fiscal years. This data can be found in Section G of the Financial Summary Tables.

For example, as shown in Table 1, the Aircraft Procurement, Army account had $5,902,609,000 available for obligation for budget year 2015. Over the next two fiscal years, Congress and the DoD made adjustments to the 2015 budget year funding in this account, totaling a net addition in funding of $455,317,000 in FY 2016 and $105,597,000 in FY 2017, as shown in Table 1. Thus, the total budget year 2015 funding for this account ended up being $6,453,523,000. This is the total available for obligation used in the denominator when calculating the actual obligation rate for each year. As shown in Table 2, the total obligations in each fiscal year of the specific budget year's funding is then divided by the total available to calculate the actual rate of obligation for each fiscal year. Actual obligation rates were only calculated for budget years from 2012 to 2015 due to the lack of complete data (i.e., final appropriated and executed amounts) for budget years 2016 through 2019.

Table 1. Aircraft Procurement, Army Budget Year 2015 Funding

Budget Year 2015	**FY 2015**	**FY 2016**	**FY 2017**	**Final**
Budget Authority	$5,799,286,000	-$25,000,000	-$15,000,000	
Balances Transferred		-$13,000,000	-$22,257,000	
Recoveries of Prior Year Obligations		$464,861,000	$72,995,000	
Reimbursable Orders	$103,323,000	$18,456,000	$69,859,000	
New Funding Available for Obligation	$5,902,609,000	$445,317,000	$105,597,000	$6,453,523,000

Table 2. Aircraft Procurement, Army Budget Year 2015 Obligations and Obligation Rates

Budget Year 2015	FY 2015	FY 2016	FY 2017
Total Obligations	$3,950,184	$1,875,308	$536,183
Overall Total Available for Obligation	$6,453,523,000	$6,453,523,000	$6,453,523,000
Obligation Rate	61.21%	29.06%	6.64%
Cumulative Obligations	61.21%	90.27%	96.91%

Comparing Projected Obligation Rates and Actual Obligation Rates

The actual obligation rates calculated in the second phase of the study were then compared to the historically-derived benchmarks for procurement accounts to determine which accounts met the 80% obligation rate goal after one year of execution and 90% after two years. The accounts were then measured against the Comptroller projected obligation rates to assess the projections' accuracy on an account by account basis.

To compare the accuracy of projections for different procurement accounts, the research team calculated the difference between the actuals and projections for each of the three fiscal years that each budget year of funding was available for obligation. Those differences were then averaged for each fiscal year of availability for an account. In addition to assessing the average projection error for each year of availability by account, the research team also aggregated the data by military department. The differences between projections and actuals were averaged by fiscal year across all accounts associated with each department rather than calculating the department average from the overall account average. The median difference by military department was also calculated to compare against the average and is located in the appendix of the report.

Analysis

Year-Over-Year Changes in Procurement Account Projected Obligation Rates

Of the 16 procurement accounts containing the complete eight years' worth of projections from budget years 2012 through 2019, one account possessed identical projections for all eight years; 12 possessed identical projections for seven of the eight years; two possessed identical projections for six of the eight years; and one possessed identical projections for five of the eight years.

It is somewhat counterintuitive that the projected obligation rates at the account level stay fairly consistent over time because the status and mix of programs within each account can vary considerably from year to year. One might expect that the procurement obligation rate would be slower for programs that are transitioning from development to procurement, are ramping up procurement, or are having contract award and negotiation issues. The fact that the projected obligation rates stay consistent from year to year suggests that these projections are not based on the execution plans of the programs within the accounts and are instead based on historical rates or aspirational obligation plans.

Moreover, the consistency of the accounts' projected obligation rates from year-to-year does not translate into alignment with the benchmarks established for budget execution. As shown in Table 3, a majority of the accounts surveyed in the study projected their obligation rate for the first year of availability would be under the 80% goal. Only 31% of the 139 budget years assessed in this study projected that the obligation rate for the first year of funding would meet or exceed 80%. The Navy was the only military department that had a majority of its first year projections achieve the goal established by the Comptroller's office.

The lack of alignment between the projected obligation rates and historically-derived benchmarks could come as a result of delays in defense appropriations. Between FY 2011 and FY 2018, appropriations for defense were delayed on average by 139 days (including days under continuing resolutions and government shutdowns). These delays in appropriations would translate into delays in obligations, leading to obligation rates under 80%. According to the IDA study, obligation rate goals were lowered for 2013 and 2014 to 62% and 66%, respectively, "in recognition of the increasing difficulties that acquisition programs have in meeting the historical execution benchmarks" (Conley et al., 2014, pp. 5–6). However, this paper did not assess the change in projected obligation rates from budget years prior to 2012 to determine whether lack of alignment with the benchmarks correlates with delays in defense appropriations because it was beyond the scope of this effort.

Table 3. Comptroller Projected Obligation Rates vs. Execution Benchmarks for First Year

Military Department	Account-Budget Years with First Year Projection ≥ 80%	Total Number of Account-Budget Years of Data	Percentage of Account-Budget Year Projections Meeting or Exceeding First Year Benchmarks
Army	11	40	27.5%
Navy	25	40	62.5%
Air Force	4	36	11.1%
Other[1]	3	23	13.0%
TOTAL	**43**	**139**	**30.9%**

As shown in Table 4, a majority of the overall account budget year projections aligned with the two-year benchmark of 90% of funds obligated. However, fewer than half of

[1] The "Other" category throughout the tables in this report include the following accounts: Joint Improvised Explosive Device Defeat Fund; Procurement, Defense-Wide; and National Guard and Reserve Equipment.

the Army's two-year projections anticipated meeting the 90% goal while four-fifths of Navy and Air Force account budget year projections met or exceeded the benchmark.

Table 4. Comptroller Projected Obligation Rates vs. Execution Benchmarks for Second Year

Military Department	Account-Budget Years with Second Year Projection ≥ 90%	Total Number of Account-Budget Years of Data	Percentage of Account-Budget Year Projections Meeting or Exceeding Second Year Benchmarks
Army	19	40	47.5%
Navy	32	40	80.0%
Air Force	29	36	80.6%
Other	18	23	78.3%
TOTAL	**98**	**139**	**70.5%**

Comparing Actual Obligation Rates to Execution Benchmarks

While the previous analysis compares the projected obligation rates to the Comptroller execution benchmarks, this section compares the actual obligation rates to the execution benchmarks. As discussed in the methodology section, actual obligation rates were only calculated for four budget years (2012 to 2015) due to the lack of complete data for subsequent years. When compared to the cumulative execution benchmark rates, the majority of the actual obligation rates failed to meet both the one- and two-year targets of 80% and 90%, respectively.

As shown in Table 5, only 13% of 68 account budget years assessed met or exceeded the targeted goal for the first year of obligations, a smaller proportion than the 29% for projected obligation rates. The Air Force and Navy had the highest number of account budget years that matched or surpassed the 80% goal with four each. None of the Army's accounts met the benchmark, while approximately 28% were projected to do so, according to the previous section.

Table 5. Actual Obligation Rates vs. Execution Benchmarks for First Year

Military Department	Account-Budget Years with First Year Actuals ≥ 80%	Total Number of Account-Budget Years of Data	Percentage of Account-Budget Year Actuals Meeting or Exceeding First Year Benchmarks
Army	0	20	0.0%
Navy	4	20	20.0%
Air Force	4	16	25.0%
Other	1	12	8.3%
TOTAL	**9**	**68**	**13.2%**

The actual obligation rates performed better against the two-year execution benchmark of 90%. Nearly half (46%) of the total 68 account budget years obligated 90% or more of their funds by the end of the second year of availability. Relative to the Army and Air Force, which only saw 15% and 38% of their respective account budget years meet the threshold, the Navy Department impressed with 70% of its 20 account budget years reaching a 90% obligation rate. If the Marine Corps is excluded, that figure improves to 88% of the Navy's budget years for procurement accounts as a service.

Table 6. Actual Obligation Rates vs. Execution Benchmarks for Second Year

Military Department	Account-Budget Years with Second Year Actuals ≥ 90%	Total Account-Budget Years of Data	Percentage of Account-Budget Year Actuals Meeting or Exceeding Second Year Benchmarks
Army	3	20	15.0%
Navy	14	20	70.0%
Air Force	6	16	37.5%
Other	8	12	66.7%
TOTAL	**31**	**68**	**45.6%**

All four account-budget years of the Air Force's "other procurement" account met both the 80% benchmark with an average obligation rate of 91.7% for the first year of availability and the 90% benchmark with an average obligation rate of 97.3% over the first two years of availability. Such high rates, which are exceptions to the rest of the Air Force's actual obligation rates, may be attributed to the large amount of classified "pass-through" funding in this account (Hlad, 2016).

Measuring the Accuracy of Projected Obligation Rates Against Actual Obligation Rates

While the actual obligation rates for the procurement account budget years surveyed underperformed against the historically-derived benchmarks, the question remains how accurate the Comptroller projected obligation rates are in comparison with the actual obligation rates. Table 7 shows the average difference between the projected and actual obligation rate over the three years of funding availability for each procurement account. On average, the difference between the projected and actual obligation rates was approximately 14% for the first year of availability, 9% for the second year, and 6% for the third. Assessed by department, the Navy had the smallest average difference between its estimates and actuals with 8% for the first year, 7% for the second, and 5% for the third. If the Marine Corps is excluded from the Navy's average, the difference drops to 7%, 6%, and 3%, respectively, for the three years of availability. It is worth noting that across the different phases of this study, the Navy's projected and actual obligation rates were best aligned with the execution benchmarks, and its projections were the most accurate overall compared to the other military departments.

The Army had the largest average difference between its projections and actuals at 16% for the first year, 10% for the second, and 7% for the third. The error was driven by a 25% average difference between the projected obligation rate and actuals for the first year of availability in the Army's missile procurement account—the largest difference of any procurement account belonging to the three military departments. The Army anticipated obligating an average of approximately 83% of its account funding in the first year of availability for budget years 2012–2015, yet only obligated 58% of funding on average.

Another comparison between the three military departments' actual obligation rates can be made by assessing the aircraft procurement accounts of each. While the three accounts are not like-for-like comparisons given they procure different platforms and possess different funding levels (e.g., for the 2015 budget year, total obligations for Army aircraft procurement were $6,361,675,000 in current dollars; $16,308,912,000 for the Navy; and $12,187,879,000 for the Air Force), they nevertheless provide some standardization in comparison. As shown in the data in Table 7, the Navy and Army had similar average differences between their projections and actuals over the three years of availability. However, the average differences for the Air Force's aircraft procurement account were more than double those of the Navy and Army for the first two years of availability.

A comparison of the average and median difference between projected and actual obligation rate for the military departments can be found in the appendix.

Table 7. Average Difference Between Projected and Actual Obligation Rate by Account, Budget Years 2012–2015[2]

Account	First Year	Second Year	Third Year
Aircraft Procurement, Army	6.76%	4.80%	3.04%
Missile Procurement, Army	24.55%	14.47%	9.78%
W&TCV Procurement, Army	19.18%	11.73%	7.41%
Ammo Procurement, Army	12.26%	4.31%	9.51%
Other Procurement, Army	15.18%	14.20%	3.78%
Total Army Procurement	***15.58%***	***9.90%***	***6.70%***
Aircraft Procurement, Navy	6.92%	4.69%	2.36%
Weapons Procurement, Navy	5.42%	5.61%	4.22%
Ammo Procurement, Navy	3.64%	4.30%	2.05%
Other Procurement, Navy	11.41%	11.10%	2.51%
Procurement, Marine Corps	12.41%	10.11%	14.50%
Total Navy Procurement	***7.96%***	***7.16%***	***5.13%***
Aircraft Procurement, Air Force	19.83%	14.77%	5.17%
Missile Procurement, Air Force	12.50%	11.07%	1.51%
Ammo Procurement, Air Force	12.26%	4.31%	9.51%
Other Procurement, Air Force	12.06%	10.27%	2.10%
Total Air Force Procurement	***14.16%***	***10.11%***	***4.57%***
JIEDDF	18.00%	15.56%	2.95%
Procurement, Defense-Wide	4.31%	3.81%	2.74%
National Guard & Reserve Equip.	37.31%	15.00%	21.55%
Total Other Procurement	***19.87%***	***11.46%***	***9.08%***
TOTAL	**13.76%**	**9.42%**	**6.16%**

[2] Averages for the overall military departments represent average of all budget years' rates associated with a particular department's procurement accounts, not an average of the account averages.

Conclusion

This paper presents a preliminary set of analysis for procurement accounts only. The final report of this project will analyze other titles of the budget, particularly RDT&E and MILCON, to determine if similar trends are evident. The full analysis will also include additional years of budget data to examine longitudinal trends in obligation rates. While this analysis examined only a subset of the budget execution data, namely procurement accounts from FY 2012 to FY 2019, it yields a number of interesting findings:

- The projected obligation rates vary little from year to year within a particular procurement account. This suggests that the services do not regularly re-evaluate the projections for accuracy, nor do they attempt to adjust projections based on the plans of programs within the account.
- Just over half (51%) of the projected obligation rates for the first and second year of funding availability meet or exceed the corresponding execution benchmarks.
- While the projected obligations rates tend to be lower than the benchmarks, the actual obligation rates tend to be even lower than the projected rates.
- The difference between projected and actual obligation rates vary considerably across accounts, with some of the largest discrepancies in Missile Procurement, Army; Wheeled and Tracked Combat Vehicles, Army; and Aircraft Procurement, Air Force.
- The difference between the actual and projected obligation rates tend to narrow in the second and third year of funding availability. This suggests that while programs may be slower than expected executing funding in the first year it is available, they tend to catch up in subsequent years.
- The actual obligation rates fall well below the execution benchmarks, with just 13.2% of accounts meeting the benchmark standard in the first year and 45.6% in the second year.
- Overall, the Navy does the best at meeting its own projected obligation rates and the Comptroller benchmarks.

A central observation from this analysis is that the Comptroller benchmarks may not be a useful way to measure program execution. This is because the services do not appear to be planning or expecting to meet the benchmarks from the outset of the appropriations process, and it is not clear who, if anyone, is using the projected obligation rates. The benchmarks, however, are used by the Comptroller and congressional staff to gauge the execution of programs. However, the data suggests that if the intention of the benchmarks is to have a common standard based on historical execution patterns by which to hold programs accountable, then the benchmarks may need to be updated to account for changing patterns in the congressional budgeting process. For example, over the past 10 years the frequency and length of continuing resolutions has increased markedly, which may be having a systemic impact on the ability of programs to obligate funding in the first year of availability (Harrison & Daniels, 2017, pp. 4–5). Moreover, a common set of execution benchmarks may not be realistic because of the wide variation observed in the actual obligation rates across procurement accounts.

References

Conley, K. M., Dominy, J. R., Kneece, Jr., R. R. K., Mandelbaum, J., & Whitehead, S. K. (2014). *Implications of DOD funds execution policy for acquisition program management* (Report No. P-5164). Retrieved from https://www.ida.org/~/media/Corporate/Files/Publications/IDA_Documents/SFRD/2014/P-5164.ashx

Defense Security Cooperation Agency. (2012). *Security assistance management manual* (Chapter 12).

Washington, DC. Retrieved from http://www.samm.dsca.mil/chapter/chapter-12

Harrison, T. & Daniels, S. P. (2017). *Analysis of the FY 2018 Defense Budget*. Retrieved from https://csis-prod.s3.amazonaws.com/s3fs-public/publication/171208_Defense_Budget_Analysis.pdf?_bMzg.Rwos033iujMRE7YyyabElygTDY

Hlad, J. (2016, June). The Air Force's budget black hole. *Air Force Magazine*. Retrieved from http://www.airforcemag.com/MagazineArchive/Pages/2016/June%202016/The-Air-Force's-Budget-Black-Hole.aspx

Office of the Under Secretary of Defense (Comptroller). (2016). *Financial summary tables: Department of Defense Budget for Fiscal Year 2017*. Washington, DC. Retrieved from http://comptroller.defense.gov/Portals/45/Documents/defbudget/fy2017/FY2017_Financial_Summary_Tables.pdf

Office of the Under Secretary of Defense (Comptroller). (2017). *OUSDC rule-of-thumb acquisition obligation and expenditure rates*. Washington, DC. Retrieved from https://www.dau.mil/tools/Lists/DAUTools/Attachments/292/OSD%20(C)%20Color%20Rule-of-Thumb%20Acq%20Obligation%20and%20Expenditure%20Rates.pdf

Schwartz, M. (2017). *End-year DoD contract spending* (CRS Report No. IF10365). Retrieved from Congressional Research Service website: https://fas.org/sgp/crs/natsec/IF10365.pdf

Schwartz, M., Sargent, J. F., Jr., & Mann, C. T. (2018). *Defense acquisitions: How and where DOD spends its contracting dollars* (CRS Report No. R44010). Retrieved from Congressional Research Service website: https://fas.org/sgp/crs/natsec/R44010.pdf

Tremaine, R. L., & Kinnear-Seligman, D. J. (2013). The challenges in meeting OSD's obligation and expenditure rate goals: A closer look at potential causal factors, their groupings, and how they modulate. *Defense Acquisition Research Journal, 20*(3), 373–400.

Unobligated balances: Freeing up funds, setting priorities and untying agency hands: Hearing before the Committee on Homeland Security and Governmental Affairs, 109th Cong. 3 (2006) (testimony of John P. Roth, Deputy Comptroller, Office of the Under Secretary of Defense).

Disclaimer

The Center for Strategic and International Studies (CSIS) does not take specific policy positions; accordingly, all views expressed in this presentation should be understood to be solely those of the author(s).

Appendix. Average and Median Difference Between Projected and Actual Obligation Rate by Military Department, Budget Years 2012–2015

Military Department	First Year	Second Year	Third Year
Army - Avg.	15.58%	9.90%	6.70%
Army - Med.	15.36%	8.60%	6.33%
Navy - Avg.	7.96%	7.16%	5.13%
Navy - Med.	5.97%	7.34%	3.37%
Air Force - Avg.	14.16%	10.11%	4.57%
Air Force - Med.	13.19%	9.54%	1.90%
Other - Avg.	19.87%	11.46%	9.08%
Other - Med.	12.54%	8.23%	3.22%
Total - Avg.	13.76%	9.42%	6.16%
Total - Med.	12.84%	8.97%	3.78%

Panel 23. Contract and Logistics Implications in Defense Acquisition

Thursday, May 9, 2019	
2:15 p.m. – 3:30 p.m.	**Chair: Elliott Branch,** Deputy Assistant Secretary of the Navy, Acquisition and Procurement ***Leveraging Contracting Strategies with Private Shipyards for Increasing Naval Fleet Operational Availability*** Joseph Bradley and Sanjeev Gupta, Patrona Corporation ***Networked Logistics & Additive Manufacturing*** Susan Sanchez, Naval Postgraduate School ***An Evaluation of Mature Performance-Based Logistics Programs*** William Lucyshyn and John Rigilano, University of Maryland

Elliott Branch—Mr. Branch is the Deputy Assistant Secretary of the Navy (Acquisition and Procurement) in the Office of the Assistant Secretary of the Navy (Research, Development and Acquisition). He is the senior career civilian responsible for acquisition and contracting policy that governs the operation of the Navy's world-wide, multibillion-dollar acquisition system. Mr. Branch is the principal civilian advisor to the Navy Acquisition Executive for acquisition and procurement matters, serves as the Department of the Navy's Competition Advocate General and is the leader of the Navy's contracting, purchasing and government property communities.

Prior to joining the Navy Acquisition Executive's staff, Mr. Branch was the first civilian director of contracts at the Naval Sea Systems Command. In that role he led one of the largest and most complex procurement organizations in the Federal government. As the senior civilian for contracting at NAVSEA, Mr. Branch was responsible for the contractual oversight of the Nation's most complex shipbuilding and weapons systems procurement programs. His duties involved the obligation and expenditure of approximately $25 billion annually.

He is a member of the Senior Executive Service (SES). Members of the SES serve in the key positions just below the top Presidential appointees. They are the major link between these appointees and the rest of the Federal work force. SES members operate and oversee nearly every government activity in approximately 75 agencies.

Mr. Branch spent time in the private sector, where he specialized in acquisition and project management education, training and consulting for the federal workforce and its associated contractors. In this role, Mr. Branch was responsible for the design, development, delivery and maintenance for a wide variety of course material ranging from project management to contract law. Mr. Branch's clients included Computer Sciences Corporation, QSS Group, BAE Systems, the Pension Benefit Guaranty Corporation, and the Departments of Defense, Energy, Justice and State.

Prior to that, he served as the Chief Procurement Officer for the Government of the District of Columbia, where he was the agency head responsible for procurement operations, policy, and for formulating legislative proposals for local and Congressional consideration. Mr. Branch led a staff of over 200 employees that supported over 40 city agencies, administered a $14 million annual operating budget, and oversaw the placement of $1.5 billion, annually, in city contracts.

Before joining the District government, Mr. Branch held various positions in the SES with the Department of the Navy (DON). In 1993, he became a member of the SES as the Director, Shipbuilding Contracts Division, at NAVSEA. He next served as Executive Director, Acquisition and Business Management for DON, responsible for policy and oversight of contract operations throughout the entire Navy. While in this position, he also served as Project Executive Officer, Acquisition Related Business Systems. In this role, he was responsible for the formulation and execution of a multi-year effort transforming the navy's acquisition system from a paper-based system into one that made use of electronic technologies and methods. In this role, Mr. Branch was directly responsible for a portfolio of projects worth more than $200 million.

Mr. Branch graduated with a Bachelor of Science Degree in Economics from the University of Pennsylvania Wharton School and completed the Executive Program at the University of Virginia Darden School. He has received the Navy Distinguished Civilian Service Medal, the David Packard Excellence in Acquisition Award, two Presidential Rank Awards for Meritorious Executive, the Vice Presidential Hammer Award for Reinventing Government, and the 2012 Samuel J. Heyman Service to America Medal for Management Excellence.

Leveraging Contracting Strategies with Private Shipyards for Increasing Naval Fleet Operational Availability

Kaitlynn M. Castelle—PhD, PMP, CSM, **is** an adjunct professor at Old Dominion University in the Engineering Management and Systems Engineering Department, teaching project management and agile development and leading workshops designed to improve organizational agility, and is a consultant in Agile transformation for local and federal government programs. [kcastell@odu.edu]

Joseph M. Bradley—PhD, PE, has had an extensive career in ship operations, maintenance, repair and organizational design. Repeatedly called upon to start new activities for the U.S. Navy, he successfully met a succession of unique challenges. He is currently an Adjunct Research Associate at Old Dominion University and President of Leading Change, LLC. [josephbradley@leading-change.org]

Sanjeev Gupta—CEO of Realization Technologies, Inc., a provider of project delivery software and services that has delivered more than $7 billion of value by helping public and private sector organizations shorten project cycle times and improve due-date performance. He has been honored with a Lifetime Achievement Award by the Theory of Constraints International Organization for his contributions to project and supply chain management. [sgupta@realization.com]

Abstract

A major rethink of NAVSEA's shipyard contracting strategy is required to support the critical need of improving our naval fleet availability. Operational Fleet material availability is reduced when various parts of the "NAVSEA production system" operate at cross-purposes. By increasing alignment between major players (i.e., NAVSEA and the private shipyards), major improvements in delivery performance, cost, and even throughput can be realized. Developing strategies and specific actions to do so is a rich field given the current state of the system. We take an analytical as well as evidence-based approach to propose strategies that can be successful given the peculiar conditions of ship repair and modernization.

Introduction

The Chief of Naval Operations (CNO) availability relies on multiple commands and supporting activities of external organizations to ensure successful planning and execution of naval vessel depot-level maintenance activities, including repair and modernization of the propulsion, electric, and auxiliary plans, and structural repairs, as directed by the CNO and scheduled according to the ship's maintenance plan (Office of the Chief of Naval Operations [OPNAV] N431, 2010; Riposo et al., 2017). Naval Sea Systems Command (NAVSEA) Contracting Directorate, as the lead technical authority, is responsible for establishing performance standards for the accomplishment of all maintenance and modernization during availability periods and ensuring commands and private companies contracted perform repairs and modernization within the authorized scope of work (Caprio & Leszcynski, 2012).

The initial Availability Work Package (AWP) consists of known maintenance actions and ship alterations to be completed during the availability period and identified by the ship crew, NAVSEA, and other supporting engineering commands. During the planning phase, additional work items identified by the crew during work discovery periods are added to the AWP, with oversight and assistance from fleet support pre-availability testing and ship deficiency identification. The final AWP contains planned and unplanned maintenance and planned modernization, as well as other work based on known or expected ship material condition.

The depot maintenance availability schedule considers the nature of the work defined in the AWP as well as the projected private or public shipyard capacity. Ships inducted for maintenance availability must adhere to tight schedules to return to the fleet or otherwise receive authorization to increase deployment length, or defer or reduce ship maintenance. Lengthened operational cycles and deployments, deferred and unplanned maintenance, and the age of the fleet have contributed to the backlog of fleet maintenance availabilities (Riposo et al., 2017). When ship maintenance availability experiences schedule growth to restore capability, training is shortened, eliminated, or deferred, also exacerbating fleet readiness (GAO, 2017).

The backlog of scheduled maintenance activities at the shipyards continues to increase, including "oversight of private-sector activities under the purview of the shipyard, continuous maintenance activity, ship alterations, nuclear equipment disposal, fleet maintenance availabilities, Nuclear Regional Maintenance Department activities, fleet technical support, availability planning activities, and process activities" (Riposo et al., 2017, p. xv). Both public and private shipyards face a number of interrelated challenges maintaining the fleet's operational availability cost and schedule performance due to unanticipated work requirements, workload-to-capacity mismatch, and workload fluctuations. Given the challenges the Navy and public shipyards face in maintaining the fleet, the case can be made that opportunity exists for leveraging contracting strategies with private shipyards to improve fleet availability. It is no secret that NAVSEA's relationships with private shipyards is critical to fleet availability, yet is often at odds with the shipyard contracting strategy. Misalignment between NAVSEA and private shipyards greatly impacts cost, schedule, and performance, and thus is a major area of interest.

Consider the following challenges of improving operational availability:

- Procurement of maintenance, repair, and overhaul (MRO) services differs from the procurement of most standard products. Production cycle times are much longer, and volume orders of magnitude lower. Low volume, high variety production complicates assessment of delivery performance as the ship repair and maintenance scope problems are often not known at the time that a contract is signed (Verma & Ghadmode, 2004).
- Uncertainty of ship maintenance and repair operations (MRO) and subsequent target costs mean firm-fixed price contracts that do not allow for scope increase for underestimated or unexpected work and result in delays due to long contract modification cycles associated with approval required for work items, such as issuing required changes to job orders or to address omission of work (SUBMEPP Commander, n.d.).
- Planning for availability to complete expected work, as far out as two years before the start of the availability, makes it impossible to foresee the magnitude of unexpected scope creep that will be identified when problems arise during the execution phase and how initially planned schedules of work will be impacted (Caprio & Leszcynski, 2012). It is also difficult to maintain the requisite infrastructure and necessary workforce capacity under this uncertainty (Buckley, 2015; Martin et al., 2017).
- When projects run behind schedule from delayed work, the amount of available overtime to accelerate completion is limited, and the lead-time necessary for outsourcing of labor with adequate skills and training is significant (Riposo et al., 2017).

- Investments in manpower and infrastructure required to support fleet availability needs are causing private shipyards to demand long-term contracts and Navy incentives (GAO, 2010).

These challenges are not new. Similar challenges affect all capital projects that rely on third party providers, including infrastructure, energy, and healthcare. In this paper, we will present an analysis of observed conflicts in the problem situation, associated assumptions arising in our current system to challenge, and how to do so, with supporting evidence from various system actors. These challenges can be consolidated as a core dilemma for NAVSEA to focus holistic resolution efforts and promote more synergistic contracts between NAVSEA and private shipyards. The research explores critical questions that address appropriate compensation to private shipyards for business and operational risk while improving fleet availability and reducing overall costs for NAVSEA.

Literature Review

Early in this project, the authors became aware of the different history stories and the need to examine a longer span of time than simply beginning at the transition from MS-MO to MAC-MO contract vehicles. Thus, we examine a wider swath of the ship maintenance processes, a review of ship maintenance contract history, human capital management, infrastructure investment, quality assurance and oversight and competition in the ship repair industry. This literature review will examine historical contracting strategies with shipbuilders and contextual aspects of the ship maintenance and repair landscape, including apparent conflicts and misalignment, preventing a more productive naval ship maintenance industry.

History of Private Sector Contracting Strategies With the Navy

NAVSEA awards approximately $24 billion in contracts annually for new construction ships and submarines, ship repair, major weapon systems, and support services (GAO, 2013). Of this, approximately $4 billion is spent annually for depot-level maintenance contracts, much of which is for repair of the nuclear fleet, performed at the four public shipyards in Norfolk (NNSY), Portsmouth (PNSY), Puget Sound (PSNSY), and Pearl Harbor (PHNSY; Riposo et al., 2008). A variety of contract fixed-price and cost reimbursement contracts are utilized, depending on the amount of responsibility each party assumes for the costs of performance risk, when risk is assumed, as well as the timing and amount of incentives offered to the contractor for achieving work at or above a specified standard (Buckley, 2015). Traditionally, cost reimbursement contracts have placed enormous risk on the government in the event that private shipyards are unable to deliver the ship on time and within budget (GAO, 2013).

Until 2004, all Chief of Naval Operations (CNO) maintenance availabilities were conducted under single-ship, firm-fixed contracts written on a ship-by-ship basis and limited by the amount of available funding for scheduled maintenance. This strategy demonstrated excessive cost growth and conflicting objectives between the Navy and industry contractors, both of which were detrimental to fostering a collaborative government-industry partnership (Duncan & Hartl, 2015).

In 2004, Multi Ship Multi Option (MSMO) contracting replaced the old system with hopes of building long-term relationships with contractors, investment in their workforce and facility modernization, and a more reliable, predictable MRO industrial base. The single, five-year contracts reimbursed all allocable, allowable, and reasonable costs expended by the contractors, plus provided the opportunity to earn an extra fee for exceptional management and technical performance. This improved collaboration and ownership, yet required close management to ensure efficiency and cost performance, which the Navy was not adequately resourced to properly provide (Duncan & Hartl, 2015).

In 2010, the Navy concluded that the surface ship force readiness was below acceptable levels, in addition to other issues about cost and schedule performance, leading it to replace MSMO with a new contracting strategy for ship repair, known as Multiple Award Contract, Multi Order (MAC-MO; GAO-17-54, 2016). According to the GAO 2016 report, several lessons were learned from pilot maintenance periods to support modification of contract processes to allow for longer planning windows to finalize work requirements for more stable requirements and pricing (GAO-17-54, 2016). MAC-MO also incorporated the input of commercial ship operators that benefit from a larger competition base (Duncan & Hartl, 2015).

In 2013, MAC-MO was officially implemented. Despite the Navy's initial optimism, cost overruns, delays in completing availabilities, and emergent maintenance issues remain a common problem in ship maintenance contracts. In the National Defense Authorization Act (NDAA) of 2019, the Committee on Armed Services House of Representatives expressed skepticism of the Navy's MAC-MO mechanism as a single-best contracting approach. Its sentiments resonate with the issues identified in the GAO report issued in 2016 (GAO-17-54) that identified apparent delays in renegotiating contracts while vessels sit idle in the yard and third party planning contractors' negligence in obtaining long lead time materials when needed. Utilization of MAC-MO in the maintenance industry implies an increased stakeholder base and a lack of systemic interfacing, which is at odds with federal internal control standards' mandate to evaluate risk responses and progress toward program objectives (GAO-17-54, 2016).

The National Defense Authorization Act (NDAA) of 2019 directed the Comptroller to produce a report to better understand the necessary adjustments to the current ship maintenance and repair process, in particular, assessment of

1. the Navy's execution of the MAC-MO strategy as it compares to the
2. previous Multi-Ship, Multi-Option strategy, with particular emphasis on cost, lost operational days, and on-time completion;
3. the effectiveness of third-party planners in the MAC-MO strategy, including their performance in developing stable, well-defined requirements during advance planning;
4. the adequacy of the Navy's structure for contract oversight;
5. the stability and viability of the ship repair industrial base, including private industry's capacity to recruit and retain critically skilled workers and maintain safe and efficient facilities; and
6. the advantages, disadvantages, or key differences between the MAC-MO and Multi-Ship, Multi-Option strategy depending on the location where the work will be performed.

(NDAA, 2019, p. 91)

Strategizing modernization and repair to quickly increase fleet operational ability as seen in previous eras is no longer possible due to the complexity of modern warships, which are equipped with rapidly evolving advanced propulsion and weapon systems, sensors and radars, and specialized materials for strength, stealth, and acoustics, among other major advancements (Barrett, 2011). The downsizing of NAVSEA in the 1990s and a shift in industry-led acquisition strategies are cited as major contributors to increased cycle time

(Keane et al., 2018). Availabilities now require more preparation, resources, coordination, and competence than ever before.

Shipyard inability to meet the Navy's demand with the current system will be explored in the next several sections, with emphasis on key aspects of the contractual relationship between the Navy and shipyards and the challenges each face.

Human Capital Management

Contract cost is comprised of four main categories: labor, material, contractor overhead, and Navy-furnished equipment, all of which are the responsibility of the Navy (GAO, 2005). A major area of budgeted costs is allocated for contractor labor, including labor hours for production, engineering and other direct support, and costs based on labor hours and the workers' labor rates (GAO, 2005). The 50/50 rule, formally known as 10 U.S.C. § 2466, requires that at least half of all Navy maintenance work be performed at a public depot, meaning shipyards are required to spend a minimum of 50% of all funding received for depot maintenance organically by their permanent workforce (Riposo et al., 2008; Porter, 2016). This limits the amount of money the private sector can receive, although typically much more than 50% is allocated to public shipyards: The average reported workforce composition in 2007 was largely government personnel (NNSY: 90%, PHNSY: 86%, PNSY: 77%, and PSNSY: 86%; Riposo et al., 2008), and more recent data continues to report around 70% (Moore, 2015).

Navy policy requires that depot maintenance be performed in a ship's homeport when possible, based on available internal capacity; otherwise the work may be solicited for open competition in the private sector. Nuclear ship repair also traditionally is allocated to the public shipyards as regulations require public depots to maintain core capabilities. While the intent is to keep the government yards operating at near capacity and to capitalize on efficiencies and other sunken facility costs and overhead, the large volume of business allocated to the public yards has detrimental effects on the ship repair industry as a whole, especially in times of imbalance between shipyard workload and capacity.

Even though the total shipyard workforce is growing, overall workforce productivity has decreased for several reasons. Inexperienced workers often replace experienced ones and represent a large portion of the total workforce: As of 2016, 32% of data on the public shipyard employees had fewer than five years of experience (GAO, 2017). To put this in perspective, personnel with one year of experience, on average, are historically only 25% as productive as those with seven years of experience and achieve approximately one-third the productivity as those with four years of experience, if given accelerated training (Riposo et al., 2017). Shipyards are limited in their ability to hire and train the numbers of people required to replace lost productivity to meet near-term peak demands that are rarely accurately forecasted, budgeted, and resourced.

Depot-level maintenance is forecasted and budgeted by the number of man-days required two years ahead of execution. Year after year, the budgeted amounts for maintenance and repair are consistently below what is ultimately required to perform the work, due to increased operational cycles, unidentified maintenance, unplanned events, and age of the fleet, which also contributes to unplanned maintenance. It is difficult to maintain the necessary workforce capacity when incorrect assumptions from the past surface into the present reality. Take for example, the evolution of the Ohio-Class Maintenance Plan, which was revised to reflect a 13% increase in the required man-days of maintenance over the life of each boat, from the 2004 estimate of approximately 406,000 man-days to the 2007 estimate of 459,000 man-days (Riposo et al., 2017). For a fleet of 18 boats, this means almost one million added man-days. In addition, the estimated time to perform the mid-life engineered refueling overhaul (ERO) was revised from 28 to 33 months (Riposo et al., 2017).

Even with experienced personnel, there remains the learning curve associated with new work services, such as CVN inactivation and support for the new generation ships; the Ford-class CVNs and the Virginia-class submarines also contribute to delays. Peak workloads and under-capacity arising from discrepancy in work packages, evolving maintenance plans, design and engineering issues, inadequate planning and scheduling, production process layout, and training continue to fall on the shoulders of the public shipyard workforce (Moore, 1996). These issues require cautious mitigation strategies, yet are not given adequate consideration in forecasting, budgeting, and contracting. Mitigation for future challenges and risks of boat shortage involves looking at how human capital and workforce planning are managed, as well as looking beyond productivity as the driver of unacceptable performance data.

Infrastructure Investment

Shipyard ability to perform repair, maintenance, and modernization depends on the extent of the available facilities and the complexity of the maintenance requirements. Contracted private shipyards must possess a Master Ship Repair Agreement (MSRA) granted by NAVSEA to perform all aspects of naval shipboard work, from minor to complex repairs and alterations. This certification demonstrates a depot's facilities, management, organization and production capabilities to repair steel, aluminum, and fiberglass hulled vessels for desired capacities. Shipyards are granted MSRA based on the ability to produce integrated (rather than individual) work packages for structural, electrical, electronics, machinery and piping work, installation, integration, and testing (Navy Regional Maintenance Center [NRMC], 2015). Shipyards are also assessed on their competence and understanding of the complex nature of machinery and systems and their ability to subcontract out and provide oversight for work not directly performed (NRMC, 2015).

High capital investment in critical infrastructure is required to maintain specialized equipment, cranes, and drydocks able to accommodate vessels of specified dimensions, rigging equipment, and various shops for structural shop fitting, machinery, piping, electrical equipment and electronics, welding/NDT, sheet metal, insulation, and painting (NRMC, 2015). Many repair activities must be performed in a drydock, such as hull inspection and repair and removal of marine growth. Larger ships, such as the DDG-51, the CG-47, and the LCS-2, are limited to the drydocks to which they may be assigned, exacerbating the already saturated supply of available drydocks, especially on the East Coast, where smaller ships may consume the submarine and carrier docks as a last resort in the future (Martin et al., 2017). The ability of the Navy's public shipyards to support the Navy's readiness needs is continually challenged by capability and capacity constraints in an environment in which public shipyards are in degraded and neglected condition.

Facilities and associated maintenance are contractor overhead expenses (GAO, 2005). For the private sector, decisions to invest are influenced by whether or not properly-structured incentives are provided to increase investment in critical infrastructure to support the needs of the future fleet (Martin et al., 2017). For that reason, the supply of drydocks available for depot maintenance are largely subject to government influence through decision making to invest in public shipyards (Martin et al., 2017). It has been suggested that the government needs to do more to directly support the domestic ship maintenance industry, given the forecast of reduced defense spending and lack of partnerships with commercial shipyards (Moore, 2015).

Quality Assurance and Oversight

Although the Navy and leading commercial buyers agree that quality is the responsibility of the contractor, the Navy routinely accepts ships not meeting quality criteria that are later addressed following delivery, resulting in increased costs and disruption to availability (GAO, 2013). Accepting ships with known deficiencies can interfere with the command maintenance plan for other maintenance and repair activities, as well as upgrades and crew training.

This issue is far less common for the commercial ship industry, where risks to quality belong to the contractor, as does the premium paid to transfer this risk (GAO, 2013). The payment structure in commercial ship buying is also structured to enforce timely correction of deficiencies. Instead, the commercial shipyards are expected to deliver a defect-free (or nearly so) vessel at delivery and are incentivized to provide timely correction of deficiencies by contract and payment structure (GAO, 2013). In contrast, the Navy typically assigns less cost risk to contractor quality problems under cost-reimbursement and fixed-price incentive contracts and pays a significantly larger proportion of the total cost upfront.

Commercial shipyards producing and maintaining oil production storage and offloading vehicles, large cruise ships, gas carriers, and offshore oil drilling ships and the buyers they serve operate in an environment vastly different from the Navy in terms of oversight and quality assurance. Commercial firms are substantially more effective in resolving quality deficiencies before delivery of ships, potentially by their differences in practice that could benefit the Navy, including (GAO, 2013)

- contracting approaches that place cost risk associated with addressing quality problems on the shipbuilder,
- incentives for timely resolution of problems,
- oversight processes with clear lines of accountability, and
- emphasis on observing in-process work.

In 2007, Supervisors of Shipbuilding, Conversion, and Repair, Puget Sound (NAVSEA's SUPSHIP Command), as the responsible authority for procurement and administration of new construction and ship repair contracts with private sector shipyards, drew to light the lack of resources needed to improve quality as identified in quality assurance inspections. In 2009, the Back to Basics effort was initiated. The program identified key quality assurance goals and developed a means to help SUPSHIP improve communication with program offices to enhance quality assurance and oversight. NAVSEA also introduced many standardized operating procedures across SUPSHIP's locations. The GAO found significant variation in SUPSHIP locations with respect to commercial practices, like the use of design drawings and random inspections, as the Navy had not defined the role they should play (GAO, 2013).

Designated classification societies surveying to monitor rules, regulations, and statutory requirements are often incorporated in commercial shipbuilding contracts to provide a robust oversight process, provide engineering services for the development and testing of new technologies, and provide technical assistance to reduce potential risks to quality such as by attending and witnessing inspections (GAO, 2013). Leading commercial ship owners reported that these services are never used as a substitute for their own oversight and quality assurance processes, as their rules may not consider their buyer-specific technical requirements.

SUPSHIP's quality assurance department is closest to the work being performed, but has limited influence on the shipbuilder and on early contracting decisions made early with respect to quality (GAO, 2013). As of 2013, a standardized quality performance standard, proposed by SUPSHIP, had not been incorporated in any shipbuilding contract (GAO, 2013). At the time of the report, the Navy had shared its intent to establish a quality team within the NAVSEA Logistics, Maintenance and Industrial Operations Directorate (previously NAVSEA 07, now 04) to promote attention to quality assurance; however, the roles, responsibilities, and authorities of the team had not been defined. The GAO suggested that the hypothetical team, if given sufficient authority and tasked with elevating SUPSHIP quality assurance concerns throughout the acquisition process, may be conducive to enhancing contractor performance and contract requirements for managing quality. It was indicated that diffused responsibility for quality in the Navy program offices, NAVSEA, SUPSHIP, INSURV, and others may be due to distraction from or supersession of competing roles and distraction with concerns for monitoring schedule, costs, or other strategic needs (GAO, 2013).

In the NDAA of 2019, the Committee directed the Under Secretary of Defense for Acquisition and Sustainment to provide a briefing by December 1, 2018, on the feasibility of the DoD's Superior Supplier Incentive Program, designed to provide contract incentives for superior contractor performance in terms of cost, schedule, performance, quality, and responsiveness (NDAA, 2019, p. 192). Previously, the Secretary of Defense was directed to conduct a review of the extent to which sustainment matters are considered in decisions related to the requirements, acquisition, cost estimating, and programming and budgeting processes for major defense acquisition programs (NDAA, 2017). The report was to include an assessment of private sector best practices in assessing and reducing sustainment costs over the lifecycle of complex systems and the organic industrial base's capabilities, capacity, and resource constraints as required by the materiel commands (NDAA, 2017). The story of how these lessons will be used to inform decisions to modify contracting policy to shape the future of defense procurement is not yet known and is constrained by intricate complexities in the relationship between the Navy and its industrial base, as discussed in the next section.

Competition

Contracting relationships also have implications for the competition within the environment. Besides the key differences between commercial and defense shipbuilding oversight and operating practices, there are also some differing contextual factors worth noting, especially with respect to competition. Measures taken for ensuring quality to protect reputation are critical to remain viable in the commercial shipbuilder's competitive environment; thus, they are pressured to meet contracted delivery dates and deliver on schedule.

The Navy's shipbuilding industrial base is vastly limited in comparison to the number of qualified commercial shipbuilders, that are also reliant on the Navy to remain in business. This codependency enables quality deficiencies and performance variances as the Navy has an interest in sustaining its limited shipbuilding base.

MAC-MO contracting has been criticized by former MSMO contract holders for introducing competition and uncertainty that could result in decisions to reduce their workforce and facilities (GAO-17-54, 2016). The previous strategy did not require the MSRA certification for noncomplex availabilities. As of March 2016, it was unclear how NAVSEA would handle the assignment of noncomplex availabilities to smaller businesses without the certification, as no indefinite delivery, indefinite quantity (IDIQ) contracts had yet been granted (GAO-17-54, 2016).

Horns of a Dilemma: Analyzing Ship Availability Contract Performance Using Goldratt's Conflict Cloud

This section discusses the current state of knowledge with respect to what has been reported as challenges in the current system, the needs of the Navy and private shipyards in addressing these challenges, and how the approaches to address needs and overcome challenges create seemingly intractable dilemmas.

Dr. Eli Goldratt, inventor of the Theory of Constraints, argued that complex systemic problems can be analyzed by understanding the dilemma that is preventing organizations from solving those problems. These dilemmas, he posited, are caused by trying to satisfy two valid underlying needs that are both necessary to achieve the overall goal but seem to conflict with each other. He suggested that instead of compromising on either of the needs, organizations should resolve the dilemma by identifying and challenging various assumptions that put those needs in conflict.

The Conflict Cloud, popularized by Dr. Goldratt, offers a way to approach complex problems through uncovering the perceived sources of the problem, examining them from the perspective of conflicts between the sources and mitigations, and then surfacing the unspoken assumptions. It is this step, the surfacing of unspoken assumptions, that frequently provides a path to a breakthrough change in the conditions and an improvement in the situation. The conflicts are often represented in a specific way, called a conflict cloud. A generic conflict cloud follows, adapted from Smith (1999 Figure 1).

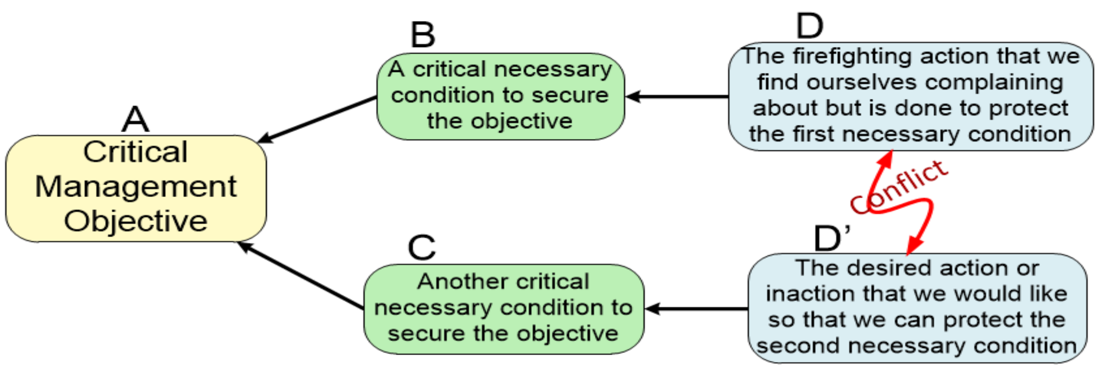


Figure 1. **Generic Conflict Cloud**
(Adapted from Smith, 1999)

Beginning on the left-most box of Figure 1 the presumed common goal is presented. The next two boxes to the right (B and C) present two apparent needs that must be accomplished to achieve the goal. The final two boxes (D and D') present two wants that must be accomplished to meet the needs. These two wants are often in opposition with each other, by mutual exclusion or due to resource contention (Andersen & Gupta, 2013). Conflict within the context of naval ship maintenance and modernization presents challenges to confront the current system and identify strategies to improve contractual relationships between NAVSEA and private shipyards.

We use the conflict cloud as a starting point to engage participants from an outsider perspective of the current landscape of the relationship between the Navy and private shipyards. From our perspective as experienced practitioners and with contextual evidence as framed in the literature review, we generated the "prototype" conflict cloud shown in Figure 2.

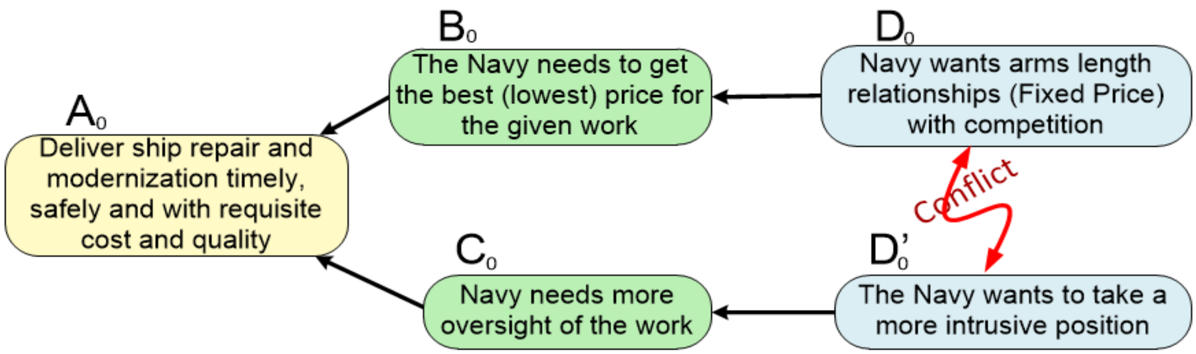


Figure 2. **Initial Conflict Cloud Perspective**

This conflict cloud was used as an entry point into semi-structured interviews, as discussed in the Research Approach section. In this example, the conflict is an internal one for the Navy, but also crosses organizational boundaries. The goal of the authors was to first identify a conflict from an external perspective, with the hope of identifying other conflicts, to be able to refine the initial interpretation of underlying conflict and articulation of the stakeholders' competing needs. The Three Cloud approach (Smith, 1999) facilitated this effort by guiding exploration of additional conflict themes that emerge in the analysis. By exposing additional conflict and assumptions, this approach supports communication among stakeholders with competing objectives without creating resistance as assumptions are surfaced. It also supports continuous improvement by analysis of the reinforcing loop where perception of an inability to change current processes is embedded.

Research Approach

Methodology

Taking a systems-based approach, we aimed to develop a rich picture from various perspectives by asking the following questions:

1. Does the perspective of the conflict as articulated make sense?
2. How would you resolve the conflict you see?
3. Do you have anecdotes that you would be willing to share?
4. What most frustrates you about the current situation?

A phenomenological research design involves a researcher's inquiry into the lived experiences of individuals about a phenomenon as described by participants (Creswell, 2013), in this case, challenges in aligning the goals of NAVSEA and private shipyards. A group of 30 individuals were solicited for input based on their experience and perspective and invited to participate in semi-structured interviews lasting from 30 to 60 minutes. Ten respondents representing both industry and the Navy responded, and eight individuals were interviewed or provided significant responses. The respondents included representation from a variety of MSRs, as well as both current and former Navy officials responsible for the execution of both the former and current contract strategies.

In order to encourage open responses, the respondents have been granted anonymity, and some discussions of the responses may be altered to prevent specific language from being used to identify the specific respondents. Responses to the questions and other input relevant to the topic were documented from each interview and processed in the analysis.

Interview Results

The respondents were remarkably open about the current situation and their individual and organization's role in the history that led to today's situation. The overwhelming fraction of respondents began by addressing their own shortfalls rather than seeking to cast blame on a different participant. Thus, quotes included,

> "We lost over half our contracting officers, and thought that was okay, and we could not replace them anyway."
>
> "The MSMO contracts assumed that our performance would improve over the contract, and to be honest, it did not."

Beginning with these comments to set the stage, we explored the circumstances around the previous contracting strategy and how it did not lead to the desired results of improving availability performance with delivery times that met the Fleet's needs. Comments here focused on the following:

- Understanding of the ship's material condition was poor.
- The ability to respond to emergent material problems was highly variable.
    - One maintainer was able to persuade the board to purchase repair parts that were expected to be consumed over several availabilities; however, this was not adopted broadly, and the accuracy of forecasts varied.
- Not all participants (Navy and industry) understood the business.
- Simplicity was favored.
- Availability costs were higher than budgeted. A variety of reasons were provided:
    - Unlike carrier or submarine programs, the engineering planning activity had been disestablished under BRAC, thus dissolving the capability to maintain class maintenance plans centrally.
    - Regional or even single port engineers could use different approaches to maintain their ships.
    - Poor cost control—loss of experienced personnel on the government side exacerbated this factor.

These comments set that stage for the transition from MSMO to MAC-MO. Those respondents that covered the transition used language like

> "The cure is worse than the disease."
>
> "The current [MAC-MO] model is win-lose."

"Hey, sorry that spec was bad, but tough, use an REA[1] at the end."

Rather than dwell here, we transitioned to developing other potential conflict diagrams. Several respondents accepted the draft conflict diagram in Figure 2, suggesting minor adjustments or clarifications. Others, especially those familiar with the cloud method, presented additional conflict diagrams.

One respondent was focused on the business arrangements and the differences that arise between the Budgeted Work Package and the Actual Work Package, especially as the availability progresses and modifications are made (i.e., contract changes) to the work package (Figure 3).

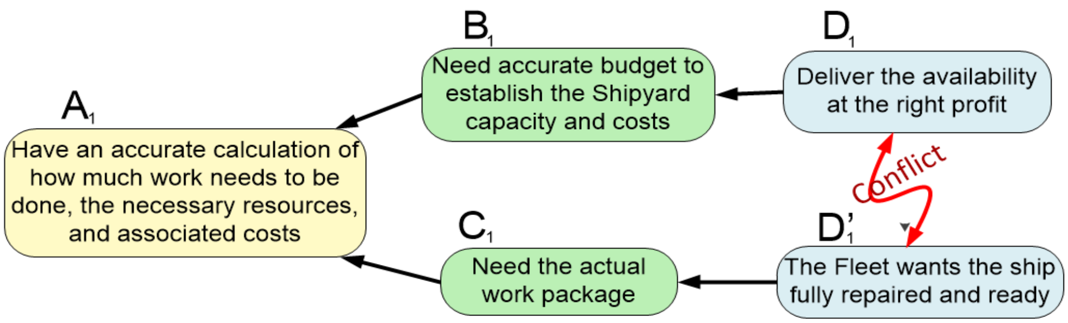


Figure 3. **Conflict Cloud Focused on Work Packages**

Another cloud (Figure 4) was developed from the perspectives of retaining/permitting a robust ship repair industry while delivering repairs at the lowest price.

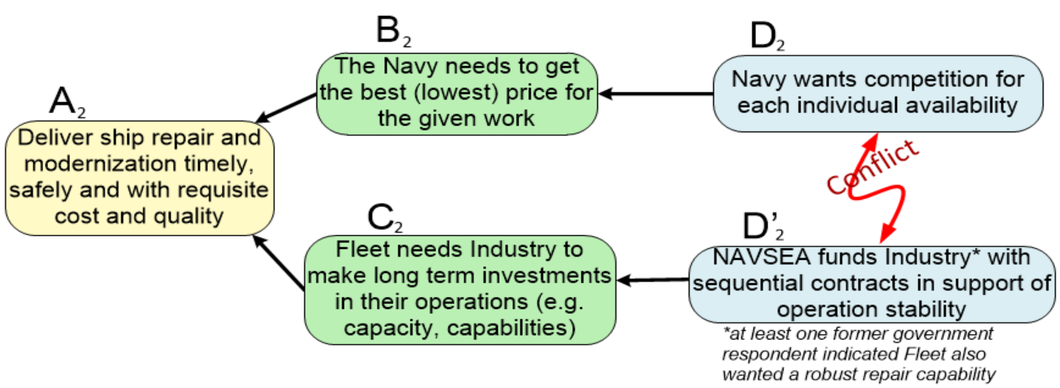


Figure 4. **Conflict Cloud Focused Between Pricing and a Robust Industry**

[1] REA (Request for Equitable Adjustment)—Under CFR and the FAR, a contractor can submit for an equitable adjustment to the terms of the contract item, in terms of costs, markups and time for completion.

Another respondent cast the conflict by capturing the different organizations that had different goals (Figure 5). This respondent noted that while the Fleet would like lower prices, there is a much greater emphasis at the Fleet on having operable ships that can respond to its changing needs.

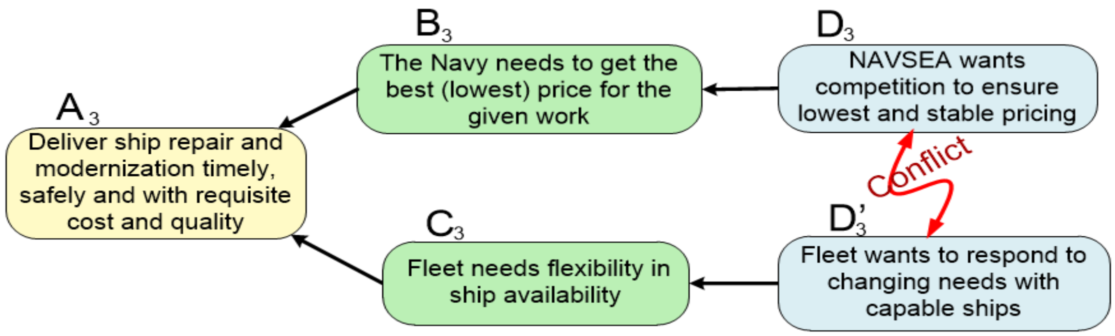


Figure 5. **Conflict Cloud Focused on Pricing and Fleet Availability**

The interviewees looked forward to evolution from the current contracting system, with improved strategies to facilitate a symbiotic relationship between the government and industry and the emergence of a more mutually beneficial procurement environment.

Core Conflict Cloud

The Three Cloud Approach (in this study, "Four Cloud") facilitated the exploration of additional conflict themes that emerged in the analysis. Using this approach, the core conflict was synthesized from four specific conflict clouds which, when combined, convey a fundamental issue leading to undesirable effects in the contracting environment. The core conflict cloud (Figure 6) shows refinement of previously identified conflicts and expresses two core wants (entities D_c and D'_c) that are prerequisite to satisfying opposing parties' core needs (entities B_c and C_c), both of which must be met to achieve the common goal (entity A_c).

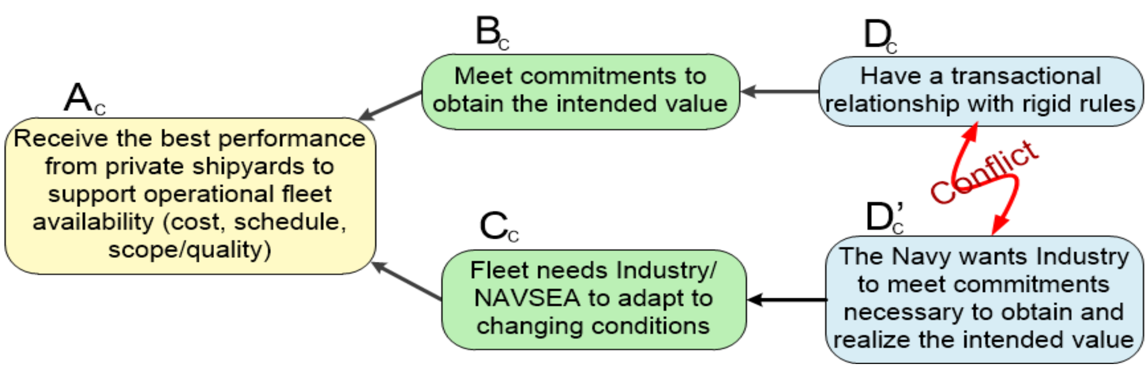


Figure 6. **Core Conflict Cloud**

Language Analysis

We used the transcripts of the conversations and the Qualitative Analysis tool Nvivo® to examine the language of the respondents. Nvivo® has the capability of analyzing language for similarity by the speakers. We limited the analysis to words five letters or longer and allowed synonyms (using a built-in dictionary) to group the respondents by the similarity of the words they used (Figure 7).

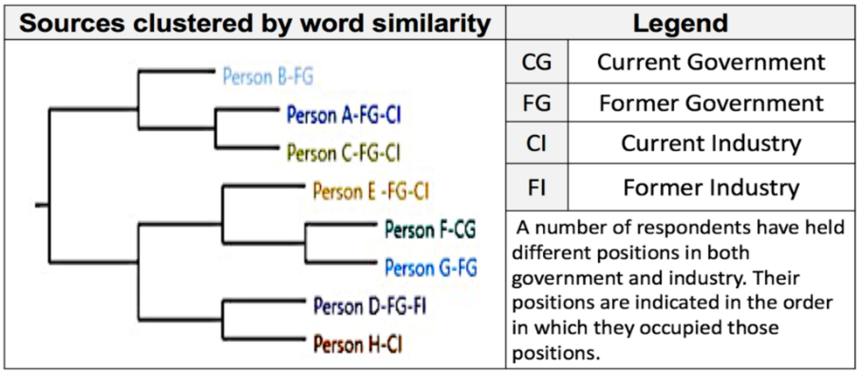


Figure 7. **Respondents Clustered by Word Similarity**

The groupings are not surprising. Person F worked for Person G, and E, F, and G are all close in their relative seniority within their community. Likewise, A, B and C are very close in their seniority and careers within the industry. As far as the authors know, D and H are unknown to each other, but use similar language. This analysis supports an evolving hypothesis that government and industry are not as far apart as might otherwise be expected.

A second part of the language analysis was to use another Nvivo® functionality and examine the frequency of word usage among the respondents. In this analysis, we searched the transcripts for words over five letters long and used synonyms to build a word cloud limited to the top 50 words, where the size of the font indicates the relative importance of the word. The word cloud and relative word frequency are illustrated in Figure 8.

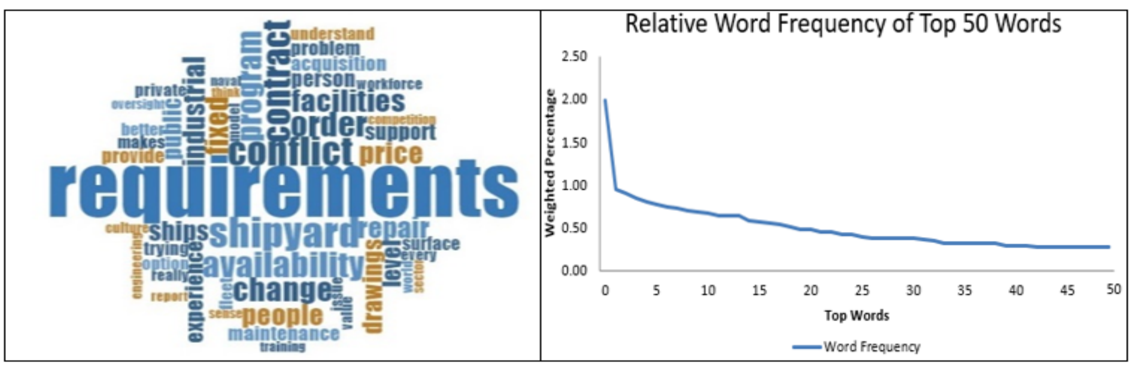


Figure 8. **Word Cloud Representing Top 50 Words in Respondent Conversations and Respective Relative Word Frequency**

The significant break between Word 1, "requirements," and all the other words led us to extend the search to consider the top five words used by the respondents, which are Requirements, Shipyard, Conflict, Contract and Availability (meaning either the ship repair period or the operational availability, as we did not distinguish the two dominant meanings in this industry). We elected to not to filter or censor words like "shipyard" and others commonly expected to arise, as it confirmed the focus of the respondents and was leveraged to develop recommendations.

Inductive Generation of Contracting Strategies

During the interviews, many respondents answered Question 2, "How would you break the conflict?" with recommendations (Table 1). The majority responded with one or two ideas based on the original conflict cloud, or a new cloud surfaced during the interview.

Table 1 Participant Recommendations for Contracting Strategy Reform

Participant and Role		Recommendation(s)
A	Former Government Current Industry	Hybrid contract—some elements of fixed price, some elements of MSMO Stability and predictability are needed, but also compromise and negotiation.
B	Former Government	(none provided)
C	Former Government Current Industry	Change perspective from transactional and optimizing locally. There can be a mutually compatible, win-win approach to value the total system.
D	Former Government Former Industry	Establish a flexible reserve managed close to the waterfront to handle new work, as part of an overall reserve, to mitigate lack of scope understanding at the front end.
E	Former Government Current Industry	Change the level of trust: allow a threshold above a fixed price to prevent 30-45 day delays to execute contract changes.
F	Current Government	Develop a mechanism to have a backlog for industry to invest; renew experience in both private and government participants; look at how to allow industrial investment for horizontal building of industrial base.
G	Former Government	Contract for a level of effort each year with Option Years for good performance (similar to Naval Shipyards) to generate a backlog to sustain workforce, training, and facility improvements.
H	Current Industry	Utilize a hybrid approach of fixed price for reasonably quantifiable work, and time and materials or cost-reimbursable for the rest. Figure out how to level playing field by using pilot projects with independent teams to identify work needed, including implied but not articulated, to improve requirements analysis.

One respondent began the interview with the statement, "Everybody has a silver bullet, and there is none!" This statement provides an excellent starting point to begin defining characteristics of the solution space. Given the complexity of the problem, the total system at large must remain the focus of the solution set, with more local or detailed elements admitted to the set based on their support of the global system rather than a focus on local optimization. The solution space must also consider constraints and underlying assumptions applied from the system environment, which may not be limited to the industry side of the equation (e.g., contracting officers).

It was generally agreed by participants that well-understood work items should be contracted as fixed price elements, with different treatment for uncertain items within the contract vehicle. An existing assumption is that a contract vehicle is required to be either transactional (fixed price) or highly collaborative (cost reimbursable). An associated assumption is that a fixed price contract improves contract performance for the Navy by transferring risk. In reality, the "contract type" is applied at the contract line item number (CLIN; Braxton et al., 2017). Contract forms can distinguish cost reimbursable items and provide caveats to incentives. There also exists the potential to develop a new contract form or utilize alternative contracting mechanisms used in other defense programs, such as multi-year procurement (MYP) or block-buy contracting (BBC), which Congress permits the DoD to use for limited programs, yielding significant reduction in cost (O'Rourke & Schwartz, 2017).

At a high level, constraint-based, pipelined availability scheduling provides stability and improves shipyard ability to provide the fleet responsiveness. Using a contract that purchases capacity over a multi-year period could include features of negotiated operating (facility, or fixed costs), material fixed price cost, a cost buffer (as discussed earlier), and a combination of "over" and "under" share ratios tied to delivery performance, known as a shareline. Use of higher tier schedules is based on constraints, an obvious one being the number of available drydocks and the unwillingness of industry to invest in expanding drydock capacity.

Associated with the change in contract form, the use of a reserve (buffer) with a share of the portion allocated to the rapid resolution of straightforward, relatively simple problems would hasten the ability of the shipyard to execute work revisions when a small problem is encountered and bypass the contract modification cycle. A data records system that allows recording, sharing and reviewing of these contract changes (Graham et al., 2018) would also enhance efficiency in these situations. A share of the cost buffer can be reserved for the more significant issues that require longer timelines to resolve, and the processes should incorporate agility as well as accuracy.

Ackroyd (2018) has defined collaborative contracting as one with a focus "on the desired outcomes of greater integration and therefore collaboration of the parties to an agreement. That subset of agreements where success for one is inextricably linked with the performance of all" [emphasis in original]. The collaborative contracting approach enables two parties with different motivations to achieve a mutually beneficial outcome. ISO 44001:2017 specifies requirements for this form of contracting to effectively identify, develop, and manage collaborative business relationships within or between organizations, facilitated by an adoption of leading behaviors to reduce barriers and build consensus (Ackroyd, 2018). The uncertainty and need for collaboration in ship maintenance procurement requires a view of contract form through multiple lenses as a key enabler of value.

The set of strategies inductively generated challenges the current contracting status quo and can be used directly to improve ship availability performance. This strategy set can also be used in other DoD environments with shared industry and government governance.

Limitations, Future Research, and Conclusion

Class maintenance plans are developed under assumptions about workloads and capacity, but without a static workload demand, are based on imperfect knowledge. "Those who are involved with executing this work believe the increase in duration is a result of executing with insufficient resources," explains Riposo et al. (2017, p. xv).

Shifting to a more collaborative, hybrid contracting regime offers the potential for long-term viability of the industry with improved performance meeting the Navy's needs and is in coherence with the Agile Manifesto's core value of *collaboration over contract negotiation* (Beck et al., 2001).

Addressing the core of conflict preventing a more collaborative environment to improve a present situation requires a critical look at three separate issues, known as "the layers of resistance" in the theory of constraints: (1) "What to change?" (2) "What to change to?" and (3) "How to cause the change?" (Goldratt, 1984). Clear identification of the problem, the solution, and how to implement the solution should be understood by the agents of change to find the necessary buy-in from key stakeholders. From a governance perspective, the repeated cycling through contract strategies implies that a system is not well-regulated. Examination of possible improvements in the complex system governance structures can assist in damping the oscillations of contract policy.

Three areas limiting this study also offer the opportunity to conduct future research on collaborative contracting:

1. The research approach was inherently qualitative, although the authors recognize that recent efforts have pushed for improved data availability to explore defense contracting geometry strategies with data-driven approaches (Braxton et al., 2017).
2. The authors' contact set did not reach two critical sets of stakeholders, contracting officer and Congressional staff. An interesting extension would be to expand the set of stakeholders and engage them for insights on expanding the core conflicts. This would offer the potential for identifying more conflicts that could be used to improve the collaboration recommendations.
3. At the start of this research, the Navy's efforts to improve the surface ship maintenance construct was largely invisible to the authors. The extensive Navy efforts, including the Private Shipyard Optimization Initiative (PSO) and the Private Sector Improvement Program (PSI) are now recognized to exist but did not inform our recommendations. As such, understanding how those efforts address the core conflicts also are of research interest.

In conclusion, the authors truly appreciate the generosity and openness of those current and former participants in the surface ship community. They shared their knowledge openly and extensively. Any errors or mischaracterizations are the fault of the authors. We look forward to learning of the success in the community in this critical endeavor.

References

Ackroyd, B. (2018). *Collaborating/relational contracting in practice*. Presentation to IPT, University of Canberra, Canberra, ACT, Australia.

Andersen, S., Gupta, M. C., & Gupta, A. M. (2012). A managerial decision-making web app: Goldratt's evaporating cloud. *International Journal of Production Research, 51*(8), 2505–2517.

Barrett, J. M. (2011). Doubling the fleet: An analysis of the causal factors behind the U.S. Navy's warship building program from 1933–1941. Fort Leavenworth, KS: Nimble Books.

Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., ... Kern, J. (2001). Manifesto for agile software development. *Agile Alliance*. Archived from the original on June 14, 2010. Retrieved June 6, 2010.

Braxton, P. J., Hetrick, K. S., Webb, K. R., Whitehead-Scanlon, I., & Ross D. J. (2017). *Contract incentives under uncertainty: Data-driven contract geometry best practices*. International Cost Estimating and Analysis Association (ICEAA).

Buckley, S. R. (2015). *A high-value best-value approach to public shipyard human capital management to improve ship availability*. Monterey, CA: Naval Postgraduate School.

Caprio, J. L., & Leszczynski, D. (2012). *A preliminary investigation into CNO availability schedule overruns*. Monterey, CA: Naval Postgraduate School.

Coleman, P. J., Hortin, J. D., & Pascioni, L. R. (2016). *The dilemma of Department of Defense business system modernization efforts: Why intended outcomes have not been fully met and what needs to change*. Monterey, CA: Naval Postgraduate School.

Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Thousand Oaks, CA: Sage.

Duncan, M. E., & Hartl, R. (2015). *Multiple award, multiple order contracts—The future of Navy surface maintenance procurement*. Monterey, CA: Naval Postgraduate School.

GAO. (2005, February). *Defense acquisitions: Improved management practices could help minimize cost growth in Navy shipbuilding programs* (GAO-05-183). Washington, DC: Government Printing Office.

GAO. (2010, July). *Defense acquisitions: Guidance needed on Navy's use of investment incentives at private shipyards* (GAO-10-686). Washington, DC: Government Printing Office.

GAO. (2013, November). *Naval shipbuilding: Opportunities exist to improve practices affecting quality* (GAO-14-122). Washington, DC: Government Printing Office.

GAO. (2017, September). *Navy ship maintenance: Action needed to maximize new contracting strategy's potential benefits* (GAO-17-809T). Washington, DC: Government Printing Office.

Graham, T., Sydor, S., Beninger, M., & Glaze, P. (2018). *Optimizing contract modifications under one universal mod*. NGAP.

Guest, G., MacQueen, K. M., & Namey, E. E. (2012). Introduction to applied thematic analysis. *Applied Thematic Analysis*, *1*, 3–18.

Keane, R. G., Jaquith, P. E., Tibbitts, B. F., Mierzwicki, T., Hough, J. J., & Arcano, J. T. T. (2018). Rapid warship acquisition: A case for fundamental change in design and acquisition policy. *Naval Engineers Journal*, *130*(4), 57–72.

Martin, B., McMahon, M. E., Riposo, J., Kallimani, J. G., Bohman, A., Ramos, A., & Schendt, A. (2017). A strategic assessment of the future of U.S. Navy ship maintenance. *Challenges and opportunities.* Santa Monica, CA: RAND.

Moore, N. A. (2015). The U.S. government's role in the ship repair industry. *Naval Engineers Journal*, *127*(4), 59–65.

Moore, R. (1996). *National* Shipbuilding Research Documentation Center *(NSRDC): Report of technology projects by NSTC survey topic* [Final report].

Naval Sea Systems Command. (n.d.). SeaPort-e harnessing power, navigating change. Retrieved January 28, 2019, from SeaPort-e Harnessing Power, Navigating Change website: http://www.seaport.navy.mil/

Navy Regional Maintenance Center (NRMC) Commander. (2015). *Master agreement for repair and alteration of vessels; Master ship repair agreement (MSRA) and agreement for boat repair (ABR)* (CNRMC Instruction 4280.1). Norfolk, VA: Department of the Navy.

Office of the Chief of Naval Operations (OPNAV) N431 (Fleet Readiness and Logistics). (2010). *Maintenance policy for United States Navy ships* (OPNAVINST 4700.7L).

O'Rourke, R., & Schwartz, M. (2017). *Multiyear procurement and block buy contracting in defense acquisition: Background and issues for Congress*. Washington, DC: Congressional Research Service.

Porter, P. H. (2016). *Organic versus contractor logistics support for depot-level repair: Factors that drive sub-optimal decisions*. Maxwell Air Force Base, AL: Air War College.

Riposo, J., Alkire, B., Schank, J. F., Arena, M. V., Kallimani, J. G., Blickstein, I., Hall, K. C., & Grammich, C. A. (2008). *U.S. Navy shipyards. An evaluation of workload- and workforce-management practices.* Santa Monica, CA: RAND. Retrieved from www.dtic.mil/get-tr-doc/pdf?AD=ADA491676.

Riposo, J., McMahon, M. E., Kallimani, J. G., & Tremblay, D. (2017). *Current and future challenges to resourcing U.S. Navy public shipyards* (No. RR-1552-NAVY). Santa Monica, CA: RAND.

Smith, D. (1999). *The measurement nightmare: How the theory of constraints can resolve conflicting strategies, policies, and measures*. Boca Raton, FL: St. Lucie Press.

SUBMEPP Commander. (n.d.). *Joint fleet maintenance manual (JFMM)* (COMUSFLTFORCOMINST 4790.3), *Volume II: Contracted ship maintenance* Portsmouth, NH: Department of the Navy.

Verma, A. K., & Ghadmode, A. (2004). An integrated lean implementation model for fleet repair and maintenance. *Naval Engineers Journal*, *116*(4), 79–90.

An Evaluation of Mature Performance-Based Logistics Programs

William Lucyshyn—is the Director of Research and a Research Professor at the Center for Public Policy and Private Enterprise in the School of Public Policy at the University of Maryland.

John Rigilano—is a Research Analyst at the Center for Public Policy and Private Enterprise at the University of Maryland.

Abstract

For more than 15 years, performance-based logistics (PBL) contracting has been used to reduce weapon system sustainment costs and increase system reliability. In its simplest formulation, PBL "explicitly identifies what is required, but the contractor determines how to fulfill the requirement." Often, the most significant improvements occur relatively early on in the PBL program. Typically, PBL programs evolve along a common trajectory. With new systems, cost-reimbursement contracts are used in order to provide the government customer and the provider with a cost baseline. Once the costs, risk factors, and system failure modes and rates have stabilized, the program generally transitions to the use of fixed-price contracts where providers are paid a fixed cost or fixed rate (e.g., per hour, per mile) so long as operational readiness is achieved at the specified level(s). Over time, the provider makes improvements to its supply chain, logistics networks, operations, and the system itself in order to reduce its costs and maximize profitability. In the "terminal stage" of its evolution, the exemplary PBL is characterized by high availability, reduced inventories, and efficient sustainment processes. This research examines three PBLs that reached this stage, including one program that reverted to the use of cost-plus contracts in an attempt to reduce costs. We found that long-running PBLs continue to deliver value, high reliability, and improved performance, and that distortions to the PBL paradigm (i.e., reverting to approaches that are more transactional) are unwarranted and may lead to unintended consequences that include higher future costs and decreased system readiness.

Introduction

Described by the Department of Defense (DoD) in 2001 as the "preferred approach to product support," performance-based logistics (PBL) represents a radical change in contracting for maintenance, sustainment, and other after-sales support services. In its simplest formulation, PBL "explicitly identifies what is required, but the contractor determines how to fulfill the requirement" (Macfarlan & Mansir, 2004, p. 40). DoD guidelines state that "the essence of performance-based logistics is buying performance outcomes, not the individual parts and repair actions. … Instead of buying set levels of spares, repairs, tools, and data, the new focus is on buying a predetermined level of availability to meet the [customer's] objectives" (Defense Acquisition University, 2005).

There is now clear empirical evidence that PBL strategies, when properly implemented, can dramatically reduce system sustainment costs while improving overall reliability and performance (Guajardo et al., 2011; Boyce & Banghart, 2012; Lucyshyn, Rigilano, & Safai, 2016). It is noteworthy, then, that PBL contracting is not being aggressively pursued across the DoD. The overall number of PBL programs has waned considerably since its peak in 2005, when there were more than 200 programs in place compared to fewer than half this number by 2012 (Erwin, 2013). In dollar terms, PBL contract obligations have gradually declined in recent years after peaking in 2013 (Hunter, Ellman, & Howe, 2017; see Figure 1).

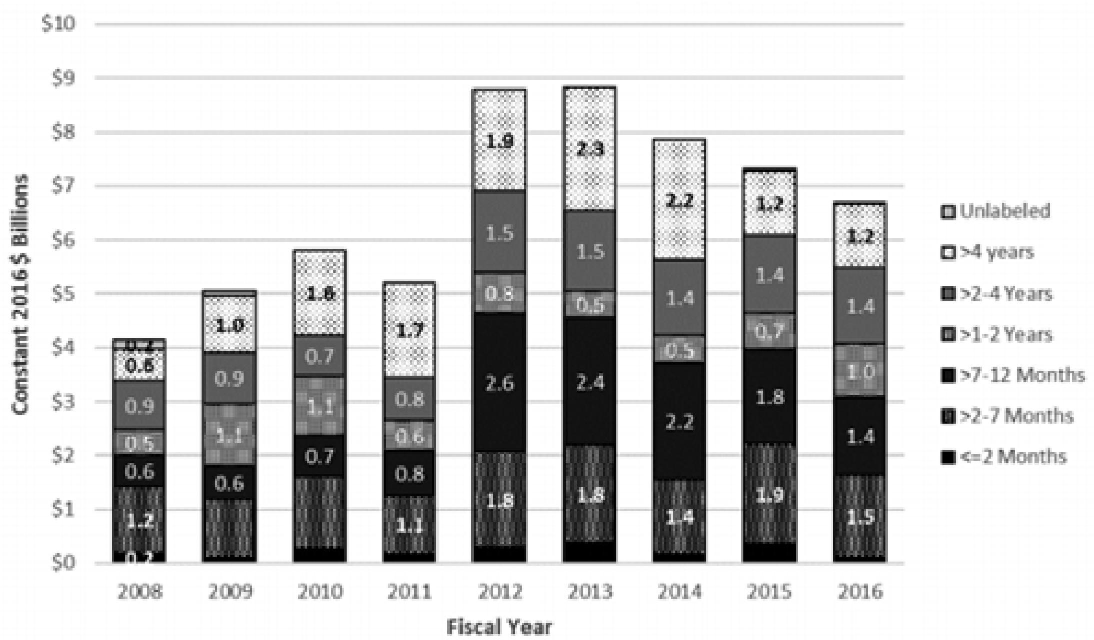


Figure 1. **DoD PBL Contract Obligations by Initial Maximum Duration, 2000–2016**
(Hunter et al., 2017)

In the early 2000s, criticism of PBL focused on contractor reliability (Gansler, Lucyshyn, & Vorhis, 2011). Critics argued that by allowing contractors flexibility, the military places itself in a dangerous position, relying too heavily on contractors who may become unreliable in the future. Others voiced concerns over whether contractors would be able to perform at the same high level during contingency and combat operations, especially if deployed in theater. Military planners feared that the “lack of control due to outsourcing could … put an entire military operation at risk” if, for example, contractors were to pull out of a war zone (Singer, 2008). To date, research indicates that these concerns are largely unfounded (Lucyshyn et al., 2016). Time and again, PBL-supported systems operating in stressful environments have met or exceeded performance requirements, contributing to mission success.

Critics, including some within government, have moved to questioning the value that is obtained through PBL, as programs mature and the benefits, in terms of both cost reduction and performance improvement, become less significant. Could it be that once the “low hanging fruit” has been picked, incremental improvements become more difficult to achieve; that reverting to traditional, transactional contracting approaches makes more sense? Selviaridis and Wynstra (2015) note that it is unclear whether “performance-based incentives in long-term contractual relationships are sustainable over time as supplier learning occurs and service improvements become marginal” (p. 3520). This report addresses this concern. Ultimately, it seeks to determine if, and how, product support contracts should be modified over time in order to provide continuous value to the customer.

PBL is still in its infancy. And given the fundamental change in functions and responsibilities—for example, the customer no longer manages (or in many cases even owns) inventory—it is not surprising that the optimal PBL contracting approach, specifically its development over the product deployment life cycle (as uncertainty in support costs change), has yet to be fully examined, let alone articulated.

Report Approach

The objective of this report is to determine whether a "steady-state" PBL—one that generates continuous value to the customer—can be achieved, and if so, how to structure the optimal arrangement. This study relies primarily on structured interviews with program personnel in both the public and private sectors; the application of the academic literature (on contracting, management science, agency theory, and transaction cost economics) to PBL; and in-depth case studies of three mature PBL contracts.

Background

Over the last two decades, the DoD has focused on reducing the cost of weapon system logistics by constructing more sophisticated contracts with more favorable terms for the government (Butler, 2013). In addition, the military services are increasingly diverting their attention to sustainment costs—which are continuing to increase across the DoD—in part because the services cannot afford to replace rapidly aging systems. The DoD has identified PBL as its preferred approach to supporting weapon system logistics.

PBL Basics

PBL contracting, when used appropriately, can reduce sustainment costs relative to traditional, transactional approaches. PBL is a logistics support solution that transfers inventory management, technical support, and the supply chain function to a provider who guarantees a level of performance at the same, or reduced, cost. Instead of buying spares, repairs, tools, and data in individual transactions, the customer purchases a predetermined level of availability in order to meet the warfighter's objectives.

The optimal PBL contract is a multi-year agreement wherein the user purchases sustainment in an integrated way, to include elements of the system's supply chain. Long-term agreements allow the provider to incur up-front investment costs in the beginning stages of a PBL contract that are later offset by future cost avoidance. Whereas traditional sustainment contracts incentivize the provider to sell parts, PBL's "pay for performance" approach aligns the objectives of the service provider, with those of the customer; and motivates the provider to reduce failures and resource consumption.

As outlined in the *Defense Acquisition Guidebook*, a PBL's performance is often measured through one or more of the following criteria.

- **Operational Availability:** Percent of time that the system is able to sustain operations tempos or is available for missions
- **Operational Reliability:** Measure of a system in meeting objectives set for mission success
- **Cost per Unit Usage:** Total operating costs divided by the individual unit of measurement for a specific weapons system (flight hour, miles driven, etc.)
- **Logistics Footprint:** Government or contractor presence required to sustain/deploy the system
- **Logistics Response Time:** Time from logistics demand sent to completion of demands (labor, support, etc.)

A successful PBL contract relies on performance metrics that are straightforward, measurable, and achievable. Additionally, these metrics must be carefully developed, implemented, monitored, and evaluated. Continuous communication between the program office and the support provider is crucial to ensure that these metrics are negotiated and executed in a manner that will ensure successful implementation of the PBL contract (Gansler & Lucyshyn, 2014).

PBL Advantages

When implemented, PBL shifts the focus of the government's efforts from transactions, to identifying performance outcomes and assigning responsibilities. The objective is to develop accountability, instead of relying on control. With PBL, active management of the sustainment process (e.g., forecasting demand, maintaining inventory, and scheduling repairs) becomes the responsibility of the support provider. Traditional logistics support dictates processes and specifications, which has the effect of constraining innovation and process improvement. Suppliers and equipment manufacturers are incentivized to sell more repair parts as opposed to developing and implementing reliability improvements. PBL changes the incentives for the supplier. The supplier is now incentivized to improve the reliability of systems and reduce inventories of spare parts, in order to increase profit.

The DoD is gradually moving away from its traditional hierarchical command and control structure and towards a more adaptive system that will provide the precise, agile support required for the distributed, network-centric operations. In this regard, there are four distinct advantages associated with the use of PBL contracting:

- **Delineates outcome performance goal.** The objective of PBL programs is to buy measurable outcomes based on warfighter performance requirements. They should, at the top level, be based on warfighter performance requirements, and include only a few simple, realistic, consistent, and easily quantifiable metrics.
- **Ensures responsibilities are assigned.** PBL metrics, when properly developed, clearly define the suppliers' responsibilities.
- **Reduces cost of ownership.** This reduction results from the decline in inventories, improved supply chain efficiency, replacement of low-reliability components, and increased system availability.
- **Provides incentives for attaining performance goal.** The PBL program should fundamentally align the interest of the supplier with that of the customer, and lead suppliers to assume greater responsibility for providing ongoing improvements to their products. PBL provides incentives for the supplier to improve design and processes and implement commercial best practices (Lucyshyn et al., 2016).

There is ample empirical data that demonstrates that PBL, when properly implemented, produces desired outcomes in the key performance areas of availability, reliability, logistics footprint, and cost. Major systems including the C-17 and F/A-18, for instance, have all reduced sustainment costs by hundreds of millions of dollars, while other systems and subsystems such as the F-22, UH-60 avionics, and F-404 engine have seen drastic improvement in availability and cycle time (i.e., logistics response and repair turnaround; Fowler, 2008). Empirical analysis has demonstrated that PBL contracts incentivize reliability improvements of 25%– 40%, compared to more traditional transactional approaches (Guajardo et.al, 2012). Other government reports (e.g., Office of the Secretary of Defense [OSD], 2009) and think-tank studies have concluded that PBL offers distinct benefits that are difficult to achieve using traditional transactional approaches.

PBL Contract Trajectory

Ensuring a PBL contract is structured properly and contains the correct incentives is crucial to its long-term success. The Center for Executive Education from the University of Tennessee (2012) identified three factors inherent to a successful PBL contract:

- **Alignment:** Both the contractor and government have embraced PBL as a new form of provider-client relationship and not merely a variant of business as usual.

