Volume 12 | Issue 1 | February 2025 ISSN: 2160-4746

Journal of Cost Analysis and Parametrics

Editor in Chief: David L. Peeler, Jr., CCEA®

ICEAA
International Cost Estimating
& Analysis Association

www.iceaaonline.com

ICEAA

Journal of Cost Analysis and Parametrics

Volume 12 | Issue 1 | February 2025

ISSN: 2160-4746

Editor's Note

David L. Peeler, Jr., CCEA®

Welcome to the continuation of the Journal of Cost Analysis and Parametrics (JCAP). Apologies to readers and authors alike for the gap between the last issue and this one. We extend much appreciation for those who have written, submitted, and revised pieces for publication, as well as those who have taken the time and mental energy to review the submissions. This issue contains articles across a wide scope of cost estimation and analysis interests.

Once again, we include author pursuits from work projects, school studies, and pieces presented to our annual ICEAA professional development and training workshop. First up, in this issue, is the best paper award from the 2023 ICEAA workshop. Nice work guys. In *Innovative Risk-Driven Contract Pricing Strategy* **Brian Flynn**, **Peter Braxton**, and **Robert Nehring** present a scoring framework that quantifies the influence of risk factors and uncertainties on contract types. Specifically, they provide a numerical basis for determining a recommended contract geometry for procurement actions.

The second article is from the graduate work of **David K. Smith**, which won the 2023 ICEAA Outstanding Air Force Institute of Technology Thesis Award. In *Human Capital Impacts in Military Acquisitions* Smith investigates the impact of program office personnel, program office location, and program Acquisition Category on cost and schedule performance.

Article three is a very interesting learning curve piece from **Brent M. Johnstone**. *Trouble With the Curve: Engineering Changes and Manufacturing Learning Curves* is a thought-provoking and applicable piece for our discipline. The piece examines the impact of engineering changes on learning curves. The article demonstrates how to analyze an engineering change by breaking it into component pieces and outlines techniques to calculate the reversionary impact on the learning curve. Brent makes learning curves reading interesting.

A wonderful foray into the cross-section of cost estimation and economics is *CSI EU (Cost Scene Investigation – European Union)* by **Douglas K. Howarth**. Our fourth article explores the decisions regarding the procurement of the Airbus A380. Specifically, the author contends his model would have informed decision-makers regarding sell quantities vis-a vis cost and price prior to decisions that precipitated significant cost growth. The article examines causes, as well as preventive information available beforehand.

What is the U.S. DoD Cost Estimation Community Saying About Agile? Find out from **Anandi Hira** and **Benjamin Kwok**, as they summarize their grey paper review of 63 papers and presentations relevant to the U.S. Government and DoD cost estimators. Of import are the various approaches to cost estimating and budgeting between commercial and defense sectors. How can DoD estimators reconcile the differences. Dig into their paper and find out.

Michael Smith, summarizes his graduate thesis in *Investigating Shifts in Engineering Manufacturing Development (EMD) Factors for Department of Defense Assets through Decadal Analysis*. He focuses on factor development for the engineering and manufacturing development phase of acquisition and uniquely investigates level two work breakdown structure elements to assess changes or trends. He shows which factors are stable, increasing, decreasing, or unpredictable over time.

We hope this overview is helpful in focusing your reading choices. Of course, we'd prefer you to read all the articles, but understand time constraints and interests. Enjoy your choices. May you find something in these pages applicable to your efforts, helpful in your professional pursuits, and personal interests. Thank you for your continued interest, renewed attention, and future support.

Innovative Risk-Driven Contract Pricing Strategy

Brian Flynn

Peter J. Braxton

Robert Nehring

Effective contract pricing strategy helps enable the achievement of best value-for-money outcomes in defense procurement. But any strategy, to be effective, must align with exigencies of the development of production effort. There's no one-size-fits-all. More specifically, selecting a contract type, such as cost-plus or fixed price; a package of incentives (such as event- or calendar-driven); and share lines (above and below target) that are appropriate for a contract requires careful consideration, measurement and assessment of risk factors and uncertainties. These include:

- Requirements Stability: Degree of firmness and completeness.
- Market Forces: Degree of competition.
- Technology Maturity. Degree to which the platform and systems push the state of the art and are technically feasible.
- Contractor Readiness: Contractor experience with the design and build of the same or similar systems.
- Price Validation: Extent to which a contract's target cost and price have been estimated independently, outside the influence of the project office or contractor.
- Schedule. Likelihood of failing to meet schedule plans and the effect of that failure.

This paper presents a scoring framework that quantifies the influence of these factors and, thus, provides a numerical basis for the determination of recommended contract geometry for upcoming procurements.

