SYM-AM-24-115

Excerpt from the Proceedings of the Twenty-First Annual Acquisition Research Symposium

Acquisition Research: Creating Synergy for Informed Change

May 8–9, 2024

Published: April 30, 2024

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

The research presented in this report was supported by the Acquisition Research Program 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).

ACQUISITION RESEARCH PROGRAM
DEPARTMENT OF DEFENSE MANAGEMENT
NAVAL POSTGRADUATE SCHOOL

Improving Comptroller Benchmarks on Program Spending

Philip S. Antón, PhD—is the Chief Scientist of the Acquisition Innovation Research Center (AIRC) in the Systems Engineering Research Center (SERC) led by the Stevens Institute of Technology. He assesses DoD needs, envisions and develops innovative research, and facilitates application to DoD practices, policies, and workforce development. For 23 years, Dr. Antón was a Senior Information Scientist at the RAND Corporation. From 2011–2016, Dr. Antón directed Acquisition Policy Analysis for USD(AT&L). His PhD and MS are in Information and Computer Science from UC Irvine, specializing in Computational Neuroscience and Artificial Intelligence. His BS in engineering is from UCLA. [panton@stevens.edu]

Odd J. Stalebrink, PhD—is an Associate Professor of Public Administration at the School of Public Affairs at the Pennsylvania State University. He specializes in public budgeting and financial management. He is a member in the Association for Budgeting and Financial Management (ABFM); the National Tax Association (NTA); the Association for Public Policy Analysis and Management (APPAM); and the European Accounting Association (EAA). He holds a PhD in public policy from the School of Public Policy, George Mason University. His master's and bachelor's degrees are in Business Administration from the Jönköping International Business School (JIBS), Sweden. [ojs10@psu.edu]

Douglas J. Buettner, PhD—is the AIRC Deputy Chief Scientist at the Stevens Institute of Technology and is an Adjunct Assistant Professor for the University of Utah's Systems Engineering Program. His doctoral research at the University of Southern California's Viterbi School of Engineering examined issues with DoD software-acquisition strategies, identifying acquisition optimizations. His research is found in Acquisition Games and Software Defects: Dynamics of the Software Developer's Dilemma. Dr. Buettner has BS and MS degrees in physics, with a minor in atmospheric science from Oregon State University, and his PhD from USC is in astronautical engineering. [dbuettne@stevens.edu]

Abstract

A common tool for overseeing program execution is to compare spending against linear benchmarks to identify programs that may be falling behind or unable to fully use their funding. These benchmarks identify candidates for further investigation and potential budgets reduced and reallocated. Pressures to meet benchmarks can drive bad behaviors, such as premature spending before good prices and intellectual property rights can be negotiated. This paper analyses business theory, program manager observations, and historical trends of DoD obligation and expenditure rates to assess ways to improve these benchmarks. Regressions of historical obligation data find that recent spending has an underlying linear trend, but temporal variables, theory, and execution realities indicate that S-shaped curves are better benchmarks. Also, benchmarks should be adjusted when Congress provides Continuing Resolutions (CRs) in lieu of full appropriations at the start of the fiscal year. Also, as expected by theory, historical expenditure patterns for Research, Development, Test, and Evaluation (RDT&E), Procurement, and Operation and Maintenance (O&M) funds follow S-shaped curves rather than the linear profiles in DoD benchmarks. Recommendations are provided, including adjustments for variable effects on obligations, S-curve profiles for improved benchmarks, and leveraging improved DoD data environments to switch to plan-based benchmarking.

Introduction

What gets measured gets managed – even when it's pointless to measure and manage it, and even if it harms the purpose of the organization to do so.

Peter Drucker

A common management tool for overseeing program execution is to track spending overtime against benchmarks to monitor progress and identify any programs that may be falling behind. In the federal government, program and contract spending involve two basic steps:

- *Obligations*, which commit funds from the U.S. Treasury for payment of goods and services, such as on a contract (see GAO, 2005, p. 70).
- *Expenditures* (also called *outlays* or *disbursements*), which are the actual financial payments (money) from the U.S. Treasury to liquidate an obligation, for example upon receipt of goods or services under a contract (see GAO, 2005, pp. 73–74).

The Benefits and Dangers of Benchmarks and Metrics

Program managers should attend to items that are important enough to measure. The problem is that we get exactly what we measure when we enforce and incentivize the metric. People will spend—one way or another, and often regardless of unintended side effects—if we measure execution against a metric and especially if we apply enforcement consequences and incentives. See, for example, anecdotal and survey evidence in Marsalis (2022) and Commission on Planning, Programming, Budgeting, and Execution (PPBE) Reform (2023).

Below is a short overview of the principles, theory, and realities of setting, using, and enforcing benchmarks and metrics. This is just a short overview, but it is important to begin by reminding ourselves what these metrics result in—good and bad.

Benefits of Benchmarks

Benchmarks can help identify performance issues. Some programs do not obligate or expend all their authorized and appropriated funding. Thus, monitoring the level of funding can be a quick way to identify programs that eventually may not fully execute their spending. With the recognition that such benchmarks are but one source of needed information, they can help focus attention on more likely candidates.

Monitored benchmarks can ensure that attention is paid to managing financial resources. When not emphasized to the extreme, monitored benchmarks like these can motivate program managers to properly plan and track spending along with associated risks.

Concerns on Using Benchmark

Untailored benchmarks may not align with program realities and plans. Programs have spending needs based on plans as well as events and decisions that need to be made during the spending period. Some programs may plan to obligate funds as soon as they are authorized and appropriated while others may have good reasons to obligate late in the spending period. Untailored (a priori) fixed benchmarks may be out of sync with such plans. Also, programs may not know in advance when they need to obligate. For example, programs, especially in Research, Development, Test, and Evaluation (RDT&E) (which often involves systems that have never been developed before), may need to address issues that arise during the spending period, address unforeseen costs or schedule challenges, or to change priorities based on new threats or technology issues and opportunities. Again, fixed benchmarks would only reflect such spending profiles from random chance.

Benchmarks can drive undesirable behaviors or effects. It is well known in business and psychology that enforced or even monitored benchmarks will drive behaviors to achieve the benchmarks despite negative consequences (Behn, 2008; Marsalis, 2022; Norden, 1970). This is a real concern in the Department of Defense (DoD) to avoid wasting resources in programs—either by not spending what could be used elsewhere, or by wasting it through less-than-prudent (but entirely understandable from an incentives perspective) spending.

Input (consumption) benchmarks lack prioritization and thus require additional information before acting on below-target programs. Finally, spending benchmarks like

these are simply input metrics (as opposed to output or outcome metrics)[1] that lack measures of the *value* of the spending. A program (and thus the DoD and taxpayers) may get better value from early spending, or it may get higher value by giving a program more time to obligate (or a contractor more time to execute). Instead, programs appear mostly driven to avoid unspent funds (which does happen in non-trivial amounts) and to identify potential resources for new urgent priorities that arise during the spending period. Thus, they can only identify potential candidates for further (deeper) assessment to understand a program's status. Without this added consideration, these metrics can devolve into blind bureaucratic taking of resources with undesirable outcomes.