- **Contract Structure:** The appropriate balance of risk and asset management is achieved, an environment is established that allows for creativity and shared success, and a pricing model is utilized that considers incentive types.
- **Performance Management:** Desired outcomes and metrics for reporting and improving are established (Hunter et al., 2018).

Typically, PBL programs evolve along a common trajectory. With new systems, cost-plus reimbursement contracts followed by cost-plus incentive contracts are used to enable the government customer and the service provider to collect sufficient data to develop a cost baseline. Once the costs, risk factors, and system failure modes and rates have stabilized, the program should transition to the use of fixed-price contracts where providers are paid a fixed cost or fixed rate (e.g., per hour, per mile) so long as operational readiness is achieved at the specified level(s). Over time, the provider makes improvements to its supply chain, logistics networks, operations, and the system itself in order to reduce costs and increase profitability. A typical PBL contract pricing structure includes three components:

- Share-in-savings, to incentivize the provider to reduce overall sustainment costs
- A fee, to reward provider for meeting performance expectations
- A fixed-price or fixed-price per operating hour contract schedule, to provide payment to provider regardless of quantity of parts or services consumed (Gansler & Lucyshyn, 2017).

In the "terminal stage" of its evolution, the exemplary PBL achieves consistently high availability, and efficient maintenance processes and supply chains. The program operates at lower risk, from both a cost and technical perspective. When this stage is reached, obtaining further performance improvements and price reductions will require increasing levels of innovation, since, presumably, the "low hanging fruit" has been picked.

Since 2000, 68% of DoD PBL contract obligations have been awarded as firm-fixed-price contracts, with cost-plus-incentive and cost plus award-fee being the next most common contract types (Hunter et al., 2018). As Figure 2 indicates, PBL contracts can be implemented at the component, subsystem, and system level.

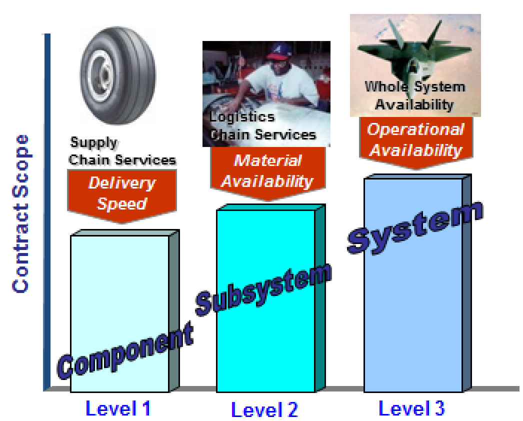


Figure 2. **Level of Implementation and Contract Scope**
(Gourley, 2014)

Risk, Profit, and Contract Type

Note that as the PBL matures, the contractor takes on more risk, which is reflected in the type of contract that is used (see Figure 3). As risk increases, so does the contractor's opportunity to increase profit.

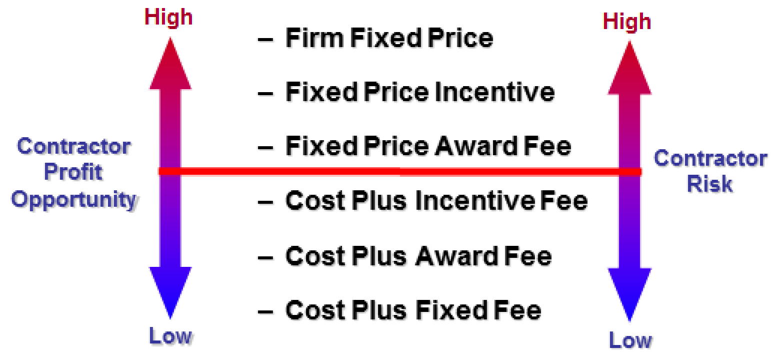


Figure 3. **Relationship Between Contract Type, Risk, and Profit Opportunity** (Gourley, 2014)

Some within government have become concerned by "excessive" profits generated by PBL contracts, even in instances where overall program costs have been reduced. This concern can manifest itself in disagreements between contracting officers (KOs) and program managers (PMs) over the type of contract that should be used—the former asserting that cost-plus contacts should be used to constrain windfall profits.

It should be noted that the KO binds the government to a contract, the legal document that specifies program requirements. In many instances, however, the KO generally does not report administratively to the PM who, of course, is responsible for overall program performance and success, including contract execution. From the KO's perspective, success is often construed narrowly. Was the contract awarded? Were protests avoided? Have costs been minimized? (National Academies of Sciences, Engineering, and Medicine, 2016). In fact, KOs, at times, dictate contract type and terms to the PM, which can lead to negative program outcomes (e.g., contracts may not take advantage of some of the flexibility available in the FAR, or be of the most appropriate length[1]). Needless to say, affordably providing the required capability to the warfighter should be emphasized over minimizing profits. As stated by the DAU (2018), "The Services' primary concern is to pay less for more when compared to their current sustainment strategy, irrespective of industry profits" (p. 30).

This is not to suggest that cost-plus contracts should be avoided altogether. As discussed, for new programs, a cost-plus contract may be essential to determining a cost baseline that can be used to develop future fixed-price contracts. In addition, when risk cannot be quantified or the cost of transferring the risk to the supplier "is more than the

[1] PBL contracts need to be long enough to enable the contractor to recover any investment made in product and process improvements. These contracts are, consequently, competed less frequently, which conflicts with guidance to compete frequently.

government can accept," cost-plus contracts are preferable (DAU, 2018). Cost-plus contacts may also be preferable, as the component or system approaches disposal and emphasis turns to containing the costs associated with wear-out and obsolescence. As a system approaches retirement, cost-plus contracts may allow the government to better balance costs, risk, and performance requirements. There is also some theoretical evidence indicating that cost-plus contracts may be well suited to certain types of product support programs, namely simpler ones for which the scope of work is limited. Kim, Cohen, and Netessine (2007) model how the customer observability[2] of two variables—the contractor's cost reduction efforts and spare parts inventory—affect optimal contract choice. They show that when the supplier and the customer are risk neutral, "which may be the case in practice if the customer and the suppliers are well-diversified corporations" the combination of a "fixed payment and a performance component" (i.e., a typical PBL contract) is optimal, provided that the contractor's cost reduction efforts and inventory levels are unobservable (Kim et al., 2007, p. 1857).

Reliability and Ownership

The reliability of a system appears to be correlated with the ownership of spare parts. That is, when the supplier owns a larger portion of spare parts, reliability is higher. Kim et al. (2011) found that "the full benefit of a PBC [performance-based contracting] strategy is achieved when suppliers are transformed into total service providers who take the ownership of physical assets" (p. 1).

Stryker: A Cost-Plus PBL

When Stryker brigades supported by a PBL contract first deployed to Iraq, Army officials reported operational readiness rates averaging 96% from October 2003 through September 2005 (GAO, 2006). In addition, the Army consistently noted that contractors were providing impressive levels of support and according to a 2006 GAO report, more knowledgeable and efficient than their military counterparts with regard to the specifics of the Stryker vehicles (GAO, 2006).

From a cost perspective, however, contract performance is less clear. In 2012, The DoD Inspector General asserted that the follow-on contract's continued use of a sole metric (readiness) in combination with a high-ceiling, cost-plus contract unduly incentivized the contractor to accumulate significant excess inventory valued at $335.9 million (DoD IG, 2012). The Army responded that the excess inventory could be attributed, in part, to contractor improvements in reliability, and that the spare parts would be used eventually, albeit at a slower pace than anticipated (DoD IG, 2012).

Given the Army's heavy reliance on Stryker during the Iraq War, changing operational tempos, and the lack of historical cost data, the use of a cost-plus fixed fee contract (as opposed to a fixed-price

When non-performance based contracting strategies (transactional contracts) are used; reliability remains low with suppliers relying more heavily on a larger inventory of customer-owned spare parts to maintain the system. When non-performance contracting strategies are used, suppliers are not incentivized to improve reliability.

[2] Kim et al. define an "observable" variable as one "that is verifiable and hence can be specified in a contract" (p. 1849).

Under fixed-price PBL contracting strategies, the optimal combination of reliability and inventory shifts away from inventory and toward improved reliability. In other words, the supplier makes investments in reliability (process, schedule, or technology), thereby obviating the need for a large parts inventory. The optimal combination shifts even farther to the right when spare parts are owned by the supplier. Cost of ownership is lowest under a PBL contracting strategy where spare parts are owned by the supplier. Kim et al. assert that when the supplier owns all spare parts, “the supply chain becomes coordinated.” They conclude that “Our analysis supports a DoD recommendation for transforming suppliers into total service providers of support services who, under the PBL arrangement, assume complete control of service functions, including asset ownership” (p. 1). At present, industry practice is for the customer to own spare assets “while the supplier decides on target stocking levels of spares and recommends to the customer a budget of spares acquisitions to achieve these levels.”

It should be emphasized that these relationships hold only when fixed-price contracts are used. PBL arrangements that use cost-plus contracts can provide suppliers with the perverse incentive to accumulate spare parts, if those parts are customer owned (see Stryker inset).

Long-Term PBLs

In this section, we provide an in-depth examination of three mature, long-running PBL programs: The High-Mobility Artillery Rocket System, better known as HIMARS, the Navy Aviation Tires Program, and the Apace helicopter’s Modernized Target Acquisition Designations Sight (M-TADS) system. The HIMARS PBL supports two major subcomponents, the Launcher-Loader Module and the Fire Control System. The Apache PBL provides subsystem-level support.

HIMARS

HIMARS is the latest addition to the military’s multiple-launch rocket system (MLRS) family. Designed with the purpose of engaging and combatting artillery, trucks, air defense, light armor, and personnel carriers; it was a lighter, more mobile variation on the MLRS M270A1, with some common components. In addition to supporting troop and supply concentrations, HIMARS has been in constant demand by both the Army and the Marine Corps (as well as foreign governments) since the production of its first prototype in 1999.

The HIMARS launcher is an impressive weapon that has continuously exceeded its operational readiness expectations. Initially developed through an advanced concept technology demonstration (ACTD) program by Lockheed Martin Missile and Fire Control in 1996, HIMARS has been referred to as “the most advanced artillery system in the U.S. arsenal.” Following their successful deployments during Operation Iraqi Freedom, HIMARS launchers have become indispensable to the arsenals of both the Army and Marines.

A Brief History

Originally conceived to meet the need for a lighter, rapidly deployable rocket launcher—HIMARS is a wheeled, agile, rocket and guided missile launcher fixed to a five-ton armored truck (Gansler & Lucyshyn, 2014). Owing to its wheeled chassis and lightweight design, the system can be easily transported by C-130, allowing it to be deployed to previously inaccessible areas at a moment’s notice (Lockheed Martin, 2011). The HIMARS has been internationally recognized for its highly efficient and innovative features, including the ability to take aim at a target in under 16 seconds, and rapidly move away from the launch site once a missile is released. In addition, its fire controls system, electronics, and communications units are interchangeable with its heavier, tracked predecessor, the M270A1.

Following the ACTD in 1996, Lockheed Martin was awarded an engineering and manufacturing development (EMD) contract for six launchers (and later an additional two launchers) in 2000 (Army-Technology, 2015). Not long after, in 2003, "the U.S. Army and Marine Corps signed a contract for the low-rate initial production (LRIP) of 89 launchers for the Army and four for the USMC" (Army-Technology, 2015). As the United States' role in overseas conflicts grew in the mid- to late 2000s, the need for HIMARS units grew (Army-Technology, 2015).

Since its introduction into the force in 1998, HIMARS has proven its value through both peacetime forcible-entry exercises and on operational deployments in the U.S. Central Command (CENTCOM) area of responsibility (Russo & Hilbert, 2008).

Program Description

The Lockheed Martin HIMARS program office is headquartered in Dallas, TX, where numerous program functions are executed; these include program management, depot repair coordination, inventory control, contracting with suppliers, design interface, and database maintenance. The program database tracks the location of each launcher, including each spare part, indicates whether the part is functional, and provides its status with regard to the repair process. The DoD's internal logistics systems rarely achieve this level of visibility for most weapon systems, often leading to ordering redundancy, misplaced orders, and an incomplete picture of program operations.

The program also employs 31 field service representatives (FSRs) that operate with deployed units stateside and overseas. In-theater maintenance work is performed primarily by soldiers, while the FSRs facilitate the supply process by overseeing numerous functions (Hawkins, 2009). These functions include the following:

- supply, receipt, storage, issue, inspecting, packaging, and shipping of subsystems and components;
- data collection and recording (maintenance actions, supply transactions, operating hours, munitions status [deployment and garrison]);
- system fault isolation using a variety of either built in or stand-alone test equipment;
- replacement of assemblies, as required;
- provision of technical assistance and support (both launcher and automotive); and
- provision of an interface for "reach back" engineering support, enabling the rapid resolution of problems.

Given the level of sophistication provided by the Lockheed Martin's database and logistics networks, the FSRs are able to streamline and simplify the repair process for launchers. As a result, early in the PBL program, Lockheed Martin was able to reduce the number of diagnostic test units provided to each battalion, from six to one. In fact, soldiers operating the system in theater need only remove and replace defective line-replaceable units.

Perhaps one of the greatest benefits of the HIMARS program is the provision of limited depot-level repair capability at each battalion, where repair work is provided by the FSR. Referred to as the capability to "Fix Forward," some 50% of all HIMARS repairs are performed on location by the FSRs, eliminating wait times and significantly reducing costs. Moreover, the FSRs are trained to test and replace circuit card assemblies (CCAs), rather than the line replaceable units (LRUs) in which they are housed, which reduces the overall logistics footprint and lowers costs —only the CCAs need to be shipped. This in-the-field repair capability has also significantly improved deployed launcher availability. According to interviews with Lockheed Martin officials, FSRs voiced few concerns over their work environments, safety, or civilian status within the battalion, with several volunteering to return for a follow-on tour.

PBL Strategy

The Army awarded the first HIMARS PBL contract to Lockheed Martin for $96 million in February 2004 (Gansler & Lucyshyn, 2006). The four-year contract (one base year and three option years), referred to as Life Cycle Contractor Support (LCCS) ended in December 2007. At this point, the Army had acquired 195 HIMARS launchers; and the Marines had acquired 40. Given its increasing inventory of HIMARS, the existence of a successful partnership between the Army and Lockheed Martin, and the cost benefits that derive from economies of scale, the Marines sought to support its launchers through LCCS upon completion of the initial contract.

Accordingly, the second contract (LCCS II), a three-year contract (one base year with two option years) worth $90 million, was awarded in January 2008 to support both the Army and Marines' systems. The shorter duration of LCCS II reflected significant risk associated with unknown launcher production quantities and price fluctuations for component spares (Gardner, 2008). A third PBL contract, for $158 million, termed Life Cycle Launcher Support (LCLS), extended HIMARS sustainment through December 2013 for services and through December 2014 for hardware.

The initial PBL strategy relied on firm-fixed-price contracts with performance incentives[3] for stateside operations, and cost-plus fixed-fee contracts for overseas contingency operations (Gardner, 2008). This strategy provided strong cost reduction incentives as well as the flexibility to meet overseas operational requirements. Moreover, the fixed-price was tied to an OPTEMPO category, with each vehicle assigned to a price category based on anticipated usage.

The LCCS/LCLS contracts tasked Lockheed Martin with the full support responsibilities for the performance-based product support of the HIMARS and MLRS M270A1 launchers' fire control systems, as well as the HIMARS launcher-loader module (Gardner, 2008). The commonality of support for the two platforms allowed the Army and later, the Marines, to take full advantage of the potential economies of scale in order to reduce costs (DoD, 2006).

The LCCS/LCLS concept represented a significant evolution from the original M270 MLRS strategy, according to which the majority of tasks (e.g., initial provisioning, inventory management, war reserve stock, repair and overhaul, depot maintenance, etc.) were provided with organic support. LCCS/LCLS, on the other hand, represents an ideal partnership; one in which the contractor assumes responsibility for providing technical support and user training in order to meet performance objectives, while at the same time

---

[3]A fee was paid to the contractor on a quarterly basis provided that the performance requirements were met.

maximizing existing Army depot and acquisition infrastructure by relying on military personnel to operate and repair the system.

Based primarily on data collection provided by Lockheed Martin during the initial contract, the LCCS team was able to make a number of changes to the LCCS II contract that would reduce future ownership costs. Notably, the team determined that the usage hours for the launchers varied significantly between active Army units and National Guard units (OSD, 2009). In an effort to reduce future costs, the less-used units were categorized under a lower operational tempo, which led to a reduction in needed support. Accordingly, Lockheed Martin and the DoD negotiated the LCCS II contract to reflect the anticipated savings derived through the reduction in operational tempo. These savings turned out to be considerable. In 2007—the final year of LCCS I—costs associated with operational tempo totaled $12.4 million; in 2009, these costs had declined to $3.8 million, for a total cost avoidance of $8.6 million.

Initially, the PBL contained three contract metrics: system readiness, response time for part delivery, and repair turnaround time. System readiness was required to be maintained at or above a specified percentage (92% for LCCS I; 90% for LCCS II); however, this requirement was not included in the third contract.[4] With regard to the second metric, the contract required that response time for mission capable parts deliveries fall within a specified range a certain percentage of the time, depending on the type of part. For overseas operations, the response time ranges were extended to provide the flexibility necessary to meet fluctuations in demand that might arise in unpredictable operating environments (DoD, 2006). The LCCS II contract, for example, required that response time be less than 48, 72, or 96 hours for U.S.-based operations, depending on the part (each of which is assigned to an Issue Priority Group), 92%, 91%, and 90% of the time, respectively (OSD, 2009). For overseas operations, the response time had to be less than 96, 120, or 144 hours (OSD, 2009; see Figure 4).

Issue Priority Group	Requirement	Percentage Required
1	48 hours (CONUS) 96 hours (OCONUS)	>92%
2	72 hours (CONUS) 120 hours (OCONUS)	>91%
3	96 hours (CONUS) 144 hours (OCONUS)	>90%

Figure 4. **Response Time Requirement for Mission Capable Parts Delivery**

[4]During this time period, the government sought generally to reduce the number of metrics used in PBL contracts to improve program outcomes and, in the specific case of HIMARS, eliminate the incentive fee tied to the readiness requirement, which was seen as redundant in light of the incentive fees tied to the other two requirements.

The third metric, repair turnaround time, specified the time period for completing LRU repairs. The contract required that LRU repairs be completed within a certain number of days a certain percentage of the time as defined by five “bands” (see Figure 5). This requirement was measured on a quarterly basis. As the figure shows, a majority of the repairs (65%) had a required repair turnaround time of less than 35 days.

Band	Repair Turnaround Time	Requirement (Percentage of Total Repairs)
Band 1	1–7 days	≥18%
Band 2	8–35 days	≥47%
Band 3	36–80 days	≤27%
Band 4	81–90 days	≤8%
Band 5	91 days	1%

Figure 5. **Turnaround Time Requirement for LRU Repair**

PBL Performance

The HIMARS PBL program achieved success early on, reaching a 99% average system readiness rate, with no launcher out of service for more than 24 hours through 2015 (Gansler & Lucyshyn, 2014). With regard to response time for mission capable deliveries and repair turnaround time, the program also performed extremely well. The CONUS average for mission capable delivery stood at 14 hours, the OCONUS average at less than one hour. Field repair turnaround time averaged 1.2 days and vendor repair turnaround averaged 34 days.

The HIMARS program also tracked reliability through mandated field analysis reports, monitoring the mean time between both system aborts (MTBSA) and essential function failures (MTBEFF). Figure 6 illustrates HIMARS units' reliability between 2005 and 2015. Note that reliability among deployed Army units, as measured by both MTBSA and MTBEFF, climbed significantly during 2009 and 2010, before stabilizing at levels that continue to exceed average reliability across all units. The peaks in reliability correspond with peaks in the number of operational hours for deployed units (i.e., 3rd quarter 2009 and 1st quarter 2010).

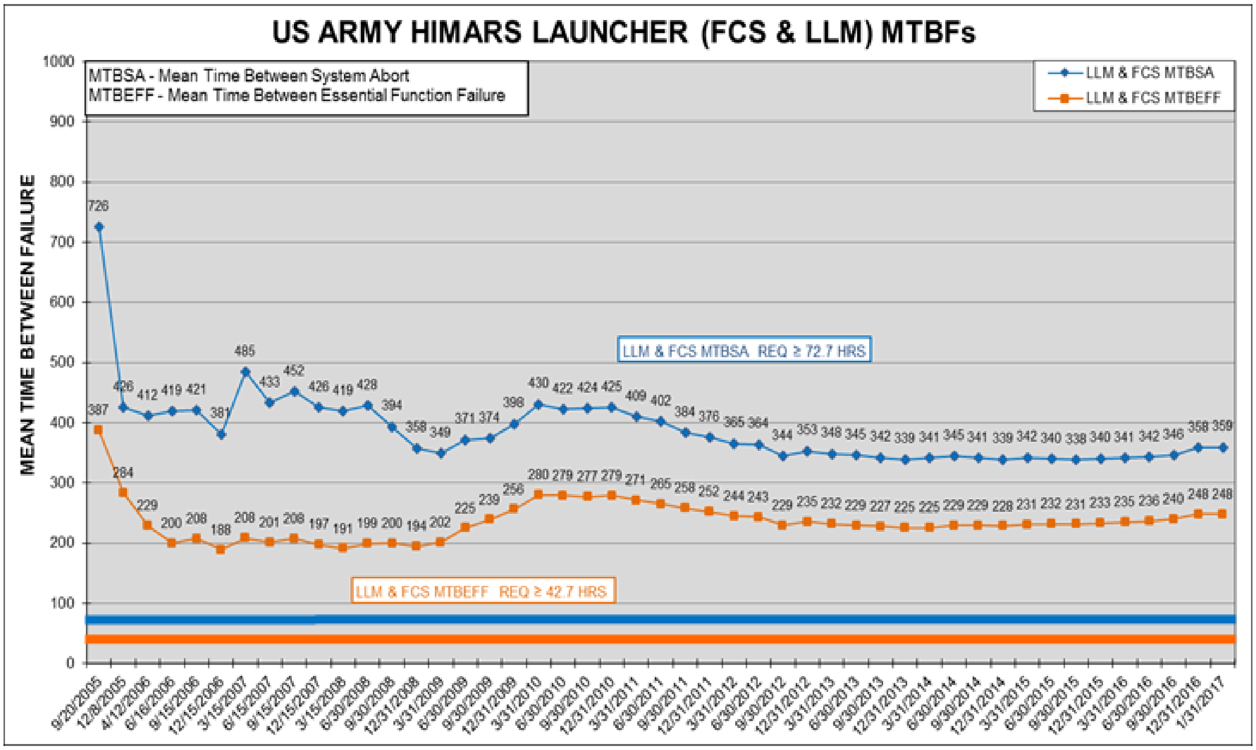


Figure 6. **HIMARS Field Reliability**
(Lockheed Martin, 2017)

Transition to Cost-Plus Contract

Despite the program's success, the DoD transitioned to the use of a cost-plus fixed fee contract in 2014, transferring much of the inventory management function from the contractor to the government, in an effort to further reduce costs through more direct government control. The five-year contract (one base year and four option years) extended support for HIMARS through 2018. Contractor personnel have suggested that the government-contracting officer pressed for the transition in an effort to constrain costs. The program continued to use the response time and turnaround time requirements. The response time (customer wait time[5]) requirement remains unchanged from the previous contract, whereas the repair turnaround time[6] requirement was modified to specify two bands as opposed to five. As with the previous contracts, 65% of repairs required a repair turnaround time of 35 days or less.

Unlike the previous fixed-price contracts, this contract specifies "stock objectives" and other inventory and operational constraints that the contractor must not exceed. This, of course, limits the contractor's flexibility to leverage economic efficiencies when buying spares, virtually eliminating the incentives to invest in program improvements and thus doing

---

[5] Customer Wait Time: The number of hours that LCLS has from the moment an FSR submits a requisition until when that item requested is in the hand of the requesting echelon.
[6] Turn Around Time: The action of repairing a LRU to Condition Code A (serviceable—issuable without qualification) within the allotted time period.

away with one of the primary benefits of performance-based contracts. Because the program shifts most of the risk back to the government, some suggest that the program is a PBL "in name only."

One of the key questions government officials must ask is whether the new arrangement satisfies objectives of reducing cost, while meeting the requirement for HIMARS availability; both in the present and in the future. It may very well be that the government is, at present, receiving sufficient value, and taking on what it considers acceptable risk. Indeed, contractor personnel stated that the government has been able to take advantage of the "residual setup," relying on the same proven processes and expertise, but in a more transactional environment in which spare parts procurement is constrained, ostensibly to reduce program costs.

The program continues to perform well; response time and turn-around time remain well above the requirement and reliability has remained consistent (see Figure 7). During the initial contracts, Lockheed Martin and its subcontractors had invested more than $10 million in design improvements, process changes, equipment and facilities to improve reliability and reduce costs. This resulted in a high level of system availability; reducing support requirements overall and enhancing mission success. The inertia from these improvements have enabled the continued high level of the programs' performance and cost reductions. According to contractor personnel, DoD costs per launcher are less in 2018 than they were in 2005; the total price of the LCLS support contract in 2018 was less than it was in 2006, even though the 2018 LCLS program supported 643 launchers, compared to the 286 launchers in 2005. The question is whether the same processes, level of detail, amount of effort, program improvements, cost reductions, and forward-looking approach can be preserved in a cost-plus environment over the long term.

Customer Wait Time						
			CONUS			
Issue Priority	**Requirement (Hours)**	**Percentage Required**	**Q1**	**Q2**	**Q3**	**Q4**
02 – 03 – 07	48	Greater than 92%	99.6 %	97.3 %	99.1 %	96.1 %
05 – 06 – 09	72	Greater than 91%	100.0 %	95.5 %	98.9 %	97.4 %
12 – 13 – 14	96	Greater than 90%	100.0 %	100.0 %	100.0 %	100.0 %
			OCONUS			
Issue Priority	**Requirement (Hours)**	**Percentage Required**	**Q1**	**Q2**	**Q3**	**Q4**
02 – 03 – 07	96	Greater than 92%	100 %	100 %	100 %	100 %
05 – 06 – 09	120	Greater than 91%	100 %	100 %	100 %	100 %
12 – 13 – 14	144	Greater than 90%	N/A	N/A	N/A	N/A
Repair Turn Around Time						
Bands	**Repair TAT in Days**	**Percentage Required**	**Q1**	**Q2**	**Q3**	**Q4**
Band 1	1-35 Days	Equal to or greater than 65%*	78.4 %	84.4 %	74.8 %	77.8 %
Band 2	> 90 Days	- 25% per occurrence	0.00%	0.00%	0.00%	0.00%

Figure 7. **HIMARS Program Results FY 2017**

One would anticipate that the contractor would be reluctant to make any additional investment in the program without a reasonable expectation of getting a return on their investment. Unsurprisingly then, contractor investment and surge capacity has indeed decreased following the transition to a cost-plus contract. According to contractor personnel, the program has not procured an LRU in five years. And, with a depleted spare parts inventory that is constrained by the contract, availability may not be able to keep pace with demand, should requirements dictate an increased operational tempo.

In addition, although the costs associated with spare parts procurement may accrue more slowly under the current contract, they will likely end up being higher when compared to previous arrangements that permit more cost-effective parts procurement (e.g., "bulk buys"). In other words, the program may no longer be able to capture economies of scale to the same extent.

The government has yet to release a Request for Proposals to continue HIMARS support beyond 2018. Contractor personnel believe that the government intends to ask for a one-year extension to bridge the existing contract as it continues to assess how support will be provided over a longer period, and an RFP for the new contract is released.

Navy Tires

In 2001, the Navy Inventory Control Point (NAVICP) had already used PBL to transform other supply chains, improving performance and reducing costs, turning their focus to aircraft tires (Mahandevia, Engel, & Fowler, 2006). NAVICP was a Command responsible for more than 400,000 items of supply, and had an inventory valued at $27 billion, with $4.2 billion in annual sales. As of July 2011, NAVICP was replaced by the Naval Supply Systems Command Weapon Systems Support (NAVSUP WSS). The mission of NAVSUP WSS is to "provide the Navy, Marine Corps, Joint and Allied Forces program and supply support for the weapons systems that keep our Naval forces mission ready" (NAVSUP, 2014). It should be noted that NAVSUP WSS only enters into a PBL contract after assessing and concluding that a PBL contract cost would be equal to or less than traditional support. Overall, NAVSUP WSS PBL contracts have reduced costs by 3.9% (The Naval Aviation Enterprise Air Plan, 2013).

A Brief History

Traditionally, NAVICP treated aircraft tires as a commodity; buying in bulk, and then storing them until they were needed. This resulted in a large on-hand inventory (approximately 60,000 tires) that may or may not have had the right mix of tires for the fleet. This inventory was maintained through small contracts for individual types of tires, which were awarded to a variety of manufacturers (OSD, 2012). The unintended consequence of this short-term acquisition process was that it sent erratic signals to the industrial base, resulting in less than optimal production runs, higher cost raw material sourcing, and longer lead-times. In addition, distribution services were provided by organic military resources, often with delays. In effect, operational units had to maintain a retail inventory. This resulted in higher overall costs to the fleet.

Program Description

The Navy developed a strategy to transition the provision of aircraft tires to a component level PBL. This strategy was implemented in 2000 and has resulted in a dramatic improvement in the availability of the required aircraft tires, with significant reduction in cost.

Initial Contract

In May 2000, NAVICP issued an RFP for a PBL contract to manufacture and deliver naval aircraft tires to all U.S. Navy, Marine Corps, and foreign military sales customers (NAVICP, 2000). A firm-fixed-price contract was competitively awarded in April 2001 to

Michelin Aircraft Tires Corporation (MATC), Greenville, SC, to manage the Navy's aircraft tire program. This contract had a five-year base with an estimated value of $67.4 million, supporting all 23 types of tires that the Navy used (NAVICP, 2001). This contract had two five-year options, and the resultant 15-year value for the contract was $261.5 million (PBL Award Summary, 2011). The first five-year option was exercised in July 2005, with an award of almost $92 million to MATC (DoD, 2005). The second five-year option was awarded in June 2010 and was valued at more than $101 million (Military Industrial Complex, 2010). This contract ended in January 2016.

This initiative was the first time the DoD contracted out for the support for new and repairable tires. MATC was prime contractor for the program as well as the manufacturer and supplier of the tires. MATC maintained responsibility for requirements forecasting, inventory management, retrograde management, storage, and transportation (Mahadavia et al., 2006). MATC subcontracted with Lockheed Martin to provide the supply chain services. These services included demand forecasting, order fulfillment, and inventory management. In addition, Lockheed Martin also managed the commercial carriers (Bland & Bigaj, 2003).

As part of their contract task, Lockheed Martin provided a service center that was available 24/7, called the Lifetime Support Command Center (LSCC). This center controlled all requisitions and maintained a real-time requisition status with web-based access, and was electronically interfaced with Michelin, the two warehouses, and through the Navy with the Naval Air Stations, Marine Corp Air Stations, carriers, and Landing Helicopter Assaults and Landing Helicopter Docks. This data, along with shipping status and product support information, was provided to Michelin to maintain their internal systems (Gansler & Lucyshyn 2006; Mahadavia et al., 2006; Bland & Bigaj, 2003).

The ambitious contract requirements were as follows:

- 95% on-time fill rate
    - 48 hours (2 days) within the continental United States (CONUS)
    - 96 hours (4 days) outside the continental United States (OCONUS)
- Reduce retail inventories to a 90-day operating level (Bland & Bigaj, 2003)
- Achieve and maintain a surge capability at a rate of up to twice the monthly demand rate of each tire type (Bland & Bigaj 2003; DoD, 2005).

The Michelin-Lockheed Martin team developed internal metrics to measure performance to achieve the 95% on-time delivery requirement. These included dock-to-stock time in warehouse, inventory accuracy, order fill time, and carrier performance (Bland & Bigaj 2003).

The program shipped its first tires on July 9, 2001. Prior to this PBL contract, tire availability was 81%. As of 2011, backorders dropped from 3,500 to zero, and logistics response time dropped from 60 days to under two days in CONUS and under four days OCONUS. As of 2011, the average customer wait time was 32.1 hours CONUS and 59.5 hours OCONUS, and on-time performance rates were 98.5%—well exceeding the contract requirement of 95% on-time (PBL Award Summary, 2011). These results were achieved during surge periods—supporting Operation Enduring Freedom and Operation Iraqi Freedom—with no reported impact to the fleet customer.

Follow-on Contract

The follow-on firm-fixed price contract was competitively awarded to Lockheed Martin by the NAVSUP WSS in February 2016. This contract had a base period of performance of three years, with two six-month options, at a total value of $131.3 million. The Navy estimated a total cost avoidance of $24.3 million under this contract. As the prime contractor, Lockheed Martin has Michelin as a subcontractor, along with other tire manufacturers, such as Goodyear, to meet specific Navy requirements. The contract requirements were consistent with the initial contract, and through 2016 Lockheed Martin exceeded the on-time delivery metric of 95% with an on-time delivery of 98.2% CONUS and 98.7% OCONUS.

This high level of material availability provided by these PBL contracts enabled the Navy to completely draw down its former stockpile of wholesale tires from 60,000 tires to zero. By eliminating the Navy's wholesale tire inventory, 280,000 cubic feet of storage space in the distribution depots were made available. This high level of availability and consistently reduced delivery timeframes significantly reduced the need for local retail customer inventory levels, which were reduced by 66%, with a value of $1.7 million. The Navy also reduced total ownership costs by handing off the responsibility for retrograde pick-ups and disposal of scrapped tires. Additionally, the quick retrograde pick-up time of 3.4 days on average eliminated the need for the labor and storage costs associated with retrograde tire management. By reducing wholesale/retail inventory and eliminating retrograde pick-up, the program demonstrated the Navy's improved inventory management.

Lockheed Martin's best-in-class logistics support system (the LSCC) also allowed the contractor to notify the NAVAIR program manager with shipment dates and serial numbers in order to locate and quarantine any tires already out of the warehouses. This program demonstrated the benefit that the Navy received from a long-term contract based on performance from the private investment in product and process improvements, that results in cost-savings and improved support to the warfighter.

AH-64 Apache

The AH-64 Apache was conceptualized as a high-powered, tank-killing, attack helicopter, capable of repelling conventional ground forces during a Soviet invasion of Europe. Still an essential part of the Army's fleet today, the primary mission of the Apache is to perform armed reconnaissance and conduct rear, close, and shaping missions, including deep precision strikes.

Since its inception, the Apache has accumulated more than 3.9 million flight hours, with operational deployments during Desert Storm, Operation Iraqi Freedom, Operation Enduring Freedom, and Operation Inherent Resolve in Iraq. Although the first AH-64 was delivered to the Army five years before the fall of the Berlin Wall, the Apache remains the Army's primary and most advanced attack helicopter. Central to the Apache's mission is the Target Acquisition and Designation Sight/Pilot Night Vision Sensor (TADS/PNVS) system, nicknamed the "eye of the Apache."

A Brief History

The first generation of the TADS/PNVS system was fielded by the Army in 1983. The system, which comprises two sub-systems, enables Apache pilots to fly at low altitudes in total darkness and poor weather. The TADS/PNVS system also provides a capability that allows the co-pilot to identify and engage hostile targets (Yenne, 2005).

In 2003, Lockheed Martin was awarded a production contract for an upgraded, modernized version of TADS/PNVS. The M-TADS/PNVS, also known as the "Arrowhead," is an "advanced electro-optical fire control system that AH-64D/E Apache helicopter pilots use for targeting and pilotage in day, night and/or adverse-weather missions" (Lockheed Martin, 2015). The updated version is projected to lower sustainment costs by 50% over the system's expected 40-year life span (Lockheed Martin, 2015).

Prior to the initial TADS/PNVS PBL contract, the sustainment cost for the Apache's sensors systems averaged $218 million per year. Product support functions were performed organically, with Lockheed Martin providing "repair and return" services on a transactional basis (DoD, 2013).

Both the original TADS/PNVS and M-TADS are designed around the concept of the Line Replaceable Module (LRM). Technicians remove and replace faulty components directly, restoring the system to service quickly. The faulty component is sent for repair off-site. The LRM concept has been shown to reduce the cost, volume, and weight of spares holdings (Curtiss-Wright, 2016). The LRM design allowed technicians to remove and replace faulty equipment on the flight line. Intermediate-level maintenance of faulty components was performed at the division or corps level, while depot-level maintenance was performed either at the then Martin Marietta depot facility in Orlando, FL, or at subcontractor facilities (Robbins & McIver, 1994).

A 1994 RAND report analyzing logistics support for the Army's high-tech weapons found that the Army overstocked certain TADS/PNVS LRMs and understocked others. The report concluded that the inefficiencies in intermediate-level maintenance would limit repair capability to only 25% of all received platforms during a large-scale operation. The report attributed this limitation to the absence of prioritization mechanisms at critical repair facilities. In an effort to improve logistics efficiency, the DoD transitioned to a PBL in 2007.

Program Description

Since 2007, Lockheed Martin has provided sustainment for the AH-64 Apache Helicopter's M-TADS/PNVS system through a series of three PBL contracts. The PBL program consists of three major functions: repair operations, logistics operations, and continuous improvement areas. Together, these functions established a system of continuous improvements supporting the Apache sensors and covered complete post-production supply chain management, including inventory management, maintenance, modifications, procurement, repairs, and spares planning of fielded systems. In 2013, the PBL supported more than 670 aircraft in 27 battalions worldwide, including multiple forward operating bases (DoD, 2013).

Repairs are performed at five special repair activities (SRAs). The largest of these is the Letterkenny Army Depot Partnership in Pennsylvania, which repairs 29 of the 53 LRMs on the M-TADS system. The partnership employs 14 personnel (six government and eight Lockheed Martin). Additional SRA locations are located in Arizona, Texas, Alabama, and Florida (Lockheed Martin, 2016).

The second function, logistics operations, comprises U.S.-based depot support facilities and contractor supply support activities (CSSAs) located at domestic and overseas U.S. military installations and within close proximity to deployed Army units. The depot support facilities oversee the following functions: management of government-owned, contractor-managed assets; distribution of repair parts to SRAs; packing, handling, shipping, and transportation; and operation of storage facilities. The CSSAs consist primarily of forward-deployed Lockheed Martin-staffed support teams. In 2013, CSSAs had a presence in Afghanistan, Iraq, Germany, South Korea, and Kuwait (Lockheed Martin, 2016). The CSSAs serve as an information conduit between Army units and Lockheed Martin's global support network. The CSSAs process repair orders, ensuring timely transportation of new parts from SRAs to deployed units.

Finally, the continuous improvement function of the PBL consists of a dedicated team of Lockheed professionals that do demand planning, obsolescence management, and work to improve reliability and maintainability. The team relies on specialized IT tools, including an asset management system that "provides data necessary to identify and implement corrective actions and proactively push improvements into the field" (DoD, 2013). Among its numerous functions, the team investigates new failure trends; reviews reliability predictions to determine current and future needs; and develops low impact, and easy-to-implement solutions to recurring or emerging logistics or technical challenges.

PBL Strategy

The PBL has relied on firm fixed-price contracts that are tied to the number of flight hours. The program has established nine flight bands, each of which is designated by a maximum number of annual flight hours. The nine bands are separated by approximately 20,000 hours; band 1 has a maximum of 87,000 hours, band 10 a maximum of 240,000 (Lockheed Martin, 2016). Thus, the Army would pay the maximum annualized value of the contract during years in which Apache flies between 220,000 and 240,000 miles. This structure is ideally suited to heavily-deployed systems, such as the Apache. It provides the contractor with the traditional incentives associated with fixed-price contracts, translating to higher levels of innovation, reliability, and availability, while at the same time offering sufficient flexibility (e.g., the Army pays for actual usage) to support changes in operational tempo and accommodates multiple deployments (for instance, by establishing new deployed CSSA locations as needed).

The first four-year contract (one base year and three one-year options) was valued at approximately $380 million. In 2012, a similar follow-on contract valued at $375 million was awarded (Lockheed Martin, 2012). A third, five-year PBL contract (one base year and four one-year options) was awarded in 2016. That contract was valued at $424 million and represents a price reduction of 10% over the previous contract (Lockheed Martin, 2016).

Program performance is measured in terms of supply availability (SA) under the Apache PBL agreement. Lockheed Martin is contractually obligated to meet a minimum availability requirement of 85%. In other words, the requested part must be received by the requesting Army unit within the required timeframe 85% of the time. This timeframe varies depending on the type of part and the location of the requesting unit. There are three issue priority groups (IPG-1 is the highest priority; IPG-3 is the lowest) and two location categories—in-country and deployed. The program relies on this matrix to meet supply availability requirements. IPG-1/deployed have the shortest timeframe requirement, IPG-3/in-country have the longest (Lockheed Martin, 2016). As with the contract structure itself, the supply availability requirement injects flexibility into the program and aligns contractor priorities with those of the Army.

Prior to awarding the 2016 contract, the Army sought to reduce costs by extending the in-country IPG-1 timeframe requirement from two to four days. Although this change resulted in cost reduction, the savings were not large. The parts inventory stayed at the same level because the lead-time to procure parts still exceeded the required timeframe, so the change only affected transportation costs.

PBL Results

Under the initial contract, Lockheed successfully slashed sustainment costs for both sensor systems and improved supply availability primarily through improvements in supply chain and obsolescence management. Lockheed has since lowered logistics and maintenance costs by leveraging data tracking for health and maintenance indicators to improve demand forecasting, determining appropriate inventory levels, and by ensuring the optimal locations of supply activities.

Between 2007 and 2013, SA for MTADS/PNVS averaged 97%, well above the 85% requirement. Figure 8 illustrates annual availability by IPG between 2007 and 2011, followed by monthly availability between January 2012 and May 2013. Notably, a high level of availability was maintained between 2011 and 2013 when Apache reached its peak OPTEMPO of over 200,000 flying hours per year. In 2012, 96,000 hours were accumulated in Afghanistan alone. The other 115,000 hours were accumulated at locations in Kuwait, Germany, Korea, and CONUS locations (Lockheed Martin, 2016). The program has prioritized the availability of deployed units, which between 2012 and 2013, averaged 99%. As of August 2018, the PBL continues to exceed the required performance, and has a proven supply availability rate of over 99%, the result of efficiencies gained in supply chain management, valued engineering services, depot level maintenance, and retrograde infrastructure.

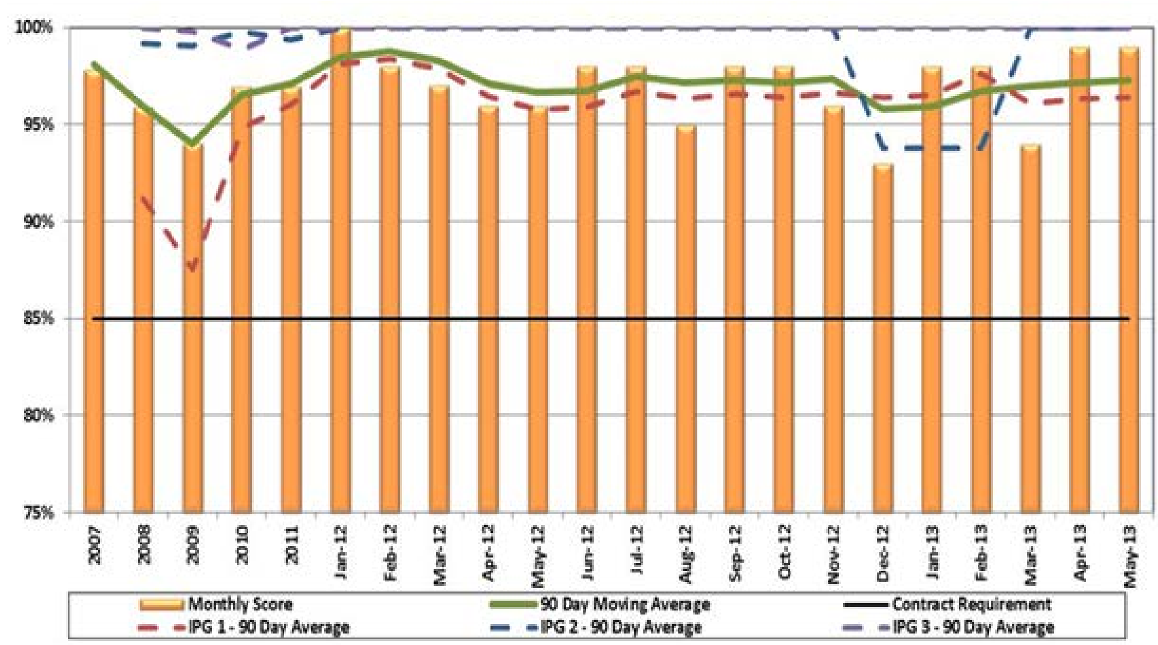


Figure 8. **M-TADS/PNVS Parts Availability**
(Breter, 2013)

Lockheed professionals working within the continuous improvement function have developed numerous solutions that have increased mean time between system failures (MTBF) by 70% compared to the pre-PBL period. Often "simple fixes" such as redesigned screws that strip less easily; a protective guard that prevents damage to exposed machinery; and improved airflow gaskets have all served to drastically improve reliability, durability, and overall performance. In addition, Lockheed has been successful in drastically increasing the annual retrograde rate—i.e., the rate at which repairable parts are transported to depots for repair, in preparation for those parts to be placed back into the supply chain—reducing the number of spares and the overall logistics footprint required to store and maintain them (see Figure 9).

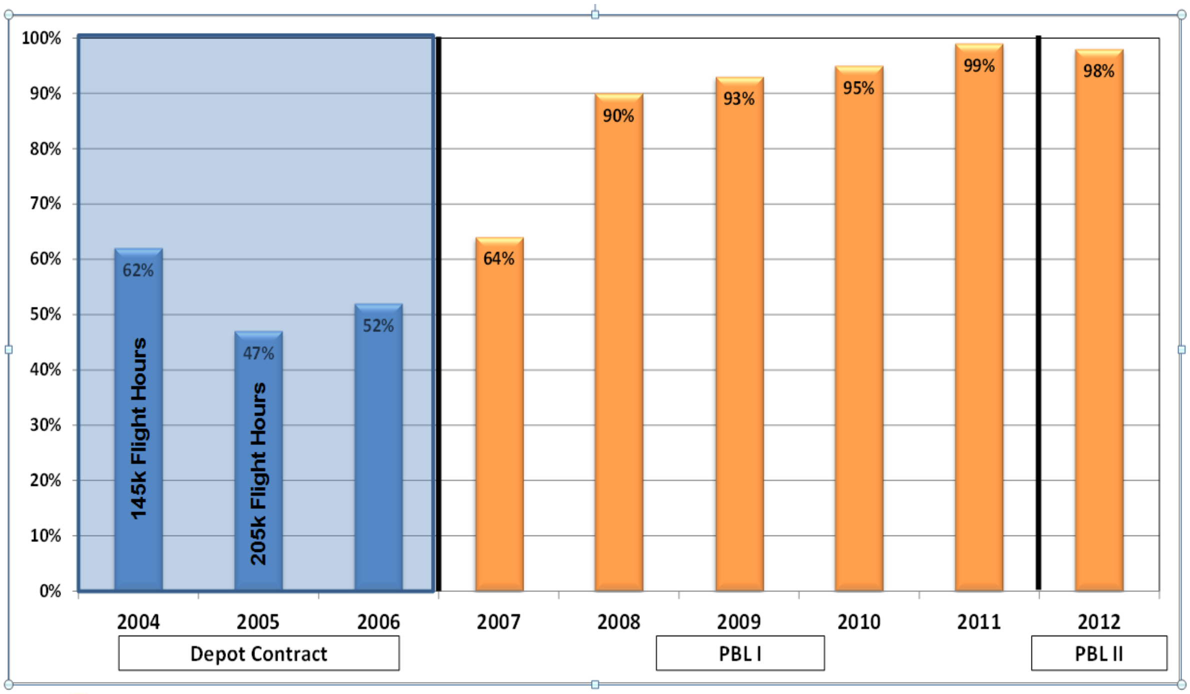


Figure 9. **M-TADS/PNVS Retrogrades by Year** **(Breter, 2013)**

The program also exceeded 99% availability for depot repair parts (see Figure 10). The PBL contract has been credited with improving fleet readiness, reducing average flying hour cost and reducing the Army's long-term inventory investment. Over the course of the initial PBL contract, depot-level repairable costs were reduced by 18%, supply inventory replenishment costs were reduced by 40%, and mean-time between maintenance actions reduced by 9.6% (OSD, 2012).

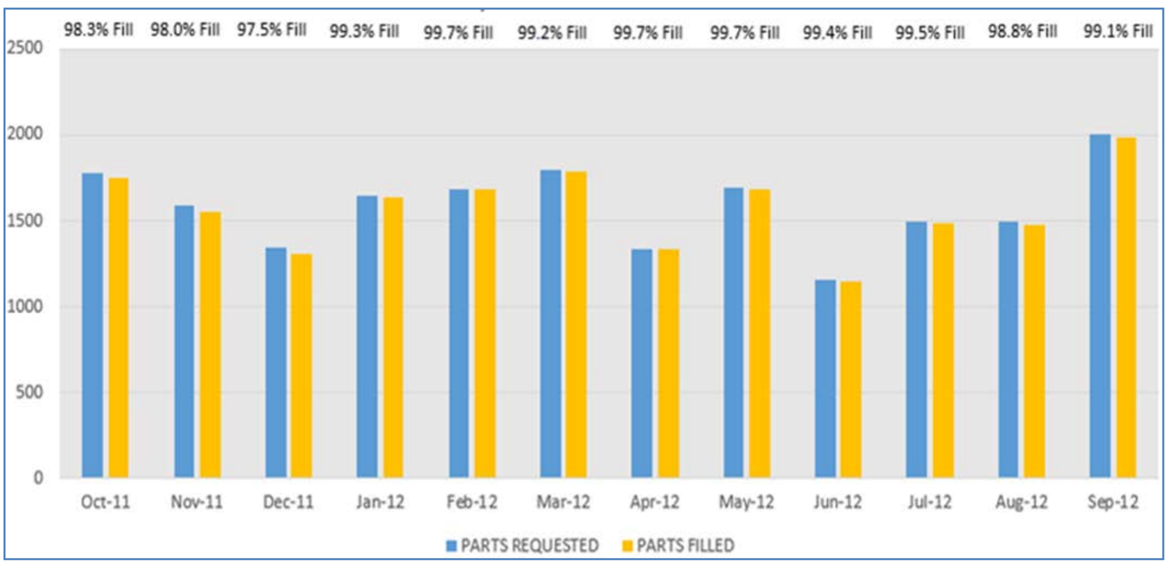


Figure 10. **M-TADS/PNVS Depot Repair Parts Availability** **(Breter, 2013)**

As mentioned previously, annual sustainment costs prior to the implementation of PBL totaled $218 million per year. In 2013, costs totaled $92 million, a drop of 58% (see Figure 11). Other accomplishments include the mitigation of 759 obsolescence and diminishing manufacturing cases since 2007, resulting in $104.2 million in cost avoidance, reduction of the maintenance support footprint, and a decrease of more than 1,000 maintenance man-hours per year through increased materiel reliability (OSD, 2012). These efficiencies enabled the government to negotiate a price reduction of approximately 10%, reflected in the most recent contract awarded in 2016. In light of the program's continued success, sustained high availability, and gains in affordability, the contractor team is optimistic about the program's future.

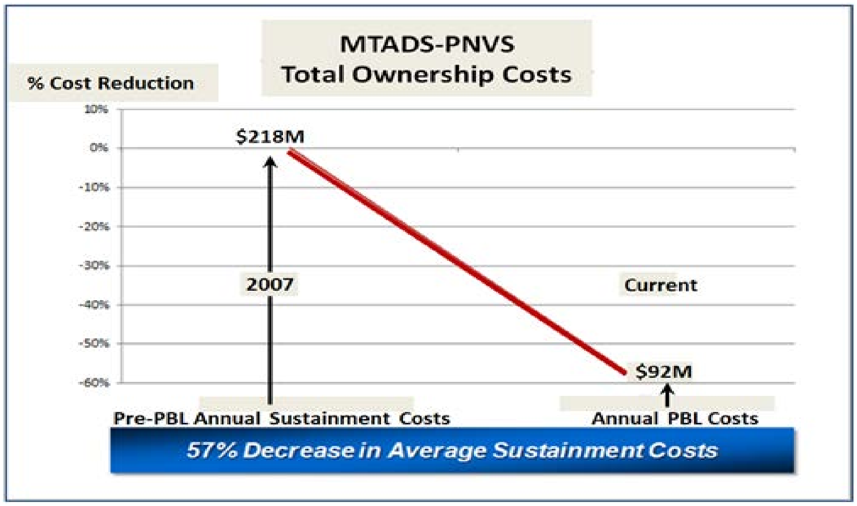


Figure 11. **MTADS/PNVS Total Ownership Costs (Breter, 2013)**

Recommendations and Conclusion

Long-running PBLs have the potential to continue to deliver value, high reliability, and improved performance. Based on our examination of the PBL construct and our evaluation of three successful PBL programs, we offer the following recommendations.

Recommendations

1. Promote the use of PBL as a proven support strategy for weapons systems throughout the life cycle.

PBLs generally perform better than traditional support mechanisms. However, support within the DoD for PBL has appeared to wane in recent years. The benefits of PBL contracts continue to accrue as systems age; even with older systems, technological refresh and modernization initiatives create new opportunities to improve products and processes and reduce costs.

PBL contracts may also be perceived as more expensive than support provided through a more traditional, transactional approach. Indeed, the price that an operational unit pays for a part may appear to increase as its reliability improves, but this is due to the operational unit's portion of the contract payment being allocated over the total number of parts provided within a given period. When aggregated at the fleet level, costs decrease as reliability improves.

The DoD should renew its commitment to the expansion of PBL in order to improve weapon system operations and reduce costs. This will require increased support from senior DoD officials and Service leaders to ensure that PBL is employed when developing product support strategy and arrangements.

2. Ensure the acquisition workforce is educated and trained to execute successful PBL contracts.

Developing and implementing successful PBL arrangement requires a different skillset than that required for contracting for transactional product support. Critics suggest, perhaps rightly, that PBL arrangements can be more challenging to develop and manage than the more traditional transactional contracts. Specifically, the acquisition workforce often does not have a thorough understanding of how to select a contract type or structure contracts with the appropriate incentives and penalties to motivate industry to provide superior support while reducing costs. Accordingly, the acquisition workforce must be trained in the appropriate use of PBL contracts, and how to structure them with suitable metrics and incentives to achieve program objectives.

3. Structure PBL contracts appropriately.

PBL contract type should be structured to reflect the current phase of the system's life cycle. When a system is mature and characterized by relatively low levels of uncertainty, both operational and technical, alignment of contractor and government objectives are optimized with fixed-price PBL contracts. These arrangements promote the greatest performance improvements and cost-reduction, higher levels of innovation, shift program risk to the contractor, and result in enhanced reliability. These contracts generally rely on a small number of performance metrics that directly support the stated outcomes, which help ensure transparency and accountability.

a. **Ensure proper alignment of government objectives with provider incentives.**

An appropriate PBL program uses the contract structure and incentives to align the objectives of the customer (the government), with those of the support provider, leading to a win-win scenario. The incentives should generally include a combination of rewards and penalties. Rewards can include financial payments and contract extensions for achieving cost and/or performance objectives. Penalties can come into play if the support provider fails to achieve the program outcomes and can include reduced fees and/or contract options that are not exercised. An inappropriate structure can create perverse incentives and result in undesired or unintended consequences.

Again, the acquisition workforce must have a good understanding of what motivates businesses to ensure that the contractual incentives will achieve the desired outcomes.

b. **Consider scalability and usage requirements in developing the product support strategy.**

There are various strategies to build some flexibility into PBL contracts to account for changes in how systems are used. If these strategies are not used, the results can be suboptimal. For example, under the previous HIMARS PBL contracts, the fixed price was tied to OPTEMPO category, with each vehicle assigned to a price category based on the customer's *anticipated* usage. In the event that vehicles are "underused," the government customer may feel as though he is overpaying. On the other hand, M-TADS, tied the fixed price to actual usage (i.e., flight hour). When possible, PBL contracts should tie price to actual system usage.

c. **Use contract length to incentivize suppliers to improve reliability and reduce costs**.

The Navy tires and M-TADS PBLs show that contracts of longer duration can incentivize suppliers to invest in reliability improvements, thereby reducing future costs. Generally, PBL contracts of shorter duration will not incentivize significant contractor investment since the contract must be long enough for the contractors to recoup their investments (otherwise they will not invest). Accordingly, future performance improvements and price reductions may not be realized.

Conclusion

As defense budgets continue to shrink, and operations and maintenance costs for weapon systems continue to rise, the DoD must heighten its focus on affordability and efficiency when it comes to new and existing weapon programs. With PBLs' vast array of benefits, when properly structured, these contracts have the potential to dramatically reduce the costs of procuring and sustaining weapon systems, while incentivizing higher levels of performance throughout the system's life cycle. As we continue to face new and evolving global threats, the demand for superior and highly reliable technology is now more crucial than ever. Although its benefits have been consistently proven throughout the years, PBL is still not being aggressively pursed throughout the DoD.

From a theoretical standpoint, the power of PBL lies in affording the provider the discretion and flexibility to select the optimal mix of inventory levels, maintenance activities, and technology upgrades in order to meet performance requirements. The case studies presented here suggest that mature PBL programs are capable of exceeding performance and cost requirements. Shifting one or more of these functions to the government customer distorts the PBL paradigm and may, over time, lead to reductions in performance, innovation, and cost savings—if not in the short term, then in later iterations of the contract.

Reference List

Army-Technology. (2015). HIMARS High-Mobility Artillery Rocket System, United States of America. Retrieved from http://www.army-technology.com/projects/himars

Bland, P., & Bigaj, L. (2003). *Implementing and managing a PBL supply chain*. Lockheed Martin.

Boyce, J., & Banghart, A. (2012). Performance based logistics and project proof point: A study of PBL effectiveness. *Defense AT&L: Product Support Issue.* Retrieved from http://dau.dodlive.mil/files/2012/03/Boyce_Banghart.pdf

Curtiss-Wright. (2016). Line-replaceable modules. Retrieved from https://www.curtisswrightds.com/technologies/line-replaceable-modules-lrm.html

Defense Acquisition University. (2005). *Performance based logistics: A program manager's product support guide*. Retrieved from http://www. dau.mil/pubs/misc/PBL_Guide.asp

Defense Acquisition University. (2018). *PBL guidebook: A guide to developing performance-based arrangements*. Retrieved from https://www.dau.mil/guidebooks/Shared%20Documents%20HTML/PBL%20Guidebook.aspx

DoD. (2005). Navy contracts for July 11, 2005. Retrieved from http://www.defenselink.mil/Contracts/Contract.aspx?ContractID=3047

DoD. (2006). *The Department of Defense awards program for excellence in performance-based logistics.* Retrieved from https://acc.dau.mil/adl/en-US/548810/file/68118/PBL%20Award%20Pkg%202006%20System_HIMARS.pdf

DoD. (2013). *The Secretary of Defense PBL awards program for excellence in PBL.* Retrieved https://acc.dau.mil/adl/en-US/679748/file/74757/Sub-system%20USA%20Apache%20Sensors.pdf

DoDIG. (2012, November 30). *Accountability was missing for government property procured on the Army's services contract for logistics support of Stryker vehicles.* Retrieved from http://www.dodig.mil/pubs/documents/DODIG-2013-025.pdf

Erwin, S. I. (2013, March). Industry and government butt heads over weapons maintenance contracts. *National Defense Magazine*. Retrieved from http://www.nationaldefensemagazine.org/archive/2013/march/Pages/IndustryandGovernmentButtHeadsOverWeaponsMaintenanceContracts.aspx

GAO. (2006). *Defense logistics: Changes to Stryker vehicle maintenance support should identify strategies for addressing implementation challenges.* Washington, DC: Author.

Gansler, J., & Lucyshyn, W. (2006). *Evaluation of performance based logistics.* College Park, MD: Center for Public Policy & Private Enterprise, University of Maryland College Park.

Gansler, J., Lucyshyn, W., & Vorhis, C. (2011). *Performance-based services acquisition.* Monterey, CA: Naval Postgraduate School.

Gansler, J., & Lucyshyn, W. (2014). *HIMARS: A high performance PBL case study.* Monterey, CA: Naval Postgraduate School.

Gardner, C. P. (2008, March). *Balancing government risks with contractor incentives in performance-based logisitics contracts.* Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a480398.pdf

Gourley, M. (2014). *The current state of performance based logistics (PBL).* Presentation at the Logistics Officer Association Symposium. Retrieved from www.logisticsymposium.org/.../975718392b4c7d40aa0a3f43995ce19288877359.ppt

Guajardo, J. A., Cohen, M., Kim, S., & Netessine, S. (2012). Impact of performance based contracting on product reliability: An empirical analysis. *Management Science, 58*(5).

Hawkins, K. (2009, December 16). HIMARS shoots high for award. Retrieved from http://www.army.mil/article/31909/HIMARS_Shoots_High_For_Award/

Hunter, A., Ellman, J., & Howe, A. (2017, March 31). Use of incentives in performance based logistics contracting*. CSIS.* Retrieved from https://www.researchsymposium.com/conf/app/researchsymposium/unsecured/file/146/SYM-AM-17-052-005_Hunter.pdf

Kim, S., Cohen, M., & Netessine, S. (2007). Performance contracting in after-sales service supply chains. *Management Science, 53*(12), 1843–1858.

Kim, S., Cohen, M., & Netessine, S. (2011). Reliability or inventory? Analysis of product support contracts in the defense industry. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.179.2945&rep=rep1&type=pdf

Lockheed Martin. (2011, September 21). Lockheed Martin delivers 400th HIMARS launcher to U.S. Army. Retrieved from http://lockheedmartin.com/us/news/press-releases/2011/september/LMDelivers400thHIMARSLaun.html

Lockheed Martin. (2012, September 24). U.S. Army awards Lockheed Martin $111 million Apache M-TADS/PNVS performance based logistics contract. Retrieved from http://www.lockheedmartin.com/us/news/press-releases/2012/september/mfc-092412-usarmyawards.html

Lockheed Martin. (2015). High Mobility Artillery Rocket System (HIMARS). Retrieved from http://lockheedmartin.com/us/products/himars.html

Lockheed Martin. (2015, June 30). M-TADS/PNVS (Arrowhead). Retrieved from http://www.lockheedmartin.com/us/products/Arrowhead.html

Lockheed Martin. (2016). *Performance based logistics for modernized target acquisition designation sight* [PowerPoint presentation slides].

Lucyshyn, W., Rigilano, J., & Safai, D. (2016). *Performance-based-logistics: Examining the successes and challenges when operating in stressful environments.* Monterey, CA: Naval Postgraduate School.

Macfarlan, W., & Mansir, B. (2004, July–September). Supporting the warfighter through performance-based contracting. *Defense Standardization Program Journal,* 38–43.

Mahadavia, D., Engel, R., & Fowler, R. (2006). Performance based logistics: Putting rubber on the ramp. *Defense AT&L Magazine*. Retrieved from http://www.dau.mil/pubscats/pubscats/atl/2006_07_08/mah_ja06.pdf

National Academies of Sciences, Engineering, and Medicine. (2016). *Owing the technical baseline for acquisition programs in the U.S. Air Force*. Washington, DC: The National Academies Press. doi:10.17226/2363

NAVICP. (2000). Statement of work (SOW) performance based logistics (PBL) of aircraft tires (N00383-00R-0040). *Commerce Business Daily*.

NAVICP. (2001). *Innovative contract saves Navy millions.* Philadelphia, PA.: Author.

Office of the Secretary of Defense (OSD). (2009). *Secretary of Defense performance based logistics awards program for excellence in performance based logistics: Summary of critical accomplishments.* Retrieved from DAU.

Office of the Secretary of Defense (OSD). (2012). *The Secretary of Defense performance based logistics awards program for excellence in performance based logistics in life cycle product support.* Retrieved from https://acc.dau.mil/adl/en-US/679748/file/74757/Sub-system%20USA%20Apache%20Sensors.pdf

Robbins, M., & McIver, D. (1994). *Precision guided logistics: Flexible support for the force-projection Army's high-technology weapons.* Retrieved http://www.dtic.mil/dtic/tr/fulltext/u2/a294987.pdf

Russo, A. M., & Hilbert, A. J. (2008, Spring). *HIMARS—Precision today and tomorrow.* Retrieved from http://sill-www.army.mil/firesbulletin/2008/Mar_Apr_2008/Mar_Apr_2008_pages_35_37.pdf

Singer, P. W. (2008). Outsourcing the fight. Retrieved from http://www.brookings.edu/opinions/2008/0605_military_contractors_singer.aspx

Yenne, B. (2005). *Secret gadgets and strange gizmos: High-tech (and low-tech) innovations of the U.S. military.* Minneapolis, MN: Zenith Press.

Panel 24. Go Fast—Rapid Acquisition

Thursday, May 9, 2019	
3:45 p.m. – 5:00 p.m.	**Chair: William Bray,** Deputy Assistant Secretary of the Navy for Research, Development, Test, and Evaluation ***Closing the Innovation Gap at SOCOM*** Greg Ingram, University of South Florida ***Agile Development Flight Center Model for the LPTA COTS Space Launch System*** Kelly Weigel, Trans Universal Energy, LLC ***Accelerate: How Programs Can Deliver Better Capabilities Faster*** Pete Modigliani, Dan Ward, and Su Chang, The MITRE Corporation

William Bray—Mr. Bray serves as the Deputy Assistant Secretary of the Navy for Research, Development, Test and Evaluation (DASN (RDT&E)) under the Assistant Secretary of the Navy for Research, Development and Acquisition (ASN (RD&A)). Mr. Bray is responsible for executive oversight of all matters related to RDT&E Budget Activities, Science and Engineering, Advanced Research and Development, Prototyping and Experimentation, and Test and Evaluation. He is also responsible for oversight and stewardship of the Department of Navy Research and Development Establishment including naval laboratories, warfare centers and systems centers.

Mr. Bray previously served as the Executive Director, Program Executive Office for Integrated Warfare Systems (PEO IWS), where he directed the acquisition and Fleet support of the Surface Navy's 155 combat systems, weapons, radars, and related international and foreign military sales programs. The mission of PEO IWS is to develop, deliver and sustain operationally dominant naval combat systems for Sailors and Marines.

Mr. Bray served as the Director, Integrated Nuclear Weapons Safety and Security (SP30) in the Navy Strategic Systems Programs Office. In this capacity, Mr. Bray served as the senior advisor to the Director, Strategic Systems Programs on nuclear weapons safety and security, and earned the Navy's Superior Civilian Service Award.

In August 2006, Mr. Bray was appointed to the Senior Executive Service (SES) and served as the Director, Integrated Combat Systems for PEO IWS, where he was the Major Program Manager (MPM) for Surface Navy Combat Systems, to include Aegis and Ship Self Defense System (SSDS) ships, and also led systems engineering activities and Surface Navy Open Architecture strategy across the PEO.

Mr. Bray started his career at the Naval Surface Warfare Center, Corona Division, California, beginning his civil service career in December 1984.

Mr. Bray graduated from The Pennsylvania State University in 1984 with a Bachelor of Science degree in Engineering, and also earned a Master of Science in Systems Management from the University of Southern California. He is DAWIA Level III certified in Program Management, Engineering (SPRDE), and Test and Evaluation.

Closing the Innovation Gap at SOCOM

Greg Ingram—is a Doctoral Candidate at the Muma Business School at the University of South Florida in Tampa, FL. In his civilian capacity, he is Director of Business Development at Forge.AI in Cambridge, MA. He is a retired Lieutenant Colonel with over 30 years of experience, primarily in Special Operations. [Ingramg@mail.usf.edu]

Abstract

Doing business with the U.S. Special Operations Command (SOCOM) can be difficult. There are a number of physical and regulatory hoops to jump through: getting on MacDill Air Force Base, getting into the SOCOM compound, and the prohibition of electronic devices on the compound, just to name a few. Not to mention the intricacies of the federal acquisition process, which could take more than a year from Broad Area Announcement (BAA) to contract award. To address these hurdles in a rapidly changing technological environment, SOCOM, in conjunction with the Doolittle Institute (now DEFENSEWERX), created SOFWERX: An unclassified facility where Special Operations operators can bring their requirements and be tested out and rapidly prototyped (https://www.sofwerx.org/). Once down selected, they can then move into a SOCOM Program of Record and be fast-tracked into the acquisition process. This compression of the Concept Development to Prototype phases vastly increases the time to product deployment.

Case Study: SOFWERX

Introduction

Doing business with the government is hard: It is bureaucratic and byzantine at best. A typical acquisition project can take a minimum of three years to get from concept development to prototype and requires detailed product specifications that may have changed by the time the product is ready for market. Given the nature of rapidly evolving technology, that process inhibits innovation and puts our service members at a disadvantage on the battlefield as their adversaries are not encumbered with such onerous bureaucratic and regulatory requirements. A perfect example is the Islamic State (or IS) using Commercial Off the Shelf (COTS) drones and re-purposing them for combat operations. Although Congress and the Department of Defense (DoD) have both recognized these shortcomings and are attempting to reform the system to make it more agile, the efforts to date have been spotty at best.

During the last 16 years of uninterrupted warfare, several agencies sprang up to address urgent battlefield requirements that were not being met. Two examples were the Joint Improvised Explosive Device Defeat Organization (JIEDO; now the Joint Improvised Defeat Agency or JIDA; https://www.jieddo.mil/) and the Rapid Equipping Force (REF; http://www.ref.army.mil/). These organizations and others set the pace for rapid prototyping and fielding of equipment and services to the warfighter, including significant innovations in protective equipment like body armor and mine-resistant vehicles. However, many of our present and future threats lie outside of the realm of conventional risks. With the advent of cloud computing, clustered Graphical Processing Units (GPUs), the proliferation of data, and the risks of cyberattacks by criminals, state, and non-state actors, the need to decrease the time from concept to market for information-related capabilities has become an imperative.

At the Joint Staff in the Pentagon, in the Under Secretary of Acquisition, Technology & Logistics (AT&L), former Secretary of Defense, Ash Carter formed three entities to try and accelerate the technology acquisition process. One entity is called the Strategic Capabilities

Office (SCO), founded in 2012. Another is the Defense Innovation Unit Experimental (DIUx and now DIU), founded in August 2015 (www.diux.mil). The third is MD5 (https://md5.net), whose mission is to build a collaborative community through "Education, Collaboration and Acceleration programs" (https://community.md5.net/md5/about). The SCO takes existing military capabilities and "makes them do something different." The DIU mission is to connect commercial and traditional defense contractors to broad defense requirements using a funding mechanism called Other Transaction Authorities (OTA), which allows for more rapid contracting, usually in 90 days or less.

Recognizing they lacked the agility to get the warfighters technology more rapidly, SOCOM created SOFWERX in October 2015. SOCOM and the Doolittle Institute created SOFWERX through a mechanism called a Partnership Intermediary Agreement (PIA), which "increases the likelihood of success in the conduct of cooperative or joint activities … with small business firms [and] institutions of higher education" (https://www.gpo.gov/fdsys/granule/USCODE-2011-title15/USCODE-2011-title15-chap63-sec3715). Located in Ybor City in Tampa, FL, this completely unclassified venue offers unfettered access for innovative companies to bring their ideas to the Special Operations Community. Cameron Hunt, former SOFWERX Chief Innovation Officer, describes the organization as being "left of requirements"[1] and "more McGyver than Q" (Hunt, 2017).

The Case

The U.S. Special Operations Command (USSOCOM) was formally created in 1987, but it ultimately resulted from lessons learned as a result of the failed attempt to free the hostages in Iran in 1980. Embarrassing failures often create opportunities for future success. USSOCOM is a Functional Combatant Command responsible for training and equipping Special Operations Forces (SOF) as compared to a Geographic Combatant Command, like U.S. Central Command (CENTCOM), who is responsible for the Middle East, South West Asia, and the Central Asian States. In 1991, SOCOM was granted service-like acquisition authority for SOF specific equipment. This gives SOCOM the ability to conduct research and development, acquire equipment, and maintain that equipment—the same capability that the Air Force, Army, and Navy have. The only caveat is that those activities must be peculiar to SOF. In addition to those authorities, Title 10, Section 167 grants SOCOM Head of Agency (HOA) status, the only COCOM with this status. This combined with the acquisition authority allows SOCOM to perform the functions of a service with a drastically reduced bureaucracy.

Since its inception in 1991, the SOCOM Acquisition, Technology, and Logistics (SOF AT&L; https://www.socom.mil/SOF-ATL/) has endeavored to be innovative and responsive to the warfighter's needs. Even though 16 years of warfare has pressured the acquisition system to field very innovative solutions, the increasingly short cycle of technology innovation has pressured the formal acquisition process to keep up with technology and to harness the efforts coming out of business incubator sites such as Silicon Valley, Boston, government research labs, and academic research institutions. As a consequence, SOCOM, under the aegis of Mr. James "Hondo" Geurts, the SOCOM

[1] A variation of the military saying, "being left of bang," i.e., being proactive and preventative.

Acquisition Executive (now Assistant Secretary of Defense for Navy Acquisition) has explored a number of ways to speed up the acquisition process. This effort is now being improved under the direction of James Smith, the current acquisition executive.

Up until this time, the barriers to entry in doing business with SOCOM were both physical and procedural. Simply gaining physical access to SOCOM requires a sponsor to allow you access to MacDill AFB, the home of SOCOM and CENTCOM. The SOCOM sponsor would need to generate a visitor request to allow access to the SOCOM campus. To compound matters, due to security requirements, no commercial computers, cell phones, etc. are allowed inside the facility. In addition, to get an audience with a program manager at SOCOM, a potential vendor must submit an idea through the SOF AT&L website, which serves as a "gatekeeper" to the AT&L offices. Traditional defense contractors understand the system and have an easier time negotiating the process.

This is the genesis of SOFWERX. SOFWERX is interested in attracting novel solutions to warfighter problem sets using both traditional solution providers, but more importantly innovative firms, who would not normally do business with the government for a wide variety of reasons. "Hondo" Geurts asked the penultimate question: "How do we keep pace with the exponential growth in our operations as well as technology, and where do you find a place where you can marry that all together?" (Gibbons-Neff, 2016). In October 2015, in response to disappointing results for the concept development of the TALOS project (Tactical Assault Light Operator Suit, also known colloquially as the "Iron Man Suit"), the SOFWERX idea came to fruition. SOCOM negotiated and signed a partnership intermediary agreement with the Doolittle Institute, and SOFWERX was born (Hunt, 2017).

A partnership intermediary agreement (PIA) gains its authority from 15 U.S.C. 3715, and is a "contract or memorandum of understanding with an intermediary that provides for the performance of services for a federal laboratory to increase cooperative or joint activities with small businesses, institutions of higher education or educational institutions" (Use of Partnership Intermediaries, 2015). The Doolittle Institute is a non-profit organization that receives base funding from SOCOM to execute a "one-to-end list" of requirements both from SOF AT&L and nominations from warfighters (Andrews, 2017). The execution of funds through a non-profit entity allows SOCOM to experiment unencumbered by the bureaucratic limitations of the Federal Acquisition Regulations (FAR), but ensures oversight through the SOF AT&L office. This allows for rapid prototyping through "challenges," hackathons, disrupters, and capability collaboration events (CCE).

SOFWERX provides a completely unclassified environment where new ideas can be tested and validated (or not) ostensibly drastically reducing the time from concept development to operational prototype. SOFWERX does this in a number of ways. Through their website (www.sofwerx.org), they announce events that open to warfighters, interagency partners, academia, and industry to provide an open environment, or ecosystem, for the free flow of information and discussion. The 10,000-square foot facility provides a laboratory where prototypes can be manufactured and tested (within a certain scale).

From an initial staffing of a handful of employees in October 2015 (all full-time SOFWERX employees are employed by the Doolittle Institute and not SOCOM), by November 2017 there were 21 Defensewerx employees, four industry fellows (including a University of South Florida fellow from the School of Engineering), an international researcher from Norway, eight core interns, six interns dedicated to the Thunderdrone effort, and two additional interns provided for by matching funds for a total of 16 interns, all paid. In addition, USF has seven paid interns (one PhD and six graduate interns) dedicated to a small satellite contract.

On November 15, 2017, SOFWERX introduced TEAMWERX, their web portal to manage the challenge process. “TEAMWERX is a prize challenge platform designed to find innovative solutions to warfighter problems” (https://www.teamwerx.org/). One example of a challenge is called City System-of-Systems Intelligence Model, with a prize of $22,500, which has a goal of “Populate a City Systems-of-Systems Intelligence (CSSI) model template from publicly available data about major cities” (https://www.teamwerx.org/city-system-of-systems-intelligence/). The beauty of these challenges (if they succeed) is that the challenge performs the function of a “fair and open competition” as defined by the FAR, and therefore the winner can be ultimately be awarded a sole source contract for production. This can be a huge time and cost saving for the government and, therefore, the taxpayer.

These types of competitions also allay concerns from non-traditional defense contractors (e.g., commercial technology companies), who are typically nervous about sharing intellectual property (IP) with the government. In this respect, SOFWERX serves as a neutral facilitator and not like a typical government research lab who wants to control IP from government-funded research. SOFWERX can exercise Research & Development Agreements that allow for the protection of IP with no expectation of results. This allows the innovators the “freedom to fail” without repercussions. This arrangement encourages innovation, unlike in the traditional government acquisition programs, which typically rely on incremental progress.

Many of the prize challenges and CCEs are technical in nature, and there is an increasing interest in data science, big data, artificial intelligence, and machine learning (see Appendix A). But many of the successes are very practical in nature. One of the recent success stories was a prize challenge to develop a “bow bumper” that attaches to the bow (front) of a light attack boat to serve as a cushion of sorts when an assault team of SEALs forcibly boards another boat or structure in the water. An existing product (inflatable) that had problems was being used. Feedback from the warfighter to SOFWERX initiated a prize challenge. Originally, eight vendors, including the incumbent, provided designs (two inflatable), and out of those, three were selected for prototype testing. Interestingly enough, the initial third place vendor became the winner of the challenge and their design is now in production (see Appendix B).

Another win for SOFWERX came when a group of SOF medics came to them with a problem they had fitting litters (military terminology for *stretchers*) into the back of their all-terrain vehicles: the Polaris MRZR (Polaris, n.d.). SOFWERX was able to take a MRZR down to their local machine shop (called DIRTYWERX), work with the medics, and fabricate the parts needed on the spot. After several iterations, the medics approved the modification, the specifications were documented, and the parts kit was put out for bid for production. This all happened in the course of a few weeks (see Appendix C).

As important as the winners are, one of the important operating principles of SOFWERX is the ability to fail. In this sense, the SOFWERX definition of failure is not finding a “workable material solution.” One example of this was a prize challenge called Fogbreaker designed to assist high speed offshore boat drivers track multiple data points through Augmented Reality while having to pay attention to existing potential navigational hazards. SOFWERX did award three winners, but the warfighter determined that none of them provided good enough solutions to proceed with further development.

One of the challenges for SOCOM AT&L has been how to integrate and leverage SOFWERX into their own Programs of Record. In January 2017, Guerts asked Kelly

Stratton-Feix to be Director of Acquisition Agility, whose purview is to manage SOFWERX. At the time, Stratton-Feix was working on a master's program that required a project that had a million dollar return on investment, so the SOFWERX position essentially killed two birds with one stone. When she stepped into the job, she had not worked with non-traditional acquisition authorities like the PIA and Other Transaction Authorities, so part of her initial work was educating herself on the strengths and limitations of these authorities. What she also found was that the program managers (PM) at SOCOM did not understand these authorities either, including how they would fit into the acquisition cycle and how they could support their acquisition efforts. These PMs are very busy people who manage very complex programs and like most busy people, they did not see the incentives of learning new methods. Although not strictly part of her job title, she felt very strongly that she needed to educate her peers. She went back through the authorities, specifically the National Defense Authorization Act of 2015 which gave birth to SOFWERX and DIUx, and also the specific authorities concerning Partnership Intermediate Agreements (PIA) and Other Transactions. She stood up a team of people to do the research and once she felt comfortable with the statutes and how it fit into the acquisition process, she launched a comprehensive education program for the acquisition workforce at SOCOM to increase awareness and to help PMs to leverage these "new" authorities to their benefit (Stratton-Feix, 2019).

One of the other issues she recognized almost immediately was the need to tie warfighter nominations to continuing funding streams in Programs of Record (POR). Serving the requirements of the warfighter is at the core of SOFWERX's mission, but if a product is developed for the warfighter is not tied to a POR that can support ongoing Operations & Maintenance (O&M) support. She had a meeting with Tambrein Bates, the SOFWERX Director, to come to an agreement on what percentage of unfunded requirements they would support from the SOFWERX budget (20%) and how they would increase the PM participation and leveraging of the SOFWERX ecosystem. Stratton-Feix and her team conducted two training sessions for all the PMs, she and Bates conducted roundtable discussions with all the PMs to get them comfortable with how these authorities can help them be successful.

As a consequence, the pace and involvement of PMs in the SOFWERX ecosystem has increased dramatically. For example, a recent Collider event held from April 9–11, 2019, had multiple PMs and SOCOM Agreements Officers participating in nine different Technical Focus Areas. The number of successful transitions has also improved, as the chart in Figure 1 illustrates (SOFWERX, 2019).

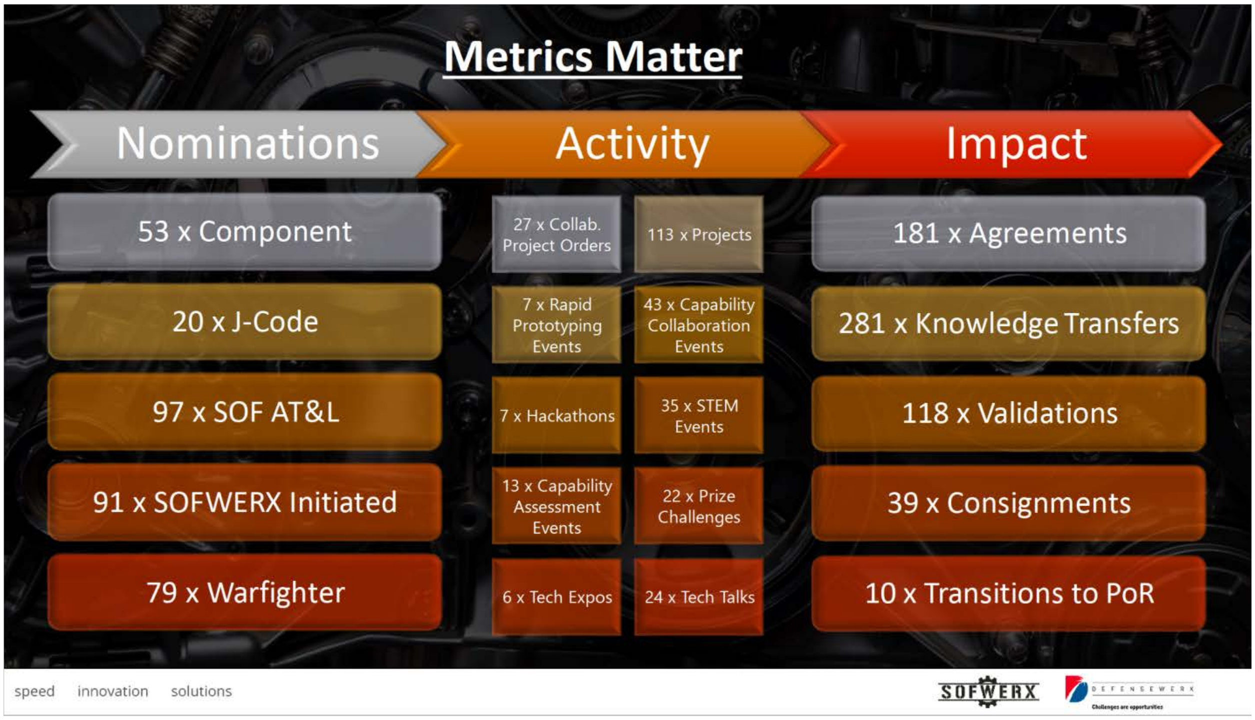


Figure 1. **Metrics Matter**

For those who are not familiar with the intricacies of the government acquisition processes and the significant hurdles and barriers to entry that exist, this case study may beg the question, “So what?” Many of these efforts to incorporate innovation into the acquisition process are considered bypass mechanisms or work-arounds that may or may not prove to be long-term solutions. In addition, as one would imagine, there exists in the community of federal acquisition professionals a considerable amount of pushback to what some consider to be an attempt to circumvent an existing system that may in fact threaten their livelihoods. While these may be legitimate concerns, the needs of the warfighter to take advantage of the rapidly changing technology development cycle as well as the potential for substantial taxpayer savings should outweigh those concerns.

Another important human factor is the rotation cycle for service members and senior executive service government employees. Sometimes an initiative started under an incumbent may not survive the incoming leadership due to the vagaries of human irrationality or the perceived lack of political capital of the program. The new SOCOM acquisition executive, James Smith, has embraced the SOFWERX concept and expanded the role of SOFWERX in the SOF AT&L process. In January 2019, SOCOM introduced a Commercial Solutions Offering (CSO) that is open for a year that covers a number of areas to support what they refer to as the “Hyper Enabled Operator” (https://www.socom.mil/SOF-ATL/Pages/JATF-CSO-CY19.aspx). This CSO is similar to the DIU model where proposers submit short white papers and quad charts. The PMs then down select and request abbreviated proposals, which then can result in an award using OTA funding. This CSO offering lowers the barrier to entry for non-traditional defense contractors who typically do not have the resources for formal proposal writing nor the accounting and finance systems required by the Defense Contract Accounting Agency (DCAA) for FAR-based contracts.

There have also been budgetary discussions in Congress as to the utility of organizations like SOFWERX and DIU, as well as the vagaries of whether or not a new administration will support the initiatives that began under the previous administration. Fortunately, both the former Defense Secretary Mattis, the Chairman of the Joint Chiefs, Gen Dunford, the CTO in AT&L, have all been very visible and vocal in supporting these innovation efforts recently in the press and in congressional testimony (Williams, 2017).

Conclusions

SOFWERX is an interesting study of the application of commercial start-up innovation practices coupled with the unusual requirements of the warfighter for more rapid access to emerging technologies. The SOFWERX model is rapidly growing and evolving and continuing to prove its value to both the warfighter community, as well as to the industrial base of companies with interesting technologies who may have never considered working with the DoD.

References

5 U.S.C. 3715 (2012). Use of partnership intermediaries. Retrieved from https://www.gpo.gov/fdsys/granule/USCODE-2011-title15/USCODE-2011-title15-chap63-sec3715

Andrews, B. (2017, November). Interview.

Carter, A. B. (2016). Remarks announcing Diux 2.0. Washington, DC: DoD. Retrieved from http://www.defense.gov/News/Speeches/Speech-View/Article/757539/remarks-announcing-diux-20

Gibbons-Neff, T. (2016, May 25). Why the U.S. military turned a hipster tattoo parlor into a Special Operations lab. Retrieved from https://www.washingtonpost.com/news/checkpoint/wp/2016/05/25/why-the-u-s-military-turned-a-hipster-tattoo-parlor-into-a-special-operations-lab/?utm_term=.6cfb88e47cbc

Hunt, C. (2017, November). Interview.

Kollars, N. (2017). SOFWERX's return on collision: Measuring open collaborative innovation. *Special Operations Journal, 3*(1), 11–20.

