SYM-AM-18-041

PROCEEDINGS
OF THE
FIFTEENTH ANNUAL ACQUISITION RESEARCH SYMPOSIUM

WEDNESDAY SESSIONS
VOLUME I

Acquisition Research:
Creating Synergy for Informed Change

May 9–10, 2018

Published April 30, 2018

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

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

Complexity in an Unexpected Place: Quantities in Selected Acquisition Reports

Gregory A. Davis—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. Except for GPS-OCX, all of them had complicated definitions of a unit. [gdavis@ida.org]

Margaret L. Giles—joined IDA's Cost Analysis and Research Division in 2016 after spending five years as a civilian with Headquarters, Department of the Army. As an Army civilian, she participated in the development of multiple Army programs through several budget and POM cycles, focusing in particular on finding greater efficiencies within the Army budget. Giles holds a master's degree in international affairs from the George Washington University and a bachelor's degree in political science from Southwestern University. [mgiles@ida.org]

David M. Tate—joined the research staff of IDA's Cost Analysis and Research Division in 2000. Since then, 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. 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]

Abstract

We have looked at the definition of units in numerous acquisition programs and discovered that the units reported are almost never simple; in some programs, no two units are the same, and almost invariably the units produced at the end of a long production run are substantially different from the early ones. We have identified three reasons why the units may differ. The first reason is changes over time, generally as system capabilities are improved. The second is due to mixed types, where units that are inherently dissimilar—such as CH-47F and MH-47G helicopters—are produced by the same program and each is called one unit. The final reason why units can differ is reporting accidents. We give examples of all three and discuss possible methods of improving the reporting requirement

Introduction

Acquisition data are primarily about a few questions: "How much funding?," "How much are we getting?," "When are we obligating the funds?," and "When are we getting what we paid for?" All of these questions are interesting, and none are straightforward. Most have been addressed elsewhere and continue to get attention. The question of "What are we getting?," however, is generally treated as though it were simple. Our experience tells us that counting quantities is often not straightforward. This report describes the findings of research that has taken us deeper into this question, showing that quantities are almost always complicated.

The Director of Performance Assessments and Root Cause Analyses (D,PARCA),[1] asked the Institute for Defense Analyses (IDA) to review the quality and utility of data used for acquisition oversight; we started with the question of quantities.

Selected Acquisition Reports

Section 2432 of Title 10 U.S.C. requires the Secretary of Defense to submit to the Congress a yearly status report for each Major Defense Acquisition Program (MDAP), known as the Selected Acquisition Report (SAR), which provides performance, schedule, and cost data. Each SAR includes separate cost estimates for several categories. Both past actual costs and future anticipated costs are reported, as well as quantity of units, for the expected life of the program (DoD, 2016).

Within the Defense Acquisition Management Information Retrieval (DAMIR) system—the repository for SAR data—the *Track to Budget* section identifies the budget program elements (PEs) and procurement line item numbers (LINs) for each appropriation associated with a program in a particular fiscal year, allowing the user to find the equivalent cost and quantity data in the President's Budget (PB) Submission prepared in the same year.[2] Reconciling SAR data with the equivalent PB Submission proves difficult, however, as cost estimates can vary between the two sources, and some PEs and LINs are shared among multiple programs in a non-transparent way. In some cases, the SAR and PB define quantities differently.

Neither the PB nor the SAR is perfect. In general, the Justification Books that the Services produce annually to support the PB contain more detail, which is good for analysis, but if it extends beyond the Future Years Defense Program (FYDP), it is as a single column labeled *To Complete*. The PB also does not include much history, with most of it in a single column labeled *Prior Years*. The SAR reports costs in both Then Year (TY) and Base Year (BY) dollars, while the PB reports exclusively TY dollars. The SARs are the Office of the Secretary of Defense (OSD)'s primary data source for analyzing MDAPs. This dataset is what analysts from many different organizations typically use, per the recommendation of the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]) staff, who describe SAR data as "the official numbers."

Why Selected Acquisition Reports Matter

The SARs are not the dataset used most often for decision-making inside the DoD. When senior leaders make large resource decisions, analysts most often assemble datasets to suit the needs of the decision-maker by pulling data from non-public systems or conducting data calls. Why then do we care about the quality of data in the SARs?

The SARs matter for two reasons: *triggering* and research. What we call *triggering* is why the SARs were created. The Services trigger investigations when they seek milestone

---

[1] PARCA is an office that was under the aegis of the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]).

[2] The President's Budget and annual SAR submissions both generally come out in the second quarter of each fiscal year. The years on a matching budget and SAR set differ by two. The budget is named for the year ahead, and the SAR is a snapshot of the program in the recent past. For example, in the second quarter of FY 2016, the FY 2017 budget was released, quickly followed by the December 2015 SAR.

authorities from the OSD. The OSD can also trigger analyses for program reviews based on the Service's annual submissions, such as the Program Objective Memorandum. Only the SARs provide regular information at the program level. For example, no other annual submission can tell OSD or the Congress about the projected procurement costs for a program that is expected to leave the development stage in five years.

Research on defense acquisition is continuously occurring in government agencies, think tanks, universities, and other organizations. In the past, researchers looking across programs have considered amount of cost growth (McNicol, 2004), setting of production rates (Rogerson, 1991), comparisons among different commodity types (Drezner, 2011), and many other subjects. This research helps the government, and SARs are the best source for comparisons across programs. While it is the nature of research that we cannot predict which research projects will yield fruitful results, we know that better quality data will yield better research results.

Nunn-McCurdy Breaches

Critical Nunn-McCurdy (N-M) breaches are established by statute. If an MDAP sustains too much cost growth, a review takes place that generally leads to either changes in the program or, occasionally, termination. PMs generally want to avoid N M breaches. Too much cost growth is defined in terms of Average Procurement Unit Cost (APUC) or Program Acquisition Unit Cost (PAUC):

- APUC = Procurement Costs/Procurement Quantities
- PAUC = Total Program Costs/(Procurement + Research, Development, Test, and Evaluation [RDT&E] Quantities)

There are four possible critical N–M breaches, two for APUC and two for PAUC. The breach calculation is performed by measuring the percentage growth in APUC or PAUC. A critical breach occurs when the variable has increased by at least 25% against the current Acquisition Program Baseline (APB) or 50% against the original APB. The original APB is the APB that was established during the Milestone (MS) B decision (formerly Milestone II).

Each SAR contains a unit cost report that compares the current APUC and PAUC estimates to the original APB and a second unit cost report comparing the estimates to the current APB if the current APB is not the same as the original one.