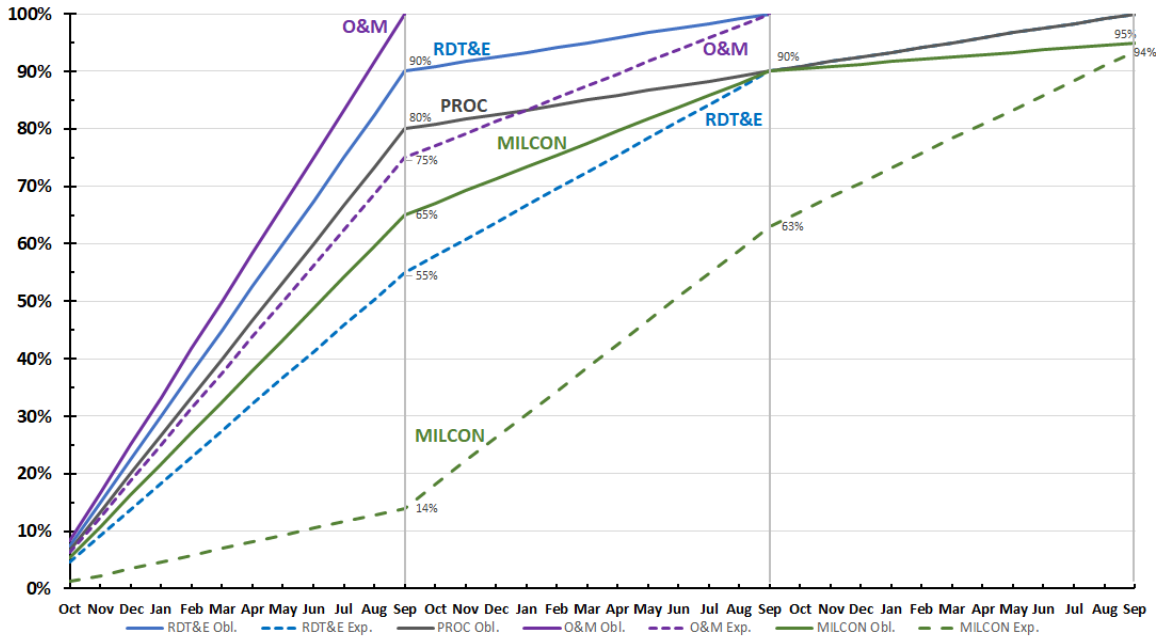
Benchmarks restrict agility at the program level. Taken together, these concerns can reduce administrative flexibilities at the program level, pressuring, and restricting spending decisions within the year(s) of execution. Increased agility in meeting DoD program outcomes requires a willingness to delegate decisions while providing clearly defined goals and objectives. This should be accompanied by appropriate accountability to program managers to fully utilize their appropriated funds or advise leadership early on if they will not be able to fully execute these funds and thus make them available for reprogramming for other purposes. Such a willingness to delegate, decentralize, and utilize administrative flexibilities at the program level could form a base-level of reform for the larger PPBE system, wherein program planning and execution agility is increased. See Stalebrink (forthcoming) for discussion of these concepts across all levels of PPBE.

Obligation and Expenditure Benchmarks in the DoD

The DoD uses linear benchmarks for each category of funding (see Figure 1). Such benchmarks can help identify programs and activities that may have issues in spending funds within the year(s) of availability and thus may be candidates for further review to have portions of their budgets reprogrammed for critical priorities that emerged in the year of execution.

[1] See, for example, National Research Council (2005) for a very useful review of the theory and application of different types of metrics to achieve desired performance and outcomes.

Figure 1. Current Comptroller Obligations and Expenditures Rule-of-Thumb Benchmarks



SOURCE: Under Secretary of Defense (Comptroller), as reported in Tomasini (2017).

NOTES: The dashed lines are the obligation (Obl.) benchmarks over time, and the solid lines are the associated expenditure (Exp.) benchmarks over time. The O&M benchmark curves rise the fastest, followed by RDT&E and Procurement (PROC). Tomasini (2017) reports that Procurement expenditures are "N/A." Exp. = expenditures; MILCON = Military Construction; O&M = Operation and Maintenance; Obl. = obligations; PROC = Procurement; RDT&E = Research, Development, Test, and Evaluation.

This paper assesses these benchmarks through quantitative analysis of DoD obligation and expenditures over time, observations from program managers, and a review of existing theory and qualitative data from experts. It also assesses the statistical effects of delayed full fiscal year (FY) appropriations associated with continuing resolutions (CRs), calendar-month effects (e.g., at the start and end of the FY), and time trends on DoD obligation rates. These analyses provide new insights into the realism of DoD obligation and expenditure benchmarks, leading to recommendations for improving these benchmarks.

Analysis of Obligation Rates: Effects of Continuing Resolutions and Other Events

We obtained data from the DoD's Advana data environment on monthly obligated dollars for separate accounts (e.g., within military services or defense wide) and categories of funding (RDT&E, Procurement, O&M, MILPERS, and MILCON) going back to FY 2011. The data also included a range of other categories, such as MILCON and smaller accounts that are not analyzed in this paper.

We aggregated these data to obtain monthly obligation dollars by category across all accounts, then calculated the percentage obligated in each month compared to the total dollars obligated by the final month. For example, if the RDT&E obligated in month 2 was $5,112,653, and the total obligated by month 24 was $71,339,247, then the month 2 percentage is about 7.17% (= $5,112,653 / $71,339,247). This yielded a series of monthly obligation rates (percentages) for each FY's authorization and appropriation out to the end of those obligations.

Using these data from Advana, we then conducted multivariate linear regressions on individual obligation categories (colors of money) to identify any variables that correlate with changes in the normal monthly obligation rates. Table 1 lists the variables tested for correlation (i.e., with a p-value no higher than 0.05). Visual examination of the monthly obligation rates

indicated that the first year of obligation behaved differently than any subsequent years, so for multi-year appropriations (RDT&E and PROC) we conducted separate regressions for the first and subsequent years.

Table 2 summarizes the statistical results with the following observations.[2] The data are well modeled by a linear obligation rate (the constant base) with adjustments for the variables shown. For example, on average, the RDT&E rate in October of the first year of obligating a FY's appropriation was 5.9% − 2.3% = 3.6%. If March of the first spending year was also the third month after the budget was passed (3 MAB), then on average the obligation rate would be about 5.9% + 6.3% + 2.3% = 14.5% (which is close to the actual value of 15.1% for FY 2012, for example).

Table 1. Variables Tested for Effects on Monthly Obligation Rates

Type	Variables
FY calendar month	October, November, December, January, February, March, April, May, June, July, August, September
Month after full budget passed	1 MAB, 2 MAB, 3 MAB, 4 MAB, 5 MAB
Time (month #)	Time

[2] Statistically, we note the following:

- While all have descent Adjusted R^2 values, the values for RDT&E (1st year), PROC, and O&M are the highest. Thus, the latter explain the variation in the data well.
- The models (the constant monthly linear contribution plus the contributions from the variables in the model) are fairly linear, with the RDT&E 1st-year for RDT&E and PROC (respectively) and O&M being very linear.

Table 2. Contributions of CR and Other Variables Affecting Obligation Rates (FY 2011–2023 Appropriations)

			RDT&E		PROC		O&M	MILPERS
			1st Year	2nd Year	1st Year	2nd–3rd Years		
Average Base	**Monthly Rate**	*Base rate:*	5.9%	1.7%	5.0%	2.0%	7.5%	7.9%
CR Effects	**1st MAB**	*If true, add:*					1.2%	
	2nd MAB	*If true, add:*	4.3%		1.8%		2.4%	
	3rd MAB	*If true, add:*	6.3%		4.6%		2.4%	
	4th MAB	*If true, add:*	3.7%		2.2%		1.6%	
Calendar Month Effects	**October**	*If true, add:*	−2.3%		−3.6%		−1.1%	
	November	*If true, add:*		0.9%		0.6%	−0.8%	
	December	*If true, add:*				0.9%		
	March	*If true, add:*	2.3%		2.6%	0.4%		
	July	*If true, add:*					2.4%	
	August	*If true, add:*					−1.8%	
	September	*If true, add:*		1.0%	4.6%	1.4%	4.1%	0.4%
Time Trend	**Time (month)**	*If true, add:*		−0.1 %		−0.1%		0.1%
Linearity	***Multiple R***		*0.80*	*0.58*	*0.77*	*0.66*	*0.84*	*0.48*
% of variation explained	***Adjusted R^2***		*63%*	*32%*	*57%*	*43%*	*68%*	*22%*

The analysis above examined RDT&E obligations together. Subsequently, we obtained DoD monthly RDT&E obligations data broken down by Budget Activity (BA) for appropriation FYs 2013, 2014, and 2017–2022. Table 3 shows the results of the linear regressions on the RDT&E monthly obligation rates as percentages of appropriation-year dollars associated variables across all of RDT&E (original Advana data) and broken down by S&T, development, and management accounts.