Machi, V. (2017, October 3). SOCOM seeking new unmanned systems capabilities through ThunderDrone. *National Defense Magazine.* Retrieved from http://www.nationaldefensemagazine.org/articles/2017/10/3/socom-seeking-new-unmanned-systems-capabilities-through-thunderdrone

Magnuson, S. (2016, May). Special Ops' "Iron Man" suit on track for 2018. *National Defense Magazine.* Retrieved from http://www.nationaldefensemagazine.org/archive/2016/May/Pages/SpecialOpsIronManSuitonTrackfor2018.aspx

Meissner, C., & Cole, A. (2016, April 16). Three ways to judge the Pentagon's tech-sector outreach. *Defense One.* Retrieved from http://www.defenseone.com/ideas/2016/04/three-ways-judge-pentagons-tech-sector-outreach/127612/

Polaris. (n.d.) MRZR D4: Transportable, agile, mobility for 2–4 (Diesel). Retrieved from https://military.polaris.com/en-us/mrzr-d4/

Sabin, J. D., & Zakner, M. K. (2016). *Analysis of expedited defense contracting methods in the acquisition of emerging technology* (Doctoral dissertation, Naval Postgraduate School). Retrieved from http://calhoun.nps.edu/handle/10945/51609

Stratton-Feix, K. (2019, April). Interview.

SOFWERX. (2019, February 5). Executive brief.

Use of Partnership Intermediaries. (2015).

Williams, L. (2017, December 7). DOD looks to DIUX for the future of acquisition. Retrieved from https://fcw.com/articles/2017/12/07/diux-sasc-acquisition-future.aspx

Appendix A: Example CCE

SOCOM Far Ridgeline Review - Artificial Intelligence/Machine Learning

13 April 2017

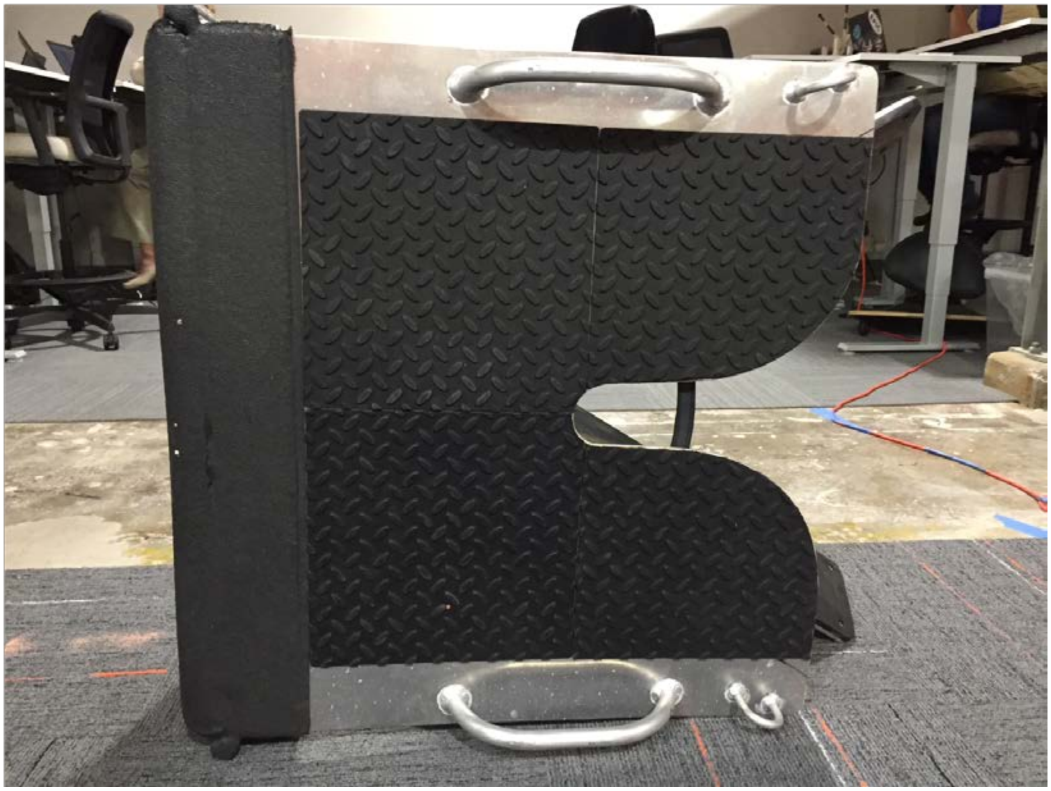
SOCOM personnel shared relevant operational vignettes and SOF use cases with the invited guests to create a shared understanding of the SOF environment. The invited guests shared their expertise in the area of ML and AI, with a goal toward understanding how the various domain knowledge of the guests could impact SOF capability. At the conclusion of the review, all participants had an increased understanding of common areas of interest, and there were a number of follow-on interactions expected to occur.

Event Outcomes

- ML/AI relies on data. SOF will need to resolve the current "stovepipe" mindset to enable effective use of ML/AI. This includes classification issues, ownership issues, proprietary issues, and the "purpose" issue (where data is used only for one purpose, then thrown away). One person noted that by not addressing this problem, we are creating our own self-imposed "denial of service" attack.
- It's important to use ML/AI for the things it's good at. Pattern recognition is one of those things, and is directly relevant to patterns of life.
- ML/AI can move us from reactive to proactive/ predictive.
- We need to collect data all the time. There was a correlation to the SOF truth that states "Competent Special Operations Forces cannot be created after emergencies occur", because data cannot be created AFTER we need it.
- AI/ML not only makes machines intelligent, it feeds human decisions.
- Culture/bureaucracy will be a LARGE impediment to leveraging ML/AI. If our incentives remain focused on reducing risk rather than implementing capability, we won't leverage the innovation.
- Three major portions of ML/AI need to be integrated to create a game-changing capability for SOF: vision, language, and contextual relevance.

SOFWERX 1320 E. 9th Ave., Ste. 100 Tampa, FL 33605 sofwerx.org

Appendix B: Bow Bumper



Appendix C: MRZR

Appendix D: Technology Focus Areas

UNCLASSIFIED

SOFWERX

Collider Event

Table of Contents

UNCLASSIFIED

For more information and to submit, visit: www.sofwerx.org/collider

Panel 25. Better Cost Estimating to Control Program Cost Growth

Thursday, May 9, 2019	
3:45 p.m. – 5:00 p.m.	**Chair: Wendy P. Kunc,** Director, Naval Center for Cost Analysis, Office of the Assistant Secretary of the Navy, Financial Management and Comptroller ***Developing Standard EMD Cost Factors for Major Defense Acquisition Program (MDAP) Platforms*** Matthew Markman, Jonathan Ritschel, Shawn Valentine, and Edward White, Air Force Institute of Technology ***A Reduced Form Model of Cost Growth of Major Defense Acquisition Programs*** David McNicol, Institute for Defense Analyses ***Eliciting Expert Opinion in Acquisition Cost and Schedule Estimating*** Gregory Brown, Air Force Institute of Technology

Wendy P. Kunc—As the Executive Director of the Naval Center for Cost Analysis (NCCA), Ms. Kunc advises Department of the Navy (DON) leadership on cost issues, develops defensible independent cost estimates and assessments for major acquisition programs, provides cost analysis tools, and performs special studies. Ms. Kunc chairs the DON Cost Review Board and the Cost Estimating Stakeholders Group that facilitates collaboration and sharing of information and best practices across the DON and the Department of Defense cost communities.

Ms. Kunc entered the Senior Executive Service in March 2005 and has 27 years of Federal Service.

Ms. Kunc was selected as NCCA's Cost Analysis Tools Division Director in May 2001. In this capacity, she served as program manager for the Naval Visibility and Management of Operating and Support Costs (VAMOSC) management information system. She also managed the Operating and Support Cost Analysis Model (OSCAM) suite as well as other automated cost analysis tools, the NCCA website, and online document library.

Prior to this position, Ms. Kunc spent 15 years with the Department of the Air Force. In June 2000, she was selected to lead the Cost Factors Branch within the Forces Analysis Division of the Air Force Cost Analysis Agency and was responsible for developing the multi-billion dollar Cost Per Flying Hour program. In 1993, Ms. Kunc served as program manager of Air Force VAMOSC for the Deputy Assistant Secretary of the Air Force (Cost and Economics) and led the expansion of VAMOSC into the more comprehensive Air Force Total Ownership Cost (AFTOC) management information system. Ms. Kunc also served as an operations research analyst at what is now the Air Force Intelligence Surveillance and Reconnaissance Agency in San Antonio, Tex. Various positions included Chief of Software Support where she developed Air Force Cryptologic Support Center software applications. Ms. Kunc began her government career as a cartographer with the Defense Mapping Agency.

Ms. Kunc holds a bachelor's degree in mathematics from the University of Missouri, Columbia, Mo. and a master's of science degree in computer information systems from St. Mary's University, San Antonio, Tex. She received a master's of science degree in national resource strategy from the National Defense University and completed the Industrial College of the Armed Forces Senior

Acquisition Course. She completed the National Defense University's CIO certification program in 2005.

Ms. Kunc received the Presidential Meritorious Executive Award in 2010, the Department of the Navy Superior Civilian Service award in 2009, and the Air Force Headquarters civilian award for Outstanding Contribution to Financial Management and Comptroller in 1998. Ms. Kunc received the OSD Comptroller team award for Innovative use of Technology in Financial Management in 1999 and 2002.

Ms. Kunc is a Certified Defense Financial Manager, is Level III certified in the Defense Acquisition Workforce, and is a member of the Acquisition Corps. Ms. Kunc is a member of the American Society of Military Comptrollers and the Society of Cost Estimating and Analysis.

Developing Standard EMD Cost Factors for Major Defense Acquisition Program (MDAP) Platforms

Capt Matthew R. Markman, USAF—graduated from the United States Air Force Academy, CO, with a Bachelor of Science degree in Business Management. Following the completion of his undergraduate degree, he was commissioned as an Officer in the United States Air Force. He completed his Master of Business Administration degree through the American Military University in July 2017. He completed his master's degree in cost analysis at the Graduate School of Engineering and Management, Air Force Institute of Technology in March 2019. Capt Markman is currently assigned to the Air Force Research Laboratory at Wright-Patterson Air Force Base.

Jonathan D. Ritschel—is an Assistant Professor of cost analysis in the Department of Systems Engineering and Management at the Air Force Institute of Technology (AFIT). He holds a BBA in accountancy from the University of Notre Dame, an MS in cost analysis from AFIT, and a PhD in economics from George Mason University. Dr. Ritschel's research interests include cost analysis, factor analysis, defense acquisitions, acquisition reforms, defense economics, public choice, and economic institutional analysis. [jonathan.ritschel@afit.edu]

Edward D. White—is a Professor of statistics in the Department of Mathematics and Statistics at AFIT. He holds a BS in mathematics from the University of Tampa, an MAS from Ohio State University, and a PhD in statistics from Texas A&M University. Dr. White's primary research interests include statistical modeling, simulation, and data analytics. [edward.white@afit.edu]

Shawn Valentine—is an Operations Research Analyst at AFLCMC Cost Staff, Wright-Patterson Air Force Base, OH. He provides cost estimate support to various AFLCMC development planning programs such as Next Gen Tanker, Combat Air Force Advanced Live Operational Training, and ISR Collect. He heads the research department, which develops and researches Air Force research initiatives at AFLCMC. He also serves as Chair for Air Force Institute of Technology student thesis topics. He has experience in financial management and operations research throughout all Air Force appropriations and acquisition phases. [shawn.valentine@us.af.mil]

Abstract

This paper creates standard cost factors that more accurately reflect observed outcomes in the development stages of major programs. Specifically, this effort creates 443 new cost factors that are delineated by five categories: commodity type, contract type, contractor type, development type, and service. The factors are developed for those elements that are "common" in a wide array of projects such as program management, systems engineering, data, or training. This paper establishes factor values at the Work Breakdown Structure (WBS) element level for each subcategory of the five identified categories. Coefficient of Variation (CV) values were found to be high (71.86% to 179.87%) in each subcategory. In a refined subset of the dataset, the CV decreased, indicating that the average percent estimating error improved when more detailed information was available. The outcome of this research is that cost estimators will have a reference tool of 443 unique factors for creating estimates and conducting the iterative process of refining cost estimates.

Introduction

Background

Cost analysts have a range of models and techniques that are utilized in a variety of ways on Major Defense Acquisition Program (MDAP) estimates. One of these tools is the application of standard cost factors. Factors are utilized as primary and as cross-check methodologies when estimating "common" cost elements such as program management, systems engineering, training, site activation, and spare costs.

Currently, the research division of the Air Force Life Cycle Management Center (AFLCMC) periodically publishes standard factor tables for aircraft Engineering and Manufacturing Development (EMD) that capture prime contractor data for a limited selection of clean-sheet design aircraft programs. Despite the utility of the AFLCMC published tables, additional data exists that can assist in refining these factors, as well as developing new factors to include Army, Navy, and Joint programs. Other identified gaps in currently published EMD factors include neglected *commodity categories* (e.g., electronic/automated software, missiles, ordnance, space, and Unmanned Aerial Vehicles [UAVs]), *development types* (e.g., modification programs), and *subcontractor* data. Each additional category of data provides estimators the ability to accomplish more in-depth analysis based on the type of program in question. Thus, the expansion and refining of factors for EMD programs will provide estimators with a more robust tool set upon which to draw from, ultimately leading to more precise estimates going forward.

Research Objectives

The purpose of this paper is to investigate the current state of EMD cost factors, refine existing standards where available, and develop and publish new cost factors for operational use by cost analysts in an array of project types. Furthermore, the conclusions from this paper help determine where future efforts should be focused towards gathering new data and/or refining existing factors. The specific objectives are to

1. Develop a suite of standard cost factors for incorporation into the current cost estimator toolkit.
2. Create a software tool for tailoring cost factors by unique characteristics such as commodity type, contract structure, or program features.

Literature Review

Cost Estimating Methodologies

The toolkit of a cost analyst consists of four primary estimating methods, as well as secondary techniques, but the use of standard factors represents a commonly utilized practice (GAO, 2009). With billions in taxpayer dollars at stake each year within the Department of Defense (DoD) acquisition budget, it is imperative that program offices, and specifically cost analysts, understand their program, draw conclusions from past programs, and leverage technology to arrive at estimates in which the American public can place their confidence and trust (GAO, 2009). Because of this responsibility, this paper aims to expand the breadth of analytical tools available, specifically with respect to the utilization of standard factors in MDAPs.

Several key documents designate and define the cost estimating methodologies utilized within the DoD, including the *Air Force Cost Analysis Handbook* (AFCAH) and the GAO *Cost Estimating and Assessment Guide*. These publications assist in setting a baseline for program offices and cost analysts to craft credible and consistent cost estimates, as well as an overarching legal requirement for the DoD to have policies in place to safeguard the billions of taxpayer dollars afforded to MDAPs each year (GAO, 2009). The four techniques outlined in the AFCAH include analogy and factor, parametric, build-up (engineering), and expert opinion (subject matter expert; Department of the Air Force, 2007). The introduction of more than one estimating technique provides cost analysts with the ability to triangulate a point estimate that considers levels of detail not fully captured by individual techniques or estimates. Furthermore, this approach serves as a crosscheck to ensure estimates do not fall too far outside the bounds of reasonableness for the given program.

Figure 1 from the AFCAH details the four cost estimating methods and shows the progression over the program life cycle.

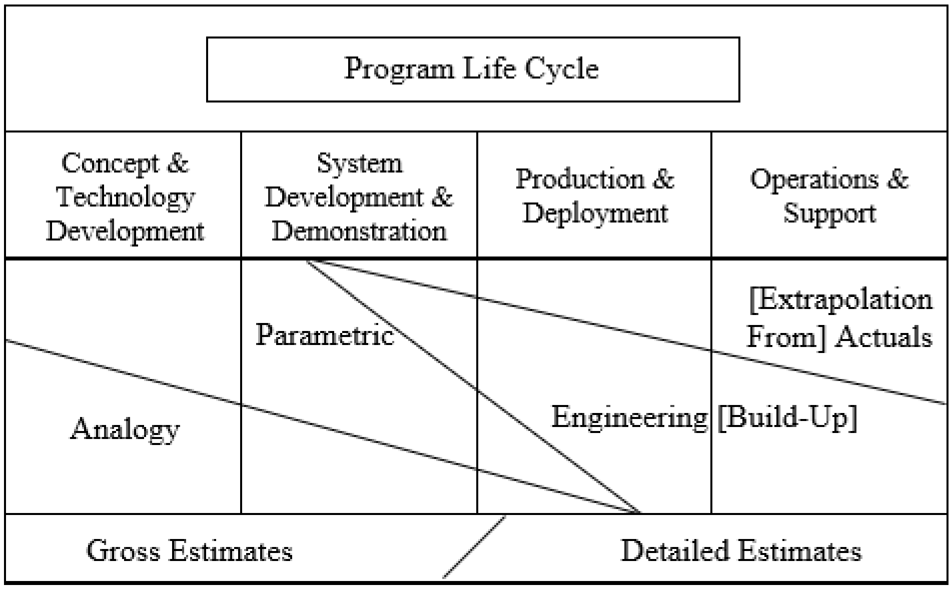


Figure 1. **Selection of Methods**
(AFCAA, 2007)

The parametric estimating technique represents an approach based upon a statistical relationship drawn between historical costs and certain characteristics (program, physical, and performance), also referred to as cost drivers (GAO, 2009). The build-up method of cost estimating consists of an exhaustive collection of lower level program element estimates followed by a roll-up of each estimate to arrive at the total program cost (Department of the Air Force, 2007). Often referred to as the engineering approach, this technique is based largely on in-depth engineering data and requires a great deal of labor and material cost information to produce a reliable estimate. The expert opinion approach to cost estimating relies on information gathered directly from subject matter experts (SME) in each area of the program, most often in instances of early concept design or development where data is scarce (Department of the Air Force, 2007). The analogy method of cost estimating takes historical data from existing similar programs or systems and applies a scaling factor (or range of factors) to account for differences in the new system and arrive at a feasible estimate (Mislick & Nussbaum, 2015). The scaling factor(s) represent disparities between the old and new programs in the context of size, performance, technology, complexity, and many others, and sets an initial estimate given the early stage of the program’s life cycle (GAO, 2009).

Elements of the Work Breakdown Structure (WBS)

The WBS concept in Major Defense Acquisition Programs (MDAP) has remained relatively constant over the past several decades (DoD, 2005). It represents a decomposition of a project into smaller, more manageable components and is sometimes referred to as the management blueprint for the project (Mislick & Nussbaum, 2015). The WBS is mandated and governed by MIL-STD-881D, ultimately fulfilling broader requirements set forth in DoD Instruction 5000.2; this DoD publication aims to maintain uniformity in definition and consistency of approach for programs developing a WBS (DoD, 2018). For the sake of consistency, the DoD has revised and updated guidance regarding the WBS only when major technological advances or changes in the acquisition process warranted such action (DoD, 2005).

The WBS consists of three primary hierarchical levels, with a fourth and fifth sometimes included in expanded forms; for this paper, only the second level is addressed. Level II of the WBS captures major elements subordinate to the system identified by level I and consists of prime mission products, including all hardware and software elements. Level II also includes combinations of system level services applicable to the program including the following elements common to most programs: integration and assembly, system test and evaluation (ST&E), systems engineering/program management (SE/PM), common support equipment (CSE), peculiar support equipment (PSE), training, data, operational/site activation, and initial spares and repair parts (DoD, 2018). These common elements at level II of the WBS are the focus for developing factors in this paper. Benefits of the WBS structure mandated by MIL-STD-881D include ease of normalization of data and information across a variety of commodity types and DoD agencies and the ability to reference past and current MDAPs to better understand and forecast their own costs, schedules, and overall program.

Previous Research on Factors in Cost Estimating

Extensive research on factors in cost estimating does not exist to the extent necessary to fully and efficiently utilize the technique, creating a gap in cost analysts' ability to employ the technique effectively. While the Air Force acquisition cost analyst community has conducted previous studies by Wren (1998) and Otte (2015) in the Engineering and Manufacturing Development (EMD) phase of the lifecycle, these were all very narrow in scope and applied solely to a limited subset of aircraft programs. Large gaps exist for additional commodity types besides aircraft, modification programs, subcontractor data, and even contract type.

The utility of factors in cost estimating extends beyond just acquisition programs, reaching across various government agencies and functions to support more efficient budgeting and execution of taxpayer dollars (Mislick & Nussbaum, 2015). With such widespread utilization of the factor method, a variety of research exists, especially within the DoD. The Naval Center for Cost Analysis (NCCA) engages in continuous research on cost estimation and publishes periodic findings to guide and strengthen cost analysis within the Navy (NCCA, 2018). In addition to this research, the NCCA conducts economic and business case analyses for a variety of issues within the Department of the Navy, creating benchmarks from which factors can be created for cost estimates (NCCA, 2018). While all military branches are governed by general DoD guidance, service-specific directives illustrate some differences in the application of certain requirements, such as cost estimation. The Air Force's use and research of the factor method extends beyond the acquisition world and is detailed in lower level directives like functional area Air Force Instructions (AFI) to better predict costs in logistics, personnel, programming, and flying hour operations (Department of the Air Force, 2018). Additionally, the Air Force publishes dozens of factor tables for personnel to utilize for estimates specific to their respective functions; these tables are updated regularly and serve as a benchmark for cost estimation within the Air Force. Another illustration of cost factors' prominence in the DoD comes from the publishing of Area Cost Factors (ACF) each year to assist in preparation and review of military construction, Army and Army Family Housing projects, and a variety of other facility related projects (PAX, 2018). These factors are the reflection of a selection of characteristics to accomplish broad levels of analysis and estimation and serve as benchmarks for estimators to then add their own individual details to modify the factors and arrive at a credible estimate (PAX, 2018).

Utility of Factors in Cost Estimating

The analogy and factor method of cost estimating is used by DoD analysts constructing estimates for MDAPs, but this approach also serves the private and public sectors in formulating cost estimates for large projects. In the case of public works projects, specifically transportation infrastructure, there is sometimes a lack of credible estimates available due to the financial interests of potential contractors and the agenda that accompanies large contract awards (Flyvbjerg, Holm, & Buhl, 2002). The issue can be at least partially relieved by the establishment of standard factors for analogous projects to protect entities (state and local governments in many cases) in need of these major services from being misled with regard to cost estimates. One issue, however, with this remedy lies in the lack of exhaustive analogy and factor studies in existence and/or available to those in need of the data (Flyvbjerg, Holm, & Buhl, 2002). While it can be argued that MDAPs pose entirely different challenges compared to large infrastructure projects, the common theme lies in the vast complexity and likelihood of changes that each type of project contains. Infrastructure projects do not represent the sole area in need of improved estimation; numerous international studies have found construction projects in general exhibit cost overruns and inefficiencies that can be traced to poor estimating practices (Baloi & Price, 2003; Elfaki, Alatawi, & Abushandi, 2014). Such widespread occurrence of inaccurate estimating necessitates a focus on the establishment of improvements in the resources available to estimators, with historical standard factors being one of those resources.

While the practice of cost estimating exists in different capacities around the world, the common theme remains the intent to arrive at an estimate that aids in the decision-making process of the project. The shortcomings of the use and structure of historical data and information are illustrated by large projects' consistent cost overruns (Riquelme & Serpell, 2013). The myriad of issues identified in projects around the world reinforces the need for additional data that will provide analysts the ability to effectively leverage historical information to arrive at a credible cost estimate. The data required to perform the necessary analysis for cost estimating requires scrutiny to ensure accuracy and applicability, but the time invested in this pursuit yields more effective estimates. The analogy and factor technique represents just one of many cost estimating methodologies, but when properly utilized in any field or environment it aids in achieving an estimate that embodies completeness, reasonableness, and analytic defensibility (Mislick & Nussbaum, 2015).

The creation and utilization of standard factors makes it possible to conduct more effective and extensive analysis at a variety of levels to construct credible cost estimates, especially in programs early in their lifecycle or with limited information regarding the central task (Mislick & Nussbaum, 2015). Several of the primary areas in which new additional analysis would be beneficial for program offices include commodity type, contractor designation (prime or sub), and contract type. These characteristics of a program serve as a starting point for data normalization, as well as more in-depth scrutiny within the structure of the WBS. The use of qualitative context factors like those dictated by the WBS format assist in the effective interpretation and use of historical information, which further strengthens the legitimacy of cost estimates that employ the standard factor approach (Riquelme & Serpell, 2013). Using the level II WBS elements as a guide, analysts have virtually every historical MDAP with relevant data at their fingertips to create factors to then extrapolate upon for their specific program. The value of a central database that encompasses all commodity types, contractor designations, and contract types lies in the ability to conduct analysis at each of these respective levels and manipulate the data to create factors for each level II element of the WBS. Through the creation of factors, cost analysts throughout the DoD can target specific analytical levels and more effectively formulate credible, defensible estimates for MDAPs.

Methodology

Data

The data gathered in this paper is from the Defense Automated Cost Information Management System (DACIMS), which exists within the Cost Assessment Data Enterprise (CADE) system. DACIMS contains Cost Data Summary Reports (CDSR), often referred to as 1921s, which contain the necessary cost data to establish factors for the MDAPs targeted for this research. EMD data was chosen as the only life-cycle phase to be analyzed based on a gap in this area identified by the literature review for this research. The dataset consists of 102 programs spanning from 1961 to 2017, representing a broad range of programs across numerous commodity types and services.

While 189 programs are available within CADE, only 102 of those programs fit the criteria for inclusion in the final dataset. Table 1 depicts the exclusion criteria and accompanying number of programs not utilized for this research.

Table 1. Dataset Exclusions

Category	Number Removed	Remaining Programs
Available Programs in CADE		189
Excluded Commodity Types	35	154
No EMD Data	25	129
1921 File Format Not .XLS	27	102
Final Dataset for Analysis		102

Programs containing only initial 1921 data were excluded. A small portion of the data came from interim 1921s. In these instances, the data contained on the interim 1921s was equal to or greater than the final contract price. There were 27 programs that contained data but lacked accessible files within CADE, resulting in the entire program’s exclusion from the dataset. These were primarily older programs with manually transcribed data from the 1980s or earlier and in many instances contained illegible data.

Differentiation between contractor type, as well as unique aspects of programs (blocks, phases, etc.) resulted in multiple factors for most programs, each with their own level II WBS elements. Table 2 provides an overview of the major characteristics of the final dataset for this research, which consisted of 443 unique factors.

Table 2. Dataset Characteristics

Category	Total
Unique Factors Created	443
Commodity Type	
Aircraft	245
Electronic/Automated Software	118
Missile	22
Ordnance	12
Space	36
UAV	10
Contractor Type	
Prime	308
Subcontractor	135

Category	Total
Development Type	
Commercial Derivative	4
Modification	135
New Design	150
Prototype	9
Subsystem	105
Variant	40
Service	
Air Force	196
Army	94
Multiple	24
Navy (includes Marine Corps)	129

Category	Total
Contract Type	
CPAF	74
CPFF	39
CPIF	66
Cost-Other	135
FFP	27
FPI	20
FPIF	19
Fixed-Other	6
Unknown	57

Factor Calculation

The cost element factors contained in this research are the ratio (percentage) of the individual level II WBS elements to a base cost. The base cost is represented by a program's Prime Mission Equipment (PME) value, which does not include the contractor's fee or miscellaneous expenses (general and administrative [G&A], undistributed budget, management reserve, facilities capital cost of money [FCCM]). An example of this ratio is the dollar value or cost of SE/PM divided by the program's PME value. After establishing cost factors for the level II WBS elements, it is possible to develop composite factors for a myriad of unique categories. Specific level II WBS elements can be examined in groupings to establish aggregate values that represent an average or percentage that can be used in formulating estimates. These groupings allow for analysis at innumerable levels, such as fixed wing aircraft, rotary wing aircraft, a specified contractor for radar modifications, a specified contractor's role in a program (prime versus sub), a specified period for a certain commodity type, and many more. An averaged cost factor represents a more accurate factor as it guards against the skewness that can result from calculations based on single data points.

Descriptive Analysis

Once the factors were established for each program, the mean, median, and standard deviation values for the various program groupings were calculated. In addition, interquartile ranges were calculated to examine variability among factors. This allowed for descriptive analysis. Similar to the innumerable amount of potential composite cost factors, there are many comparisons that can be performed using this dataset. This research highlights five major categories: service, commodity type, contractor designation, contract type, and development type. Table 3 lists the categories and respective sub-categories for which factors were established in this research.

Table 3. Categories for Comparison Analysis

Category	Subcategories
Contractor Designation	Prime, Sub
Service	Army, Navy (includes Marine Corps), Air Force, Multiple
Commodity Type	Aircraft, Electronic/Automated Software, Missile, Ordnance, Space, UAV
Development Type	Modification, New Design, Prototype, Subsystem, New MDS Designator, Commercial Derivative
Contract Type	CPAF (Cost Plus Award Fee), CPFF (Cost Plus Fixed Fee), CPIF (Cost Plus Incentive Fee), Cost-Other (Other than CPAF, CPFF, CPIF), FFP (Firm Fixed Price), FPI (Fixed Price Incentive), FPIF (Fixed Price Incentive Firm Target), Unknown

Results and Analysis

Systems Engineering/Program Management (SEPM)

The SEPM element of the WBS represents one of the more prominent factors in this analysis in several ways. First, SEPM had the fewest amount of blank values of any WBS element, with only 19 blanks, or 4.29%. SEPM values ranged from 0.43% to 4768% of Prime Mission Equipment (PME), indicating potential reporting anomalies and/or additional issues in the extreme upper values. To establish meaningful exclusion criteria, the distribution of all SEPM values was computed using JMP software. Analysis of the distribution resulted in values above 150% of PME being removed from the dataset for all remaining SEPM analysis. These excluded values represented only 4.06% of the dataset, were more than three standard deviations from the mean, and in most cases were part of a Major Defense Acquisition Program (MDAP) with a total PME of less than ten million dollars. Table 4 shows the distribution of SEPM values after exclusions were made and provides descriptive statistics utilized in further analysis.

Table 4. SEPM Descriptive Statistics

Max	1.4655	**Mean**	0.3884
75%	0.5319	**Std Dev**	0.3015
Median	0.3038	**N**	406
25%	0.1643		
Min	0.0043		

The resulting distribution for the SEPM WBS element is characterized by many data points, as well as a high standard deviation value. The distribution's central points lie between 0.25 and 0.4, which is reinforced by the mean and median values of 0.38 and 0.30, respectively. Table 5 displays an example of the descriptive statistics broken out by category for the SEPM WBS element. The detailed analysis displayed in Table 5 for subsequent WBS elements (Training, Data, PSE, CSE, Site Activation, Other, and Spares) is not provided in this paper due to space constraints but is available upon request.

Table 5. SEPM Summary Table

	Mean	Std Dev	N	Max	75%	Median	25%	Min
Service								
Air Force	0.3685	0.2755	177	1.324	0.4894	0.2972	0.159	0.0043
Army	0.508	0.3372	91	1.3453	0.6989	0.4426	0.2514	0.0098
Navy	0.3393	0.3039	115	1.4655	0.465	0.2551	0.1421	0.0105
Multiple	0.3142	0.2053	23	1.0007	0.4047	0.2699	0.1626	0.0903
Development Type								
Modification	0.3484	0.2555	124	1.3191	0.4954	0.2845	0.1539	0.0043
New Design	0.4738	0.3472	131	1.4655	0.6582	0.3759	0.219	0.0053
Prototype	0.1906	0.1472	8	0.39	0.3417	0.1783	0.0627	0.0126
Subsystem	0.373	0.2816	101	1.324	0.5343	0.2793	0.161	0.0105
New MDS Designator	0.3249	0.2924	39	1.3619	0.3887	0.2517	0.1154	0.0445
Commercial Derivative	0.184	0.1011	3	0.2676	0.2676	0.2128	0.0716	0.0716
Contractor Type								
Prime	0.3849	0.3068	284	1.3619	0.4896	0.2947	0.1609	0.012
Subcontractor	0.3966	0.2898	122	1.4655	0.5613	0.3336	0.1724	0.0043
Commodity Type								
Aircraft	0.3025	0.2385	227	1.3619	0.4115	0.2292	0.1421	0.0105
Electronic/Automated Software	0.5463	0.3511	107	1.4655	0.7816	0.4875	0.2568	0.0098
Missile	0.5014	0.3297	20	1.2822	0.7695	0.3897	0.2682	0.0576
Ordnance	0.3426	0.1737	11	0.6117	0.5007	0.285	0.2439	0.0811
Space	0.3825	0.3093	31	1.3191	0.4972	0.3109	0.1488	0.0043
UAV	0.4913	0.3217	10	1.324	0.5435	0.3655	0.303	0.2617
Contract Type								
CPAF	0.4128	0.2641	66	1.2792	0.5792	0.3649	0.2206	0.0337
CPFF	0.5189	0.3896	37	1.3453	0.7022	0.4233	0.2387	0.0053
CPIF	0.3905	0.2987	61	1.2924	0.522	0.2729	0.18	0.0276
Cost-Other	0.4082	0.3103	126	1.4655	0.5874	0.3175	0.1767	0.0043
FFP	0.2457	0.2531	25	1.0786	0.3494	0.156	0.0871	0.0105
FPI	0.2118	0.2232	17	1.0081	0.2349	0.1694	0.0729	0.0484
FPIF	0.4203	0.2811	19	1.2822	0.5578	0.3931	0.2218	0.0675
Fixed-Other	0.572	0.2327	4	0.8384	0.8026	0.5427	0.3707	0.3643
Unknown	0.3131	0.2573	51	1.3144	0.4426	0.243	0.1275	0.0385

System Test & Evaluation (ST&E)

ST&E contained the second largest amount of datapoints for analysis. Only 57 rows, or 12.87%, of the total factors were blank values for ST&E. Values for ST&E ranged from below 0.1% to as high as 1485% of PME, indicating potential reporting anomalies in the upper extreme values. ST&E values below 0.1% of PME were excluded as they represented trivial dollar amounts (less than $16K in most cases). On the high end of the distribution, ST&E values above 150% of PME were excluded, and in all five instances the PME dollar amount for the MDAP was less than ten million dollars. The upper and lower exclusions of ST&E values make up only 2.71% of the dataset. Table 6 depicts the ST&E distribution as well as its accompanying descriptive statistics. Table 7 displays an example of the descriptive statistics broken out by category for the ST&E WBS element.

Table 6. ST&E Descriptive Statistics

Max	1.0776	**Mean**	0.2144
75%	0.2999	**Std Dev**	0.2027
Median	0.1611	**N**	374
25%	0.0658		
Min	0.0012		

Table 7. ST&E Summary Table

	Mean	Std Dev	N	Max	75%	Median	25%	Min
Service								
Air Force	0.2251	0.2074	166	0.9641	0.328	0.1672	0.0668	0.0013
Army	0.2157	0.1915	80	1.0575	0.2784	0.1992	0.0793	0.0012
Navy	0.2201	0.215	105	1.0776	0.3083	0.1582	0.0697	0.0032
Multiple	0.1059	0.1027	23	0.3312	0.1821	0.0642	0.0207	0.0021
Development Type								
Modification	0.2155	0.2193	119	1.0776	0.2986	0.1396	0.0623	0.0013
New Design	0.2143	0.188	114	1.0575	0.304	0.1817	0.0611	0.0016
Prototype	0.2673	0.1028	9	0.4561	0.325	0.282	0.1792	0.1177
Subsystem	0.1744	0.1883	89	0.8523	0.2378	0.1038	0.0428	0.0012
Variant	0.2934	0.2281	39	0.9436	0.4288	0.2456	0.0987	0.0083
Commercial Derivative	0.1804	0.1432	4	0.3659	0.328	0.1585	0.0548	0.0388
Contractor Type								
Prime	0.2294	0.2019	274	1.0776	0.3089	0.1838	0.0754	0.0012
Subcontractor	0.1733	0.2001	100	1.0575	0.2396	0.0999	0.0305	0.0016
Commodity Type								
Aircraft	0.2498	0.2139	225	1.0776	0.3515	0.2036	0.021	0.0013
Electronic/Automated Software	0.1702	0.1924	88	1.0575	0.2199	0.1038	0.0348	0.0012
Missile	0.2041	0.1772	18	0.7363	0.2615	0.1842	0.0619	0.0243
Ordnance	0.1513	0.0998	11	0.3389	0.2468	0.0961	0.0704	0.0596
Space	0.0778	0.0879	23	0.3797	0.1157	0.0448	0.021	0.003
UAV	0.2068	0.1273	9	0.3924	0.3266	0.1893	0.0887	0.0444
Contract Type								
CPAF	0.1802	0.1964	63	1.0575	0.2761	0.1072	0.038	0.0025
CPFF	0.1671	0.2095	31	0.8523	0.2213	0.0791	0.0253	0.0016
CPIF	0.2586	0.22	55	1.0677	0.3796	0.1997	0.0829	0.0021
Cost-Other	0.1824	0.1748	113	0.9641	0.2618	0.1277	0.0474	0.0012
FFP	0.1777	0.1503	20	0.4561	0.3426	0.13	0.0588	0.0118
FPI	0.3907	0.1991	20	0.9436	0.5222	0.3267	0.2803	0.1276
FPIF	0.2876	0.2168	17	0.7307	0.3371	0.2167	0.1233	0.0226
Fixed-Other	0.2714	0.2483	4	0.6104	0.5283	0.2227	0.0632	0.0298
Unknown	0.2248	0.2163	51	1.0776	0.2416	0.1608	0.0968	0.0044

Despite the high value for standard deviation displayed by the ST&E WBS element, the resulting mean and median values lie within close proximity to one another in the distribution. ST&E also exhibited a large number of available data points, with only 15.5% of the entire dataset excluded for analysis.

Training

The Training WBS element showed a sharp decline in reported data, with more than half of the dataset containing no values for Training. Despite 235 (53.05%) of the rows being blank, this element still contains ample data for analysis. The vast majority (85.4%) of the Training data comes from the aircraft and electronic/automated software commodity types. Distributional analysis resulted in the threshold for inclusion in the analysis of this element being set at values above 0.05% of PME. This resulted in the exclusion of 14 (3.16%) data points, the majority of which were less than $100K amounts in multi-million-dollar MDAPs. Also, two Training values above 80% were excluded, which amounted to less than 0.5% of the total dataset. These extreme upper values of 82% and 2275% represented a commercial derivative program and a likely reporting anomaly, respectively. Table 8 shows the distribution and descriptive statistics for the 192 values analyzed for the Training WBS element. The detailed analysis similar to Table 5 for Training is available upon request.

Table 8. Training Descriptive Statistics

Max	0.4237	**Mean**	0.0342
75%	0.037	**Std Dev**	0.0648
Median	0.0101	**N**	192
25%	0.0031		
Min	0.0006		

The Training WBS element contained data for less than half of the entire dataset. Its standard deviation value was high in relation to the calculated mean value, due in part to several data points in the right tail of the distribution. The Training data resided largely between the values of 0.01 and 0.04.

Data

The Data WBS element lacked 176 values, or 39.73% of the total dataset. Data is similar to Training with respect to its concentration of information within the aircraft and electronic/automated software commodity groups. It surpasses the characteristics of Training, with 87.3% of the dataset for the Data WBS element coming from these two commodities. Data represented the lone element with no additional exclusions beyond blank values, as the distribution was much more concentrated than other elements. Table 9 provides a look at the descriptive statistics for the Data WBS element. The detailed analysis similar to Table 5 for Data is available upon request.

Table 9. Data Descriptive Statistics

Max	0.3935	**Mean**	0.0364
75%	0.0367	**Std Dev**	0.0568
Median	0.0186	**N**	267
25%	0.0074		
Min	<0.0001		

While the Data WBS element offered values for over 60% of the entire dataset, its distribution is characterized by a high standard deviation value and numerous values well beyond three standard deviations from the mean of 0.03.

Peculiar Support Equipment (PSE)

PSE contained only 149 values of data. Blank PSE values make up 64.56% of the entire dataset. Upper and lower exclusions add another 1.8% to the amount excluded. The upper exclusions made were only two values, one of which was nearly 300% of PME, indicating likely reporting anomalies, and the other well above three standard deviations and part of a multinational development effort. The concentration by commodity type is similar to the Training and Data WBS elements, with 65.8% of the dataset coming solely from the aircraft commodity type. Table 10 shows the descriptive statistics for PSE. The detailed analysis similar to Table 5 for PSE is available upon request.

Table 10. PSE Descriptive Statistics

Max	0.44	**Mean**	0.0584
75%	0.0629	**Std Dev**	0.0867
Median	0.0217	**N**	149
25%	0.0074		
Min	0.0001		

The PSE WBS element displays a concentration of data points between the values of 0.01 and 0.05. Beyond that concentration, the data is spread as far as five standard deviations from the mean. The 149 data points for PSE account for only 33.6% of the entire dataset.

Common Support Equipment (CSE)

CSE represented a sharp decline of available data, resulting in only 50 values for analysis. The CSE WBS element is also made up primarily by the aircraft commodity type (62%), and then evenly distributed between each of the remaining types. Only two values (0.45%) were excluded from the CSE analysis, both of which were beyond three standard deviations and indicative of reporting anomalies based on their extremely high values. The distribution for CSE lacks any major shape with data points spread several standard deviations from the mean value of 0.015. Full descriptive statistics for the CSE WBS element are shown in Table 11. The detailed analysis similar to Table 5 for CSE is available upon request.

Table 11. CSE Descriptive Statistics

Max	0.1272	**Mean**	0.0151
75%	0.0115	**Std Dev**	0.0291
Median	0.0019	**N**	50
25%	0.0006		
Min	<0.0001		

Site Activation

Site Activation mirrored the limited availability quality of CSE, offering only 47 data points, or 11.29% of the total factors, for analysis. The 47 data points exclude three upper extreme values beyond three standard deviations. The majority of the values (78.7%) for the Site Activation WBS element are comprised of the aircraft and electronic/automated software commodity types. The Site Activation descriptive statistics are summarized in Table 12. The detailed analysis similar to Table 5 for Site Activation is available upon request.

Table 12. Site Activation Descriptive Statistics

Max	0.3464	**Mean**	0.0386
75%	0.0432	**Std Dev**	0.0706
Median	0.004	**N**	47
25%	0.0005		
Min	<0.0001		

Almost 90% of the dataset was excluded from the Site Activation WBS element's analysis, and such a small sample size yielded a distribution devoid of a dominant shape. The standard deviation value was nearly double the value of the mean and data points encompassed a range that exceeded four standard deviations.

Spares

The Spares WBS element exhibited a low number of data points. Only 84 values were analyzed after removing the 358 blanks and one upper extreme value that was above 100% of PME. The concentration by commodity type for the Spares WBS element is similar to the Training, Site Activation, and Other WBS elements with 86.9% of the data points coming from aircraft and electronic/automated software. The descriptive statistics and distribution for Spares is shown in Table 13. The detailed analysis similar to Table 5 for Spares is available upon request.

Table 13. Spares Descriptive Statistics

Max	0.226	**Mean**	0.0362
75%	0.0574	**Std Dev**	0.0436
Median	0.0174	**N**	84
25%	0.0035		
Min	<0.0001		

Less than 20% of the dataset was available for analysis for the Spares WBS element. Its values were not characterized by large disparities like several other WBS elements' values, with a standard deviation just slightly higher than the mean. Its data points were concentrated between 0.01 and 0.05.

Timeframe Specific Analysis

Recall from the initial dataset exclusion criteria in Table 1, 27 programs were excluded due to inaccessible files or illegible data entries (largely programs from the 1980s or before). To determine whether this exclusion of these older programs had an effect on the factors developed, a timeframe specific analysis on a subset of the data spanning the past two decades was accomplished using 1998 as the cut-off date. Table 14 displays the descriptive statistics for the SEPM WBS element for the original dataset, as well as the revised dataset spanning the most recent 20 years.

Table 14. SEPM Descriptive Statistics Comparison

Commodity	Original Mean	1998-Pres Mean	Original Median	1998-Pres Median	Original CV	1998-Pres CV
Aircraft	0.3025	0.3433	0.2292	0.2727	78.84	71.78
Electronic/Automated Software	0.5463	0.5479	0.4875	0.4875	64.27	66.76
Missile	0.5014	0.5014	0.3897	0.3897	65.77	65.77
Ordnance	0.3426	0.3484	0.285	0.3409	50.7	52.22
Space	0.3825	0.4059	0.3109	0.3109	80.86	83.38
UAV	0.4913	0.5154	0.3655	0.3887	65.49	64.32

The descriptive statistics of the subset of data for SEPM are similar in most cases, and identical in some, to the original dataset. The consistency displayed between the subset and original dataset leads to the conclusion that the 27 programs excluded due to inaccessible files or illegible entries would likely not affect the descriptive statistics or statistical analysis conducted in this research.

Analysts should always be as specific as possible when establishing estimates, especially for the SEPM WBS element. However, for the majority of the remaining WBS elements, analysts can include a broader dataset to arrive at an estimate, at least until greater levels of detail are available.

Purpose Specific Analysis

The distributions and descriptive statistics of the values for each WBS element reveal large Coefficient of Variation (CV) values (standard deviations divided by mean) in each category. Table 15 shows the CV means for each WBS element.

Table 15. Coefficient of Variation Summary

WBS Element	Collective Mean	Collective Std Dev	CV
SEPM	0.3802	0.2732	71.86%
ST&E	0.2117	0.1822	86.07%
Training	0.0295	0.0503	170.51%
Data	0.0331	0.0477	144.11%
PSE	0.0538	0.0749	139.22%
CSE	0.0149	0.0268	179.87%
Site Activation	0.0307	0.0526	171.34%
Spares	0.0787	0.1375	174.71%

Because the standard deviations are so large for this dataset, statistical analysis will likely not identify differences in certain instances where a cost analyst may identify differences through practical analysis. An example scenario is provided to demonstrate the utility of filtering data down to lower levels through utilization of program-specific information in a hypothetical initial cost estimate.

Scenario Example

This scenario pared the dataset down to only prime contractor data for Army MDAPs in the electronic/automated software commodity type. The development type category was examined, looking only at the SEPM WBS element. Through knowledge of the unique program characteristics, the analyst is able to reduce the CV in this illustrative example by more than 20% for the SEPM element. This is just one example (of numerous) in which program-specific knowledge can utilize the factors developed here to create more accurate estimates.

Conclusions

This research resulted in 443 new cost factors created from a multitude of diverse programs. Factors were developed by *development type* (commercial derivative, modification, new design, prototype, variant, and subsystem), *contractor type* (prime and sub), *Service* (Air Force, Army, Navy, and Multiple), *contract type* (various) and *commodity type* (aircraft, electronic, missile, ordnance, space, and UAV).

The descriptive statistics were examined for each category, as well as each level II WBS element. This revealed large standard deviation values and large CV values, pointing to the conclusion that each MDAP presents unique characteristics that must be explored and understood to make the inclusion of its data truly meaningful in the context of constructing a cost estimate. The practicality of achieving an in-depth understanding of each program utilized for a factor and analogy cost estimate is not realistic in many cases. Thus, the "preliminary" nature of many factor and analogy estimates. These generic composite factors represent a starting point for analysts in instances where MDAP characteristics may be unrefined (i.e., broad capability deliverable(s) with undefined processes). Given the fluid nature of estimates at this stage of developing requirements, a robust dataset remains appropriate. Once a program's requirements have been solidified and the manner in which they will be accomplished is well-defined, analysts can begin to refine their dataset to MDAPs with direct application to their program. The intent of this research is to make the dataset utilized for analysis available to DoD analysts to enable an approach to factor creation that can be tailored to the needs of the individual.

Practical analysis provides a valuable approach to understanding the data utilized for an estimate. In the context of factor cost estimating, practical analysis offers the ability for estimators to examine a dataset and determine logically which data points to include or exclude. The practical analysis can be in addition to or in place of statistical analysis, depending on the situation. This research serves as a precursor to statistical analysis to be conducted on this dataset. An analyst constructing an estimate for a new cargo aircraft engine for the Air Force may find no statistical difference between SEPM values for a dataset of 100 factors. However, if the analyst learns the program will likely award some type of fixed contract, the dataset can be refined to exclude inapplicable MDAP factors. The dataset becomes smaller but more precise and the potential for statistical differences between the smaller set of subcategories must be examined. The ability to establish both general and specific estimate values strengthens the defensibility of the estimate by displaying a range of values and explicit reasoning for the merits of each one.

Significance of Results

This paper represents one of the largest DoD factor studies for MDAPs in the EMD phase conducted to date. Previous efforts within the Air Force Lifecycle Management Center

(AFLCMC; Wren, 1998; Otte, 2015) established factor values for specific purposes and System Program Offices (SPOs), whereas this effort is intended for wider-access distribution accessible to analysts across the DoD to accomplish individualized analysis. The compilation of EMD data contained in 443 separate Cost Data Summary Reports (CDSRs) into a single location provides DoD analysts the ability to streamline estimate formulation while also increasing the breadth of data from which estimates are based. The descriptive statistics for each WBS element and accompanying summary tables provide analysts the ability to create an initial estimate quickly. With this estimate as a placeholder, the analyst can then incorporate statistical and/or practical analysis to arrive at a more accurate estimate. These steps can be performed as an iterative process as more details emerge, further refining the estimate.

Summary

This paper utilized available data from the CADE system to centralize CDSRs for 102 MDAPs and create 443 unique factor values across numerous commodity types, development types, contract types, and services for each WBS element. The factor approach to cost estimating hinges upon the availability of meaningful data, and the centralization of over 50 years of MDAP data allows cost estimators in the DoD to efficiently access and refine a broad dataset to create estimates for their respective programs. Furthermore, the dataset provides a starting point to perform the iterative process of refining the data and practical analysis to arrive at a defensible estimate. The importance of efficient and effective cost estimating in the acquisition workforce within the DoD is evident based on budgetary restrictions, political climate, and many other factors. Thus, the importance of this research lies in the analyst’s ability to expand their estimating toolset by quickly and efficiently accessing a compilation of hundreds of relevant data points that previously existed in hundreds of distinct locations.

Bibliography

Air Force Cost Analysis Agency (AFCAA). (2007). *Cost risk and uncertainty analysis handbook*. Washington, DC: AFCAA.

Baloi, D., & Price, A. D. F. (2003). Modeling global risk factors affecting construction cost performance. *International Journal of Project Management, 21*(4), 261–269.

Department of the Air Force. (2007). *Air Force cost analysis handbook*. Washington, DC: Department of the Air Force.

Department of the Air Force. (2018). *Financial management: US Air Force cost and planning factors* (AFI 65-503). Washington, DC: HQ USAF.

DoD. (2005). *Work breakdown structures for defense materiel items* (MIL-HDBK-881A). Washington, DC: Author.

DoD. (2007). *Cost and software data reporting (CSDR) manual* (DoD 5000.04–M–1). Washington, DC: Author.

DoD. (2011). *Work breakdown structures for defense materiel items* (MIL-STD-881C). Washington, DC: Author.

Elfaki, A. O., Alatawi, S., & Abushandi, E. (2014). Using intelligent techniques in construction project cost estimation: 10-year survey. *Advances in Civil Engineering, 2014,* 107926, 1–11. http://dx.doi.org/10.1155/2014/107926

Flyvbjerg, B., Holm, M. S., Buhl, S. (2002). Cost underestimation in public works projects: Error or lie? *Journal of the American Planning Association, 68*(3), 279–295.

GAO. (2009). *Cost estimating and assessment guide* (GAO-09-3SP). Washington, DC: U.S. Government Printing Office.

Mislick, G. K., & Nussbaum, D. A. (2015). *Cost estimation: Methods and tools.* Hoboken, NJ: John Wiley & Sons.

Naval Center for Cost Analysis (NCCA). (2018). NCCA reference information. Retrieved July 19, 2018, from https://www.ncca.navy.mil/references.cfm

Otte, J. (2015). *Factor study September 2015*. Wright-Patterson Air Force Base, OH: Air Force Lifecycle Management Center Research Group.

PAX. (2018, April 24). DoD area cost factors (ACF). *Programming Administration and Execution System Newsletter (PAX), 3.2.1*, 1–19.

Riquelme, P., & Serpell, A. (2013). Adding qualitative context factors to analogy estimating of construction projects. *Procedia—Social and Behavioral Sciences, 74*, 190–202. Retrieved from https://ac.els-cdn.com/S1877042813004667/1-s2.0-S1877042813004667-main.pdf?_tid=ab4ccb38-10d8-4a26-8874-5d00d10fc6a1&acdnat=1531852501_491c91b1a45e3da1f9fc5d5141441d48

Wren, D. (1998). *Avionics support cost element factors*. Wright-Patterson Air Force Base, OH: Aeronautical Systems Center.

Acknowledgments

This material is based upon work supported by the Acquisition Research Program at the Naval Postgraduate School. The views expressed in written materials or publications, and/or made by speakers, moderators, and presenters, do not necessarily reflect the official policies of the Department of Defense nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.

A Reduced Form Model of Cost Growth of Major Defense Acquisition Programs

David L. McNicol—joined the Department of Defense (DoD) in 1982. From 1988 until 2002, he was a Deputy Director of Program Analysis and Evaluation (PA&E). Earlier, Dr. McNicol taught at the University of Pennsylvania and the California Institute of Technology. He holds a BA in economics from Harvard and an MS in management and a PhD in economics and finance from MIT. Employed at the Institute for Defense Analyses (IDA) since his retirement from the DoD, he became director of the Cost Analysis and Research Division in 2006. Still at IDA, McNicol stepped down in 2012 to return to his previous role as a Research Staff Member. [dmcnicol@ida.org]

Abstract

This paper considers a problem posed implicitly by comparing a basic assumption typically used in quantitative analyses of cost growth of major defense acquisition programs (MDAPs) with that used in David L. McNicol, *Acquisition Policy, Cost Growth, and Cancellations of Major Defense Acquisition Programs*, IDA Report R-8396, September 2018 (hereafter *Acquisition Policy*). An analysis in the traditional mold mainly uses program characteristics (such as the maturity of key technologies) to explain cost growth. *Acquisition Policy* instead uses a categorical variable for funding climate, categorical variables marking major changes in acquisition policy, and measures of program duration. At first glance, these two approaches seem to adopt radically different theories of the causes of cost growth in MDAPs. In fact, they do not. The paper demonstrates this by deriving the model of *Acquisition Policy* from a more complete model in which the traditional model is a structural equation. In terms of the more complete model, that of *Acquisition Policy* is the reduced form representation of the traditional model.

Introduction

I wrote this paper to answer a question I was asked after a presentation I made to the 15th Annual Naval Postgraduate School (NPS) Acquisition Research Symposium (McNicol, 2018b). My presentation concerned a model that related cost growth on major defense acquisition programs (MDAPs) to changes in acquisition policy, funding climate (which is a proxy for the intensity of competition among MDAPs for funding at Milestone [MS] B), and measures of program duration. The question asked was: Why did you not include as explanatory variables any program characteristics—for example, the degree of concurrency between Engineering and Manufacturing Development (EMD) and procurement? I had anticipated this question and had an answer, but it was clear to me as I gave it that my answer was inadequate. On reflection, I concluded that I had not fully thought through the issue. This paper is the remedy offered.

The following section identifies relevant previous studies, and states in a general form a model in which program characteristics are used to explain cost growth. The Funding Climate-Policy Model section briefly sketches the model of McNicol (2018b). The next section, A More Complete Model and the Reduced Form Relationship, uses a more complete high-level model of cost growth on MDAPs to show that the Program Funding Climate-Policy Model is drawn from an underlying theory consistent with analyses that employ program characteristics to explain cost growth. The final section states my revised and, I hope much improved, answer to the question that led to this paper.

The Program Characteristics Model of Cost Growth

During the 1960s and 1970s, several papers produced by the RAND Corporation considered whether the changes in acquisition policy and process made by Robert McNamara in the early 1960s, and further changes introduced by David Packard in July 1969, had improved MDAP outcomes, particularly with regard to cost growth (Dews et al., 1979; Perry, 1975; Perry et al., 1971; Perry et al., 1969). Perry et al. (1971) contained a section on the causes of cost growth; it attributed cost growth to three factors—technical uncertainty, scope change, and cost estimating error (see also Srull, 1998, Chapter 1).[1] These papers did not suggest that program characteristics were causes of cost growth.

It is unclear when or why thinking shifted, but the idea that cost growth of MDAPs can to a large extent be explained by program characteristics seems to have entered the literature through two studies that appeared in the early 1990s—Tyson et al. (1992) and Drezner et al. (1993). Tyson et al. (1992) is cast as an evaluation of the effects on cost growth and schedule slips of six policy changes, each of which is embodied in a program characteristic. Drezner et al. (1993) states that they are using program characteristics (and also changes in DoD-level funding) to explain cost growth. Each of these studies took the program characteristics they considered as a given. In contrast, Tyson, Harmon, and Utech (1994) attempted to derive from the analysis the set of characteristics that are most important for cost growth. Lorell, Payne, and Mehta (2017) provided a clear and compelling study with a broadly similar intent.

The following are representative examples of the program characteristics linked to cost growth by studies that have appeared since the mid-1990s:

- Realism of the MS B EMD schedule
- The maturity of the technologies employed
- Whether the program involved a full-scale prototype prior to MS B
- The degree of concurrency between development and production
- The appropriateness of the contract type used
- Whether program requirements are technically feasible and remain stable
- Funding stability
- Whether the MS B cost estimate is realistic
- Test assets in the program
- The amount of computer code that will be reused (i.e., taken from a legacy system)
- The overhead rate

Until fairly recently, no two studies adopted (or derived) the same set of program characteristics as the main causes of cost growth. Since 2010, however, root cause analyses sponsored by the Office of Program Assessment and Root Cause Analyses

[1] Perry and his colleagues thought that cost estimating errors were by a considerable margin the least important of the three sources of cost growth they identified. In 1970, David Packard, then Deputy Secretary of Defense, identified unrealistically optimistic MS B cost estimates as the main source of cost growth.

(PARCA) have gone a considerable distance towards establishing a taxonomy of the proximate causes of cost growth. Examples of root cause analyses sponsored by PARCA are Blickstein et al. (2011), Blickstein et al. (2012), and Diehl, Gould, and Lo (2012).

Generally accepted conclusions have been reached on only a few topics. For example, there is a consensus that Total Package Procurement (TPP) and Fixed Price Development contracts are associated with high cost growth. There also seems to be a consensus that average cost growth of MDAPs has not increased or decreased across the past half century. More generally, there is widespread, although not unanimous, agreement that unrealistic assumptions embedded in MS B baselines are the largest source of cost growth. On many topics, however, the cost growth literature leaves considerable room for debate.

The term *Program Characteristics Model* is used here as a label for the idea that program characteristics are a major source of cost growth.[2] A general representation of this idea is the following:

$$\mathrm{Ch_{PAUC}} = \frac{\mathrm{h}(\mathrm{x_1}, \dots, \mathrm{x_n})}{\mathrm{C_{MSB}}}.$$

The dependent variable ($\mathrm{Ch_{PAUC}}$) is the percentage change in Program Acquisition Unit Cost (PAUC), which is defined in the following section. C_{MSB} is the MS B estimate of acquisition cost. The numerator [$h(x_1,\dots,x_n)$] is the actual cost the model projects based on program characteristics $X = (x_1,\dots,x_n)$.

In practice, studies of the extent to which program characteristics influenced cost growth require large amounts of often difficult to acquire data. They inevitably are imperfect because of gaps in the data and analytical issues. These problems are not important in the context of this paper, however.[3] What we need is simply the representation of the idea that program characteristics drive a significant part of cost growth.

The Funding Climate-Policy Model

The model adopted by McNicol (2018b) is drawn from McNicol (2018a) (hereafter referred to as *Acquisition Policy*), which accepts the premise of the Program Characteristics Model: the proximate causes of a large portion of cost in MDAPs lie in unrealistic assumptions embedded in the MS B baseline. Viewed from this standpoint, the model in *Acquisition Policy* is placed one step upstream from previous cost growth studies. It

[2] Not all studies that fall under the heading “cost growth” were concerned with the links between program characteristics and cost growth. Some were concerned with the more modest problem of describing the main features of cost growth. Is cost growth markedly higher in one of the Services? Has cost growth increased over time? Others examined whether changes in acquisition policy and process led to improvements in MDAP outcomes over time, e.g., lower cost growth.

[3] It is worth noting that estimating a Program Characteristics Model statistically is effectively impossible because of the huge data requirements. Drezner et al. (1993) seems to be the only example of an attempt to do so. That study, however, used only six program characteristics and a measure of budget growth, and did not report the estimated equation. McNicol (2004) might be regarded as another example; however, it uses a hybrid of the Program Characteristics Model and what the study calls the Speeding Model of cost growth, plus several other variables inspired by cost analysis considerations.

examines root cause (i.e., causes of causes), where the Program Characteristics Model is concerned with the proximate causes of cost growth.

For convenience, I will refer to the model developed in *Acquisition Policy* as the Funding Climate-Policy Model. This model is a version of the Speeding Model of cost growth introduced in McNicol (2004). The Speeding Model posits that all "drivers"—program managers (PMs) and the components who "own" the MDAPs—have some propensity to speed, that is, to adopt unrealistic assumptions about the performance of the system or unrealistic assumptions that reduce its apparent cost and/or EMD schedule. The other side of the Speeding Model is external constraints on speeding—speed limits backed up by the police, fines, and the courts. In the context of major system acquisition, that primarily means acquisition policy and Office of the Secretary of Defense (OSD)–level oversight.

From this line of thought, the Funding Climate-Policy Model distills two sets of variables to characterize cost growth due to Errors of Inception: funding climate (the surrogate for the intensity of competition for acquisition funds at MS B), and changes in acquisition policy. The model takes an ad hoc approach to the other two main sources of cost growth—Errors of Execution and Program Changes.

The equation estimated is

$$Ch_{PAUCi} = a_0 + a_1Climate_i + a_2DSARC_i + a_3PCDSARC_i + a_4DAB_i + a_5AR_i + a_6T_{boomi} + a_7T_{busti} + e_i.$$

PAUC is acquisition cost (the sum of EMD and procurement cost) divided by the number of fully configured units acquired. PAUC growth is computed by comparing the MS B baseline value of PAUC—which can be thought of as a goal or a prediction—to the actual PAUC reported in the final Selected Acquisition Report (SAR) for the program. Both the MS B baseline and the final value[4] of PAUC are stated in program base year dollars. The actual value is restated on the basis of the MS B baseline quantity by moving up or down the cost progress curve as appropriate. The ratio of the MS B baseline value of PAUC to the quantity-adjusted actual value is an estimate of what PAUC growth would have been had the MS B baseline quantity been acquired.

Table 4 defines the categorical variables used in the study. The first of the acquisition policy bins (McNamara-Clifford) does not appear explicitly in the model because it is used as the reference category. *Acquisition Policy* identifies the factors used to establish the break points between bust and boom climates and the acquisition policy bins.

[4] For a program that is still underway, the most recent estimate (as reported in the SAR) of the final value was used.

Table 4. Categorical Variables of the Funding Climate-Policy Model

Variable	Short Name	Period (Fiscal Years)
Climate	bust climates	1965–1982, 1987–2002
	boom climates	1983–1987, 2003–2008
McNamara-Clifford	McNamara-Clifford	1965–1969
Defense System Acquisition Review Council	DSARC	1970–1982
Post-Carlucci DSARC	PC DSARC	1983–1989
Defense Acquisition Board	DAB	1990–1993, 2001–2009
Acquisition Reform	AR	1994–2000

Finally, T_{boomi} and T_{busti} are the numbers of years the i^{th} program spent in boom and bust years, respectively. These provide a rough and ready way to capture PAUC growth due to Errors of Execution and Program Changes, which the model cannot distinguish. The term e_i is a random variable that is assumed to have a constant mean and variance.

Table presents the estimated parameter values and their associated p-values.[5] It is difficult to find anything to complain about in these results. Each of the estimated coefficients has the expected sign, and the estimated magnitudes are reasonable. All the coefficients from the Speeding Model are statistically significant at the 1% level or less, which is the most striking feature of the results. The estimated coefficient of T_{boom} is significant at about the 2% level. The estimated coefficient for T_{bust} is insignificant, which is consistent with prior expectations. About 26% of the variation in PAUC growth over the sample is accounted for by the model, which, for panel data without any lagged variables, is remarkably high. Of course, results like this never “prove” a model to be valid but, as in this case, they may fail to reject it.

[5] The p-value in this instance provides a test of the statistical significance of the estimated coefficients of the regression equation. The null hypothesis is that the true value of the coefficient is zero. The p-value then is the probability of obtaining the estimate from a sample if its true value is zero. For example, the estimated coefficient of T_{boom} is 3.8%/yr. and the associate p-value is 0.021. This means that the odds of observing a coefficient for T_{boom} as large as 3.8%/yr. are about 2 in 100. Consequently, we reject the null hypothesis that the true value of the coefficient is zero. The border for statistical significance is generally set at 5% or sometimes 10%. Thus, an estimate coefficient with a p-value of 0.05 or less would be said to be “statistically significantly different from zero at the 5% level.”

Table 5. Estimated Coefficients and p-Values for a Model That Includes the Effects of Post-MS B Funding Climate and Duration †

	Coefficients	p-value
Intercept	73.1%***	< 0.001
Errors of Inception—Intensity of Competition for Funds		
Funding Climate	-28.7%***	0.009
Errors of Inception—Acquisition Policy		
DSARC	-56.7%***	< 0.001
PC DSARC	-50.3%***	0.001
DAB	-59.5%***	< 0.001
AR	-80.2%***	< 0.001
Errors of Execution and Program Changes		
T_{boom}	3.8%/yr**	0.021
T_{bust}	0.59%/yr	0.515

***Statistically significant at less than the 1% level

** Statistically significant at less than the 5% level

R-Squared = 0.26, F = 7.02 (P < 0.001), N= 149. Estimated using OLS. Four programs that passed through two boom periods and the three mid-1980s MDAPs acquired using TPP-like contracts were omitted. Wald's test for the equality of the estimated coefficients of the categorical variables for acquisition policy periods with the Bonferroni correction yields F= 1.43, p = 0.0.946.

† Adapted from Table 16, page 38, of *Acquisition Policy*

A More Complete Model and the Reduced Form Relationship

The Program Characteristics Model and the Funding Climate-Policy Model were developed to answer different questions, so their differences may be tolerable. Still, it is awkward to have two models that address related questions, have the same dependent variable (cost growth of MDAPs), and different explanatory variables, a situation that cries out for an explanation. This section extracts one from a more complete high-level model of cost growth on MDAPs. A couple of pages are required to sketch the model. After that is done, the argument can be completed very quickly.

The first relationship in the model describes the results of the PM's judgment of what the cost for the program must be for it to be funded. (The PM's superiors may be involved, but that fact is not important for the purposes of this exercise.) Note that the context is a specific program coming up for MS B review in a particular Program Objective Memorandum (POM) cycle. No assumption about how the PM makes their judgment is required, but it is worth noting that the problem is intrinsically one of constrained optimization. The PM wants the highest cost that will provide a solid chance that the program will be funded. The cost that the PM decides is needed is denoted by C^*. The variable marking the intensity of competition is denoted by W, and the restrictions (that is, acquisition policies) that the PM believes must be observed are denoted by R. We assume that

$$C^* = f(W, R)\,, \qquad (1)$$

and assume further that C* decreases as competition for funds (W) becomes more intense. The question of how to measure W is set aside here.

The next part of the model represents the choice of program characteristics to be changed as necessary to get apparent cost down to C*. Recall that the relevant program characteristics are denoted by $X = (x_1,\ldots,x_n)$. By departing from realistic values for any of the program characteristics, the PM creates some risk for the program. The problem is to select values for the program characteristics that reduce cost to C* at minimum risk, but this may not be a hard problem. Suppose that the PM, with the assistance of the program office staff, can assign the x_is to risk bands—say low, medium, and high. The assignments would be made in terms of the perceived risks to the program of departing from the realistic or best practice value of the characteristic. The reduction in the apparent cost of the program also is associated with each of the x_is. The least risk solution is then found by reading down the list until the cost estimate for the program cost reaches C*.

This approach assumes that the risk of setting one program feature at an unrealistic level is not affected by the choices made for other program features. For example, the assumption would be that the risk entailed by procuring an unrealistically small quantity of initial spares is not affected by assuming an unrealistically early start for operational testing. In fact, these two elements of risk are intertwined.[6] Consequently, most would model this decision as a problem of picking the values of program characteristics X to minimize some measure of program risk M given C, and subject to the interactions of program risks and restriction R imposed on the program office. The solution to this problem is a relationship (known as an "efficient frontier") between the risk measure M and the cost achieved, C. Each point on the curve of this relationship is associated with a particular set of program characteristics that achieves the cost C at minimum risk, given R and the interdependencies of program risks. In this simple model,[7] the bundle of program characteristics accepted is that which gives C*. This solution can be written:

$$x_i^* = g_i(C^*, R),\ i = 1,\ldots, n\ , \qquad (2)$$

where x_i^* is the value of the i[th] program characteristic given by the solution to the optimization problem.

Many will balk at the apparent implication that PMs and their staffs literally solve the optimization problem sketched above. Especially during the early years of a program, the volume of work that a program office must do and the rapid pace of events are such that spending the time required to optimize any one decision probably would be, well, not optimal for a program office. Consequently, on many decisions that must be made, PMs and program offices live in the land of "good enough." Of course, PMs and program office staffs are professional and knowledgeable and work at solving problems, so the solutions they develop generally are sound. The essential assumption, however, is not that the decisions made are near-optimal. Rather, in the context of the model, the essential assumption is that the PM's decisions on program characteristics respond to changes in external events—

---

[6] The simple approach may still be viable if the interdependencies are few enough and simple enough.

[7] The first of many refinements of the model would replace Equation (1) with a relationship that characterizes the PM's willingness to trade off two categories of risks: (a) risk that the program will not be funded because it is perceived as being unaffordable; and (b) latent risks to the program created by adoption of unrealistic values for some program characteristics.

especially the intensity of competition for funds and restrictions that they must observe—in the same way as the optimal solution. The statistical analysis does not "see" departures from optimality. What it sees are the responses to changes in relevant external conditions—funding climate and acquisition policies—and the model is rejected if these responses depart significantly from what it predicts.

The final relationship in the model is just the Program Characteristics Model of cost growth:

$$C_f - C_{MSB} = h(x_1,\ldots,x_n)\ . \qquad (3)$$

C_f is what the acquisition cost of the program finally turned out to be (adjusted to the MS B quantity and stated in program base-year dollars), excluding cost growth due to Errors of Execution and Program Changes. C_{MSB} is the acquisition cost projected at MS B (which always tacitly assumes no Errors of Execution or Program Changes). Note also that the x_is are consistent with the MS B baseline and the CARD (which ideally are consistent with one another).

The remainder of the argument is just a matter of substituting Equation (1) into Equation (2) and the result into Equation (3). The first of these steps yields:

$$x_i^* = g_i(f(W,R),\ R) \equiv G_i(W,R), \quad i = 1,\ldots,\ n\ . \qquad (4)$$

Note that $G_i(W,R)$ is simply a renaming adopted to cut down on notational clutter. Substitution of Equation (4) into the Program Characteristics Model, Equation (3), gives:

$$C_f - C_{MSB} = h(G_1(W,R),\ \ldots\ ,\ G_n(W,\ R)) \equiv H(W,R)\ . \qquad (5a)$$

Now divide by the MS B PAUC and use the original form of the Program Characteristics Model (Equation (3)):

$$\frac{C_f - C_{MSB}}{C_{MSB}} = Ch_{PAUC} = \frac{h(x_1,\ldots,x_n)}{C_{MSB}} = \frac{H(W,R)}{C_{MSB}}\ . \qquad (5b)$$

It is obvious in Equation (5b) that the Funding Climate-Policy Model is simply the reduced form of the Program Characteristics Model.

An elaboration of the model sketched here—for example, incorporation of uncertainty—is unlikely to change the result just stated. What could change it is incorporation into the model of an additional feature of the acquisition process. Thinking along these lines, the first place to look would be the OSD-level acquisition review process. The policy variable R was defined as the set of acquisition policies that the PM believes must be observed. That is, the model tacitly assumes that the PM knows with certainty which policy restrictions require compliance. A PM, of course, never knows for sure how rigorously the applicable policies will be enforced.

A surface read of this observation is that it points to an elaboration of the model. The real point, however, is that the Funding Climate-Policy Model largely is irrelevant unless there is significant porosity in the OSD-level oversight process. Within the logic of the model, more intense competition for funds is an incentive for PMs to propose programs that have unrealistically optimistic and unreasonably risky characteristics. But to the extent that OSD-level reviews lead to the rejection of unrealistic elements in proposed programs, the programs that emerge from the review are realistic and are risky only within the bounds of existing policy. All of the x_i^* of Equation (4) are then determined by policy restrictions (R), and funding climate has no effect. For funding climate to have an effect, it must be that a PM can, at some risk, violate some of the rules some of the time or that there are major gaps in the rules.

Conclusion

To repeat, the question that motivated this paper was: Why did I not include program characteristics as variables in the Funding Climate-Policy Model? One answer is that doing so would contradict the specifications of both models. A better answer is that including program characteristics in a Funding Climate-Policy model would answer no question. The studies that employ the Program Characteristics Model of cost growth are intended to provide good housekeeping guidance on how to structure MDAPs. The Funding Climate-Policy Model is concerned with explaining why the DoD does not always follow the dictates of policy and prudence in laying out major acquisition programs. Including program characteristics in a Funding Climate-Policy model would produce results that, regardless of the estimated test statistics, cannot be interpreted in terms of the question either model is intended to address.

Some might respond that it is reasonable to test the Funding Climate-Policy Model against alternatives. There is of course nothing wrong with doing that. It is not accomplished, however, by simply including one or more program characteristics in a Funding Climate-Policy Model. It would be necessary to formulate carefully the two models to be compared, and design a good way to distinguish them. Certainly the most direct—and probably the best—way to test the model sketched here is to estimate Equation (4) for several program characteristics over an interval of time long enough to include both bust and boom funding climates and some significant changes in acquisition policies.

References

Blickstein, I., Boito, M., Drezner, J., Dryden, J., Horn, K., Kallimani, J. G., … Wong, C. (2011). *Root cause analyses of Nunn-McCurdy breaches*, Vol. 1 (MG-1171/1-OSD). Retrieved from https://www.rand.org/pubs/monographs/MG1171z1.html

Blickstein, I., Drezner, J., Libicki, M. C., McInnis, B., McKernan, M., Nemfakos, C., … Wong, C. (2012). *Root cause analyses of Nunn-McCurdy breaches,* Vol. 2 (MG-1171/2-OSD). Retrieved from https://www.rand.org/pubs/monographs/MG1171z2.html

Dews, E., Smith, G. K., Barbour, A., Harris, E., & Hesse, M. (1979). *Acquisition policy effectiveness: Department of Defense experience in the 1970s* (R-2516-DR&E). Retrieved from https://www.rand.org/pubs/reports/R2516.html

Diehl, R., Gould, B., & Lo, T-N. (2012). *Root causes of Nunn-McCurdy breaches—A survey of PARCA root causes analyses, 2010–2011* (IDA Paper P-4911). Alexandria, VA: Institute for Defense Analyses.

Drezner, J. A., Jarvaise, J. M., Hess, R. W., Norton D. M., & Hough, P. G. (1993). *An analysis of weapon system cost growth* (MR-291-AF). Retrieved from https://www.rand.org/pubs/monograph_reports/MR291.html

Lorell, M. A., Payne, L. A., & Mehta, K. R. (2017). *Program characteristics that contribute to cost growth: A comparison of Air Force major defense acquisition programs* (RR 1761). doi:10.7249/RR1761

McNicol, D. L. (2018a). *Acquisition policy, cost growth, and cancellations of major defense acquisition programs* (IDA Report R-8396). Alexandria, VA: Institute for Defense Analyses.

McNicol, D. L. (2004). *Cost growth in major weapon procurement programs*. Alexandria, VA: Institute for Defense Analyses.

McNicol, D. L. (2018b). Further evidence on program duration and unit cost growth. *Proceedings of the 15th Annual Acquisition Research Symposium* (Vol. I; SYM-AM-18-032; pp. 89–105). Retrieved from https://calhoun.nps.edu/handle/10945/58481

Perry, R. (1975). *Reforms in system acquisition* (P-5482). Retrieved from https://www.rand.org/pubs/papers/P5482.html

Perry, R. L., DiSalvo, D., Hall, G. R., Harman, A. J., Levenson, G. S., Smith, G. K., & Stucker, J. P. (1969). *Systems acquisition experience* (RM-6072-PR). Retrieved from https://www.rand.org/pubs/research_memoranda/RM6072.html

Perry, R., Smith, G. K., Harman, A. J., & Henrichsen, S. (1971). *System acquisition strategies* (R-733-PR/ARPA). Retrieved from https://www.rand.org/pubs/reports/R0733.html

Srull, D. (Ed.). (1998). *The Cost Analysis Improvement Group: A history*. McLean, VA: LMI.

Tyson, K. W., Harmon, B. R., & Utech, D. M. (1994). *Understanding cost and schedule growth in acquisition programs* (IDA Paper P-2967). Alexandria, VA: Institute for Defense Analyses.

Tyson, K. W., Om, N. I., Gogerty, D. C., & Nelson, J. R. (1992). *The effects of management initiatives on the costs and schedules of defense acquisition programs*, *Vol. I: Main report* (IDA Paper P-2722). Alexandria, VA: Institute for Defense Analyses.

Acknowledgments

This paper was reviewed by David A. Sparrow, Matthew S. Goldberg, and Gregory A. Davis, all of the Institute for Defense Analyses. The author is grateful for their comments on this paper and their continuing support of the project of which this paper is a part.

Eliciting Expert Opinion in Acquisition Cost and Schedule Estimating

Major Gregory E. Brown, USAF—serves as an Instructor of Financial Management and Data Science for the Air Force Institute of Technology (AFIT) School of Systems and Logistics. As an instructor, he develops and teaches continuing education courses on acquisition cost analysis, statistics, and data analytics. Prior to becoming an AFIT instructor, Maj Brown worked in multiple cost-analysis positions across the Air Force Lifecycle Management Center, most recently serving as the Cost Analysis Chief for the Fighters and Bombers Directorate. He holds a BA in economics and BS in business-finance from Colorado State University, a graduate certificate in applied statistics from Penn State University, and an MS in cost analysis from the Air Force Institute of Technology. [gregory.brown@AFIT.edu]

Abstract

Despite the emphasis on data and analytics in acquisition cost and schedule estimating, many estimating situations still require eliciting expert opinion from a subject matter expert. This is problematic, as a 2007 RAND report concludes that there is no standard model for seeking expert input for acquisition estimates. Per the report, the DoD's "elicitation methodologies are largely ad hoc, in that they are seldom based on or derived from references to the elicitation literature" (Galway, 2007). In this paper, a popular and commonly cited elicitation model—the Stanford Research Institute (SRI) elicitation model—is presented and adapted to the cost and schedule estimating process. It is posited that the consistent application of a formal model would reduce expert biases and improve the acquisition community's risk and uncertainty analyses. This paper also provides the results of an original meta-analysis of published experiments that examine expert elicitation for business and engineering problems. The data reveals that experts are overconfident and struggle to identify the true range of outcomes for both business and engineering problems. However, using a structured elicitation model, training the expert prior to the elicitation, and providing the expert with feedback are shown to decrease expert overconfidence.

Introduction

Even with ongoing efforts to improve acquisition databases, sometimes the historical data we need for cost and schedule estimates is simply unavailable. In other instances, historical data is available but requires adjustment to account for radical changes in a technology or manufacturing process (Kitchenham et al., 2002). In these instances, analysts may turn to the opinion of experts, using an interview process known as elicitation. Through elicitation, it is possible to tap into the knowledge and experience of engineers, logisticians, and programmers. Utilizing expert elicitation carries risks, however. Without proper guidance, experts may fall victim to cognitive biases, resulting in predictions that are both inaccurate and overconfident.

Academic research has long since recognized the problem of expert biases and began designing elicitation protocols to guard against them in the 1960s. Experiments have shown that by following a structured elicitation framework and providing feedback, the quality of elicitation may be improved. Regrettably, no standard elicitation model currently exists across the DoD cost and schedule estimating community. Instead, RAND notes that the DoD's "elicitation methodologies are largely ad hoc, in that they are seldom based on or derived from references to the elicitation literature." Simply put—analysts are learning to conduct elicitation by trial and error, rather than being guided by a structured model. To compound this problem, RAND notes that elicitations are poorly documented within DoD cost estimates, resulting in elicited estimates that cannot be reviewed and reexamined after the initial estimate is completed (Galway, 2007).

Based on these revelations, it is evident that we—the cost analysis community—need a change in attitude towards elicitation and elicitation training. We wouldn't expect an analyst to construct a parametric model without first receiving education on linear regression methods, so why is the expectation for elicitation any different? We must stop viewing elicitation as an ad hoc art, and instead adopt a more structured, scientific process. Rather than novice analysts learning elicitation through improvisation, we should educate new analysts using those methods that are validated by decades of research from the fields of psychology, behavioral economics, decision analysis, and Bayesian statistics.

To initiate this change, this author proposes a five-step model first introduced by decision analysis researchers at the Stanford Research Institute (SRI). The model is provided within this research paper in a subsequent section. As a caution, it would be unwise to jump directly to the elicitation model without first understanding the fundamentals that shape the model. Thus, this research paper is divided into five sequential sections, with each section building upon knowledge from the prior. To begin, this paper provides a definition for expert elicitation and background on the advantages and disadvantages of elicitation. The next section describes common expert biases so that the cost analyst may better learn to recognize them. Then this author examines whether the Joint Agency Cost Schedule Risk and Uncertainty Handbook (JA CSRUH) heuristic of treating expert intervals as encompassing only 70% of uncertainty is accurate and defendable. Moreover, strategies are provided for controlling expert overconfidence. Next, the paper outlines the SRI elicitation methodology that will serve to further reduce expert bias while also promoting improved documentation of elicitations. The next section introduces methods for adapting the SRI model to elicitations with multiple experts. Finally, the last section provides a summary and recommendations for future elicitation research and change efforts.

When taken in aggregate, it is the author's hope that the research cited within this paper will help promote a change in attitudes toward expert elicitation in the community, so that expert predictions are treated in a similar manner to traditional data and statistical models. Rather than blindly accepting or rejecting expert predictions, analysts should instead adopt a more structured approach that will allow the expert's opinion to be afforded the same level of review and validation that we would demand for any other cost or schedule model.

Background

Defining "Expert" and "Elicitation"

So, what is an expert? An *expert* is defined as an individual who has mastered the specialized skills or bodies of knowledge relevant to a particular subject. While the expert doesn't know everything about a subject, it is expected that his or her prediction on a problem is more likely to be correct than that of the public at large. However, being an expert in one field does not make an individual better qualified in unrelated fields. Research finds that experts—even at the PhD level—are no better at predicting outcomes in fields unrelated to their expertise than the general population (McKenzie et al., 2008; Nichols, 2017). Thus, we would not expect a chemical or nuclear engineer to be particularly skilled at estimating lines of code if he or she had never worked in software engineering. Finding the right expert for a given estimate is paramount.

Conversely, *elicit* means “to call forth or draw out (as information or a response)” (“Elicit,” 2017). Thus, in an expert elicitation, the cost estimator is asking the expert call forth information from his or her area of expertise. The term *elicit* and *elicitation* are preferred to synonyms such as *interview*, as *elicit* is the most commonly favored term in academic research, beginning with usage in early Bayesian statistics research (e.g., Winkler, 1967) as well as the earliest RAND Delphi Method study (i.e., Brown, 1968).

Relevant Cost Estimating Methodologies

Many cost estimating methodologies are cited across DoD literature: extrapolation from actuals, parametric, analogy, bottom-up engineering, and expert opinion. In this author’s experience, at least three of these methodologies will typically require elicitation of an expert. In instances where the cost estimator has no historical data to leverage, the estimator may directly elicit an *expert opinion* from the expert. Alternatively, when only a few historical data points are available—insufficient for a parametric model—the estimator may seek the expert’s help in identifying the best *analogy*, to which the expert may subsequently apply a scaling or complexity factor (AFCAH, 2008). *Parametric* models require elicitation as well, as the inputs to the parametric model are seldom known with certainty at the beginning of a project. For example, when employing a parametric software cost estimating model, variable inputs such as source lines of code (SLOC) and code re-use are typically estimated by a technical expert (Jorgenson, 2007). Because these inputs are uncertain during the early phases of a program, applying relevant elicitation protocols can improve the accuracy of the expert’s elicited inputs to the model.

Why Elicit an Expert’s Opinion?

The DoD recently introduced the Cost Assessment Data Enterprise (CADE), an online database intended to significantly increase the cost analyst’s access to cost, schedule, and technical acquisition data. As the CADE platform matures and access to data improves, less time will be spent gathering data for cost and schedule estimates, allowing for the adoption of more innovative and accurate modeling techniques (Watern, 2016). Given the availability of CADE, is the elicitation of experts still relevant to cost estimating?

Elicitation will likely remain relevant for several reasons. Firstly, CADE is focused on collecting data for Acquisition Category (ACAT) Level 1 programs, currently defined by 10 U.S.C. 2430 as having a development budget greater than $480 million or procurement budget greater than $2.79 billion. As a result, smaller programs are not well represented in CADE, and when they are, they will typically have less collected data to leverage for future estimates. Thus, constructing a parametric model for a minor systems modification may not always be feasible. Secondly, changes in technology mean that available historical data may not always be relevant to the current estimating task and may require adjustment by the expert (Kitchenham et al., 2002). For example, a parametric schedule model based on software using a waterfall strategy may require recalibration by an expert before it is used to estimate a project with an agile strategy. Thirdly, even when sufficient analogous data is available to establish a parametric model, meta-analysis suggests that in certain scenarios, experts are just as accurate as parametric models in estimating outcomes. These scenarios are explored next.

Accuracy of Experts Compared to Models

Do models always outperform expert predictions within cost estimates? Jorgenson (2007) reviewed 16 software cost estimating studies that directly compare the accuracy of formal parametric models with that of experts. After aggregating the studies, Jorgenson found that the average accuracy of the expert-derived estimates was higher than for the model-provided estimates for 10 of the 16 studies. Jorgenson’s finding contradicts the belief—held by some—that parametric models will always outperform the expert in the context of cost estimating.

When and why might one method outperform the other? Sanders and Ritzman (1991) theorized that models are superior for prediction when using data which is “stable.” As an example from the medical field, a meta-analysis of 136 individual medical studies find that statistical models are more likely to correctly diagnose a medical condition than medical experts (Grove et al., 2000). One particularly notable study is Nashef et al. (1999), who proposed the European System for Cardiac Operative Risk Evaluation (EuroSCORE) model, in which age, gender, pulmonary disease status, and a multitude of cardiac lab values are able to more accurately predict the likelihood of post-operative death or complication than an experienced heart surgeon. In this setting, the data is stable, in that the human body is not significantly changing or evolving. The same predictive relationships built on the initial sample of patient demographics and lab values are expected to remain valid over time. Almost two decades later, the EuroSCORE model remains in use in the United States, Europe, and Japan, and the model continues to be validated by using the populations of different countries (e.g., Shen et al., 2018).