Subprograms

An MDAP's baseline may indicate that it has multiple subprograms to increase visibility into the program's activities. If so, each unit produced and each dollar spent is assigned to one of the subprograms. Subprograms have been used to distinguish variants of a system, such as two similar but not identical missiles, or to look at different parts of a system, such as engines and airframes. Each year, each subprogram has its APUC and PAUC calculated and compared to the baseline. According to the N-M Act (10 U.S.C. § 2433), if any subprogram exceeds its thresholds, an N-M breach is declared for the entire program, not just the subprogram that exceeded its baseline.

The popularity of subprograms has changed through the years, as can be seen in Figure 1. The total number of programs each year did not change much, but declaring subprograms became less common from 1998 to 2009, when a rebound started. It is not clear what has caused these changes.

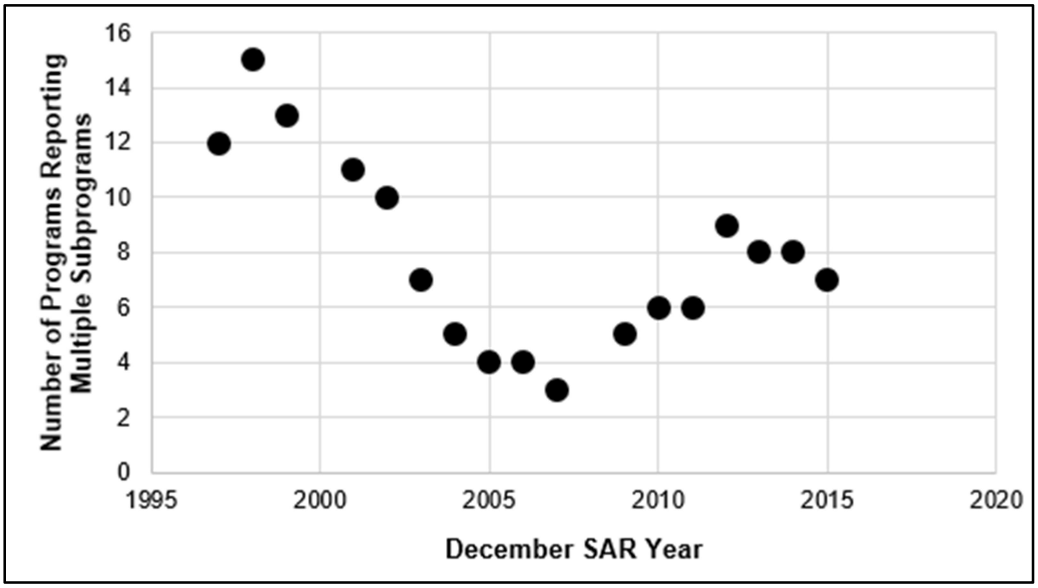

Figure 1. Subprograms in MDAPs 1997–2015

An Example of Budget and SAR Discrepancy: Gray Eagle

Quantity reporting in the SAR is the focus of this report. We begin with a few illustrative examples. The Army's MQ-1C Gray Eagle program acquires unmanned aerial systems (UASs). In the *Track to Budget* section of its 2015 SAR, the program identifies the following LINs within the *Aircraft Procurement*, *Army* appropriation:

- A00005 (MQ-1 UAS);
- A01001 (MQ-1 Payload, which includes funding for other programs); and
- A01005 (Common Sensor Payload Full Motion Video (CSP FMV), a sub-Line Item Number to A01001).

Both A01001 and A01005 are listed as *shared*. The quantities and costs found in these LINs in PB 2017, however, differ from those in Gray Eagle's 2015 SAR, as shown in Table 1. Note that both sources project the program to finish in FY 2018.

Table 1. Gray Eagle Program SAR and Budget Comparison

	FY 2015		FY 2016		FY 2017		FY 2018	
Data Source	Q	Cost (TY $)	Q	Cost (TY $)	Q	Cost (TY $)	Q	Cost (TY $)
PB 2017	19	$246,490 K	17	$355,445 K	0	$60,117 K	0	$10,806 K
Dec 2015 SAR	2	$246,400 K	3	$322,200 K	0	$60,200 K	0	$15,200 K
Difference	**17**	**$90 K**	**14**	**$33,245 K**	**0**	**-$83 K**	**0**	**-$4,394 K**

Note: K – thousand; Q – quantity.

The cost differences in FY 2015 and FY 2017 are minimal, but there is no obvious explanation for the more significant differences in FY 2016 and FY 2018 costs. In PB 2017, the unit of accounting for this program is one unmanned airplane. However, the capability is also dependent on how many ground assets for operating the systems are acquired and on the differences between aircraft, as they are not all the same. In the SAR, the quantity is measured in companies, each of which contains several aircraft with different configurations and some amount of ground equipment. There is a standard measure for what a company is, but not all companies fit the standard description. While the SAR does include a great deal of detail in various written sections, this makes it difficult to use the quantities in the data for quantitative analysis.

A Complex Example: The CH-47F Chinook Program

The Army's CH-47F Improved Cargo Helicopter program demonstrates challenges that can occur when counting quantities across years in both the PB and SAR. This program builds Chinook helicopters, which are easy to count, yet there are serious questions when looking at the data.

First, the CH-47F program's definition of one *unit* has changed over time. In the early days of the program (as reflected in the original June 1998 SAR), the plan was to SLEP[3] 300 existing CH-47D helicopters to an updated configuration, which would be called the CH-47F. In PB 2005, the plan was to SLEP 287 CH-47D helicopters to the CH 47F configuration and 50 MH-47E Special Operations helicopters to a new MH-47G configuration. The definition of a unit had changed to include both CH-47D/F conversions and MH-47E/G conversions, which produce distinct end items and have different expected costs.

The Army's February 2007 budget justification forms expanded the set of planned activities to include all of the following:

- SLEPs of CH-47D to CH-47F
- SLEPs of MH-47E to MH-47G
- New builds of CH-47F from scratch for active duty Army units
- New builds of CH-47F in a different configuration for National Guard units

The reported and projected unit costs for these activities were all different. More to the point, the definition of a unit now included not only a remanufactured existing helicopter, but also a newly built helicopter of the same design. While these may be functionally identical from an operational point of view, there are reasons why an analyst would want to know how many of each were to be built—and at what cost. To further complicate matters, the helicopters produced (both SLEP and new build) employ a mix of mission subsystems, some of which could be repurposed from a remanufactured helicopter or other existing decommissioned helicopter, and some of which had to be built (and purchased) new. The type and number of repurposed subsystems continued to vary from year to year, so that the production inputs (and price) even for new build active component CH-47Fs were different from year to year.

The end result of these changes is that any given unit produced by the CH-47F program might have any one of the MH-47G, CH-47F Army, or CH-47F National Guard configurations. A CH-47F unit might be remanufactured or built new. Whether remanufactured or new, it might include some unspecified mix of government-furnished (free) and contractor-furnished (at a price) mission subsystems. For example, as of the 2013 PB submission (February 2012), 43 new build units had been produced at an average cost of $15 million, of which $1.1 million per unit was for government-furnished equipment (GFE).