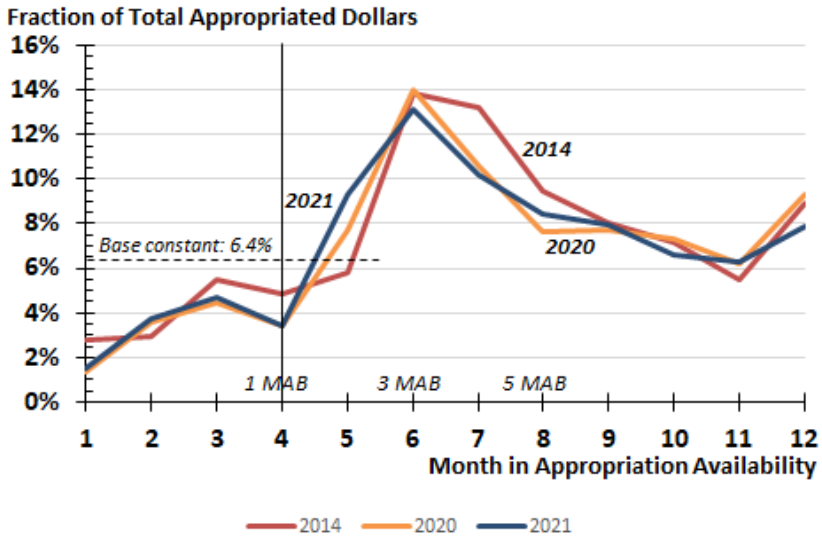
Table 3. Contributions of CR and Other Variables Affecting RDT&E Obligation Rates (FY 2013–2014, 2017–2022 Appropriations)

		1st Year of Availability				2nd Year of Availability			
		All	S&T	DEV	Mgt	All	S&T	DEV	Mgt
Avg. Base		5.9%	6.4%	6.0%	6.3%	1.7%	2.3%	1.6%	2.5%
Add CR Effects	CR		-1.0%		-1.4%				
	1 MAB		-1.6%						
	2 MAB	4.3%		3.0%	2.1%				
	3 MAB	6.3%	4.8%	7.1%	2.1%				
	4 MAB	3.7%	4.3%	4.8%					
	5 MAB		1.5%						
	6 MAB		1.8%						
Add Calendar Month Effects	Oct.	-2.3%	-3.5%	-1.9%				0.40%	-0.63%
	Nov.		-1.1%			0.91%		0.52%	
	Mar.	2.3%	1.6%	1.7%					0.68%
	Sept.		2.4%		2.9%	1.0%	1.3%	0.84%	1.3%
Time Trend	Time (mo.)					-0.10%	-0.15%	-0.10%	-0.10%
Adj. R2		63%	84%	60%	42%	32%	48%	50%	33%

MAB = month after budget is passed; CR = month under a continuing resolution (the months before 1 MAB); BA = Budget Activity; S&T = Science and Technology (BA-1, BA-2, and BA-3 combined); DEV = development (BA-4, BA-5, and BA-7 combined); Mgt = Management [Support] (BA-6); mo. = month; Oct. = October; Nov. = November; Jan. = January; Mar. = March; Aug. = August; Sept. = September.

In our data sample, appropriations from had their first MAB in January. Figure 2 illustrates the CR and month effects for three appropriation years (FYs 2014, 2020, and 2021) for which January was the first month after the budget was passed (the 1 MAB Here, months 1–3, October through December) operated under a CR and thus were below the base constant of 6.4%. The model showed that October was an additional 3.5 percentage points low and November an additional 1.1 percentage points lower than the baseline minus the CR effect. This can be seen in the lower left of the figure. The first MAB is 1.6 percentage points lower in the model, but rises significantly in MABs 3–8, then returning closer to the base constant with a final increase in September. The example shows that the expected cumulative obligations should be about 84% by September; the actuals for FYs 2014, 2020, and 2021 are 88.0%, 83.3%, and 83.1%, respectively (85% on average—close to 84% from the model).

Figure 2. Monthly S&T Obligation Rates for Appropriation FYs with Full Budget Passed in January (FY 2014, 2020, and 2021 Appropriations)



Comparison of Historical Data Against Current and Proposed DoD Obligation and Expenditure Benchmarks

The following figures graphically show the differences between historical obligations and expenditures against current and recommended benchmarks (e.g., with S-curves for RDT&E, Procurement, and O&M along with historically patterned benchmarks).

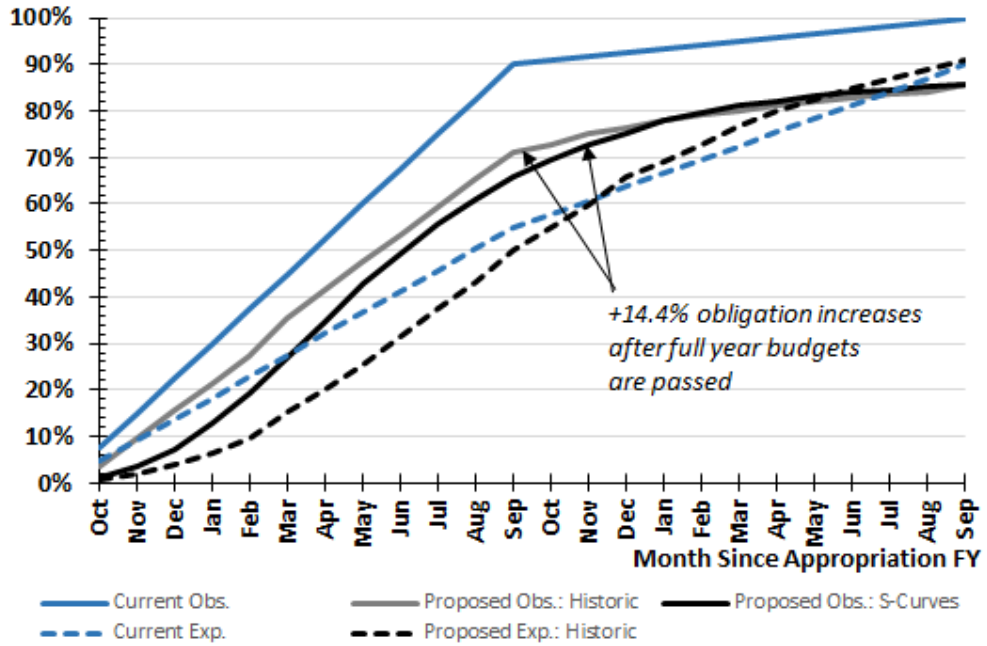
Figure 3 compares the current and recommended RDT&E benchmarks. The recommended obligation curves show the cumulative amounts, not counting the bumps that would be added in after the full FY budgets are passed. Figure 4 illustrate how the MAB obligation effect would add based on what month the full budget is passed. The first figure shows the effect when lowering the initial portion to further strengthen the s-curve effect while the second figure shows the effect based solely on the historical values from FY 2011–2022 actual obligations. These plots illustrate the strong effect on obligations of delayed final appropriations for the FY.

Figure 5 compares the current and recommended Procurement benchmarks. Again, the recommended obligation curves show the cumulative amounts not counting the bumps that would be added in after the full FY budgets are passed. There is no current expenditure benchmark for Procurement. The recommended benchmarks reflect insights from analyzing actual procurement benchmarks. Further analysis is needed to reflect the different lengths of different procurement accounts.

Figure 7 compares the current and recommended O&M benchmarks. Again, the recommended obligation curves show the cumulative amounts, not counting the bumps that would be added in after the full FY budgets are passed.

Finally, Figure 9 compares the current and recommended MILCON benchmarks.

Figure 3. Cumulative Fraction of RDT&E Obligations and Expenditure by Month for Current and Proposed Benchmarks (FY 2011–2022 Appropriations)



NOTE: See Figure 4 for how the 14.4% obligation increases are added depending on when the final FY appropriations are passed.