Conversely, Sanders and Ritzman (1991) theorized that experts are superior at prediction in unstable, changing conditions, as one might face when estimating the cost for a new technology with changing cost drivers. In describing the results of his software meta-analysis, Jorgenson (2007) concluded that in research and development, “the technology, the types of software produced, and the production methods, change frequently” (p. 460). This lack of stability, combined with small data sets, makes it difficult to build an accurate statistical model that is not overfitted to the historical data. Unlike the model, the expert is not limited to considering only a few variables, but instead may utilize decades of cumulative experience as well as all available context about the program being estimated. Thus, in some cost estimating scenarios, the expert may have the advantage “in that they typically possess more information and are more flexible in how the information (or lack of information) is processed.”

Given that neither parametric models nor experts are always the best, some researchers suggest employing an “ensemble” approach, whereby output from the parametric model and output from the expert are combined (i.e., averaged) to reduce estimating error. Over time, theory states that an ensembled estimate will have greater accuracy than either the parametric model or expert alone, assuming that both estimates are unbiased and capture different information. As evidence that ensemble models can be successfully employed in cost estimating, Li et al. (2008) tested the application of Optimal Linear Combining (OLC) to software cost estimating, with the estimates from a parametric software model and expert each weighted based on their expected accuracy. On average, the OLC ensemble increases the accuracy of software cost estimates when compared to the parametric model or expert alone.

Problems with Utilizing Elicitation in Cost Estimating

Despite evidence from Jorgenson (2007) that experts can be as accurate—or more accurate—than models in cost estimating, many decision-makers remain hesitant to make decisions using elicited estimates without traditional data. Why is this?

- Recognizing that experts are prone to both motivational and cognitive biases, the decision-maker may view all elicitations as biased or inaccurate.
- Due to overconfidence, experts have historically been overly precise when estimating prediction intervals, leading the decision-maker to accept more uncertainty and risk than he or she was briefed. However, it is currently not

known how overconfident experts are (i.e., what percentage of uncertainty is actually captured by the expert assisting with DoD cost estimates?) Based on a 1976 study, the JA CSRUH recommends treating the expert's input as only capturing 70% of outcomes.

- No standard elicitation methodology exists within the DoD. As a result, RAND observes that elicitations are often poorly documented within cost and schedule estimates, and it is difficult for more senior reviewers or cost agencies to validate the inputs provided by the expert (Galway, 2007). For decision-makers, the credibility of an elicitation is only as good as the documentation and justification surrounding the expert's estimate.

However, each of these potential problems may be overcome by the research presented in this paper. By following a consistent protocol—such as the SRI elicitation model—and documenting the rationale behind the elicited estimate, it is possible to regain the trust of the decision-maker.

Expert Biases

In the previous section, the definition of expert elicitation was given, and evidence was provided that expert opinion can be as accurate as data-driven models. However, employing expert opinion can be problematic, as experts can be affected by biases—both intentional (i.e., motivational bias) and unintentional (e.g., optimism). These biases may drive the expert to be less accurate within a given estimate. Moreover, biases may cause the expert to consistently underestimate or overestimate a requirement across multiple estimates, resulting in entire product portfolios that are underfunded or overfunded. Although not an exhaustive list, six cognitive biases commonly encountered when eliciting an expert's opinion are summarized.

Motivational Bias

Motivational bias is driven by the expert's desire to influence the decision to his or her own benefit. As notional examples of motivational bias, a program manager may benefit from understating the cost of a new effort in order to secure initial funding or milestone approval. Conversely, an engineer may benefit from overstating the costs for a proposed technical solution that he or she does not support.

Optimism

Individuals assess that they are better than others and less likely than others to experience negative events or outcomes. These individuals will focus on what can "go right" in a project, while believing that nothing could "go wrong." Often, this is driven by a false sense of control over events. As a result, experts who succumb to optimism bias will consistently underestimate task completion times and costs, even when presented with the information that the vast majority of similar tasks have run over both schedule and budget (Flyvbjerg, 2011).

Availability

Availability says that individuals are more likely to recall information that is either recent or made the most significant impression on that individual, while ignoring less impressionable information. As a consequence, experts may base their elicitation on the information that is easiest to recall, rather than taking into account the full range of observations and experience.

Anchoring

Anchoring states that individuals will often use readily available information (e.g., an analogous project) as the initial basis for an estimate, before making further adjustments to account for differences (Spetzler & Stael von Holstein, 1975). However, research experiments have shown that on average, individuals tend to make insufficient adjustments to the initial basis, resulting in the response being “anchored” to the basis (Kahneman & Tversky, 1974). As a result, when using an analogy as basis for an estimate, the expert may fail to fully adjust for the change in complexity between the historical analogy and the new effort.

Unstated Assumptions

The unstated assumptions mode of judgment says that individuals will naturally condition their estimate on unstated assumptions. As a consequence, the elicited distribution will often ignore events which the expert believes he or she is not responsible for considering.

> For example, a cost estimate might be made with the implicit assumptions that the base design will not change. However, the same person, when questioned about the likelihood of the base design’s changing, might think such a possibility very likely. (Boyd & Regulinski, 1979)

While assumptions are necessary for a cost estimate, it is important that these assumptions are clearly verbalized by the expert, documented by the cost analyst, and later briefed to the decision-maker.

Overconfidence

Overconfidence states that individuals will believe their point estimate to be a better and more reliable estimate than it really is. As a consequence, the expert will generally understate the uncertainty about a quantity, resulting in a prediction interval that is smaller than it should be.

Eliciting Uncertainty From Experts

Background

When using parametric-based cost estimating relationships, uncertainty about a prediction is calculated in the form of a prediction interval. Assuming the assumptions necessary for linear regression are met (e.g., equal variance of errors and normal distribution of errors at each value of the predictor), there is generally no need to adjust the prediction interval, as it is unbiased. For example, given that a future observation comes from the same population as the sample used to build the parametric model, a 95% prediction interval is expected to contain the future observation 95% of the time.

However, when relying on expert opinion as the basis for an estimate, the analyst faces the added challenge of generating a prediction interval with the help of the expert. Due to overconfidence, the expert’s elicited interval will generally be smaller than the interval representing the true state of the expert’s knowledge. Overconfidence can be lessened using techniques that drive the expert to consider the full range of outcomes, but experiments show that these techniques will not completely resolve overconfidence. Moreover, due to “unknown unknowns,” it is often not feasible for the expert to imagine all possible outcomes. It is therefore necessary to account for additional uncertainty when modeling inputs elicited from an expert.

The JA CSRUH recognizes this problem and recommends treating the expert's interval as encompassing only 70% of the range of uncertainty. The handbook's 70% heuristic is derived from Capen (1974), who concluded that experts rarely account for more than 60%, and never account for more than 70% of the possible range of outcomes. Capen arrived at his conclusion by surveying 1,000 petroleum engineers who were asked to estimate prediction intervals for 10 generic, encyclopedia-type questions, such as "What is the area of Canada in square miles?"

Based on this author's experience, however, some program managers and decision-makers may question the validity of a heuristic which requires the application of additional uncertainty, increasing the cost or schedule of a program. In turn, the heuristic may be difficult for the analyst to defend, due to the research's age (over 40 years old) and the reality that Capen was asking the engineers to estimate intervals for encyclopedia problems, and not problems directly related to their area of engineering expertise. If the petroleum engineers had instead been asked to generate prediction intervals related to petroleum engineering, would they show less overconfidence and provide more realistic intervals? To help resolve this question and provide the analyst with relevant research to cite when defending their estimate, a meta-analysis is conducted.

Meta-Analysis of Expert-Elicited Intervals

To re-validate Capen's earlier findings, this author searches for additional research that utilizes surveys to assess the accuracy of expert prediction intervals. To best align with the problem types typically encountered in acquisition cost estimating, the search query is designed to capture studies in which either business or engineering experts provide intervals directly related to their field or industry. Studies involving undergraduate students are excluded, but studies involving graduate students (e.g., Goldenson & Stoddard, 2013) are included if it is documented that the graduate students have prior industry experience in their field. To increase the meta-analysis's relevance to cost and schedule estimating, only studies involving the prediction of continuous ranges are included; studies in which experts are asked to estimate probabilities of discrete events (e.g., True or False) are excluded.

After applying the inclusion and exclusion criteria, a total of five studies encompassing 17 total surveys and 21,000 individual predictions are identified. The following are descriptions of the studies:

- Russo and Schoemaker (1992) asked corporate business managers to provide prediction intervals for technical questions related to the managers' own firm and industry (11 aggregated surveys; 7,660 total predictions).
- McKenzie et al. (2008) asked information technology (IT) professionals to provide prediction intervals for IT industry questions (one aggregated survey; 1,720 total predictions).
- Ben-David et al. (2010) asked Chief Financial Officers of major companies to provide prediction intervals for S&P 500 market returns for the following year; the survey is repeated annually over a nine-year period (one aggregated survey; 11,600 total predictions).
- Goldenson and Stoddard (2013) asked graduate students with industry experience to provide prediction intervals for source lines of code (SLOC) and effort in person-years for previously completed software projects based on a description of the software requirements, team size and programming language (three aggregated surveys; 290 total predictions).

- Bar-Yosef and Venezia (2014) asked experienced brokerage analysts to use accounting data for a company to provide prediction intervals for net income, earnings per-share, and share price (one aggregated survey; 30 total predictions).

The results are plotted in Figure 1, and weighted averages for each available increment of confidence are summarized in Table 1. The full results are provided in Appendix A. In viewing Figure 1, the requested prediction interval for the survey is plotted on the x-axis, while the percentage of experts whose response contained the true answer is plotted on the y-axis. The diagonal dashed line represents the calibration line, where a well-calibrated group of experts should fall. For example, if a 50% prediction interval is requested, then approximately 50% of the experts should provide an interval that contained the true response. Additionally, a simple linear regression model is fit to the data and represented by a black line, while a weighted regression model—with survey sample size assigned as the weight—is fit and plotted as a solid blue line. The models are designed to measure whether the confidence level requested from the expert impacts the percentage of correct responses, and whether the confidence level requested from the expert impacts the degree of observed overconfidence.

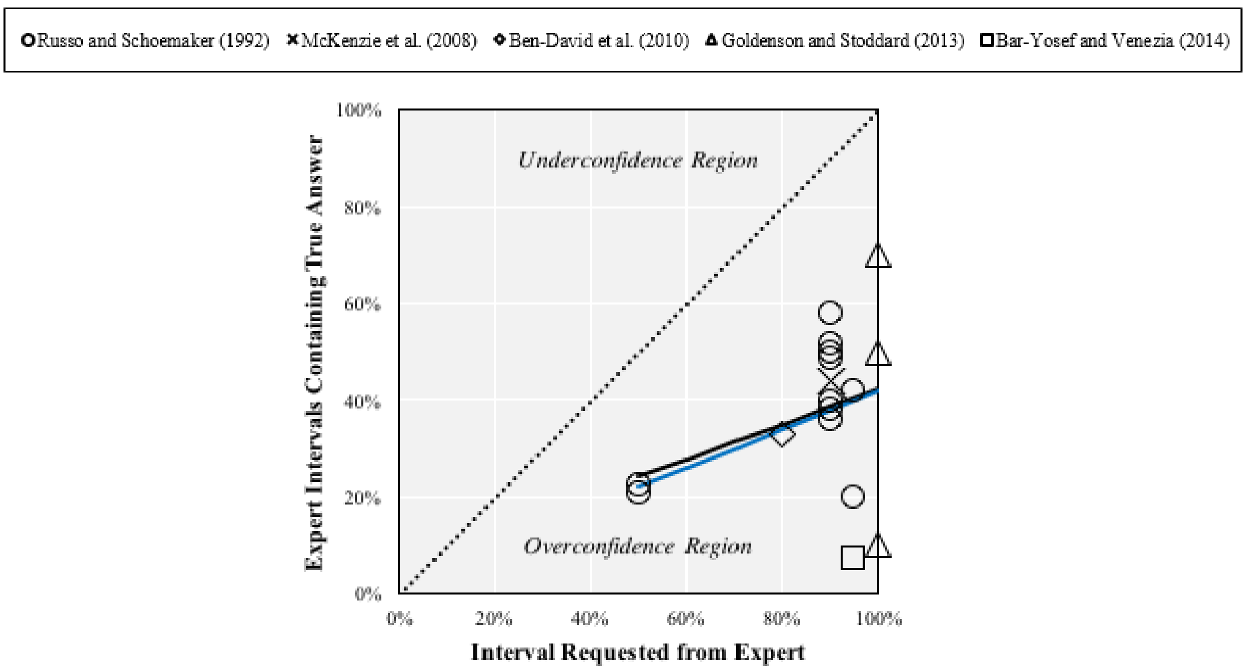


Figure 1. **Meta-Analysis Results**

After assessing Figure 1, it is observed that for each of the 17 aggregated surveys, the percentage of experts with correct intervals falls below the calibration line. This indicates that experts are overconfident, on average. Moreover, it is observed that at no time do more than 70% of experts in a given survey predict the true response, even when asked to provide a 100% interval, as in Goldenson and Stoddard (2013). Thus, the meta-analysis validates Capen's finding that experts never identify more than 70% of the possible range of outcomes.

Digging deeper, both the simple linear regression and weighted regression models are examined, with the goal of determining if a linear relationship exists between the interval requested and the percentage of experts with the correct answer. Although a positive slope is calculated for the simple linear regression model, it is not statistically significant (p-value = 0.22). The weighted regression model is statistically significant (p-value = 0.04); however, upon closer inspection, this is the result of the model overfitting (i.e., passing directly through) a few influential data points with relatively larger sample sizes. Due to the limited number of surveys, transformations and non-linear methods are not considered. Thus, when considering results from across studies, no regression-based conclusions are drawn about the relationship between the requested expert interval and actual expert interval.

Table 1. Weighted Average of Correct Expert Estimates at Confidence-Level Intervals

Interval Confidence-level Requested from Expert	50%	80%	90%	95%	100%
Expert Intervals Containing Truth (Weighted Avg.)	21%	33%	45%	31%	36%
Aggregated Sample Size	1,600	11,600	5,200	2,610	290
Total Number of Surveys	2	1	8	3	3

However, in viewing the data more subjectively, the experts do appear to drift further away from the calibration line as the confidence-level of the requested interval increases. As indicated in Table 1, at higher confidence levels, a 30% adjustment would be insufficient to capture the true range of uncertainty in all but one case. This raises the question: should the analyst instead be adding greater than 30% uncertainty to the expert's elicited range? This author's experience indicates that an even greater adjustment would face resistance in the acquisition community. Such an extreme adjustment to the expert's cost or schedule estimate may not be palatable to decision-makers and risks offending the expert who provided the elicited input. Thus, strategies for naturally reducing the expert's overconfidence are explored next.

Strategies for Decreasing Overconfidence—Feedback and Formatting

The meta-analysis result raises the concern that assuming that the expert captures 70% of the true responses is itself optimistic. This phenomenon—when observed in other studies—has led researchers to conclude that most individuals have a poor understanding of statistics and prediction intervals (Kahneman, 2011). However, research shows that it is possible to improve the calibration of expert's prediction intervals by focusing on two areas: feedback and elicitation formatting.

As an example of feedback, this author examines Goldenson and Stoddard (2013), a software estimating study previously cited in the meta-analysis. The study consisted of three rounds. During the first round, only 10% of experts provided a prediction interval containing the true requirement, despite being asked to provide a 100% prediction interval. Following round one, feedback was provided to experts that they were overconfident. In turn, 50% of experts identified the true response in round two, and after post-round two feedback, 70% identified the true response in round 3. Thus, providing ongoing feedback to experts does appear to significantly improve the calibration of expert's prediction intervals. Simply making the expert aware that they suffer from overconfidence results in more accurate prediction intervals.

Other elicitation studies show that focusing on the formatting of the elicitation questions can decrease overconfidence. In experiments involving non-experts, the following strategies have been shown to decrease overconfidence:

- Ask for the high and low outcomes prior to asking for the most likely. This format has been shown to decrease overconfidence related to the anchoring bias (e.g., Soll & Klayman, 2004; Speirs-Bridge et al., 2010).
- Allow the expert to select the confidence level of the interval they would like to provide (e.g., 70%). This format decreases overconfidence compared to confidence levels that are pre-specified by the analyst (Teigen & Jorgenson, 2005).
- When appropriate, allow experts to provide intervals based on lower confidence levels. For example, individual experts providing answers corresponding to lower-confidence prediction intervals (e.g., 50% confidence) show less overconfidence than those providing higher-confidence prediction intervals (e.g., 90% confident) (Teigen & Jorgenson, 2005). Using standard formulas from the JA CSRUH's Table 2-8, a lower-confidence interval may be adjusted outward by the analyst to capture 100% confidence.
- For experts who struggle to conceptualize the prediction interval concept, manually walk the expert through the creation of the prediction interval. For example, simplify the prediction interval concept by asking the expert "could the requirement exceed 1,000 hours?" or "what is the probability that the requirement exceeds 1,000 hours?" (Teigen & Jorgenson, 2005).
- Finally, after recording the initial prediction interval, verify the expert's answer by asking the expert to consider why they may be wrong. For example, ask the expert to consider that the true requirement is greater than the upper bound of the prediction interval. Ask: what are a few reasons this could be? What assumptions or considerations may be wrong? Given these erroneous assumptions, was the initial estimate too low? Lastly, ask the expert if they wish to revise the upper bound. Herzog and Hertwig (2009) followed a similar line of questioning in their experiments and discovered that simply questioning the individual's initial conclusion prompts the individual to consider knowledge that was previously overlooked or assumed to be true when constructing the prediction interval.

Elicitation Model for a Single Expert

Many unique interview models have been proposed for gathering expert opinion. However, this author elects to utilize the *Stanford Research Institute (SRI) Elicitation Process* model. The model originated with Spetzler and Stael von Holstein (1975), decision analysis researchers from Stanford University. The SRI model is cited in numerous subsequent research efforts and is regarded by Morgan et al. (1990) as the most popular and influential elicitation model. As presented in Figure 2, the SRI model consists of five sequential phases: motivating, structuring, conditioning, encoding, and verifying.

In viewing Figure 2, it should be stressed that Spetzler and Stael von Holstein did not consider documentation as a separate step. Instead, documentation should be a continual process that takes place throughout each phase of the elicitation model. When writing documentation, the analyst should always strive to communicate the assumptions, rationale, and analogies used to estimate an outcome. This will serve two purposes. Firstly, this will help the analyst better explain the estimate to decision-makers in the event that the expert is not present. Secondly, it is rare that the same expert will be available for consultation throughout a program's life-cycle. Thus, recording the reasoning for the expert's estimate will be useful if a different expert is assigned to the program in the future and the estimate is revisited.

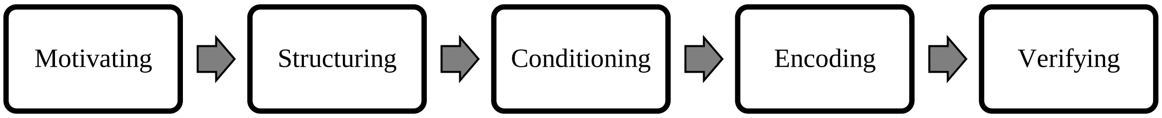


Figure 2. **SRI Elicitation Model**
(Spetzler & Stael von Holstein, 1975)

Phase 1: Motivating

Unlike data or models, which are at the control of the cost analyst, elicitation requires human interaction. Thus, the motivating stage is intended to introduce the expert to the purpose of the elicitation and establish rapport (Spetzler & Stael von Holstein, 1975, p. 352). Although this phase may seem superfluous, it should not be disregarded. Galway (2007) noted that in the DoD, elicitations are often rushed due to the time constraints and shortages of available experts. Galway's assessment matches with this author's own experience. For most experts, assisting with a cost or schedule estimate is a secondary duty, which takes them away from their primary duty. To motivate the expert and generate "buy in," it is imperative that the expert is made aware of the purpose of their inputs early in the process. Whether the end goal is a major milestone brief or the budgeting of future funding, how will the expert's input help the integrated product team succeed?

This author asserts a secondary focus of the motivating stage should be the education of the analyst, with a focus on achieving a basic technical understanding of the requirement to be estimated, thereby limiting the risk of *hypocognition*. Wu and Dunning (2017) wrote that hypocognition exists when one operates outside his or her conceptual landscape. Hypocognition is problematic as it can limit the ability of two individuals to exchange information. It is therefore imperative that the analyst makes an effort to develop a basic understanding of the requirement or technology to be estimated, as well as its associated terminology, as it will later direct the course of the conversation between the analyst and expert. Without knowing said terminology, the individual will have difficulty receiving and communicating the ideas advanced by the expert during the elicitation process. Moreover, individuals cannot use concepts they do not have or understand to explain phenomena (Levy, 1973). During the briefing stage of the estimate, the analyst—if not conceptually familiar with the requirement or technology being estimated—will risk misrepresenting or distorting the basis of the expert's elicited estimate. Thus, the analyst should ensure they have a working knowledge of the requirement or technology prior to entering the later phases of the elicitation.

Phase 2: Structuring

The purpose of the structuring phase is to define the uncertain quantity (or quantities) that requires expert input. If necessary, the "structure should be expanded as necessary so that the subject does not have to model the problem further before making each judgement" (Spetzler & Stael von Holstein, 1975, p. 353). The typical human can only hold about seven separate pieces of distinct information in their working memory at a time (Miller, 1956). Thus, by simplifying the problem into components or subcomponents, the cost analyst reduces the number of factors that the expert must mentally model when providing an estimate.

The extent of structuring—or breaking down the effort into distinct pieces—should be driven by the basis of the expert's knowledge and any supporting data. Just as estimating at too high a work breakdown structure (WBS) level may reduce precision, attempting to estimate at too low of a WBS level may also insert unnecessary bias or error. Moreover, the cost analyst should generally avoid structuring an estimate so that the expert must provide his or her answer in dollars. Instead, the cost analyst should ask the expert what unit of measure he or she prefers, so that the expert does not have to go through the mental exercise of converting units. The cost analyst may discover that the expert prefers to estimate in hours, full-time equivalents, or SLOC, rather than in dollars.

Phase 3: Conditioning

The conditioning phase strives to head off biases and condition the expert to "think fundamentally about his judgement" (Spetzler & Stael von Holstein, 1975 p. 353). In their original paper, Spetzler and Stael von Holstein provided limited detail on the conditioning phase, aside from suggesting that the analyst ask the expert to describe how they go about assigning probabilities. However, later researchers have supplemented the SRI model by recommending that the expert is provided training on cognitive biases and probability distributions. Other authors have suggested putting the expert through a series of warm-up exercises to allow for calibration (Morgan et al., 1990). Based on this author's experience, however, setting aside sufficient time for warm-up exercises or demonstrations may not be feasible for DoD cost and schedule estimates, particularly for routine estimates or smaller programs.

However, this author has found success with utilizing the conditioning phase to introduce the concept of the probability distribution and provide a preview of what to expect in the encoding phase. This author's conditioning protocol consists of three steps. First, the analyst should begin every session with a brief overview of the triangular or beta-PERT distribution, making mention that the expert will be asked to separately provide a low, high, and most likely estimate. Second, explain to the expert that he or she will later be asked to quantify the confidence interval percentage captured by the given low and high estimate. Finally, emphasize that as the expert considers the low, high, and most likely estimate, he or she should verbalize the assumptions and conditions that would lead to the provided outcome.

Phase 4: Encoding

As introduced previously, research suggests we should first ask the expert for the low and high values to avoid the anchoring effect. It is therefore recommended that the analyst first ask the expert for the "low" value, followed by the "high" value. After each value is provided, ensure the expert is verbalizing both the assumptions and events that could lead to that value. If the expert is not being clear—or their response is not understood—the cost analyst should continue to ask "why?" until the analyst is confident that the estimate is justified and can be explained. Only after obtaining the extremes—and their justification—should the cost analyst ask for the "most likely" value. Once again, the most likely value should be accompanied by a rationale that would lead to the most likely outcome.

After recording the range and most likely, ask the expert how confident they are in their low and high values. What percentage of outcomes will fall within the provided range or what percentage will fall outside the range? Alternatively, the analyst can ask the expert what percentage of outcomes will be greater than the high and what percentage will be less than the low. Then use the provided low and high probabilities to calculate the absolute minimum and maximum via the equation in the JA CSRUH's Table 2-8. Going forward, take care to distinguish the expert-provided low and high from the calculated minimum and maximum, which have been expanded to capture 100% confidence.

Phase 5: Verifying

Finally, having recorded the high, low, and most likely values and the rational for those values, the analyst should verify that the expert's judgement has remained consistent. For example, it is possible that during the course of discussion, the expert recalled additional information that may lead him or her to adjust the high and low bounds provided earlier in the elicitation session.

Begin by showing the expert the minimum and maximum values that were calculated when the provided low and high were adjusted to encompass 100% of confidence. Ask the expert whether there are conceivable scenarios that could lead to a value outside of the calculated minimum and maximum bounds. If the expert concedes that a scenario exists, ask if he or she would like to adjust the absolute minimum and maximum. If the expert would prefer not to adjust the bounds, then ask for the probability of an outcome outside of the minimum and maximum bounds and use the provided probability to further adjust the bounds outward. If necessary, repeat this step until the expert is satisfied with the calculated minimum and maximum.

When the expert is satisfied with the absolute minimum and maximum interval, the initial elicitation is completed. At this point, consider applying an additional 30% uncertainty to account for bias and "unknown unknowns." Even when following the SRI protocol, biases will exist in the expert's answer, as it is not possible for the expert to consider all possible outcomes and scenarios, especially those that fall outside of their area of expertise. When determining whether additional uncertainty is warranted, consider that parametric cost models tend to have coefficient of variation (CV) values between 0.15 and 0.35 (Naval Center for Cost Analysis, 2015, p. 32). Also consider which acquisition milestone the expert's estimate is supporting. Elicitations occurring early in the development production life-cycle will have greater uncertainty than those occurring later. Carney (2013) found that at the program level, estimates have CVs of 0.41 to 0.74 at Milestone A, 0.45 to 0.61 at Milestone B, and 0.23 to 0.32 at Milestone C. Thus, an expert-derived prediction interval with a coefficient of variation lower than 0.25 is likely overconfident and could benefit from the inclusion of additional uncertainty.

Adapting the Model for Multiple Experts

Although the SRI model is initially presented as a model for eliciting opinion from a single expert, the encoding phase of the SRI model may be easily adapted to allow for multiple experts. Prior to beginning the elicitation, the analyst must decide how much interaction to allow between experts. Although interaction is beneficial in allowing for the exchange of ideas and assumptions, it also contributes to *groupthink*, a cognitive bias not yet introduced. In groupthink, the position of a few experts leads the entire group to a consensus that does not represent the individual experts' private opinion.

What does literature recommend for controlling groupthink? At one end of the interaction spectrum, an analyst may allow the group to openly discuss the low value, high value, most likely value, and corresponding confidence level without any structure until a consensus is reached for each. However, groupthink is most likely to occur in this scenario. Nearer the other end of the spectrum is the Delphi method, in which experts exchange anonymous written inputs and justification until a consensus is reached. By allowing for anonymous inputs, the risk of groupthink is significantly reduced. Even more extreme, some authors propose not allowing any communication between experts, and instead taking a simple average or weighted average of the experts' individual inputs. In this case, each elicitation is conducted separately with no interaction between the experts, thereby preventing any groupthink.

A more moderate method is the nominal group technique, in which each expert is forced to establish an estimate prior to interacting with other group members. After the initial estimate, each expert presents his or her position, the rationale behind the position, and all relevant assumptions. After the initial positions are revealed, differences between individual estimates are openly discussed in an attempt to reach a consensus. If consensus is not reached, then the divergent position of each expert is averaged (Gustafson et al., 1973). A similar technique, known as "Planning Poker," is commonly utilized for estimating requirements in the agile software community. In planning poker, software engineers assign difficulty to a user story (i.e., software requirement) by simultaneously revealing a poker card with the difficulty number. Each expert then defends the rationale behind his or her initial poker card estimate. After discussions among the experts, subsequent poker rounds—in which each expert may update his or her estimate—are conducted until the estimates converge to the same assigned difficulty value (Cohn, 2012). For both of these methods, the most important step is that each expert is forced to commit to an initial estimate prior to discussions beginning. Committing to an initial estimate prior to group discussion prevents the group from anchoring off the first expert's response and promotes the open exchange of assumptions and ideas across the group.

Conclusion & Recommendations

Research shows that expert opinion can be as accurate as parametric-based methods (Jorgensen, 2007). However, when not properly guided, experts are prone to biases, and liable to be overconfident when estimating the uncertainty surrounding an estimate. To achieve more consistent and accurate results with elicitation, this author advocates for the adoption of a structured elicitation model, such as the SRI model. The model integrates methods—such as first asking the expert for the low and high estimate—that are shown in experiments to naturally reduce human overconfidence. Moreover, adopting a common model will promote more rigorous documentation, so that the expert's opinion may be subjected to the same extent of senior analyst review and verification as traditional data. To further improve the quality of elicitation, two additional recommendations are provided.

As the first recommendation, this author advocates increasing formal training and education on elicitation for new cost and schedule analysts. Most analysts today have learned elicitation via a trial-and-error or ad hoc approach, and not a formal education program. We would not expect an analyst to construct a parametric model without first learning the fundamentals of learning regression, so why are our expectations any different for elicitation? Every new analyst should be given at least a rudimentary introduction to elicitation and provided with a common framework. To assist in guiding new analysts, a checklist that this author has personally used is included in Appendix C. Readers are encouraged to further adapt and improve the checklist for their own uses.

As a second recommendation, further research is needed to determine the accuracy and CV of elicitation-based cost and schedule estimates. The Air Force Life-Cycle Management Center's Cost Research Branch is currently undergoing a project that will examine historical cost growth within program office estimates, with cost estimating methodology being a recorded factor (S. Valentine, personal communication, March 13, 2019). Such a study will prove valuable, as it will establish a statistically-based CV range for expert-elicited estimates, so that we will no longer have to strictly rely on rules-of-thumb, such as adding an additional 30% to expert-estimated uncertainty intervals.

In closing, this is an exciting time to be a cost or schedule analyst. CADE and other acquisition databases are increasing our access to data, allowing for more innovative estimates and analyses. However, even with more data, situations will continue to arise in which we must seek the opinion of an expert. By inserting more structure and discipline into the elicitation process, we can avoid the most common pitfalls of expert opinion, thereby leading to more accurate and reliable cost and schedule estimates.

References

Air Force Cost Analysis Agency. (2007). *Air Force cost risk and uncertainty analysis handbook (CRUH)*. Washington, DC: Author.

Bar-Yosef, S., & Venezia, I. (2014). An experimental study of overconfidence in accounting numbers predictions. *International Journal of Economic Sciences*, *3*(1), 78–89.

Ben-David, I., Graham, J., & Harvey, C. (2010). *Managerial miscalibration*. Retrieved from https://faculty.fuqua.duke.edu/~charvey/Research/Working_Papers/W105_Managerial_miscalibration.pdf

Boyd, D. W., & Regulinski, S. G. (1979). *Characterizing uncertainty in technology cost and performance* (Report 1114). Los Altos, CA: Decision Focus.

Brown, B. B. (1968). *Delphi process: A methodology used for the elicitation of opinions of experts*. Santa Monica, CA: RAND Corporation.

Burgman, M. A. (2016). *Trusting judgements: How to get the best out of experts*. Cambridge, England: Cambridge University Press.

Capen, E. (1976). The difficulty of assessing uncertainty (includes associated papers 6422, 6423, 6424, and 6425). *Journal of Petroleum Technology*, *28*(8), 843–850. doi:10.2118/5579-pa

Carney, S. T. (2013). *Investigation into risk and uncertainty: Identifying coefficient of variation benchmarks for Air Force ACAT I programs* (Master's thesis). Air Force Institute of Technology.

Cohn, M. (2012). *Agile estimating and planning*. Upper Saddle River, NJ: Prentice Hall PTR.

Elicit. (n.d.). In *Merriam-Webster*. Retrieved October 19, 2017, from http://www.merriam-webster.com/dictionary/elicit.

Flyvbjerg, B. (2011). Over budget, over time, over and over again. *Oxford Handbooks Online*.

Galway, L. A. (2007). *Subjective probability distribution elicitation in cost risk analysis: A review*. Santa Monica, CA: RAND Corporation.

Goldenson, D. R., & Stoddard, R.W. (2013). *Quantifying uncertainty in expert judgment: Initial results*. Pittsburgh, PA: Carnegie Mellon Software Engineering Institute.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*(1), 19–30. doi:10.1037/1040-3590.12.1.19

Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind. *Psychological Science, 20*(2), 231–237.

Jorgensen, M. (2007). Forecasting of software development work effort: Evidence on expert judgement and formal models. *International Journal of Forecasting, 23*(3), 449–462. doi:10.1016/j.ijforecast.2007.05.008

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus, and Giroux.

Kitchenham, B., Pfleeger, S. L., McColl, B., & Eagan, B. (2002). An empirical study of maintenance and development estimation accuracy. *Journal of Systems and Software*, *64*(1), 57−77.

Levy, R. I. (1973). *Tahitians: Mind and experience in the Society Islands*. Chicago, IL: University of Chicago Press.

Li, Q., Wang, Q., Yang, Y., & Li, M. (2008). Reducing biases in individual software effort estimations. In *Proceedings of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM 08)*. doi:10.1145/1414004.1414041

McKenzie, C. R. M., Liersch, M. J. & Yaniv, I. (2008) Overconfidence in interval estimates: What does expertise buy you? *Organizational Behavior and Human Decision Process*, *107*, 179–191.

Morgan, M., Henrion, M., & Small, M. (1990). *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge, England: Cambridge University Press.

Nashef, S., Roques, F., Michel, P., Gauducheau, E., Lemeshow, S., & Salamon, R. (1999). European system for cardiac operative risk evaluation (EuroSCORE). *European Journal of Cardio-Thoracic Surgery, 16*(1), 9–13.

Naval Center for Cost Analysis. (2015). *Joint agency cost schedule risk and uncertainty handbook (JA CSRUH)*. Washington, DC: Author.

Nichols, T. (2017, March & April). How America lost faith in expertise. *Foreign Affairs*, *96*(2), 60–73.

Patterson, R. E., Pierce, B. J., Bell, H. H., & Klein, G. (2010). Implicit learning, tacit knowledge, expertise development, and naturalistic decision making. *Journal of Cognitive Engineering and Decision Making, 4*(4), 289–303.

Russo, J. E., & Schoemaker, P. (1992). Managing overconfidence. *Sloan Management Review*, *33*(2), 7–17.

Sanders, N. R., & Ritzman, L. P. (1991). On knowing when to switch from quantitative to judgemental forecasts. *International Journal of Operations & Production Management, 11*(6), 27–37. doi:10.1108/01443579110005523

Shen, L., Chen, X., Gu, J., & Xue, S. (2018). Validation of EuroSCORE II in Chinese patients undergoing coronary artery bypass surgery. *The Heart Surgery Forum, 21*(1), 036.

Soll, J., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 299–314.

Speirs-Bridge, A., Fidler, F., McBride, M., Flander, L., Cumming, G., & Burgman, M. (2010). Reducing overconfidence in the interval judgements of experts. *Risk Analysis*, *30*(3), 512–523.

Spetzler, C. S., & Staël Von Holstein, C. S. E. (1975). Probability encoding in decision analysis. *Management Science, 22*(3), 340–358.

Teigen, K. H., & Jorgensen, M. (2005). When 90% confidence intervals are 50% certain: On the credibility of credible intervals. *Applied Cognitive Psychology*, *19*(4), 455–475.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Watern, K. (2016). Cost Assessment Data Enterprise (CADE). *Air Force Comptroller Magazine*, *49*(1), 25.

Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects. *Journal of Experimental Psychology: General*, *125*(4), 387–402.

Winkler, R. L. (1967). The assessment of prior distributions in Bayesian analysis. *Journal of the American Statistical Association*, *62*(319), 776–800.

Appendix A: Meta-Analysis Results

Author (Year)	Interval Requested	Intervals Containing Truth	Number of Predictions	Expert	Estimating Task
Russo & Schoemaker (1992)					
Advertising 1	90%	39%	750	Corporate Managers	Advertising Industry Knowledge
Advertising 2	50%	22%	750	Corporate Managers	Advertising Industry Knowledge
Computers 1	95%	20%	1,290	Corporate Managers	Computer Industry Knowledge
Computers 2	95%	42%	1,290	Corporate Managers	Computer Firm Knowledge
Data Processing 1	90%	58%	252	Corporate Managers	Data Processing Industry Knowledge
Data Processing 2	90%	38%	261	Corporate Managers	General Business Knowledge
Money Management 1	90%	50%	480	Corporate Managers	Financial Industry Knowledge
Petroleum 1	90%	50%	850	Corporate Managers	Petroleum Industry and Firm Knowledge
Petroleum 2	50%	21%	850	Corporate Managers	Petroleum Industry and Firm Knowledge
Pharmaceutical 1	90%	51%	390	Corporate Managers	Pharmaceutical Firm Knowledge
Security Analysis 1	90%	36%	497	Corporate Managers	Security Industry Knowledge
McKenzie et al. (2008)	90%	44%	1,720	IT Professionals	IT Industry Knowledge
Ben-David et al. (2013)	80%	33%	11,600	Chief Financial Officers	Stock Market Return (S&P 500)
Goldenson & Stoddard (2013)					
Battery 1	100%	10%	140	Graduate Students	Software Development Effort
Battery 2	100%	50%	80	Graduate Students	Software Development Effort
Battery 3	100%	70%	70	Graduate Students	Software Development Effort
Bar-Yosef & Venezia (2014)	95%	7%	30	Brokerage Analysts	Financial Forecasts

Appendix B: Meta-Analysis Regression Model Outputs

Simple Linear Regression Model Output

```
Call:
lm(formula = Expert.Intervals.Containing.True.Answer ~ Interval.Requested)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.33484 -0.02667  0.00333  0.11333  0.27700 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.05973    0.24994   0.239    0.814
Interval.Requested  0.36328    0.28229   1.287    0.218

Residual standard error: 0.1683 on 15 degrees of freedom
Multiple R-squared:  0.09943,	Adjusted R-squared:  0.03939 
F-statistic: 1.656 on 1 and 15 DF,  p-value: 0.2176
```

Weighted Linear Regression Model Output

```
Call:
lm(formula = Expert.Intervals.Containing.True.Answer ~ Interval.Requested, 
    weights = Sample.Size)

Weighted Residuals:
    Min      1Q  Median      3Q     Max 
-7.1281 -0.4152  0.3115  2.5454  3.5386 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)  
(Intercept)         0.02155    0.14655   0.147   0.8851  
Interval.Requested  0.39675    0.17652   2.248   0.0401 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.819 on 15 degrees of freedom
Multiple R-squared:  0.2519,	Adjusted R-squared:  0.2021 
F-statistic: 5.052 on 1 and 15 DF,  p-value: 0.04007
```

R Code

```
# Import Data
Experts <-
data.frame(Interval.Requested=c(0.90,0.50,0.95,0.95,0.90,0.90,0.90,0.90,0.50,0.90,0.90,0.90,0.80,1.
00,1.00,1.00,0.95),
Expert.Intervals.Containing.True.Answer=c(0.39,0.22,0.20,0.42,0.58,0.38,0.50,0.50,0.21,0.51,0.36,0.
44,0.33,0.10,0.50,0.70,0.07),Sample.Size=c(750,750,1290,1290,252,261,480,850,850,390,497,1720,
11600,140,80,70,30))
attach(Experts)

# Compute simple linear regression and weighted linear regression models
simple.lm <- lm(Expert.Intervals.Containing.True.Answer~Interval.Requested)
weighted.lm <- lm(Expert.Intervals.Containing.True.Answer~Interval.Requested,
weight=Sample.Size)

# Outputs
summary(simple.lm)
summary(weighted.lm)
```

Appendix C: Elicitation Checklist for Cost and Schedule Analysis

Phase 1: Motivating

- ☐ *Analyst*: Familiarize yourself with the requirement needing expert elicitation. Begin formulating questions, and gather data that may be relevant to the expert.
- ☐ *Tell the Expert*: The purpose of this cost estimate is to estimate _____ in support of _____.

Phase 2: Structuring

- ☐ *Ask the Expert*: Should we break down the estimation of the requirement into smaller components?
- ☐ *Ask the Expert*: Would you feel most comfortable estimating the unknown quantity in person-hours, full-time equivalents (FTEs), SLOC, or another unit?
- ☐ *Ask the Expert*: What ground rules and assumptions are you making about the requirement being estimated?

Phase 3: Conditioning

- ☐ *Tell the Expert*: Today I will ask your assistance in constructing the triangular or Beta-PERT distribution that best represents your state of knowledge. I will begin by asking for your low outcome, followed by your high outcome, and lastly I will ask for the most likely outcome.
- ☐ *Tell the Expert*: Next, I will ask you for the probability (or likelihood) that the costs will be lesser/greater than your estimated low and high.
- ☐ *Tell the Expert*: When providing your response for low/high/most likely, please explain the assumptions, rationale, mental model, or analogy used to estimate each outcome. This will help us defend the estimate to decision-makers, and will be useful if the estimate is later revisited.

Phase 4: Encoding

- ☐ *Ask the Expert*: What is the low outcome? Why?
- ☐ *Ask the Expert*: What is the high outcome? Why?
- ☐ *Ask the Expert*: What is the most likely outcome? Why?
- ☐ *Ask the Expert*: Could an outcome be less than your low estimate? If so, what is the probability? What scenario could cause this to happen?
- ☐ *Ask the Expert*: Could an outcome be more than your high estimate? If so, what is the probability? What scenario could cause this to happen?

Phase 5: Verifying

- ☐ *Analyst*: If the expert responded that the interval had a confidence interval of less than 100%, adjust the expert's low and high using JA CSRUH Table 2-8 so that a 100% confidence level is reached. These values are the distribution's absolute min and max.

- ☐ *Ask the Expert*: Are there any conceivable scenarios that could cause the outcome to be less than the minimum? If so, what is the probability? What scenario could cause this to happen?
- ☐ *Ask the Expert*: Are there any conceivable scenarios that could cause the outcome to be more than the maximum? If so, what is the probability? What scenario could cause this to happen?
- ☐ *Ask the Expert*: Does the distribution require any further adjustments? Does it best represent your current state of knowledge?
- ☐ *Analyst*: The elicitation is complete. Thank the expert for their time. Compute the elicited distribution's coefficient of variation (CV), and consider adding 30% additional uncertainty if the CV is low (less than 0.25). Note that the expected CV will vary depending on the requirement being estimated and the milestone that the estimate is supporting.

Panel 26. Acquisition Reform: IT Systems, Foreign Military Sales, and Set-Based Design

Thursday, May 9, 2019	
3:45 p.m. – 5:00 p.m.	**Chair: Michael McGrath,** Consultant and Senior Technical Advisor, McGrath Analytics LLC ***Information Technology Acquisition Best Practices*** Terrence Leary, Erin Schultz, and Ginny Wydler, The MITRE Corporation ***Implementing Set-Based Design in DoD Acquisitions*** Norbert Doerry and Philip Koenig, NAVSEA ***The Effects of Exporting on Defense Acquisition Outcomes: A Quantitative Look at FMS Contracting—Preliminary Findings*** Samantha Cohen, Andrew Hunter, and Greg Sanders, Center for Strategic and International Studies

Michael McGrath—Dr. McGrath is an independent consultant. As a former Vice President at Analytic Services Inc. (ANSER), he led business operations in Systems and Operations Analysis. He previously served as the DASN (RDT&E), where he was a strong Navy proponent for improvements in technology transition, modeling and simulation, and test and evaluation. In prior positions, he served as: Vice President for Government Business at the Sarnoff Corporation (former RCA corporate lab); ADUSD for Dual Use and Commercial Programs in the Office of the Secretary of Defense (OSD), with responsibility for industrial base and commercial technology investment programs; Program Manager at the Defense Systems Research Projects Agency (DARPA), where he managed manufacturing technology programs; and Director of the DoD Computer-aided Acquisition and Logistics Support program, automating the interface between DoD and industry for technical data interchange and access. His early government career included positions in Logistics Management at Naval Air Systems Command and in Acquisition Management in OSD. He has served on Defense Science Board and National Academies studies, and is an active member of the National Defense Industrial Association (NDIA), the National Materials and Manufacturing Board, the Board on Army Science and Technology, and several university and not-for-profit advisory boards.

Dr. McGrath holds a BS in Space Science and Applied Physics and an MS in Aerospace Engineering from Catholic University, and a doctorate in Operations Research from George Washington University (where he also served as adjunct faculty).

Information Technology Acquisition Best Practices

Terry Leary—Principal Analyst at The MITRE Corporation, has more than 25 years of experience in program management, acquisition, and contracting. He is a former Air Force Program Manager for major systems acquisitions. He holds an MS in Astronautical Engineering from George Washington University and BS in Astronautical Engineering from the United States Air Force Academy. He is DAWIA Level II certified in both Program Management and Systems Planning Research and Development. [tleary@mitre.org]

Erin M. Schultz—Center for Acquisition Management and Sciences Technical Director at The MITRE Corporation, provides subject matter expertise in the areas of contracting, source selection, and acquisition strategies. She has over 25 years of contracting experience with both federal government and industry, specializing in information systems. [eschultz@mitre.org]

Virginia L. Wydler—Chief Acquisition Scientist at The MITRE Corporation, has over 25 years of experience in acquisition and contracting. She holds an MS in Acquisition Management, Naval Postgraduate School, and a BS in Business Administration, University of Maryland. She is a CPCM and Fellow with NCMA, and DAWIA Level II certified in Contract Management. [vwydler@mitre.org]

Executive Summary

There is a plethora of best practices and strategies for information technology (IT) systems implementation. This abundance of information can overwhelm government acquisition professionals when trying to select the most appropriate path to execute IT modernization and may lead to sub-optimum decisions and disappointing outcomes.

Mistakes can be very expensive, especially when shifting from legacy systems to modernized technology. Recent legislation addresses the cost of inefficiency:

> The Federal Government spends nearly 75 percent of its annual information technology funding on operating and maintaining existing legacy information technology systems. These systems can pose operational risks. … These systems also pose security risks, including the inability to use current security best practices … making these systems particularly vulnerable to malicious cyber activity. (The Modernizing Government Act of 2017)

Government and industry need to create a process for efficient and cost-effective approaches to manage IT acquisition. Unfortunately, very few approaches are grounded in practical and tactical applications. This report provides recommendations for best practices, frameworks, and models that will improve IT acquisitions and modernization efforts for network services. The report will allow federal program managers and acquisition professionals to implement IT acquisition strategies that appropriately fit their situation on the acquisition lifecycle spectrum.

Background

Twenty-first century computing power is very different from 20th century computing power. Although computers were introduced in the 20th century, it is in the 21st that computers have become a major instrument used in everyday life. Everyone from children to large corporations and governments rely on computers. Information from email to phone numbers to bank account numbers to missile launch codes are stored in computer systems and processed by software applications.

The combination of all the hardware, software, processes, and protocols that enable information sharing is now commonly referred to as *information technology* (IT). The dictionary defines IT as “the study or use of systems (especially computers and telecommunications) for storing, retrieving, and sending information.” The problem with the dictionary definition is that in today’s world, IT encompasses so many products and services that the use of the term *IT* is often susceptible to conflicting interpretations within programs and federal agencies or between industry and government.

Initially, the federal government viewed IT as a mission support function. As IT grew to be part of everyday life, agencies realized it needed to be more secure and robust. Unfortunately, the federal government has been playing catch up because of its inability to procure new IT and keep pace under existing federal regulations.

The *January 2017 State of Federal IT Report* places the importance of improving IT acquisitions in perspective. It identifies 41 federal agencies with annual IT spending greater than $500 million. The report also states that as of September 2016, the Federal IT Dashboard listed more than 4,300 IT programs in 780 major IT investments, and 43% of the projects were listed as over budget or behind schedule.[1]

Government program managers (PMs) struggle not only to buy new IT systems, but to modernize existing systems and avoid high operations and maintenance (O&M) costs of legacy systems. New laws, regulations, and guidance have created a plethora of frameworks, models, and methods for procuring and managing IT, and PMs can become overwhelmed with choices that are available to them. There is a need for analysis of the current literature and a gleaning of best practices and checklists that shed light on a path to successful IT procurement and modernization.

Scope

The term IT acquisition encompasses a broad range of products or services (e.g., end-user products, storage, compute, data centers, applications and software development, networks, transport, network management, help desk). Two areas of IT acquisition have already received heavy focus: (1) applications and software development (especially agile development), and (2) cloud services. Because it would be too extensive to research and address best practices for the full range of IT products or services, the focus of this report was narrowed to concentrate on the acquisition of network services (operations). Network services were selected because of its technical complexity, high cost, high risk, and highly sought-after requirements. The literature shows numerous federal acquisition professionals have requested or are requesting help with acquisition strategies and request for proposal (RFP) development for network services. Therefore, the goal of this report is to

- sort through the voluminous literature and highlight specific recommendations/best practices that federal agencies and program offices should be implementing for a successful network service acquisition;
- analyze existing models, frameworks and taxonomies that can be applied to IT acquisitions;
- develop and highlight checklists of the most important and applicable best practices; and

[1]CIO-Council-State-of-Federal-IT-Report-January-2017; p. Pol-4; p. A-2.

- reference exemplars and templates from other IT contracts (evaluation factors, service level agreements (SLAs), statements of work, incentive and/or payment plans, etc.).

Even though the focus for this report is on network services, the checklists and exemplars that have been applied as best practices can be applied to other types of IT acquisitions.

The Federal Government Legislative Initiatives

Diverse IT laws, regulations, policy, and guidance have been developed by the federal government to support the ever-expanding need for computing services. An initial scan of the available literature revealed thousands of recommendations concerning IT acquisition best practices since 2000. The Government Accountability Office (GAO) alone produced 803 recommendations between fiscal years (FY) 2010–2015. This large number of recommendations is a deterrent for acquisition professionals hoping to leverage the lessons learned by others.

Two of the most impactful policies were Section 5202 of the Clinger-Cohen Act of 1996 and Section 39.103 of the Federal Acquisition Regulations (FAR). Both recognized the potential benefits of modular contracting to control large systems implementation, and both state that agencies should, to the maximum extent practicable, use modular contracting for an acquisition of a major system of IT. Other key items of recent legislation include the following:

- Federal IT Acquisition Reform Act (FITARA), Dec. 19, 2014
- Modernizing Government Technology Act of 2017 (MGT Act)
- FITARA Enhancement Act of 2017
- TBM Council TBM Taxonomy v2.1, March 1, 2018
- President's Management Agenda, March 19, 2018

Some of the references used for this research provide detailed lists of legislative actions and major policies that impact the management and oversight of IT acquisitions. For more information on historical statutes and policy changes since Clinger-Cohen, see the following:

- History of IT Acquisition Reform—JCM05, Sept. 2015; pp. 91–103
- GAO-17-8-IT: Workforce Key Practices for Strong IPT, Nov. 30, 2016; pp. 5–8
- CIO Council State of Federal IT Report—Jan. 2017; pp. A-2–A-19

Frameworks and Taxonomies: How to Define IT Services

IT is many things to many people. How should it be categorized so federal agencies are talking about the same thing and making like comparisons? What is the difference between storage, compute, data center, transport, layer 2/3, local area networks/wide area networks, networking, desktops, end-user devices, etc., and does their use mean the same thing when employed by different federal agencies? The adoption of frameworks and taxonomies is needed to help answer these questions early in a program. A model or framework provides a basic structure, and a taxonomy provides a scheme of classification (lower level details).

Government programs typically assign a PM and use an Integrated Program Team (IPT) to execute the acquisition and operation of their IT system. One of the first problems encountered by any IPT working on an IT acquisition is to identify a solution that will fit their organization goals and mission. The IPT has many choices, including which framework or taxonomy to adopt as they develop requirements and secure funding. Historically, they are not focused on adopting a common framework or taxonomy standardized across the federal government. However, that is exactly what is needed to provide transparent IT costs, consumption, and performance and enable cross-agency analysis and data sharing.

An evolution of frameworks and taxonomies are available for both industry and government to use in developing and deploying IT systems. Industry was the driver for many years, and the government adopted many of the ideas to embrace commercial best practices and adopt open system architectures. However, the government has created their own frameworks and taxonomies in recent years to bring both business and financial discipline to their requirements development processes. This led to a range of options for what should or could be used.

The following is a short summary of the key frameworks that appear to be most prevalent. The frameworks range from an engineering solution, to a business management solution, evolving to an investment management solution, and finally a framework that reflects a cost focus.

Open Systems Interconnection Model—An IT Engineering Solution

The Open System Interconnection (OSI) model defines a networking framework to implement protocols in seven layers. If you ask a network engineer to work on an IT acquisition, they almost always refer to the OSI 7-layer model, which developed from commercial industry standards.

The OSI 7-layer model was published in 1984 by International Standards Organization (ISO) as standard ISO 7498 and by the Telecommunications Standardization Sector of the International Telecommunication Union as standard X.200. It divides network communication into seven layers. Layers 1–4 are mostly concerned with moving data around. Layers 5–7 contain application-level data. Networks operate on one basic principle: “pass it on.” Each layer takes care of a very specific job and then passes data on to the next layer.

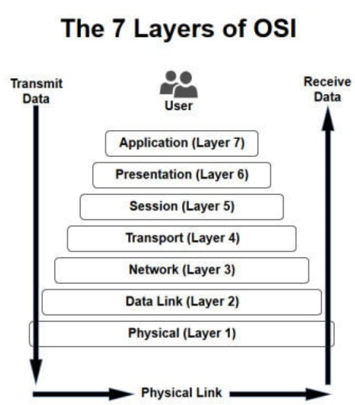


Figure 1. **OSI 7 Layer Model**

IT Infrastructure Library—Aligning IT Engineering With Business Management

IT Infrastructure Library (ITIL) is the set of detailed practices for IT Service Management (ITSM) whose primary purpose is to align IT services with business needs. By the early 1980s, the United Kingdom (UK) Central Computer and Telecommunications Agency (CCTA) saw that government and industry contracts were developing their own IT management practices and recognized the need for a standard. To meet this need, CCTA published the first ITIL in 1989. ITIL is a description of processes, procedures, and checklists used to establish a baseline allowing organizations to plan, implement, and improve. ITIL was built around a Plan-Do-Control-Act process model for controlling and managing operations. There were several updates from 1989–2005, followed by new releases of ITIL Version 2 in 2006, ITIL 2007, and the current version, ITIL 2011. A depiction of ITIL 2011 is shown in Figure 2.

Figure 2. **ITIL 2011**

Information Technology Investment Management Framework—Linking Business Management to IT Investment

In March 2004, the GAO released Information Technology Investment Management (ITIM) as an update to the exposure draft published in 2000. ITIM is a framework to measure the maturity of an organization's investment management processes that was built around the Select/Control/Evaluate approach outlined in the 1996 Clinger-Cohen Act.

ITIM identifies critical IT investment processes, establishes the presence or absence of these critical processes in an organization, assesses an organization's IT investment management capability and maturity, and offers recommendations for improvement. Used in this way, ITIM can be a valuable tool that (1) supports organizational self-assessment and improvement and (2) provides a standard against which an evaluation of an organization can be conducted.[2]

The ITIM Stages of Maturity with Critical Processes

Maturity stages	Critical processes
Stage 5: Leveraging IT for strategic outcomes	- Optimizing the investment process - Using IT to drive strategic business change
Stage 4: Improving the investment process	- Improving the portfolio's performance - Managing the succession of information systems
Stage 3: Developing a complete investment portfolio	- Defining the portfolio criteria - Creating the portfolio - Evaluating the portfolio - Conducting postimplementation reviews
Stage 2: Building the investment foundation	- Instituting the investment board - Meeting business needs - Selecting an investment - Providing investment oversight - Capturing investment information
Stage 1: Creating investment awareness	- IT spending without disciplined investment processes

Figure 3. **ITIM Stages and Critical Processes**

ITIM defines five successive stages of maturity that an organization can achieve in relation to IT and the critical processes that must be in place to achieve each stage. ITIM was developed as a tool to assess the maturity of an agency's IT investment management process and to identify areas for improvement. ITIM was not designed to define specific IT services or to link the business and finance aspects of IT acquisitions.

Technology Business Management Framework—An IT Business, Investment, and Cost Solution

Technology Business Management (TBM) was first released in 2016 after a year-long IT Cost Commission that partnered private and public businesses with the federal government. The Commission identified 21 recommendations for improving IT expenditures. Version 2.0 was released in October 2016 and Version 2.1 was released in March 2018. Chief information officers in industry and academia recently adopted TBM v2.1 as a value-management framework enabling technology leaders and their business partners to collaborate on business aligned decisions.

TBM is not only a framework but provides a standard taxonomy to describe cost sources (cost pools), technologies, and IT resources/applications/services (IT towers). The TBM taxonomy provides the ability to compare costs, technologies, resources, applications, and services to peers and third-party options (e.g., public cloud). Just as businesses rely on generally accepted accounting principles (or GAAP) to drive standard practices for financial

[2] GAO-04-394G-IT Investment Management (ITIM) Guide, Mar 2004

reporting—and thus comparability between financial statements—the TBM taxonomy provides a generally accepted way of reporting IT costs, capabilities and other metrics.3

The TBM taxonomy is shown in Figure 4.4

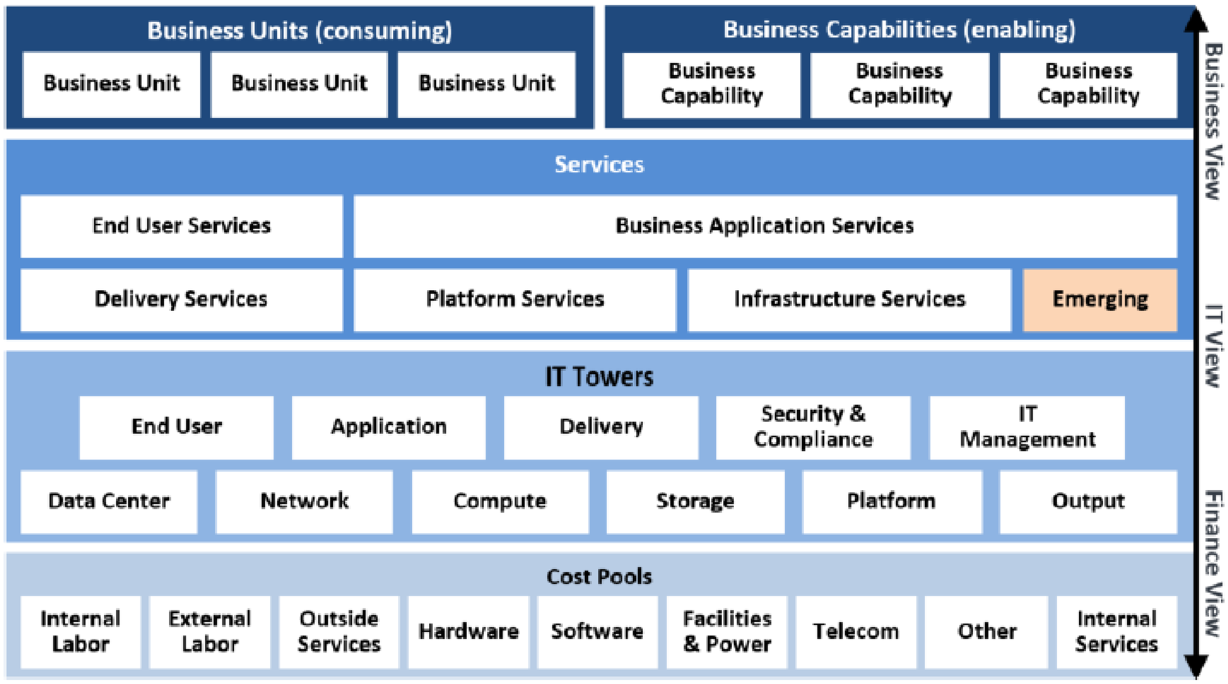


Figure 4. **TBM Taxonomy Overview**

The OSI model, ITIL practices, and ITIM frameworks all serve useful functions, but none of them provide the taxonomy needed to define and classify IT services and expenses that facilitate business-aligned decisions and enable cross-organizational analyses. Only, TBM provides enough detail to begin attempting to fully realize these goals.

Current Federal Guidance

The Office of Management and Budget (OMB) provides annual guidance on how federal agencies must assemble their IT budget exhibits for the president's budget submission to Congress. After ITIM's release in 2004, the OMB adopted its principles in its annual guidance, but it was not fully achieving its goal to understand and compare agency IT expenses on development and operations and maintenance (O&M)

Starting in 2016 (for the FY2018 budget), the OMB began the roll out of TBM to align with best practices in industry and academia. Currently, the OMB has adopted TBM for use by all federal agencies. As stated in the FY2020 IT Budget—Capital Planning Guidance (CPG),

[3] TBM Taxonomy, TBM Council, Version 2.0, Oct. 31, 2016; pg. 1.
[4] TBM Taxonomy, TBM Council, Version 2.0, Oct. 31, 2016; pg. 1.

> OMB is leveraging this widely adopted open source taxonomy, which is used within private, public and academic sectors and generating significant value. Leveraging a taxonomy that provides a standard business model for IT and is proactively managed by a non-profit organization also alleviates some of the burden for the government to identify, define, and achieve consensus on the standards and terms used. OMB is following an incremental process to roll out these changes.[5]

The CPG is incorporated into Circular A-11, Presentation, Submission and Execution of the Budget, which details how federal agencies are to submit their budget. The reference includes a figure that identifies the phased approach by year for full TBM implementation.

Federal adoption of TBM was initiated by the OMB during the Obama administration and it receives continued support during the Trump administration. The President's Management Agenda (PMA), released in March 2018, identifies Cross Agency Priority (CAP) Goal 10 as Improving Outcomes Through Federal IT Spending Transparency. It highlights that in the FY2018 President's Budget, 84% of the IT budget is categorized as "other" instead of being mapped to a specific IT category and spend. CAP Goal 10 states, "The Federal Government will adopt TBM government-wide by FY2022. This approach will improve IT spending data accountability and transparency, empowering agency executive suite leadership from across the enterprise to drive mission value and improve customer experience through technology."6

Based on OMB guidance and PMA direction, the TBM taxonomy is now mandated for all federal agencies. The timelines may vary between the OMB-mandated roll out and adoption by program offices, but the bottom line is that agencies are required to adopt the TBM taxonomy in IT acquisitions. All federal IT programs should adopt the TBM taxonomy and map the various aspects of their program to the appropriate IT towers and IT cost pools as defined by TBM taxonomy (see Figure 4). The IT towers should be reflected in the requirements and procurement documents. The IT cost pools should be reflected in either the Contract Data Requirements Lists (CDRLs) or the contract line item number (CLIN) structure for all acquisitions.

IT Acquisition Best Practices

This report provides best practices in four areas. The next section creates two new checklists that are based on best practices gathered from the literature. The references reviewed to support the research of IT acquisition best practices are identified in alphabetical order in the references section. In the sections preceding that, other existing well-defined best practices that are cited in this document for further consideration are provided; also, artifact references to locate examples of recent IT acquisitions that could be adopted for new requirements or competitions are provided.

IT Best Practices for Organizations and IPTs

The current IT literature offers numerous best practices. However, because they are embedded in the documentation, reviewing the literature can be overwhelming. It may be

---

[5] FY2020 IT Budget Capital Planning Guidance A-11 Draft, May 16, 2018
[6] President's Management Agenda, March 19, 2018

very difficult for a PM or IPT members to determine which items best suit their needs. This report analyzed each of the references to identify the most critical (and often repeated) best practices, sort them into logical groupings, and synthesize the information into an easy-to-use format. Most of the IT best practices fit in one of four areas: organizational readiness at the strategic and operational level; execution details at the tactical level; agile software development; or cloud services. The last two areas are specialized types of IT services, are the most mature forms of IT acquisitions, and already have a significant amount of research supporting best practices. Therefore, this research concentrated on the first two areas: organizational readiness and execution details.

The format chosen to present the analysis is a checklist with associated reference(s). This approach allows the reader to quickly identify the best practice and do targeted follow-up research on the details and intent behind each recommendation.

The two checklists that consolidate the numerous best practices are as follows:

- Appendix A, **IT Acquisition Best Practices—Organizational Readiness Checklist,** is an assessment of an agency or organization's readiness for large IT contracts.
- Appendix B, **IT Acquisition Best Practices—Pre-RFP Checklist for the IPT,** includes practical and tactical items the IPT can employ in developing the RFP.

The Organizational Readiness Checklist is targeted for agency leadership and provides valuable insight and risk assessment for these continuously repeated best practices. If the agency or organization has not assessed its own readiness to tackle major IT acquisitions and assigns it to an IPT, then the IPT needs to conduct that assessment and identify the gaps and risks to senior leadership. Failing to do that, the IT acquisition risks not meeting schedule or being over budget as have many other programs. The Pre-RFP Checklist for the IPT is a grouping of all the best practices that could be controlled or influenced by the IPT as they prepare the acquisition strategy and RFP. Neither of the checklists is exclusive to network services and can be equally useful to any IT acquisition.

IT Best Practices for Other Areas

There are several other areas of IT acquisition where best practices are needed. The resources in the references to this report have already developed well-defined criteria or checklists for these areas. These checklists are summarized below and should be used as presented in the source material.

IT Workforce Taxonomy and Labor Skill Best Practices

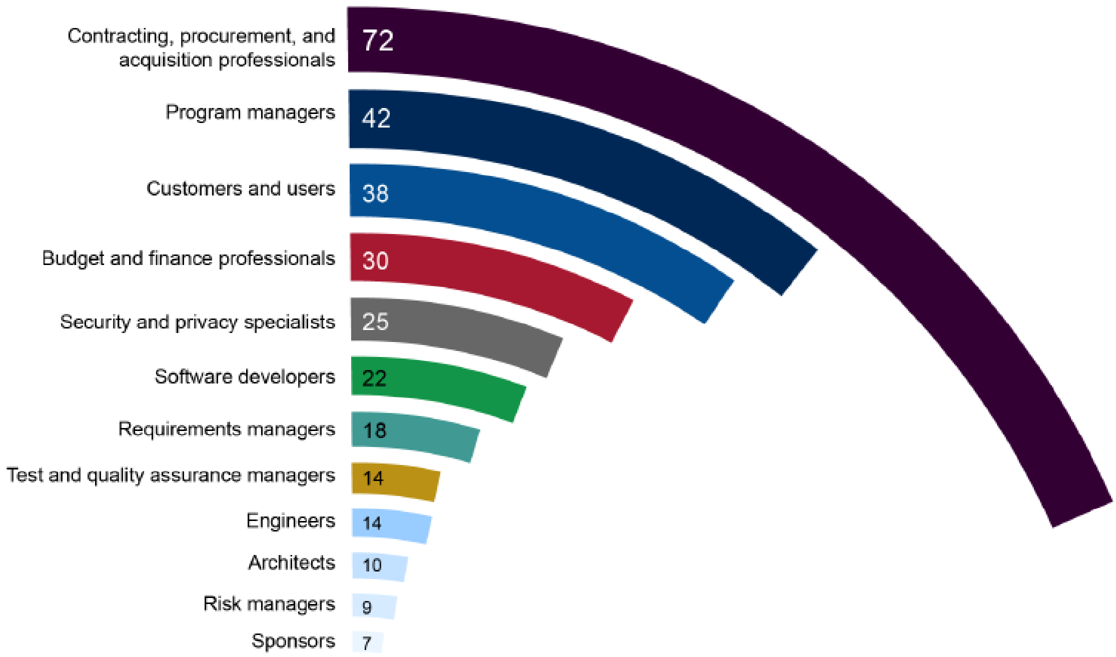
Properly staffing the IPT is critical. Several GAO reports point to this issue and provide additional information on workforce capabilities and skills:

- GAO-14-183T-IT: Leveraging Best Practices
- GAO-17-251SP-IT: Opportunities for Improving Acquisitions and Operations
- GAO-17-8-IT: Workforce Key Practices for Strong IPT, Nov. 30, 2016

GAO-17-8 offers the most explicit information on IPT skillsets and summarizes the proposed make-up, shown in Figure 5.[7]

[7] GAO-17-8-IT: *Workforce Key Practices for Strong IPT*, Nov. 30, 2016; p. 25

The first five core disciplines shown in the GAO report are critical to any IPT, but inclusion of the remaining disciplines should be considered and depends on the scope and size of the IPT.



Source: GAO analysis of relevant guidance from recognized sources including MITRE, the Project Management Institute, Inc., and the Software Engineering Institute; and data from selected agency chief information officers and program offices from the Departments of Commerce, Defense, Veterans Affairs, and the General Services Administration. | GAO-17-8

Figure 5. **Frequency of Core Disciplines in IPT**

Requirements Management Checklist

Good requirements management practices help organizations to better manage the design, development, and delivery of IT systems within established cost and schedule timeframes. GAO Report 18-326, *DoD MAIS: Adherence to Best Practices*, May 24, 2018, provides a list of best practices that can help an IPT through the requirements management process.[8] The report includes these practices, with additional detail for developing requirements:

- developing an understanding with the requirements providers of the meaning of the requirements;
- obtaining commitment to requirements from project participants;
- managing changes to requirements as they evolve during the project;
- maintaining bidirectional traceability among requirements and work; and
- ensuring that project plans and work products remain aligned with requirements.

[8] GAO-18-326: *DoD MAIS: Adherence to Best Practices*, May 24, 2018; p. 18

Risk Management Checklist

GAO Report 18-326, *DoD MAIS: Adherence to Best Practices*, May 24, 2018, provides a list of IT best practices that can help an IPT through the risk management process.[9] An effective risk management process includes the following leading practices:

- determining risk sources and categories;
- defining parameters used to analyze and categorize risks and to control the risk management effort;
- establishing and maintaining the strategy to be used for risk management;
- identifying and documenting risks;
- evaluating and categorizing each identified risk using defined risk categories and parameters, and determining its relative priority;
- developing a risk mitigation plan in accordance with the risk management strategy; and
- monitoring the status of each risk periodically and implementing the risk mitigation plan as appropriate.

Solicitation and Contract Exemplars

In addition to checklists, one of the most useful items an IPT can leverage in developing an IT acquisition are exemplars from programs of similar size and scope. Several federal websites contain materials that can be used to develop an IT requirement and conduct a source selection:

- **GSA Websites**: General Services Administration (GSA) Technology Products and Services website has examples of statements of work for various IT functions. See https://www.gsa.gov/technology/technology-products-services/how-to-get-help/sample-technology-statements-of-work. GSA also provides the Acquisition Gateway, which includes a detailed document library and project center with specific exemplars. The Acquisition Gateway can be accessed at https://www.gsa.gov/tools/supply-procurement-etools/acquisition-gateway.
- **DAU Website**: The Defense Acquisition University (DAU) has developed several templates, exemplars, and guides to support IT acquisition planning and contract vehicle decisions. They have an IT/Software (SW) CoP that provides a forum that is focused on improving the performance of the Defense IT/SW workforce. This community is focused on collaborating with the IT/software acquisition workforce to ensure engineer, design, develop, and sustain world-class IT/software acquisition practices. This community touches on all aspects of IT/software acquisition to facilitate better, faster, cheaper software solutions for all DoD personnel. Their link is https://www.dau.mil/cop/it/SitePages/About.aspx. They provide more than 30 documents that offer lessons learned on various military IT systems. They also have a tools section that offers best practices in assessing risk and compliance.
- **MITRE Website**: MITRE has created a public website, AiDA (Acquisition in a Digital Age), that provides extensive references for acquisition guides and

[9] GAO-18-326: *DoD MAIS: Adherence to Best Practices*, May 24, 2018; pp. 18–19

acquisition templates that are helpful to PMs and IPTs: https://aida.mitre.org/references/. The site also has a section specific to IT acquisition (https://aida.mitre.org/references/it/) that offers extensive information on IT, with policies, guides, reports, and articles.

The Future

IT Acquisition Trends

Although modernizing the federal government infrastructure and technologies has been a focus of acquisition and source selection, several new approaches are starting to gain traction within the federal programs. Many federal agencies are already purchasing cloud services and agile software development through Managed Service Providers (MSPs). These contracts are usually performance based and often use a fixed price contract pricing arrangement. Agencies are now starting to try to extend this MSP approach to other areas of IT acquisitions, including network services, cybersecurity, and end-user devices. MSPs use the commonly-applied terms such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (Saas). However, the literature shows that the new term of IT as a Service (ITaas) is trending.

On the commercial side, industry is driving towards providing commercial software as a subscription. This trend has been described in several trade magazines and IT publications. Some companies estimate that software as a subscription could grow to as much as 80% of current licensed software by as early as 2020.10

Further Research

There is more that can be done to analyze the current thinking around IT modernization and network services. Some recommendations for continued research include the following:

- Expand the collection of exemplars. Add more materials from resources supporting IT acquisitions, such as GSA Alliant, and provide an analysis and summary of the exemplar content so it can be more readily applied.
- Conduct more analysis on recommended considerations for choosing contract types and incentives for IT acquisitions.
- Continue analysis of IT acquisition related reports for updates and additions to the existing checklists. This would include, but not be limited to, the following highly relevant reports that came out after the literature review was complete:
    - GAO-18-42-IT: Agencies Need to Involve CIOs
    - GAO-18-234T: Further Implementation of FITARA
    - GAO-18-460T-IT: Further Recommendations for Acquisitions
    - GAO-18-566T-IT: Continued Implementation of High-Risk Recommendations

[10] Christy Pettey, Lessons learned from IT leaders who successfully moved to a SaaS-based business model, Gartner Group website article, May 30, 2018; p. 1

- Evaluate means to integrate IT modernization efforts with change management. Recent literature shows that change management can be holding back policymakers from achieving the full benefits of IT modernization.[11]

Conclusion and Recommendations

There are many best practices that exist within both government and industry to efficiently and effectively manage IT systems and their modernization. However, it is difficult for any program manager to grasp the full breadth and scope of the information and then select a model or practices that best fits the program. Several recommendations can help alleviate the situation:

- **Select the appropriate model, framework, and/or taxonomy:** This report reviewed the latest approaches to buying and managing IT systems as they evolved from the 1980s until now. More than one may apply; however, the TBM taxonomy appears the be the most comprehensive model.
- **Adopt the TBM taxonomy** and map the various aspects of the program to the appropriate IT towers and IT cost pools as defined by TBM taxonomy. The IT towers should be reflected in the requirements and procurement documents. The IT cost pools should be reflected in either the Contract Data Requirements Lists (CDRLs) or the contract line item number (CLIN) structure for all acquisitions.
- **Utilize checklists to help guide the Agency and IPTs through the process:** The agency or organization needs to assess its own readiness to tackle major IT acquisitions. The IPT needs to conduct that assessment and identify the gaps and risks to senior leadership before starting the project. It is recommended that the Agency and IPT utilize the checklists shown in Appendixes A and B of this report as they prepare the acquisition strategy and execute the source selection.

References

1. CIO-25-Point Implementation Plan to Reform Federal IT, 12 Sep 2010
2. CIO Council State of Federal IT Report, Jan 2017
3. Comp Econ How to Evaluate IT Procurement Contracts, Nov 2008
4. Comp Econ Network Operations Outsourcing Trends, Sep 2017
5. Contracting Guidance to Support Modular Development, 14 Jun 3012
6. Federal Acquisition Regulation (FAR) 39.103
7. Federal Computer Week, Five Myths of Change Management, 28 Mar 2019
8. FY2020 IT Budget Capital Planning Guidance-A-11-Draft, 16 May 2018
9. GAO-04-394G-IT Investment Management (ITIM) Guide, Mar 2004
10. GAO-14-183T-IT-Leveraging Best Practices, 13 Nov 2013

[11] Dixon, Marglin, and Mason, Five Myths of Change Management, *Federal Computer Week*, March 28, 2019

11. GAO-17-8-IT Workforce Key Practices for Strong IPT, 30 Nov 2016
12. GAO-17-251SP-IT Opportunities for Improving Acquisition and Operation, 11 Apr 2017
13. GAO-17-291-Defense Contracting-Incentives, 11 Jul 2017
14. GAO-17-317-High Risk Series, 28 Mar 2017
15. GAO-17-494T-Implementation of IT Reform Law and Related Initiatives, 28 Mar 2017
16. GAO-18-326-DoD MAIS Adherence to Best Practices, 24 May 2018
17. History of IT Acquisition Reform-JCM05, Sep 2015
18. IT Procurement Practices That Clients Apply and the Best Practices That Gartner Recommends, 2009
19. Modernizing Government Technology Act of 2017
20. OMB FAC C Digital Services, 18 May 2018
21. OMB Guidance for specialized acquisition cadres, 13 July 2011
22. OMB FAC PPM, revised, Dec2013
23. Presidents-Management-Agenda, 19 Mar 2018
24. TBM Council-TBM-Taxonomy-v2.1, 01 Mar 2018

Disclaimer

Appendix A. IT Acquisition Best Practices—Organizational Readiness Checklist

Item	Best Practice	Reference
1.	Assess IT Investment Management maturity using ITIM	• GAO-04-394G-IT: Investment Management (ITIM) Guide, Mar 2004; p. 1-19
2.	Analyze the IT workforce for Skill Gaps and develop a plan to fill them.	• GAO-17-8-IT: Workforce Key Practices for Strong IPT; p. 5-10, 45 • GAO-17-494T: Implementation of IT Reform Law and Related Initiatives Can Help Improve Acquisitions; p. 1, 7-10
3.	Program staff has necessary knowledge and skills.	• GAO-14-183T-IT: Leveraging Best Practices; p. All • GAO-17-8-IT: Workforce Key Practices for Strong IPT; p. 2, 45-68
4.	Properly Staff Integrated Product Team (IPT).	• GAO-14-183T-IT: Leveraging Best Practices; p. All • GAO-17-8-IT: Workforce Key Practices for Strong IPT; p. 25 • Presidents-Management-Agenda, 19 Mar 201; p. 20 • OMB Guidance for specialized acquisition cadres, 13 July 2011; p. A-2 to A-4 • IT Procurement Practices That Clients Apply and the Best Practices That Gartner Recommends_2009; p. 7-12
5.	Program staff are consistent and stable.	• GAO-14-183T-IT: Leveraging Best Practices; p. All
6.	Senior Department and Agency executives support the program.	• GAO-14-183T-IT: Leveraging Best Practices; p. All • GAO-17-251SP-IT: Opportunities for Improving Acquisitions and Operations; p. 7, 13 • GAO-18-326-DoD MAIS: Adherence to Best Practices; p. ES, 7-11 • Procurement Practices That Clients Apply and the Best Practices That Gartner Recommends_2009; p. 11-12
7.	CIO role is following FITARA. CIO: 1) has significant role in decision making for IT budgeting, 2) approves IT budget for agency, 3) certifies compliance with OMB incremental dev guidance, 4) reviews and approves IT contracts, 5) approves appointment of any agency employee with title of CIO.	• GAO-17-494T: Implementation of IT Reform Law and Related Initiatives Can Help Improve Acquisitions; p. 4-5, 11-18 • GAO-17-251SP-IT: Opportunities for Improving Acquisitions and Operations; p. 4-7
8.	Program receives sufficient funding.	• GAO-14-183T-IT: Leveraging Best Practices; p. All
9.	Agency uses IT spend plans to improve budgets.	• GAO-17-251SP-IT: Opportunities for Improving Acquisitions and Operations; p. 9
10.	Agency develops and maintains an IT Strategic Plan. 1) Use a strategic approach for legacy system migration. 2) Migrate more services to cloud. 3) Buy more and develop less. 4) Mitigate impacts on jobs when closing data centers or purchasing services.	• GAO-17-251SP-IT: Opportunities for Improving Acquisitions and Operations; p. 15, 20-21
11.	Agency provides oversight for IT purchased as a service.	• GAO-17-251SP-IT: Opportunities for Improving Acquisitions and Operations; p. 15
12.	Attract and invest in IT workforce.	• GAO-17-251SP-IT: Opportunities for Improving Acquisitions and Operations; p. 16, 19

Appendix B. IT Acquisition Best Practices – Pre-RFP Checklist for the IPT

Item	Best Practice	Reference
1.	Use the TBM Taxonomy from the outset; map the scope of your effort to the appropriate IT Towers and Cost Pools.	• Presidents-Management-Agenda; p. 40 • FY2020 IT Budget Capital Planning Guidance-A-11-Draft; p. 5-8 • TBM Council-TBM-Taxonomy-v2.1; p. All
2.	Develop a modular contracting approach consistent with requirements of FAR 39.103-104.	• FAR 39.103-104 • Contracting Guidance to Support Modular Development, 14 Jun 3012; p. 3-7
3.	Use performance-based outcomes specified through SLAs. 1) SLAs developed by IPT (to include contracting), reviewed by legal.	• IT Procurement Practices That Clients Apply and the Best Practices That Gartner Recommends_2009; p. 9 • GSA-18-326-DoD MAIS: Adherence to Best Practices; p. 7
4.	Identify and actively engage with the stakeholders/users throughout the acquisition (especially in development of requirements).	• GAO-14-183T-IT: Leveraging Best Practices; p. ES, 4, 6, 13 • GAO-18-326-DoD MAIS: Adherence to Best Practices; p. ES, 7-11 • IT Procurement Practices That Clients Apply and the Best Practices That Gartner Recommends_2009; p. 11
5.	IPT manages and prioritizes requirements.	• GAO-14-183T-IT: Leveraging Best Practices; p. ES, 4, 6, 13 • GSA-18-326-DoD MAIS: Adherence to Best Practices; p. ES, 7-11
6.	Focus on cybersecurity; convey IT and cyber issues early to leadership.	• GAO-17-251SP-IT: Opportunities for Improving Acquisitions and Operations; p. 6-8, 21
7.	Work more closely with the procurement (contracting) organization.	• GAO-17-251SP-IT: Opportunities for Improving Acquisitions and Operations; p. 9-11 • CIO-Council-State-of-Federal-IT-Report-January-2017, p. Rec-9
8.	Determine if program is ready for a Managed Service Provider (MSP) approach or plan. 1) Requires detailed understanding of current systems and performance metrics. If not ready, consider a hybrid contracting strategy 2) Determine readiness for Firm Fixed Price (FFP); • If FFP, consider an outcome-based payment plan; only pay for services delivered • If not FFP yet, investigate use of Incentive Fees (IF) over Award Fees (AF)	• DoDI 5000 Series for Major Weapons Systems • Contracting Guidance to Support Modular Development, 14 Jun 3012; p. 10-14 • Comp-Econ-How to Evaluate IT Procurement Contracts, Nov 2008; p. 1-11
9.	Leverage common contracting templates	• IT Procurement Practices That Clients Apply and the Best Practices That Gartner Recommends_2009; p. 9
10.	Leverage common evaluation factors	• IPT guidance from Contracting Officer
11.	Buy as an Enterprise 1) Leverage existing contract vehicles (GWACs, MAC, Agency, IDIQ, etc.). 2) Get a free scope evaluation (e.g., GSA Alliant 2)	• CIO-Council-State-of-Federal-IT-Report-January-2017, p. F-5 • Presidents-Management-Agenda, 19 Mar 201; p. 34 • GSA Website, Agency websites
12.	Implement a strong risk management program	• GSA-18-326-DoD MAIS: Adherence to Best Practices; p. ES, 7–18

Implementing Set-Based Design in DoD Acquisitions

Norbert Doerry—is the Technical Director of the NAVSEA Technology Office (SEA 05T). In addition to leading special projects, Dr. Doerry facilitates the transition of technology from industry and academia into naval warships. He retired from active duty in the U.S. Navy as a Captain with 26 years of commissioned service. He currently is focused on developing Medium Voltage DC (MVDC) Integrated Power and Energy Systems (IPES) for future warships, institutionalizing Set-Based Design in the U.S. Navy, and facilitating the introduction of flexibility and modularity features in future U.S. warships. [norbert.doerry@navy.mil and doerry@aol.com]

Philip Koenig—is the Director of the NAVSEA Industrial and Economic Analysis Division (SEA 05C1). His current work focuses on projections of cost, capability, and capacity in the U.S. shipbuilding industrial base. His previous experience was in the Future Ship and Force Architecture Concepts Division (SEA 05D1), the Office of Naval Research, NSWC Carderock, Chevron Shipping Company, and Vickers Shipbuilding Group Ltd. He teaches shipbuilding courses at the University of British Columbia and at MIT. [philip.koenig@navy.mil and pckoenig@mit.edu]

Abstract

Set-based design (SBD) is a technical and managerial approach that is increasingly being used to improve quality and responsiveness in U.S. naval ship design projects. It was employed on the Ship-to-Shore connector, the Amphibious Combat Vehicle (ACV), and the Small Surface Combat Task Force, and is being applied in ongoing surface combatant and submarine design studies. In contrast to iterative point-based design approaches, SBD projects arrive at a design solution by systematically eliminating regions of the design space rather than by selecting a solution early and iterating it through a design spiral to make it work. This paper reviews the fundamentals of SBD and discusses implementation strategies to reduce technical, schedule, and market risk; accelerate design convergence; enable distributed design teams; and improve cost estimates. We discuss how SBD enables early identification and resolution of knowledge gaps, enabling quicker design progress. The role of SBD in organizational learning and the ability to re-use knowledge products across acquisition programs is highlighted.

Introduction

In the design of many types of complex engineering systems, requirements and technical attributes are subject to considerable uncertainty. In this environment, organizing and managing the design workflow and decision making process to ensure that the optimal design is produced is difficult. In the past, complex design projects have been run in a point-design-based paradigm, but that approach has some weaknesses. Set-based design (SBD) is a comparatively new method that has gained traction in recent years in naval ship early-stage design. It has been applied to ship-to-shore connectors (Mebane et al., 2011), amphibious combat vehicles (Burrow et al., 2014; Doerry et al., 2014), surface combatants (Garner et al., 2015), submarines (Parker et al., 2017), and other programs.

The SBD method is conventionally described as a process of generation and elimination. First, a range of possible design solutions is generated. Each is described in terms of a set of design variables. The ranges of each variable are combined to define an *n*-dimensional design space. Through a process of elimination, infeasible or highly dominated regions of the design space are discarded and the design space becomes more restricted.[1] Design decisions are deferred to the latest possible point in the project schedule, thus keeping the maximum extent of the design space available for consideration until the latest possible moment.

There are pitfalls that arise in applying this method; the way certain details are handled can determine the success (or otherwise) of the design outcome. For example, delaying decisions confers no intrinsic benefit of its own; value is created only when such a delay is designed to generate lead time to gain specific types of additional information needed to make a better decision. Otherwise, delay is merely procrastination, which reduces focus and dissipates momentum.

SBD Fundamentals

The SBD concept dates back to Toyota's approach to automotive design as described in benchmarking studies done in the 1990s (Ward et al., 1995a; Ward et al., 1995b). Sobek, Ward, and Liker (1999) set forth the general principles as follows:

1. Map the design space.
   a. Define feasible regions.
   b. Explore trade-offs by designing multiple alternatives.
   c. Communicate sets of possibilities.
2. Integrate by intersection.
   a. Look for intersections of feasible sets.
   b. Impose minimum constraint.
   c. Seek conceptual robustness.
3. Establish feasibility before commitment.
   a. Narrow sets gradually while increasing detail.
   b. Stay within sets once committed.
   c. Control by managing uncertainty at process gates.

Singer et al. (2017), working in naval ship design, characterized SBD as follows:

1. Communicating broad sets of design values,
2. Developing sets of design solutions,
3. Evaluating sets of design solutions by multiple domains of expertise,
4. Delaying design decisions to eliminate regions of the design space until adequate information is known, and
5. Documenting the rationale for eliminating a region of the design space.