---

[3] SLEP is the acronym for "Service Life Extension Program" and is often used as a verb in defense circles. A SLEP can be funded with either procurement or Operations and Maintenance dollars.

The estimated cost to complete the new build program was $2.19 billion for 112 units, or $19.5 million per unit, of which $2.4 million per unit was expected to be GFE. This reflects the expectation that units authorized through FY 2013 would use recovered avionics suites from existing aircraft, but that half of the new build units after that would require new-build (contractor-furnished) avionics. There were clearly anticipated differences in components and cost between units produced up to that point and units expected to be produced in the future.

Furthermore, there are inconsistencies between the SAR and the PB submissions regarding which units comprise the CH-47F program. How new builds versus SLEPs are counted in different years is unusual and is described in detail in the Accidents chapter.

Organization of This Paper

We have divided the common differences among unit definitions into three buckets: changes over time, mixed types of units, and reporting accidents. It is not uncommon for more than one category to apply to a given program; the Chinook has all three. The next three chapters describe what each of these categories means, how confusions arise, and what analysts should do when trying to use cost reporting data. In Suggested Adjustments to Reporting, we make some modest recommendations for modifications to acquisition data reporting that could help make the data more useful for many sorts of analyses. As part of those recommendations, in A Thought Experiment: JLTV, we consider the Joint Light Tactical Vehicle (JLTV) program—how its reporting might have been done differently and what the ramifications of those differences might have been.

Changes Over Time

Implicit in the concept of a *unit* is that every instance of the unit should be identical. Every inch should be the same length, every second should have the same duration, and every run scored in a baseball game should count equally. As noted above, this is often not true of procurement units in MDAPs. One reason that non-identical units might arise is that the product may evolve over time. Even when counting quantities is simple, such as when counting helicopters or ships, the units procured at different times are usually different in both cost and capability. In our full report (Davis, Giles, & Tate, 2017), we detail changes over time in ships, tactical aircraft, and tactical land vehicles. In this excerpted report, we look only at one program, the Air Force's AIM-120 Advanced Medium Range Air-to-Air Missile (AMRAAM) program.

The AMRAAM program was established at a Defense Systems Acquisition Review Council MS I Review in November 1978. After an extended development period, an acquisition baseline of 24,320 units was set in December 1988. The first production units were authorized under the FY 1987 budget and fielded in 1991. The acquisition target was reduced to 16,427 missiles in a 1992 re-baselining that also doubled the expected per-unit cost.

The AIM-120 is still in production. The Air Force now intends to buy a total of 12,851 missiles, and the Navy an additional 4,461 missiles, for a total of 17,312. The final unit is projected to be authorized in FY 2025, almost 40 years after the first unit.

The explanation for the continued utility of the AIM-120 is that the missiles being produced today are nothing like the missiles that were produced in the early 1990s. Figure 2 shows the history of average unit cost by annual production lot for AMRAAM missiles, with filled shapes showing historical data and open shapes, projections. After a typical initial learning curve, it is clear that there have been major changes to the program over its history. In fact, many upgrades, modifications, and wholesale redesigns of the missile have occurred over time; the Teal Group reports seven (Teal Group Corporation, 2014, p. 133). Some were simply improvements, while others had new functions, such as the AIM-120C3, designed with smaller control surfaces to fit inside the weapons bay of an F-22 Raptor, and the AIM 120D, which includes many new features such as Global Positioning System navigation and a two-way datalink.

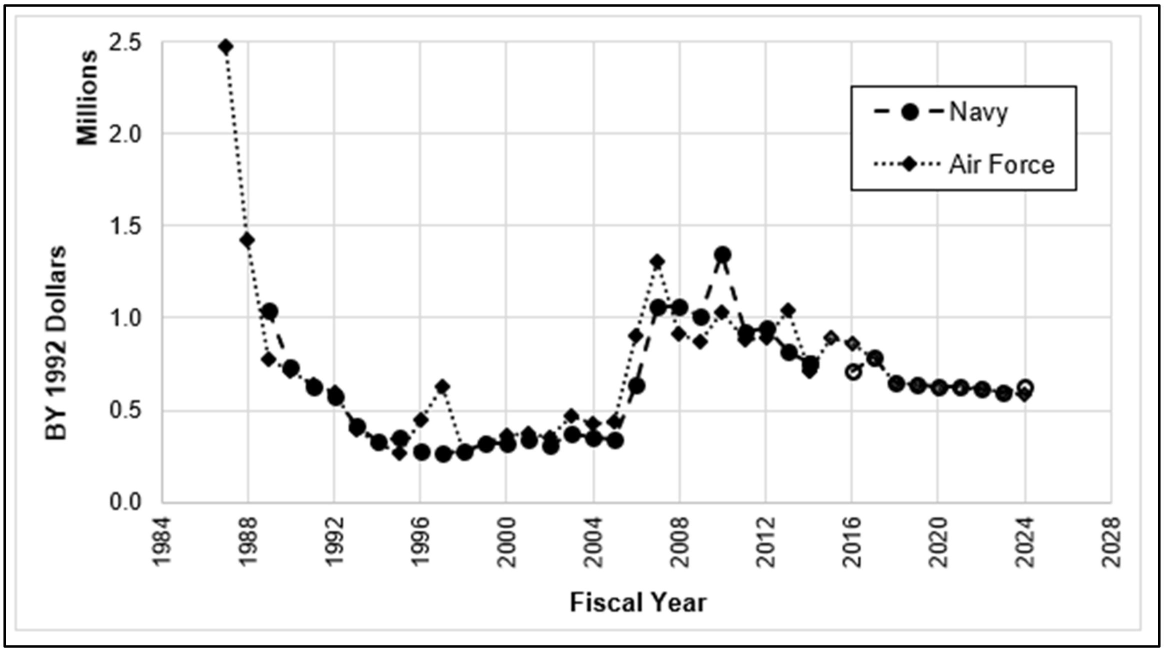

Figure 2. Annual Average Unit Cost for AMRAAM Missiles

There is no sense in which an AIM-120D is “the same thing” as an AIM-120A, or even an AIM-120C7. This is a clear instance in which the implicit assumption that units are interchangeable has been violated.

Of course, within the AMRAAM program, there is no confusion about the kinds of missiles that are currently being produced, their capabilities, or plans for future improvements. The question, then, is how the program might adjust its data reporting to enhance transparency for planners, analysts, and oversight bodies.

Mixed Types

Program offices often procure different end items at the same time. These items are usually similar to each other but substantially different; yet, for quantity reporting purposes, each is considered one *unit*. This often comes about because of different missions or end users. Sometimes, the types are completely different. To illustrate this concept, in this paper we look at an electronics suite.