Figure 4. Current and Proposed S-Curve Benchmarks: Cumulative Fraction of RDT&E Obligations by Month for (FY 2011–2022 Appropriations)

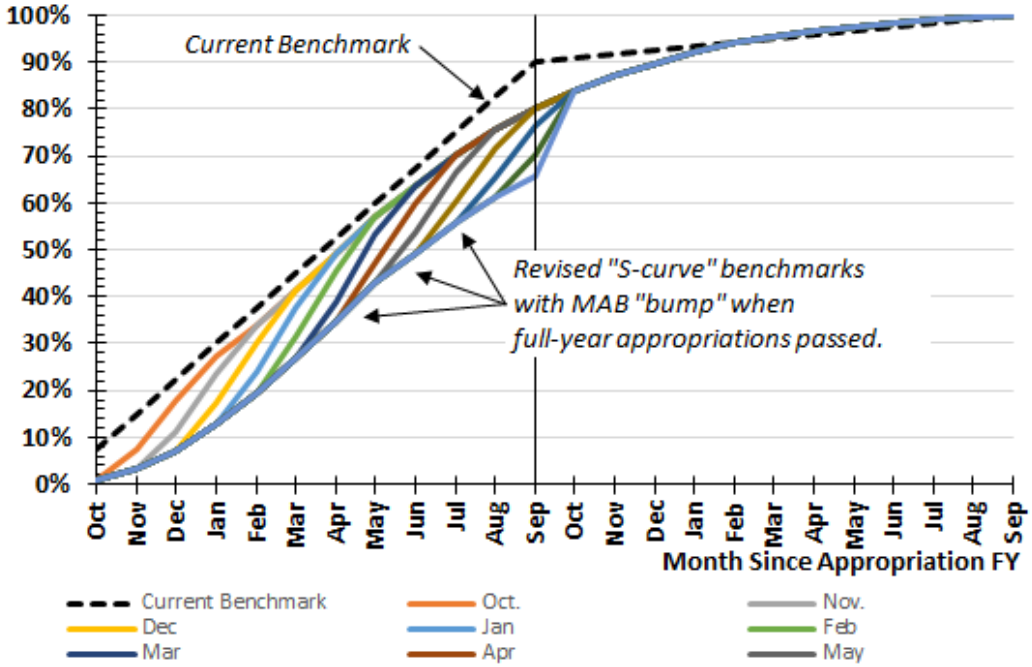
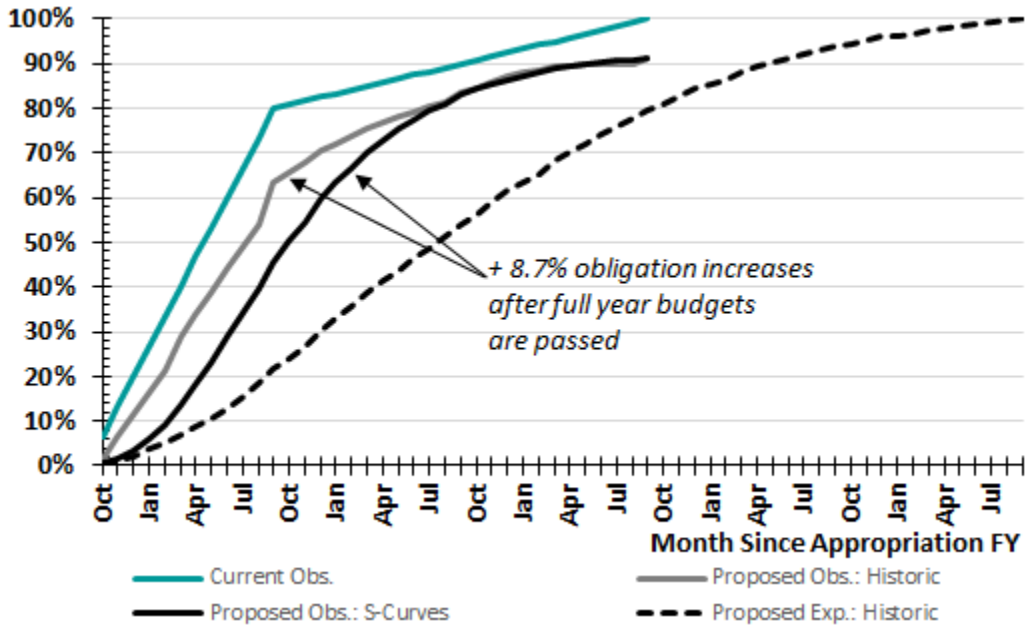


Figure 5. Cumulative Fraction of Procurement Obligations and Expenditure by Month for Current and Proposed Benchmarks (FY 2011–2021 Appropriations)



NOTE: See Figure 6 for how the 8.7% obligation increases are added depending on when the final FY appropriations are passed.

Figure 6. Current and Proposed S-Curve Benchmarks: Cumulative Fraction of PROC Obligations by Month for (FY 2011–2021 Appropriations)

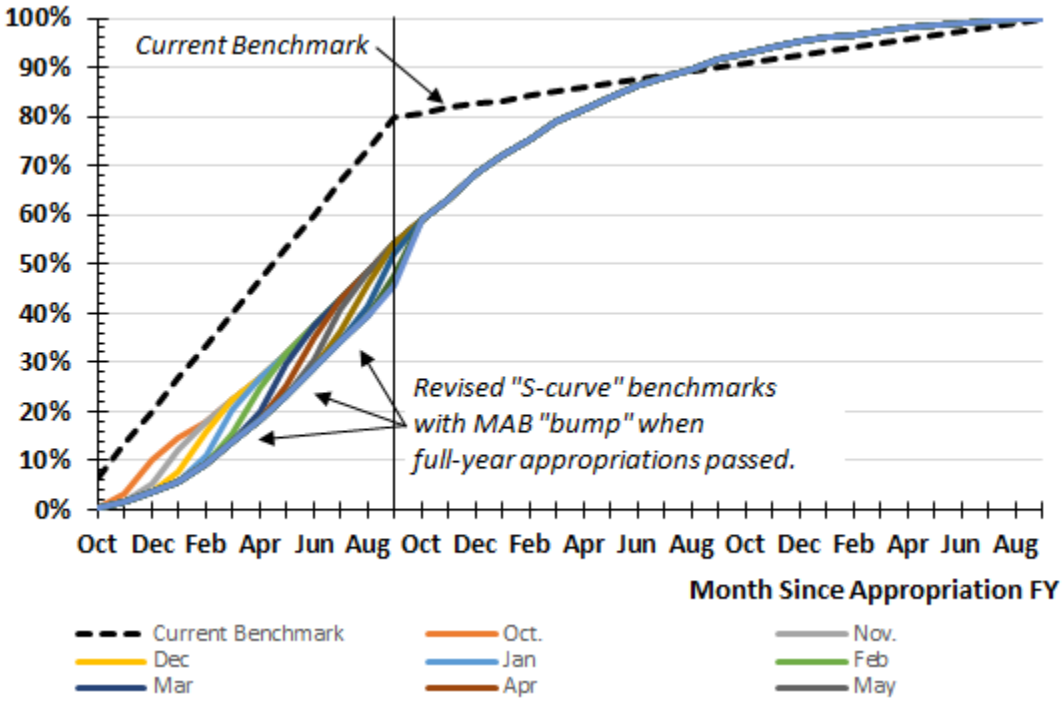
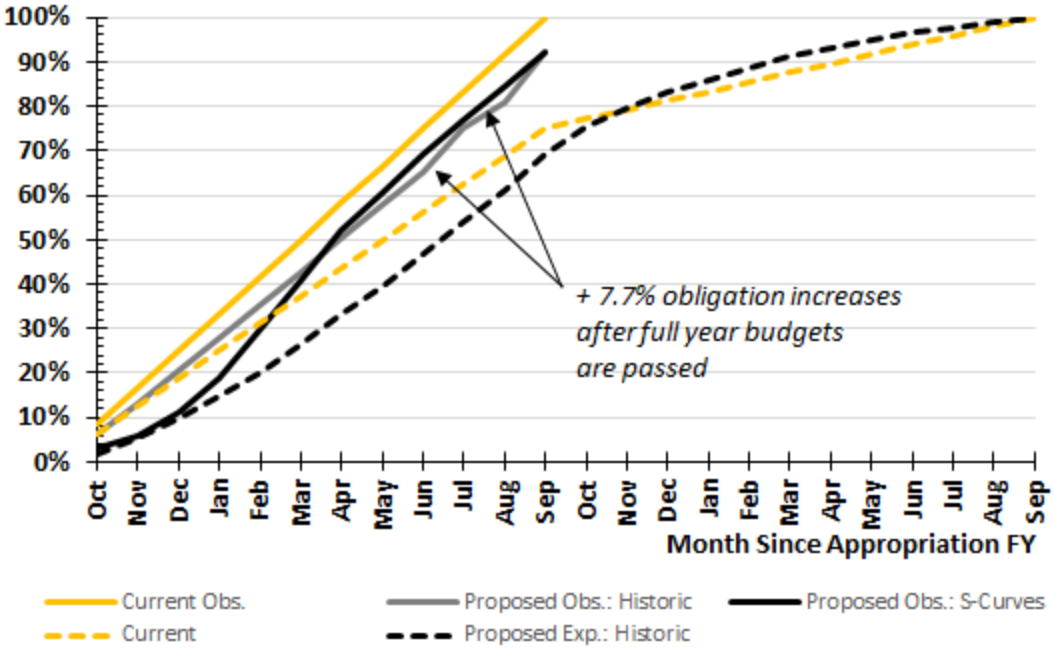