Starting with a characterization of the design space that is large enough to ensure with high probability the inclusion of the best solution of a design problem, SBD systematically eliminates infeasible and highly dominated regions of that design space. SBD thus arrives at a design solution largely through a process of elimination. A region of the design space is infeasible if there is a high confidence that a solution to the design problem

[1] In the design space, feasible solutions are points (or regions encompassing many points) that satisfy the criteria of all design domains (disciplines) (e.g., hydrostatics, speed, range, military effectiveness, cost). Highly dominated regions of the design space are those in which there is another region that is superior by every metric.

does not exist within the region. A region is highly dominated if the key metrics of interest in another feasible region are all better, even when considering uncertainty.

The process of eliminating a region of the design space is called a set reduction. Early on, set reduction is generally accomplished by determining that a region is not feasible. While determining that a region of a design space is feasible requires a considerable amount of information because every domain (sometimes called a design discipline) must evaluate with high confidence that the region is feasible, determining that a region is not feasible only requires one domain to conclude with confidence that the region is not feasible. In this way, an SBD design process can proceed cumulatively as each domain adds new knowledge. Thus, in traditional point-based design (PBD) methods that concentrate on evaluation of the feasibility of a design concept, the activities of the many domains must be coordinated; the schedule is impacted by the slowest domain. The asynchronous nature of SBD relaxes the need for tight coordination among the domains, reducing the dependency of the project schedule on any one domain.

In SBD, focus is placed on identifying key knowledge gaps, conducting experiments and analyses to resolve the knowledge gaps, and deferring associated design decisions until the knowledge gap has been resolved. As described by Cloft, Kennedy, and Kennedy (2018), this was the method employed by the Wright brothers to beat all others in becoming the first to achieve heavier-than-air flight with a relatively small budget. The Wright brothers identified three knowledge gaps:

- “the construction of the sustaining wings”
- “the generation and application of the power required to drive the machine through the air”
- “the balancing and steering of the machine after it is actually in flight”

To close the gaps, the Wright brothers systematically performed experiments to understand the impact of different design options on each of the knowledge gaps. They constructed a wind tunnel to test hundreds of different wings and produce trade-off curves in a short time. Their newly gained understanding of wings enabled them to design an efficient propeller which in turn reduced the power required from the engine. Cloft, Kennedy, and Kennedy (2018) cite the Wright brothers’ approach to engineering design, based on an organized approach to the obtaining and application of knowledge, as an early example of effective SBD.

Figure 1 illustrates an example of a process for set-reduction. Initially, the entire integrated design space is considered feasible because none of the regions have been shown to be not feasible. Domains 1 and 2 begin work to create new knowledge to determine what parts of the design space are feasible (green), not feasible or highly dominated (red), or uncertain (yellow) from the domain’s perspective. Domain 2 illustrates a good practice of starting with low fidelity analysis that can quickly and inexpensively categorize much of the design space as feasible, not feasible, or highly dominated, but still leaves a considerable amount of the design space uncertain. Follow-on higher fidelity work, which takes longer and is more expensive, can concentrate on the uncertain region. As each Domain completes its analysis, its results are incorporated into the integrated design space as part of a set reduction. Note that since Domain 3 started after set reductions had taken place, it need not consider regions of the design space that had already been eliminated.

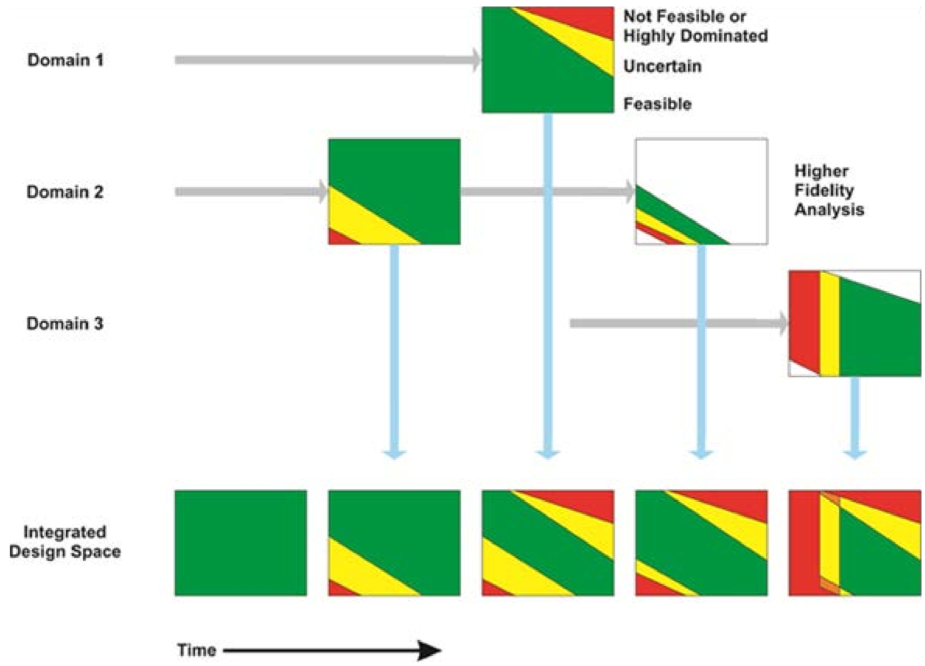


Figure 1. **Representative Implementation of Set Reduction**
(Singer et al., 2017)

The uncertainty of the analytic processes and test procedures should be well understood and considered in deciding to characterize a region of a design space feasible, not feasible or highly dominated. The goal is that for a given domain, no new information would result in a feasible region being considered not feasible or highly dominated, or a region not feasible or highly dominated being considered feasible. The uncertainty of the analytic process and test procedure results should be used to determine the boundaries of the remaining uncertain region from the perspective of the domain.

Figure 1 can also be used to distinguish between feasibility and viability. The green area in the integrated design space denotes feasibility. A region of the design space is feasible if analysis or testing to date has not shown that region to be uncertain or not feasible. Input from new domains could result in additional regions of the design space becoming uncertain or not feasible. A region is viable if all future analyses and testing (including verification testing) show that configurations exist that meet all requirements.

During design, while the feasibility of any one configuration can be determined based on analysis and testing performed to date, the viability of the configuration cannot be determined with confidence because the complete set of analysis and testing will not have been performed. If, however, the set of feasible configurations that correspond to a design space region are different enough from one other such that the probability that all of the configurations currently evaluated as feasible prove not to be viable is very small, then we can conclude that a viable configuration exists in that feasible design space region. Identifying the configurations within a feasible design space region that are viable or not feasible becomes the objective of future work.

Cost Estimating

Extending ship concept design cost estimating processes to the SBD environment is a work in progress. In SBD, the design variables defining the integrated design space do not always completely define a configuration; they are generally those that have a strong interaction between two or more domains. Design parameters that do not have a strong interaction, but are only an influence within a single domain, are typically treated independently by the domain teams. Hence a single point in the *n*-dimensional design space (with specified design variables) may reflect a large number of configurations corresponding to the multitude of combinations of individual domain design parameters that can be mapped to that single point.

Since cost estimating is one of the design domains, cost drivers should be design variables. However, practical difficulties arise due to the many design parameters, not all of whose cost implications are well enough understood to be incorporated into cost estimating relationships. See Cooper and Koenig (2018) for a discussion of this issue. Furthermore, there are some cost variables whose implications are not yet adequately built into the ship design solution generation process. An example of the latter would be industrial base capacity utilization, which is a very sensitive driver of naval ship cost. Work remains to be done to develop methods for incorporating that (and other) cost drivers into the design set generation process.

Figure 2 depicts an integrated design space consisting of a set of configurations intended to meet a specific set of requirements. The y axis is associated with one design variable with a hard constraint that separates feasible points (blue) from points that are not feasible (red). If all the feasible configurations were to prove viable, the best cost to assign to this point in the integrated design space would be the least expensive point (blue point furthest to the left); this configuration achieves the stated goals at the lowest cost. However, this feasible configuration may not prove viable once additional analysis and testing is conducted. Hence this cost is a lower bound with considerable cost risk. A higher cost estimate for a point in the integrated design space is associated with more configurations with a cost estimate equal to or below the higher cost estimate. For some cost above the lower bound, the probability will likely be low that all of the feasible configurations with a cost estimate below the specified cost are shown to be not viable. The lowest cost where this condition is met should be used as the cost estimate for that particular point in the integrated design space. Doerry (2015) details a method based on a diversity metric for determining this cost point.

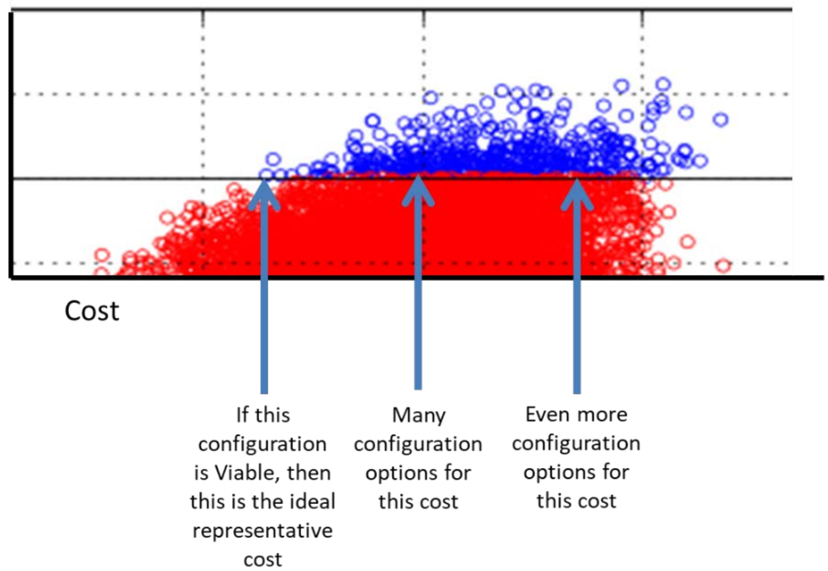


Figure 2. **Cost Estimating in SBD**

This process evaluates the cost for a configuration meeting a specific set of requirements based on the set of individual configuration cost estimates and not on the cost estimate of a particular configuration. If all the feasible configurations share one or more common failure modes, and the particular set of requirements associated with the design space is of great interest, then work and analysis should be performed to resolve whether the failure modes are failures or not.

The difference between estimating the cost of a single configuration and the cost associated with a group of possible configurations can be illustrated with an options analysis. SBD inherently incorporates the concept of an option. An option is the right or ability to do something in the future for a specified cost, but not the obligation to do so. The cost of acquiring an option is compared to the potential value it will bring in the future when more information is available to make a decision.

As an example, consider a project with designs for two configurations, one which includes widget A and one with widget B. Widget A costs $1,000 and is certain to work. Widget B costs $300, but there is only a 70% chance that it will work as planned. If it does not, there will be an estimated $2,000 of rework. The cost estimate for the configuration with Widget B may incorporate the $300 to account for Widget B and assume the change-order pool will be sufficient to cover the possibility that Widget B does not work. Alternately the cost estimate could include the rework:

$$C_B = 0.70 \times \$300 + (1 - 0.70) \times (\$2{,}000 + \$300) = \$900 \quad (1)$$

where the minimum cost would be $300 and the maximum $2,300. If, however, preserving the option to install Widget A or B is incorporated at a cost of $100, the cost estimate would be

$$C_{A\ or\ B} = 100 + \ 0.70 \times \$300 + (1 - 0.70) \times \$1{,}000 = \$610 \quad (2)$$

where the minimum cost would be $400 and the maximum $1,100. Without the option, Widget B would likely be selected because its expected cost of $900 is less than the expected cost of $1,000 for Widget A. Preserving the option to use either Widget A or Widget B reduces the down-side risk as compared to option B alone ($1,100 instead of

$2,300) as well as the expected cost ($610 instead of $900 for Widget B alone or $1,000 for Widget A alone).

This inherent incorporation of options within SBD is one of its strengths. Expected costs can be reduced at the same time schedule delays due to rework can be avoided. Incorporating Widget B without incorporating the probability of rework would make it a program risk. Incorporating the option to use Widget A or Widget B effectively transforms the risk associated with Widget B into an opportunity. The value of this opportunity can be incorporated into the cost estimate.

Implementation Strategies

While the basic concepts of SBD are known, how to implement SBD for a particular design problem is not always clear. Key issues include

1. Defining the design problem
2. Organization of the design team
3. Specifying the design domains or disciplines
4. Identifying the variables that define the design space
5. Setting the initial boundaries of the design space
6. Establishing feasibility metrics
7. Establishing dominance metrics
8. Determining the types of analyses needed and scheduling them
9. Making a design choice once the design space has been narrowed to that which is feasible from the perspective of all domains

Defining the Design Problem

A design problem should be defined specifically enough to enable the design team to focus its efforts, but not so specific to require redefinition, as knowledge is gained during design activities. Many times, a set of requirements is provided by the customer, but these requirements may not be firm. Typical reasons for requirements not being finalized include (Singer et al., 2017) the following:

a. Some known requirements may require study to determine appropriate values or measures.
b. Some specified requirements may be relaxed once the cost impact is fully understood.
c. The need for some requirements may not be known because of a lack of understanding of the design space.
d. The need for some requirements may not be known because of evolving exogenous factors.

One of the first tasks of a design team should be to clearly define the initial set of requirements and characterize the uncertainty of these requirements. The uncertainty of requirements can be evaluated as part of a requirements risk review (Singer et al., 2017).

Where a requirement has uncertainty, it should be bounded within a range as part of the requirements risk review. For these requirements, the work and timeframe necessary to establish the threshold requirement should be defined. The system design must be affordably flexible to handle the range of requirement values until the requirement is finalized. Note that the requirement may never be finalized or may change over the product's service life, in which case a modularity or flexibility based approach towards meeting the requirement may be required.

Organization of the Design Team

Figure 3 shows one way of organizing a design team for accomplishing SBD in product development (such as preliminary design for a ship). The stakeholder board includes those with vested interests in the product's technical characteristics, schedule, cost and contribution to an overall portfolio of products. Often, the stakeholder board approves major set reductions and if necessary, selects the final configuration from the remaining feasible design space. One of the values of SBD is in helping the stakeholder board understand the design space and gain an organizational consensus on the way forward. Because SBD starts with broad boundaries for the design space and systematically eliminates regions of the design space based on evidence, the impact of late "did you consider X, Y, or Z?" questions is minimal because the answer will generally be "Yes, we considered X, Y, and Z and eliminated them for the following reasons …" In a traditional PBD, the design team either expends additional (and probably unplanned for) resources to address X, Y, and Z, or risks the political consequences of ignoring the interests of a stakeholder.

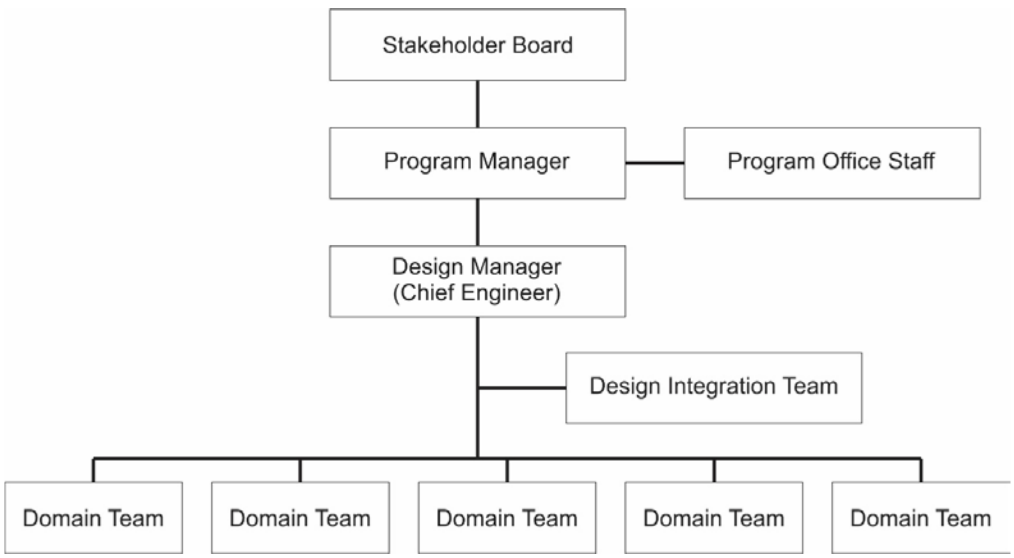


Figure 3. **Design Team Organization**
(Singer et al., 2017)

It is not unusual for stakeholders to have their own favorite solution (sometimes called a pet rock) prior to the start of the design effort. Ideally these pet rocks fall within the initial design space; hence, if they are eliminated, they are eliminated based on solid data and on consensus of the overall stakeholder board. In some cases, when presented the data, stakeholders will themselves advocate for the set reduction that eliminates their own pet rock. This is in contrast to a traditional PBD, which rarely includes all of the stakeholder's pet rocks in its initial set of configurations. Even if a pet rock is included and then eliminated because another configuration is evaluated as "optimal," the pet rock owner may not be satisfied because of a disagreement in the formulation of the optimization utility function. Note that SBD does not require the formulation or use of a utility function.

The responsibilities of a program manager and program office staff do not change whether the design is conducted using SBD or point-based methods. In many programs, the program manager concentrates on external interfaces such as the stakeholder board, Congress, and Department of Defense (DoD) organizations as well as program management activities such as contract management and financial management.

The director of the overall design effort is known as the chief program engineer or design manager. The design manager, supported by the design integration team, develops the overall plan for conducting SBD, provides tasking to each of the domain teams, coordinates domain team activities, presents major set-reductions to the stakeholder board to concur with set-reductions, documents set-reductions, manages the requirements, and manages the integrated design space. The design manager is also responsible for the integration and production of the specifications for the following detail design and construction contract. These specifications describe either the final solution or remaining design space.

Each domain, or design discipline, will have its own domain team. The exact number and definition of the domains and domain teams will vary somewhat project to project. Ideally the majority of the members of any one domain team would be co-located, but the collection of design teams need not be co-located. In some cases, it may be beneficial for one or more members of a domain team to also be a member of the design integration team to facilitate overall communication and coordination.

Specifying the Design Domains or Disciplines

Domains, or design disciplines, are typically defined based on the structure of the design organization. In many design organizations, design team members are provided by functional organization to form a large project team. In other design organizations, a small centralized design integration team assigns design tasks to the functional organizations; the functional organizations may not provide dedicated team members. SBD can function in both design organization constructs.

Another consideration for determining the boundaries of a domain is the ability of the domain to work independently and in parallel with the other domains. A design structure matrix (Eppinger & Browning, 2012) may prove useful for capturing the relationships between proposed domains and determining the degree of coupling among them. Ideally, a domain would require few key design variables to analyze.

When SBD is implemented for the design of a product, the domains should include not only those that define the product, but also those that evaluate the product. For preliminary ship designs (Technology Maturation and Risk Reduction), the definition domains typically are aligned with the traditional design disciplines:

- Hull
- Propulsion
- Electric plant
- Auxiliary systems
- Habitability
- Communications systems
- Weapon systems and combat system
- Aviation
- Arrangements
- Topside design

The evaluation domains typically are defined for assessments that have strong dependencies on multiple definition domains. Assessments that are strongly dependent on a single definition domain are typically accomplished by the definition domain. Typical preliminary ship design evaluation domains include the following:

- Weight management
- Signatures
- Producibility
- Cost
- Survivability
- Operational effectiveness
- Reliability, maintainability, and availability
- Human systems integration and manpower assessments
- Environmental, safety, and occupational health compliance
- Requirements management/traceability

For a ship concept study, conducted as part of the Material Solution Analysis, there may be only one “Ship Design” definition domain, and the evaluation domains could be limited to cost, survivability, operational effectiveness, and requirements management. Because many evaluation domains are not considered, a concept study should not result in a point design, but rather a design space which can be further reduced during preliminary design. The design space should be diverse in that it includes a variety of design approaches and/or features such that the likelihood that the un-evaluated domains will render the entire design space not feasible is small. Doerry (2015) provides methods for calculating diversity metrics for a design space. In SBD, Requirements Management should include tracking the uncertainty of requirements over time.

Identifying the Variables That Define the Design Space

While there are many thousands of design decisions that must be made to fully define a complex product, many of these decisions have impact entirely within one definition domain. On the other hand, some design decisions have significant ramifications across multiple definition and evaluation domains. The design variables associated with these significant cross-domain impacts should be used to define the overall product design space. The impact of design variables with small cross-domain impacts should be captured in uncertainty analysis; evaluation domains should consider the range of these small impact design variables when establishing the region of the design space categorized as uncertain.

In some domains, it may prove advantageous to apply SBD recursively within the boundaries of the domain.

For many products, the design manager and the leaders of each domain will collectively have sufficient insight to identify the design variables to use for defining the product design space. A design structure matrix (Eppinger & Browning, 2012) may prove useful for capturing the relationships among the domains.

Setting the Initial Boundaries of the Design Space

The initial range of values for design variables should be broad enough so that the resulting design space includes the global optimal solution to the design problem. Of course, if one doesn’t know which combination of design variable values results in the global optimal solution, then it is hard to have confidence that any restricted range will encompass the optimum. The way out of this dilemma is to take advantage of constraints, requirements, and the expertise of the domains.

For many acquisitions, the constraints with the greatest impact on design space boundaries are time, cost, and technical maturity. For example, immature technologies, such as those with a low Technology Readiness Level (TRL; Office of the Assistant Secretary of Defense for Research and Engineering [OASD(R&E)], 2011) that cannot realistically transition to mature products in time to support the acquisition can be safely eliminated. The consideration and elimination of these technologies should be documented.

One way to identify the boundaries is to start with the high priority requirements such as Key Performance Parameters (KPPs), Key System Attributes (KSAs), and Additional Performance Attributes (APAs), as defined in DoD (2015). Next, have the domains use ideation methods to develop sets of approaches for achieving these high priority requirements. Shah, Vargas-Hernandez, and Smith (2003) list a number of ideation methods as well as provide metrics for evaluating the number and variety of alternative ways of meeting an objective. An initial assessment of feasibility may be useful to eliminate options that cannot meet constraints. Documentation of this initial set-reduction is key to enable rapid reassessment of an eliminated solution approach should a constraint be relaxed or new, unanticipated information is obtained.

Based on the combined sets of approaches from all the domains, each domain should be able to translate them into a proposed initial set of boundaries. For some domains, the approaches will impact derived requirements (such as electrical and cooling demand). These boundaries should incorporate uncertainty as evaluated by the domains.

Establishing Feasibility Metrics

Early on, many immature technologies and products can be eliminated if they clearly cannot support the acquisition schedule, even if moderate delays in the acquisition schedule are accommodated. The evaluation of immaturity should be based on conversations with the industrial base or other hard evidence. Assuming a product will not be available because it currently is not available may result in a premature set-reduction. If an emerging technology has substantial benefit but cannot meet current schedule constraints, this should be conveyed to the customer to determine if delaying the schedule is warranted, or whether modularity and flexibility features should be incorporated to enable technology insertion when it is ready.

Physics based modeling and simulation should be employed as much as possible. Singer et al. (2017) define a Feasibility Element to be the output of analysis expressed as one of three values:

1. Feasible: high confidence that the configuration is feasible with respect to the analysis
2. Uncertain: low confidence that the configuration is either feasible or not feasible
3. Not Feasible: high confidence that the configuration is not feasible

Initially, low fidelity modeling can be used by each domain to classify the design space into feasible (green), uncertain (yellow), and not feasible (red) regions from their perspective. The integration team combines the design space evaluations from the different domains to create an integrated design space based on the following rule set (Singer et al., 2017):

1. Feasible: All feasibility elements are feasible.
2. Uncertain: All feasibility elements are either feasible or uncertain, with at least one uncertain.
3. Not Feasible: At least one feasibility element is not feasible.

As the design progresses, compound integration risk can be captured by considering regions where more than "n" feasibility elements are uncertain as Not Feasible under the assumption that the likelihood that all of the uncertain feasibility elements will eventually prove feasible is low.

Using the three colors to indicate the feasibility assessment values helps considerably in visualizing the impact of set-reductions. As regions of the design space that are red are eliminated as part of a set-reduction, each of the domains can concentrate of the remaining regions within their domain design spaces that are evaluated as uncertain. In this way, higher fidelity modeling can be focused on the regions of uncertainty rather than over the entire design space.

As the design progresses and the design space is better understood, the uncertainty associated with constraints can be reduced based on discussions with the customer. These constraints will further restrict the feasible region of the integrated design space.

In some cases, the lack of time or resources may require assessment of feasibility values for a particular feasibility element to be made qualitatively based on expert input. Documenting the rationale for the expert assessment is critically important to developing a recovery strategy if the assessment later is determined to be incorrect. Where possible, the uncertain region should be explored with quantitative analysis, and the feasible and not feasible regions selectively verified through quantitative analysis.

Establishing Dominance Metrics

One of the advantages of delaying decisions in SBD is that one can identify and pick the lowest cost option for which one has confidence will work. In point-based methods, options are often selected early when both feasibility and cost are not known with any degree of certainty. Within SBD, as more is known of the cost and feasibility of options, certain options can be eliminated because although they will work, other solutions will with high probability also work and will also cost less. A set-reduction can therefore be made based on dominance if the set-reduction does not have a significant impact on either the risk of feasibility or on the projected cost.

Determining the Types of Analyses Needed and Scheduling Them

Early on, priority should be given to analyses that can quickly and inexpensively eliminate as much of the integrated design space as possible. Regions eliminated need not be analyzed by other domains, thereby reducing the amount of work required. For example, Garner et al. (2015) reported that logic and initial appraisals led to the quick elimination of nearly 96% of the initially defined design space. The remaining analyses could focus on the remaining 4%, confident that the "best solution" did not reside in the eliminated 96%.

If a possibility exists that a feasible design space does not exist at all, testing limiting conditions may be of great value to prevent costly analysis of a concept that is fatally flawed. For example, during the concept exploration of the Amphibious Combat Vehicle (ACV), Burrow et al. (2014) reported that a baseline study was conducted to see if an ACV could be devised that met less than acceptable performance requirements at a reasonable price. One of the purposes of this study was to ensure that it made sense to conduct the more detailed and expensive analysis. If the unacceptable performance was not feasible, or if its cost was excessive, then it didn't make sense to continue the study. Any additional capability would cost more, and achieving feasibility would be more difficult. As it turned out, the unacceptable performance was feasible and not at an unreasonable cost; further analyses continued.

For many domains, there is great value in initially using fast, low fidelity, but well understood, models to screen the remaining design space with high confidence into feasible, not feasible, and uncertain regions. Higher fidelity modeling can then focus on resolving the remaining uncertain regions that have not been eliminated by other domains.

Some domains rely heavily on model testing. Ideally these model tests should not be used in a confirmation role as is typical in PBD, but rather to validate digital simulation models that are scalable across the remaining design space. The choice of parameters for the model should be based on maximizing contributions to model validation and not to ensure a point design meets program requirements. Maximizing learning usually has greater value than simple requirements verification. Furthermore, because the model parameters do not depend on the final outcome, their parameters can be chosen early, enabling the fabrication and testing of the models to occur earlier, thereby enabling earlier application of the testing insights gained.

Making a Design Choice Once the Design Space Has Been Narrowed to That Which Is Feasible From the Perspective of All Domains

The end game for SBD depends on the acquisition strategy and to some degree on the views of the stakeholder board. One possible outcome is a specification for the next acquisition stage that defines the remaining feasible design space. Industry is allowed to propose a configuration of their choosing that resides in the feasible design space. The government then selects the proposal using traditional source selection criteria.

Another possible outcome is to let the stakeholder board negotiate among itself to pick a single point or smaller set within the remaining feasible design space. This outcome recognizes that the optimal solution from typical utility functions may not be acceptable to enough stakeholders. In the end, it is enough for the stakeholders to form a consensus on what the single point or smaller set is, without having come to an agreement as to why the result should be chosen. Different stakeholders may support the same outcome for very different reasons. This outcome doesn’t preclude using utility functions and traditional optimization techniques to help the stakeholder board better understand the remaining feasible design space.

Another outcome is to analyze the remaining risks and select a region of the feasible design space that is robust to the consequences of the risks being realized or not. Conduct additional analyses of this region, while at the same time conduct work to resolve the risks. As risks are resolved (i.e., determine that the consequence will or will not happen with certainty), adjust the boundaries of the selected region of the design space accordingly. In this way, the design progresses with a high degree of risk tolerance.

Organizational Learning

If one of the goals of a design endeavor is to minimize the cost and amount of time to complete the design, then a logical approach is to have conducted as much of the analyses as possible prior to the start of the design. If previous work enables an immediate set reduction, then convergence to a final solution can happen faster. While opportunistic applications of previous work should always be pursued, even more benefit can be obtained by instantiating formal organizational learning techniques. Companies such as Toyota which have implemented effective organizational learning have been able to reduce product development time even when the complexity of their products has increased (Cloft et al., 2018). These techniques can include the following:

1. *Document Set Reductions*. Since SBD calls for good documentation for set-reductions, if the generalized knowledge and the resulting rationale for set reduction from a previous study is still valid for a current study, then the set reduction can occur with little or no additional work. In this way, there is great value in making generalizable conclusions within a set-reduction and properly documenting the assumptions and conditions associated with these conclusions. This documentation must be accessible to future design teams.
2. *Conduct pre-studies to characterize the design space*. Often studies are conducted prior to the start of a design activity to develop point designs to understand the "art of the possible." Unfortunately, the conclusions that can be drawn from these point designs are limited to the assumptions and tasking of the particular study, which can differ considerably from the current study. Studies of greater value provide more general insight that is intended to be applied to future studies instead of attempting to provide recommendations based on analysis of one or a few point designs. Historical examples of this generalized knowledge include the development of standardized series such as the Taylor Standard Series for hull resistance predictions (Gertler, 1954) and NACA wing section series for lift and drag predictions for foils (Abbott & Von Doenhoff, 1959). Other historical examples include the accumulation and publication of data such as the Hoerner (1965) manuscript on fluid-dynamic drag and the Hoerner and Borst (1985) manuscript on fluid-dynamic lift. Within the Naval Sea Systems Command, this type of knowledge is captured in design practices and criteria manuals (DPCs) which were previously called design data sheets (DDSs). The key is that these documents capture knowledge and insight rather than documenting a particular solution. Understanding the reasons for why potential solutions should be avoided is just as valuable (if not more so) than being presented with recommended solutions (where the recommendation may depend on many unstated assumptions). Tasking statements for pre-studies must emphasize the desired goal is generalized insight rather than point recommendations for a specific notional design. A process should exist for incorporating the knowledge gained from the pre-studies into the applicable DPCs or equivalent documents.
3. *Capture feedback from production and operations.* One of the challenges with documentation such as DPCs is keeping them up to date with lessons learned once the design has transitioned to the shipbuilder for detail design and construction. The value of capturing this critical information was recognized by Toyota. In 1995, Ward et al. reported that Toyota engineers would document in their lessons-learned books the positive and negative aspects of their designs once they transitioned to manufacturing. This insight enabled the designers to improve their future designs with respect to manufacturability without constant interactions with manufacturing engineers. Similarly, feedback from the operators should also be captured in lessons-learned documents.
4. *Capture knowledge in algorithms and data sets for design tools*. Automated design tools are very useful for systematically exploring a design space. These design tools must reflect in their algorithms acceptable design criteria and practices that necessarily evolve as technology advances and more is learned about a given discipline. Furthermore, most design tools require validated data sets to function. Since ship designs don't occur frequently, the data associated

with each ship design should be captured for re-use on following ship designs. This process needs to be well thought out, resourced, and institutionalized.

5. *Ensure the design workforce is trained and understands the design space, design tools, and supporting data sets*. One way of accomplishing this has been proposed by Jons and Wynn (2009) as part of Continuing Collaborative Concept Formulation (C3F). As Jons and Wynn observed in 2009,

> Continuous concept formulation forges an effective ship design and warfare analysis community, shortens the time to respond to emerging requirements, and produces system cost estimates based on solid engineering. Collaboration enables rapid ways and means tradeoffs for a broad set of possible future environments.

Compared to point-based design (PBD) methods where a baseline concept is chosen early and modified over time, SBD promises to arrive at better designs quicker without a cost penalty. Singer et al. (2017) list the following benefits of SBD:

1. Rework is minimized because decisions are delayed until there is sufficient knowledge to make robust decisions. This is in contrast to other design methods where decisions are made early based on the best (but incomplete) information available at that time.
2. Decisions are made based on a good understanding of the overall design space, not just on the analysis of one or two options.
3. Decisions can be made on partial information. If one domain of expertise finds a region of the design space to be infeasible, that region is infeasible independent of what other domains discover.
4. The different domains of expertise can work semi-autonomously. This enables design teams that are geographically dispersed. Additionally, the overall schedule is less likely to slip if one domain of expertise is late.
5. New information, including changing requirements, can be more readily incorporated into the design process. Good documentation of set reduction decisions can quickly identify the impact (if any) of new information.
6. With the right organization, tools, and experienced workforce, the design process can be accomplished faster than traditional designs.
7. Because options are not selected until proved feasible, the end product should have less technical risk as compared to traditional designs.

Conclusions

If properly implemented, SBD can improve design decisions and the quality of designs in less time than conventional PBD. This paper described the basic method and how it can be applied to design problems. It highlighted a number of points that should be considered in planning and executing SBD. With the information provided, a design team should be able to successfully plan and execute an SBD based design process.

References

Abbott, I., & Von Doenhoff, A. (1959). *Theory of wing sections*. New York, NY: Dover.

Burrow, J., Doerry, N., Earnesty, M., Was, J., Myers, J., Banko, … Tafolla, T. (2014, October 23). *Concept exploration of the amphibious combat vehicle.* Presented at the SNAME Maritime Convention, Houston, TX.

Cloft, P., Kennedy, M., & Kennedy, B. (2018). *Success is assured: Satisfy your customers on time and on budget by optimizing decisions collaboratively using reusable visual models.* Productivity Press.

Cooper, D., & Koenig, P. (2018, November 14–15). Cost estimating concepts and approaches for modular adaptable ship designs. ASNE Design Sciences Series Symposium: Modular Adaptable Ships, Washington, DC.

Office of the Assistant Secretary of Defense for Research and Engineering (ASD[R&E]). (2011, April). *Technology readiness assessment (TRA) guidance*. Washington, DC: Author.

DoD. (2015, February 12). *Manual for the operation of the joint capabilities integration and development System (JCIDS)* (including errata as of 18 Dec 2015). Washington, DC: Author.

Doerry, N. (2015, March 4). *Measuring diversity in set-based design*. Presented at ASNE Day 2015, Arlington, VA.

Doerry, N., Earnesty, M., Weaver, C., Banko, J., Myers, J., Browne, D., Hopkins, M., & Balestrini, S. (2014, September 25). *Using set-based design in concept exploration*. Presented at the SNAME Chesapeake Section Technical Meeting, Army Navy Country Club, Arlington, VA.

Eppinger, S., & Browning, T. (2012). *Design structure matrix methods and applications*. Cambridge, MA: MIT Press.

Garner, M., Doerry, N., MacKenna, A., Pearce, F., Bassler, C., Hannapel, S., & McCauley, P. (2015, November 3–7). *Concept exploration methods for the small surface combatant*. Presented at the World Maritime Technology Conference 2015, Providence, RI.

Gertler, M. (1954, March). *A reanalysis of the original test data for the Taylor Standard Series* (Report 806). Washington, DC: Navy Department, The David W. Taylor Model Basin.

Hoerner, S. (1965). *Fluid-dynamic drag.* Published by the author.

Hoerner, S., & Borst, H. (1985). *Fluid-dynamic lift*. Published by Liselotte Hoerner.

Jons, O., & Wynn, S. (2009, April 8–9). *Continuing collaborative concept formulation (C3F): A proposal to improve ship acquisition and to invigorate naval engineering*. Presented at ASNE Day 2009, National Harbor, MD.

Mebane, W., Carlson, C., Dowd. C., Singer, D., & Buckley, M. (2011, September). Set-based design and the ship to shore connector. *Naval Engineers Journal*, *123*(3), 79–92.

Parker, M., Garner, M., Arcano, J., & Doerry, N. (2017, June 14–15) *Set-based requirements, technology, and design development for SSN(X)*. Presented at

Warship 2017: Naval Submarines & UUVs, The Royal Institution of Naval Architects, Bath, UK.

Shah, J., Vargas-Hernandez, N., & Smith, S. (2003, March). Metrics for measuring ideation effectiveness, *Design Studies*, *24*(2), 111–134.

Singer, D., Doerry, N., & Buckley, M. (2009). What is set-based design? *ASNE Naval Engineers Journal*, *121*(4), 31–43.

Singer, D., Strickland, J., Doerry, N., McKenney, T., & Whitcomb, C. (2017). *Set-based design*. SNAME T&R, 7–12.

Sobek, D., Ward, A., & Liker, J. (1999, Winter). Toyota's principles of set-based concurrent engineering. *Sloan Management Review,* 67–83.

Ward, A., Liker, J., Christiano, J., & Sobek, D. (1995a). The second Toyota paradox: How delaying decisions can make better cars faster. *Sloan Management Review*, *36*, 43–61.

Ward, A., Sobek, D., Christiano, J., & Liker, J. (1995b). Toyota, concurrent engineering, and set-based design. In J. K. Liker, J. E. Ettlie, & J. C. Campbell (Eds.), *Engineered in Japan: Japanese technology management practices* (pp. 192–216). New York, NY: Oxford University Press.

The Effects of Exporting on Defense Acquisition Outcomes: A Quantitative Look at FMS Contracting—Preliminary Findings

Andrew Hunter—is a senior fellow in the International Security Program and director of the Defense-Industrial Initiatives Group at CSIS. From 2011 to 2014, he served as a senior executive in the Department of Defense, serving first as Chief of Staff to Under Secretaries of Defense (AT&L) Ashton B. Carter and Frank Kendall, before directing the Joint Rapid Acquisition Cell. From 2005 to 2011, Hunter served as a professional staff member of the House Armed Services Committee. He holds an MA in applied economics from Johns Hopkins University and a BA in social studies from Harvard University.

Samantha Cohen—worked as a research associate with the Defense-Industrial Initiatives Group at CSIS. Her work focuses on managing and analyzing data to identify relationships among policies, defense spending, and the related impacts on the United States and national security. Her recent research focused on the design and management of international joint development programs, new entrants' survival rates and business graduation in the market for government contracts, and defense acquisition trends. Cohen holds a BS in economics from American University in Washington, DC, and an MS in economics from Katholieke Universiteit (KU) in Leuven, Belgium.

Greg Sanders—is a fellow in the International Security Program and deputy director of the Defense-Industrial Initiatives Group at CSIS, where he manages a research team that analyzes data on U.S. government contract spending and other budget and acquisition issues. In support of these goals, he employs SQL Server, as well as the statistical programming language R. Sanders holds an MA in international studies from the University of Denver and a BA in government and politics, as well as a BS in computer science, from the University of Maryland. [gsanders@csis.org]

Nicholas Halterman—is an intern with the Defense-Industrial Initiatives Group at CSIS. His current work focuses on statistical analysis and machine learning in Python and R. He holds a BA in politics, philosophy, and economics from Oxford University, where he focused on quantitative political methods and political theory. Halterman's previous research was on civilian displacement and charitable foundation behavior. [nhalterman@csis.org]

Contributors:

Maggie X. Wang—was an intern with the Defense-Industrial Initiatives Group (DIIG) at CSIS. In addition to her work at DIIG, which focused on arms transfers data, she has conducted open source research for non-proliferation organizations and political risk consultancies. She is pursuing a BA in government and data science at Smith College and is currently studying at the Graduate Institute of International and Development Studies in Geneva.

Andrew Howe—was an intern with the Defense-Industrial Initiatives Group (DIIG) at CSIS. In addition to his work with international military transfers data, he also works on developing interactive data visualizations. He holds a BA in human evolutionary biology from Harvard University and an MPP from Georgetown University's McCourt School of Public Policy.

Abstract

This paper studies how approaches to security cooperation as well as the characteristics of foreign military sales (FMS) recipients influence defense acquisition outcomes. A review of the literature finds that the lower asset specificity for internationally traded goods, the strength and history of the security relationship, and the quality of partner institutions all are likely influencers of performance. This project has created a unique contract-level FMS dataset, cross-referenced other sources to evaluate the quality of contract reporting, and used to validate economic research regarding the influence of international sales transaction cost.

Introduction

Foreign military sales (FMS) have grown markedly in recent years with major agreements announced during the prior administration, followed by a broad-based push to accelerate and increase FMS by the present administration, which includes revisions to the Conventional Arms Transfer Policy. This trend, shown in

, is more prominent in agreements than deliveries, although the latter have been increasing and jumped in 2017. This FMS drive has multiple sources, not the least of which being a greater emphasis on working by, with, and through partner nations in the National Defense Strategy potentially overlapping with the economic challenges of the global financial crisis and subsequent U.S. spending reductions that reduced defense industry revenues. These twinned motivations are important because arms exports, as recognized by U.S. law, are political and a form of security cooperation, while at the same time having economic and industrial base implications. The political challenges of arms exports have been thrown into sharp relief by the ongoing debate over U.S. support for the Saudi led war in Yemen, as opponents of the war have sought to cut back FMS as a way of adding to the pressure they seek to apply to the Saudi regime.

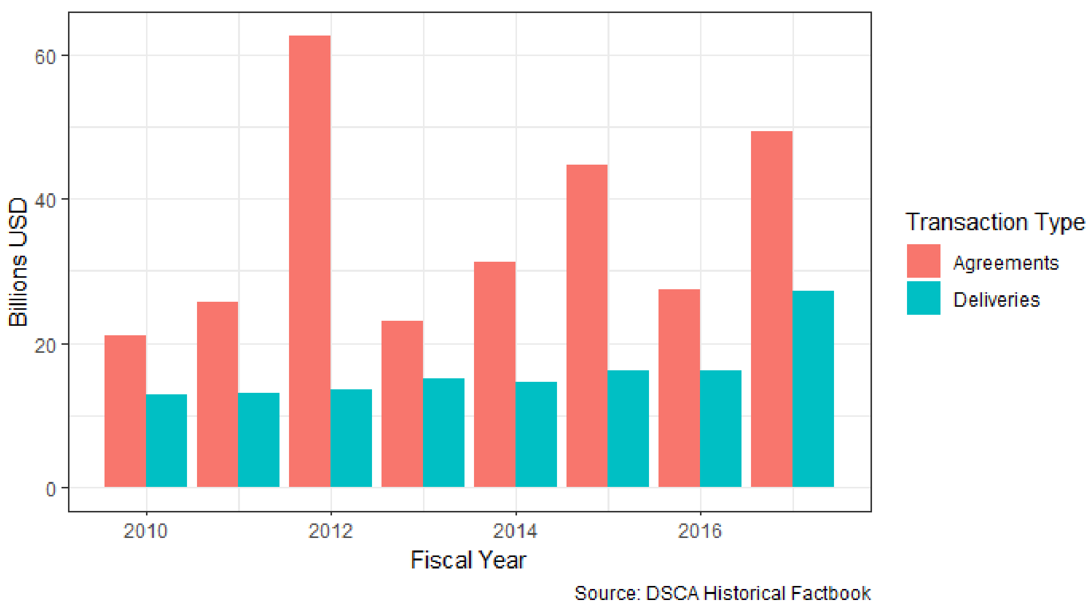


Figure 1. **FMS Agreements and Deliveries by Fiscal Year**

Given FMS's utilization of the U.S. defense acquisition system, and in keeping with laws emphasizing foreign policy considerations in all arms exports, those emphasizing economic and industrial base factors tend to also posit that expanding FMS furthers broad U.S. national security goals. Likewise, those emphasizing deliberation and caution point to the risks of poorly considered deals falling apart, and of the possible proliferation of closely held U.S. technological developments, potentially undermining U.S. national security goals.

The interaction of these considerations means that when considering the acquisition effects of FMS specifically, a wide range of potential influences come into play. On the one hand, the effects of sequestration have incentivized industry to look abroad for revenue growth, and program managers have looked to capitalize on budget savings from overseas sales that can result in lower production costs and shared support costs. On the other hand, arms exports are inherently challenging due to the risk of complications when meeting foreign requirements, instability in international demand, blocks by Congress or the executive branch, and the risk of adverse technology proliferation. This project seeks to evaluate the performance of federal contracts that incorporate FMS. This paper presents the work done to identify the appropriate literature and hypotheses and to build a curated dataset of federal contracts that utilize FMS.

To evaluate the performance of contracts that utilize FMS, the study team first references the existing body of literature that analyzes contract performance and investigates if any papers specifically looked at contract performance for FMS. While the body of contract performance literature is extensive, there are no pieces that empirically analyzed contract performance for FMS contracts. This is likely because the publicly available contracting data from the Federal Procurement Data System (FPDS) is incomplete in indicating whether a transaction was FMS. Thus, a large portion of work done for this paper involves curating a dataset using other fields in FPDS to identify FMS transactions that were mislabeled or unlabeled. This labeling effort includes both application of rules based on transaction funding account and an experiment with labeling using machine learning detailed later in this paper.

While previous work has not examined FMS contracts in particular, the existing bodies of literature provide guidance on theorizing about and measuring contracting performance. Work on security assistance details some inherent challenges of arms exports in meeting foreign requirements and the risk of adverse technology proliferation. Existing work on transaction cost theory provides a foundation to build models that estimate the effects of FMS contract characteristics on FMS contract performance outcomes. Several authors have found that transaction costs, and in particular asset specificity, are a driving force behind acquisition outcomes for services and products (Williamson, 1981; Brown & Potoski, 2003; Adler et al., 1998). Expansion to international markets may reduce asset specificity, as well as creating other economies of scale. Other research, however, has examined how transaction costs are exacerbated in the context of international business (Berghuis & Butter, 2017). This paper will draw on these theories to explore contract performance in the context of FMS.

Scope

To guide the research for this project, the study team posed these four research questions, the first two of which are answered in this paper:

1. How can contracts that utilize FMS be better identified in FPDS using information from other fields?
2. How does FPDS foreign funding data align with the Defense Security Cooperation Agency’s FMS data?
3. Do FMS contracts perform better than non-FMS contracts? This question was subsequently expanded to cover projects incorporating FMS and not just FMS transactions.
4. What variables contribute to the performance of FMS contracts and in what direction and magnitude?

The remainder of this paper expands on the issues raised above by discussing the FMS process, its important role in the defense industrial base, and how contract theory informs the analysis for this project. It also details the methodology used to identify FMS contracts in FPDS, the resulting database, and provides a descriptive overview of the distribution of contract performance metrics for FMS and non-FMS contracts.

Literature Review

Security Assistance and Cooperation

FMS is intended as a U.S. foreign policy tool for strengthening the security of the United States and promoting global security. FMS is authorized under Section 3 of the Arms Export Control Act (AECA) where it is considered as security assistance. The *DoD Security Assistance Management Manual* (DoD 5105.38-M) has a list of eligible countries and organizations who can participate in FMS. An FMS process begins when a foreign customer determines that its military and security needs require a U.S. defense article or service. That foreign government or organization then alerts the U.S. government of its intent to participate in FMS through submitting a letter of request (LOR). From there, the U.S. government organization that is both relevant to the requirement and authorized to receive and process LORs, otherwise known as the implementing agency, works through an interagency process to determine whether the LOR requestor is an eligible participant of the security assistance process under AECA. If so, the implementing agency moves forward in determining an appropriate letter of offer and acceptance (Defense Security Cooperation Agency, 2012, C5.1).

Export Controls and End Use Monitoring

Monitoring and evaluation are essential components of any form of security sector assistance. Throughout security assistance relationships, the United States is able to calculate return on investment, identify and prevent abuse of military resources, and enforce any forms of conditionality on assistance (Dalton et al., 2018, p. 9). In addition to its strategic importance, monitoring is statutorily required under the Leahy laws, which mandate vetting of individuals and units before they receive training or equipment, thereby preventing security sector assistance from going to foreign security forces that commit gross violations of human rights. Beyond the Leahy requirement for end-use monitoring, the Arms Export Control Act (AECA) and International Traffic in Arms Regulations (ITAR) place substantial restrictions and requirements on both foreign military sales and direct commercial sales, including requirements about the eligibility of potential recipient countries and eligibility of platforms and technologies (Gilman, 2014, p. 4). Two separate programs exist to provide end-use monitoring for transfers of military equipment: Blue Lantern and Golden Sentry. Blue Lantern operates under the Department of State’s Directorate of the Defense Trade Controls and monitors use of equipment from direct commercial sales, while the Golden Sentry program is administered by the Defense Security Cooperation Agency and monitors FMS (Fergusson & Kerr, 2017, p. 6). Golden Sentry provides oversight for recipient security and handling of materials, reports any misuse or illegal transfer of equipment, and performs physical inspections and inventory management in some cases (Little, 2017).

Golden Sentry and other end-use monitoring are essential to reducing the risks of transfers by “ensuring that they are not misused and remain within the security force to which they are assigned” (Dalton et al., 2018, p. 10). Alongside concerns about human rights violations and potential proliferation of weapons beyond intended recipients, FMS can increase the risk of harmful strategic behavior by recipients. Capability transfers and the perception of U.S. support create moral hazards for recipient regimes, leading to opportunistic behavior like coup-proofing and power consolidation, both of which can ultimately degrade military capacity and undermine U.S. goals in security assistance (Boutton, 2018, pp. 8–10). These risks, and the monitoring needed to mitigate them, can significantly complicate security assistance and impose meaningful transaction costs.

Despite an increasing emphasis on the economic aspects of FMS in current political discourse, it remains the case that "arms transfers are a foreign policy tool and cannot be wholly separated from U.S. security cooperation policy" (Dalton, 2018, p. 38).

Defense Institutional Capacity

Defense Institution Building (DIB) is an element of security cooperation which has received increased attention in recent years. It seeks to improve security outcomes and mitigate risk of material misuse by increasing institutional capacity in recipient countries to combat the dangers of instability, weak oversight, and poor governance (Dalton et al., 2018, p. 19). DIB is stipulated as an integral part of any security cooperation agreement, as part of the FY 2016 NDAA. The growing focus on DIB and on recipient-country institutions more broadly highlights the fundamentally political aspect of successful security assistance, including FMS. While FMS programs may not themselves involve significant DIB activities, the presence (or lack) of institutional capacity in recipient countries remains a central driver of risk.

Interoperability

The 2018 National Defense Strategy expressed a clear desire to increase interoperability, noting that the ability to "act together coherently and effectively to achieve military objectives requires interoperability" (p. 9). While interoperability includes elements of communication and operational concepts, material overlap between forces can also be a significant contributor to interoperability. As De Vore (2011) argues, "States equipped with the same weapons can support, reinforce, repair, and resupply each other's armed forces without advanced warning" (p. 628). Combined with the shared training and logistic integration that can accompany arms transfers, FMS can provide the material foundation for increased interoperability between U.S. forces and recipient-nation forces. This line of reasoning is echoed in the 2018 National Defense Strategy, which includes as part of its plan for increasing interoperability the need to "prioritize requests for U.S. military equipment sales" (p. 9).

Economics for International Cooperation

The rise in foreign military sales observed at the beginning of this report has been driven not just by security assistance concerns, but also economic factors. The Great Recession put pressure on defense budgets in the United States and Europe while expenditures increased for "several countries—particularly in East Asia, South Asia, the Middle East, and South America" (Gilman, 2014, p. 1). The present U.S. administration prominently featured economic ends in the April 19, 2018 National Security Presidential Memorandum Regarding U.S. Conventional Arms Transfer (CAT) Policy. That document made it a policy of the executive branch to

> increase trade opportunities for United States companies, including by supporting United States industry with appropriate advocacy and trade promotion activities and by simplifying the United States regulatory environment; strengthen the manufacturing and defense industrial base and lower unit costs for the United States and our allies and partners, including by improving financing options and increasing contract flexibility; facilitate ally and partner efforts, through United States sales and security cooperation efforts, to reduce the risk of national or coalition operations causing civilian harm. (Trump, 2018)

At the announcement briefing Dr. Peter Navarro, Assistant to the President for Trade and Manufacturing Policy, discussed these rationales and, when asked about the desire by some buyers for offsets and technology transfer, went further to make the case for jobs and industrial promotion saying, “The organizational culture of the Trump administration is: buy American, hire American” (U.S. Department of State, 2018). With regards to the U.S. industrial base, the most explicit discussion of how economics and industrial issues tie into larger U.S. defense goals is the 2018 Report to the President by the Interagency Task Force in Fulfillment of Executive Order 13805, otherwise known as the Defense Industrial Base Review (IBR). This document points to concerns that prominently feature the first and second order effects from the Budget Control Act of 2011 (BCA) and sequestration (which will be referred to as the defense drawdown henceforth), which helped prompt a greater emphasis on foreign military sales.

The Defense Industrial Base Review

A combination of the 2008 financial crisis, 2011 debt-ceiling crisis, 2011 closing of the Iraq War, and BCA led the domestic demand for defense items to decline. Specifically, the budget caps mandated by the BCA from fiscal year (FY) 2012–2021 were significantly lower than requested funding levels which triggered sequestration in 2013. A previous CSIS study found that the decline in budget carried over to the defense industrial base, which experienced decreased revenue across all platform portfolios:

> CSIS analysis showed that buried within the substantial decline in defense contract obligations were significant variation from sector to sector, with declines varying from catastrophic (Land Vehicles), to steep (Facilities and Construction, Space Systems), to relatively modest (Ships & Submarines). Other sectors suffered a whipsaw effect in which solid business growth suddenly switched to sharp decline (Aircraft). (McCormick, Hunter, & Sanders, 2017, p. VI)

Moreover, medium and large federal vendors experienced the most variance in defense market share and the top companies working with the DoD saw their portfolios shift from R&D to products and services (McCormick et al., 2018). The IBR (2018) also found that sequestration has led to lower defense spending compared to the levels projected before it was triggered.

The IBR has deemed sequestration as one of the five macro forces behind the risks that threaten the U.S. industrial base. The IBR discusses multiple ways in which sequestration causes risks to the industrial base: “inconsistent appropriations, uncertainty about future budgets, macro-level ambiguity in U.S. Government expenditures, and the effects of the Budget Control Act” (IBR, 2018). The IBR argues that successful markets are dependent on predictability, where industries can invest and plan based on informed decisions. That said, Harrison and Daniels (2018) note that while the budget caps drove a gap between Obama administration budget proposals and actual results, the challenges in relying on the DoD’s Five Year Defense Plan (FYDP) long predated the BCA:

> While the FYDP is useful for planning purposes, in the past, it has been a poor indicator of where the budget is headed. As shown in Figure 2, the FYDPs submitted by the Reagan administration greatly exceeded the actual level of funding appropriated by Congress, and the Reagan FYDPs continued projecting growth even when the budget was declining. In the 1990s, the Clinton administration repeatedly projected a lower defense budget than Congress ultimately appropriated. (Harrison & Daniels, 2018, p. 4)

The challenges of predicting did not go away even during the period of single party control of the Congress and the Executive Branch during the 114th Congress. As Harrison and Daniels (2018) note, "While the NDS calls for a 'more resource-sustainable approach' to fund this modernization effort, the unclassified summary of the strategy fails to delineate how it plans to fund its ambitions" (p. 1).

However, all aspects of the present difficulty in predicting the demand for defense goods and services are familiar from prior eras. As noted by the Interagency Task Force's IBR, the spending uncertainty caused by sequestration often results in "peaks of surge and valleys of drought," that disrupt scale production because suppliers can be left with excess capacity during the valleys of drought (p. 21). This can lead to long-term market distortion.

Lastly, the fluctuations in demand caused by the BCA have had rippling effects across defense industry supply chains where companies have struggled in their abilities to hire and retain the necessary skilled workforces for their products and services. While McCormick found that the U.S. subcontracting data was inadequate to fully examine supply chain questions, he did find "the market shock of sequestration and the defense drawdown had a disproportionate effect on Small and Medium-sized vendors" (McCormick et al., 2018, p. 17). The IBR adds that, "Without correcting or mitigating this U.S. Government-inflicted damage, DoD will be increasingly challenged to ensure a secure and viable supply chain for the platforms critical to sustaining American military dominance" (p. 21).

Transaction Cost

Transaction cost theory, as a general approach to understanding economic behavior, lays the foundations for analysis of contracts. As defined by Williamson (1981), transaction cost theory measures transaction costs along three dimensions: frequency, uncertainty, and asset specificity; with asset specificity especially relevant to defense contracting. Minimizing transaction costs is a main driver of municipal governments' decisions to contract services or products, and the type of transaction cost specific to a product or service plays a role in determining contract type. Thus, they are a strong driver of contract design and behavior (Brown & Potoski, 2003; Adler et al., 1998). In the context of military sales, FMS may raise costs for specific transactions due to the difficulties of international transfers, but it may also reduce transaction costs for overall projects by increasing economies of scale and reducing asset specificity. These effects are discussed in turn below.

International Supply Chains

Berghuis and Butter (2017) studied transaction costs in the context of international supply chains and found that international contracting has characteristics that result in high "intangible" transaction costs that require contracts that are more detailed, complete, difficult, expensive, and that need higher-trust relationships. A previous CSIS study found that international acquisition programs "exhibit a greater level of inherent organizational complexity, which poses a range of obstacles…international programs encourage participants to behave opportunistically, face collective tradeoffs that result in sub-optimal end products for individual nations, and experience competing factors within their structure" (Sanders & Cohen, 2017). The study also found cases where the desired benefits were outweighed by adverse effects of international cooperation resulting in negative cost, schedule, and end-product outcomes. Berghuis and Butter (2017) note that these effects vary greatly based on the strength of the relationship between international partners, raising the possibility of measures of "relational contracting" which may result in superior performance.

Offsets

Offsets are a central and contentious aspect of international defense sales. Offsets are accompanying agreements to defense sales which require sellers to provide some economic value to the purchasing country as part of the terms of service. They may be direct, such as a requirement for the seller to purchase components from the buyer country, or indirect, such as a requirement for the seller to purchase or invest in goods or services unrelated to the military sale (Petty, 1999). For military sales conducted through FMS, federal policy is that the "DoD does not encourage, enter into, or commit U.S. firms to FMS offset arrangements" (Acquisitions for Foreign Military Sales, 48 U.S.C. §225.7306). This policy does not, however, prevent U.S. firms from negotiating offsets as part of an FMS sale without direct DoD involvement.

Offsets in international defense sales raise potential issues for domestic economic benefits. Offset agreements may shift economic gains from production to host countries via local co-production or components restrictions, reduce competitiveness through technology and capacity transfers, and ultimately reduce or outweigh some of the economic benefits of FMS (Petty, 1999). Recently, the DoD's stance on offsets in FMS has grown more supportive, including a reduction in oversight of offsets negotiated between contractors and foreign customers (Censer, 2018). Overall, both the transactional burden of negotiating offsets and the potential economic harms to U.S. production pose a theoretical challenge to the economic benefits of FMS.

Asset Specificity

While both international transaction costs and offsets pose challenges to the benefits of FMS, one strong argument for its benefits is the potential effect on asset specificity. For most procurement contracts, producing the final product requires significant investment in capital infrastructure, both physical and informational. Asset specificity refers to the level of specialization for that infrastructure (Williamson, 1981). When infrastructure can be used after contract completion to produce products for the open market or other contracts, the effective cost of investment for the supplier is decreased. When the infrastructure is specific to the current contract, as is frequently the case in the defense industry, the full cost of investment is borne by the supplier for that specific contract. Capital useful for post-contract production is effectively subsidized by that future revenue, while fully-specific infrastructure is not. The degree of asset specificity is therefore a crucial determinant of both contract price and degree of supplier investment. Where asset-specificity is high, infrastructure investment by the supplier is costlier and is thereby disincentivized. This can lead to under-investment and sub-optimal contracts or require costly monitoring and incentives to ensure adequate investment (Schmitz, 2001).

FMS offers a potential boon to the United States government by reducing asset specificity. Asset specificity is high in defense contracting because it is typically a monopsony and requires highly specialized technical capacity, typically leaving suppliers with expensive infrastructure that cannot be reused after a defense contract expires. We should expect this to significantly increase prices: Defense contracts experience high costs to infrastructure investment and require significant incentives (and accompanying monitoring) to overcome those costs and achieve an optimal product (Schmitz, 2001). FMS, however, alleviates the effects of monopsony, and allows for potential asset-reuse after a U.S. government procurement contract expires. While the infrastructure remains specific to a technological product, it becomes viable for use in multiple contracts with multiple recipients. In short, the infrastructure may only be useful for producing F-15s, but producing F-15s for the United States, United Kingdom, and so forth, effectively reduces asset specificity by increasing the applications for the infrastructure.

Notably, this relies on the supplier *expecting* these future contracts. When firms know that FMS will occur, their estimates of asset-specificity should decrease, leading to increased investment and superior outcomes (Schmitz, 2001). This theory suggests that contracts including FMS from the outset with defense exportability features should have lower costs and superior outcomes to equivalent contracts that do not, and that the earlier in the process that FMS is included the stronger these effects should be.

Advantages of Scale

Alongside asset specificity, increases in scale can improve acquisition outcomes through other mechanisms. While asset specificity helps improve outcomes by increasing incentives to invest in up-front capital and training, high production output can help reduce per-unit costs of investment and training. Holding up-front costs constant, each additional unit of production reduces the average per-unit cost until it approaches the marginal cost of each new unit. This economy of scale is central to the effects of monopolies, in which potential harms of market consolidation must be weighed against the benefits of decreasing per-unit cost with increasing scale (Peltzman, 1997). Alongside the declining per-unit cost of infrastructure, increased scale carries benefits through learning curves. As production occurs, involved workers gain experience and tend to discover more efficient techniques, leading to a declining *marginal* cost to production, on top of the declining *average* cost to production experienced for physical infrastructure (Sanders & Huitink, 2019). Unlike in the case of domestic monopolies, FMS does not clearly reduce domestic competition in order to achieve gains in scale, but effectively creates new customers by expanding the potential pool of buyers to foreign governments. This may allow FMS to achieve economies of scale for defense industrial producers without making the traditional tradeoffs to competition experienced in domestic situations.

Economic benefits from decreasing unit cost last beyond the time of purchase. When FMS and domestic procurement run concurrently, economies of scale and learning curve benefits can extend to maintenance, upgrades, and other lifecycle costs, particularly as many FMS products require additional service and parts from the United States after the initial sale. In general, Kirkpatrick (2004) finds that lower per-unit costs are associated with lower lifecycle costs, indicating not only a direct economy of scale to maintenance and parts, but a follow-on effect from reducing initial unit costs. Taken together, these effects offer a theoretical case for FMS lowering per-unit and lifecycle costs, both of which could drive superior acquisition outcomes for programs and platforms which include FMS.

Finally, FMS transactions do not only affect the immediate production cycle, but may have lasting effects on communication, infrastructure, and future projects. The IBR highlights the importance of maintaining and growing defense cooperation agreements with partners and allies to achieve economies of scale and scope as well as interoperability. Specifically, the IBR mentions the FY2017 NDAA's addition of Australia and the United Kingdom to the National Technology Industrial Base (NTIB) as an opportunity to jointly work on industrial base challenges (Interagency Task Force, 2018). The FMS process may help establish and grow defense cooperation by providing U.S. produced materials, ultimately creating the conditions for joint development, DCS, or other forms of security sector cooperation which may carry economic benefits for the U.S. defense industrial base.

Conceptual Framework and Hypotheses

This paper posits that a range of considerations from the security cooperation and assistance domain, as well as traditional economics and contracting literature, have a relationship with foreign military sales contracting outcomes. On both the positive and negative sides of the ledger, strategic and political considerations by the United States and the purchaser nation may influence the level of support for the program and whether it completes delivery at all. Transaction costs literature, organizational complexity, and traditional considerations of scale provide a possible mechanism for these non-economic considerations to influence outcomes as the purchase quantities, supporting institutional infrastructure, and alignment of U.S. and recipients' interests all depend on a variety of factors that can be better measured at the country-level rather than being specific to any given project.

Before testing these hypotheses, the study team had two falsifiable premises to test, which are the focus of this paper. These two premises directly relate to the study's research questions and must be confirmed before the study team can have confidence in the dataset produced as part of this project.

> P_1: Foreign Military Sales data identifiable in the Federal Procurement System correlates with and captures a majority of the spending reporting from other sources.

As will be discussed in subsequent sections, the official labeling of FMS contracts is radically incomplete prior to 2016. The results section includes comparisons of FPDS data with that of the Defense Security Cooperation Agency (DSCA) and the Stockholm International Peace Research Institute's (SIPRI) Arms Transfer database. While both are inexact comparisons, this cross validation is critical to establishing appropriate level of confidence and caveats for use of the FPDS dataset.

> P_2: The pricing mechanism selected for FMS contracts will vary significantly from comparable domestic contracts.

The transaction cost literature emphasizes that acquirers respond to different transaction cost context with different forms of contracts. This observed property provides a useful way to validate the relevant of transaction costs considerations. In addition, even when an FPDS transaction is correctly identified as containing FMS funding, that does not necessarily mean that the entirety of the transaction, let alone the contract, are for an international audience. FMS is tightly integrated into the U.S. acquisition system and practitioners have noted that international customers may only be one funder among many in a large bundled buy.

Transaction Costs

The first hypothesis comes directly from the economics literature and the asset specificity theory that if there is a perceived greater and more widespread potential demand for a product, this incentivizes a variety of investments with positive implications for acquisition performance through decreased transactions cost.

> H_1 **Lower Specificity:** As the number of export agreements for a project increase (decrease), the likelihood of ceiling breaches and terminations decrease (increase).

In exploring this variable, the study team intends to incorporate, where possible, controls relating to other parts of transaction cost. For example, if there is sufficient data on the use of international supply chains, or, less likely, latitudinal data on offsets, then these variables would be worth including to help distinguish asset specificity from other positive or negative influences on transaction cost.

Security Partnership

The next three hypotheses require identification of recipient countries, which the FPDS dataset has not yet achieved. The project team intends to apply machine learning techniques to transaction descriptions. However, that capability is still untested and, even with a hybrid approach including manual identification, may not prove sufficient to the task. The project's back-up plan for recipient identification is to limit the dataset for these variables to those with easily identified projects (e.g., major defense acquisition programs), where the recipients can be more easily determined through other primary and secondary sources including DSCA and SIPRI. In either case, one challenge with this approach is that a given transaction may have a one-to-many relationship with recipients. Once better data is produced, the study team will determine a means of aggregation (e.g., average rating for recipients or looking at the minimum score to identify the weakest link) and will apply this approach consistently across all hypotheses.

The next hypothesis draws more on economic literature than wider foreign policy concerns.

> H_2 **Past Deliveries:** As the number of past bilateral deliveries increase (decrease), the likelihood of ceiling breaches and terminations decrease (increases).

This hypothesis posits that more interactions with the U.S. acquisition system will smooth the path for subsequent cooperation, both in terms of building out the bilateral relationship and improving country proficiency with the sometimes arcane U.S. system. This hypothesis intentionally emphasizes the number of transactions rather than the value of those transactions to put small and large countries on similar footing and also because more routine cooperation, even for less valuable items, may show more about the relationship than high profile projects.

The third hypothesis looks at similar questions but through more of a security assistance lens.

> H_3 **Alliance Status:** As the recipient's integration into the alliances with the seller increases (decreases), the likelihood of ceiling breaches and terminations decrease (increases).

While formal alliances are clearly demarcated, there are some measurement challenges with this variable. For example, Egypt is a major non-NATO treaty ally but is not part of a formal mutual defense pact with the United States while the Rio Treaty includes a score of Western Hemisphere countries including Cuba, which is not known for its security cooperation with the United States (U.S. Department of State, Treaty Affairs, n.d.). That said, the NATO alliance in particular incorporates both collective defense measures and acquisition related provisions and thus some level of differentiation should be possible, perhaps along separate treaty commitment and defense acquisition arrangement axes.

The last hypothesis has perhaps the strongest theoretical justification in the security assistance literature, but will also be a challenge for measure identification.

> H_4 **Institutional Health:** As the health of the recipient's institutions increase (decrease), the likelihood of ceiling breaches and terminations increases (decreases).

This hypothesis has multiple justifications. In political economy terms, more robust institutions reduce the risk of corruption and suggest greater capacity on the recipient's part and a lower risk of process breakdown. Second, institutional strength may prove directly relevant to the Leahy Laws, that restrict arms transfer to units in purchasing countries with a history of human rights abuses. The most direct justification comes from Andrew Boutton (2018), who argues that "that in uncertain political environments—where regimes are driven by internal power struggles and institutions are underdeveloped—military aid can create a dangerous moral hazard" (p. 7). Recipients who believe that their relationship with the provider ensures their security may engage in coup-proofing behavior that undermines the effectiveness of military institutions and may exacerbate grievances within their country.

Data and Methods

Data Sources and Structure

Identifying the Datasets

This paper identifies three primary datasets for studying FMS. The first of these is the Defense Security Cooperation Agency's (DSCA's) *Historical Facts Book*, which provides country-level overviews for arms transfers (2017). This data is available in PDF form, which our team scraped to assemble a dataset tracking country-year level data for FMS agreements and deliveries from 2010 to 2017. The DSCA data does not provide data on individual transactions.

The second dataset is the Stockholm International Peace Research Institute's Arms Transfer database (SIPRI, 2019). SIPRI provides as complete a record as possible of all international arms transfers, including transfers performed via direct commercial sales and transfers from providers other than the United States. SIPRI does not include services. Unlike DSCA, SIPRI provides information on individual transfers, including platform and delivery date. Importantly, due to the variability in pricing between identical platforms, SIPRI does not attempt to provide transaction size in U.S. dollars, but uses a custom Trend Indicator Value (TIV) metric. TIV captures the military significance of the hardware transferred, and is intended for capturing general trends in transfers, not for measuring the discrete dollar size of the transfer. This limits the ability of SIPRI data to be directly integrated with other sources, but it provides extremely valuable directional data on transfers at both the country and platform level.

The final and most substantial dataset is the Federal Procurement Data System's database of all acquisition transactions which use the federal procurement system. FPDS offers extremely granular data on transactions, allowing for detailed breakdowns along types of contract structures, platforms, level of competition, and similar variables. Whether or not a transaction is FMS is recorded in the "foreign funding" field which "indicates that a foreign government, international organization, or foreign military organization bears some of the cost of the acquisition" (USA Spending, 2019). While FPDS provides by far the most granular data on transactions, it suffers two major drawbacks. First, it does not provide explicit information on recipient countries, although some degree of country-attribution may be extracted from plaintext descriptions. Secondly, as shown in ,Figure below, "foreign funding" is only reliably recorded in recent years, with a majority of data before 2015 unlabeled.

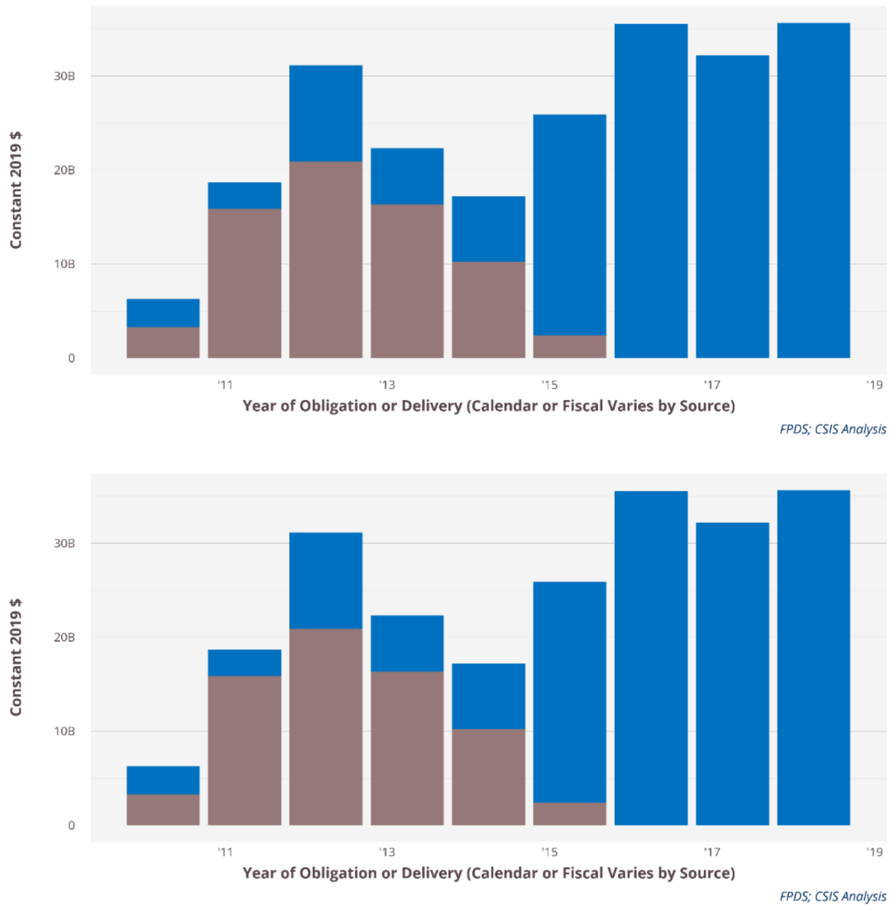


Figure 2. **Limitations of Labeling of Foreign Funding**

Machine Learning

Working with the FPDS data for analyzing FMS involved a significant challenge with missing data. While 2016, 2017, and to a lesser extent 2015 were all reliably coded for foreign funding, in previous years coding was sparse or non-existent. In order to extend any analysis prior to 2015, it will be necessary to create some form of classification process, in which unlabeled FPDS transactions can be classified as either FMS or non-FMS.

To classify the unlabeled historical data, we tested three different classification approaches. The first was a simple set of handwritten rules, in which transactions were labeled based on their agency and treasury account codes. The remaining two classifiers were both machine learning approaches, in which a machine learning algorithm was trained on several million labeled transactions to learn patterns to use in classifying new unlabeled observations. The first machine-learning approach uses a random forests algorithm, which creates a large number of decision trees and aggregates their predictions (Breiman, 2001). The second approach uses deep learning, which creates a series of artificial "neurons"' capable of learning complex patterns (LeCun, Bengio, & Hinton, 2015).

All three classification methods were developed using labeled data from FY 2016 and FY 2017, and were then tested on the entirety of the labeled FY 2015 data. The performance from the hand-rules and random forest models are shown below. Precision captures the rate of true positives from classification (i.e., when the classifier predicts something as FMS, what percentage of the time is it correct?). Recall measures the number of cases captured by the classifier (i.e., what percentage of all FMS transactions did the classifier correctly predict were FMS?). F1-score is the harmonic mean of precision and recall and is a standard overall measure of classifier performance. In all three cases, a score of 1 indicates a perfect classifier.

As shown in Figure 3, the random forest and hand-coded classifiers have similarly strong performance, with manual coding showing slightly greater precision while random forest performs better in recall. Both models correctly label the majority of cases, particularly measured by dollars. So far, deep learning has failed to generalize to the 2015 data, but strong results in the initial 2016/2017 test data indicate potential for improved performance. Existing literature has found that deep learning models outperform random forest models for high-cardinality datasets like FPDS (Guo & Berkhahn, 2016).

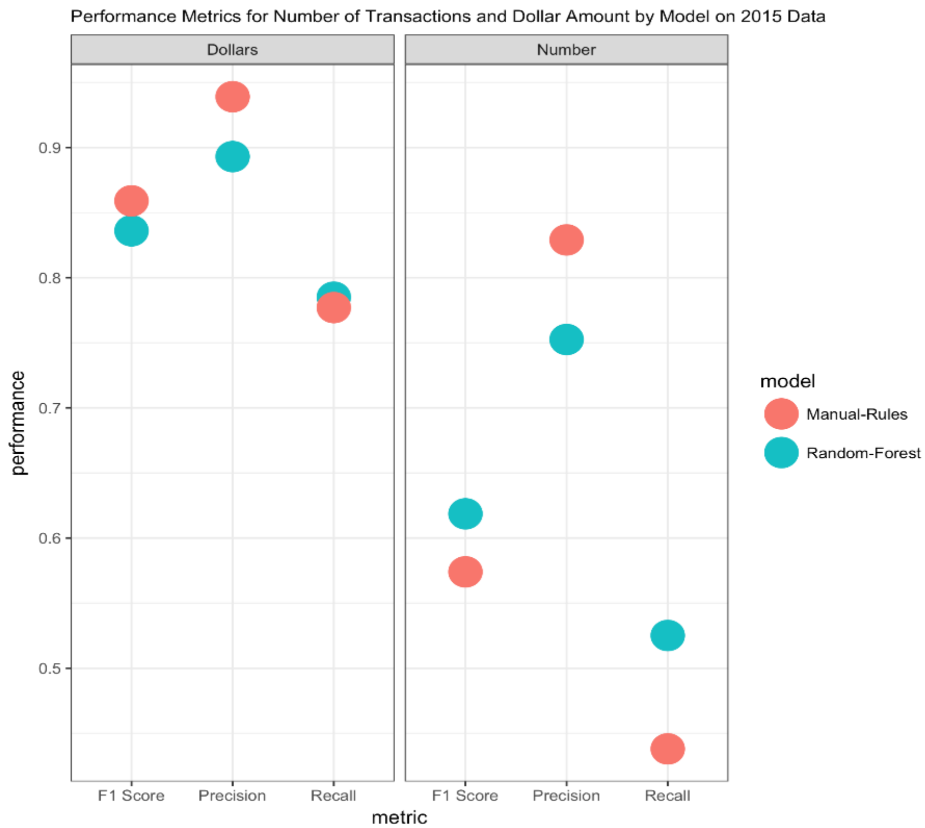


Figure 3. **Classifier Performance**

One additional complication of the FPDS data is the substantial variation in transaction size. Single transactions can range from thousands of dollars to billions of dollars. Because of this variation, it is generally more useful to perform statistical analysis with total dollar figures, rather than transaction counts. However, machine learning algorithms and performance metrics typically operate at the level of observations, not at aggregated values from those observations. In practice, this means that the classifiers discussed here train and measure success based on the number of transactions they correctly classify, not the number of dollars they correctly classify. This volatility in transaction significance makes a strong case for a human-machine hybrid approach, in which a machine learning algorithm is used in a first pass to classify all transactions, and then a human verifies the few largest transactions by hand. In practical terms, this is only possible for transactions that the classifier identifies as FMS; because the vast majority of all transactions are classified as non-FMS, an inordinate amount of hand-vetting would be needed to cause any meaningful change in performance. For positive predictions, however, limited human-checking of the largest predictions can prevent costly false positives.

All three classification methods offer distinct strengths moving forward. The handwritten rules and random forest models already provide strong performance in classifying both transactions and dollars. The handwritten rules have the advantage of being simple and fast to implement, but they are also inflexible, incapable of using more complicated forms of data, and offer limited room for future improvement. Both the random forest and deep learning methods are more complex and more difficult to implement but offer considerable flexibility and room for future improvement. Both models have the capacity to incorporate plain-text descriptions of contracts and add other variables as desired. Deep learning models offer greater flexibility in incorporating text and are able to capture more complex relationships but are less interpretable than the random forest models and thus far have delivered worse performance on historical data. In their current state, random forests deliver the best overall combination of performance, flexibility for future improvement, and interpretability of results.

Future work on classification strategies will include the incorporation of plain-text descriptions of contracting requirements, which should improve performance. In addition, the current handwritten rules rely entirely on treasury account information, which is not available prior to 2012 in a usable format. Both machine classifiers can generate predictions without using that information, though issues of changing offices and similar new-data problems increase for any classification strategy as it moves back in time. Ideally, either random forest or deep learning models will be able to reliably extend FMS classification significantly past 2012. Together, these classification strategies provide methods for significantly expanding existing FMS datasets and enabling granular historical analysis of FMS transactions.

Measurement of Independent and Dependent Variables

While all three datasets have limitations, between them they offer a number of valuable measures for analyzing FMS. DSCA, SIPRI, and classifier-extended FPDS allow for analysis of high-level trends in FMS expenditures over the last several years. These trends are measured via dollar value of total obligations for FPDS, dollar value of FMS delivers for DSCA, and in TIVs for SIPRI.

Both SIPRI and FPDS record the type of arms being transferred, allowing for platform-level breakdowns of trends. However, the two datasets use a different taxonomy of platforms; for instance, SIPRI includes engines as a separate category while FPDS does not. Our team assembled a crosswalk from SIPRI to FPDS by coding the individual weapons platforms in the SIPRI categories which did not match FPDS portfolios. This makes it possible to breakdown SIPRI data into FPDS portfolios, allowing comparison between SIPRI and FPDS at the platform level. Additional work will be required to enable FPDS-to-SIPRI translation.

To analyze the characteristics of FMS transactions, we use a simplified version of FPDS's "Type of Contract" field, which indicates the use of fixed fees, incentives, and so forth. This allows for comparison between FMS transactions and other non-FMS DoD transactions in terms of which pricing mechanisms they use. Future work on contract performance may draw on FPDS's measure of terminations, cost-ceiling breaches, and extent of competition.

Preliminary Results

Comparing Sources

All three datasets exhibit some level of agreement at the year level on general trends in FMS. However, there is a large discrepancy between FPDS and DSCA totals, with FPDS figures generally substantially exceeding DSCA figures. This is especially surprising given the lower quality of labeling in the FPDS dataset. This may be in part due to partially-FMS transactions being labeled as wholly-FMS by FPDS, though hand-verification of the largest FMS transactions in FPDS did not find any mixed transactions. There is also some issue of timing: FPDS, for instance, contains several large FMS transactions in 2012 due to obligations to produce a number of aircraft, while DSCA deliveries may smooth those obligations out as the aircraft are delivered over several years and tend to occur later in time.

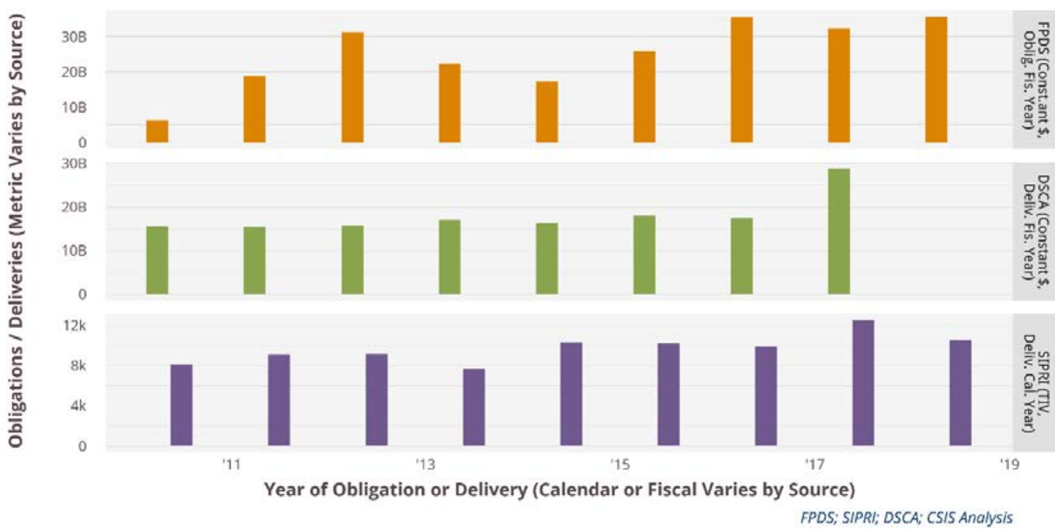


Figure 4. **Annual Funding by Source**

While DSCA and FPDS show some agreement on trends, if not precise amounts, SIPRI appears to deviate from both DSCA and FPDS in year-to-year trends. Some of this is due to the nature of how SIPRI calculates TIVs, however. As shown in Figure 5 breaking out FPDS and SIPRI by platform shows a much greater degree of agreement. Both FPDS and SIPRI show that aircraft dominate U.S. arms exports. They show similarly low and relatively steady rates for ships and submarines and land vehicles. Both FPDS and SIPRI show a sharp and steady increase in Ordinance and Missile exports, though they disagree on the trends for sensors. On the whole, the platform-level analysis reveals a high level of agreement between SIPRI and FPDS, with the disagreements on the aggregate level appearing to be primarily a result of different calculations of aircraft value, possibly due to TIV calculations for the Joint Strike Fighter.