The Navy's Integrated Defensive Electronic Countermeasures (IDECM) program acquires electronics suites to protect the various F/A-18 aircraft from radio frequency guided missiles. IDECM achieved MS II approval in October 1995, although it was too small at the time to be an MDAP. Because of changes, it became an MDAP in March 2008 and issued its first SAR in June 2008. The *Mission & Description* section of the December 2015 SAR describes the blocks as follows:

- IDECM Block 1: A federated suite, consisting of the ALQ-165 On-Board Jammer (OBJ) and ALE-50 expendable decoy
- IDECM Block 2: An integrated suite, consisting of the ALQ-214 OBJ and ALE-50 expendable decoy
- IDECM Block 3: An integrated suite, consisting of the ALQ-214 OBJ and ALE-55 Fiber Optic Towed Decoy
- IDECM Block 4: A Hardware Engineering Change Proposal to the ALQ-214 OBJ to render it suitable for operation on F/A-18C/D aircraft, while retaining all functionality, when installed on F/A-18E/F

The SAR contains two subprograms: IDECM Blocks 2/3 and IDECM Block 4. The December 2015 SAR reports an APUC of $2.502 million for Block 4 and a far lower APUC of $0.090 million for Block 2/3. This is because the quantities are so different. Block 4 has a quantity of 324, roughly the number of airplanes they will be protecting. Block 2/3 has a quantity of 12,805, although the Navy bought fewer than 600 F/A-18E/Fs in total. Eighty-five of the 12,805 were purchased with *1506 Navy Aircraft Procurement* funds and the balance were or will be bought with *1508 Procurement of Ammunition, Navy and Marine Corps* funds. We presume that those purchased with ammunition funding are only the disposable decoys. The unit costs based on the *End Item Recurring Flyaway* column in each year are presented in Figure 3.

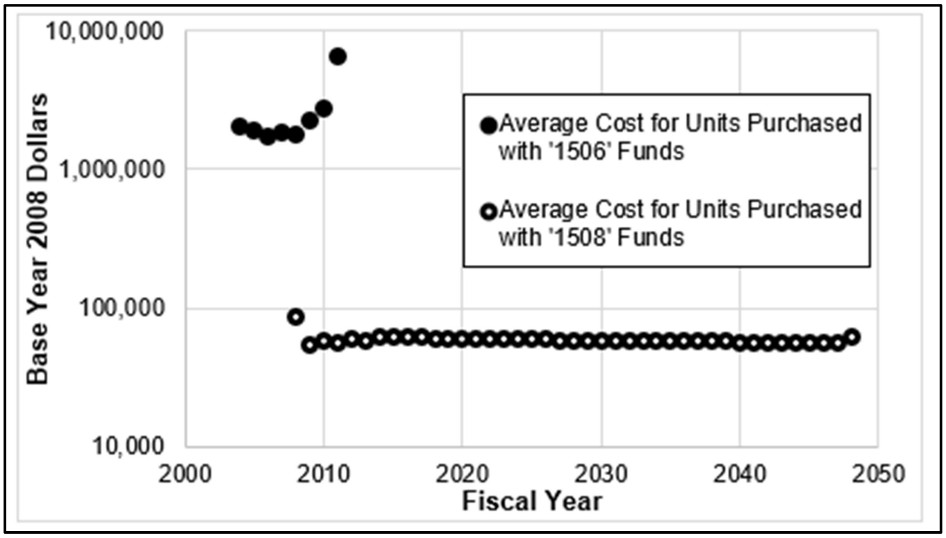

Figure 3. IDECM Block 2/3 Annual Unit Cost by Appropriation Type From the December 2015 SAR

Even though showing the two unit costs on the same chart requires plotting them on a logarithmic scale, the two are both considered *units* for the official unit cost calculation. Just within the more expensive *1506* units, it is clear that there have been significant changes, as the cost there does not follow a typical learning shape, which would be expected to slope down.

While the details have changed with time, the IDECM program has used this reporting system since it issued its first SAR in June 2008.

Accidents

The confusions described above generally come about because of some decision by leadership about how the data should be presented;[4] this category, in contrast, is about cases in which it seems there were also outright errors in how the quantity numbers were put together. We do not know how frequently this happens, but we know that it does happen and can sometimes persist for several years. We do not suggest that any of the cases described below involve intent to confuse analysts, but they did have that effect.

We used the term *accidents* (as opposed to *mistakes* or *errors*) because it was the term a government official in AT&L applied to reporting anomalies for programs like Chinook. We identified three in the December 2015 SARs: Chinook helicopters, the Evolved Expendable Launch Vehicle, and the ICBM Fuze modification programs. It is quite possible there are more. We only present the Chinook situation in detail, as the others were about how dollars were assigned within the program.

[4] Or, more precisely, leadership makes a decision about how the program should be managed and what systems it should produce, possibly without considering the impact this will have on the coherence and consistency of quantity or unit cost reporting.

As described in the section titled A Complex Example: The CH-47F Chinook Program, the CH-47F Chinook Improved Cargo Helicopter program made a number of changes to its definition of *unit* over the course of the program. In the December 2015 SAR, however, the program apparently lost track of how it had been defining a unit and submitted quantity and cost forecasts that did not include all of the units identified in the simultaneous PB submission.

Figure 4 shows the discrepancy between predicted future quantities in the December 2015 SAR and the corresponding 2017 PB. Through FY 2017, the total quantities match perfectly. Beginning in 2018, units described as SLEP units in the PB are missing from the SAR forecast. As a result, the projected cost of these units is not included in the SAR calculations of APUC, PAUC, APUC growth, or PAUC growth.