Figure 7. Cumulative Fraction of O&M Obligations and Expenditure by Month for Current and Proposed Benchmarks (FY 2011–2023 Appropriations)



NOTE: See Figure 8 for how the 7.7% obligation increases are added depending on when the final FY appropriations are passed.

Figure 8. Current and Proposed S-Curve Benchmarks: Cumulative Fraction of O&M Obligations by Month for (FY 2011–2023 Appropriations)

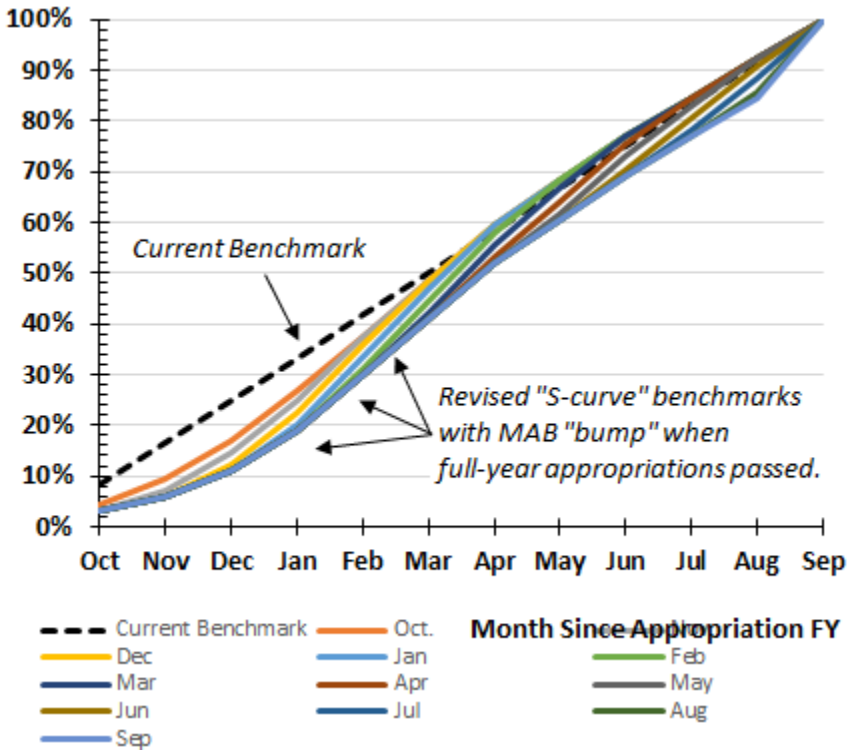
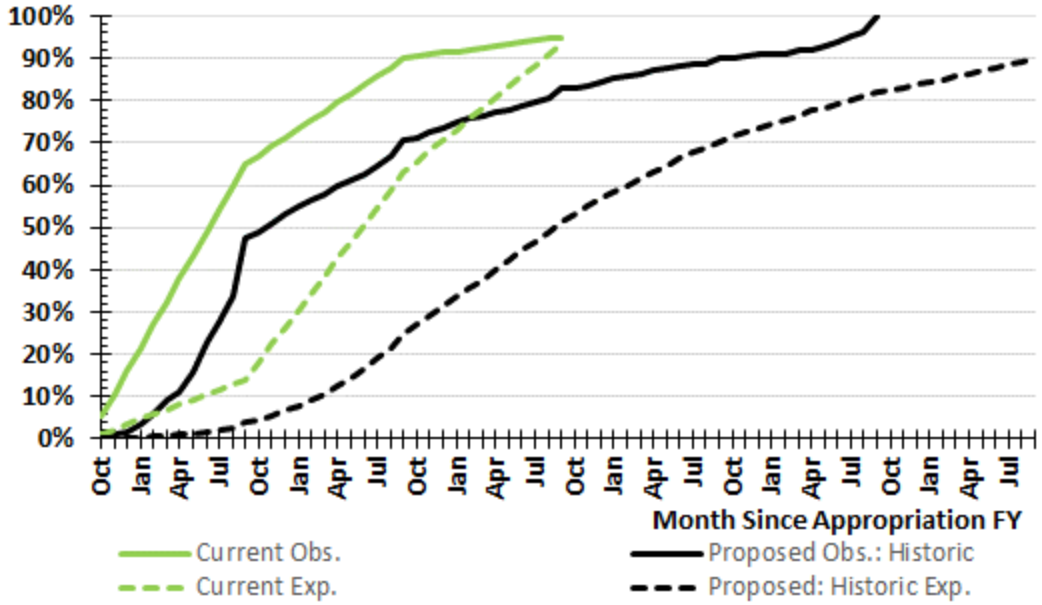


Figure 9. Cumulative Fraction of MILCON Obligations and Expenditure by Month for Current and Proposed Benchmarks (FY 2011–2021 Appropriations)



Observations on Variables Affecting Monthly Obligation Rates

Here are the significant effects uncovered by these analyses:

- Obligation rates are higher in the 2 to 6 months after the full budget is passed (MAB; i.e., once managers know their authorized spending). Thus, CRs delay a portion of funding into later in the FY.
- S&T and Management Support within RDT&E have a significantly lower obligation rate during CR months that other types of funding did not exhibit.
- Obligation rates are often lower the first October in the spending cycle, possibly reflecting assertions in the literature that it takes time to delegate spending authorization to program managers.
- Obligation rates are often higher the first March in the spending cycle (i.e., the month before the midyear spending reviews).
- Obligation rates for some types of funding are higher in September.
- While each category of funding has a general underlying linear trend, MILPERS obligations are linear with slight upward trend.
- RDT&E and Procurement dollars obligate the first year on a fairly linear basis but then inflect to a reduced, curved basis. Thus, obligations are modeled well by linear models with these variate effects.
- Military Construction (MILCON) shows a significant upward curve in the first year rather than the straight line in the benchmark but becomes fairly linear afterwards. Also, a significant fraction of MILCON obligations occurs after year 3, which is not in alignment with the benchmark targets.