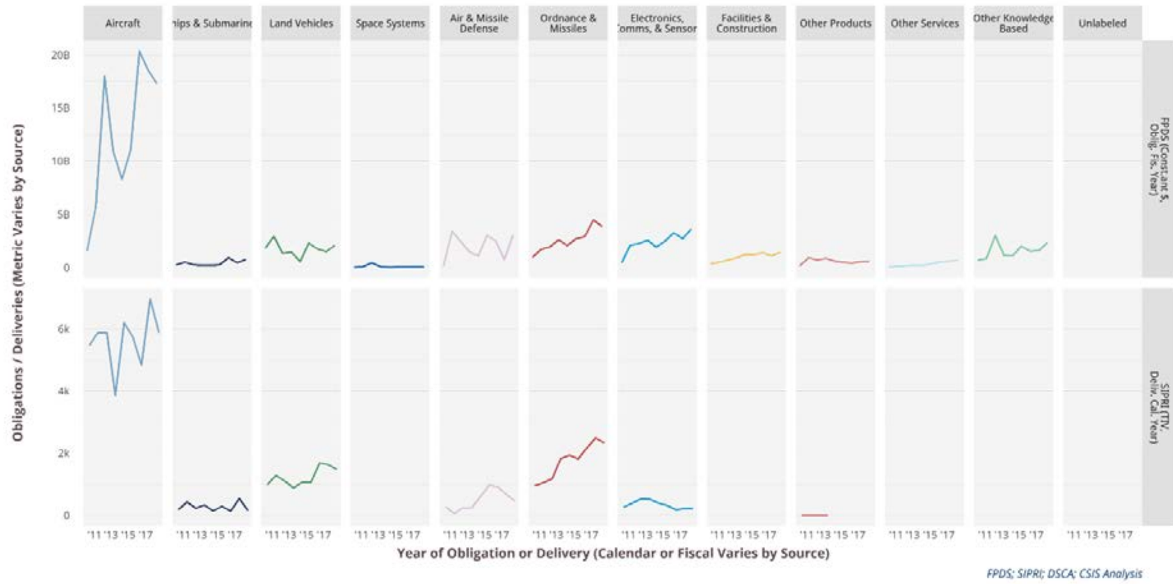


Figure 5. **Comparing FPDS and SIPRI**

Contracting Approaches

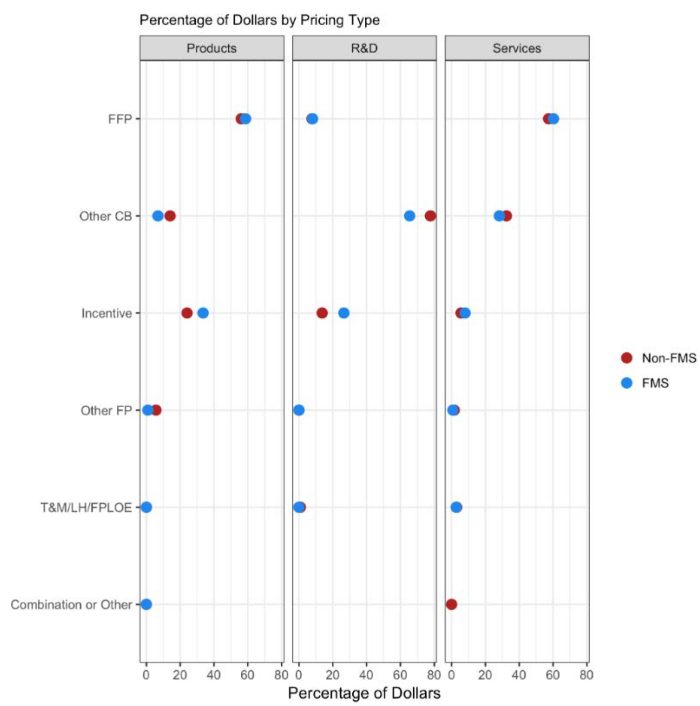
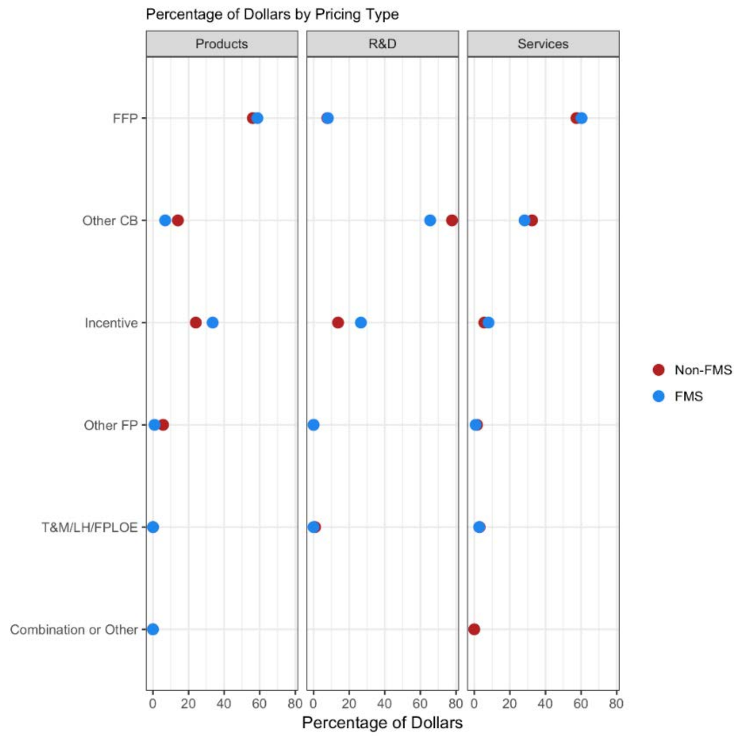


Figure 6 shows the breakdown of contract pricing types for FMS and non-FMS DoD transactions. FMS and non-FMS pricing structures are similar in many ways, especially for service provision. For both products and R&D however, there is meaningful divergence in contract structure in keeping with expectations from the theoretical literature. FMS transactions tend to use incentive-based contracts, specifically fixed-price incentive fee, more frequently than non-FMS transactions. That approach was favored, where appropriate, by the Better Buying Power initiatives and would be in keeping with the use of higher-incentive contracts in the presence of reduced monitoring capacity and higher transaction costs as may be the case in international transfers. Interestingly, the higher use of incentives by FMS contracts does not result in a drop of firm-fixed-price contracts. Instead, FMS transactions tend to use other cost-based mechanisms less often than non-FMS transactions, which may suggest differences in monitoring capacity or degree of trust for domestic sales as opposed to FMS.



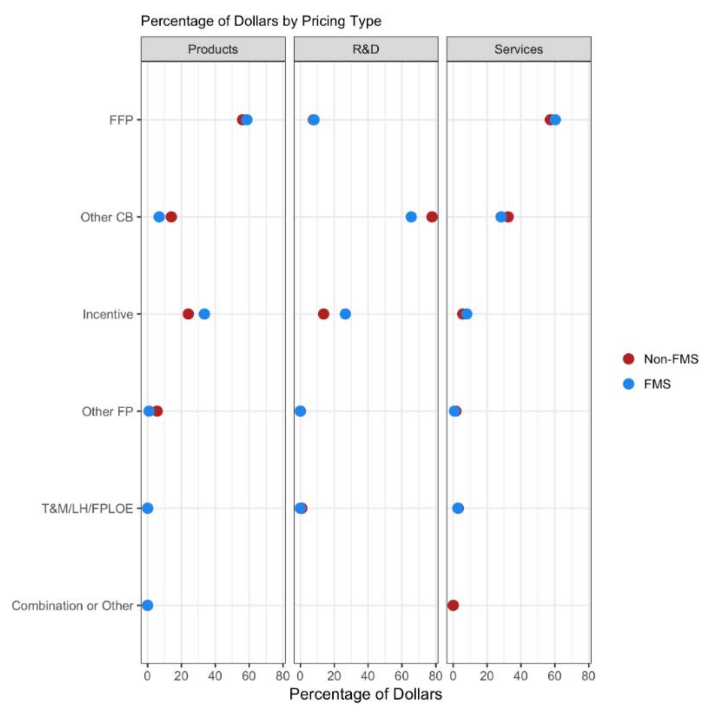


Figure 6. **Contract Pricing for FMS Versus Non-FMS Contracts**

Discussion and Next Steps

The limitations on tracking FMS spending in FMS significantly impede not just the research questions raised in this paper, but a range of other pertinent questions regarding this important and controversial subset of defense contracting. For security sector assistance in particular, assessment, monitoring, and evaluation are watchwords. More rigorous data enables anyone seeking to understand the benefits and risks of present FMS. The biggest surprise thus far in the results is that both in the years that are best labeled, as well as in those that are likely missing some FMS contracts, FPDS obligation levels exceed the deliveries reported by DSCA. The study team will look closely at this issue, examining issues of the timing of obligations versus deliveries, as well as the bundling of FMS and domestically-funded transactions.

This project still has important steps ahead, particularly in the further integration of FPDS, SIPRI, and DSCA data across country and project lines where possible. Taking those steps will better enable the analysis of hypotheses, improves the study team's ability to validate FPDS obligation levels, and enables future researchers and practitioners seeking to better understand the interconnected and high stakes issues surrounding FMS.

References

Acquisitions for Foreign Military Sales, 48 U.S.C. §225.7306.

Adler, T. R., Sherer, R. F., Barton, S. L., & Katerberg, R. (1998). An empirical test of transaction cost theory: Validating contract typology. *Journal of Applied Management Studies, 7*(2), 185–200.

Boutton, A. (2018). The dangers of U.S. military assistance to weak states. *Policy Roundtable: The Pros and Cons of Security Assistance, Texas National Security Review,* 7–18.

Breiman, L. (2001). Random forests. *Machine Learning*, *45,* 5–32.

Brown, T. L., & Potoski, M. (2003). Transaction costs and institutional explanations for government service production decisions. *Journal of Public Administration Research and Theory, 13*(4), 441–468. Retrieved from https://www.jstor.org/stable/3525657

Censer, M. (2018, June 29). Pentagon issues final rule on FMS offsets. Retrieved from https://insidedefense.com/insider/pentagon-issues-final-rule-fms-offsets

Dalton, M. (2018). The risks and tradeoffs of security cooperation. *Policy Roundtable: The Pros and Cons of Security Assistance, Texas National Security Review,* 35–41.

Dalton, M., Shah, H., Green, S. N., & Hughes, R. (2018). *Oversight and accountability in U.S. security sector assistance: Seeking return on investment*. Retrieved from Center for Strategic and International Studies website: https://csis-prod.s3.amazonaws.com/s3fs-public/publication/180207_Dalton_OversightAccountability_Web.pdf

Daniels, S. P. (2018). *Show me the money: Assessing the fiscal reality of the National Defense Strategy's ambitions.* Retrieved from Center for Strategic and International Studies website: http://defense360.csis.org/wp-content/uploads/2018/09/Daniels_Endgame_D360.pdf

Defense Security Cooperation Agency. (2012). *Security assistance management manual: Chapter 5–FMS case development.* Retrieved from https://www.samm.dsca.mil/chapter/chapter-5

Defense Security Cooperation Agency. (2017). *Historical facts book.* Retrieved from https://www.dsca.mil/sites/default/files/historical_facts_book_-_30_september_2017.pdf

Gilman, D. (2014). *Foreign military sales*. Retrieved from Defense Security Cooperation Agency website: https://www.dsca.mil/sites/default/files/final-fms-dcs_30_sep.pdf

De Vore, M. R. (2011). The arms collaboration dilemma: Between principal-agent dynamics and collective action problems. *Security Studies, 20*. Retrieved from http://www.tandfonline.com/doi/pdf/10.1080/09636412.2011.625763

Fergusson, I. F., & Kerr, P. K. (2017). *The U.S. export control system and the export control reform initiative.* Retrieved from Congressional Research Service website: https://fas.org/sgp/crs/natsec/R41916.pdf

Guo, C., & Berkhahn, F. (2016). *Entity embeddings of categorical variables* (arXiv:1604.06737v1). Retrieved from https://arxiv.org/abs/1604.06737

Harrison, T., & Daniels, S. P. (2018). *Analysis of the FY 2019 defense budget.* Retrieved from Center for Strategic and International Studies website: http://defense360.csis.org/wp-content/uploads/2018/09/180917_Harrison_DefenseBudget2019_WEB_FINAL_FINAL.pdf

Interagency Task Force. (2018). *Assessing and strengthening the manufacturing and defense industrial base and supply chain resiliency of the United States: Report to President Donald J. Trump by the Interagency Task Force in fulfillment of Executive Order 13806*. Retrieved from https://media.defense.gov/2018/Oct/05/2002048904/-1/-1/1/ASSESSING-AND-STRENGTHENING-THE-MANUFACTURING-AND%20DEFENSE-INDUSTRIAL-BASE-AND-SUPPLY-CHAIN-RESILIENCY.PDF

Kirkpatrick, D. (2004). Trends in the cost of weapons systems and the consequences. *Defence and Peace Economics*, *15*(3), 259–273. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/1024269032000123203

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521,* 436–444.

Little, T. D. (2017, September 5). End-use monitoring is the key to success in foreign military sales. *Army Sustainment Magazine*. Retrieved from https://www.army.mil/article/192447/end_use_monitoring_is_the_key_to_success_in_foreign_military_sales

McCormick, R., Hunter, A. P., & Sanders, G. (2017). *Measuring the impact of sequestration and the drawdown on the defense industrial base.* Retrieved from the Center for Strategic and International Studies website: https://csis-prod.s3.amazonaws.com/s3fs-public/publication/180111_McCormick_ImpactOfSequestration_Web.pdf?A10C65W9Qkx07VaJqYcJguCH.7EL3O7W

Peltzman, S. (1977). The gains and losses from industrial concentration. *Journal of Law and Economics,* 20. Retrieved from https://www.nber.org/papers/w163

Petty, F. S. (1999). *Defense offsets: A strategic military perspective*. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/a363929.pdf

Sanders, G., & Cohen, S. (2017). *Designing and managing successful international joint development programs.* Retrieved from the Center for Strategic and International Studies website: https://csis-prod.s3.amazonaws.com/s3fspublic/publication/170125_Sanders_IntlJointDevelPrograms_Web.pdf?6CJdjahdu6eaSqqv0vZnfioPn.nrHcgr

Sanders, G., & Huitink, Z. (2019). *Evaluating consolidation and the threat of monopolies within industrial sectors.* Retrieved from the Center for Strategic and International Studies website: https://csis-prod.s3.amazonaws.com/s3fs-public/publication/190208_Sanders_Monopolies_WEB_v2.pdf

Schmitz, P. W. (2001). The hold-up problem and incomplete contracts: A survey of recent topics in contract theory. *Bulletin of Economic Research, 53*(1).

Stockholm International Peace Research Institute (SIPRI). (2019). *Sources and methods.* Retrieved from https://www.sipri.org/databases/armstransfers/sources-and-methods#Contents

Trump, D. J. (2018, April 19). *National security presidential memorandum regarding U.S. conventional arms transfer policy* [Memorandum]. Washington, DC: White House. Retrieved from https://www.whitehouse.gov/presidential-actions/national-security-presidential-memorandum-regarding-u-s-conventional-arms-transfer-policy/

USA Spending. (2019). *Data dictionary.* Retrieved from http://files.usaspending.gov/docs/DATA+Transparency+Crosswalk.xlsx

U.S. Department of State, Bureau of Public Affairs: Office of Press Relations. (2018, April 19). *Briefing on updated conventional arms transfer policy and unmanned aerial systems (UAS) exportpolicy* [Press release]. Retrieved from https://www.state.gov/r/pa/prs/ps/2018/04/280613.htm

U.S. Department of State, Treaty Affairs. (n.d.). U.S. collective defense arrangements. Retrieved from https://www.state.gov/s/l/treaty/collectivedefense/

Williamson, O. E. (1981). The economics of organization: The transaction cost approach. *American Journal of Sociology, 87*(3), 548–577. Retrieved from http://www.jstor.org/stable/2778934

Additional Papers

	Investigating Possible Locations of Undisclosed Subcontractors through Data Analytics of Employment and Federal Contract Records M. Eduard Tudoreanu, Keith Franklin, Ningning Wu, and Richard Wang University of Arkansas Little Rock ***Simulation Modeling for Testing of an Undersea Rescue System*** CDR Nathan Luther, CDR Benjamin Pollock, and Dashi Singham, Naval Postgraduate School ***International Acquisition Best Practices*** Virginia L. Wydler, Erin Schultz, and Kelly Hough, The MITRE Corporation ***A New Way to Justify Test and Evaluation Infrastructure Investments*** Patricia F. Bronson and Paul F. Piscopo, Institute for Defense Analyses ***Issues With Access to Acquisition Information in the Department of Defense: Considerations for Managing Program Data in the Emerging Acquisition Environment*** Jeffrey A. Drezner and Megan McKernan, RAND ***O&M Cost Modeling for the Department of Defense*** Brian G. Gladstone and Karen L. Johnson, Institute for Defense Analyses ***Accelerating Defense Acquisition: Faster Acquisitions Produce a Stronger Force*** Peter J. Modigliani and Daniel B. Ward, The MITRE Corporation ***Networked Logistics and Additive Manufacturing*** Susan M. Sanchez, Gregory E. Lynch, Claudia Luhrs, and Mary L. McDonald, Naval Postgraduate School

Investigating Possible Locations of Undisclosed Subcontractors through Data Analytics of Employment and Federal Contract Records

M. Eduard Tudoreanu—is Professor of Information Science at University of Arkansas Little Rock. Professor Tudoreanu has expertise in human-computer interaction, information quality, advanced visualization of complex data, and virtual reality. He worked on visual data analysis, and has extensive experience in software development and user interface design. Professor Tudoreanu was the founding Technical Director of the Emerging Analytics Center. He has been the keynote speaker at ABSEL 2010, and served as a panelist for the National Science Foundation and Missouri EPSCoR. He earned his Doctor of Science degree in Computer Science in 2002 from the Washington University in St. Louis. [metudoreanu@ualr.edu]

Keith Franklin—is a PhD candidate in Information Science at University of Arkansas Little Rock. He has over 20 years' experience in project management, IT, and Quality. He holds certifications in Project Manager (PMP), Information Security Manager (CISM), Risk and Information Systems Controls (CRISC), Quality Engineering (CQE), and Business Improvement (BIA). He has held management positions at Johnson & Johnson and SIMS Industries Medical, Pfizer (formerly Warner-Lambert division). He earned his Master of Business Administration from City University of Seattle and was honorably discharged from the U.S. Navy. [kxfranklin@ualr.edu]

Ningning Wu—is Professor of Information Science at the University of Arkansas at Little Rock. She received a BS and MS degree in Electrical Engineering from the University of Science and Technology of China and PhD in Information Technology from George Mason University. Dr. Wu's research interests are data mining, network and information security, and information quality. She holds certificates of the IAIDQ Information Quality Certified Professional (IQCP) and the SANS GIAC Security Essentials Certified Professional. [nxwu@ualr.edu]

Richard Wang—is Director of the MIT Chief Data Officer and Information Quality Program. He is also the Executive Director of the Institute for Chief Data Officers (iCDO) and Professor at the University of Arkansas at Little Rock. From 2009– 2011, Wang served as the Deputy Chief Data Officer and Chief Data Quality Officer of the U.S. Army. He received his PhD in information technology from the MIT Sloan School of Management in 1985. [rwang@mit.edu]

Abstract

This paper analyzes data from employment and federal contracts, and it provides a characterization of how contracts affect local employment. It is a continuation of the research presented at last year's Acquisition Research Symposium on searching for undisclosed subcontractors by exploiting the linkages between publicly available employment data and cuts to federal contracts. Many federal and DoD contracts are performed by a team of contracting entities, where some prime contractors rely on subcontractors to execute specific parts of the contract. For many reasons, including national security, privacy, or competitive advantage, some of these subcontractors are not publicly disclosed and have the potential to be unmitigated single stress points in the acquisition process. The paper focuses on gaining a deeper and data-verified understanding of the interactions between federal awards and employment numbers, particularly on the boost to local employment that the start of a large contract may provide. A process for analyzing large federal contracts side by side with employment information is presented. The result of the analysis has found that locations of large Navy awards rank above 70% of other locations in the country in terms of the magnitude of employment changes, under certain industry classification reporting methods.

Introduction

The ecosystem of DoD contracts often involves multiple entities, both large enterprises and small businesses, who work together to achieve the results needed by the DoD. Federal contracts can be awarded to single entities, but also to multiple contractors. Furthermore, contractors rely on other entities, henceforth *subcontractors*, to perform specific parts of the contract. For reasons of security, confidentiality, or competitive advantage, some of these subcontractors are not publicly disclosed. One overarching research question is whether enough publicly-available data exists to allow for the discovery of undisclosed subcontractors.

The open society and transparent government of the USA, through contributions from local and federal government as well as private companies, makes information available on a wide variety of topics, from air quality to social interactions, from federal contracts to employment status. It is unlikely that a single such data repository would allow information about hidden subcontractors to be determined, though the combination of multiple data sources might do just that, especially if the same undisclosed contractor is participating in multiple contracts.

The data-driven approach taken in this paper is to analyze a large number of contract events, specifically the start, end, and any modifications. The reasoning is that one such event—for example, securing a contract to perform work for the federal government—may lead, at least in some cases, to a boost in employment in those locations closely related to the performance of the work. Similarly, the ending of an award or a negative modification may result in a drop in employment. When a large number of events are processed together, it may become possible to hone in on the location of an undisclosed contractor, as shown in Figure 1. This process could potentially be used to determine the likelihood of a location to be home to an undisclosed contractor for a given industry. A number of factors will influence the ability to detect the correlation between awards and employment, but two are worth considering next. First, the larger the award or award modification, the more likely it is to produce an effect on the employment data. Second, employment variations can be better detected in smaller cities and rural areas than in large metropolitan regions.

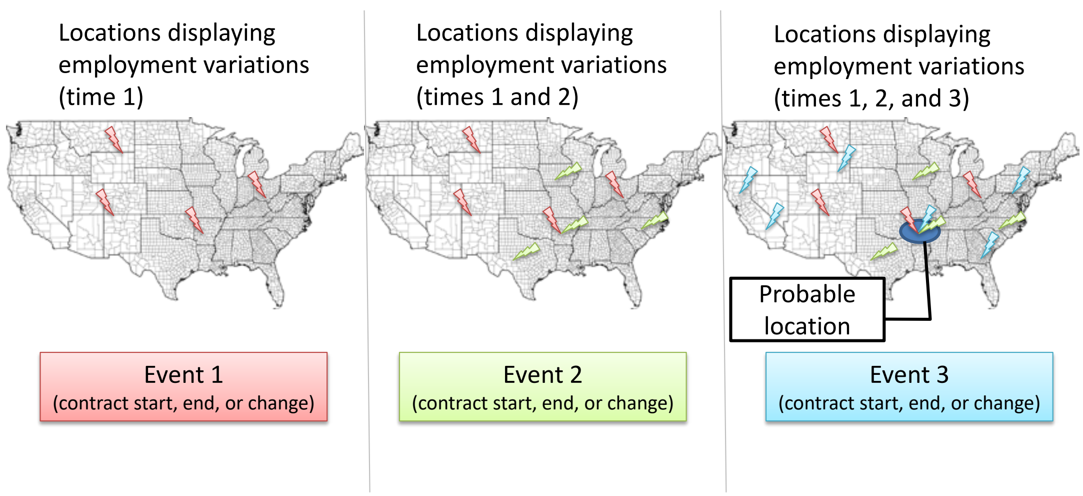


Figure 1. **Determining Possible Locations for Subcontractors Based on Repeated Correlations Between Contract Events and Employment Variations**

Note. Undisclosed contractors who participate in many awards, and thus pose more risk to acquisition, are more likely to be discovered.

The paper focuses on the narrower issue of determining the relationship between known contract events (beginning, end, modification) and employment variations at or around the same time, both

a. at the disclosed locations from the contract, which provides a known relationship that can be reliably assessed, and
b. in the rest of the country, which will provide a measure of the existing noise in the data.

That noise can be further reduced by examining only industries related to each individual contract being studied and by considering employment trends at a local level in the context of the country as a whole.

The paper presents a process to analyze two large open data sources, one about federal contracts and the other about employment in the USA. Based on the datasets derived through the process, this research determined that a promising metric that relies on the magnitude of employment changes at a location and on a classification of industry type. On the datasets for the year 2016, the metric places locations that benefit from large contracts above 70% of other locations in terms of employment. Thus, through repeated elimination, it may be possible to narrow the location of an undisclosed contractor to a manageable number of places in the country, which can then be searched manually for web or social media presence of businesses capable of being an undisclosed contractor.

The rest of the paper covers related work in the next section, followed by an explanation of the methodology used to obtain and process the data. Results, conclusion, and future work are the topic of the last two sections.

Related Work

This paper expands our previous ARP Symposium work (Tudoreanu, Franklin, Rego, Wu, & Wang, 2018) that used a manual correlation process and revealed that reductions (modifications that cut the amount of money originally allotted) in contracts, which were large relative to the contractor's size, were correlated to a drop in employment in more than two-thirds of the cases. Location quotient (LQ), which provides a relative measure of a region's employment in an industry sector relative to the nation, showed a correlation for 75% of the studied contracts' reduction. The main difference between this research and the past is the addition of positive contract modifications (including contract start date) into the analysis, the use of data science to be able to handle and process tens of thousands of Navy contracts, and the examination of the noise present in employment data.

Previously, policy makers and researchers have recognized the need to employ data as a multifaceted means of increasing the agility of the acquisition process (Krzysko & Barney, 2012). To this end, research has looked at automatic means of dealing with the heterogeneous acquisition data sources from text processing (Zhao et al., 2015), systems engineering (Cilli et al., 2015), and business (Gaither, 2014) perspectives. Our paper is different both in content and in the approach: in content, because we seek to characterize the hidden flow of funds in the supply network, and in the approach, because our expertise in data quality and data science provides a more value-based perspective.

A recent approach used the data collected by government agencies. The objective was to use administrative data, that is, state unemployment insurance information that is from covered wages and salaried workers based on the workers quarterly earnings. In studying domestic outsourcing, the primary constraint was how existing data is limited in estimating the number of workers who are employed by contractor companies or who provide services to firms as independent contractors. This data "can be used to document employment in contractor firms and the number of independent contractors, link contractor industries with the firms using their services" (Houseman & Bernhardt, 2017). The paper also recommends using earnings information that includes wages, salary employment, and self- employment tax data that can be provided by the IRS.

A second paper "uses linked worker-firm administrative tax data from U.S. tax returns to explore the changing relationship between firms and independent contractors" (Miller, Risch, & Wilking, 2017). A dataset was constructed, which used digitized tax filings from the IRS for the tax years of 1997 to 2015. Individuals were linked to their employer via reports Form 1040, Schedule C and Schedule SE, and information reports W-2, 1099-MISC, and 1099-K.

There have been other researchers who successfully integrated publicly available data with private information, but their goal was radically different than the scope of this paper. Such data integration efforts have been employed to reconcile corporate names (Gayo et al., 2013) and to provide a tools for accessing unified corporate names (Llorens, Rodriguez, & Vafopoulos, 2015). Another approach involved government data enhanced with information services from outside sources to provide additional, generalized context into the data (Felten et al., 2009).

Methodology

Two data sources were used in our analysis, Federal Procurement Data Systems–Next Gen (FPDS) and Bureau of Labor Statistics (BLS). The procurement database provides a list of federal contracts awarded over the years, including any modifications to an award. The labor statistics includes employment data on a monthly and quarterly basis, both in absolute numbers and relative to the rest of the country, through the location quotient (LQ, 2008). Due to the limited amount of time and computational resources, the primary focus was limited to the year 2016, yet even this approach resulted in some operations involving tens of millions of entries and long processing times. The methodology described here is applicable to the analysis of additional years.

Data from the two sources was copied to a local database hosted on a relatively powerful server powered by an 8-core, dual Intel Xeon processor with 48 GB of memory. Java programs and MySQL queries were used to populate the database.

The data processing steps are listed below, along with an explanation of the types of data involved in those steps. The end-goal is three-fold:

a. to determine whether a large contract correlates with an increase in employment given the contract's industry and location,
b. to determine the overall behavior of employment in the entire United States, and
c. to compare the employment at the contract's location to all the other locations in the country at the same time (of the year) and industry in order to be able to determine a metric that has the potential to uncover hidden subcontractors.

To this end, steps 3 and 4 are split on two separate pathways, one to examine employment behavior strictly related to FPDS contracts, and one to study the employment trends in the entire country.

Step 1: Select Contract Data From FPDS

The analysis of contract data started with FPDS records for the Department of the Navy, and it underwent two selection steps. First, only contracts and modifications with a start or end date in 2016 were considered. Second, the analysis filtered "small" contract events (begin, end, or modification events), which in this case were events with dollar amounts in the range of –$99,000 and $499,000. That is, only changes to a contract that either reduced the amount awarded by more than $99,000 or had more than $499,000 were considered because they have a larger chance of producing layoffs or hiring. This step resulted in over 23,000 contracts (and modifications) being selected.

While a wealth of information is available for each contract, the relevant data to be used for the rest of the analysis includes the following:

- dollar amount;
- zip code of the principal place of performance, which can provide the local area in which to examine employment;
- start date, used as the effective date when the contract could start affecting employment;
- NAICS code, which is a standard way of classifying various types of industry in the United States and Canada (NAICS, 2017). The codes in FPDS are six digits long. NAICS uses a somewhat hierarchical structure in which related sub-industries share the first few digits of the larger industry type. For example, a four-digit ABCD code is generally a sub-type of the three-digit ABC industry.

Step 2: Employment Data From Bureau of Labor Statistics (BLS)

Employment data included both monthly and quarterly metrics. Due to the nature of our search, the year 2016 was selected because it was already in our systems. The last quarter of 2015 and the first of 2017 were also considered for some of the steps below. The number of quarterly entries in the year 2016 alone is over 14 million (that computes to over 52 million monthly entries) for all covered locations and industry codes.

The following fields were deemed important:

- month of the year to determine the time frame;
- industry type, which was provided as NAICS (2017) codes. Three-, four-, five-, and six-digit industry codes were used (see step 1 for an explanation of the NAICS code length);
- U.S. county and state; and
- number of people employed in the county per industry type and month. Note that not all counties have employment for all NAICS codes.

Step 3: Contract Pathway: Filter Relevant BLS Data Based on Large Contracts

From this point, two pathways were pursued in the processing: one that focused on employment relevant to existing large contracts, henceforth termed *Contract Pathway*, and one focused on the overall employment situation in all locations in the United States, named *All Locations Pathway*.

These steps reduced the size of employment data (step 2) to only those locations that appear in at least one of the large contracts selected at step 1. The zip code of the place of performance of a contract was mapped to county and state. All monthly employment information about that location was copied in a working data store regardless of NAICS or time of year. All of 2016 and the last quarter of 2015 and first of 2017 went through this transformation. For the year 2016, only about 2 million quarterly (that is 6 million monthly) entries were excluded.

Step 3: All Locations Pathway: Filter Relevant BLS Data Based on Large Contracts

While it is important to analyze the employment trend in the country as a whole, there is little reason to consider NAICS codes that are completely unrelated to any of the large contracts in the other pathway because employment performance in one part of the economy (say, agriculture) is not directly correlated to performance in another type of industry (say, information technology).

We designed an easily measurable definition for what it would mean for two industries to be related. The FPDS data provides a six-digit NAICS code for the product that is the subject of the contract. BLS uses the NAICS code slightly differently in that BLS identifies the type of employer. It is possible that one employer may have products in different, yet related industry types. Our definition of related industries relies on the hierarchical nature of NAICS. Thus, in addition to the six-digit code known from FPDS, we also considered the often more-encompassing five-digit code obtained by removing the last digit, as well as the four- and three-digit ones. Thus, for each FPDS contract, we created a bundle of four related industries by the process of removing last digits of the code.

Given our definition of related industry types, this step removed all employment information for NIACS codes that do not belong in an industry bundle from at least one of the large contracts. Note that this automatically excluded codes with fewer than three or more than six digits. The result was a dataset of about 4.8 million quarterly (14.4 million monthly) employment entries. No data from years 2015 nor 2017 was included.

Step 4: Contract Pathway: Calculate Employment Before and After Contract Start Date

The start date of the contract may be a possible event that can lead to observable changes in employment levels. It is unlikely that any business would start hiring before the contract begins, and thus before funds become available. One good temporal point to be considered as the reference employment is the month right before the start date. We will refer to it as the *before* employment level. Similarly, after the contract is awarded, the business may hire additional people to perform the work. The hiring may actually take some time, so we decided to examine the contract start month and two more months after for any changes in the employment. Positive dollar amounts might lead to higher employment; therefore, we use the maximum employment level reported in any of the three months (start month and the immediately following two months) as our *after* employment metric.

The before and after levels were computed for each contract and for each NAICS code from the related industry bundle (see step 3 All Location Pathways for an explanation of the bundle). An entry in this dataset consists of the following:

- contract data (dollar amount, location, start date);
- NAICS code (either a three-, four-, five-, or six-digit code). Not all combinations of location, code, and month have reported employment data;
- local before and after employment levels.

The dataset has over 53,000 entries, which include the multiple industry bundles, with about 16,400 unique contracts/modification. Some of the contracts from step 1 were not included because the employment information was not available for that location and time of the year either before the contract start date, or after, or both.

Step 4: All Locations Pathway: Calculate Employment Before and After Each Month

A similar technique was used to calculate employment levels in every county in the United States for months starting in February and ending in October 2016. The before level is the one from the month before the month being calculated (thus the first one is February), and the after level is the maximum of the three months following, and including the target month (thus the last possible target month is October because it requires November and December data).

The dataset has the following fields, with more than 3 million entries. Note that it is possible that employment level was not reported for some locations, NAICS, and month combinations:

- month;
- NAICS code;
- county;
- before and after local employment levels.

Step 5: Combining the Two Pathways

This final step of data processing involves comparing each contract entry from step 4, Contract Pathway, to each possible location in the United States (step 4, All Locations Pathway). All possible combinations were generated as long as they had the same NIACS code and the same month. The total exceeds 52 million combinations. In additional to being able to perform a pair-wise comparison, we also determined the national average before and after employment levels at the time of each contract.

Results

This section presents two broad metrics for employment levels. First, it just examines whether the employment registered an increase as shown by the *before* level being smaller than the *after* numbers. Second, the magnitude of the change, whether increase or decrease, from *before* to *after* is presented. For this section, we focus on positive contract events, particularly a contract's start date and any positive modification to the dollar amount. The majority of the contracts, around 13,900, from step 4, Contracts Pathway, fall into this category.

The country overall is experiencing an increase in employment during the studied period. Based on step 4, All Locations Pathway, in 65.2% of the counties and NAICS codes over the year, the employment records show an increase. For the locations, times, and NAICS codes related to contracts, that number is only slightly larger at 68.8% (from dataset produced at step 4, Contracts Pathway). The results are depicted in Figure 2.

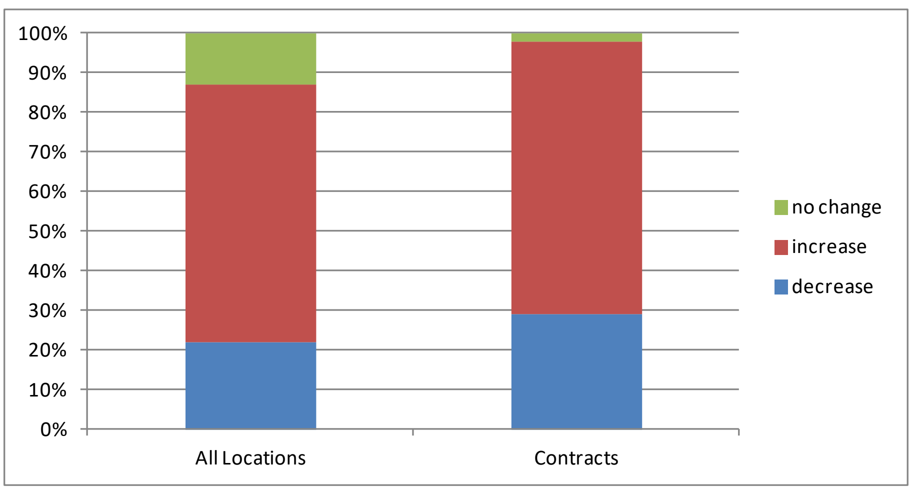


Figure 2. **Relative Changes in Employment as Recorded in the Datasets Produced in Step 4**

Note. Both the country as a whole over the year 2016 and those locations affected by awarded contracts show similar behavior in the number of instances where an increase is recorded (as a percentage of all entries in each dataset).

A change in employment in itself does not seem to be sufficient to allow the discovery of an undisclosed contractor. Thus, a second metric, the magnitude of the employment change, was considered. Formally, the magnitude is obtained by subtracting the *before* number from the *after*. The magnitude can be negative if the employment level drops. Using the data from step 5, the magnitude in the counties related to a contract, at the start date of the contract, and for the relevant industry bundle was compared to the magnitude of change for the same industries and at the same time for each and every county in the US.

Using the dataset from step 5, the locations of an awarded contract have a larger magnitude of employment increase when compared to all other U.S. locations (about 35 million out of the 52 million possible pairs). Figure 3 shows the results broken down by the length of the NAICS code, and it can be seen that for the three-digit NAICS code, the relative percentage between larger contract locations (red) and lower ones (blue) is the highest at 70.3%.

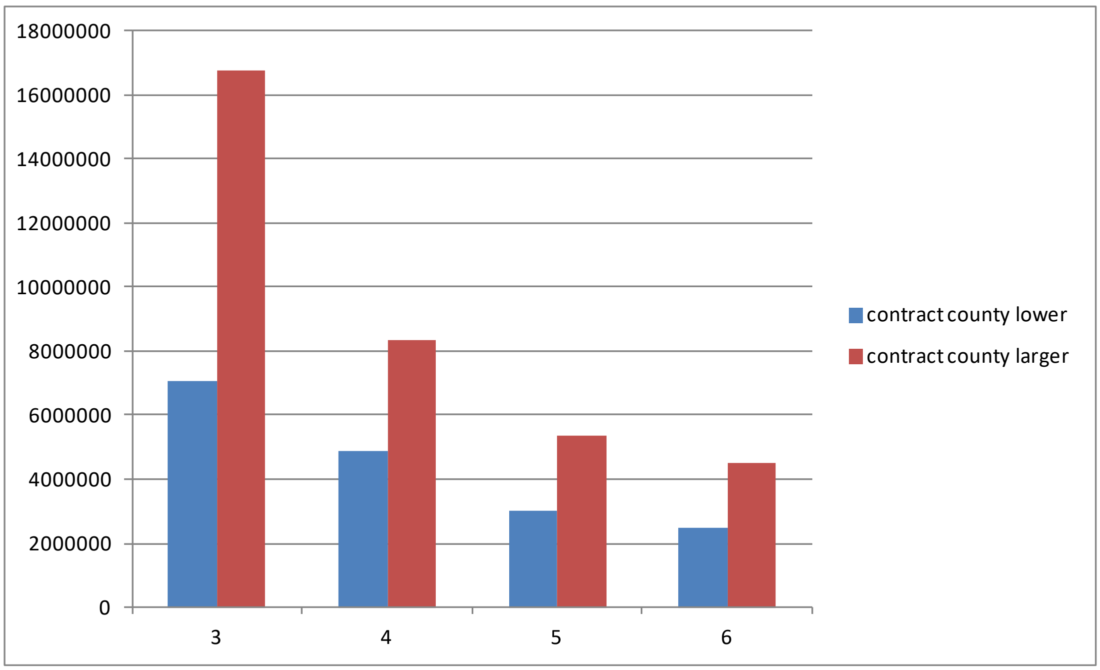


Figure 3. **Comparing the Magnitude of Employment Changes on Contract Locations, Times, and NAICS Code With All Other Recorded Locations in the U.S.**

Note. The bars' height shows the number of pairs in which the magnitude of employment change in the contract county is larger (red) or lower (blue) than some other county's. The x-axis breaks down the numbers by length of NAICS code.

Conclusion and Future Work

The paper employed large data analysis and found that the magnitude of employment changes is higher in 70% of places of performance for a federal award as compared to the rest of the country for three-digit NAICS codes. This finding is promising as a metric to help uncover hidden contractors because it has the potential to eliminate more than two-thirds of U.S. locations. It can be used repeatedly for many contracts that use products and services in related industries to assign a probability for various potential locations of undisclosed subcontractors. Such undisclosed contractors are typically employed in contracts with a larger scope to achieve confidentiality, security, or a competitive advantage. Depending on the situation, acquisition experts may need additional planning to protect such hidden contractors if security is desired, or may rely on data science to identify these contractors and avoid them becoming a weak link in the acquisition process.

The main contributions of this work are the development of a data science process for joining large acquisition and employment datasets and the development of a potential metric based on using the magnitude of employment changes in three-digit NAICS code industry. Future work will focus on three main thrusts:

a. running simulations where one contract is picked at a time, and the most likely locations of that contract are determined based on the three-digit NAICS code metric presented in this paper;
b. considering additional years, especially one where the employment is declining, which would be the opposite of 2016, a good year for employment; and
c. determining additional probability metrics for possible locations for contractors. Good candidates may be combinations of NAICS codes and location quotients (LQ, 2018).

References

Cilli, M., Parnell, G. S., Cloutier, R., & Zigh, T. (2015). A systems engineering perspective on the revised defense acquisition system. *Systems Engineering*, *8*(6), 584–603.

Felten, E. W., Robinson, D., Yu, H., & Zeller, W. P. (2009). Government data and the invisible hand. *Yale Journal of Law and Technology*, *160*.

Gaither, Carl C. (2014). Incorporating market based decision making processes in defense acquisitions. *International Journal of Defense Acquisition Management*, *6*, 38–50.

Gayo, J. E., Pablos, P. O., Rodriguez, J. M., & Vafopoulos, M. (2013). Towards a stepwise method for unifying and reconciling corporate names in public contracts metadata. The CORFU technique. Seventh Metadata and Semantics Research Conference Track on Metadata and Semantics for Open Repositories, Research Information Systems and Data Infrastructures. Thessaloniki, Greece.

Houseman, S., & Bernhardt, A. (2017). *Memorandum: Data needs for research on domestic outsourcing in the United States*. Retrieved from https://research.upjohn.org/staff_publications/1

Krzysko, M., & Baney B., (2012). The need for “acquisition visibility.” *Journal of Software Technology*, *15*(1), 4–9.

Llorens, L., Rodriguez, J. M., & Vafopoulos, M. (2015). Enabling policy making processes by unifying and reconciling corporate names in public procurement data. *The CORFU Technique* (Vol. 41, pp. 28–38). Madrid, Spain. Computer Standards & Interfaces.

Location Quotient (LQ). (2008). *Bureau of Economic Analysis.* Retrieved from https://www.bea.gov/faq/index.cfm?faq_id=478

Miller, A., Risch, M., & Wilking, E. (2017). Independent contractor or employee? The changing relationship between firms and their workforce and potential consequences for the U.S. income tax.

NAICS. (2017). North American Industry Classification System. U.S. Office of Management and Budget. Retrieved from https://www.census.gov/eos/www/naics/2017NAICS/2017_NAICS_Manual.pdf

Tudoreanu, M. E., Franklin, K., Rego, A., Wu, N., & Wang, R. (2018). Searching hidden links: Inferring undisclosed subcontractors from public contract records and employment data. In *Proceedings of the 15th Annual Acquisition Research Symposium* (Vol. 2; pp. 491–499). Monterey, CA: Naval Postgraduate School.

Zhao, D., MacKinnon, D., Douglas, J., Gallup, D., & Shelley, P. (2015, September). *Lexical Link Analysis (LLA) application: Improving web service to Defense Acquisition Visibility Environment* (NPS-AM-15-129). Monterey, CA: Naval Postgraduate School.

Simulation Modeling for Testing of an Undersea Rescue System

CDR Nathan Luther—Master of Systems Analysis Program, Naval Postgraduate School, Monterey, CA

CDR Benjamin Pollock—Master of Systems Analysis Program, Naval Postgraduate School, Monterey, CA

Dashi Singham—PhD, Operations Research Department, Naval Postgraduate School [dsingham@nps.edu]

Abstract

Undersea Rescue Command (URC) can mobilize its people and equipment worldwide to conduct a rescue of personnel from a disabled submarine stranded on the sea floor up to depths of 2000 feet of seawater. In 2019, URC anticipates reaching Initial Operating Capability on a new system, Submarine Rescue System-Transfer Under Pressure, which allows survivors to remain under pressure throughout the process of being rescued. Modeling and simulation provides an opportunity to validate the procedures for the rescue before URC implements them in the real world. This study tested current URC procedures and offers recommendations for when to use different decompression policies, and analyzes the types of rescue delays to expect under the new system.

Introduction

Simulation modeling can be an effective way of testing the performance of potential systems before they are implemented. One main benefit is that the performance of many potential system configurations can be estimated using computer modeling, while it may be difficult or expensive to test such configurations on the actual system. Simulation has been used to provide analysis in numerous sectors, for example, healthcare, energy, defense, financial, and technology. Computing resources continue to become increasingly available, and simulation is becoming more popular as a tool for conducting analysis.

In particular, simulation is being increasingly used in test and evaluation for defense systems (Giadrosich, 1995; Marine Corps Operational Test & Evaluation Activity [MCOTEA], 2013). Simulation can be used in the prototyping stage to determine potential configurations with good performance. It can also be used in developmental test and evaluation to troubleshoot and determine whether the system will meet test requirements. Simulation can be used to determine the potential feasibility of a system without resorting to expensive physical testing. Even if operational tests are eventually required to ensure the system performs as expected, simulation can be used as a precursor to identify potential problems or improvements to be made.

This paper describes a research project that employs simulation to model the complex process of undersea rescue. In particular, the simulation model studies a new proposed system and compares different policies for operating the system. The research team worked directly with experts and operators of undersea rescue processes to build and evaluate the simulation model, and then used statistical analysis methods to evaluate different policies to answer research questions set out by the undersea rescue community.

Undersea rescue, like many other defense processes, can involve a high degree of uncertainty. The goal is to find the best policy that performs well given uncertainty in how specific model components may perform. Stochastic simulation programs are specifically designed to incorporate uncertainty, and we employ analysis methods here to compare different operational policies given this model uncertainty.

Undersea Rescue Process Simulation Model

Undersea Rescue Command (URC) can mobilize its people and equipment worldwide to conduct a rescue of personnel from a disabled submarine (DISSUB) stranded on the sea floor up to depths of 2000 feet. In 2019, URC plans to reach Initial Operating Capability on a new system, the Submarine Rescue System-Transfer Under Pressure (SRS-TUP) system. SRS-TUP will allow survivors from the submarine to remain under pressure throughout the process of being rescued from the time they exit their submarine into the rescue vehicle, up to the deck of the surface ship where URC's two submarine decompression chambers (SDCs) are located. They are transferred from the DISSUB to the rescue ship and into these chambers using a Pressurized Rescue Module (PRM). This pressurized transfer reduces the likelihood that survivors will suffer from decompression sickness or other decompression-related complications.

URC has initial procedures for its use, and cannot yet conduct real world testing on the system to validate that its procedures minimize expected rescue delay times and maximize overall rescue effectiveness. This study helps to verify these procedures by performing modeling and simulation of rescues at a wide variety of depths and DISSUB internal pressures.

During a rescue, there are two main policy options to consider. The PRM can bring 16 survivors up from a DISSUB per sortie, but the SDC can hold up to 35 people. URC decision makers must decide whether to start decompression after each rescue vehicle sortie or whether to wait until another sortie arrives before starting decompression. At higher internal pressures in the DISSUB, the decompression timeline becomes the limiting factor in the rescue, making it more critical to maximize the number of survivors in each decompression. Current procedures state that decision makers should expect decompression after each sortie method to result in no delays in the overall rescue unless internal pressure on the DISSUB exceeds 60 feet seawater (fsw).

The goal is to build a simulation to model the process of rescuing survivors from a pressurized disabled submarine. There are constraints on the number of survivors that can be transported at a given time. The URC will likely provide several rescuers on-board the DISSUB to assist with the rescue, known as a DISSUB Entry Team (DET). Additionally, the PRM requires two attendants for operations who breathe the same pressurized air and require decompression. Based on the length of the attendant's exposure, they may be able to conduct more than one sortie, but require a "clean time" between decompression and recompression, and there are limits on the number of sorties, or amount of pressure they can be exposed to more than once. There are also aspects of the model that are highly variable which are modeled in the simulation. One aspect is the time for different events to take place, like loading/unloading personnel from the modules, or the time to transport survivors from the DISSUB to the surface ship. Incorporating this uncertainty in a simulation model allows for different policies to be tested to see which ones perform best under unpredictable conditions.

Model Objectives

There are two possible decompression policies to consider when there are two available SRS-TUP chambers, and the analysis in this research guides when to use each of these policies:

- Alternate use of the two SDCs after each sortie. As each sortie arrives, it will unload its survivors in one of the SDCs, alternating between the two, and decompression will commence after each sortie.
- Alternate use of the two SDCs once each is full. As each sortie arrives, it will unload its survivors in one of the SDCs, with subsequent sorties unloading to the same SDC until it is near or at capacity, at which point decompression will commence.

The goal of the study was to determine what resources or policies are needed to execute a successful rescue as quickly as possible. The following were the two key research questions:

- When should each decompression policy (decompress immediately after each sortie, or only after the SDC is full) be used? How does that vary for different DISSUB internal pressures?
- How many PRM attendants are required to meet manning requirements to avoid creating any significant rescue delays?

These questions lend themselves to a simulation-based analysis because there are multiple options for employing the SDCs depending on expected sortie and decompression times. URC has procedures for SRS-TUP employment but lacks data demonstrating that those procedures are likely to produce the best rescue outcomes. As this specific system has yet to be fielded, there is no existing data set to analyze. Additionally, modeling and simulation provide a much larger data set over a range of DISSUB depths and internal pressures than could reasonably be achieved through real world testing.

Experimental Setup

The simulation model can vary two types of variables: decision variables and noise variables. Decision variables are those that must be chosen by the analyst in operating the system, and usually the analyst is trying to optimize the choice of decision variables. For example, the analyst may be using the simulation model to determine how many people to staff at a given station, or which routing pattern to use for aircraft or vehicles. Noise variables are uncertain variables that are uncontrollable by the analyst but must be modeled because they affect the performance of the model.

In this study, our decision variable is the decompression policy choice (alternate the use of SDCs after each sortie, or alternate after one if full). There are two major noise variables modified to test how the policies perform under different settings. The first is the depth of the DISSUB. Depths of 250, 1000, and 2000 fsw are considered. The second noise variable is the internal pressure of the DISSUB. This parameter was varied at values of 25, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 110, 120, and 132 fsw.

In order to assess the performance of the system, three measurers of effectiveness are considered. The first is the total time the rescue is paused while awaiting chamber availability. This compares the overall time to complete a rescue to a rescue with unlimited decompression capacity. The second metric is the average time for an individual survivor to complete rescue from start of the simulation, which correlates to the time survivors are waiting in the queue to be decompressed. The third metric is the number of required PRM attendants to complete rescue without delay.

Simulation Model Description

We used discrete event simulation to build a model for the rescue process from start to finish. Discrete event simulation is used to model stochastic and dynamic systems, and is an appropriate methodology for this problem to model the state of the rescue over time to keep track of operating personnel and survivors. Building simulation models can help answer questions about the system before it has been built, and can incorporate uncertainty in the model logic to help predict a range of possible outcomes. Because there is uncertainty in how long it will take the new SRS-TUP system to perform different functions, discrete event simulation can incorporate probability distributions for these times to ensure that the decision-maker does not overestimate the performance of the system by assuming deterministic values.

This project used Simio simulation software to model a rescue process and used aspects of the software to help answer the research questions. Simio is a state of the art discrete-event simulation modeling tool that is used in academia, industry, and government applications. Its strength is that it provides not only a clear framework for modeling discrete-event systems, but it also incorporates sophisticated analysis methods to allow the models results to be analyzed statistically.

The discrete-event framework in Simio can primarily be applied to queueing systems, which were adapted to model a submarine rescue. Survivors were modeled as entities which are transported through the different components of the rescue using vehicles which represent the PRM. The decompression process is modeled as a server with a processing time. A series of add-on processes are used to model custom logic unique to this problem that could not be modeled using standard objects. Add-on processes have options to implement coding logic such as if/then statements, update state variable values, and transfer entities or objects to new locations. Thus, Simio can be used to model complex systems without requiring specific coding knowledge by the user. For a detailed guide to Simio and simulation modeling, see Smith et al., (2017).

Additionally, Simio can implement state of the art simulation techniques, like ranking and selection (Kim & Nelson, 2001), to determine the best system configuration. Another advantage is that different policies can be directly compared using the same model as a baseline. For example, different decompression rules or clean time limits can be implemented by tweaking parameters in the model. Simio allows for simultaneous runs of the same model with different parameters which means manual changes do not need to be made. There is a tool called the Subset Selection Analyzer that can be used to statistically compare scenarios to choose the best policy. Finally, Simio makes it easy to run multiple replications quickly by taking automatic advantage of multiple cores on the same machine.

In order to obtain the best validation possible, the team compiled a document describing all the details of the rescue process that were modeled in the simulation program. This document was sent to the URC leadership for feedback on whether the parameters and system dynamics modeled were realistic. The simulation program itself could not be transferred due to licensing and computing restrictions, thus we made the effort to ensure that the model details were communicated without needing to train or explain the details of the simulation modeling program to others.

Then, the members of our team verified the simulation model was working as expected by comparing the details from the project description with the simulation model code. The simulation model was built with ongoing debugging to ensure all components were working.

Experimental Results

This study found that while URC's procedures are generally correct, there are two potential issues to consider to achieve better results. Current policy suggests 60 fsw as the threshold for the internal pressure beyond which the decompression policy should switch from after every sortie, to waiting until the SDC is full. Simulation model results show that the crossover point at which decision makers should switch policies and fill a SDC with two sorties of survivors before starting decompression is lower, at 45 fsw.

This study also recommends that URC update their procedures for SRS-TUP to base the decision on the decompression rules based on the expected decompression time. When the expected decompression time is less than 12 hours (the approximate required time for two sorties), decompression should occur after each sortie because there is enough time to make the chamber available for the next sortie. When the expected decompression time is longer than 12 hours, decompression should occur only after the SDC is full.

Experimental designs and statistical methods are becoming increasingly important in assessing the performance of systems in test and evaluation (Ortiz & Harman, 2016; Hill, 2017). The simulation model was run under a variety of conditions, varying the number of survivors, rescue depth, and DISSUB internal pressure. In the end, most experiments involved 155 survivors to simulate a worst-case rescue with a large number of people to be transported. Initially, an experimental design was developed using a Nearly Orthogonal Latin Hypercube (NOLH) model (Cioppa & Lucas, 2007; Sanchez, 2011). This model chooses experimental design points to span the space of possible variables efficiently, rather than testing every possible combination of noise factors. A few key conclusions can be drawn from these results. In particular:

- With two chambers available, when the average decompression time is more than twice the average sortie time, delays in the rescue will be incurred for chamber availability. This is intuitive because there are two chambers and the sorties may arrive faster than the decompressions can occur. However, rescues involving fewer than two decompressions (due to a small number of survivors) will not incur delays.
- The average decompression time is largely a function of DISSUB internal pressure. This varies from just 2.7 hours to over 55 hours over the range of pressures evaluated and has the most significant impact on rescue delays.
- The average sortie time is largely a function of DISSUB Depth, but varies little over the range of data. With depth ranging from 264 to 2000 feet, the sortie time only changed from 5.07 to 5.81 hours. This effect was small compared to the decompression time.

To analyze the performance of the system, we consider three specific quantities that are measured in the simulation model.

- Average rescue delay per sortie (the total time the rescue is delayed due to SDC unavailability divided by the number of sorties)
- Time from first to last rescue (total time taken to complete the rescue)
- Average decompression time for survivors across the entire rescue

Two factors account for the rescue delays more than any other: the time required for decompression, and the decision variable of this decompression policy to use. We present each of our performance metrics according to these two factors. Figure 1 shows results with the average rescue delay per sortie displayed against the time required for decompression under each decompression policy. In each of the two policies, the decompression cycle time accounts for over 85% of the variability in the total delay in the rescue. Since decompression cycle time is driven by the DISSUB internal pressure, this pressure is the most significant factor in determining which decompression policy to use in a rescue.

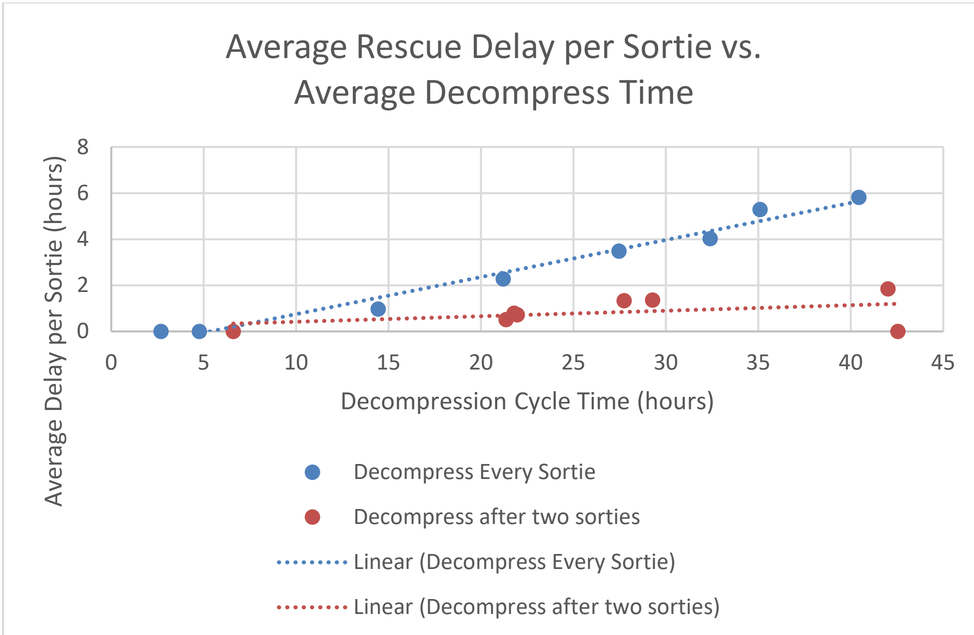


Figure 1. **Average Rescue Delay per Sortie vs. Average Decompression Time**

Next, rescues were simulated at DISSUB internal pressures from 25 to 132 feet of seawater. We measured the cumulative delay over the rescue (the total time for the rescue from start to finish) under each of these conditions for each of the two decompression policies. Plotting the cumulative delay against the DISSUB internal pressure (Figure 2), a clear distinction can be seen, with no delays in the rescue up to pressures of 40 ft sw. At pressures above 45 fsw, the expected decompression time became over twice than the sortie time, which warranted holding decompression until the chamber was full.

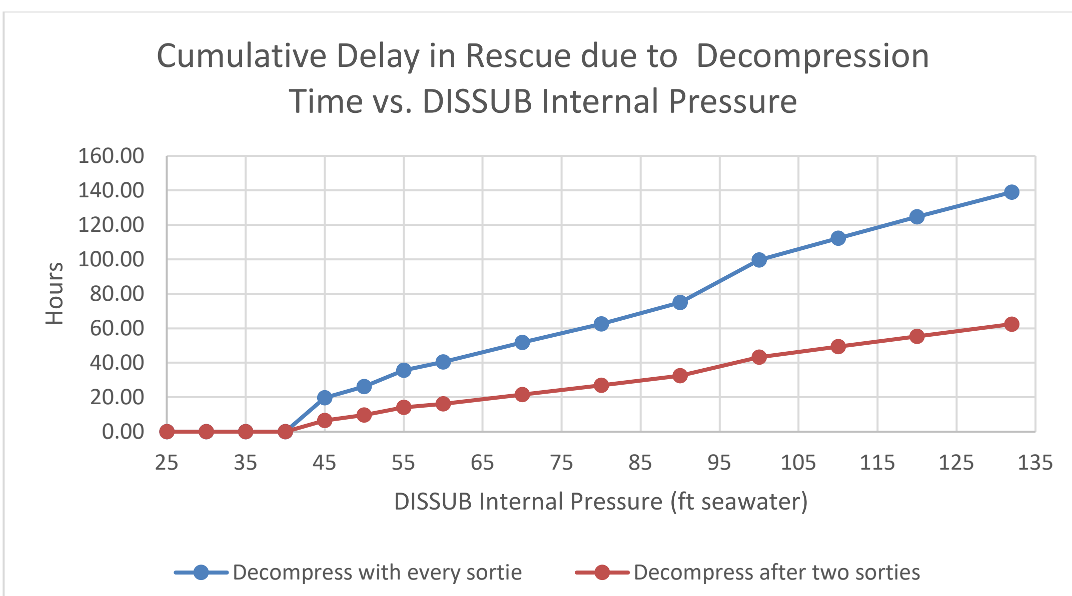


Figure 2. **Cumulative Delay in Rescue due to Decompression Time vs. DISSUB Internal Pressure**

However, just looking at the cumulative delay at the aggregate level fails to capture the effect on individuals. During rescues with short decompression times, survivors may be left waiting unnecessarily to decompress, increasing their risk of complications. We also looked at the average time for an individual survivor to complete decompression from the start of the rescue, which is graphed in Figure 3. For DISSUB internal pressures below 45 fsw, there is a slight efficiency advantage for decompressing after each sortie. Additionally, using only a single SDC to decompress survivors after every sortie for these lower pressures provides the flexibility of having the other SDC available for treatment of any survivors experiencing decompression complications. This is already captured in the URC's procedures.

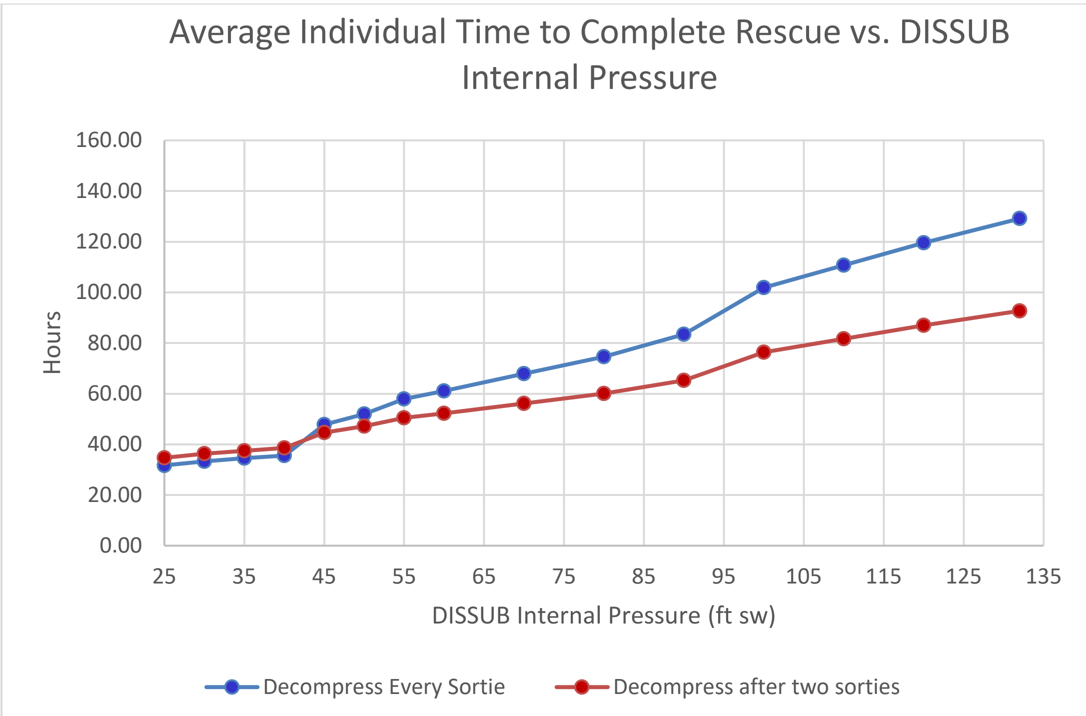


Figure 3. **Average Individual Time to Complete Rescue vs. DISSUB Internal Pressure**

Next, we study the number of attendants needed for the rescue to be performed successfully. URC must maintain enough qualified attendants on staff and ready to execute a rescue at all times. Conducting a rescue with fewer than the required number of attendants could result in pausing the rescue operation while waiting for attendants to complete decompression or their post-decompression clean time. Our model accounted for all attendants either on watch or otherwise unavailable and determined the maximum number of attendants needed for a given rescue.

With the first sortie, the PRM will bring two attendants that will stay on-board the DISSUB to assist the crew in the rescue. Based on the internal pressure in the DISSUB, these attendants will have a limited stay time, and need to be replaced by fresh attendants on a future sortie. Their return, however, takes away seats from the survivors, so this process effectively adds additional survivors that require rescue. For our simulations, we assumed that the URC will provide continuous coverage of two DISSUB attendants and swap them out as required. We use the model to determine how many total attendants will be needed.

The PRM attendants, who remain on the PRM through the rescue can either conduct watch turnover after every sortie or stay with the PRM for two cycles. After a single sortie, the PRM attendant will have been pressurized for less than four hours, so has not reached saturation. The attendant will be eligible for a reduced decompression timeline, and after waiting a "clean time" at atmospheric pressure will be available for a follow-on sortie. If any attendant were to stay on for a second sortie, they would remain exposed to the DISSUB pressure through that second cycle and require the same decompression cycle that the survivors entail. At this point, the attendant would not be available for additional sorties.

We ran our experiment varying the DISSUB pressure from 25 to 132 fsw using both options for attendants (a single sortie per attendants, and a dual sortie per attendants). Since we were trying to find the worst-case rescue situation, we used the design specifications of a 2000 fsw rescue depth and 155 survivors. The results are shown in Table 1 for the Time to Last Rescue (TTLR) under each option along with the number of attendants required not to delay the rescue.

Table1. Attendants Required for Rescues Under Various DISSUB Internal Pressures

				1 Sortie Per Attend.		2 Sorties Per Attend.	
DISSUB Attend	Total People (including 155 survivors)	DISSUB Internal Pressure		TTLR	Attendants Required (Average)	TTLR	Attendants Required (Average)
2	157	25		61.421	10	61.421	10
2	157	30		63.021	10	63.021	10
2	157	35		64.221	10	64.221	10
2	157	40		65.321	10	65.321	10
4	159	45		80.8515	10.08	80.8515	10
4	159	50		86.0462	11.2	86.0462	10
4	159	55		93.1536	11.12	93.1536	10
6	161	60		100.716	11.2	100.716	12
8	163	70		109.427	11.08	109.427	12
10	165	80		121.546	11.6	121.546	12
12	167	90		129.206	11.12	129.206	12
16	171	100		1732.97	20	145.775	12
20	175	110		1740.68	20	155.484	12
26	181	120		1754.23	20	171.711	12
30	185	132		1761.28	20	182.815	12

All rescues below the pressure of 55 fsw could be conducted without a delay for a sortie to enter the SDC with only 10 qualified attendants. The worst-case scenario, from a depth of 2000 ft, with 155 survivors pressed to 132 fsw, will require 12 qualified attendants. This number could increase if some sorties carry fewer than 16 survivors, which could happen if a stretcher needs to be used to carry an injured survivor, or if there are additional personnel on a sortie for medical or other reasons.

Conclusions

This study validated the URC's current policy for the SRS-TUP that the best policy is generally to decompress a chamber after two sorties when it is full, rather than decompressing immediately after each sortie. The current threshold policy for using this decompression policy when the internal pressure is higher than 60 fsw could instead be lowered to 45 fsw. Higher pressures result in longer decompression times, and thus decompressing after each sortie may result in delays for the next sortie that arrives. The current policy in use calls for decompressing after every sortie when the decompression time is less than the length of a sortie.

Our results show that decompressing after every sortie can lead to longer delays than waiting to fill an SDC before decompressing and that total delays in the rescue may range from 20 hours at 45 fsw internal DISSUB pressure to 140 hours at 132 fsw. The difference in URC's assumptions and the simulation results is most likely due to the simulation modeling 5% of survivors encountering some difficulty during decompression and requiring a longer decompression cycle. We selected the 5% value for the model after consulting with URC. It is also possible that there are numerous other causes for delay that are not predicted by the model, so we recommend URC allows a buffer time for unexpected problems.

Using a simulation model for the entire rescue process, we demonstrate the effects of two possible decompression policies on the time to complete a rescue. We incorporate uncertainty in the time to complete various aspects of the rescue, as well as vary the possible conditions (pressure, depth) associated with a scenario to find a robust policy that is preferred under extreme or poor conditions. In addition to determining which decompression policy to use, the study provides guidance on the number of attendants needed to complete the rescue, and the overall time to complete a rescue successfully. The results of the study were made available to URC for their planning purposes.

References

Cioppa, T. M., & T. W. Lucas. (2007). Efficient nearly orthogonal and space-filling Latin hypercubes. *Technometrics*, *40*(1), 45–55.

Giadrosich, D. L. (1995). *Operations research analysis in test and evaluation*. Reston, VA: American Institute of Aeronautics and Astronautics.

Hill, R. (2017). The test and evaluation workforce and a base of sand issue. *The ITEA Journal of Test and Evaluation*, *38*(2).

Kim, S. -H., & Nelson, B. L. (2001). A fully sequential procedure for indifference-zone selection simulation. *ACM Transactions on Modeling and Computer Simulation*, *11*(3), 251–273.

Marine Corps Operational Test & Evaluation Activity (MCOTEA). (2013). *MCOTEA Operational Test and Evaluation Manual.* Quantico, VA: United States Marine Corps.

Ortiz, F., & Harman, M. (2016). DOE in DT: The place to be! *The ITEA Journal of Test and Evaluation, 37*, 241–245.

Phillips, S. P., et al. (2016). *JHU/APL Submarine Rescue Capability Study*. Laurel, MD: Johns Hopkins University Applied Physics Laboratory.

Sanchez, S. (2011). NOLH design spreadsheet. Retrieved May 24, 2018, from http://harvest.nps.edu/

Smith, J. S., Sturrock, D. T., & Kelton, W. D. (2017). *Simio and simulation: Modeling, analysis, and applications* (4th ed.). Simio LLC.

U.S. Department of the Navy, Naval Sea Systems Command, Supervisor of Diving and Salvage. (2017). *U.S. Navy Submarine Rescue System Decompression Plan* (Rev 0).

U.S. Department of the Navy, Naval Sea Systems Command, Portsmouth Naval Shipyard. (2017). *Submarine Rescue System mission scenarios, operating checklists, & system overviews 0A* (Release 2-8).

International Acquisition Best Practices

Virginia L. Wydler—Chief Acquisition Scientist at The MITRE Corporation, has over 25 years of experience in contracting. She holds an MS in acquisition management, Naval Postgraduate School, and a BS in business administration, University of Maryland. She is a Certified Professional Contract Manager (CPCM) and Fellow with National Contract Management Association (NCMA). [vwydler@mitre.org]

Erin Schultz—Center for Acquisition Management and Sciences Technical Director at The MITRE Corporation, provides subject matter expertise in the areas of contracting, source selection, proposals, and acquisition strategies for a variety of customers. She has over 25 years of contracting experience with both federal government and industry, specializing in information systems. She has been an instructor for acquisition and contracting courses for several associations, including NCMA. She has taught internal acquisition and source selection courses for The MITRE Institute. [eschultz@mitre.org]

Kelly Hough—Acquisition Strategy and Management Department Head at The MITRE Corporation, provides acquisition and contracting expertise for federal government sponsors. She is a former Contracting Officer and North Atlantic Treaty Organization (NATO) PM, and a Certified Acquisition Professional with more than 20 years of experience in executing U.S. Defense domestic and international acquisition programs. [khough@mitre.org]

Acknowledgements

The MITRE Corporation Contributing Authors:

Carla Brown

Jim Bui

Eugene Cullen

TJ Restivo

Josh Semat

Executive Summary

United States (U.S.) program offices need help implementing complex international acquisitions, as their missions expand into the global use of U.S. defense weapons. This research documents best practices that will allow federal program managers (PMs) to implement international acquisition strategies that fit their situation on the acquisition lifecycle spectrum. The research reflects a comparison of international acquisition best practices by U.S. agencies, foreign entities, and commercial industry for military systems and space exploration that program managers can adopt to advance international acquisition strategies.

Background

National defense has become a global business as United States (U.S.) national interests are closely intertwined with those of the rest of the world. Globalization has driven the U.S. to seek opportunities to collaborate with U.S. allies and partner nations, which creates increasingly complex procurement strategies. The U.S. continues to balance national concerns with partner desires. To support global interest in U.S. products, the Department of Defense (DoD) sells military capabilities to coalition partners through Foreign Military Sales (FMS) under the Federal Acquisition Regulation (FAR) and executed in accordance with rigorous and complex U.S. regulations. Conversely, U.S. industry is also permitted to sell military capabilities directly to foreign governments under the Direct Commercial Sales (DCS), following rules for export control set by the Department of State, Department of Commerce, and other federal agencies. These two processes create a dynamic environment for program managers (PM).

International Acquisition Issues

According to a February 2019 General Accounting Office (GAO) report, the DoD received 3,038 FMS-related requests "in fiscal years 2014 through 2018 from 93 countries across 6 geographic regions." FMS and DCS have their own unique set of procurement rules, especially dealing with export control, international coalitions, and foreign governments. A significant number of key players are involved in the FMS and DCS approval processes. This mix of entities and differing procurement rules can be viewed by foreign partners as protracted and cumbersome.

Space exploration in the U.S. is shifting from an era of government control to that of multi-national coalitions and commercial investments. Commercial space ventures have increased, with Space-X, Blue Orbit, and other private companies investing in space travel. At the same time, the National Security Council has directed the National Aeronautics and Space Administration (NASA) to execute a new lunar mission. When established in 1958, NASA was directed to pursue cooperation "with other nations and groups of nations." This principle of international cooperation is still important today. Such collaboration will be essential in addressing the inherently global and interrelated space race (NASA, 2014).

Spurred by U.S. coalition partners and U.S. industry, the current administration has shown a renewed interest in adjusting the arms transfer policy. The White House is driving a review of the acquisition and contracting processes used to execute arms transfers under the Conventional Arms Transfer Policy, released on April 9, 2018. There has been dramatic growth in the level and dollar value of U.S. arms exports, and changes in the world market. To execute arms transfers reliably, PMs need to understand rules governing arms exports and their impacts and need a thorough understanding of the overall acquisition and contracting process.

Global Influences—Arms Transfers

The Center for Strategic International Studies (CSIS) hosted a forum on August 8, 2018, titled "U.S. Arms Transfer Policy: Shaping the Way Ahead." The forum chair, Andrew Philip Hunter, stated that change is afoot in the world of U.S. arms transfers. He stated that

> our arms exports and security cooperation more generally are a major focus of the strategy, both in the National Security Strategy and in the National Defense Strategy. And no doubt related to that there's a huge leadership focus on it … that is unparalleled in the last year and a half. (Hunter, 2018)

Ambassador Tina Kaiako, Acting Assistant Secretary of State for Political-Military Affairs, U.S. Department of State (DOS), reported that a major new policy, which updated the previous CAT policy from 2018, is designed to (1) shift from a reactive to a proactive approach to CAT to boost the U.S. defense industrial base, (2) secure resources to execute the shift in approach; and (3) develop a broad engagement plan with U.S. stakeholders including Congress, etc. A CAT Implementation Plan (with classified sections) was promulgated July 13, 2018.

Laura Cressey, Deputy Director, Office of Regional Security and Arms Transfers, U.S. DOS, addressed the need to decrease (FMS) cycle time. She stated,

> We want to increase US competitiveness by building in exportability. … It takes 300+ days for major FMS acquisition execution. DoD is overwhelmed

> with missions and FMS has become a 3rd priority. That is an opportunity potentially with the administration to look at carving out, with the Congress' support, unique federal acquisition regulation procedures for FMS contracting to have a truly rapid process. That's going to take whole-of-community support, and we at the chamber are willing to help with that. (Cressey, 2018)

Based on the issues raised in the CSIS forum, U.S. agencies should review the acquisition and contracting processes used to execute the transfers. Now, more than ever, it is critical to reduce cycle times and improve the quality of the acquisition process.

Foreign Military Sales Contracting

The sale of U.S. weapons and military systems to foreign governments is complex, and, for a major weapon system sale, the process could last for many years. The U.S. infrastructure supporting FMS does not represent a stand-alone arrangement, but instead utilizes the existing DoD acquisition structure. The diverse laws, regulations, policies, and guidance that govern U.S. procurements also govern international acquisition, with some exceptions.

Defense FMS Policy, Regulation, and Guidance

Under the Foreign Assistance Act (FAA) and the Arms Export Control Act (AECA), and in accordance with Executive Order 13637, the Secretary of State is responsible for the supervision and oversight of Security Assistance (SA) programs. SA refers to the collection of programs authorized under Title 22 U.S. Code (U.S.C) wherein the U.S. provides defense articles, military education and training, and other defense-related services to foreign nations by grant, loan, credit, cash sales, or lease, in furtherance of national policies and objectives.

Security Cooperation (SC) comprises all activities undertaken by the DoD with foreign defense security establishments, including all DoD-administered SA programs. Title 10 U.S.C. Section 301 defines security cooperation programs and activities of the DoD as "any program or interaction of DoD with the security establishment of a foreign country to build capabilities, to provide access or to build relationships." The DoD administers many of these FAA- and AECA-authorized security assistance programs using the *Security Assistance Management Manual* (SAMM).

The FAR does not make specific references to FMS, since FMS is a Defense Department function to control the procurement of weapons from industry to sell to foreign governments. However, the FAR does provide an exception to full and open competition under FAR 6.302-4, International Agreement.

The Defense Federal Acquisition Regulation Supplement (DFARS), Subpart 225.73, Acquisitions for FMS, provides policies and procedures for the acquisition of FMS, and authorizes the DoD to enter into contracts for resale to foreign countries or international organizations. All the Military Departments (MILDEPs) have issued further supplements to the DFARS to aid contracting personnel in implementing FAR and DFARS provisions.

The DoD does not maintain a separate acquisition infrastructure for FMS; instead, the DoD supports FMS utilizing the pre-existing infrastructure established to support U.S. acquisition and logistics needs. The DoD Instructions 5000 series, which provides mandatory policies and procedures for all Major Defense Acquisition Programs (MDAPs) and Major Automated Information System (MAIS) Acquisition Programs, requires acquisition managers to pursue international cooperation in acquisition to the maximum extent feasible, and consistent with core business practices and the overall political, economic, technological, and national security goals of the U.S.

The Defense Institute of Security Assistance Management (DISAM) publication, *The Management of Security Cooperation*, commonly referred to as the “green book” due to its green cover, covers the full range of security cooperation activities and is the basic textbook employed by the Defense Institute of Security Cooperation Studies. It is considered the authoritative source for FMS guidance. DISAM, Chapter 1, states that the FMS program is a non-appropriated program administered by the Defense Security Cooperation Agency (DSCA), through which eligible foreign governments purchase defense articles, services, and training from the government. The purchasing government pays costs associated with the sale.

The Letter of Acceptance (LOA) is commonly referred to as a “case” and is assigned a unique case identifier for accounting purposes (Defense Institute of Security Cooperation Agency, 2018, Chapter 1). The LOA is a bilateral agreement whereby the U.S. commits to provide the approved goods or services and the foreign government agrees to the terms, conditions, and payment schedule. The U.S. agreement is caveated as “best effort,” meaning the U.S. cannot be considered in default of the agreement if product performance levels will not be achieved. Pursuant to LOA, the U.S. initiates the acquisition process and awards a contract on behalf of the foreign government. The U.S. executes the acquisition of behalf of the foreign entity. The FMS party indemnifies the U.S. and agrees to absorb all financial risk.

Acquisition Planning

The FMS Case process is executed in three phases—Pre-Case Development, Case Development, and Case Execution. The primary planning activities fall into pre-case development and case development. The pre-case development phase can go indefinitely as the parties discuss national requirements, sources, affordability, as well as potential political implications and an assessment of available U.S. technologies. The U.S. does not start tracking the contract execution timeline until a requirement transitions to the case development phase, which commences with the receipt of a Letter of Request. During this phase, the U.S. issues a formal offer in the form of an LOA, which documents the terms of agreement between the U.S. and a foreign government. This phase concludes with the countersignature of the LOA and the initial deposit to Defense Finance and Accounting Services.

DISAM Chapter 5 (Defense Institute of Security Cooperation Agency, 2018, p. 5-2) provides a detailed look at the FMS Case process. It depicts a very long process from the time the LOA is requested to acceptance of the LOA, which is typically at least 250 days. This timeline assumes that every milestone or activity executes as expected and does not consider the pre-case development time.

The FMS process can apply to both competitive and directed source acquisitions. If a foreign country states in official written direction, such as an LOA, that the contract (or subcontract(s)) is to be awarded to a specific firm, the procuring office must process a written justification and approval as described in FAR 6.303 and 6.304, along with applicable DFARS regulations and any service-specific or local instructions. Other than this provision for a foreign customer directing award of the contract and/or subcontract(s) to a specific firm, agencies meet FMS requirements in accordance with the “normal” acquisition process as prescribed in the FAR, DFARS, Service-specific regulations, and any local instructions.

Standard FMS funds provided by the acquiring nation are not constrained by fiscal year (FY) limitations and do not expire with the end of the FY. Conversely, U.S. foreign military funding (FMF) does expire and must be obligated prior to the end of that FY. This funding is typically not available to the procuring agency until the third or fourth quarter of a FY.

On June 28, 2018, the DoD issued a Class Deviation effective immediately that stated that when determining contract type for FMS procurements, contracting officers shall comply with Section 830 of the National Defense Authorization Act (NDAA) for FY 2017. The policy expressed in the DoD procurement memorandum dated June 28, 2018, *Negotiations of Sole Source Major Systems for U.S. and U.S./FMS Combined Procurements*, and the guidance provided at DFARS Procedures, Guidance and Information (PGI) 216.403-l(l)(ii)(B). This new policy states that contracting officers shall use firm fixed price (FFP) contracts for FMS unless one of the following exemptions applies:

- The FMS customer has established, in writing, a preference for a different contract type, or has requested, in writing, that a different contract type be used for a specific FMS.
- The contracting officer requests a waiver on a case-by-case basis when a contract type other than FFP is in the best interests of the U.S. and American taxpayers.

The determination of best interest must be made on a case-by-case basis and be approved by the Chief of the Contracting Office.

Negotiations and Award

Although the FMS host nation is the final customer, it is not a party involved in the contract negotiations and final award. The U.S. is the legal entity with which the country has contracted. DFARS 225.73 encourages FMS customer participation in discussions with industry regarding development of technical specifications, establishment of delivery schedules, special warranty provisions, varying alternatives, quantities, and options needed to make price-performance tradeoffs. Restrictions regarding foreign national participation in negotiations apply if:

- The contract includes requirements for more than one FMS customer.
- The contract includes unique U.S. requirements.
- Contractor proprietary data is a subject of negotiations.

DFARS 225.7304(c) states that no proprietary data, including cost or pricing data, can be released to the FMS customer unless the contractor has authorized it. Further, DFARS 225.7304(d) states that customer participation in contract negotiations is left to the discretion of the contracting officer after consultation with the contractor. In FMS situations, contractors may be less willing to provide enough insight into the basis of estimate for their proposed technical approach or costs, given that they know there is no competition and not all technical and cost information will be shared with the end customer.

Offset Agreements

One aspect of FMS contracting that does differ from any other contracting efforts is “offsets.” Whereas offset arrangements are a typical element of many international procurements, the DoD does not encourage, enter into, or commit U.S. firms to FMS offset arrangements. From an FMS perspective, the decision to engage in offsets, and the responsibility for negotiating and implementing offset arrangements, resides with the companies involved.

DFARS 225.7303-2(a)(3) defines an offset agreement as the contractual arrangement between the FMS customer and the U.S. defense contractor that identifies the offset obligation imposed by the FMS customer that has been accepted by the U.S. defense contractor as a condition of the FMS customer's purchase. These agreements are independent of the LOA and FMS contract. DFARS 225.7303-2 provides insight as to both direct and indirect offsets:

- A direct offset involves benefits, including supplies or services, that are directly related to the item being purchased.
- An indirect offset involves benefits, including supplies or services, that are unrelated to the item being purchased.

Table 1 provides the most common types of direct, direct/indirect and indirect offsets.

Table 1. Offset Categories

Direct	Direct/Indirect	Indirect
Co-Production Subcontracts	Technology transfer Training Licensed Production Foreign Direct Investment, Credit Assistance and Financing	Export Assistant Purchases Offset Swapping (compensation of offset obligation through reciprocal abatement)

Offsets can take many forms. Offset requirements should be negotiated by the FMS customer and industry prior to the FMS contract. Fully executed agreements prior to award rarely occur due to the complexity of the offset agreements and the extended negotiation timelines required.

Direct Commercial Sales (DCS) to Foreign Governments

DISAM defines DCS as the "export of defense articles, services, and training licensed under the authority of Section 38, AECA, made by U.S. defense industry directly to a foreign government." Basically, DCS is any sale to a foreign government that is not executed through the FMS/FMF system. The U.S. is not a party to a DCS contract. Many large-scale DCS will often have a corresponding and significantly smaller FMS case to accommodate items requiring government-to-government transfer. Other than government-to-government transfers, the required controls are implemented through licensing by the DOS, specifically, the Directorate of Defense Trade Controls (DDTC). Execution of DCS programs is governed by the International Traffic in Arms Regulation (ITAR) under 22 Code of Federal Regulation (CFR) 120-130.

The primary difference between DCS and FMS is that DCS removes the U.S. from its role as the "middleman" and allows the foreign government(s) to interact directly with U.S. industry to determine/execute requirements and to assemble an overall package that best fits the partner nation's needs and budget. DCS is not subject to U.S. procurement regulations and is often subject to the foreign government's procurement rules.

Although sometimes perceived as less cumbersome than FMS (LMD Defense, n.d.), DCS is subject to the same ITAR regulations and export restrictions as FMS. With DCS, the responsibility for ensuring compliance rests with the vendor. The penalties for violation can be severe, ranging from debarment to imprisonment and/or the levying of significant fines. As of August 1, 2016, ITAR violations may result in monetary penalties of up to $1.09 million (per violation). Civil penalties apply to each individual violation. A single violation of noncompliance can be broken down into multiple violations, resulting in penalties in the range of tens of millions of dollars (Export Rules, n.d.).

A summary of best practices applied across Defense International acquisition is shown in Table 2.

Table 2. Defense International Acquisition Best Practices

U.S. Defense International Acquisition Best Practice	Reference
Use an integrated product team (IPT) to integrate international requirements into the program cost, schedule and performance.	https://www.dau.mil/cop/iam/_layouts/
Utilize the DoD Acquisition Strategy Template, April 2011 version, International Acquisition, and Int'l Acqn and Exportability (IA&E).	https://www.dau.mil/cop/iam/Pages/Documents.aspx
Attend DAU International Acquisition Learning Path Courses: ACQ 120 Fundamentals of Int'l Acqn, ACQ 230 Int'l Acqn Integration, ACQ 340 Advanced Int'l Mgmt Workshop, ACQ 380 Int'l Acqn Management, CL Module 048 Export controls.	https://www.dau.mil/
Utilize OUSD(AT&L) Defense Acquisition Guide (DAG) update Feb 2017. Provides comprehensive guidance on IA&E. Use Job Support Tools (JSTs): IA&E Assessment, Acquisition Strategy, International Considerations, Defense Exportability Integration, International Cooperative Programs (ICPs), FMS Systems Acquisition, International Business Planning.	Chapter 1) IA&E Considerations (para 4.2.8) Substantial IA&E Supplement. https://www.dau.mil/tools |
Become a member of DAU IA&E CoP for potential collaboration.	https://www.dau.mil/cop

Defense FMS Military Activity

The DoD Military Services use the FMS process to sell weapons systems to foreign governments. Each Service has unique capabilities to sell and uses different U.S. industries.

U.S. Navy (USN)

The USN engages with partner nations around the world to deliver sea and air-based maritime capabilities to foreign partners. Capabilities include Command, Control, Communications, Computers and Intelligence (C4I), aircraft and airborne weapon systems. ships and submarines and their combat systems, and corresponding logistical supplies and services. The management of these technology transfers plays a key role in shaping the USN's approach to global partnerships and achieving the goals of the maritime strategy. USN manages and implements International Security Assistance programs, Cooperative Development programs, and Technology Security policy. In total, the Navy is tracking ~3,800 open FMS cases with an associated value of ~$118 billion; in FY 2019 alone, there are 432 active cases totaling $5.1 billion. As a reporting unit to the Assistant Secretary of the Navy for Research, Development and Acquisition, the Naval acquisition centers support Regional Combatant Commanders' and Navy leadership's efforts in building long-term relationships with our maritime security partners around the world. By teaming with a wide network of U.S. defense industry and security community product and service providers, PMs, policy makers, and technical and regulatory agencies, they support the defense requirements of our friends, allies, and coalition partners.[1]

U.S. Army

The U.S. Army often shares military capabilities of tank and helicopter warfighting technology with multiple nation-allies around the world to ensure joint military readiness against shared adversaries. The U.S. Army Security Assistance Command (USASAC) is known as the "Army's face to the world," maintaining relationships with more than 150 countries through its role in FMS. USASAC was located at Fort Belvoir until September 2009, when it became the first flag-level command to move to Redstone Arsenal in 2011, a full two years ahead of the Base Realignment and Closure (BRAC) schedule. The relocation to Redstone Arsenal keeps USASAC in close proximity to Headquarters, U.S. Army Materiel Command, its parent command, and Army Security Assistance Enterprise partners, such as the Program Executive Office/PM community, which also have a presence or connection to Redstone.

The Security Assistance Enterprise includes the security assistance management directorates of each of the AMC life cycle management commands, which ensure the Army supports each FMS case. The technical specifications and costs for specific items, such as helicopters, that are requested by a country must be developed by the SAMD, such as AMCOM (Aviation and Missile Command), and coordinated with the PEO, such as PEO Aviation. Another example would be a tank, which could be coordinated through U.S. Army Tank-Automotive and Armaments Command's (TACOM's) (TACOM Life Cycle Management Command's) SAMD (Gillespie, 2011).

U.S. Air Force (USAF)

The USAF engages with partner nations around the world to deliver aircraft and C4I capabilities. The Air Force brought home $27 billion in foreign military sales in FY 2017—nearly 213% more than the previous year—amid several changes meant to reduce cycle times, according to the service's security assistance and cooperation director.