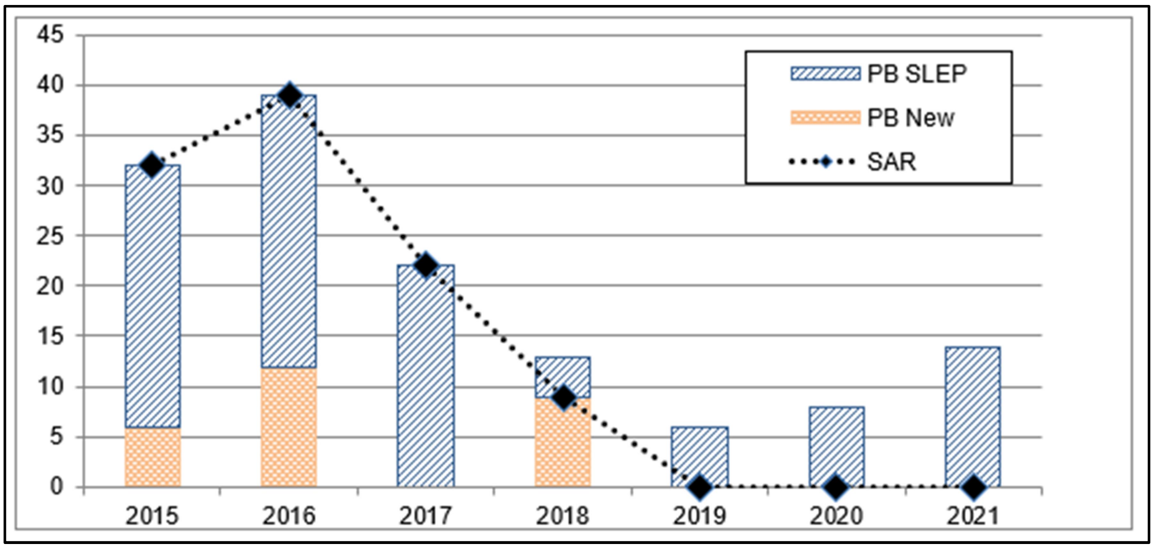

Figure 4. SAR Versus PB Production Quantities for CH-47F

Since the SAR and the concurrent PB are required by law to agree on costs and quantities within the FYDP, this is clearly an accident.

There is also an ongoing mismatch between the SAR and PB with regard to the past quantity produced. In the SAR, every past unit produced is counted, regardless of whether it was a SLEP unit or a new build. In the PB, in the early years of the program, there were no top-level quantities reported, presumably on the basis that upgrades to CH 47F configuration were just one of many ongoing upgrades in the Army's helicopter fleet. Typically, programs that perform multiple types of upgrades, but are not applying all of them to every legacy platform, report the number of each type of upgrade performed separately. They do not typically roll these up into a total quantity for the program's LIN because the individual upgrades are not comparable and the number of platforms modified does not match the total number of any one type of modification.

When the decision was made to build CH-47F helicopters as new builds, the program began reporting a total quantity of units at the line item level and chose to include both SLEP and new build units in this total. However, they never looked back to include previously produced SLEP units in the *Prior Quantity* total. As a result, each new SAR and PB submission disagree both on how many helicopters have been procured and on how many will have been procured in total when the program is finished.

Suggested Adjustments to Reporting

One possible response to the issues described previously is to tell programs never to change what they are buying; once the baseline is set and the program is approved, the plan should be followed and the systems should not change. This assumption is implicit in the data reporting process. And yet, this has never been government practice, and we do *not* recommend that it be adopted. Our military goes to great lengths to provide our warfighters with the best possible equipment, and we should not forbid that just to make bookkeeping easier. We do offer some modest proposals that could make the reported data more useful, but first we need to be careful about incentives.

Data and Incentives

Data recording systems provide incentives. "You get what you pay for" is an old expression. In 2007, Dr. H. Thomas Johnson wrote, "Perhaps what you measure is what you get. More likely, what you measure is *all* you get" (2007). If the acquisition system's data requirements are not aligned with the system's goals, suboptimal performance may follow. This is exacerbated when penalties are associated with data reporting. Generally, people would prefer to report accurate data, but when the data will be used to justify punishment, the reporters are incentivized to either change the facts that lead to the data—possibly in creative and unproductive ways—or to provide incorrect data.

The statute that defines the N-M breach specifies PAUC and APUC thresholds that influence program behavior. Since N-M reviews impose costs on programs and can trigger cancellation of a program, many people in defense acquisition, including program managers (PMs), try to avoid them. This likely accounts for some of what we see in data reporting today.

Any changes made to the system need to be considered in this light. If people's careers will depend on what data they report, at times those data are more likely to reflect what is needed to satisfy the checker rather than reality. Furthermore, people will bend reality to make the data look "right" even if that will not yield the best actual result for national security.

Monitoring Changes Over Time

If we accept that the units produced during the course of a program will change over time, PMs should be given useful and standardized ways to describe (and ideally quantify) those changes, both for past units produced and planned future production.

The current taxonomy of SAR Variance Categories recognizes seven possible reasons for cost growth. Cost growth due to design changes must always be categorized as "Engineering," lumping together planned and unplanned changes, as well as optional versus necessary changes. For oversight and analysis, it would be useful to be able to distinguish at least three sub-categories of "changes over time":

- Pre-planned product improvements (P3I),
- Unplanned changes (necessary and unnecessary), and
- Block upgrades or evolutionary acquisition.

Pre-Planned Product Improvements (P3I)

P3I is a form of *spiral* acquisition, in which the first units produced do not include all of the capabilities that the procuring Service has identified as being required. The reasons for delaying might be budgetary, technical, operational, or some combination. The key is that the program has a plan from the beginning to add specific known improvements and has developed cost and schedule estimates for those improvements. This allows P3I costs to be included in the SAR and other program submissions.

In the current SAR, or even the more detailed PB, it is difficult to report current or anticipated P3I costs in a transparent fashion. The additional costs beyond what the program would cost if the improvements were not made will be a mix of RDT&E costs (for developing and testing the new design), nonrecurring costs (for things like new documentation and tooling), end-item recurring flyaway (EIRF) costs (for actual production of the improved units), possible non end-item recurring flyaway (NEIRF) costs (if improvements are made to non-end-item systems), and support costs (if the cost of support and/or spares for the new design is not exactly the same as for the original design).

For the marginal cost of improvements to be visible in the SAR, reporting would need to explicitly include P3I costs. One way to do this would be as follows:

- If the planned improvements are small in number and to be done at a few discrete times during the production run, treat them like Block Upgrades (see Block Upgrades, Evolutionary Acquisition, and Agile Development section).
- If the planned upgrades are more numerous and continuous, establish a Planned Upgrades subprogram and report the RDT&E and Procurement costs associated with planned changes to the original design under that subprogram. For each year in the SAR Annual Funding report, the program should report the following:
    - Under the main end item subprogram, report the quantity produced or planned and the estimated cost if those units had been made to the original design.
    - Under the Planned Upgrades subprogram, report zero quantity and the additional marginal procurement cost for the lot due to design changes. This additional cost should be split among EIRF, nonrecurring, and support costs in the usual way.
    - Report RDT&E costs for the original design under the primary end item subprogram.
    - Report RDT&E costs associated with planned design changes in the Planned Upgrades subprogram.

This system would allow analysts to clearly understand how much of the price change over time was driven by planned improvements and how much was unexpected. It would support meaningful learning curve modeling and also provide some progress tracking of new capability insertions. The narrative portions of the SAR would describe the capability enhancements obtained to date, the plan for future insertion of new capabilities, and the unexpected changes made to the base program.