These statistical models align somewhat with linear obligation rate targets set by the DoD Comptroller and are compatible with anecdotal assertions that when told to obligate, programs do. This does not account for any changes in DoD priorities given new threats or technological

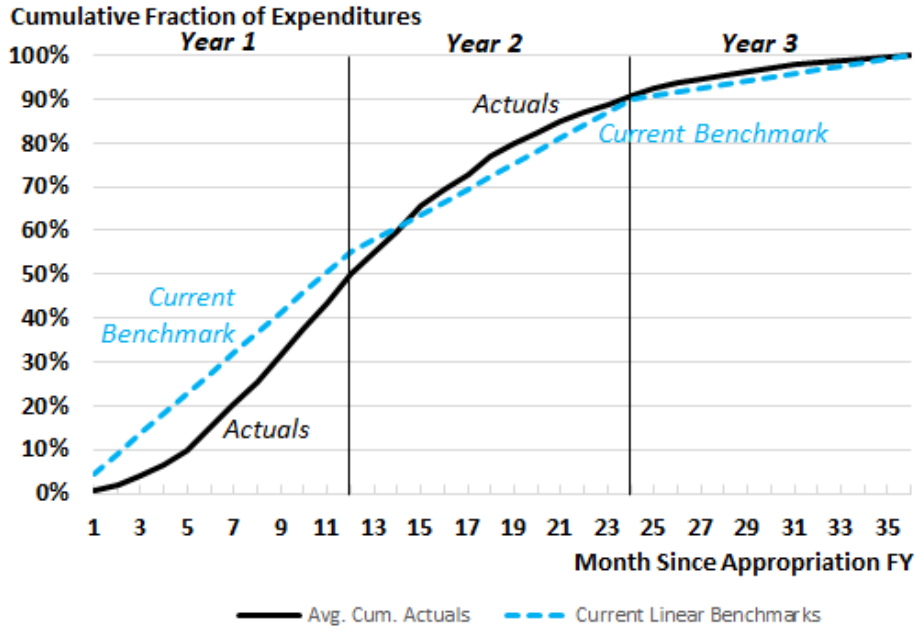
opportunities since the budgets were first drafted early in the PPBE process, but when told to spend or risk losing their funds, individuals across the DoD appear to do so to a large extent.

Expenditure Rates

Analysis of DoD expenditure data shows that RDT&E, Procurement, and O&M expenditures followed an S-curve shape rather than the linear profiles in the DoD's benchmarks. This aligns with over 50 years of data and theory in the literature on program execution profiles.[3]

While the S-curve for RDT&E meets the Comptroller's linear benchmarks at the 24-month point of 90%, but the average 6-month value of 15.5% is well below the benchmark of 27.5% and the 12-month value was also lower than the benchmark (see Figure 10). Thus, the DoD's linear RDT&E benchmark poorly informs the midyear and first-year execution review for RDT&E. Similar profiles were seen for O&M and MILCON (see Figure 11 and Figure 12).

Figure 10. Average Cumulative RDT&E Expenditures Versus Benchmark as Percentage of Month 36 Obligations (FY 2011–2021 Appropriations)



NOTE: Month 1 is October of the FY in which the appropriations were made.

[3] See, for example, Behn, 2008; Burgess et al., 2014; Brown et al., 2015; Davis, 2008; Davis et al., 2009; Gallagher & Lee, 1996; Lee, Hogue, & Gallagher, 1993; Lee, Hogue, & Hoffman, 1993; Norden, 1970; Schiavoni, 2019; Watkins, 1982.

Figure 11. Average Cumulative O&M Expenditures Versus Benchmark as Percentage of Month 12 Obligations (FY 2011–2022 Appropriations)

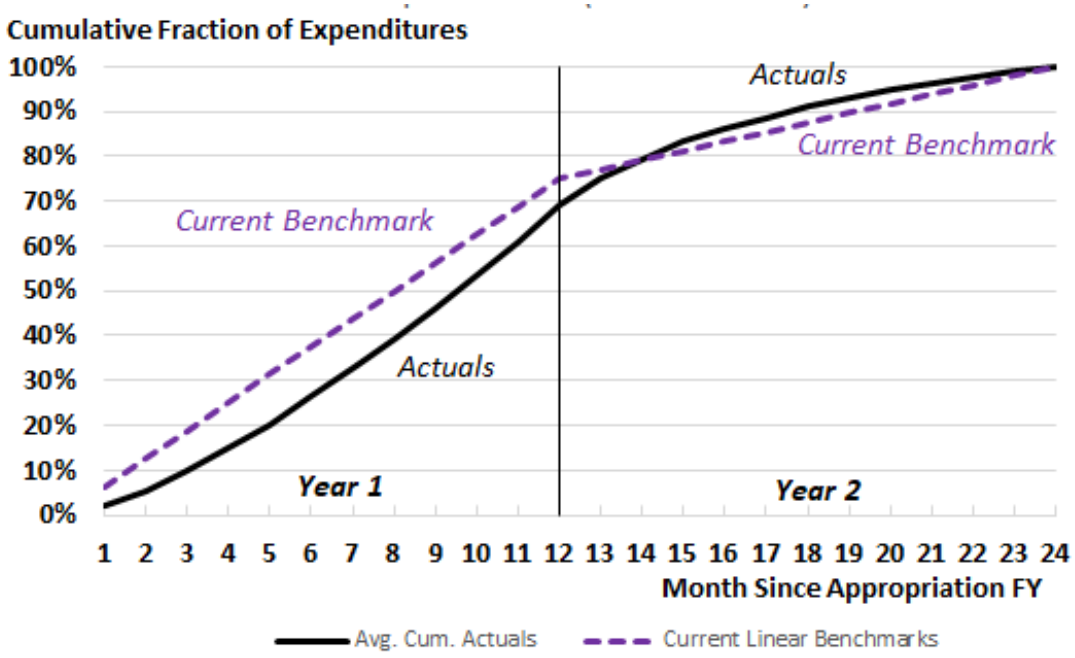
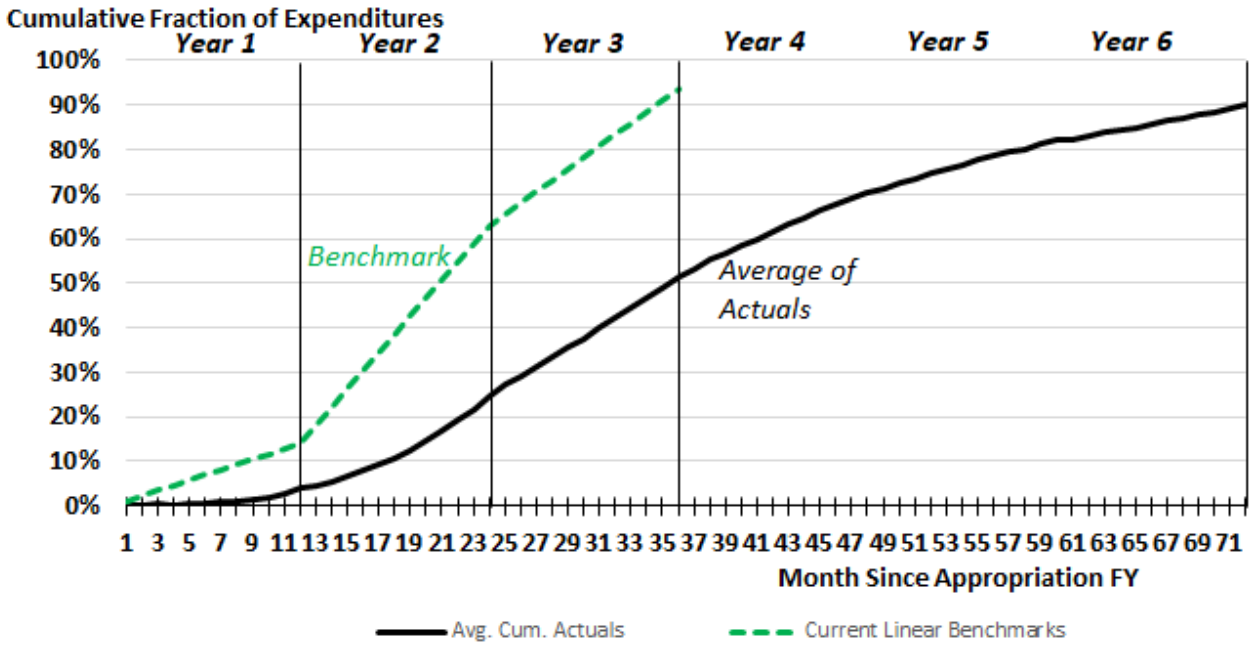
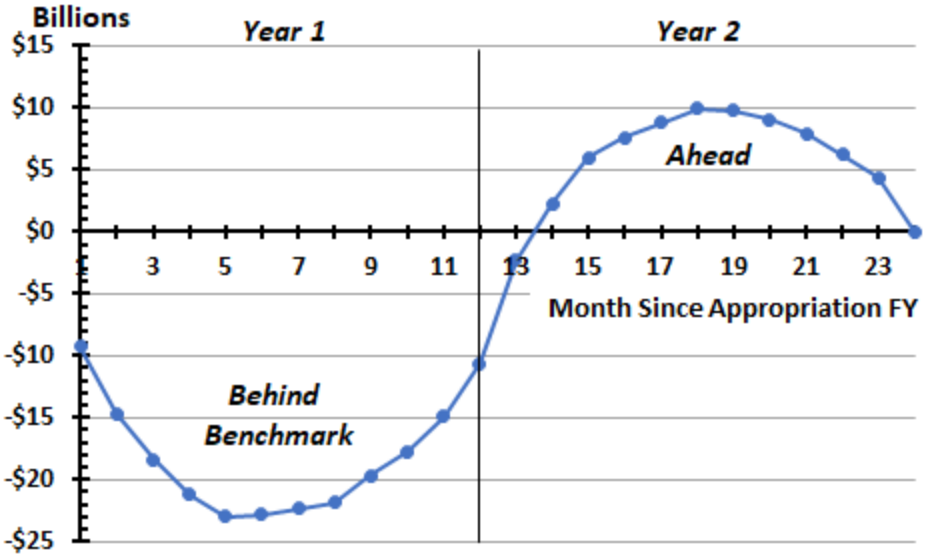