Brig. Gen. Gregory Gutterman, who leads the directorate that handles FMS sales to 109 foreign allies as part of the Air Force Lifecycle Management Center at Wright-Patterson Air Force Base, said the service usually sells an average of $9–$10 billion per year. "Twenty-seven billion, that's a great number," he told *National Defense* on December 15. "If you look at the Fortune 500, McDonald's sold $24 billion worth of hamburgers last year, and we brought in $27 billion worth of military revenue. That's a significant contribution to our gross domestic product here in our nation," he added.

One major factor was Qatar's decision to purchase 36 F-15 fighter jets and related services for $12 billion, he said, noting, "That was really the reason for such a record year." The other top two drivers were F-35 deliveries to Israel and sustainment costs related to Iraqi F-16 fighter jets. Finally, Gutterman noted that the Air Force security assistance and cooperation directorate, or AFSAC, typically sells about $1 billion worth of supply chain–related costs per year (Machi, 2017).

---

[1] See https://www.secnav.navy.mil/nipo/Pages/mission.aspx

A summary of best practices applied across FMS is shown in Table 3.

Table 3. FMS Contracting Best Practices

Foreign Military Sales Contracting Best Practice	Reference
Anticipate lengthy process to accommodate arms export control restrictions. Set expectations early on timelines.	DISAM, Chapter 9
Avoid utilizing Undefinitized Contract Actions (UCA) not in the best interests of the Government. Negotiate FMS contracts up front, technical and cost terms clearly defined.	FAR Part 6.303
Just in time training on the FMS case process and FAR contracting process provided to the FMS sponsor.	https://www.dau.mil/
Anticipate different fund sources. FMF does expire; FMS funds don't expire at FY. Use Agency Comptroller for expenditures.	Funding Source and appropriation rules
Develop comprehensive Life Cycle Cost Estimate with high confidence factor and matured risk model, fully funded at program initiation. Maximize use of existing cost model data.	Funding Sources
Comply with new FMS policy on FFP contract type. Inform FMS customer of policy change; review LOA for terms.	NDAA FY2-18; DPAP memo dtd 28 Jun 18
Contracting officer's representative (COR) and/or the case manager (CM) interface with the contractor, monitoring performance to control scope changes and any resulting changes to LOA.	Program Team structure
Incorporate Earned Value Management (EVM) or EVM-like practices to monitor cost, schedule, and technical performance.	Navy practice
Account for all costs for in-country personnel. Use DOS site and actual experiences to identify costs for in-country personnel.	Navy practice
Establish strong communication and information sharing with contractor and host nation; host nation cannot give direction to contractor; contract is between contractor and PMO.	Navy practice
Ensure that any offsets are clearly defined up front by all parties. Offsets are a mix of direct and indirect contributions.	Air Force practice

Foreign Acquisition Processes

This section presents a comparison assessment of U.S. DoD acquisition system processes and those of its international allies, to include France, Germany, the United Kingdom (UK), and Australia. It also summarizes comparisons of acquisition practices, since the U.S. and other countries have increased their focus on warfare.

Foreign Government Acquisition Systems

Different countries often use different processes and procedures for the acquisition of defense systems. Research shows that the form of government, cultural norms, new age initiatives, industrial base, and ability to innovate across the marketplace all play a pivotal role in how a country's acquisition processes and systems are shaped and organized. As a result, there is no exact or standard method for comparing the efficiency of acquisition systems between the U.S. and other nations. However, this assessment was conducted at a high level to measure the relationships between several objective variables: government structures, policy and oversight, acquisition phases, technology utilization, FMS, acquisition workforce, training, and the industrial base.

Since 2010, the U.S. has heightened its focus on cyber warfare, bringing together cyber capabilities from partner countries along with those of the Army, Navy, Air Force, and Marine Corps. The last several years have witnessed an increased uptick in the expansion of cyber capabilities, training, and expertise across the world-wide governmental workforce. Several countries have adapted to ever-changing cybersecurity procedures and methods by increasing investment levels and coordinating cyber-incident responses. The research analyzed a wide-variety of government resources, cybersecurity market reports and other open sources for reporting both similarities and differences. However, there is inconsistency across nations on how to approach cyber warfare or policy that integrates cyber and acquisition. Many foreign governments are interested in modeling the U.S. with their agile processes that can result in shorter acquisition cycles.

A summary of best practices applied across foreign government acquisitions is shown in Table 4.

Table 4. Foreign Government Acquisition Best Practices

Foreign Government Acquisition Best Practice	Reference
Use Integrated Project Teams for acquisition activities.	All countries
Increase investment on acquisition management training.	All countries
Express socio-economic concern for health of defense industry.	All countries
Negotiate budgets internally within defense organization.	All countries
Establish formalized acquisition structures for weapons systems from conception to disposal.	All countries
Reform the acquisition system continuously.	U.S. and UK
Delegate significant project management powers to an international armaments organization—the Joint Organization for Cooperation in Matters of Armament (OCCAR).	France, Germany, Italy, and the UK
Intellectual property rights are treated under the United Nations Arms Transparency resolution.	France, Australia
Integrate the defense market, including the formation of two organizations—the Western European Armaments Organization (WEAO) and the OCCAR—to improve armament cooperation, which are integral to European countries.	Multiple European governments
Cybersecurity is treated differently in multiple countries. Very little standardization in practice when adopting or accommodating cybersecurity across foreign governments.	Multiple Foreign Governments

NATO Acquisition Process

North Atlantic Treaty Organization (NATO), also known as the North Atlantic Alliance, was formed on April 4, 1949, when 12 countries signed the North Atlantic Treaty (also known as the Washington Treaty). To date, the original goals have not fundamentally changed, nor has the Treaty been rewritten. The only “amendments” have been the inclusion of accession protocols added as new members join. With the addition of Montenegro in 2016, NATO membership has grown to 29 countries. Then, as now, the Treaty commits members to the shared risk, responsibilities, and benefits of collective defense. Moreover, this treaty and its NATO members form a unique “community of values committed to the principles of individual liberty, democracy, human rights, and the rule of law.”

There is no singular set of procurement rules for NATO, nor is there a central organization responsible for procurement. The rules and methods of procurement are dependent on the funding sources, the host nation involved, the type of goods and services required, and the degree of urgency (*Navigating NATO Procurement*, n.d.). A host nation is defined as the participating country or NATO Agency responsible for implementing a project. Thus, procurement is undertaken by different entities (countries or NATO Agencies) on behalf of NATO. Notwithstanding, NATO has issued a series of directives and policies that govern the majority of NATO procurement.

The Strategic Command (Bi-SC) Procurement Directive (Bi-SC Directive Number 6-70) dated December 22, 2004, is not rooted in law. The Bi-SC directive is comparable to the U.S. FAR in that it provides overarching acquisition policy guidance to NATO acquisition communities and organizations. Like the FAR, the Bi-SC provides governing principles, roles and responsibilities, procurement policies, and procedures that govern the acquisition of most goods and services.

The AC/3-D/221 (1996 Edition) NATO Security Investment Programme (NSIP)—Procedures for International Competitive Bidding, provides the basic procedures for competition of NATO NSIP projects and is comparable to FAR Part 6. These procedures focus primarily on establishing roles and responsibilities for the host nation in pursuit of maximining competitive opportunities for eligible nations’ industries. Ultimately, the host nation plays perhaps the most significant role in the overall procurement process.

The NATO acquisition process requires a significant amount of time and pre-coordination, typically 18–24 months. Once a program or project has been approved, by consensus, the Nations have agreed to fully fund the requirement over its intended period of performance. Once transferred to the host nation, the Nations’ financial contributions become no year, no color money. Once approved, a NATO program becomes a fully funded program, like a multi-year procurement in the U.S.

A summary of best practices applied across NATO is shown in Table 5.

Table 5. NATO Best Practices

North Atlantic Treaty Organization Best Practice	Reference
Use consensus decision-making, not voting, and decision is acceptable to all member countries.	NATO Alliance of April 1949
Negotiation is rapid. Members know positions in advance.	NATO Alliance of April 1949
Funding is provided by all nations, according to an agreed cost-sharing formula. All funding decisions by consensus, unanimous.	NATO Security Investment Program
NATO has overarching directives, procurement procedures by host nation. Comparable to the U.S. FAR—however, not law.	Bi-SC and AC/3-D/221
Consider options for cooperative development to reduce overall development costs for participants. Terms of cost share agreement may reduce schedule impact from ITAR restrictions.	NATO practice

Space Exploration Acquisition

International space exploration has moved well beyond the era when the U.S. government was the only heavy investor. The U.S. now collaborates with other countries in space exploration. However, these business relationships and coalitions can take longer and cost more than commercial investments by private firms. A 2012 comparison study by Aerospace between NASA and European Space Agency (ESA) development durations showed that ESA space programs take 30% longer than NASA programs. However, commercial vendors are launching space products even faster, leveraging technology and open systems.

NASA Acquisition and European Space Agency Acquisition Process

NASA's mission is to reach for new heights and reveal the unknown so that all that can be learned will benefit all humankind. NASA typically utilizes the expertise of multiple Centers to address the technical challenges that projects may face. By contrast, ESA's purpose is to provide for and promote, exclusively for peaceful purposes, cooperation among European States in space research and technology and their space applications. All Member States contribute to these programs on a scale based on their Gross Domestic Product (GDP), and provide the necessary expertise to ensure mission success. The other programs, known as optional, are only of interest to some Member States, which are free to decide on their level of participation.

Joumier, Freaner, Bitten, and Edmonds (2012) presented a paper comparing ESA and NASA acquisition approaches and the potential effects on science mission development duration and schedule changes at the joint International Society of Parametric Analysis and Society of Cost Estimating and Analysis Conference in 2012. Their study contrasted and compared the acquisition approaches of NASA and ESA science missions to identify differences and assess the development durations to identify any significant differences in schedule lengths and changes.

ESA and NASA acquisition phases are similar in terms of Phase A Conceptual Design, Phase B Preliminary Design, and Phase C/D Detailed Design and Implementation. Primary differences are in Phase B and Phase C/D. ESA Phase B comprises a competitive Phase B1 and separate Phase B2. ESA Phase B2 is like NASA Phase B. ESA Phases C/D are similar in content to NASA's, but ESA's contracts are typically FFP. Additionally, role/sharing must be agreed upon by all ESA partner Member States. NASA often serves in the integrator role for science missions, while ESA typically has the prime contractor serve in the integrator role.

The Joumier study compared average schedule durations for ESA Phase B2/C/D versus NASA Phase B/C/D for 32 NASA missions and 21 ESA missions. The findings showed longer schedule durations for ESA missions when compared to NASA missions. The average for NASA non-Earth-orbiting missions was 56.3 months versus ESA's 72.7 months. For Earth-orbiting missions the average for NASA was 70.1 months versus ESA's 91.8 months.

The study did not analyze cost data, but schedule is a proxy for cost and in many instances is proportional to cost. An extension in schedule will result in cost increases depending on the amount of personnel (both government and contractors) involved in the program. Due to the work sharing agreement among ESA Member States, ESA programs overall are more complex to manage, cost more, and take longer than NASA programs.

A summary of best practices applied across NASA and ESA is shown in Table 6.

Table 6. NASA and ESA Best Practices

NASA and International Space Agency Best Practice	Reference
Use streamlined requirement process, limiting requirement growths and reduce time develop to design to meet user needs.	Defense Management System College report, *A Comparison of the Acquisition System of France, Great Britain, Germany and the U.S.*; GAO report, *Briefing on Commercial and DoD Requirements and Acquisition Practices*; and SIA, *Smart Buying—Improving SATCOM Procurement*.
Partner with industry for evolutionary product development to achieve stability, reduce risk; enable short program schedules by limiting new design elements, reduce test and integration.	
Use new procurement techniques, contract type, and incentive fees tied to performance to encourage good contractor behavior.	
Reduce development and procurement schedules by streamlining test approval processes and reduce reporting requirements included in Contract Data Requirements List.	
Use a single IPT including users, stakeholders, and industry to empower PMs to make decisions with minimum oversight and in a timely manner with information from the IPT.	
Leverage commercial by procuring items commercially available to the maximum practical extent including off-the-shelf.	

Commercial Space Acquisition

Several studies have addressed space systems commercial acquisition practices that the DoD could adopt to reduce costs. These practices apply to international space system acquisitions as well. A 2010 GAO report discussed commercial practices that could benefit the DoD, including the recommendation to acquire mature critical technologies prior to program start achieving a high level of technology maturation prior to program initiation. This approach helps to (1) ensure resources and requirements match, and (2) avoid concurrently developing technologies, finalizing designs, and demonstrating manufacturing processes, which can lead to cost and schedule inefficiencies. Other recommendations included using evolutionary product development, tying contract incentives to performance, and empowering PMs.

The Satellite Industry Association (SIA) published a report on improving DoD satellite communications acquisitions that included the best practices that were very similar to what has been recommended in the past (SIA, 2014). These included performing integrated planning, leveraging commercial capabilities, and establishing polices that underpin a robust supply chain.

The Air Force Studies Board concluded in 2015 that using open standards and purchase data rights at the beginning of the program would shorten the lifecycle. Open standards allow the government to execute modularized functionality upgrades to future spirals without starting a new development. Purchasing data rights at the beginning allows the government to "own the technical baseline" for lowering the cost of sustainment and future upgrades (Air Force Studies Board, 2015, p. 4).

A summary of best practices applied across commercial space acquisition is shown in Table 7.

Table 7. Commercial Space Acquisition Best Practices

Commercial Space Acquisition Best Practice	Reference
Acquire systems that do not require research, development, test, and evaluation (RDT&E) (e.g., acquiring existing satellite bus from prime contractor developed for commercial customer).	Defense Management System College report, *A Comparison of the Acquisition System of France, Great Britain, Germany and the U.S.*; GAO report, *Briefing on Commercial and DoD Requirements and Acquisition Practices*; and SIA, *Smart Buying—Improving SATCOM Procurement*.
Use open standards and purchase data rights at the beginning of the program. Allow the government to execute modularized functionality upgrades to future spirals without starting a new development. Purchasing data rights at the beginning to allow the government to "own the technical baseline" for sustainment.	
Streamline requirements process. Freeze all requirements after authority to proceed. Requirements creep rare in commercial space. Reduce the amount of documentation and CDRLs that are required. Cost will be reduced by staffing resources.	
Streamline the decision-making process to a few key decision makers. Form a decision-making board for key milestones.	
Reduce oversight on the program. Once contract is awarded, let the contractor execute to well-defined requirements and system performances. Reduce the amount of reviews with contractors.	
Use FFP contracts to acquire space systems instead of cost-plus. FFP contracts are commonly used for commercial acquisitions.	

Summary and Conclusion

Globalization intertwines U.S. national security interests with those of the rest of the world. Additionally, spurred on by U.S. coalition partnerships, U.S. industry is expanding into new markets. All this makes for a multi-faceted environment.

- There are many players in the process—Congress, DOS, Commerce, Defense DSCA, Defense Industry, and Commercial Space Industry, each with a different role.
- There are many layers of regulations and agreements—ITAR, EAR, Offset Agreements, Trade Agreements, Treaties, FAR, and other Federal Agency–level regulations and guidance. The timelines driven by the sheer number of players, each with their own set of policies and regulations, drives significant delays.
- There are many investment options for both government and industry—Defense Research and Development (R&D), NASA R&D, Defense Industry Independent R&D (IR&D), and Commercial Space Industry IR&D.
- There are many conflicting Business Rules (e.g., Ownership of Intellectual Property Rights, Cost Sharing of Operations, and Logistics Support). Not all information is releasable to the foreign entities, further slowing the process and international sales.

The White House is driving a review of the acquisition and contracting process that is used to execute arms transfers. Thus, there may be potential changes in arms transfer policies. Program offices need a thorough understanding of the acquisition and contracting process to soundly execute arms transfers under these new policies. Also, the DoD, DOS, and Department of Commerce need a cohesive and collaborative approach to the military systems and arms control.

FMS is still big business. According to the GAO (2019), the DoD reported more than $55 billion in FMS for FY 2018 alone. Although the DoD has undertaken various initiatives intended to make the FMS program more responsive and better able to meet customers' expectations, the FMS process is still perceived as cumbersome and unable to keep pace with foreign governments' demands.

FMS is still a recognized acquisition process. However, DCS is on the rise as U.S. industry seeks to expand markets and sales. Acquisition policy and practice changes may be needed to help minimize the adverse impact of FMS and FMS/DCS hybrid programs. It may be necessary to sustain FMS cases as a viable option in arms transfers.

Based on the research and analysis in this report, there are best practices across multiple processes, organizations, and systems that could be applied. The following is a summary of the key best practices that are consistently applied across organizations:

- Program offices use integrated product or process teams (IPT) to integrate the requirements for international acquisition in executing their programs and provide appropriate oversight.
- Acquisition organizations use various decision-making approaches to achieve consensus or streamline the decision-making process.
- Acquisition and program offices train their staff in international acquisitions.
- Contracting officers and industry factor in the lengthy process for acquisition approvals, including import–export requirements.

- Program offices use commercial products where possible to reduce development, test, and integration issues when selling to foreign customers.
- Contracting officers use fixed price contracting arrangements, where possible, to reduce risk and manage scope changes.

References

Air Force Studies Board, Division on Engineering and Physical Science, National Research Council. (2015). *Owning the technical baseline for acquisition programs in the Air Force* [Workshop report].

Arms Export Control Act (AECA).

Cressey, L. (2018, August 8). Presentation at the CSIS Forum. Retrieved from https://www.csis.org/analysis/us-arms-transfer-policy-shaping-way-ahead

Defense Acquisition University (DAU). International Acquisition Management Website: https://www.dau.mil/cop/iam/Pages/Default.aspx

Defense Federal Acquisition Regulation Supplement (DFARS) 225.73, Acquisitions for Foreign Military Sales (FMS).

Defense Institute of Security Assistance Management (DISAM).

Defense Institute of Security Cooperation Agency. (2018, May). *Defense Institute of Security Assistance Management Publication*.

Export Rules. (n.d.) Penalties for ITAR noncompliance. Retrieved from http://www.exportrules.com/itar/penalties-for-itar-noncompliance.html

Federal Acquisition Regulation (FAR), 48 C.F.R. ch. 1 (2019).

Federal Acquisition Regulation (FAR) Circular May 29, 2014.

Foreign Assistance Act (FAA).

GAO. (2010, January 14). *Briefing on commercial and Department of Defense requirements and acquisition practices* (GAO-10-315R). Washington, DC: Author.

GAO. (2019, February). *Foreign Military Sales: Observations on DOD's approach to developing price and availability estimates for foreign customers* (GOA-19-214). Washington, DC: Author.

GAO. (2019, February). *Observations on DOD's approach to develop price and availability estimates for foreign customers*. Washington, DC: Author.

Gillespie, K. (2011, November 1). Foreign Military Sales: Army supports international relations through military partnerships. *Soldiers Magazine*.

Hunter, A. P. (2018, August 8). U.S. arms transfer policy: Shaping the way ahead. Presentation at the CSIS Forum. Retrieved from https://www.csis.org/analysis/us-arms-transfer-policy-shaping-way-ahead

Joumier, Freaner, Bitten, & Edmonds. (2012). *Comparison of ESA and NASA acquisition approaches and their potential effects on science mission development duration and schedule change*. Aerospace Corporation.

LMD Defense. (n.d.). Direct commercial sales. Retrieved from http://lmdefense.com/direct-commercial-sales/

Machi, V. (2017, December 18). Air Force foreign military sales skyrocket to $27 billion. *Air Force News.*

NASA. (2014). Global reach: A View of NASA's international cooperation. Retrieved from https://www.nasa.gov/connect/ebooks/global_reach.html

NATO. (1996). *NATO Security Investment Programme—Procedures for International Competitive Bidding* (AC/3-D/221).

NATO. (2004, December 22). Bi-Strategic Command Procurement Directive (Bi-SC Directive No. 60-70).

Navigating NATO procurement: UK Joint Delegation to NATO, Department for International Trade. (n.d.). Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/653378/Navigating_NATO_Procurement.pdf

Satellite Industry Association. (2014). Smart buying—Improving SATCOM procurement.

A New Way to Justify Test and Evaluation Infrastructure Investments

Patricia F. Bronson—is currently a research staff member at the Institute for Defense Analyses (IDA) specializing in requirements analyses, cost estimation, resource planning, and root cause analysis for large, complex, defense systems and systems-of-systems known as major defense acquisition programs. Prior to joining IDA, Dr. Bronson was an operations research analyst for the Cost Analysis Improvement Group within the Office of the Secretary of Defense. Her work products inform investment decisions by senior leadership. Her portfolio includes naval destroyers, combat and missile systems, military aircraft, unmanned aerial vehicles, and space systems. [pbronson@ida.org]

Paul F. Piscopo—provides consultant services to the Institute for Defense Analyses. Before retiring from the DoD in 2004 with 37 years of service, he served as Special Assistant to the Director, Defense Research and Engineering (DDR&E) and Executive Director of the National Aerospace Initiative (NAI) Washington Office. Prior to accepting that position, he served as Chief of Staff for the President's Commission on the Future of the United States Aerospace Industry—and before that, as Associate Director for the Weapon Systems Directorate in the Office of the Deputy Under Secretary of Defense for Science and Technology. [ppiscopo.ctr@ida.org]

Abstract

In 2013, the Congress directed that a study be conducted on the ability of the national test and evaluation infrastructure to effectively and efficiently mature technologies for defense-related hypersonic systems development through 2030. It further required that a report be submitted to the Congress on the study results, along with a plan identifying the capability needs and proposed defense-related investments. The Institute for Defense Analyses (IDA) supported both congressionally directed efforts and was subsequently tasked with providing a business case analysis for the proposed investments. This article describes the IDA-developed methodology used to successfully justify and secure full funding for the proposed five-year, $350 million Department of Defense Test and Evaluation infrastructure investment augmentation.

Introduction

State-of-the-art test and evaluation (T&E) capabilities are essential for successful development of new aerospace products, as well as for the upgrading of currently fielded products. Despite the unarguable fact that system development programs require a robust and continuing investment in research, development, test and evaluation (RDT&E), including the T&E infrastructure, the Department of Defense (DoD) still must justify additional test infrastructure investments needed to effectively and efficiently develop and field future aerospace systems. This has proven to be a major challenge for facility owners and operators.

The Hypersonic T&E Infrastructure Working Group (IWG), established to respond to congressional direction regarding adequacy of the DoD's T&E infrastructure for the development of hypersonic missiles, found capability gaps in the DoD's wind tunnel infrastructure. Their analysis established the need for $350 million in improvements at several facilities. The Institute for Defense Analyses (IDA) was asked to develop a Business Case Analysis (BCA) to support the investment.

IDA proposed using an approach that values the potential programmatic cost savings that could reasonably be expected to accrue during system development from funding proposed T&E capability enhancements.

Background

Figure 1 shows an operational concept for notional hypersonic boost glide vehicles (left) and a scramjet-powered cruise missile (right). Both the strategic and tactical boost glide vehicles share an operational concept for delivering a payload. The Strategic Boost Glide (SBG) vehicle is delivered by a multi-stage ballistic missile, has an extended glide phase inside the atmosphere, and ends in a terminal dive. The Tactical Boost Glide (TBG) vehicle is launched from an aircraft, employs a rocket engine to boost it to hypersonic speeds, has an extended glide phase inside the atmosphere, and ends in a terminal dive.

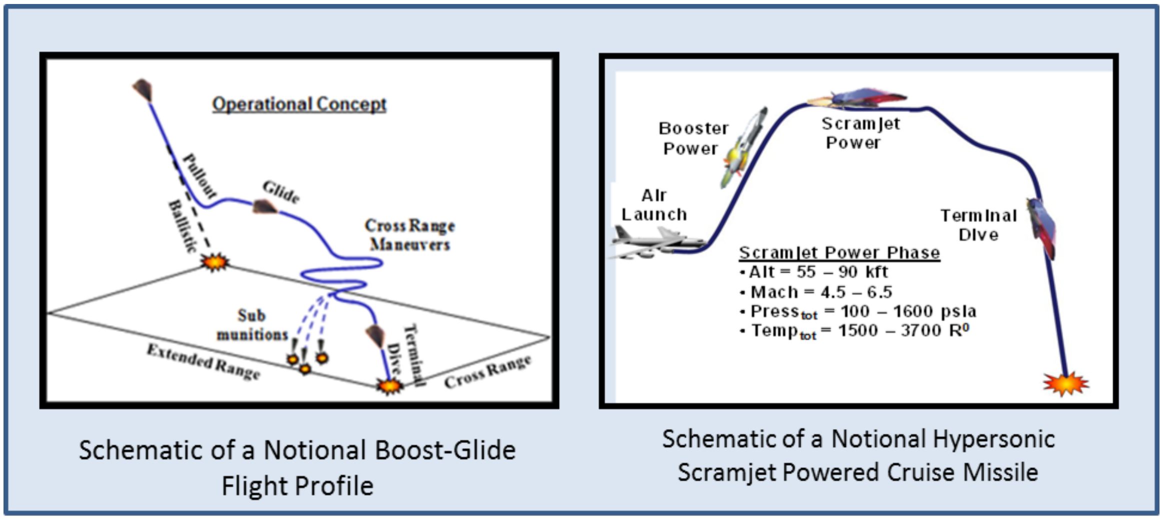


Figure 1. **Conceptual Hypersonic Weapons**

Methodology

First, the research team described a generic hypersonic development program that assumed the capability gaps in the hypersonic T&E infrastructure associated with that design were closed before the program started. Three successful missile Major Defense Acquisition Programs (MDAPs) were used to develop a generic resource-loaded schedule model. Second, the team estimated schedule delays the three conceptual programs might encounter if the capability gaps were not closed. The T&E Infrastructure subject matter experts (SMEs) identified the value of closing the capability gaps in terms of additional flight tests needed during development, based upon flight test failures in their experience. Third, the research team introduced random schedule delays and added resources to the resource-loaded schedule to estimate the final state of the programs. Estimated savings were taken as the difference between the initial and final states. The team created a computer model to simulate the growth over a range of initial conditions. The results reported are the average of 1,000 runs.

Results

Joint Air-to-Surface Standoff Missile (JASSM), Phased Array Track Radar Intercept of Target (PATRIOT) Advanced Capability 3 (PAC-3), and Terminal High Altitude Area Defense (THAAD) were chosen as reference programs. As a group, they bracketed the expected development challenges (each faced a technology readiness challenge) and costs the conceptual hypersonic missile system programs would likely face.

- JASSM is a subsonic stealthy cruise missile that is used to attack surface targets. It is powered by an air-breathing turbojet engine that provides sustained flight in the atmosphere and accomplishes target recognition and terminal homing via infrared (IR) imaging.

- PAC-3 is a tactical, hypersonic, ballistic missile that can achieve speeds of Mach 5+ and intercepts at altitudes of approximately 20 kilometers (km). It was the first MDAP that delivered hit-to-kill technology.

- THAAD is a hypersonic hit-to-kill ballistic missile that employs divert and attitude control technology and an advanced guidance, navigation, and control (GN&C) system to achieve its end-game mission. THAAD pushed the range (approximately 200 km) and altitude (150 km) envelopes beyond the PAC-3 missile.

Table 1 compares the developmental challenges of the reference programs to the three conceptual conventional hypersonic programs.

Table 6. Characteristics of Analogous MDAPs and Conceptual Hypersonic Programs

MDAP	MDAP Attributes	Parallel Conceptual Programs Analogy
JASSM	• Stealthy cruise missile • Sustained subsonic flight in the atmosphere • Air-breathing turbojet engine • Target recognition/homing via IR imaging • Designed to hit surface targets	• Sustained hypersonic flight in the atmosphere • Air breathing scramjet engine • Target recognition and terminal homing • Designed to hit surface targets
PAC-3	• Tactical missile (Mach 5+) • Powered by a solid propellant rocket • Hit-to-kill technology • GN&C/Divert and attitude control	• Tactical missile (hypersonic) • GN&C/autonomous end-game
THAAD	• Hypersonic ballistic missile interceptor • Hit-to-kill technology • GN&C/Divert and attitude control • Extensive flight path (THAAD has an estimated range of 200 km and can reach an altitude of 150 km)	• Hypersonic vehicle • GN&C/autonomous end-game • Extensive flight path/similar altitudes

Figure 2 shows a breakout of development costs for the JASSM, PAC-3, and THAAD programs in billions of dollars ($B) adjusted to FY 2014. (All cost values in this study were in FY 2014 dollars unless otherwise stated.) The cost values were derived from each program's Selected Acquisition Reports (SARs) and Contractor Cost Data Reports (CCDRs). The THAAD system program comprised two major development efforts: the ground radar and the THAAD missile. Only the portion associated with the missile development was used to inform this cost estimate. Cost Estimating Relationships (CERs) were derived from these cost data. Spacing on the horizontal axis is the average Munition Recurring Unit Cost (MRUC) reported during the development phase.

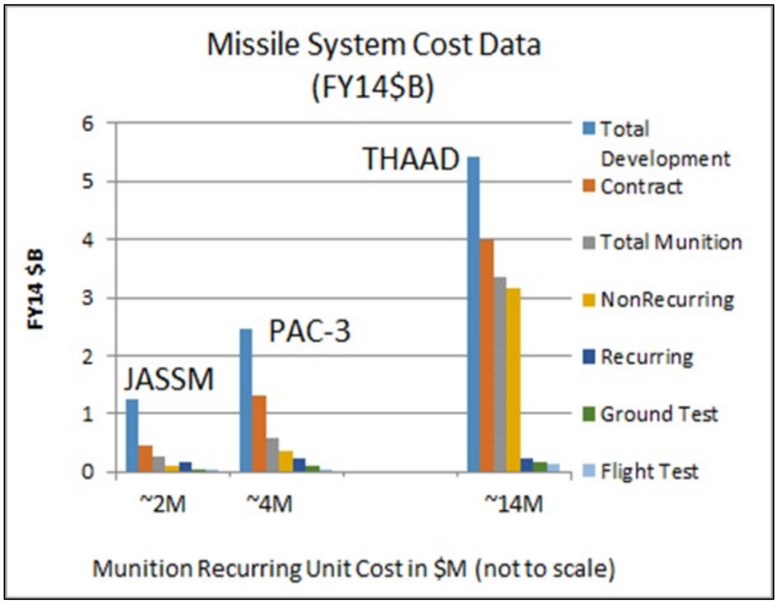


Figure 2. **Actual JASSM, PAC-3, and THAAD Development Costs**

Figure 3 shows the initial estimated and final actual time intervals between Milestone (MS) A and MS C for JASSM, PAC-3, and THAAD as a function of MRUC. These data show initial schedules ranging from five to 10 years and final (as executed) schedules ranging from eight to 17 years. They also show actual schedule delays ranging from two to seven years. The straight lines suggest empirical relationships between development time for MDAPs and the MRUC.

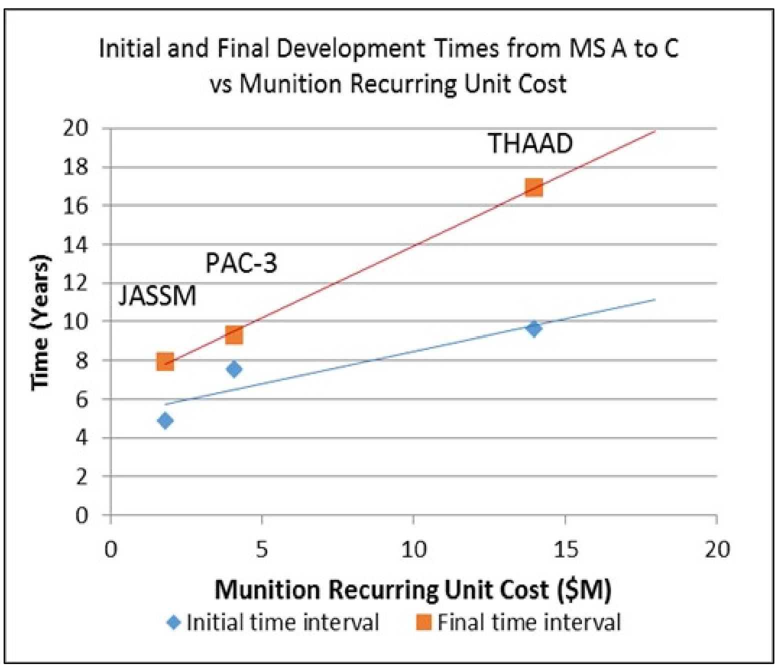


Figure 3. **Actual Initial/Final MS A-to-C Time Intervals**

Figure 4 shows the actual number of flight tests flown as a function of MRUC (calculated from the development CCDR). The number of flight tests displayed in this chart was compiled from actual data gathered from the JASSM Risk Reduction and EMD phases, PAC 3 and its predecessor Flexible Lightweight Agile Guided Experiment and Extended Range Interceptor programs, and the THAAD Program Definition and Risk Reduction (PDRR) and Engineering and Manufacturing Development (EMD) phases. The straight line represents an empirical relationship between the number of flight tests executed on MDAPs and the MRUC.

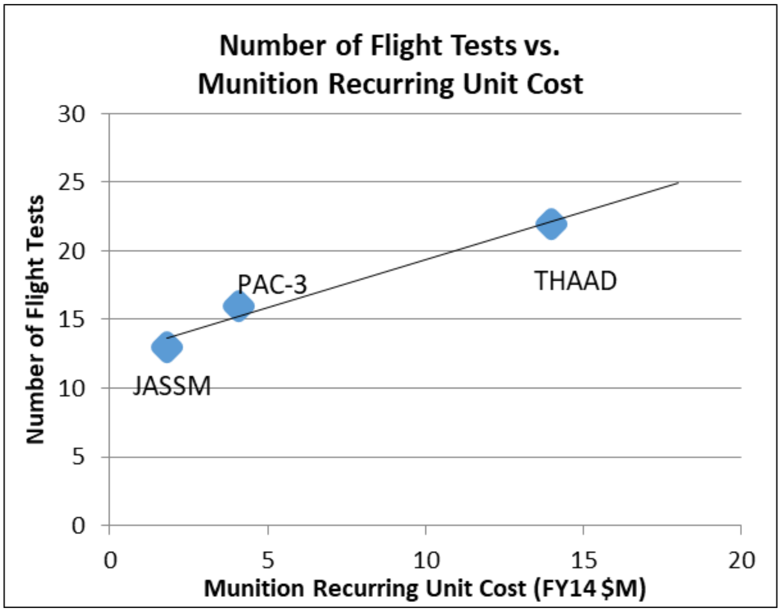


Figure 4. **Actual Number of Fight Tests on JASSM, PAC-3, and THAAD**

Figure 5 presents a frequency histogram of the time between flight tests (known as test centers) for the JASSM, JASSM Extended Range (JASSM-ER), PAC-3, PAC-3 Missile Segment Enhancement (PAC-3 MSE), and THAAD programs, as executed. The IDA research team used these data to inform its flight test schedules. According to these data, 90% of all flight test centers were below 12 months (with design flaws and schedule delays included).

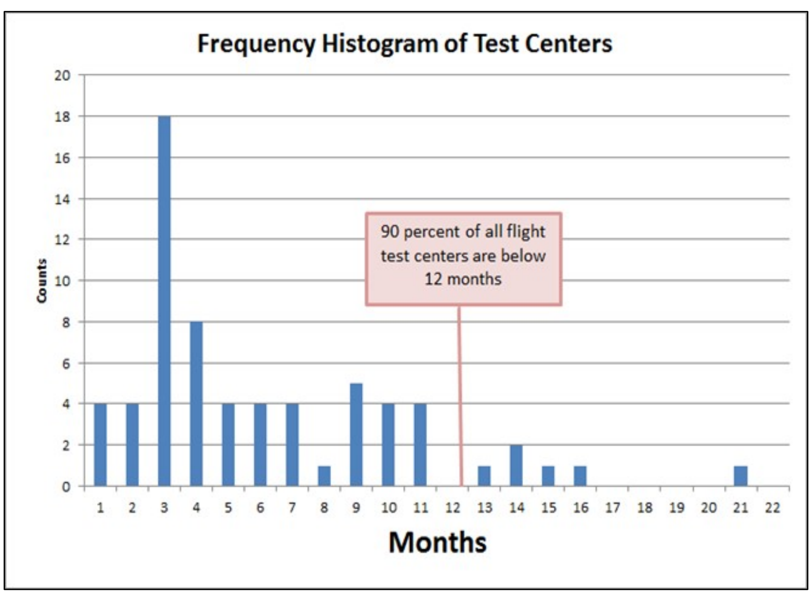


Figure 5. **Actual Flight Test Centers**

Figure 6 depicts a sample resource-loaded schedule for a program executing with adequate T&E infrastructure. The different color bands represent the various elements of cost (as shown in Figure 2). The program depicted has three years of development and ground testing after MS A approval and prior to the first flight test. The flight test program executes with an average of four months between flight test centers. Since this schedule is populated with cost data from a model built with JASSM, PAC-3, and THAAD program data, it includes any design flaws, flight test failures, redesign efforts, and schedule delays inherent in those programs.

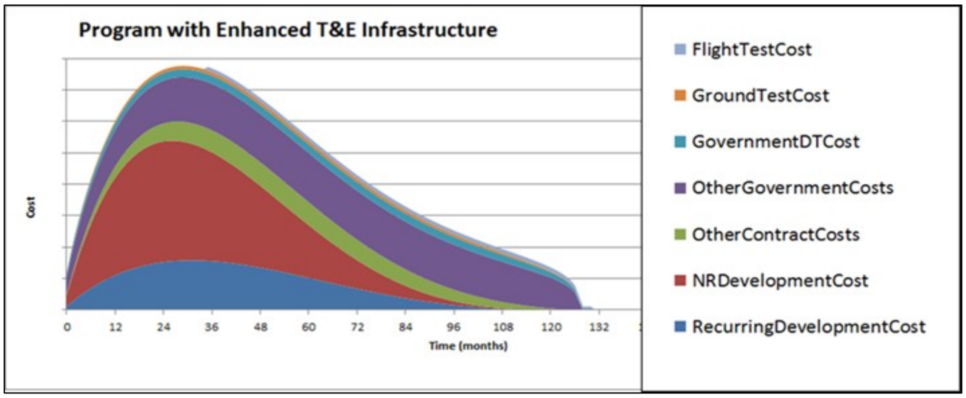


Figure 6. **Sample Initial Resource-Loaded Schedule for a Program With an Enhanced T&E Infrastructure**

The T&E SMEs characterized the design and development problems each of the development programs might expect to encounter if the hypersonic T&E infrastructure were not enhanced prior to MS A, and translated them into an estimated number of additional unanticipated design flaws that would persist past the critical design review. Table 2 shows the SME-generated analysis for the conceptual TBG program; it shows five undetected design flaws in the lower right two columns.

Table 2. SME-Generated Analysis of Estimated Undetected Design Flaws for the Conceptual Boost Glide Program

Conceptual System A (with Enhancements)											
Test Type	Test Objectives Addressed	Est Test Cost ($K)	Est Test Time (weeks)	Number of Ground Tests			Total Cost ($K)	Experimental (Supplements Data)		Undetected Design Flaws (Possible F/T Failures)	
				Pre-MS A	MS A-B	Post MS B		MS A-B	MS B-C	MS A-B	MS B-C
Aero	1.1-to-1.5	4,000	8	2	2	0	16,000	baseline	baseline	baseline	baseline
Aerotherm	2.1-to-2.7	1,000	4	1	1	0	2,000	baseline	baseline	baseline	baseline
Materials	3.4-to-3.11	2,000	26	2	1	0	6,000	baseline	baseline	baseline	baseline
Propulsion	4.2-to-4.3	5,000	12	2	2	0	20,000	baseline	baseline	baseline	baseline
Stage/Store	5.1	500	2	0	2	8	5,000	baseline	baseline	baseline	baseline
Weather	6.1-to-6.3	2,500	12	0	2	2	10,000	baseline	baseline	baseline	baseline
GNC	7.5-to-7.7	2,000	8	0	2	2	8,000	baseline	baseline	baseline	baseline
Lethality	8.1	1,000	8	0	1	2	3,000	baseline	baseline	baseline	baseline
Conceptual System A (without Enhancements)											
Test Type	Test Objectives Addressed	Est Test Cost ($K)	Est Test Time (weeks)	Number of Ground Tests			Total Cost ($K)	Experimental (Supplements Data)		Undetected Design Flaws (Possible F/T Failures)	
				Pre-MS A	MS A-B	Post MS B		MS A-B	MS B-C	MS A-B	MS B-C
Aero	1.1-to-1.5	5,000	10	3	2	1	30,000			1	1
Aerotherm	2.1-to-2.7	2,000	8	2	1	0	6,000				
Materials	3.4-to-3.11	2,500	34	2	1	0	7,500			1	
Propulsion	4.2-to-4.3	7,000	18	2	2	1	35,000			1	
Stage/Store	5.1	500	2	0	2	12	7,000				1
Weather	6.1-to-6.3	2,500	12	0	3	3	15,000	2	4		
GNC	7.5-to-7.7	2,000	8	0	2	3	10,000				
Lethality	8.1	1,000	8	0	1	3	4,000				

Table 3 shows the SME-generated estimates of the capability gaps and design flaws for the three conceptual programs.

Table 3. Resulting Additional Major Design Flaws Resulting From Infrastructure Capability Gaps

Hypersonic Weapon System Type	Number of T&E Infrastructure Capability Gaps	Estimate of the Number of Additional Major Design Flaws
Scramjet Cruise Missile (CM)	10	9
Tactical Boost Glide (TBG)	7	3
Strategic Boost Glide (SBG)	9	5

Figure 7 shows the resource-loaded schedule (from Figure 6) with schedule delays due to the number of design flaws.

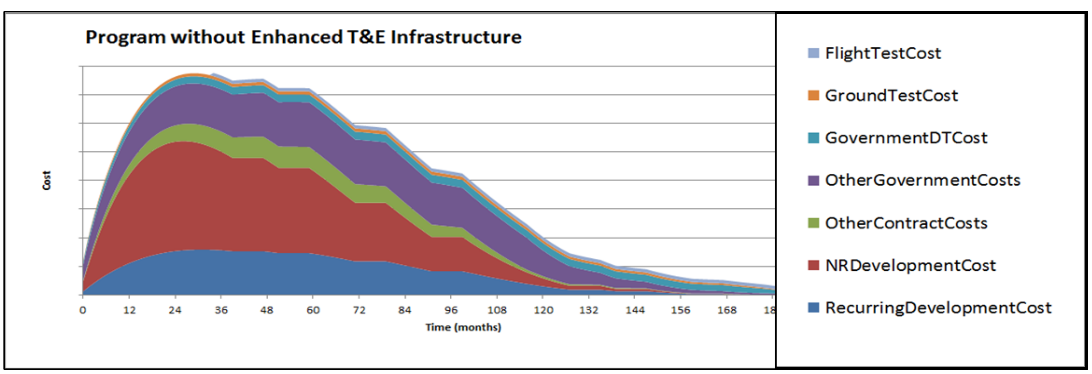


Figure 7. **Sample Resource-Loaded Schedule With Added Schedule Delays**

Table 4 shows the estimated savings for the range of development program costs from $1.3 to $2.9 billion. For reference, the IDA team included the initial RDT&E schedule in years (line 2), the number of flight tests (line 3), and the savings to the three conceptual programs if the unanticipated design flaws are avoided (lower right quadrant).

Table 4. Study Results: Estimated Savings Over a Range of Development Costs

Estimated Savings Over a Range of Development Costs					
Range of Development Costs ($M)		1,300	1.800	2,400	2,900
Initial RDT&E Schedule (Years)		9	10	10	10
Number of Flight Tests		18	21	23	23
	Number of Additional Design Flaws	**Savings if the Design Flaws are Avoided ($M)**			
TBG	3	100	150	200	270
SBG	5	150	240	310	400
CM	9	240	380	530	690

Table 5 shows the calculated (discounted) net savings over the range of estimated development costs from $1.3 billion to $2.9 billion analyzed for the three conceptual systems: Scramjet CM, SBG, and TBG. Each entry in Table 5 is the amount of the cost avoided by making the investment (i.e., the numbers from Table 4 less the $350 million investment). While there was no compelling evidence to make the investment based on the costs avoided for either the TBG or SBG programs, should the DoD decide to pursue both (Table 5, bottom line), the investment option became more attractive.

Table 5. Study Results: Net Savings With Enhanced Hypersonic T&E Infrastructure

Net Discounted Savings				
	Range of Development Costs ($M)			
	1,300	*1,800*	*2,400*	*2,900*
	Savings ($M)			
TBG	-250	-200	-150	-75
SBG	-200	-125	-50	50
Scramjet CM	-125	25	175	325
Both TBG and SBG	-100	25	150	300

Conclusion

The IDA-developed methodology was used successfully to justify and secure a five-year, $350 million T&E infrastructure investment augmentation for the DoD. Potential users of this process, however, are reminded again that it takes substantial time and effort—and success is not guaranteed. In the hypersonic missile arena, preparing the pathway and developing the plan took over three years to complete and required substantial effort not only by the core IDA research team, but also by an extensive support team of government and industry SMEs who provided information and counsel on the key capability needs, the capability gaps, the impacts of not closing the gaps, and the proposed investment plan.

References

National Defense Authorization Act (NDAA) for Fiscal Year 2013, Pub. L. No. 112-239, §1071, 112th Cong. (January 2013).

Office of Management and Budget (OMB). (2015). 2015 discount rates for OMB Circular No. 94 (Memo M-15-05). Washington, DC: OMB.

Piscopo, P. F., Ahmed, H. F., Bronson, P. F., Hallion, R. P., Hong, J. S., Lewis, M. J., & Wu, L. (2016). 2015 Hypersonic T&E investment planning study Phase II final report (IDA Paper P-5313). Alexandria, VA: Institute for Defense Analyses. Draft, not authorized for further distribution.

Piscopo, P. F., Marren, D. E., & Dunn, S. C. (2018). A value-based justification process for aerospace RDT&E capability investments. In Proceedings of American Institute of Aeronautics and Astronautics Science and Technology Forum. https://doi.org/10.2514/6.2018-0388

Test Resource Management Center. (2015). Report on the ability of the U.S. test and evaluation infrastructure to effectively and efficiently mature hypersonic technologies for defense systems development and plan for proposed improvements and modernization. Washington, DC: Office of the Secretary of Defense, TRMC.

Acknowledgments

The authors acknowledge and thank the DoD Test Resource Management Center for its sponsorship and financial support of the research efforts presented herein. In addition, we acknowledge and thank Dr. John Hong, Ms. Hiba Ahmed, and Ms. Linda Wu from IDA, as well as the myriad of SMEs from the Arnold Engineering Development Complex and other government organizations, for their substantial and significant technical efforts and contributions in the development and application of this capability-based valuation approach.

Issues With Access to Acquisition Information in the Department of Defense: Considerations for Managing Program Data in the Emerging Acquisition Environment

Jeffrey A. Drezner—is a Senior Policy Researcher at RAND. He has over 34 years of professional experience conducting policy analysis on a wide range of issues, including planning and program management, analyses of cost and schedule outcomes in complex system development programs, aerospace industrial policy, and defense acquisition policy and reform. His research continues to emphasize mixed qualitative and quantitative approaches to analyze issues associated with technology development, organizational behavior, and program management. Drezner received his PhD in political science from Claremont Graduate University. [zner@rand.org]

Megan McKernan—is a Senior Defense Researcher at RAND. She has more than 14 years of experience conducting DoD acquisition analyses. She is co-leading research examining DoD acquisition data management. She has also conducted other defense acquisition analyses: prototyping; IT acquisition; Industrial Base considerations; tailoring the acquisition process; program manager tenure; and root causes of Nunn-McCurdy unit cost breaches. She uses a variety of methods in conducting research including case studies, interviews, and literature reviews. She holds an MA in international trade and investment policy from The George Washington University and a BA in economics from William Smith College. [mckernan@rand.org]

Jerry M. Sollinger—other co-contributors from RAND.

Abstract

Acquisition data lay the foundation for decision-making, management, insight, and oversight of the Department of Defense's acquisition program portfolio. Recent statutory changes to organizational structures, as well as to roles, responsibilities, and authorities, have introduced new challenges and opportunities for the collection, storage, and use of acquisition information. This research identifies and describes some of the issues and challenges related to managing acquisition program information in this emerging acquisition environment and suggests options for addressing these challenges and opportunities.

Introduction

Acquisition data lay the foundation for decision-making, program management, insight, and oversight of the Department of Defense's (DoD's) acquisition program portfolio. Recent statutory changes to organizational structures, as well as to roles, responsibilities, and authorities (RRAs), have introduced new challenges and opportunities for the collection, storage, and use of acquisition information. These changes—which we collectively refer to as the emerging acquisition environment—may have an impact on acquisition program data governance and management, and what data are needed for acquisition program information in support of program management, analysis, and oversight.

Research Objective and Approach

The objective of this research was to identify and concisely describe some of the issues and challenges associated with managing acquisition program information in the emerging acquisition environment. The intent was to provide timely information to inform some of the policy design and implementation decisions the DoD must make in response to recent changes.

Our approach consisted of three main steps. First, the study team identified and described recent changes to DoD acquisition RRAs. This step was fundamentally descriptive in nature and was accomplished by reviewing relevant legislation and acquisition policy changes, and by interviewing DoD leadership in charge of developing policy to guide or implement the changes. Second, the study team identified a set of specific challenges for acquisition data that may arise from the changes in RRAs. The topics were chosen with approval of the sponsor but were informed by six earlier studies on Issues With Access to Acquisition Data and Information in the Department of Defense (Riposo et al., 2015; McKernan et al., 2016, 2017, 2018; unpublished 2018 and 2019 research by Jeffrey A. Drezner, Megan McKernan, Badreddine Ahtchi, Austin Lewis, and Douglas Shontz, Ken Munson, Devon Hill, Jaime Hastings, Geoffrey McGovern, Marek Posard, and Jerry Sollinger). Several topics were ultimately selected:

- General data governance and management issues associated with the emerging acquisition environment;
- Specific data challenges associated with the implementation of the Middle Tier of Acquisition for Rapid Prototyping and Rapid Fielding; and
- Implications of termination of the Selected Acquisition Report (SAR).

Third, the study team identified implications, potential opportunities, and risks for acquisition data for each of the identified topics, as well as general guidelines to consider when strategically managing data. Implications were developed on the basis of published best practices for data management and an understanding of how those practices are currently implemented in the DoD acquisition system. Where possible, the study team also identified how current DoD policies and practices may need to change to become consistent with the emerging and future acquisition environment (in terms of roles, responsibilities, and structure) and identified options for mitigating the challenges.

Key Scoping Assumptions

Recent changes in acquisition RRAs prompted a question about what acquisition data are required for the Under Secretary of Defense for Acquisition and Sustainment (USD[A&S]) to execute its evolving acquisition responsibilities. Ultimately, acquisition program data requirements are a decision for USD(A&S) that depends on how USD(A&S) intends to use acquisition program data (i.e., the "use cases") and on the financial costs and potential managerial and administrative burdens the DoD is willing to accept to collect, manage, store, share, and govern acquisition program data and information relative to the benefits of having the data. This research did not address this basic question, which bears on broader questions of acquisition policy; instead, we assume that USD(A&S) will continue to need acquisition program data to support a broad set of use cases. These use cases include the following:

- Statutory and regulatory reporting
- Tracking program cost, schedule, and performance outcomes against an established baseline
- Providing program insight and oversight to anticipate, understand, and mitigate the factors affecting adverse cost, schedule, and performance outcomes
- Conducting portfolio analyses, including both traditional (i.e., by service or weapon system type) and new analyses (i.e., mission-focused capabilities and kill-chains)

- Understanding the performance of the overall acquisition system, or any specific pathway within that system (e.g., traditional, tailorable DoD Instruction [DoDI] 5000.02; middle tier) to inform improvements in policy and process design and implementation

This assumption scopes our analysis, since ultimately USD(A&S) may decide that some of these use cases (or their specific instantiations) are no longer needed in the new environment, or that the costs and potential burdens associated with collecting, managing, storing, sharing, and governing acquisition program data cannot be justified given their benefits. Analyses of such trade-offs are left for future work.

The topics listed above address only a few of the challenges associated with acquisition program data governance and management due to the recent Office of the Secretary of Defense (OSD) reorganization, change in the milestone decision authority (MDA) for major acquisition programs, and other changes in RRAs. They represent a sample of challenges the DoD will need to confront in the emerging acquisition environment. This paper presents a summary of these challenges and opportunities.

Summary of Recent Changes to Acquisition Roles, Responsibilities, and Authorities

The Fiscal Year (FY) 2016 and FY 2017 National Defense Authorization Acts (NDAAs) included changes in the roles, responsibilities, and organizational structure of service and OSD organizations managing and overseeing acquisition programs. Section 825 of the FY 2016 NDAA delegated decision-making to the service acquisition executives (SAEs) for new major defense acquisition programs (MDAPs); the SAE is now the default milestone decision authority (MDA) for new MDAPs, and the service chiefs have an increased role in acquisition decision-making, including requirements and program management decisions.

Section 901 of the FY 2017 NDAA changed the structure of acquisition organizations within OSD. Beginning in February 2018, the USD(Acquisition, Technology, & Logistics) was dissolved, and two new under secretariats were created: Acquisition and Sustainment (A&S) and Research and Engineering (R&E). In addition, a chief management officer (CMO) position was established in OSD and given responsibility for

> establishing policies on, and supervising, all business operations of the Department, including business transformation, business planning and processes, performance management, and business information technology management and improvement activities and programs, including the allocation of resources for business operations, and unifying business management efforts across the Department. (National Defense Authorization Act for Fiscal Year 2017)

Some overlap in acquisition program information management may now exist within the DoD among these three positions—USD(A&S), USD(R&E), and CMO—and their accompanying organizations.

Congress also directed the DoD to reemphasize the use of prototyping and reduce acquisition timelines. Section 804 of the FY 2016 NDAA directs the creation of a "middle tier of acquisition for rapid prototyping and rapid fielding," and Section 806 of the FY 2017 NDAA establishes additional processes and reporting on prototyping within the services. In the FY 2018 NDAA, Congress repealed the submission of a SAR for each major acquisition program to Congress, effective December 31, 2021.

At the time of this writing (March 2019), the services and OSD have implemented these structural changes but are still working through some policy and implementation details (DoD, 2017). One area that will be affected by these policy and implementation decisions is that of acquisition program data and other associated acquisition information. Such information is currently generated, collected, stored, accessed, and used by a wide range of organizations in the Services, OSD, and external organizations (e.g., Congress, academic researchers, and federally funded research and development centers [FFRDCs]). Implementing these changes in policy, organizational roles, responsibilities, and structure will necessarily impact the generation, collection, storage, and use of acquisition data. In particular, the changes may obfuscate the authoritative source of specific data, disrupt collection, and limit access and use. As responsibilities move to the Services, their staff may need to develop new or expanded capabilities, particularly in terms of oversight and portfolio management.

General Data Governance and Management Issues Associated With the Emerging Acquisition Environment

As with any large complex organization, the DoD faces challenges related to data access and management. Prior to the current reorganization and statutory changes, the challenges affecting acquisition information included complex security policies regulating information systems; cultural and technical barriers to accessing and sharing information; and lack of awareness of the breadth and depth of information available to DoD leaders and staff. A rich set of information is available to support acquisition insight, analysis, and decision-making, but the full extent to which this information is used remains unknown. In addition, no common data environment exists for all acquisition information, and there is no agreement on all data needs and definitions across the DoD: Both issues result from decentralized governance and management. While most of the underlying data used for program management and oversight/insight are similar across OSD and the services (at least for Acquisition Category [ACAT] I programs), specific metrics and uses differ. For example, all organizations use program cost, schedule, and performance data measured against a baseline; however, OSD tracks only those schedule events contained in the baseline, while the Services tend to provide that information as well as a more complete and integrated picture of schedule. Finally, introducing changes to rules regarding controlled unclassified information (CUI) will further complicate management, sharing, and use of acquisition information.

Key questions senior acquisition leaders need to consider include the following:

- What information does OSD and the Fourth Estate need and why? In particular, what does USD(A&S) need to execute the USD's statutory responsibility to advise the SAEs on acquisition decisions, to inform policy-making, to inform the Secretary of Defense and Deputy Secretary of Defense for program status and portfolio analyses, and to report to Congress?
- Is it possible to have decentralized program execution and oversight while maintaining OSD insight on policy effects, institutional performance, and key program status and outcomes?
- How will portfolio performance be monitored and improved in this decentralized structure, especially with respect to integrated mission and kill-chain capabilities?
- How can data and insight improve the execution of programs?
- What data capabilities will be lost if some information flows stop?
- What information is no longer needed (or of low value)?

- What critical new information is needed?
- Can and should acquisition program data be standardized across the DoD enterprise and across different services and types of programs? Which data?
- What are the military departments doing with their information flows as their organizations change?
- What costs and burdens are associated with collecting, managing, storing, sharing, and governing acquisition program data?

To address these challenges, USD(A&S) could begin by creating a strategic management plan for acquisition information that identifies what acquisition program information is needed by whom to accomplish enterprise-wide objectives without overburdening the military departments. Creating this plan will require elaborating on the acquisition data use cases. Given such a strategic management plan, USD(A&S) and the military department leadership could then work together to standardize a core set of data elements, data definitions, authoritative sources, and management approaches. This effort would facilitate information sharing and understanding; align data governance and management across organizations, use cases, and program types; and be an important substantive step toward a common acquisition data framework. This effort could start with the existing data governance and management framework that has enabled standardized data reporting for ACAT I programs over the last several decades.

Specific Data Challenges Associated With the Implementation of the Middle Tier Acquisition Pathway

The new Middle Tier acquisition pathway illustrates many of the challenges just described. The Middle Tier pathway—consisting of both rapid fielding and rapid prototyping—is an alternative acquisition process intended to accelerate the delivery of capabilities to the warfighter. It provides a blanket waiver to both the traditional acquisition (DoD Directive 5000.01) and requirements (Joint Capabilities Integration and Development System) processes. Implementation of the Middle Tier pathway requires program data to inform both programmatic and policy decisions. Interim guidance from the USD(A&S) provided parameters regarding information requirements for the Middle Tier (USD[A&S], 2018a, p. 3; USD[A&S], 2018b; USD[A&S], 2019). It also identified an initial set of core information that should be collected regarding these efforts (at a minimum) and discussed a data-driven collaborative policy-making process that will draw on lessons learned from the initial implementation. The Navy and the Air Force released guidance in April 2018, with the Air Force following up with additional detailed guidance in June 2018 (Assistant Secretary of the Navy [Research, Development and Acquisition], 2018, pp. 1–3; Assistant Secretary of the Air Force for Acquisition, Technology and Logistics, 2018; Assistant Secretary of the Air Force [Acquisition, Technology & Logistics], 2018, pp. 7–8; Assistant Secretary of the Army for Acquisition, Logistics and Technology, 2018). One major similarity between the Navy and Air Force guidance is the emphasis on tailoring current statutory and regulatory information requirements and seeking waivers as needed to minimize information requirements and help maintain schedule, making tailoring a key tool that program managers will need to use. Service guidance suggests that tailoring should be driven by the unique characteristics of the Middle Tier efforts and by the decisions being made by the milestone decision authority.

Middle Tier acquisition will need to address and resolve many of the challenges that have faced traditional acquisition processes in the past. These challenges include the following:

- Determining exactly what data are reported for a middle tier "program," at what frequency, and how. While the USD(A&S) and service guidance memoranda address this issue, they do not resolve it.
- The service guidance memoranda reflect a lack of standardization across organizations in terms of what should be reported, relying instead on tailoring data reporting to reflect the characteristics of each program. No guidance is provided on how to tailor or how to determine what is appropriate for a given middle tier activity.
- The objective of the Middle Tier pathway is speed. There is a risk that the process could become overburdened by reporting requirements, slowing it down.

The Middle Tier acquisition pathway also illustrates how the existing data infrastructure (information systems, data collection conventions, common data definitions) can support and adapt to new acquisition authorities and processes. While adjustments and refinements of Middle Tier data collection will occur as experience is gained with the new processes, the existing IT infrastructure and data environment in OSD and the services could be adapted to support the information needs of the Middle Tier pathway, while maintaining some degree of alignment and consistency across the traditional acquisition pathways and across organizations.

Implications of Termination of the Selected Acquisition Report (SAR) to Congress

The submission of a SAR for each major acquisition program to Congress was repealed by the FY2018 National Defense Authorization Act (NDAA), effective December 31, 2021. While this change was part of Congress' broader effort to ease the DoD's reporting burden, the change creates an opportunity for the DoD to review and propose a revised reporting structure that satisfies Congress's need for detailed, transparent performance information but in a way that the DoD finds more efficient and effective. The SAR has been a bedrock of transparency and data on the cost, schedule, and performance of MDAPs for oversight and analysis at the program, portfolio, and policy levels—both immediate and longitudinally. Analyses using SAR data have been useful to improving and informing weapon system acquisition strategies and policymaking in the DoD and Congress for decades. Here we discuss some of the consequences of terminating the SARs.

The SAR has been used for about 50 years to understand and track MDAP cost, schedule, and performance. SARs are important because collectively they provide a structured and relatively consistent mechanism for informing Congress on the performance of major investments, are useful for management and oversight, and are one of the only sources of longitudinal, standardized program information supporting program, portfolio, and process analysis for MDAP investments. The data included in the SAR constitute a starting place for developing common acquisition program data management across all program levels, program types, and components. The data also serve as a source of useful information for the development of acquisition strategies and system life-cycle management processes, as well as independent cost estimates.

If not replaced with another reporting construct that provides consistent longitudinal data across programs, the elimination of this information source by Congress could, in turn, eliminate many of the benefits that have accrued from its use over time. Of particular concern is the potential loss of common data standards and definitions for measuring program performance and a source for cost, schedule, and performance data for independent program milestone assessments and policy analysis. Without these common data standards and definitions (i.e., a common data framework) institutionalized over decades of SAR creation and submission, the military departments' performance measurements (i.e., definitional standards) may drift over time, leading to reduced transparency and inefficiencies if additional work is required to reconcile disparate data during analysis. Also, the statutory status of the SAR serves an important enforcing function for compliance.

However, opportunities may also exist to improve on the SAR for future reporting constructs while still retaining some of the key data elements. Some data elements of the SAR are useful for information or analytic purposes while others could be improved, streamlined, or eliminated. Below are some example opportunities, core elements, and hidden needs. These examples highlight known uses that further analysis could refine to improve SAR-like reporting to multiple user communities.

Streamlining and integration with other information sources. The DoD could review and integrate sources of similar information to reduce burden and increase efficiency rather than creating pieces solely for the SAR. For example, the SAR's Executive Summary—an authoritative source of program history, status, purpose, and plans—could be sourced from or integrated with other similar sources.

Revision of certain elements. Some elements are known to be either problematic or particularly burdensome with little value, while others are valuable but require significant additional work to prepare. For example, the SAR Cost Variance section is known to have theoretical issues in how cost change types are allocated to statutory bins. However, some of this information has been useful for informing (in part) analysis of trends in cost variance and root cause analyses. The DoD could develop and propose an alternative approach that is less confusing and more informative. Operating and Support (O&S) cost data are valuable to those who are seeking to understand high-level O&S costs, but the data included in the SAR do not provide insight into how these costs, their uncertainties, and changes over time may be due to external factors beyond the control of the acquisition system. One possible improvement might include adding data on elements that drive sustainment costs (e.g., more consistent provision of reliability information and enriched information on maintainability).

Key elements for retention (including some that seem obscure and burdensome). Some elements are useful, but their utility may not be apparent, given the amount of work involved in preparing them. Two examples are provided here. Schedule events—and how well the program is doing against their baseline thresholds and objectives—can be used to help understand program timelines from Milestone B through C. They allow analysts to identify how long acquisition takes (cycle time) and any schedule growth. Unit Costs are used to directly identify whether programs have breached congressionally mandated Nunn-McCurdy cost thresholds and the associated reporting, review, restructuring, or cancellation activities required by law (10 U.S.C. 2433). The SAR record for a program also allows one to identify what baseline is used for a program's reported unit cost growth.

USD(A&S) could take the opportunity presented by Congress to reassess, improve, and streamline the current information contained in the SAR, the structure of the SAR itself, and the process by which this information is reported to Congress and DoD. The SAR itself does not necessarily need to be preserved, but the program data it contains need to continue to be collected and disseminated to both internal DoD and external stakeholders. The core data requirements for a range of use cases—from Congressional reporting to portfolio analysis—are supported by the current set of data elements contained in the SAR.

Principles for Moving Forward

Based on prior research, we offer the following four guidelines to ensure that requirements and processes associated with Middle Tier program data and other acquisition information are as efficient and effective as possible:

- Let decision-making drive data requirements. Data and information must not be generated for its own sake but must support important decision-making about policy, process, programs, and integrated capability outcomes. As a starting point, USD(A&S) can describe data requirements by specifying important acquisition use cases that must be supported.
- Minimize reporting requirements and costs more generally. Information and documentation requirements should be austere, with minimal data reporting. Historically, successful rapid prototyping and fielding activities have had austere information requirements. Guidance appears to recognize this by emphasizing tailoring.
- Standardize where possible. A common acquisition program data framework should be developed for a core set of program data. The existing data framework reflected in the legacy SAR provides a strong foundation from which to start.
- Capitalize on existing structures. One way to minimize costs and burdens (including ad hoc data calls) is by using existing data frameworks, information systems, and organizations to the maximum extent practical, especially when such data are shared automatically between systems.

References

Assistant Secretary of the Air Force for Acquisition, Technology, & Logistics. (2018, April 10). Seven steps for incorporating rapid prototyping into acquisition [Memorandum]. Washington, DC: Department of the Air Force.

Assistant Secretary of the Air Force for Acquisition, Technology, & Logistics. (2018, June 13). Air Force guidance memorandum for rapid acquisition activities [Memorandum]. Washington, DC: Department of the Air Force.

Assistant Secretary of the Army for Acquisition, Logistics, & Technology (ASA[ALT]). (2018, September 25). Office of the Assistant Secretary (Acquisition, Logistics, & Technology) of the Army middle-tier acquisition policy [Memorandum]. Washington, DC: Department of the Army.

Assistant Secretary of the Navy (Research, Development and Acquisition). (2018, April 24). Middle tier acquisition and acquisition agility guidance [Memorandum]. Washington, DC: Department of the Navy.

DoD. (2017, August). Report to Congress restructuring the Department of Defense Acquisition, Technology and Logistics Organization and Chief Management Officer Organization, in response to Section 901 of the National Defense Authorization Act for Fiscal Year 2017 (Public Law 114-328). Washington, DC: DoD.

McKernan, M., Moore, N. Y., Connor, K., Chenoweth, M. E., Drezner, J. A., Dryden, J., … Szafran, A. (2017). Issues with access to acquisition data and information in the Department of Defense: Doing data right in weapon system acquisition (RR-1534-OSD). Santa Monica, CA: RAND Corporation.

McKernan, M., Riposo, J., McGovern, G., Shontz, D., & Ahtchi, B. (2018). Issues with access to acquisition data and information in the Department of Defense:

Considerations for implementing the Controlled Unclassified Information Reform Program (RR-2221-OSD). Santa Monica, CA: RAND Corporation.

McKernan, M., Riposo, J., Drezner, J. A., McGovern, G., Shontz, D., & Grammich, C. (2016). Issues with access to acquisition data and information in the Department of Defense: A closer look at the origins and implementation of controlled unclassified information labels and security policy (RR-1476-OSD). Santa Monica, CA: RAND Corporation.

National Defense Authorization Act for Fiscal Year 2017, Pub. L. No. 114-328, § 901.

Riposo, J., McKernan, M., Drezner, J. A., McGovern, G., Tremblay, D., Kumar, J., & Sollinger, J. (2015). Issues with access to acquisition data and information in the Department of Defense: Policy and practice (RR-880-OSD). Santa Monica, CA: RAND Corporation.

Under Secretary of Defense for Acquisition and Sustainment. (2018a, April 16). Middle tier of acquisition (rapid prototyping/rapid fielding) interim authority and guidance [Memorandum]. Washington, DC: DoD.

Under Secretary of Defense for Acquisition and Sustainment. (2018b, October 9). Middle tier of acquisition (rapid prototyping/rapid fielding) interim guidance [Memorandum]. Washington, DC: DoD.

Under Secretary of Defense for Acquisition and Sustainment. (2019, March 20). Middle tier of acquisition (rapid prototyping/rapid fielding) interim guidance 2 [Memorandum]. Washington, DC: DoD.

O&M Cost Modeling for the Department of Defense

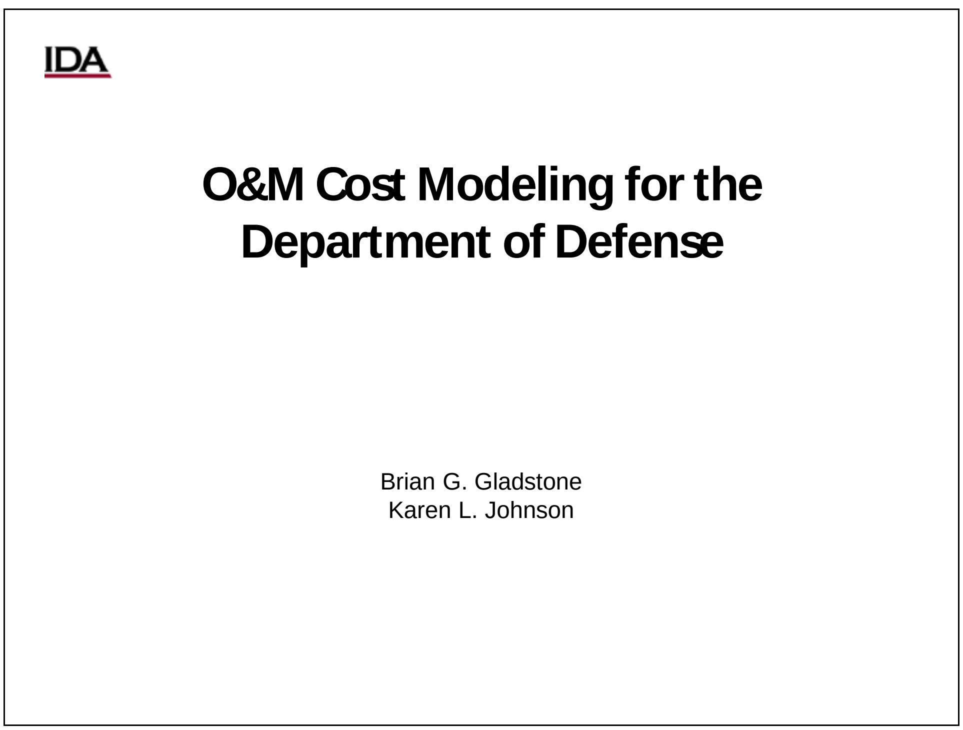
Brian G. Gladstone—is a Research Staff Member at the Institute for Defense Analyses (IDA) whose expertise includes operations research analysis, resource analysis, cost analysis of Major Defense Acquisition Programs, cost analysis of operation and sustainment costs, acquisition policies, industrial base matters, and evaluating corporate costs. He previously worked as an analyst in the Office of the Under Secretary of Defense for Acquisition, Technology, & Logistics and the OSD Office of Cost Assessment and Program Evaluation in the Pentagon. He holds an MS in chemical biology from Harvard University and a BS from the University of California, Los Angeles. [bgladsto@ida.org]

Karen L. Johnson—has worked for IDA for over 26 years and holds a BS in computer science from Clemson University and a Master of Science in Computer Information Systems from American University. She has a wide range of expertise and experience designing and developing cost model applications, relational databases, system architectures, and technical solutions for various projects. She has gained a great deal of cybersecurity expertise maintaining the security postures of SIPRNet applications. She has also contributed to Homeland Security's SAFETY Act, performing reviews of applications for anti-terrorism technology SAFETY Act certifications. [kjohnson@ida.org]

Abstract

The Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics asked the Institute for Defense Analyses to evaluate and update their Operation and Maintenance (O&M) cost model. This document presents the projections of O&M expenditures from revised O&M models for the Department of Defense and the Services, updated for the fiscal year 2015 President's Budget (PB15). This report shows differences between model projections and PB15 requests for O&M for three of four Services as well as for the Department at large.

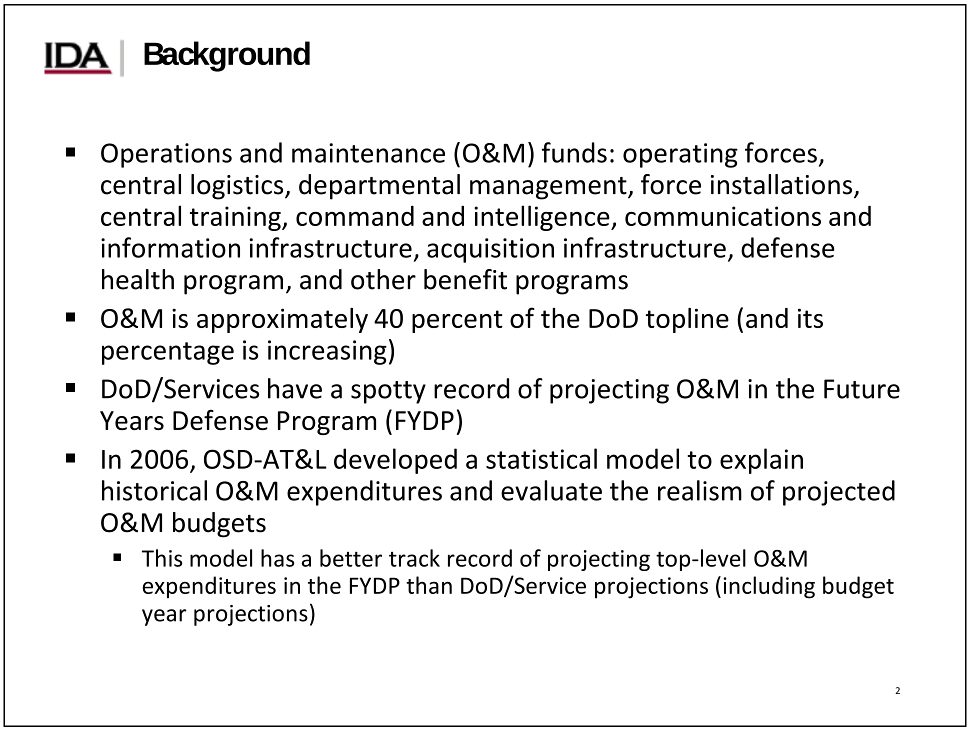
O&M Cost Modeling for the Department of Defense



Slide 1. Presentation Cover Slide

This document is an annotated version of a briefing prepared by the Institute for Defense Analyses (IDA) for the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]). The briefing was delivered to the sponsor, the OUSD(AT&L) office in Acquisition Resources and Analysis (ARA), on May 6, 2014. It presents Operation and Maintenance (O&M) cost estimates for the Department of Defense (DoD) and the Services as well as a cost model to evaluate future O&M costs.

Background



Slide 2. Background

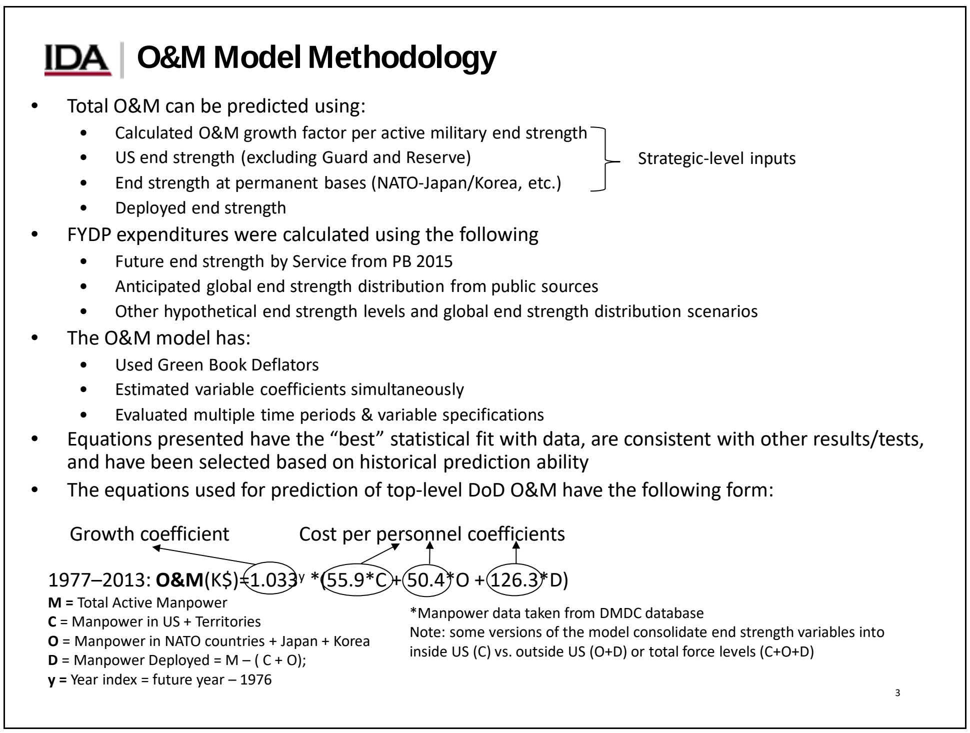
The OUSD(AT&L) asked IDA to evaluate and update a model for projecting future O&M expenditures and develop an automated O&M estimating tool for use by AT&L staff. O&M expenditures are of particular interest to the OUSD(AT&L) because they are a large and growing component of the DoD topline. In the President's Budget for fiscal year 2015 (PB15), O&M expenditures are the single largest funding title, accounting for 45% of the total DoD budget in Fiscal Year (FY) 2014. This is up from 40% of the DoD topline throughout much of the 2000s.

O&M funding supports a wide variety of day-to-day activities, such as individual and military unit training, equipment maintenance, base operations and facilities sustainment, personnel acquisition and management, and certain administrative and Service-wide activities. Despite the size and importance of the O&M account, expenditures were consistently greater than cost projections during the past two decades of Future Years Defense Program (FYDP) submissions. This behavior results in risk to military readiness, continuity of core DoD activities, and investment programs (i.e., procurement and research, development, test, and evaluation [RDT&E]), since all of these accounts have been historical "bill-payers" for O&M shortfalls when DoD budgets are decreasing and supplemental funding is scarce.

Recently, O&M shortfalls have been corrected in supplemental budget requests. During Operation Enduring Freedom (OEF), Operation Iraqi Freedom (OIF), and its successor Operation New Dawn (OND), supplemental funding for Overseas Contingency Operations (OCO) has funded a significant percentage of total O&M expenditures. As these contingency operations come to a close, and military personnel return to the United States, the availability of OCO or other supplemental funding is uncertain. Additionally, recent fiscal pressures on the DoD topline, and the federal government, constrain budgetary flexibility. These current conditions necessitate accurate DoD base budget O&M cost estimates to ensure military readiness, continuity of core DoD activities, and minimal disruptions to investment programs.

In 2006, the OUSD(AT&L) developed O&M cost models that use military end strength and global posture (as proxies for operations tempo [OPTEMPO]) to estimate Service and DoD O&M costs. Historically, this model's FYDP O&M estimates were more accurate than Service and DoD FYDP cost projections in predicting O&M expenditures over the FYDP.

O&M Model Methodology



Slide 3. O&M Model Methodology

The O&M model described on Slide 3, as well as its AT&L predecessors, estimates O&M expenditures based upon two types of variables: an O&M per person growth factor and the end strength/global posture of the active duty military. Both end strength and global posture variables serve as OPTEMPO proxies since they respond to the demands on the military in a similar manner. For most Service O&M cost estimates, and the top-level DoD O&M estimate, these variables statistically explain more than 90% of the historical variability in O&M expenditures since the 1970s (shown later).

The O&M per active duty military end strength growth factor is the first component of the O&M model. Surprisingly, the long-run average macroscopic DoD growth factor has been relatively stable at about 3.5% per year per active duty end strength, when global posture is also considered (see regression results on Slide 5). Substantial annual historical O&M cost changes are mathematically explained by this real-growth factor; however, the underlying causes of this growth are complex. Example components of this cost growth factor may include changes in (1) military technology (old versus new); (2) military benefits; (3) military readiness; (4) DoD business practices; (5) external markets; (6) accounting and budgeting practices; (7) the cost or amount of equipment per active duty military end strength; and (8) changing military practices (i.e., conversion to an all-volunteer force, use of contractors in lieu of military personnel [e.g., contractor logistics support], etc.). Perceived real O&M growth per active duty military end strength can also be the result of errors in proscribed inflation indexes.

In addition, this growth can arise from both "beneficial" and "detrimental" changes to military operations, investments, personnel, and more. For example, manpower reductions resulting from a more efficient military (perhaps due to more complex and expensive equipment) can cause the O&M rate (O&M cost/military end strength) to increase because military manpower (end strength) is shrinking in the denominator of this factor, and the cost of the more complex equipment (which usually has greater O&M costs) is increasing in the numerator. However, this O&M cost growth may be offset by total cost savings elsewhere (such as in Military Personnel [MILPERS]), and may be considered "beneficial" to the Department. Conversely, O&M cost growth due to maintenance costs for aging facilities and equipment, new and more expensive equipment (that does not have an impact on end strength), or other reasons may be considered "detrimental" by some. This report does not evaluate benefits or harms that accrue from increasing or decreasing O&M expenditures per man. Nor does it evaluate the relative magnitudes of beneficial versus detrimental O&M cost changes; it does not indicate whether the level of past, present, or future O&M funding was optimal for the DoD. Further investigations could examine these underlying causes and impacts in depth, but such analysis is beyond the scope of this paper.

Active duty military end strength and global posture variables are the second component of the O&M model. End strength, global posture, and OPTEMPO are all logically likely to change as a result of major military conflicts and tensions and are expected to be O&M cost drivers. In fact, the timing of actual O&M increases and decreases during 1962–2014 correlates well with wartime and peacetime, respectively (not shown). Each of these variables is also expected to cause an enterprise-wide effect on O&M costs. Although top-level measurements of end strength and global posture correlate with major military conflicts, aggregate OPTEMPO is significantly more difficult to measure at macroscopic levels over time. In addition, end strength and global posture are strategic input variables to the DoD's plans that are recorded, tracked, and predicted. They are measurable at any time and are not subjective.

Although the general concept of estimating O&M with end strength and global posture variables remains from 2006, the composition of the variables has evolved. Previous versions of the O&M model estimate O&M using a growth factor and one of the following:

- End strength in the United States and end strength abroad
- 3–5 geographical regions (United States, Europe, Asia, Middle East, Other)
- By geographic combatant command

This paper's version of the model uses three force distribution and end strength variables: (1) active duty military end strength in the United States, (2) active duty military end strength in permanent overseas bases (which include North Atlantic Treaty

Organization [NATO] countries, Japan, and Korea), and (3) other active duty end strength deployed outside of the United States or permanent overseas bases.

These force global distribution variables are selected because OPTEMPO, which is responsible for a large portion of O&M costs/person, should be different, on average, for each category. For instance, the enterprise-wide O&M cost per person of troops stationed in the United States and overseas bases should be less than for deployed forces because OPTEMPO is lower, and logistics and supporting infrastructure (headquarters, bases, etc.) are defined and less demanding. It is unknown whether U.S. O&M cost per person is greater than permanent overseas base O&M cost per person, because the impact of host-nation support is not clear. Each of these hypotheses is generally supported by model estimates of O&M per-person cost coefficients (shown on Slide 3). It should be noted that the U.S. O&M cost per person and the permanent overseas bases O&M cost per person coefficients are not statistically different from one another (their 95% confidence intervals overlap and their coefficients are similar).

Historical actual data for force global posture is available from Defense Manpower Data Center (DMDC) to develop the O&M models. Because DMDC does not provide global posture forecasts over the FYDP, force global posture estimates during 2015–2019 used in this paper are derived from publicly available sources, including statements by DoD officials, budget or planning documentation, and accounts in the news media. The O&M cost model uses these data sources to produce logical estimates of future O&M costs. In addition, hypothetical end strength and force distribution scenarios during 2015–2019 are model inputs, used to perform “what if” O&M cost estimates. For instance, the lowest number of military deployments in recent times occurred in 1997. Using the end strength and global posture from 1997 as a model input to each year of the FYDP predicts a possible lower bound O&M cost estimate, and assumes that DoD behaviors, active military end strength, and worldwide distribution are similar to 1997 in the future.

The O&M models selected for this briefing are displayed on Slide 5 through Slide 7). Numerous versions of the O&M models were evaluated in this work. We selected models that had logical coefficients and promising descriptive statistics, and appeared to make robust forecasts over time.

The first model characteristic we evaluate is the length and duration of the O&M cost modeling time period. A cost modeling methodology is developed by systematically running multiple regressions using data from different historical time periods, comprising varying numbers of years. The 1977–2013 time period is selected because it is the era of the all-volunteer force, requires fewer data adjustments and standardizations than earlier time periods, and is robust in its forecasts.

Different force distribution variables are also evaluated, including:

- “United States” and “abroad” (two variables)
- “Deployed” and “non-deployed” (two variables)
- “Ashore” and “afloat” (two variables)
- Total DoD end strength (one variable)

It should be stressed that the O&M models described do not predict an optimal level of DoD or Service O&M funding. These O&M models forecast the Services’ and DoD’s likely O&M spending based on the relationship of historical O&M costs to historical global posture, active duty military head count, and O&M growth/person. These historical time periods include times of O&M funding abundance and times of suboptimal O&M funding practices (e.g., the hollow force era after Vietnam and the “procurement holiday” era after the Cold

War). In addition, these O&M models do not consider factors such as readiness and other variables that may change (e.g., Active Duty to Guard and Reserve ratios, changing readiness levels, new equipment, or concepts of operations). Thus, judgment that considers "real time" knowledge of the DoD's future plans must be used when interpreting, using, or modifying O&M estimates.

These O&M cost models are useful as starting points to evaluate official or proposed FYDP base-budgets or wartime O&M projections under various force structure and global posture scenarios. If there are significant differences between the O&M projections derived from these cost models and the DoD's or the Services' estimates, it is useful to question assumptions to gain an understanding of why the future O&M/person relationships are expected to be different from the historical O&M/person relationships.

This paper evaluates whether future DoD or Service O&M cost projections, given a specified level of active duty military end strength and global posture (as a proxy for OPTEMPO), are consistent with expectations based on history.