On the other hand, this system introduces a potentially onerous new type of reporting—namely, the hypothetical cost of the units if they had all been made to the original design. This is not information programs currently possess, and there are potential pitfalls and perverse incentives in how programs might choose to compute and report these counterfactual costs. In particular, cost growth due to design changes that might have been necessary in the base program (e.g., for safety reasons, to meet threshold requirements, or due to diminishing manufacturing sources) could be allocated either to the base subprogram or to the P3I subprogram, whichever seemed least likely to risk an N-M breach.

For N-M purposes, several regulatory changes might be beneficial. First, the primary end item and the Planned Upgrades should be treated as separate triggers. The primary end item would use the usual PAUC and APUC thresholds. The Planned Upgrades subprogram might only have limits based on total cost growth, or perhaps time-phased cost growth (e.g., average cost per year, rather than average cost per unit). Ideally, a breach on the Planned Upgrades subprogram would *not* imply a breach on the base subprogram (although the reverse would not be true).

Under this system, the main temptation for struggling programs would be to mischaracterize some of their core program cost growth as P3I, so as to avoid an N-M breach on the primary end item. By shedding planned improvements, the program could avoid having an N-M breach on either subprogram. This is not necessarily a bad thing. The oversight challenge would be to align operational test criteria with the phased capabilities to be produced.

Unplanned Changes

It is not uncommon for systems already in production to incorporate significant design changes that were not foreseen by the program. Reasons for this can include urgent operational needs from the field, correction of defects discovered post-fielding, implementation of Value Engineering proposals, or response to changes in the adversary/threat environment.

It is clearly unreasonable to require programs to report things they are not yet planning to do. For unplanned changes, the challenge is how to report them as they are discovered and after the fact, in ways that transparently describe the reasons for any corresponding cost and schedule changes.

It would be ideal if SAR reporting of unplanned changes distinguished clearly between design changes driven by new performance requirements and changes required to meet the original program requirements. One possible way to accomplish this would be to add a new category, “Requirements,” to the list of SAR variance categories. Cost changes due to design changes required to meet original program requirements (as of the current APB) would be classified as engineering variances. Cost changes due to new or modified performance requirements would be classified as requirements changes. For a program with a P3I subprogram, the base program and P3I subprogram would report separate cost variances, using the new category where appropriate.

Unfortunately, it is unlikely that programs would report these categories accurately. Not only are there strong incentives to categorize all cost growth as being due to new requirements, but there is often genuine confusion within the program office about which requirements are part of the baseline and which have been added during the course of development and production. In theory, the Cost Analysis Requirements Document and other mandatory acquisition documents establish the baseline requirements assumed by the baseline cost estimate. In practice, this is not as clear, especially for programs that have been re-baselined at some point.

Block Upgrades, Evolutionary Acquisition, and Agile Development

Some programs know in advance that they intend to upgrade or replace the initial design with an improved future design, but do not yet know what those changes will be or what they will cost. They may not know which attributes will be enhanced, since that decision will be based on developments in the future. If multiple changes are made to the weapon system design at a few discrete points in time, these are often termed *block upgrades*. If many changes are made on an ongoing basis as their usefulness becomes known, this is sometimes referred to as *evolutionary acquisition*. In the special case of software programs doing repeated rapid insertion of new features in close collaboration with the users of the software, it is called *agile development*.

In each of these cases, the reporting challenge is that the program knows that they intend to spend money in the future, but they do not know what they will be spending it on, what it will cost, or when it will happen. The challenges for oversight and management are obvious—especially when a program being managed in this way is shoehorned into a reporting system designed for unchanging units. This is part of what happened to the RQ 4B Global Hawk program, which was intended from the beginning as an evolutionary acquisition but was required to guess both content and schedule of future upgrades as part of its original acquisition baseline. Those guesses were then treated as firm requirements by the acquisition system, even after Air Force leaders had changed their minds about both priorities and threshold performance.

In the case of block upgrades, one possibility is to simply declare a new program for each block. This is the approach taken by the AIM-9, AIM-9X, AIM-9X Block II missile programs; the F/A-18C/D and F/A-18E/F fighter aircraft programs; and the UH-60L and UH-60M Blackhawk helicopter programs (among many others).

Other programs have treated successive blocks as distinct official subprograms. This approach was taken by the Joint Air-to-Surface Standoff Missile (JASSM) program. The original program had no subprograms and developed the AGM-158 missile. During that development, the Air Force studied possible improvements to the missile and decided to develop a second variant with longer range. The original AGM-158 was redesignated AGM-158A, and the new "JASSM-ER" (Extended Range) was designated AGM-158B. The program was split into two subprograms for reporting purposes, with JASSM-ER schedule, development costs, and production costs (and cost variances) reported separately. The Navy went even further with the new AGM-158C (LRASM) variant, deciding to make it a distinct program[5] rather than creating a new subprogram within the JASSM program. This may be because the new program is Navy-only, while JASSM is an Air Force program.[6]

An advantage of these approaches is that they isolate the unit cost of the new block from the past, rather than computing an average over all past blocks. It would defeat the purpose of the N-M legislation if 50% APUC growth in what is essentially a new weapon system became invisible because it was being averaged together with thousands of past units of completely different design.[7] A second advantage is that the block upgrade is clearly identifiable as design changes to meet new requirements, as opposed to design changes to overcome technical difficulties in achieving the original requirement.

One disadvantage of the subprogram approach, as currently implemented, is that an N-M breach by any block triggers a mandatory review of every subprogram, as is discussed later.

---

[5] PNO 449, "Offensive Anti-Surface Warfare Increment 1 (Long Range Anti-Ship Missile)," abbreviated as "OASuW Inc 1 (LRASM)"

[6] The "J" in JASSM stands for "joint," and at one point there was consideration of mounting this weapon on Navy aircraft. However, that has not happened, and all of the funds in the SAR are reported from Air Force appropriations.

[7] This is what has happened with the AIM-120 AMRAAM program, as described in the Unplanned Changes section.

A disadvantage of both subprograms and separate programs is the difficulty of accounting for shared RDT&E, nonrecurring, and support costs, such as for testing equipment or software that is used by multiple blocks. For example, the RQ-4B Global Hawk family all uses a common ground station. If this program had used separate subprograms for each distinct aircraft design, it would be inappropriate for the original RQ 4A subprogram to bear the cost of all upgrades to the ground station systems and software, given that all blocks benefit from those upgrades.

A logical response to this problem would be for the Global Hawk program to make the ground station systems a separate subprogram. The difficulty with this is that it would create the possibility of an N-M breach due to cost growth in a subprogram that accounts for only a small fraction of total program cost. A more reasonable approach would be for programs to be able to declare a single subprogram responsible for procurement of items other than end items. This subprogram would only be liable for an N-M breach if its estimated total cost (RDT&E + Procurement) grew to exceed a threshold percentage of the estimated PAUC for the overall program, which would require new legislation from the Congress.