Figure 12. Average Cumulative MILCON Expenditures Versus Benchmark as Percentage of Month 36 Obligations (FY 2011–2018 Appropriations)



Overall, RDT&E, O&M, and MILCON expenditure differences between actuals over the last decade and the current linear benchmarks can be as large as $10 billion, $23 billion, and $3 billion, respectively (see, for example, Figure 13).

Figure 13. Dollar Difference Between Average Cumulative O&M Expenditures and Current Benchmark (FY 2011–2022 Appropriations)



Aligning Obligation and Expenditure Benchmarks with Theory and Data

This paper reached the following conclusions based on the review of theory and analysis of available data (see also Anton and Buettner, forthcoming, for further details).

At the least, benchmarks should be adjusted to reflect realities evident in recent years. DoD obligation and expenditure data consistently show statistically significant differences between average actuals and simple linear benchmarks. If benchmarks are not adjusted, then benchmarks are less effective at identifying potential issues. When average (normal) actuals are behind the benchmark, then too many programs may be undergoing subsequent deep-dive performance reviews. Likewise, when average actuals are above the benchmarks, then too few programs may be undergoing subsequent deep-dive performance reviews. Thus, these are indicators that updating benchmarks may improve the effectiveness and efficiency of performance reviews by helping to focus on programs that may be behind. For example, Figure 13 shows that O&M expenditures are, on average, as much as $23 billion below benchmarks in months 5–6 (right before mid-year reviews) and as much as $10 billion over benchmarks by month 18. This indicates potentially significant inefficiencies given limited oversight resources.

The best shape of obligation benchmark curves ultimately comes down to intent and theory. While our analysis shows that managers in the DoD have tended to obligate at rates that generally align with current linear obligation benchmarks, there are good reasons to reconsider these profiles. First, even with pressures to obligate on a straight line, actual data show startup delays as well as reductions due to CR effects. Also, RDT&E inherently involves engineering uncertainty and surprises, so it may be more effective for the DoD and the country to target more obligations in the second year than in the first. In addition, shifting more obligation targets for RDT&E and Procurement into the second year would give DoD managers more time to make investments when needed (earlier or later), negotiate better deals (e.g., prices, intellectual property rights, and deliverables), and fully assess contractors' execution, subcontracting, and supply-chain plans and risks.

Benchmarks should be adjusted for CR and financial-management realities. Regardless of the basic shape of the benchmarks, the statistical analysis in this paper shows real-world effects that should be considered for RDT&E, Procurement, and O&M. CRs result in obligation bumps after full budgets are passed as well as reductions during CRs for S&T and

Management Support. Obligation rates in the first month (October of the first year) are lower than the current benchmarks (probably from the time it takes for the financial management system to allocate spending authority to program managers). These CR effects introduce some level of S-curve patterns into actual obligation rates.

S-curves for obligation benchmarks may be beneficial for RDT&E, PROC, and O&M. While actual obligations have underlying linear bases, shifting to an S-curve profile for obligations would allow more time for improved performance and deals, addressing the points above.

Benchmarks can be useful but require additional due diligence. When combined with further due diligence, benchmarks can help the DoD and Congress identify funds that could be reprogrammed to address higher-priority threats and needs that emerge during the spending periods. The combined effects of these benefits can improve DoD mission outcomes by identifying badly needed resources. However, the emphasis here is on proper use and due diligence to ensure a balance between the benefits and issues. The use of benchmarks alone does not provide insight into the practical realities and issues in execution. Anecdotes indicate that DoD and Congressional leadership do not rely solely on benchmarks to identify from whom to take money for new urgent priorities that arise during the year of execution. However, other anecdotal evidence indicates that program managers believe otherwise, adding to the concern that these managers may prioritize spending to benchmarks over more prudent uses of financial resources, leading to undesirable or unforeseen negative side effects.

Avoid unforeseen negative consequences from managing to benchmarks. Finally, metrics drive behavior. This concern is well documented in the literature[4] and can be seen in the increased obligation rates in March immediately before the midyear reviews that identify programs spending below the benchmark rates for potential budget reprogramming to other programs and needs. While management metrics can be useful tools for insight, management pressures will drive behavior to the exclusion of other factors. Forcing people to spend to a curve will get spending to that curve whether or not that spending results in the best use of taxpayer dollars and the best results for national security. This axiom also applies to other potential uses of these benchmarks, such as adjusting Office of Management and Budget (OMB) apportionments based on changes in benchmarks.

Suggestions for Future Research

Future research in the following areas may lead to additional recommendations:

- Piloting modified benchmarks.
- Identifying expenditure benchmark profiles for Procurement.
- Assessing obligation and expenditure rates at the account level within each category.
- Assessing sources of obligation and expenditure data errors.

Recommendations

In this paper we provided the bulk of our theoretical work from Anton and Buettner (forthcoming) and Stalebrink (forthcoming). Based on these observations, we recommended that the DoD Comptroller consider modifying their benchmarks. Four optional variants are discussed in the report and are provided in Table 4. The preferred option includes adding additional S-curve ramp-up elements on top of historical obligation behaviors and recommends replacing linear expenditure profiles with historical S-curve profiles. Table 5 summarizes our recommendations.

[4] See, for example, National Research Council, 2005; Behn, 2008.

Anton and Buettner (forthcoming) provide candidate benchmark tables that better reflect recent history as well as the correlative effects of month, CR, and time.

In addition to aligning expenditure benchmarks to actual data and theoretical objectives, such changes could help eliminate the negative side effects cited in theory and the literature that program managers may seek expenditures prematurely just to meet comptroller benchmarks at the expense of other program and department objectives of prudent use of the resources (see, for example, Commission on PPBE Reform, 2023, p. 33; Marsalis, 2022). Slight delays in switching to S-curves with their lower initial expenditure benchmarks should give program managers more time to get good deals for the program, the DoD, and taxpayers rather than having to rush negotiations and contracting to meet somewhat arbitrary benchmarks or risk losing their funding.

There would be some cultural and process adjustments for both Congress and the DoD (and industry) to adjusting the obligation and expenditure benchmark profiles over time, but the benefits could be improved performance given the financial resources provided by Congress and the taxpayers to the DoD. In the end, keep in mind the following insightful quote.