Baseline DoD Future End Strength Distribution

IDA | **Baseline DoD Future End Strength Distribution**

- End strength total from PB 15
- Force distribution from public sources

Derived End Strength for FY 13-19							
	Personnel Distribution						
	FY 13	FY 14	FY 15	FY 16	FY 17	FY 18	FY 19
United States	1,084,881	1,057,392	1,043,941	1,017,441	993,241	973,541	963,741
NATO, Japan, Korea	146,917	144,502	142,087	142,087	142,087	142,087	142,087
Deployed	150,547	121,406	121,172	121,172	121,172	121,172	121,172
Total	1,382,345	1,323,301	1,307,200	1,280,700	1,256,500	1,236,800	1,227,000

- Changes in end strength levels from PB 2015:
  - Army: reduction from 532,043 to 420,000 through 2019
  - Navy: reduction from 323,951 to 315,718 through 2019
  - Marine Corps: reduction from 195,657 to 174,000 through 2019
  - Air Force: reduction from 330,694 to 303,852 through 2019
- Alternative scenario shifts "deployed" end strength in excess of the minimum historical deployment level (which occurred in 1997) to the U.S. for all future years

4

Slide 4. Baseline DoD Future End Strength Distribution

Slide 4 shows DoD FY 2013–FY 2019 end strength and global posture estimates that are used to project O&M during the FYDP in this paper. Adjustments are made to DMDC total end strength data because they are not consistent with total military end strength reported in PB15. Specifically, DMDC's 2013 percentage of total end strength in the "United States," "Permanent Overseas Bases," and "Deployed" categories are applied to the historical 2013 total military end strength reported in PB15. Because DMDC has not yet reported global posture information for 2014, the ratios from 2013 were applied to the 2014 total end strength reported in PB15. The number of troops in the "United States," "Permanent Overseas Bases," and "Deployed" categories are then adjusted based on expectations reported in publicly available sources. End strength reductions projected in

PB15 and the return of forces from Afghanistan to the United States constitute most of the change to future end strength projections. These same end strength and global posture adjustments are applied to each of the four Services (data not shown). The expected Service end strength changes reported in PB15 are also shown on this slide.

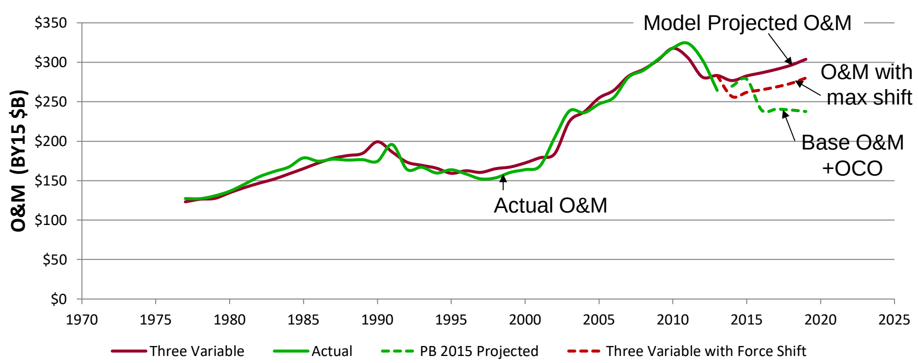
As described previously, forecasts of O&M costs throughout the FYDP are also generated using 1997 end strength and global posture (~35,000 troops deployed) actuals. This develops "lower bound" O&M cost forecasts that are consistent with recent historical relationships between O&M, end strength, and global posture (and thus OPTEMPO).

The automated O&M model enables analysts to develop models and perform their own "what if" analyses by choosing regression eras and providing end strength/global posture forecasts.

Total DoD O&M Projection Comparison

IDA | Total DoD O&M Projection Comparison (includes Defense-wide O&M)

Model ($B)	FY13	Delta	FY14	Delta	FY15	Delta	FY16	Delta	FY17	Delta	FY18	Delta	FY 19	Delta	FY14-19 Total Delta
Total O&M w/ OCO	264	(19)	270	(7)	279	(4)	239	(47)	241	(50)	240	(57)	238	(66)	(232)
Model Estimate	283		277		283		287		291		296		304		



- Model predicts a $232 B shortfall in O&M budget starting in 2014 ($426 B w/out OCO)
- Shifting deployed forces to United States reduces shortfall to $100 B (including OCO)
- More than half of the shortfall is attributable to Defense-wide O&M

5

Slide 5. Total DoD O&M Projection Comparison (Includes Defense-Wide O&M)

Slide 5 shows actual top-level DoD O&M (solid green line) costs from 1977 through 2013 in billions of BY 2015 dollars, along with the model's cost estimate (solid red line). DoD PB15 (dashed green line) and model forecasts (again displayed as a solid red line) are shown for the FYDP ending in 2019. FYDP estimates are based on expected global posture from adjusted DMDC data (described previously) and PB15 total end strength.

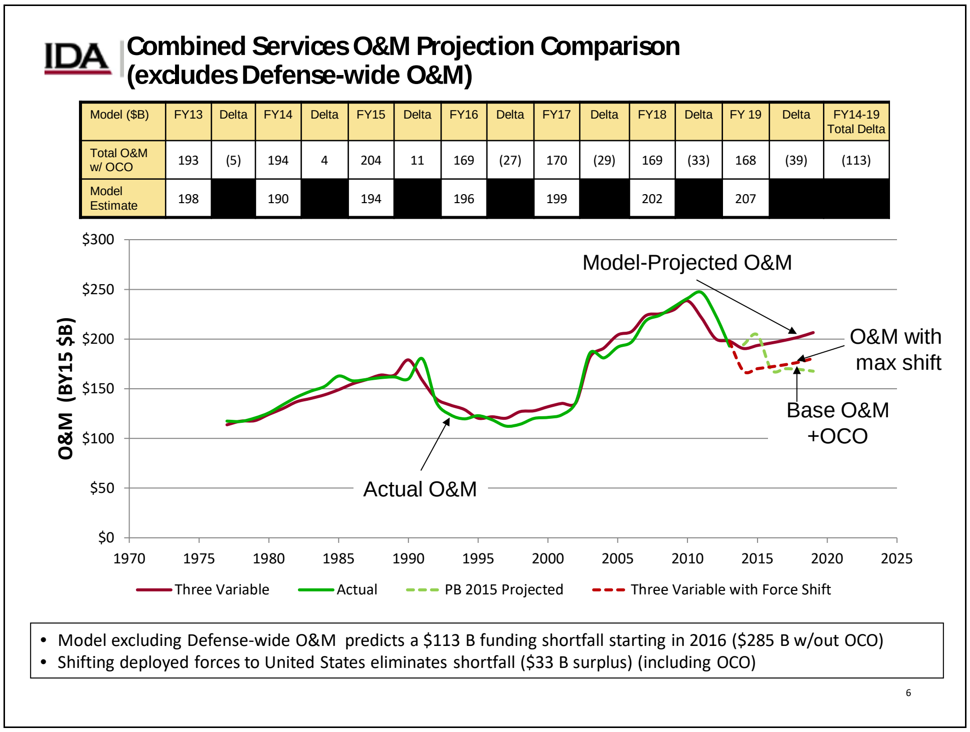
The "lower bound" O&M cost forecast (dotted red line), which is consistent with recent historical relationships between O&M, end strength, and global posture in 1997 (which had the minimum number of deployments in recent times), is also shown.

Actual DoD O&M cost and modeled O&M cost during 1977–2013 are similar; however, both the baseline O&M cost model and the "lower bound" O&M cost model are higher than PB15 O&M requests (including OCO) in the FYDP. The O&M model forecasts that O&M will cost $232 billion more than PB15 requests (with expected OCO) over the

FYDP; excluding the DoD’s expected OCO increases this difference to $426 billion. The lower-bound estimate is $100 billion more than DoD’s PB15 request (including OCO).

On Slide 5, much of the difference in O&M estimates during the FYDP is attributable to Defense-wide O&M, which is difficult to model using end strength and global posture variables. Slide 6 depicts a model that removes Defense-wide O&M and re-estimates O&M expenditures from 1977 through the FYDP.

Combined Services O&M Projection Comparison



Model ($B)	FY13	Delta	FY14	Delta	FY15	Delta	FY16	Delta	FY17	Delta	FY18	Delta	FY 19	Delta	FY14-19 Total Delta
Total O&M w/ OCO	193	(5)	194	4	204	11	169	(27)	170	(29)	169	(33)	168	(39)	(113)
Model Estimate	198		190		194		196		199		202		207		

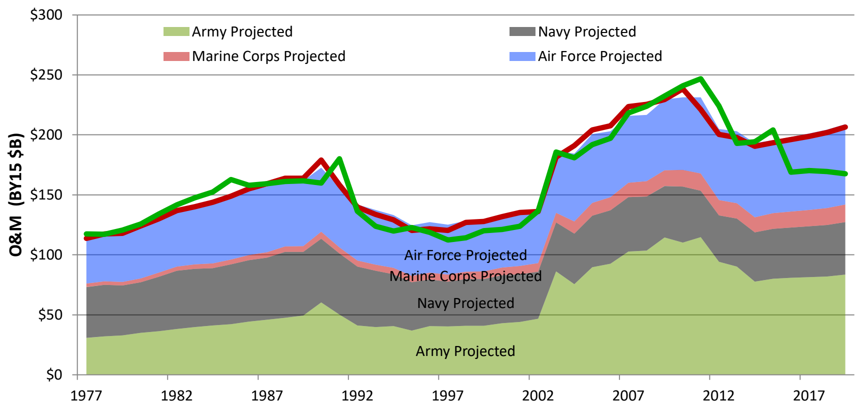
Slide 6. Combined Services O&M Projection Comparison

Slide 6 shows that historical actuals and model-calculated O&M (both excluding Defense-wide O&M) track reasonably well. In the FYDP, model forecasts of Service-only O&M are $113 billion larger than Service-only O&M in PB15 (with requested OCO added). This increases to $285 billion if OCO is removed from expected Service-only O&M. In the alternative scenario, deployments fall to a minimum historical level, and the model projects O&M costs $33 billion lower than those contained in PB15 (with requested OCO). It should be emphasized, however, that the alternative scenario is inconsistent with publicly available information on the DoD’s future end strength and global posture.

Service-Level Buildup O&M Projection Comparison

IDA | Service-Level Buildup O&M Projection Comparison

Model ($B)	Army	Delta	Navy	Delta	Marine Corps	Delta	Air Force	Delta	Defense-wide (not in chart)	Delta	FY14-19 Total Delta	Delta less Defense-wide
Total O&M w/ OCO	417	(69)	267	14	63	(19)	327	(54)	431	(198)	(325)	(128)
Model Estimate	486		253		82		381		629			



- IDA's projections for the Services aggregate to levels consistent with top-level history and projections
- Army, Air Force, and Marine Corps individually project shortfalls against O&M in PB15
- Shifting deployed forces to United States eliminates shortfall ($27 B surplus) (inc. OCO)

7

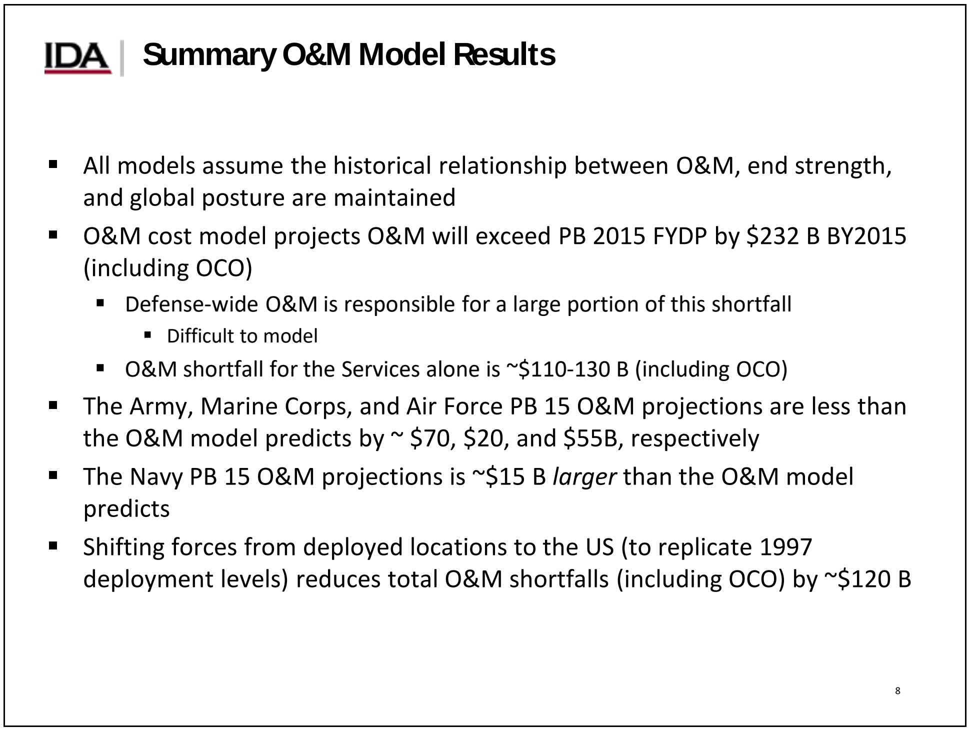
Slide 7. Service-Level Buildup O&M Projection Comparison

Slide 7 depicts O&M model estimates from each individual Service-specific O&M cost model that was generated for this paper (colored areas). These estimates were generated using historical and PB15 projected military end strength data by Service and estimates of each Service's global posture from publicly available sources, cited previously. In addition, the O&M model estimates from the combined Services model (discussed on Slide 6, and excluding Defense-wide) is shown with a red line. The solid green line represents actual combined-service O&M expenditures and requests over the FYDP from PB15.

The sum of O&M cost estimates from individual Service cost models are very close to the projection from the combined Services O&M model for the entire time period evaluated. Both estimating methodologies are larger than PB15 combined Service O&M requests by over $100 billion in the FYDP, and both modeling methodologies project surpluses of ~$30 billion in the "lower bound" scenario. Thus, the models remain consistent.

Slide 7 documents PB15 O&M requests, corresponding O&M cost model projections, and the difference between them. Three of the four Services' (Army, Marine Corps, and Air Force) PB15 O&M requests are lower than the corresponding Service-specific O&M model projections (not shown). Only the Navy PB15 O&M request is larger than the O&M model forecasts, but, in practice, this surplus would be retained by the Navy and not used to offset the shortfalls of the other Services.

Summary O&M Model Results



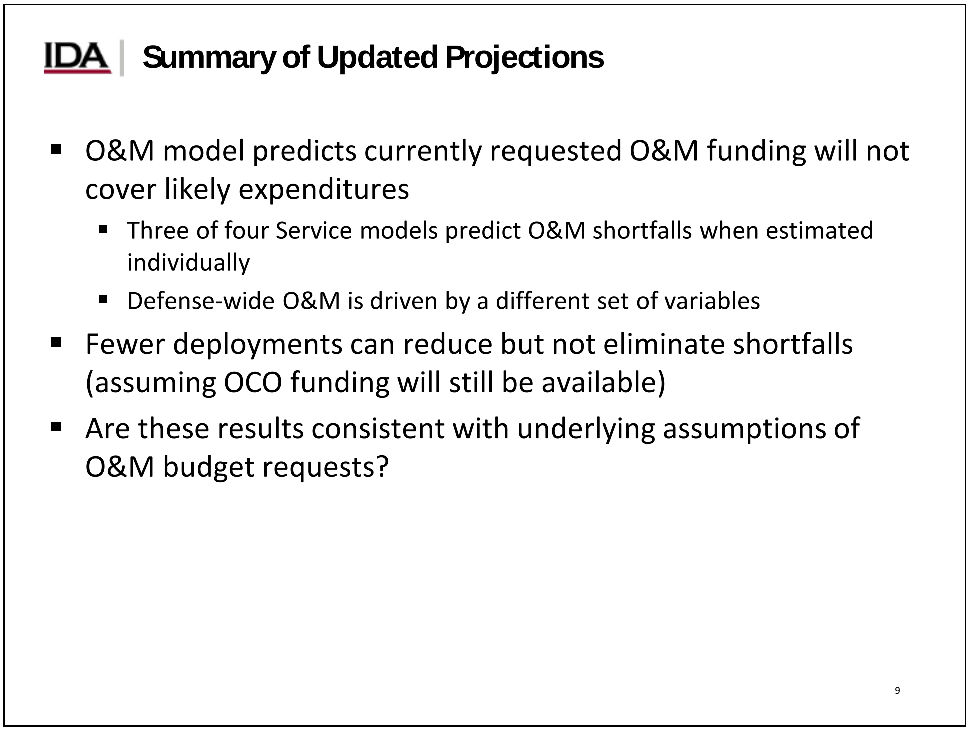
Slide 8. Summary O&M Model Results

Slide 8 summarizes this paper's O&M cost projections using PB15 data. Estimates are generated at the DoD level, the combined Service level, and for the individual Services. Except for the Navy, the O&M forecasts in this paper are significantly higher than those reported during the FYDP in PB15. The top-level DoD O&M model in this paper predicts that the DoD will spend ~$230 billion more on O&M than is in the PB15 FYDP. Much of this difference (~50%) is attributable to Defense-wide O&M, which is difficult to model. Removing Defense-wide O&M from total DoD O&M reduces the projected shortfall over the FYDP by nearly half, to $110–$130 billion.

We also generated O&M cost models for each of the Services individually. The Army, Marine Corps, and Air Force FYDP O&M levels are lower than the O&M model predictions by $70 billion, $20 billion, and $55 billion, respectively. In contrast, the Navy cost model in this paper estimates lower O&M costs than the Navy FYDP O&M level by ~$15 billion. When aggregated (with the Navy overage canceling part of the shortfalls of the other Components), combined individual O&M cost model projections for each of the four Services are consistent with a single cost model that evaluates total DoD O&M, excluding Defense-wide O&M.

What-if analyses demonstrate that bringing most of the deployed troops back to the United States (1997 deployment levels) would decrease the difference in DoD O&M at top level by more than half. Such an aggressive redeployment scenario would also eliminate shortfalls in the combined Services cost model and reduce the shortfalls in the individual Service models. These projections, of course, include projected future OCO under the current PB15 estimates, which may not materialize if overseas military commitments shrink to historically low levels.

Conclusions



Slide 9. Summary of Updated Projections

The O&M models in this paper indicate that requested O&M funding in the FYDP is less than historical relationships would suggest. This is true both at the top level of the DoD and for three of the four individual Services. These trends will likely continue even if the military reduces its deployments quickly. Only at historically minimal levels of deployments, far different from the global posture at the time of this study, do O&M levels requested by the DoD match model-projected O&M costs. We do not have a robust model to make a projection of future O&M expenditures for Defense-wide O&M.

The estimating methodology discussed in this paper could allow the OUSD(AT&L) analysts to evaluate O&M costs using just a few strategic variables. In addition, it enables “what if” analyses with respect to military end strength and global posture. O&M models and their projections can elucidate inconsistent cost-driving assumptions and help to highlight them for discussion and analysis in the budget process.

References

Defense Manpower Data Center. (n.d.). Active duty military personnel by service by region/country. Retrieved from https://www.dmdc.osd.mil/appj/dwp/reports.do?category=reports&subCat=milActDutReg

Feickert, A. (2013). Army drawdown and restructuring: Background and issues for Congress. Washington, DC: Congressional Research Service.

Greene, D., Carberry, S., & Bowman, T. (2013, April 5). Afghans want U.S. to clarify troop level post 2014. Retrieved from https://www.npr.org/2013/04/05/176300494/afghan-want-u-s-to-clarify-troop-level-post-2014

Office of the Under Secretary of Defense (Comptroller). (2014). Operation and maintenance programs (O-1) revolving and management funds (RF-1): Budget amendment to the fiscal year 2015 President's Budget request for overseas contingency operations.

Roulo, C. (2014, May 29). Hagel: Post-2014 troop level to protect Afghan progress. Retrieved from http://www.defense.gov/news/newsarticle.aspx?id=122362

Wong, K. (2013, June 25). Army lays out plans to remove 80,000 soldiers; base communities brace for impacts. Washington Times.

Acknowledgments

Thank you to the original technical reviewers, Daniel L. Cuda and R. Royce Kneece, for their review of IDA Document D-5279, O&M Cost Modeling for the Department of Defense, the source document for this paper.

Accelerating Defense Acquisition: Faster Acquisitions Produce a Stronger Force

Peter J. Modigliani—is a Senior Defense Capability Accelerator within The MITRE Corporation. He works with acquisition executives and CIOs across defense, intel, and civil agencies on enabling speed, agility, and innovation in their enterprises and major programs. [pmodigliani@mitre.org]

Daniel B. Ward—is a Principal Systems Engineer at The MITRE Corporation. He specializes in rapid innovation and is the author of two books: The Simplicity Cycle and F.I.R.E.—How Fast, Inexpensive, Restrained, and Elegant Methods Ignite Innovation. [dbward@mitre.org]

Abstract

U.S. military superiority is at risk due to rapid technological advancements changing the character of war in an increasingly complex global security environment. Department of Defense (DoD) executives fear the DoD has lost its ability to go fast. They stress the need to increase speed and agility in defense acquisition. MITRE researched organizations across the DoD, government, and industry that delivered capabilities faster than comparable organizations to understand their keys to success. Based on this initial research, strategies were identified across five major areas to enable practitioners to accelerate deliveries to the warfighters. These include creating a culture of speed, managing requirements, systems design, program execution, and contracting.

Accelerating deliveries starts with leadership creating a culture of speed, agility, and innovation to deliver capabilities to users for mission success. Effectively scoping a program and managing requirements enables rapidly delivering an initial solution and iterating vice overly defining requirements prematurely. Designing systems faster requires embracing principles of user focus, reuse, simplicity, modularity, and open systems. Executing programs faster includes tailoring and streamlining acquisition processes, documentation, and reviews with delegated decision authorities. Contracting should be part of a holistic business strategy to leverage existing vehicles and exploring the full range of FAR and non-FAR strategies available. These strategies, along with new rapid acquisition pathways enable the DoD to deliver better solutions faster.

The Need for Speed

What keeps me up at night is not North Korea,
but the fear that the U.S. has lost its ability to go fast.
— Gen John Hyten (2017), USSTRATCOM Commander

The United States has enjoyed significant technical advantages over its adversaries in most conflicts over the last 100 years. That may not be the case for future conflicts, as the 2015 novel Ghost Fleet demonstrated. In this fictional account of a near-future World War III between the United States and China, America's military superiority was eroded by an adversary able to quickly outmatch and undermine the nation's most advanced technical systems. Several of the U.S. military's most advanced weapons systems were defeated by comparable enemy systems based on designs and technologies stolen from U.S. defense companies, then fielded in a fraction of the time it took the United States.

If such a thing were to happen in real life, future military leaders in the United States would look to the current DoD research and acquisition enterprises for the source of their difficulties. As they reflect on what could have been done differently, they could reasonably point to the slow pace of acquiring and delivering military capabilities as a major contributor to America's losses. This is hardly a new hypothesis, of course. As far back as 1986's

Packard Commission report, the acquisition community has known that “an unreasonably long acquisition cycle … is a central problem from which most other acquisition problems stem” (Packard, 1986).

As the 2018 National Defense Strategy (NDS) says, the United States has entered a new era of great power competition with the rise of China and a resurgent Russia (Mattis, 2018). Further, the military technical advantage the U.S. military has long maintained over its competitors is steadily eroding as the nation’s competitors have the same access to the globalized technology marketplace driving innovation. Commercially driven breakthroughs in new technologies—artificial intelligence, advanced autonomy, robotics—are changing the very character of war. That competitors have access to these same technologies risks eroding the conventional overmatch to which the U.S. military has grown accustomed.

The NDS acknowledges that the DoD is in a race to develop and integrate cutting-edge technologies before its competitors do the same. Yet, the DoD’s bureaucratic structure, lengthy processes, and risk-averse culture inhibits timely adoption of new technologies. The reality is that competitors can iteratively field new systems in faster cycles, rapidly eroding our military, economic, and technical superiority.

There are parts of the acquisition community that move faster than others. Special Operations Command (SOCOM), for example, is well known for its ability to quickly deliver affordable, effective new weapon systems. The Navy Acquisition Executive (when with USSOCOM) James “Hondo” Geurts famously said, “Velocity is my combat advantage” (Clevenger, 2016). His use of operationally-focused language (“combat advantage”) instead of administrative language (“buying power”) is not an accident. It reflects his perspective on why the military develops technology in the first place. In a similar vein, Dr. Will Roper, the Air Force Acquisition Executive, coined “Celerity!” as a mantra to encourage the Air Force acquisition workforce to go faster (Roper, 2018).

Of course, delivering real battlefield advantage requires more than just raw speed. It also requires a nuanced capacity for agility, the ability to rapidly adapt to change—particularly when facing the emergence of new capabilities or an adversary’s new way of operating. That means the DoD requires agile systems, organizations, and strategies. As most major weapon systems are increasingly software intensive, the DoD must employ modern software development practices such as Agile DevOps. Agile typically entails small, frequent releases; valuing working software over documentation; being responsive to changes; and active user involvement throughout development (Mitre, n.d.-a). DevOps is the set of practices to integrate and automate processes between software development teams and operations to deliver software faster. Adopting Agile DevOps practices extends beyond writing software code and requires deeper changes to program structure, requirements, security, contracting, testing, systems engineering, and culture.

In researching successful organizations, programs, and initiatives across the DoD, other federal agencies, and industry, MITRE identified the following set of specific practices to enable speed and agility. MITRE is working with many federal agencies to apply these practices to accelerate their acquisition programs and enable adoption of Agile development practices. The team is relentlessly focused on shorting the time from “idea to IOC”—Initial Operational Capability.

> *Success goes to the country that … better integrates technology and adapts its way of fighting. Our response will be to prioritize speed of delivery, continuous adaptation, and frequent upgrades.*
> *—National Defense Strategy (Mattis, 2018)*

Researching Acceleration

Recognizing the urgency by defense executives, acquisition professionals, and warfighters to deliver innovative solutions faster, MITRE embarked on a research project to understand how to successfully accelerate capability deliveries. The team researched exemplar organizations across the DoD to include the Special Operations Forces Acquisition Technology and Logistics and the Air Force Rapid Capability Office to understand their keys to success.

In analyzing the schedules of major DoD programs, the team imported schedule and cost data on the DoD's major defense acquisition programs and major automated information systems into the Tableau analytics platform visualization tool. It allowed the team to identify direct correlations between acquisition costs and schedules for the Engineering and Manufacturing Development (EMD) phase.

Data was sorted and filtered by service, program category, a new start vs. modernization, and decade of program start. Tableau provided visual summaries of the data with box plots on the center 50 and 75% of timelines between Milestones A and B and B and C for each grouping of programs. It allowed the team to identify edge case programs—those who were able to deliver much faster than comparable programs as well as those that took considerably longer. This led to pursuit of initial research with individual programs to understand how their environment, constraints, and strategies impacted schedule.

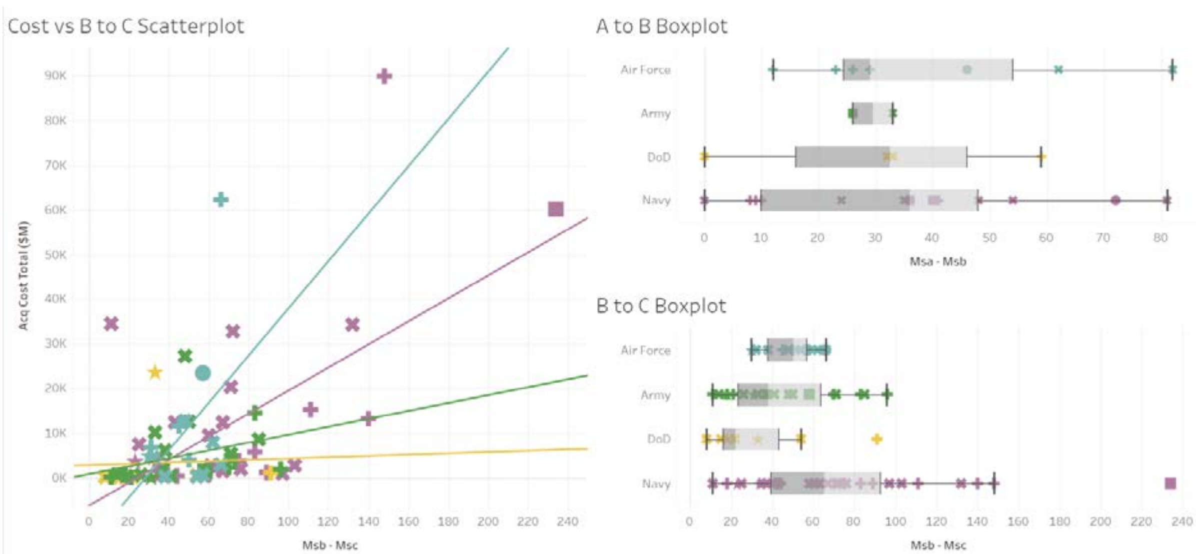


Figure 8. **Sample Tableau Visualization of Schedule Milestone Analysis**

Furthermore, the team researched commercial industry, including high-tech startups. The focus was to understand their leading strategies to rapidly exploit leading technologies for commercial solutions and how they can be applied within the defense acquisition enterprise.

The purpose of this research is to enable acquisition professionals to deliver better solutions faster. The team curated dozens of practical, field-tested strategies and tactics to apply to programs to accelerate IOC. These strategies span five initial areas: leadership and culture, requirements, system design, program execution, and contracting. Each strategy has curated content, videos and graphics, actions that programs can take, and links to dozens of references. The research is published as part of an Accelerate initiative on MITRE's Acquisition in the Digital Age (AiDA) website: https://aida.mitre.org/accelerate/.

Leadership and Culture

According to the research, culture is a key determinant of organizational performance, particularly in acceleration, agile, and innovation. Culture refers to a wide range of beliefs, behaviors, and standards that influence an organization's activities and outcomes. The norms and behaviors of a team are strongly influenced by the organizational culture in their parent organizations.

Leaders have the opportunity—and the responsibility—to influence their team's culture. One simple way to do this is to develop a strategic plan for establishing specific norms and behaviors related to agility. A leader might help foster a culture of experimentation and rapid learning by providing training and tools that support such behaviors. Leaders can further reinforce a culture of speed by delegating decision authorities to those closest to the action. Since rapid project teams often encounter resistance, ranging from passive skepticism to open opposition from key stakeholders, leaders could provide public support and recognition for acceleration to help overcome the resistance. Executives are recognizing teams and individuals that embrace agile methods, flexible contracting, and aggressively tailoring the 5000 policy.

Acceleration introduces new risks to a program while reducing others. While the net change is generally positive, leaders and staff must be mindful of the overall risk profile associated with acceleration. Ironically, being risk averse may be the biggest risk of all. Spending too much time perfecting the program analysis, documents, and briefings simply transfers the risks to the operational community. Operating at a rapid pace also often requires the team to acquire new skills. Fortunately, there are many training sources available across the DoD and industry that organizations can leverage and tailor for their environment.

Scope and Requirements

Another key determinant according to the data is effectively scoping a program, increment, or release is a critical element to being able to deliver capabilities in a timely manner. The key is to scope the work that leverages mature technologies, is affordable within the available budget, and can realistically be delivered within the needed timelines. To help meet expected delivery dates, some degree of flexibility is needed in the requirements. The operational command should convey requirements via high-level objectives. The acquiring organization can then iteratively deliver capabilities based on budgets, schedules, risks, threats, and other factors.

The DoD can accelerate delivery of innovative solutions by designing acquisition portfolios that deliver an integrated suite of smaller capabilities, rather than monolithic stand-alone systems. Operational commands should consider authoring a Capstone Portfolio Requirements document to cover a broad mission or capability area rather than that of a single program. Acquirers and developers should focus on rapidly delivering a Minimum Viable Product (MVP) to accelerate learning and rapidly iterate capabilities (Brikman, 2016). An MVP is the smallest possible product that is valuable, usable, and feasible. This replaces the DoD's traditional approach of elaborate planning, intuition, and big-bang upfront design. MVPs and iteration practices favor experimentation, customer feedback, and iterative design.

One key to iterative design is requirements that are iteratively defined. These requirements can be managed via dynamic program, release, and sprint backlogs rather than through formal requirements documents. The DoD must give up the fallacy of defining all the requirements for a system upfront. As the NDS stressed, "a rapid, iterative approach to capability development will reduce costs, technological obsolescence, and acquisition risk" (Mattis, 2018). A close partnership and active collaboration between users, acquirers, and materiel developers is critical to delivering mission impactful solutions.

System Design

Accelerating the pace of delivery is not about simply "turning the crank faster." The DoD should also take a fundamentally different approach to the way it designs systems in the first place. The discipline of Design Thinking (and its related discipline, Human Centered Design) is an important enabler of speed. It combines empathy for users, immersion in the problem, creativity in the generation of insights and solutions, and a data-based experimental approach to assess the quality of solutions. The related discipline of Systems Thinking balances holistic thinking and reductionist thinking. It enables programs to arrive at effective solutions sooner and avoid unnecessary delays and re-work.

Prototyping, experimentation, and rapid deliveries of MVPs in the early phases of the acquisition life cycle should shape requirements and system design. Agile and iterative developments value putting capabilities in the hands of users and shaping future releases based on performance and feedback. Implementing a modular open systems approach enables innovation, interoperability, and technology refresh from a variety of competing vendors. Trimming is an iterative technique for removing unnecessary elements from technical designs, system architectures, process diagrams, communications products, and organizational structures (Mitre, n.d.-b).

Documentation and Reviews

Our research found documentation historically consumed a significant amount of a program's schedule. Thus, to effectively accelerate a program, the team should constrain the amount of time spent developing, reviewing, and approving documents. In 2015, the GAO reported that acquisition programs spent over two years on average completing numerous information requirements for their most recent milestone decision, yet acquisition officials considered only about half of the requirements as high value (Sullivan, 2015).

One example of a sound approach to documentation comes from the Agile Manifesto (2001). The Agile software approach emphasizes working software over comprehensive documentation and offers this perspective: "Simplicity—the art of maximizing the amount of work not done—is essential." While writing documentation is important, not writing documentation is also important. Aim to only produce the documents that are useful and needed to manage the program, rather than writing "compliance only" documents which exist only to satisfy the interests of headquarters staffs.

Similarly, programs should apply the concept of Minimum Effective Dose to their documentation. This concept comes from the medical community, where doctors and nurses recommend patients take the least amount of medicine that delivers the desired effect. Acquisition programs should adopt a similar Minimum Effective Documentation strategy, aiming to produce as little as possible, as much as necessary.

This involves MDAs and functional leaders clearly identifying what information is required and developing the minimum set of documents that can capture the required information. While a functional oversight organization may expect a functional document, a program office may merge the content of that document with others to minimize the number of documents to coordinate. Communicating the intent of this tailored approach in advance helps increase the buy-in from reviewers and other stakeholders.

Streamlined documentation coordination and approvals are equally critical to accelerate schedules. Exemplar organizations identify upfront the minimum set of officials to coordinate and approve each program document. Many have leveraged IT tools and business rules (e.g., no response within 10 days signifies concurrence) to streamline coordination across multiple organizations. Many milestone decision authorities have delegated approval of various program documents to lower level officials.

Program reviews should be focused on the highest risks, open issues, and provide the oversight officials with the key information to decide if the program is ready to proceed. Weeks and months can be lost with pre-briefs and rework to debate elements of the program's strategy and refine the messaging for leadership. Like documentation, reviews should be kept to an absolute minimum. Program reviews should provide the highlights of the strategy, with the details in the program documentation. Successful MDAs and functional leaders set clear on the key information required for each review and ensure subordinate reviews are minimized.

Contracting

Finally, data shows contracting is often one of the longest lead-items in the acquisition life cycle, and one of the riskiest. Traditional contracting methods can take 18 months to three years to compete and award a contract. This increases the risk of the program delivering products that are operationally irrelevant, technologically obsolete, or both.

Successful acquisition organizations approach contracting as a holistic business strategy where program managers partner with their contracting officers early to develop and shape the strategies. They work together to achieve the mission objectives within the environmental constraints. Far too many acquisition organizations separate contracting from the program offices to "process the paperwork," which leads to lengthy timelines and poor contract strategies. They also regularly collaborate with industry to communicate the government's needs, approach, and timelines and to solicit feedback on issues with their strategy.

Leveraging existing contracts to award a task or delivery order saves significant time over developing and awarding a new contract. Programs should first look to the array of existing contracts to see if the scope of work and pool of vendors meet their needs. The Under Secretary of Defense for Acquisition and Sustainment's (USD[A&S]) Contracting Cone outlines 23 different Federal Acquisition Regulation (FAR) and Non-FAR contract strategies (USD[A&S], 2018a). The online tool offers insights on common applications, pros/cons for use, restrictions, and references.

Program Executive Officers should establish multiple-award contracts to cover a broad portfolio area. If established correctly, these contracts have aggressively streamlined processes with standardized language, terms, and metrics to enable rapid orders. Similarly, a portfolio can establish their own Other Transaction Authority (OTA) Consortium to tap a pool of non-traditional vendors focused on their portfolio capabilities (USD[A&S], 2018b). These portfolio vehicles enable each program and project to aggressively cut contracting timelines.

There is a wide array of FAR and non-FAR contracting strategies available to the acquisition community with ample flexibilities on their use. Instead of a traditional, lengthy FAR Part 15 approach, many use OTAs, Broad Agency Announcements, Federal Supply Schedules, Simplified Acquisition, and Commercial Items to reach contractors in a fraction of the time. The FAR explicitly encourages speed, agility, and innovation, yet many interpretations assume that a lengthy approach is safer. The FAR also stresses using modular contracting to the maximum extent practicable, by dividing large efforts into a series of smaller efforts.

Contracting officers are the key linchpins to a successful government–contractor partnership, which is critical to success. They can identify the key levers (e.g., progress payments and bonuses for cash flow) to incentivize contractors for speed to delivery. In addition to contracting, many rapid organizations have empowered, experienced, forward-leaning professionals from other functional areas to include legal, test and evaluation, and finance.

Summary

The current operational environment demands acquisition professionals accelerate their capability deliveries. The culture has begun to shift over the last two years from controlling costs to accelerating schedules. There are proven strategies and tactics throughout the acquisition life cycle to lean the acquisition and requirements processes to achieve IOC sooner. The current leadership in the Pentagon are strong champions of speed and agility. Congress has also been a strong proponent of speed, offering a series of new authorities and flexibilities to go faster to include the popular Middle Tier Acquisition (Mitre, n.d.-c). There are additional opportunities to accelerate other major schedule drivers across the acquisition life cycle to include test and evaluation. The time is ripe for acquisition professionals to lean forward and accelerate deliveries of innovative solutions. See more at https://aida.mitre.org/accelerate/.

Reference List

Agile Manifesto. (2001). Retrieved from http://agilemanifesto.org/

Brikman, J. (2016). A minimum viable product is not a product, it's a process. Retrieved from https://blog.ycombinator.com/minimum-viable-product-process/

Clevenger, A. (2016). SOCOM acquisition chief embraces change, adaptability. Defense News. Retrieved from https://www.defensenews.com/industry/2016/01/20/socom-acquisition-chief-embraces-change-adaptability/

Hyten, J. (2017). Remarks from Air Force Association 2017 Symposium. National Harbor, MD. Retrieved from https://www.youtube.com/watch?v=3m691gCJWME&feature=youtu.be&t=25m47s

Mitre. (n.d.-a). MITRE's acquisition in the Digital Age (AiDA) agile development practices. Retrieved from https://aida.mitre.org/models/agile-acquisition/agile-overview/

Mitre. (n.d.-b). MITRE's AiDA just keep trimming. Retrieved from https://aida.mitre.org/blog/2019/03/26/just-keep-trimming/

Mitre. (n.d.-c). Middle tier of acquisition. Retrieved from https://aida.mitre.org/middle-tier/

Mattis, J. (2018). National Defense Strategy. Retrieved from https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf

Packard, D. (1986). President's Blue-Ribbon Commission on Defense Management Report. Retrieved from https://usacac.army.mil/cac2/CSI/docs/Gorman/06_Retired/01_Retired_1985_90/07_86_PackardCommission_FinalReport/01_PackardCommission_FinalReport.pdf

Roper, W. (2018). Seven steps for incorporating rapid prototyping into acquisition memo. Retrieved from http://acqnotes.com/wp-content/uploads/2018/05/Air-Force-Incorporating-Rapid-Prototyping-into-Acquisition.pdf

Sullivan, M. (2015). DoD should streamline its decision-making process for weapon systems to reduce inefficiencies (GAO-15-192). Retrieved from https://www.gao.gov/assets/670/668629.pdf

USD(A&S). (2018a). Contracting cone. Retrieved from https://aaf.dau.mil/contracting-cone/

USD(A&S). (2018b). Other transaction guide. Retrieved from https://aaf.dau.mil/ot-guide/

Networked Logistics and Additive Manufacturing

Susan M. Sanchez—PhD, is a Professor at the Naval Postgraduate School in the Department of Operations Research, with a joint appointment in the Graduate School of Business & Public Policy. She is Co-Director of NPS's SEED Center for Data Farming and specializes in advancing the theory and application of simulation experiments and efficient designs to address complex problems in defense and national security. Her awards include recognition as a Titan of Simulation and an INFORMS Fellow. [ssanchez@nps.edu]

Gregory E. Lynch—is a Major in the U.S. Marine Corps and a graduate student in the Department of Operations Research at the Naval Postgraduate School. He is a logistician and has served in Iraq, Afghanistan, South America, and various locations throughout the Pacific Region. [gelynch@nps.edu]

Claudia Luhrs—PhD, is a Professor at the Naval Postgraduate School in the Department of Mechanical and Aerospace Engineering. Her research interests focus on functional materials. Recent projects include the use of additive manufacturing techniques. Her work aims to understand the correlations between structural features and the material's reactivity, thermal, electrical and mechanical properties. Application projects with naval relevance include the generation of materials for sensors, batteries, supercapacitors, and structural components. She received the NPS Carl E. and Jessie W. Menneken Faculty Award for Excellence in Scientific Research in 2014 and was a Fulbright awardee in 2016. [ccluhrs@nps.edu]

Mary L. McDonald—is a Faculty Associate–Research in the SEED Center for Data Farming at the Naval Postgraduate School. She supports a wide range of applied projects whose overall goal is to advance the collaborative development and use of simulation experiments and efficient designs of experiment in order to provide decision-makers with timely insights on complex systems and operations. [mlmcdona@nps.edu]

Abstract

Additive manufacturing has the potential to fundamentally change how military expeditionary operations are conducted. By manufacturing spare parts in remote sites, rather than relying on lengthy and extensive supply chains or remaining tethered to an "iron mountain" of logistics support, the expeditionary units have the potential to be more agile, to maintain their readiness at high levels while deployed, and to extend their operational reach. We describe how the additive manufacturing capability can be combined with a networked logistics approach for the U.S. Marine Corps. The ultimate goal is to develop a logistics support structure that is more resilient to enemy attacks and provides improved support to the forward units.

Introduction

Additive manufacturing (AM) has enjoyed success in a number of specialty fields. Potential benefits for expeditionary units include achieving higher readiness at lower cost, because deployed units can use AM to create replacement parts at or near the point of demand, rather than either relying on carrying large quantities of spare parts or dealing with long lead-times for replacements. Another potential benefit is the ability to reduce wastage of the materials used in the three-dimensional (3D) printing process and subsequent post-treatments by only producing what is needed. Finally, if the same compounds can be used to manufacture a variety of parts, AM could help forward-deployed units maintain a high level of readiness while dramatically reducing their logistics footprint.

To realize this potential, program managers have several decisions to make. They must determine how best to acquire AM capabilities, what classes of components are suitable for AM, whether the resulting structural stability and reliability are comparable for components made using AM and current methods, and how differences in reliability may affect the supply chain and readiness levels. If the suitability and reliability are not factored into the decision-making process, then AM may end up being a costly and largely redundant logistics system running in parallel with the current supply chain, rather than being a transformative capability.

AM is integrally tied with Department of Defense (DoD) acquisition programs in several ways. First, the capability for AM must be procured. Rather than setting specifications and requirements for parts or component parts, the program managers must set requirements for AM processes that are capable of 3D-printing and subsequent finishing operations in order to produce items that meet the necessary specifications for the parts or component parts. Second, although the flexibility of AM is often touted, the issues of the quality and reliability of the resulting parts are not generally considered—or, these characteristics are considered in isolation, rather than via their effects on the supply chain and operational effectiveness. An implicit assumption in much of the literature is that the resultant parts will be as capable when produced using AM as they are when produced using standard manufacturing techniques. Third, the current roadmap for employing AM in DoD operations is incremental in nature. For example, the U.S. Army's phases for AM are (1) determining how AM can be used to repair or replace existing parts, (2) using AM to produce a single part rather than assembling multiple component parts, and (3) using AM to create parts that do not currently exist (U.S. Army, 2017). The Marine Corps and Navy have similar guidance (Department of the Navy, 2017; Department of the Navy, Headquarters United States Marine Corps, 2017).

In this paper, we present a model-based framework for a transformative rather than an incremental approach for incorporating AM technologies within the DoD. We do so by creating a simulation model of networked expeditionary logistics operations—a concept of operations that now may be possible. Because stockpiles of spare parts are no longer the only way of ensuring that the combat logistics element is fully supporting the expeditionary units, we can explore the simulation model's behavior to gain insight about other alternatives.

Background

We begin with a short overview of several key areas that motivate this research. Our discussion is meant to be illustrative, not exhaustive.

Additive Manufacturing: Previous Research Themes

There has been a rapid escalation of additive manufacturing research and applications in recent years. It has already demonstrated success in specific industries, where computer-controlled 3D printing using a variety of compounds has opened up new customization possibilities for manufactured parts. For example, the medical field has enjoyed success in customizing polymeric parts, such as right-sizing cardiovascular stents rather than relying on a limited number of sizes. Custom-sized biodegradable stents reduce the risks of complications that arise if an ill-fitting stent moves and ultimately fails, and additional surgery is required to repair or replace the stent (Hodsden, 2016); they can be quite beneficial for infants and children who need temporary assistance while they are growing (Fessenden, 2013). Other successful applications have been reported in areas ranging from sports equipment (Graziosi et al., 2017) to spare parts for air-cooling ducts of the environmental control system for F-18 fighter jets (Khajavi et al., 2013) to 3D-printed jet engines (Sturmer, 2015).

Previous research related to AM falls into a few general categories. The first is research related to the AM process itself, including the polymeric, metal alloy, or composite materials used in the 3D printing part of the process, along with the post-treatment operations required for the materials to attain their structural capabilities (Frazier, 2014). Post-treatment operations, once the printing process is complete, can include various types of heat treatment to reduce porosity or induce the desired microstructures and properties such as annealing or hot isostatic pressing.

A second stream of research involves studying the logistics supply chain, contrasting AM versus traditional manufacturing for producing spare parts. This has been accomplished in different ways. Case study approaches have been used as part of an inductive research approach, such as the work by Oettmeier (2016), who conducts and describes semi-structured interviews for three focal firms, suppliers, and customers for AM devices in the medical industry. Oettmeier concludes that the effects of AM technology adoption on the supply chain configuration are context-specific and depend on a number of exogenous and supply chain-related factors. Mellor et al. (2014) also use a qualitative case study approach to create a normative structural model of AM implementation, including factors related to the technology and supply chain, as well as other structural and strategic aspects of the organization. See, for example, Silva and Renzende (2013) for further discussion of logistics implications of additive manufacturing for a number of different industries.

Other research examines the life-cycle cost of AM relative to traditional manufacturing techniques. For example, Westerweel et al. (2018) develop an analytic cost model and conduct a full factorial experiment involving seven factors, each at three levels, to gain some managerial insights. They conclude that logistics savings can occur because of the reduced production lead time inherent in AM. They also find that large investments in AM are attractive if there are large numbers of systems with long life cycles, and if the reliability of the AM parts is quite close to that of the parts produced by the original equipment manufacturer (OEM). Still others have looked at the supply chain for the powders used in AM applications, rather than focusing on the supply chain associated with OEM parts (Dawes et al., 2015).

Additive Manufacturing for Expeditionary Operations

With regard to military operations, U.S. Army Chief of Staff General M. A. Milley stated, "The convergence of new developments such as ubiquitous information technology and personal communications, proliferation of precision guided weapons, robotics and on-site 3D printing, and rapidly growing urbanization all augur a very different era of warfare" (Barno & Bensahel, 2017). AM may be beneficial for legacy systems as well, if the original parts are no longer being manufactured but custom AM parts can be made as needed.

An example of a simulation-based assessment of AM for military operations appears in Moore et al. (2018). They create forecasts of replacement parts for the M109A6 Paladin self-propelled 155mm Howitzer, based on data obtained from the U.S. Army during the initial stage of Operation Iraqi Freedom (OIF). They use these data-driven forecasts as inputs to a simulation model to assess the feasibility of integrating AM into the Army's supply chain for 48 different combinations of the three factors: the echelon at which the AM is placed, the printing speed, and the available volume of metallic compounds for printing the metal parts. They recommend that "the Army needs to continue experimenting with AM facilities in the field under realistic demand rates and operating environments," and also suggest that AM should most likely start with small items where quality control requirements are not so onerous. Other nations are also intrigued by the prospect of incorporating AM into military logistics support (Ng, 2018).

Some AM approaches are more suitable for harsh and variable environments than others, in part due to their safety requirements. Zelinski (2019) describes how metallic 3D printing that involves arc-welding metallic compounds deposited by solid wire feed, or by high-velocity cold spray of metal powder, can be relatively safe. In contrast, the safety requirements for setting up and using laser melting systems may prohibit those forms of AM in some operational environments.

Logistics for Expeditionary Operations: Current System

The resources contained in these stockpiles are critical to the survival of a military force and directly contribute to their mission success. An adversary capable of destroying these stockpiles or significantly deteriorating the supply distribution process can seriously disrupt or even halt military operations.

A graphical representation of an iron mountain logistics approach appears in Figure 1 based on a scenario from Lynch (2019), who considers expeditionary operations at the Marine Expeditionary Unit (MEU) level. In this graphic (not to scale), we show how the current system often works. The seabase is treated as an essentially unlimited floating warehouse of supplies and fuel. The ultimate goal is meeting the logistics needs of the supported units—in this scenario, two infantry units and a Forward Arming and Refueling Point (FARP). Each infantry unit represents a standard Marine Corps Infantry Rifle Company. FARPs do not have a standard size, but the FARP in this scenario is roughly equivalent to the size of an infantry platoon. Most supplies are moved by ship-to-shore connectors from the seabase to a fixed, fortified, onshore position—the so-called “iron mountain,” although jet fuel is typically delivered to the FARP by air assets. The supported units each generate requests for several different types of supply items. Some supplies—such as Meals Ready to Eat (MREs), bottled water, and general fuel consumption—are used at rates proportional to the number of personnel in the unit. Ammunition and missile usage are less predictable and depend on the operational tempo. The convoys tend to make regular deliveries over long distances and are comprised of many logistics vehicles (LVs) as well as security vehicles for added protection. The black boxes notionally represent the amounts of supplies at various points in the system. For example, the seabase is typically assumed to have (or have access to) unlimited inventory; the iron mountain has a very large supply on hand; the convoys carry large amounts in each delivery; and the supported units must keep enough on hand for sustainment between convoy arrivals. Of course, this does not capture the full complexity of logistics support in real-world military operations.

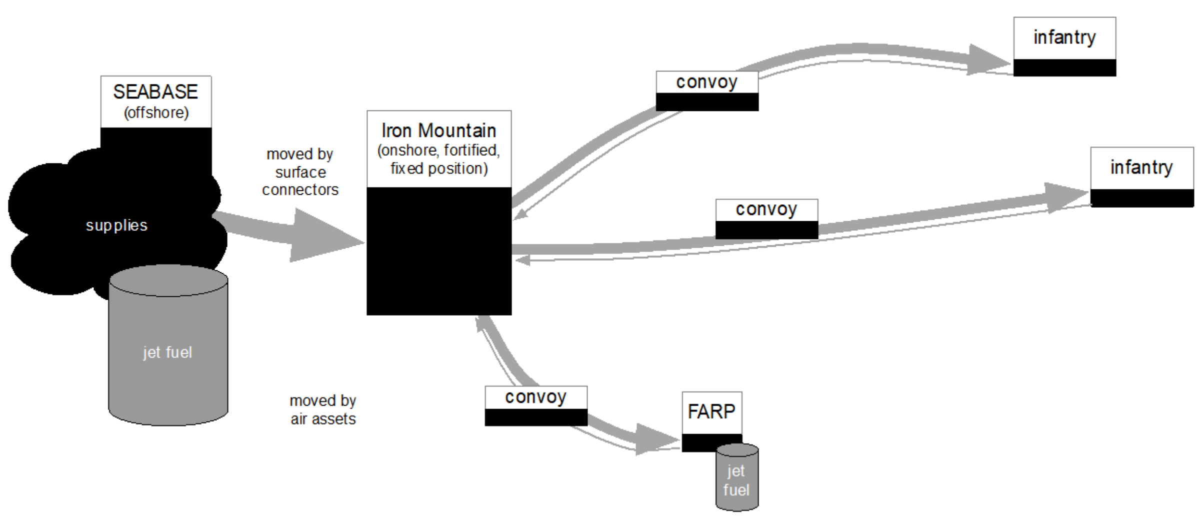


Figure 1. Current Logistics Supply Movement in Expeditionary Operations

As mentioned earlier, plans for using AM in expeditionary operations in the near future focus on its potential for repairing or replacing existing parts. By injecting this capability into an existing logistics chain, the primary benefits are those of reducing the storage capacity and lead time for replacement parts without adversely affecting readiness. Parts that can be easily manufactured using AM technologies can range from those used steadily throughout the operations, to rarely-needed replacement parts for mission-critical items. In either situation, it may require less storage volume to ship bulk raw materials (e.g., metallic powders) and manufacture the parts as needed, than to store and access completed parts. Easy access can be particularly problematic in very high-density storage systems such as the hold of a ship (Gue, 2006), or in limited staging areas where containers transferred from ships may await other transportation to their final destinations (Gue & Kang, 2001).

For the logistics system of Figure 1, there are three places in which adding AM capability might be beneficial: the seabase, the iron mountain, and the supported units themselves. Each has benefits and drawbacks. The seabase is often considered the most secure, and on larger ships, it may be possible to set up dedicated AM facilities (including appropriate post-treatment stations) with access to a ready supply of bulk raw materials. Lead-times for 3D printing replacement parts at the seabase may be less than lead-times for receiving them from the United States or other regional supplier. The iron mountain has similar capabilities, although there may be less control of some environmental characteristics (heat, humidity, dust, vibration) that might affect the AM production schedule or the resulting quality and reliability of the parts. Lead times for 3D-printed parts from the iron mountain might be less than those of 3D-printed parts from the seabase, particularly if small numbers of items are needed. Adding AM capability directly to the supported units has both potential benefits and potential drawbacks. On one hand, it may reduce the lead time for replacement parts even further. On the other hand, it may be the most likely to be adversely impacted by weather conditions, and long post-treatment requirements may either reduce the unit’s mobility or result in less reliable replacement parts.

Networked Logistics for Expeditionary Operations

Headquarters U.S. Marine Corps recently published the Marine Corps Operating Concept (MOC), which states the Marine Corps must “[redesign] our logistics to support distributable forces across a dynamic and fully contested battlespace—because iron mountains of supply and lakes of liquid fuel are liabilities and not supportive of maneuver warfare” (Department of the Navy, 2016, p. 9).

During the wars in Iraq and Afghanistan, the insurgent forces were incapable of conducting an attack on the scale required to destroy a large base containing massive quantities of supplies. Attacks such as those on Camp Bastion and Camp Shorabak caused damage and some casualties (to Afghan troops) but did not pose a serious threat for the viability of the entire bases and their operations (Shah et al., 2019; Snow, 2019). As the United States has transitioned to preparing for a conflict with a near-peer adversary, this is no longer true: an iron mountain is a very enticing target.

Even in situations where enemy actions are not a concern, iron mountains can still be liabilities. The 2010 fire in the Supply Management Unit lot in Camp Leatherneck, Afghanistan, is one such example: Although the fire was eventually contained with no casualties, most of the inventory was destroyed—including construction materials, medical supplies, and repair parts (Pelczar, 2010). In this way, a networked logistics structure may add resilience.

Consequently, a new method of providing logistics support to expeditionary forces is needed. Lynch (2019) creates a simulation model intended to help analysts explore the function of a networked logistics force. A simplified graphical representation appears in Figure 2. Instead of consolidating and distributing supplies from a large, stationary iron mountain, the supplies are redistributed to smaller logistics support nodes that occasionally move around the battlefield. There are three types of units that require support from these logistics nodes: infantry units and a Forward Arming and Refueling Point (FARP) as in Figure 1, as well as a shore-based missile. The shore-based missile unit is based on a platoon from the Army High Mobility Artillery Rock System (HIMARS) Battalion, because providing shore-based missile support is “a relatively new concept for the Marine Corps” (Lynch, 2019).

The supported units each generate requests for several different types of supply items. Some supplies—such as Meals Ready to Eat (MREs), bottled water, and general fuel consumption—are used at rates proportional to the number of personnel in the unit. Ammunition and missile requests are randomly generated, providing an implicit rather than explicit representation of their use during combat operations. The supplies are loaded on logistics vehicles (LVs) for delivery. For the current model instantiation, each LV can be considered a truck that carries supplies on pallets.

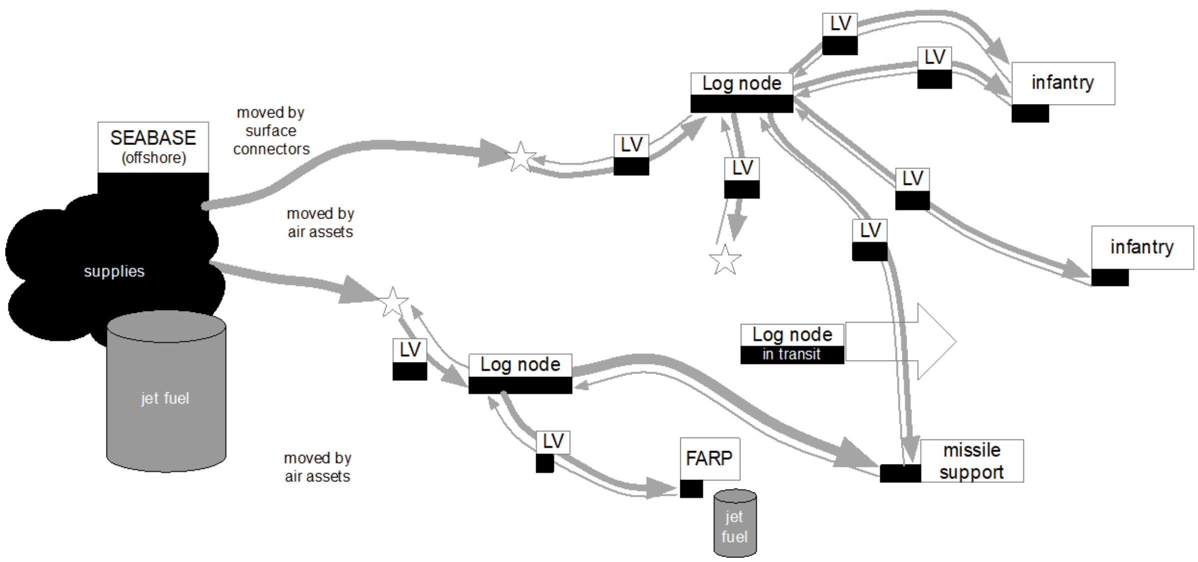


Figure 2. Logistics Supply Movement in Expeditionary Operations

The networked logistics of Figure 2 is clearly more complex than that of Figure 1, as can be seen by the larger number of potential routes taken by LVs. The networked logistics structure is highly dynamic, as the supported units, logistics nodes, and rendezvous points can all change over time. Ideally, this dynamic structure will enhance the maneuverability of the forward units and make the force less vulnerable to attacks by a near-peer adversary. There are other differences as well. The black boxes that represent the amounts of supplies at various locations in the network tend to be much smaller than those in Figure 1. This has the potential to increase the agility and extend the operational reach of the supported units. However, care must be taken to ensure they are mission-capable despite their decreased logistical footprint.

Table 1 provides a brief comparison of the iron mountain and networked expeditionary logistics structures.

Table 1. Differences Between Current and Networked Expeditionary Logistics Systems

Element	Assumptions: Iron Mountain Logistics	Assumptions: Iron Network Logistics
Seabase	Offshore, invulnerable, infinite capacity	Offshore, invulnerable, infinite capacity
Supported units	Two infantry, one FARP	Two infantry, one FARP, one shore-based missile support
Onshore logistics element	Iron mountain: immobile, heavily fortified, very large capacity, regularly resupplied from seabase	Logistics nodes: mobile, self-sufficient, use their own logistics vehicles (LVs) to change logistics node locations in a single trip, resupply from fixed or ad hoc rendezvous points
Seabase -> onshore	Large deliveries to fixed location at fairly regular intervals	Smaller deliveries to LVs at both fixed and ad hoc rendezvous points
Convoys	Large, heavily armed, long and regular trips to supported units	Single LVs travel faster, make frequent short trips to supported units, less predictable transit routes

There are several places where AM capabilities might be added to the iron network: at the seabase, at the logistics nodes, or at the supported units. The same basic benefits and drawbacks apply. Stationary units (i.e., the seabase) can be heavily protected, and AM may reduce storage and lead time requirements. Still, the mobility of the logistics nodes and the use of single LVs rather than large convoys may affect the way AM is implemented, or vice versa. If AM is added at logistics nodes, then those nodes must wait to change locations until all post-printing treatments are complete. Both the duration of AM operations and timing of logistics node moves are decisions that must be made.

Lynch (2019) implements the networked logistics simulation model using the Ruby programming language. Each run of the simulation represents 180 days of operation, beginning from a time where all supplies have arrived at the seabase and each logistics node and supported unit has the supplies it needs to begin operations. Logistics nodes and supported units all use supplies over time; the logistics nodes are handled as internal requests that require no transportation as long as the node has the requested supplies on hand. Each logistics node will move its location after filling a specified number of requests. The LVs begin moving to a requesting unit as soon as either the LV is nearly full (e.g., seven or more of eight pallet spaces filled), or the request has been waiting a sufficiently long time. Logistics nodes place requests for resupply to the seabase whenever their inventories drop below specified levels but can also receive direct shipments by air if needed.

A few other modeling choices deserve mention. LVs can encounter breakdowns or enemy attacks at random times during transit. If an LV suffers a maintenance breakdown, that delays the delivery process by a relatively short amount of time (hours to days). If an enemy attack occurs, there is some probability that the LV wards it off successfully and then continues on after a short delay; there is also some probability that the attack is successful and the LV and its inventory are all destroyed. In the latter situation, new requests are automatically generated for all destroyed items.

Another key assumption is that inventory levels are visible to all players in the simulation. This is essential because in the situation in which one logistics element cannot provide support requested by a unit, it must then pass that request to another logistics node. Trust in the logistics structure is also critical in practice (Spangenberg, 2017). Without that trust, each unit has incentives to hoard items or make larger requests than necessary, which may keep the logistics footprint large or reduce the agility of the force.

For further details of this networked logistics simulation model, see Lynch (2019).

Research Methodology: Data Farming

Headquarters U.S. Marine Corps recently published the Marine Corps Operating Concept (MOC), which states the Marine Corps must "[redesign] our logistics to support distributable forces across a dynamic and fully contested battlespace—because iron mountains of supply and lakes of liquid fu

Data farming and data mining are different! Lucas et al. (2015) compare and contrast these metaphors as follows:

> Miners seek valuable nuggets of ore buried in the earth, but have no control over what is out there or how hard it is to extract the nuggets from their surroundings. As they take samples from the earth they gather more information about the underlying geology. Similarly, data miners seek to uncover valuable nuggets of information buried within massive amounts of data. Data-mining techniques use statistical and graphical measures to try to identify interesting correlations or clusters in the data set.
>
> Farmers cultivate the land to maximize their yield. They manipulate the environment to their advantage using irrigation, pest control, crop rotation, fertilizer, and more. Small-scale designed experiments let them determine whether these treatments are effective. Similarly, data farmers manipulate simulation models to their advantage, using large-scale designed experimentation to grow data from their models in a manner that easily lets them extract useful information. … [The output data sets] also contain better data, in the sense that the results can reveal root cause-and-effect relationships between the model input factors and the model responses, in addition to rich graphical and statistical views of these relationships. (p. 297)

The building blocks of data farming are a collaborative approach to rapid scenario prototyping, modeling platform development, design of experiments, high performance computing, and the analysis and visualization of the output—all with the intent of providing decision-makers with timely insights (NATO, 2014). Of these, design of experiments is key: it is the only way to break the so-called "curse of dimensionality." For example, suppose our simulation has 100 inputs (i.e., factors), each factor has two levels (low and high) of interest, and we decide to look at all combinations. A single replication of this experiment would require over 178 millennia on the world's fastest supercomputer (the Summit at Oakridge National Laboratories), even if each of the 2100 (roughly 1030) simulation runs consisted of a single machine instruction! Yet efficient experiment designs enable us to run interesting simulation models with dozens or hundreds of factors on a modern laptop or small computing cluster in a matter of days to hours, taking the study from the realm of the impossible to the realm of the practical.

A data farming approach is useful for the networked logistics study because the simulation model has a large number of potential factors, and the ways in which they affect the system performance are complicated and not (yet) well understood. Running a designed experiment and analyzing the results (both statistically and graphically) provides a quantitative basis for trade-off analysis.

Preliminary Results

We now present some preliminary results from an initial experiment. These are intended to be illustrative of the types of analytic products and insights that can result; more detailed explorations and analyses are needed before developing actionable recommendations. Table 2 lists the factors, their descriptions, and the low and high values over which we vary their levels. In all, we vary 13 factors for a total of 1,025 factor combinations, called design points (DPs). Our design is based on a nearly-orthogonal Latin hypercube with 65 design points (Cioppa & Lucas, 2007), so each factor can be explored at up to 65 different levels. For comparison purposes, a brute-force approach to studying 13 factors at 65 levels each would require over 3.6 septillion design points! Because our model is stochastic, we replicate each design point 20 times to reveal the variability in the system. It took roughly 8 hours to complete the 20,500 runs (simulated 180-day operations) on a single laptop.

Table 2: Factors and Factor Ranges for Initial Experiment

Factor	Description	Low level	High level
external resupply time	Wait time for logistics node resupply (days)	2	10
max wait time	Maximum time logistics vehicles wait before departing (days)	0.5	3.0
number of LV	Number of vehicles per logistics node	8	20
log node min	Triangular distribution minimum value	0.5	1.5
log node max	Triangular distribution maximum value	2.5	3.5
log node mode	Triangular distribution mode	1.5	2.5
onload mean	Mean time (days) to load vehicle (gamma distribution)	0.25	0.65
onload shape	Shape parameter for loading vehicle (gamma distribution)	8	12
offload mean	Mean time (days) to unload vehicle (gamma distribution)	0.1	0.5
offload shape	Shape parameter for unloading vehicle (gamma distribution)	8	12
enemy attack	Probability of an enemy attack	0.01	0.1
enemy kill	Probability of an attack resulting in destruction of the logistics vehicle	0.01	0.03
maintenance	Probability of an unscheduled maintenance issue	0.5	0.25

There are many potential measures of performance (MOPs) or measures of effectiveness (MOEs) that can be examined. For illustrative purposes, we focus on a single one: the average number of requests in queue awaiting processing. This time-weighted average is calculated for every simulation run. We then average the results over the 20 replications for each design point to obtain 1,025 values of MOE = Mean(avg requests in queue).

Figures 3 and 4 show the results of applying stepwise regression to fit a response surface metamodel for this MOE. The stepwise algorithm used considered main effects, quadratic terms, and two-way interactions for inclusion in the final model. Figure 3 indicates that the regression model fits reasonably and explains 90% of the variability in the data. Figure 4 contains a numeric and visual display of the model's terms, all statistically significant (p < 0.0001). The horizontal bars indicate the direction and magnitude of each term. The "tornado" visual results from having sorted the terms from greatest impact to least impact. We see that the top two influential factors are the number of LVs and external resupply time—both of which are decision factors over which the Marine Corps can exercise control.

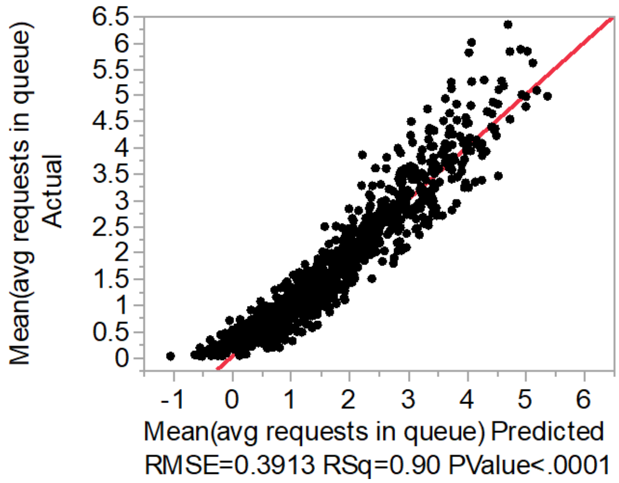


Figure 3: Regression Summary Results

Term	Estimate	Std Error	t Ratio
number of LV	-0.245	0.0034	-71.30
external resupply time	0.209	0.0051	40.97
enemy attack	9.4267	0.4581	20.58
onload mean	1.8833	0.1037	18.17
log node min	-0.709	0.0409	-17.35
enemy kill	28.717	1.7055	16.84
(external resupply time-6.06244)*(number of LV-14.0312)	-0.023	0.0015	-16.02
log node mode	-0.652	0.0409	-15.96
log node max	-0.63	0.0409	-15.39
(number of LV-14.0312)*(number of LV-14.0312)	0.0142	0.0011	13.04
offload mean	0.9917	0.0979	10.13
(enemy attack-0.055)*(enemy kill-0.02016)	636.52	65.015	9.79
max wait time	0.1568	0.0166	9.47

Note. p-value < 0.0001.

Figure 4: Important Regression Terms, All Statistically Significant

As it is sometimes difficult to understand the interactions and quadratic effects from just the regression coefficients, we find that an interaction plot can be a useful graph. Figure 5 contains the set of plots for the regression's two-way interaction terms. The presence of non-parallel lines is a visual indicator that two factors interact with each other, meaning that the effect of each factor on the response depends on the value of the other factor. We now give an interpretation of the interaction that occurs between external resupply time and number of LV. There are two complementary visual representations of this interaction. We will describe the interaction plot that appears in the fourth row, third column. When the number of LVs is at its highest value in the experiment (=20 and represented by the blue line), then decreasing external supply time over its range decreased average requests from (approximately) 0.7 to 0.5. This is not a huge difference because requests awaiting fulfillment were kept fairly low regardless of supply time due to the luxury of the larger number of vehicles. However, when the number of LV was at its lowest value (=8 and represented by the red line), then decreasing supply time had a much larger effect—in this case, average requests decreased from close to 5 to under 2. In other words, supply time has a greater impact when there are fewer LVs; in fact, decreasing the supply time can help mitigate the problems associated with having fewer vehicles.

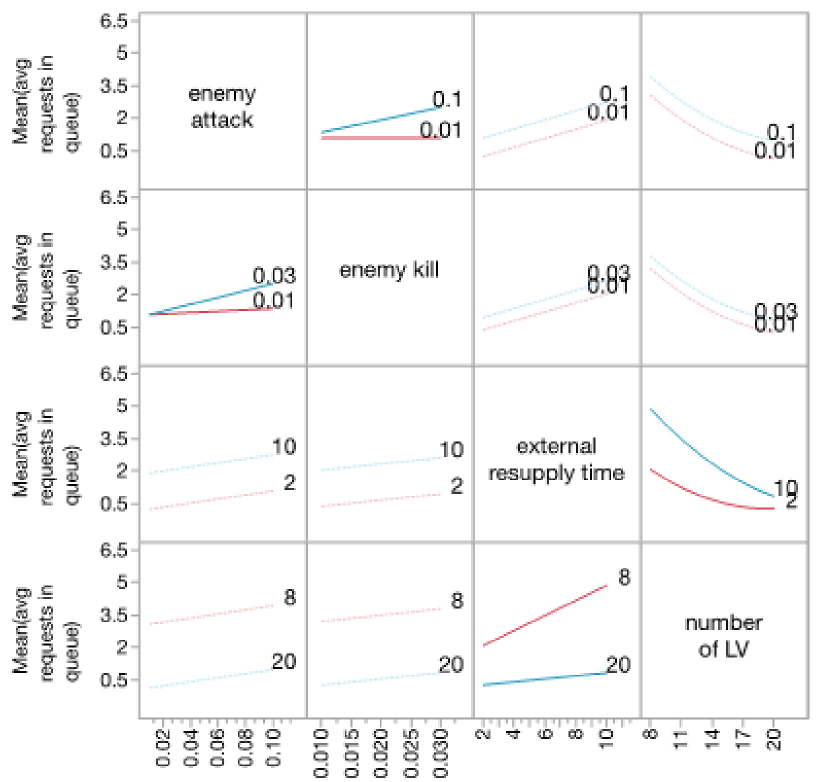


Figure 5: Interaction Profiles

Finally, regression is not the only type of metamodel possible. An alternative is the partition tree, as shown in Figure 6. The recursive partitioning technique that is used to produce the tree identifies factors and cut points that best predict the response. Recursive partitioning nicely complements the use of regression modeling because (1) it is a nonparametric approach (so it does not require any assumptions of the underlying data); (2) it reveals insight about interesting thresholds or cut-points that may be associated with jumps or discontinuities in the response; and (3) the “decision tree” structure that results is generally easily communicated and intuitively understood, even by those without a technical background.

We first notice that the two factors that appear in the tree are the same top two influential factors from the regression. The first (top-level) split in the tree occurs on the number of LVs, so this is a highly influential factor. The left and right nodes of this split indicate that when the number of LVs is greater than or equal to 13, the average number of requests is 1.08, but when the number of LVs is less than 13, then the average number of requests increases to 2.71. As a subsequent split reveals, when we are able to increase the number of LVs to 16 or more, the average number of requests drops to 0.82. The other factor that appears in the tree is external resupply time. In the cases where 16 or more LVs are available, dropping resupply time to under five days is able to reduce the average number of requests to 0.39 (close to zero). However, cases associated with fewer than 13 LVs and resupply time in excess of six days led to 3.3 requests awaiting fulfillment, on average.

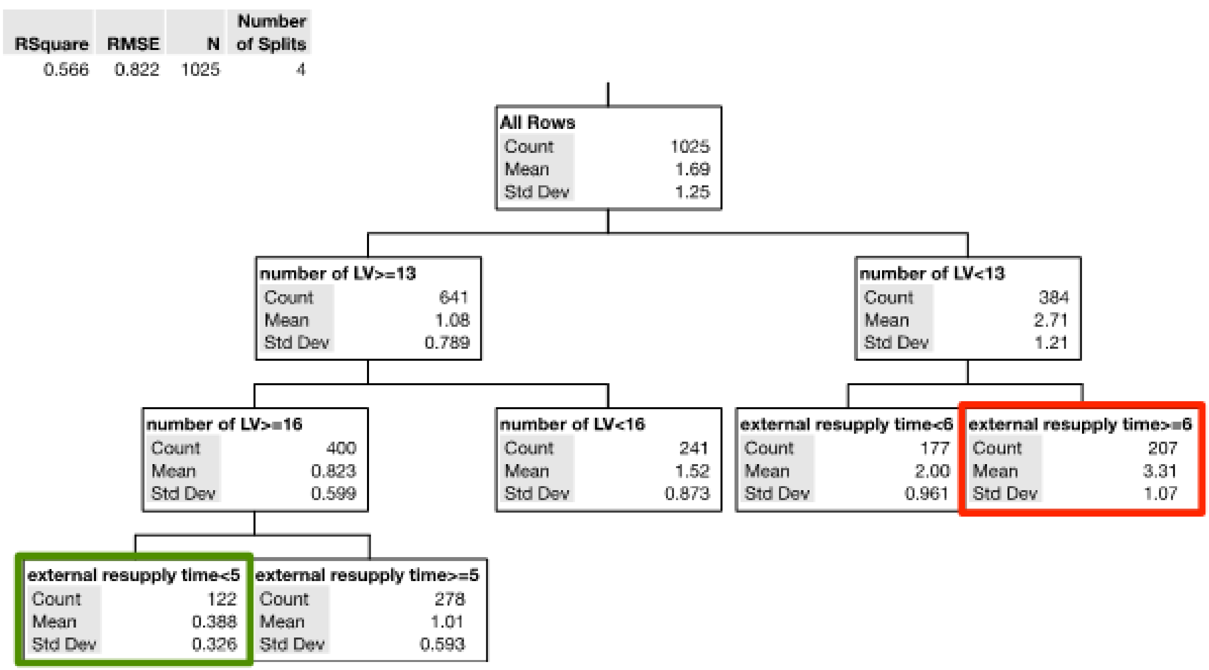


Figure 6: Partition Tree for Mean (Avg Requests in Queue) After Four Splits

With metamodeling approaches such as regression and partition trees, we can gain insight into which factors and interactions are most important, interesting cut points or “knees in the curve,” and which factors have little effect. Some of these findings may be contrary to initial intuition. However, there are other insights that can be gained as well from experimentation. As one example, we could identify which DPs (alternatives) met user-defined constraints on the response. Additionally, though we considered mean performance for these initial insights, we could go further and apply a loss function that captures both mean performance (does it meet a defined threshold?) as well as variability over random replications or uncontrollable/noise factors (more variability translates to higher risk). We may also consider multiple objectives, identifying if a trade-off is involved. We may reduce our set of alternatives that we need to consider by removing those that are dominated by others, leaving only those on the interesting Pareto optimal frontier. Finally, visual analysis through plots and graphs should always accompany and precede metamodel fitting, though for brevity we do not include it here.

We finally mention that all of this rich analysis is enabled through the application of efficient and flexible designs of experiment. Efficient designs enable experimentation over a number of factors and levels that is simply not possible through brute force (i.e., exhaustively sampling every possible combination). Efficient designs allow us to break the curse of dimensionality.

For more information and examples of data farming, and links to software and spreadsheets for constructing designs, see Lucas et al. (2015), Sanchez and Sanchez (2017), Sanchez et al. (2012, 2018), or the NPS SEED Center for Data Farming website at https://harvest.nps.edu.

Ongoing Work

In the second phase of this research, we plan to refine our simulation model and conduct further experiments, which could include a more detailed treatment and examination of how the quality, reliability, and time required to produce AM parts influence networked expeditionary logistics. Expertise about various properties and characteristics of different materials used in AM (e.g., polymeric materials, composite materials, and metal or alloy composition powders), as well as post-treatments required for the parts to achieve their final structural and physical characteristics, may guide the simulation model factors related to quality, reliability, and lead-times required for manufactured parts. Additionally, with an increased focus on the refinement of the manned-unmanned teaming concept (Department of the Navy, 2016, p.16), we plan to introduce the use of one or more unmanned vehicles into the model. Our second phase of research may then be guided by the following questions:

- How does the use of unmanned vehicles and AM affect the readiness of an expeditionary unit? What are the primary readiness drivers? Under what conditions do these either increase or decrease overall readiness?
- If parts made using AM differ in their characteristics from those made using current manufacturing processes, what are the ranges (or distributions) of the suitability of the parts for their intended use? Under what conditions are parts made using AM likely to be either more reliable, equally reliable, or less reliable than current parts?
- How does AM affect the life-cycle cost? What are the primary cost drivers? Under what conditions does AM either increase or decrease life-cycle cost?
- Are there win-win conditions where AM increases readiness while reducing costs? Are there lose-lose conditions where AM should be avoided because it reduces readiness while increasing costs? Are there conditions where trade-offs must be made between readiness and cost?

Summary

Our interest is in investigating the impact of AM on military logistics and life-cycle costs for Marine Corps expeditionary operations. We view AM as a potential transformative capability, but to realize its full potential for expeditionary operations, the Marine Corps logistics concept of operations must change.

Our work provides a template for augmenting the acquisition decision process by using simulation analytics—specifically, a data farming approach. Many characteristics of an AM-capable expeditionary operational unit can be explicitly studied as factors within large-scale simulation experiments. Consequently, we can identify which sources of data (e.g., demand patterns, reliability, quality, printing and processing time, lead-time) or their interrelationships are the key drivers of readiness and performance. This might help program managers set initial requirements, determine what should be monitored most closely as AM programs are rolled out, or assist in estimating the potential benefits as new AM compounds or processes become available over time.

References

Barno, D., & Bensahel, N. (2017). Three things the Army Chief of Staff wants you to know. *War on the Rocks*. Retrieved from https://warontherocks.com/2017/05/three-things-the-army-chief- of-staff-wants-you-to-know/

Cioppa, T. M., & Lucas, T. W. (2007). Efficient nearly orthogonal and space-filling Latin hypercubes. *Technometrics*, *49*(1), 45–55.

Dawes, J., Bowerman, R., & Trepleton, R. (2015). Introduction to the additive manufacturing powder metallurgy supply chain. *Johnson Matthey Technology Review*, *59*(3), 243–256.

Department of the Navy. (2017). *Department of the Navy (DoN) additive manufacturing (AM) implementation plan v2.0.* Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/1041527.pdf

Department of the Navy, Headquarters United States Marine Corps. (2016). *The Marine Corps Operating Concept: How an expeditionary force operates in the 21st century.* Retrieved from https://www.mccdc.marines.mil/Portals/172/Docs/MCCDC/young/MCCDC-YH/document/final/Marine%20Corps%20Operating%20Concept%20Sept%202016.pdf?ver=2016-09-28-083439-483

Department of the Navy, Headquarters United States Marine Corps. (2017). *Headquarters Marine Corps procedural guidance update on the management and employment of additive manufacturing* (MARADMINS Number 594/17). Retrieved from https://www.marines.mil/News/Messages/Messages-Display/Article/1353764/headquarters-marine-corps-procedural-guidance-update-on-the-management-and-empl/

Fessenden, M. (2013, September 1). 3D printed windpipe gives infant breath of life. *Scientific American*, *309*(3), 22. doi:10.1038/scientificamerican0913-22a

Frazier, W. E. (2014). Metal additive manufacturing: A review. *Journal of Materials Engineering and Performance*, *23*(6), 1917–1928.

Graziosi, S., Rosa, F., Casati, R., Solarino, P., Vedani, M., & Bordegoni, M. (2017). Designing for metal additive manufacturing: A case study in the professional sports field. *Procedia Manufacturing 11*(2), 1544–1551.

Gue, K. R. (2006). Very high density storage systems. *IIE Transactions*, *38*(1), 79–90.

Gue, K. R., & Kang, K. (2001). Staging queues in material handling and transportation systems. In B. A. Peters et al. (Eds.), *Proceedings of the 2001 Winter Simulation Conference* (pp. 1104–1108). Piscataway, NJ: IEEE.

Hodsden, S. (2016, October 21). 3D-printed biodegradable stents bypass traditional stent complications. MED DEVICE online. Retrieved from https://www.meddeviceonline.com/doc/d-printed-biodegradable-stents-bypass-traditional-stent-complications-0001

Khajavi, S. H., Partanen, J., & Holmstrom, J. (2014). Additive manufacturing in the spare parts supply chain. *Computers in Industry*, *65*(1), 50–63.

Lucas, T. W., Kelton, W. D., Sanchez, P. J., Sanchez, S. M., & Anderson, B. L. (2015). Changing the paradigm: Simulation, now a method of first resort. *Naval Research Logistics*, *62*, 293–305.

Lynch, G. E. (2019). *Networked logistics: Turning the iron mountain into an iron network* (Master's thesis). Naval Postgraduate School, Monterey, CA (in process).

Mellor, S., Hao, L., & Zhang, D. (2014). Additive manufacturing: A framework for implementation. *International Journal of Production Economics*, *149*, 194–201.

Moore, T. A., McConnell, B. M., & Wilson, J. R. (2018). Simulation-based evaluation on integrating additive manufacturing capability in a deployed military environment. In M. Rabe et al. (Eds.), *Proceedings of the 2018 Winter Simulation Conference* (pp. 3721–3729). Piscataway, NJ: IEEE.

Ng, M. (2018, August 24). Additive manufacturing: Taking the "iron mountain" out of military logistics. Land Power Forum, Australian Army Research Centre. Retrieved from https://www.army.gov.au/our-future/blog/logistics/additive-manufacturing-taking-the-iron-mountain-out-of-military-logistics

Oettmeier, K., & Hofmann, E. (2016). Impact of additive manufacturing on supply chain network structures: An exploratory case study analysis. *Journal of Manufacturing Technology Management, 27*(7), 944–968.

Pelczar, K. (2010, May 25). 7th ESB Marines awarded for "heroic actions" during Leatherneck fire. 1st Marine Logistics Group News. Retrieved from https://www.1stmlg.marines.mil/News/News-Article-Display/Article/543570/7th-esb-marines-awarded-for-heroic-actions-during-leatherneck-fire/

Sanchez, S. M. (2018). Data farming: Better data, not just big data. In M. Rabe et al. (Eds.), *Proceedings of the 2018 Winter Simulation Conference* (pp. 425–439). Piscataway, NJ: IEEE. Retrieved from https://www.informs-sim.org/wsc18papers/includes/files/035.pdf

Sanchez, S. M., & Sanchez, P. J. (2017). Better big data via data farming experiments. In A. Tolk, J. Fowler, G. Shao, & E. Yucesan (Eds.), *Advances in modeling and simulation—Seminal research from 50 years of Winter Simulation Conferences* (pp. 159–179). Springer.

Sanchez, S. M., Sanchez, P. J., & Wan, H. (2018). Work smarter, not harder: A tutorial on designing and conducting simulation experiments. In M. Rabe et al. (Eds.), *Proceedings of the 2018 Winter Simulation Conferences* (pp. 237–251). Retrieved from https://www.informs-sim.org/wsc18papers/includes/files/022.pdf

Sanchez, S. M., Lucas, T. W., Sanchez, P. J., Wan, H., & Nannini, C. J. (2012). Designs for large-scale simulation experiments, with applications to defense and homeland security. In K. Hinkelmann (Ed.), *Design and analysis of experiments, Volume 3: Special Designs and Applications* (pp. 413441). New York, NY: Wiley.

Shah, T., Faizi, F., & Zucchino, D. (2019, March 1). Taliban target key Afghan base on eve of peace talks. *New York Times*. Retrieved from https://www.nytimes.com/2019/03/01/world/asia/taliban-peace-talks-attack.html

Silva, J. V. L., & Rezende, R. A. (2013). Additive manufacturing and its future impact in logistics. In *Proceedings of the 6th IFAC Conference on Management and Control of Production and Logistics* (pp. 277–282). Fortaleza, Brazil: The International Federation of Automatic Control.

Snow, S. (2019, March 1). Taliban fighters try to storm base in Helmand province housing U.S. Marines and Afghan forces. *Marine Corps Times*. Retrieved from https://www.marinecorpstimes.com/news/your-marine-corps/2019/03/01/taliban-fighters-try-to-storm-base-in-helmand-province-housing-us-marines-and-afghan-forces/

Spangenberg, K. M. (2017, February 6). Rethinking logistics. Retrieved from http://www.defensefellows.org/?p=418

Sturmer, J. (2015, February 25). 3D printing: Australian researchers create jet engine, breakthrough captures attention of Airbus and Boeing. ABC News. Retrieved from https://www.abc.net.au/news/2015-02-26/australian-researchers-create-first-3d-jet-engine/6262462

U.S. Army Research, Development and Engineering Command. (2017). Additive manufacturing. Retrieved from https://www.army.mil/standto/2017-08-08

Westerweel, B., Basten, R. J. I., & van Houtum, G.-J. (2018). Traditional or additive manufacturing? Assessing component design options through lifecycle cost analysis. *European Journal of Operational Research*, *270*, 570–585.

Zelinski, P. (2019). Metal 3D printing in a machine shop? Ask the Marines. Retrieved from https://www.additivemanufacturing.media/blog/post/metal-3d-printing-in-a-machine-shop-ask-the-marines

Acquisition Research Program
Graduate School of Business & Public Policy
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net