Possible Methods for Handling Mixed Types

As the examples in Mixed Types show, many solutions have been found to the mixed type problem, but all of them have drawbacks.

Subprograms

For some programs, subprograms have provided an elegant solution. For example, the Army's original Multiple Launch Rocket System (MLRS) program distinguished two subprograms: the mobile rocket launcher and the tactical rocket it would fire. This allowed the program to accurately track unit cost growth for both of the fully configured end items being developed and produced. The launcher was produced within its original cost estimate; the rocket experienced a critical N-M breach.[8] Similarly, the Army's PAC-3 suite of upgrades to the Patriot missile system was (after several schedule breaches in the first few years of development) divided into subprograms for the Missile Segment and the Fire Unit.

The fact that a unit cost breach in any subprogram triggers a breach in the whole program discourages their use, even where it seems like an obvious solution. A program without subprograms often has more leeway to do things that will make the cost growth look smaller. For example, if the MLRS program had not defined subprograms, but had treated the rockets as the end-item units, they would have shown a lower percentage cost growth for the combined program than was seen for just the rocket subprogram. In addition, the program could have decided to produce fewer launchers than originally planned, reducing both PAUC and APUC without changing the official number of *units* being produced. Doing so might have avoided the N-M breach, at the cost of greatly reduced transparency regarding cost growth and reduced capability.

Making subprograms more appealing would require congressional action, possibly in an annual authorization bill, which seems possible if some way to maintain program cost accountability could be devised. The Congress might be willing to allow the Milestone

---

[8] Unfortunately, the program did not similarly distinguish the variant rockets being produced or the later conversion of the entire system from an unguided rocket launcher to a guided missile launcher.

Decision Authority to designate alternative triggers for programs with subprograms, especially if some of the subprograms involve far fewer dollars than others.

In theory, SAR reporting could be expanded so that each program could report simultaneously on multiple distinct end items without declaring subprograms. The principal distinction between this approach and subprograms would (presumably) be the mechanisms for deciding cost and schedule breaches. As with subprograms, it would be important in implementing this change to avoid creating perverse incentives to PMs. In particular, accurately defining multiple end items should not increase a program's chances of experiencing an N-M breach.

Multiple Programs

If a Service is planning to buy a mix of different end items in response to a given set of mission needs, they have some flexibility in deciding how to group those efforts into programs. It is not always obvious which grouping would best serve the needs of both the Service and the oversight community.

At one (unfortunate) extreme, the Army decided to make Future Combat Systems a single program with literally hundreds of different physical products ranging in size and complexity from light tanks down to man-portable UASs, along with many tens of millions of lines of software implementing communications, mission command, and networked fires. The official units for that program were Brigade Sets, of which 15 were to be produced. A prime "lead systems integrator" contract was awarded, with authority to reconsider the mix and capabilities of systems to be developed and procured in each Brigade Set. This offered no useful insight into the program's activities or progress.

At the other extreme, the Army decided to split procurement of their new AH-64E Apache helicopters into two separate programs, one for remanufactured aircraft and the other for new builds. A 2008 acquisition decision memorandum signed by the Army Acquisition Executive contains the following language.

> As a recently delegated Acquisition Category IC program, the AH-64E Apache program is comprised of two separate programs, the Remanufacture program and the New Build program. Each of these programs are separate and distinct with respect to the Acquisition Program Baselines (APB), and their funding lines; however, they have identical configurations and are produced on the same production line. (Shyu, 2013)

The choice to create two MDAPs creates challenges for both the Army and OSD because it adds extra reviews and recordkeeping. Having multiple programs, as with subprograms, creates two triggers for an N-M breach, but it also means that any breach would affect only one of the two programs, whereas creating two subprograms would expose the entire program. It also splits what naturally feels like one program—indeed, the previous language refers to it both as one program and as two in the same paragraph. Since both programs produce identical new AH-64E helicopters, why should they be separated? Although distinct for reporting purposes, they have common goals and management. They share a PM and a production contract,[9] but only the remanufacture program reports any RDT&E costs. Even within Apache, both programs list "Other Support" funds in their SARs, and since the two programs are producing identical helicopters, it is unclear how the Army decides whether a given support purchase will be credited to one program or the other. One cannot understand what is going on in either program without considering the other, which would seem to violate the notion of what constitutes a program. Where there is only one distinct end item, having multiple programs is questionable.

Multiple programs should only be considered as an option in the case of block upgrades to an existing program (as discussed in the section titled Block Upgrades, Evolutionary Acquisition, and Agile Development), or when the set of things to be procured by a proposed new program involves the following:

- Significantly different product types with different acquisition risks,
- Multiple independent contracts with no real synergies, and
- Few significant interoperability requirements among systems.

In general, it is rarely appropriate to split a new proposal into multiple programs.

An example of a program that perhaps should have been split into multiple programs is the Stryker (originally "Interim Armored Vehicle") program. This program procured eight specialized variants of an existing non-developmental armored vehicle. Of these eight variants, six were relatively minor modifications of the existing design, while two[10] required extensive engineering changes to the original. An appropriate program management strategy would have been to make the six "minor modification" variants a single program (with six subprograms), and the two major redesigns either a second program with two subprograms, or two additional separate programs. That would have isolated the development risks of the two most risky projects from the more straightforward projects and would have given the OSD and the Congress better visibility of how the various projects were progressing. As it happened, the Stryker program experienced a significant (but not critical) N-M breach, driven entirely by problems in the two major redesign vehicles.

Different Cost Categories

Using the different cost categories in current SAR reporting can give some visibility into what is happening in a program, but generally does not allow better identification of different unit types. The distinction between *end items* and *non-end items* was not designed to capture differences among multiple distinct end items.

The Air Force's UAS MQ-9 Reaper program plans to procure 347 units, where each unit is an aircraft. The total procurement cost for the program is $9.2 billion in BY 2008 dollars, but only 52% of that is EIRF. Another 22% is categorized as NEIRF, and the remaining 26% is Total Support. This information is useful for cost analysts, although this

[9] The December 2015 SAR for the remanufacture program lists four procurement contracts and two RDT&E contracts. The new build SAR only shows one contract, which is one of the four procurement contracts in the remanufacture program.

[10] The two were the M1128 Mobile Gun System (MGS), which attempted to mount a tank-like 105 mm direct fire cannon on a relatively light wheeled vehicle; and the M1135 Nuclear, Biological, Chemical, Reconnaissance Vehicle (NBC RV), which required a suite of sophisticated environmental sensors and a positive-overpressure internal environment.

distribution has no impact on N-M reporting.[11] The aircraft quantity can be compared to the EIRF to understand those units, but there are no quantities reported for ground stations, so an analyst can only know what has been spent on them in total, not what each costs. In this case, NEIRF is something like a subprogram for the ground stations, but it is less transparent than actual subprograms would be.