Tell me how you measure me, and I will tell you how I will behave. If you measure me in an illogical way … do not complain about illogical behavior.

Eliyahu Moshe Goldratt

Table 4. Benchmarks Options: Elements and Ranking

	Obligations		Expenditures	RDT&E	Rank
	Base Shape	***Variables***			
Option 1	S-curves on historic	CR, MAB, Calendar, and Time Effects	Historic (S-curved)	Separate S&T, DEV, Mgt	1 (Preferred)
Option 2	S-curves on historic			Combined	2
Option 3	Historic (linear base)			Separate S&T, DEV, Mgt	2
Option 4	Historic (linear base)			Combined	3
Option 5	As-is (arbitrary lines)	None	As-is (arbitrary lines)	Combined	**4**

Table 5. Recommendations for Improving Obligation and Expenditure Benchmarks

Obligations	Expenditures

• Reduce obligation benchmarks for the first 1–2 months for RDT&E, PROC, and O&M to reflect process delays in allocating spending authorities.
• Consider changing benchmarks to S-curves instead of straight lines.
• Consider allowing more time in benchmarks for later spending to give time to get better negotiated deals and address surprises.
	• Change benchmarks to S-curves for RDT&E, PROC, and O&M.
	• At a minimum, if the benchmarks are not changed to S-curves, consider:
	  ○ Reducing expenditure benchmarks for the first 3 months.
	  ○ Changing benchmark shapes to straight lines across all years for multi-year funds rather than front-loading in the first year.
	• Add predictive metrics to identify more likely spending shortfalls.

• Explore switching to plan-based benchmarks instead of fixed benchmark curves, using Advana to collect plans from program offices. • Ensure proper due diligence along with spending relative to benchmarks before taking program funds. • Use needs, plans, and priorities for budgeting—not just spending. • Avoid overly enforcing benchmarks and other metrics. Keep these as information tools. • To avoid slowing down DoD acquisition, do not use obligation and expenditure benchmarks as a guide to OMB apportionments—instead inform apportionments based on the distribution data of recent actual obligations and expenditures. • Pilot these changes before pursuing more aggressive shifts to lower benchmarks in earlier years to understand better the effects (if any) on changes in unobligated and unexpended funds at the end of normal availability.	

Acknowledgements

We would like to thank the Commission on Planning, Programming, Budgeting, and Execution (PPBE) Reform for their sponsorship and support throughout this research. In particular, the questions, discussions, and feedback from the Commission chairs and staff were extremely helpful: the Hon. Robert Hale, the Hon. Ellen Lord, as well as the commission staff, especially Lara Sayer (our sponsor of record), Annie Crum, Elizabeth Bieri, Caroline Bledsoe, Kelle McCluskey, Brooks Minnick, and Soleil Sykes. We would also like to thank John Stedge in the Chief Digital and Artificial Intelligence Office (CDAO), for guiding us on how to pull the Obligations and Expenditures data from Advana. Thanks to Scott Lucero and Nathan Self from Virginia Tech for support with extracting USASpending.com data from their data repository. Finally, we would like to acknowledge and thank our Stevens staff: Technical Editor, Nancy Méndez-Booth for her editorial review comments and suggestions, and Tara Kelly for her project management skills, edits, and overall support.

The views, findings, conclusions, and recommendations expressed in this material are solely those of the authors and do not necessarily reflect the views or positions of the United States Government (including the Department of Defense (DoD) and any government personnel) or the Stevens Institute of Technology.

References

Anton, P. S.,& Buettner, D. J. (Forthcoming). *DoD obligation and expenditure rates: More-realistic benchmarks and the effects of continuing resolutions and other events on*

obligation rates. Acquisition Innovation Research Center, Systems Engineering Research Center UARC, Stevens Institute of Technology.

Behn, R. D. (2008). On why, to improve performance, measurement is rarely enough. *Bob Behn's Public Management Report*, *5*(9).

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. https://doi.org/10.1080/1941658X.2015.1096219

Burgess, E., Smirnoff, J., & Wong, B. (2014, June). *Weibull analysis method* [Briefing]. ICEAA Annual Symposium, Denver, CO.

Commission on Planning, Programming, Budgeting, and Execution (PPBE) Reform. (2023, August). *Interim report*. https://ppbereform.senate.gov/wp-content/uploads/2023/08/PPBE-Commission-Interim-Report-Final.pdf

Davis, D. (2008, April 15). Early warning model for acquisition program cost and schedule growth. *2008 SCEA-ISPA Joint Annual Conference and Training Workshop*. https://www.iceaaonline.com/wp-content/uploads/2017/09/EVM04.pdf

Davis, D., Christle, G., & Abba, W. (2009, January). *Using the Rayleigh model to assess future acquisition contract performance and overall contract risk: The future isn't what it used to be. (Yogi Berra) (Volume I)* (CRM D0019289.A2/Final). Center for Naval Analysis. https://www.cna.org/archive/CNA_Files/pdf/d0019289.a2.pdf

Department of the Air Force. (2022, February). *Planning, Programming, Budgeting, and Execution System training program: Reference manual*. https://afacpo.com/AQDocs/PPBE.pdf

Gallagher, M. A., & Lee, D. A. (1996, June). Final-cost estimates for research & development programs conditioned on realized costs. *Military Operations Research*, *2*(2), 51–65. https://www.researchgate.net/profile/Mark-Gallagher-5/publication/233620264_Final-Cost_Estimates_for_Research_Development_Programs_Conditioned_on_Realized_Costs/links/55956b8808ae5d8f3930f1ff/Final-Cost-Estimates-for-Research-Development-Programs-Conditioned-on-Realized-Costs.pdf

GAO. (2005). *A glossary of terms used in the federal budget process* (GAO-05-734SP). https://www.gao.gov/assets/gao-05-734sp.pdf

Lee, D. A., Hogue, M. R., & Gallagher, M. A. (1993, September 1). *Determining a budget profile from a development cost estimate*. https://apps.dtic.mil/sti/pdfs/ADA275864.pdf

Lee, D., Hogue, M., & Hoffman, D. (1993). *Time histories of expenditures for defense acquisition programs in the development phase*. Annual Meeting of the International Society for Parametric Analysis.

Marsalis, J. R. (2022). *The negative impacts of the benchmarks*. Senior Acquisition Course, National Defense University. https://nps.edu/documents/105938399/0/Negative+Impacts+of+Benchmarks_Marsalis.pdf/fc216616-7b4c-45e6-d4fa-c4bc48a6b71f?t=1652994212754

National Research Council. (2005). *Thinking strategically: The appropriate use of metrics for the climate change science program*. The National Academies Press. https://doi.org/10.17226/11292

Norden, P. V. (1970). Useful tools for project management. In W. Voris & M. K. Starr (eds.), *Management of Production* (pp. 71–101). Penguin.

Schiavoni, M. L. (2019, May 14–17). *Adaptive curve fitting: An algorithm in a sea of models* [Briefing]. ICEAA Professional Development & Training Workshop, Tampa FL. https://www.iceaaonline.com/wp-content/uploads/2019/06/AM05-Adaptive-Curve-Fitting-Schiavoni.pdf

Stalebrink, O. J. (Forthcoming). *A budget-theoretical perspective on DoD PPBE reform*. Acquisition Innovation Research Center, Stevens Institute of Technology.

Tomasini, R. (2017). *OUSDC rule-of-thumb acquisition obligation and expenditure rates*. Defense Acquisition University. https://www.dau.edu/sites/default/files/Migrated/ToolAttachments/OSD%20%28C%29%20Color%20Rule-of-Thumb%20Acq%20Obligation%20and%20Expenditure%20Rates.pdf and https://www.dau.edu/tools/ousdc-rule-thumb-acquisition-obligation-and-expenditure-rates.

Watkins, H. (1982). *An application of Rayleigh curve theory to contract cost estimation and control* [Master's thesis, Naval Postgraduate School]. https://apps.dtic.mil/sti/pdfs/ADA118213.pdf

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net