Reducing Accidents

When humans carry out activities, accidents are inevitable. Reducing accidents requires good processes. We have not analyzed the process for generating SARs or PB submissions. In principle, that could (and perhaps should) be done from a quality assurance point of view.

We were also told that AT&L/Acquisition Resource and Analysis (ARA) performs the OSD's checks on Service-submitted data, and they do not have enough time to do it thoroughly. All of the draft SARs arrive at the OSD in the same season. About a week after the data arrive, ARA meets with each program for about one hour, at which time ARA can ask questions. They feel that this process is insufficient and clearly there are changes that could reduce the accident rate.

The best way to improve ARA's review is probably not only to add more time. While more time might help, ARA would probably also benefit from specialized tools to help them analyze the draft SAR data and quickly compare them to budget submissions, prior year SARs, and general rules about how acquisition programs typically behave. Proposing improvements to that process is beyond the scope of this report.

A Thought Experiment: JLTV

To illustrate the kind of reporting that would be necessary to improve both oversight and data utility for cost analysts, we looked at the JLTV program. The full analysis is in our full report (Davis et al., 2017). We determined that at the beginning of the program, seven subprograms might be appropriate, as indicated in Table 2.

Table 2. Suggested Initial JLTV Subprograms

Number	Subprogram	Description
1	Utility	Base design Utility Vehicle
2	General Purpose (GPV)	Base design GPV
3	Heavy Guns Carrier (HGC)	Base design HGC
4	Close Combat Weapons Carrier (CCWC)	Base design CCWC
5	P3I – Common	P3I common to all variants
6	P3I – HGC and CCWC	P3I specific to HGC and CCWC
7	Support equipment	Trailers, armor kits, etc.

This would not be practical if an N-M breach could be triggered by any one of them.

[11] One could imagine the Air Force lowering the ratio of ground stations to aircraft, not for operational reasons but rather because they want to control APUC.

Conclusions

The default assumption for any acquisition program is that all of the units it produces are identical and interchangeable. This is seldom true—consider asking an F 35A to land on a ship. Any analysis that assumes interchangeable units is making an unwarranted assumption that can lead to mistaken conclusions. The importance of these mistakes will vary, both with the details of the program and the nature of the analysis. We hope that this work can lead to two kinds of changes: one for analysts using acquisition data, and a second for policy makers defining reporting requirements for programs.

For analysts, the primary message is "Beware." It is not uncommon for invisible differences between units to be important to an analysis, as we saw with previous IDA studies of hedonic price indices for aircraft and tactical vehicles discussed in the Changes Over Time chapter. Without additional data from non-SAR (and sometimes non-PB) sources, it is often impossible to understand the relationships between price, cost, and quantity in many programs. Such additional data are, unfortunately, not always available. Analysts need to know the limits of what can be inferred from the existing data.

For policy makers, there are many opportunities to improve data reporting requirements and guidance, and these come in three varieties. First, there ought to be explicit acknowledgment that not all units are identical, and some effort should be made to quantify unit-by-unit or lot-by-lot differences for analysis and oversight. Second, the rules need to encourage the desired behaviors. The current N-M rules are an excellent example of how rules incentivize behavior in ways that may be counterproductive. For example, IDECM's unit costs could be reduced by purchasing more towed decoys. When designing new reporting requirements, policy makers need to keep this in mind. Finally, the quality assurance processes applied to official data ought to be studied and improved. While some accidents are inevitable, the system today probably lets through more than it should. SARs are much like custom manufactured parts. Each one is unique, but good processes could still make them more uniform and useful.

References

10 U.S.C. § 2432. Selected Acquisition Reports. Retrieved March 28, 2017, from http://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title10-section2432&num=0&edition=prelim

10 U.S.C. § 2433. Nunn-McCurdy Act. Amendment to the FY 1982 National Defense Authorization Act. Retrieved from https://www.congress.gov/bill/97th-congress/senate-bill/815

Davis, G. A., Giles, M. L., & Tate, D. M. (2017). *Complexity in an unexpected place: Quantities in selected acquisition reports* (IDA Paper P-8490). Alexandria, VA: Institute for Defense Analyses.

Defense Acquisition University. (n.d.) *Glossary of defense acquisition acronyms and terms*. Defense Acquisition Management Information Retrieval (DAMIR) System. Retrieved from https://dap.dau.mil/glossary/pages/1742.aspx

DoD. (1992). *Cost analysis guidance and procedures* (DoD 5000.4 M). Washington, DC: Assistant Secretary of Defense (Program Analysis and Evaluation). Retrieved from https://biotech.law.lsu.edu/blaw/dodd/corres/pdf/50004m_1292/p50004m.pdf

DoD. (2016). *Department of Defense selected acquisition reports* (Release No. NR-106-16). Retrieved from https://www.defense.gov/News/News-Releases/News-Release-View/Article/703529/department-of-defense-selected-acquisition-reports-sars

Drezner, J. A., Arena, M. V., McKernan, M., Murphy, R., & Riposo, J. (2011). *Are ships different? Policies and procedures for the acquisition of ship programs.* Santa Monica, CA: RAND. Retrieved from https://www.rand.org/pubs/monographs/MG991.html

Feickert, A. (2017). *Joint Light Tactical Vehicle (JLTV): Background and issues for Congress* (CRS Report RS22942). Washington, DC: Congressional Research Service. Retrieved from https://fas.org/sgp/crs/weapons/RS22942.pdf

Johnson, H. T. (2007). Lean dilemma: Choose system principles or management accounting controls—not both. In J. Stenzel (Ed.), *Lean accounting: Best practices for sustainable integration*. New York, NY: John Wiley & Sons.

McNicol, D. L. (2004). *Cost growth in major weapon procurement programs* (IDA Paper P-3832). Alexandria, VA: Institute for Defense Analyses.

Rogerson, W. P. (1991). Excess capacity in weapons production: An empirical analysis. *Defence Economics, 2*(3), 235–249. doi:10.1080/10430719108404695

Shyu, H. (2013). *Acquisition decision memorandum for the AH-64H Apache new build program* [Memorandum for Program Executive Officer, Aviation]. Washington, DC: Office of the Assistant Secretary of the Army.

Teal Group Corporation. (2014). *World missile briefing*.

Acknowledgments

Thank you first to Dr. Peter Eggan of PARCA for supporting us throughout this project, even as we shifted gears so significantly. We also want to express our gratitude to Kathy Conley, Strider McGregor-Dorsey, and Brian Rieksts, all of IDA, whose technical reviews kept us moving (and complaining), leading to a better final product.

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CA 93943

www.acquisitionresearch.net