1.0 Introduction

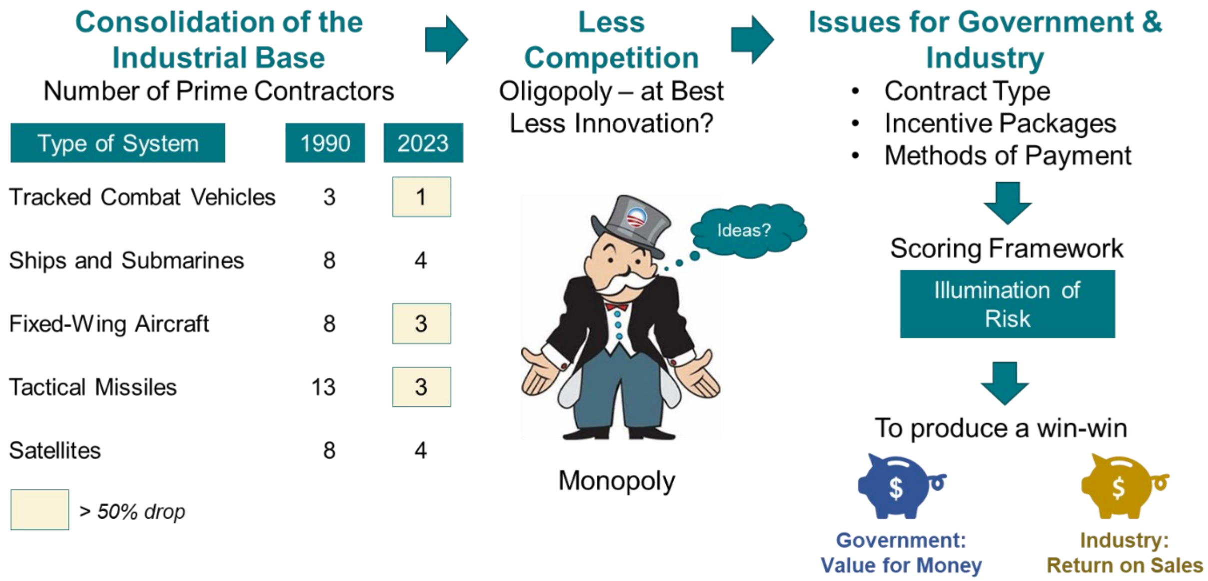
The U.S., the U.K., Canada, Australia, and other NATO and alliance partners routinely promulgate guidance on the establishment of contract parameters and contract geometries [1]. The guidance almost exclusively focuses on cost-based pricing where the defense marketplace is defined by oligopoly or even monopoly on the seller side and monopsony on the buyer side [2]. This is the reality of defense procurement today, internationally. The tenets of the guidance from whichever nation tend toward statements of general principles or intent, such as: use cost-plus contracts for high-risk procurements, use fixed-price incentive contracts for those of moderate risk, and use firm-fixed price for those of low risk. These generic recommendations make perfect sense.

But questions remain in implementing the guidance. For example, what precisely drives risk? Which factors influence risk and to what degree? What makes risk high versus only moderately high? How

---

1. Examples include the U.S. DoD's Contracts Price Referencing Guide and the Federal Acquisition Regulation (FAR), particularly Part 15, "Contracting by Negotiation," and Part 16, "Types of Contracts." Another example is the Australian Government's Contract Management Guide, Procurement Policy Branch, Commercial and Government Services, Department of Finance, Australia, December 2020.

2. Market-based pricing holds under conditions of robust, competitive procurement, which are increasingly rare. As noted by the U.S. Undersecretary of Defense for Acquisition and Sustainment, [USD(A&S); January 2022], "When markets are competitive, the Department reaps the benefits through improved cost, schedule, and performance for the products and services needed to support national defense. During initial procurement, incentivizing innovation through competition drives industry to offer its best technical solutions at a best-value cost and price. During contract performance, the expectation that contractors will have to compete against other firms in the future encourages them to perform effectively and efficiently."



Type of System	1990	2023
Tracked Combat Vehicles	3	1
Ships and Submarines	8	4
Fixed-Wing Aircraft	8	3
Tactical Missiles	13	3
Satellites	8	4

Figure 1: The Need for an Improved Contract Risk Model

do you tell the difference? Is there a magic metric to employ? What's the impact of any one factor in driving overall contract and procurement risk? Which specific elements of risk should be addressed in the contract – to incentivize vendor performance?

This paper attempts to provide the answers, through an analysis of U.S. Navy programs, contracts, and Contract Line Item Numbers (CLINs), and develops a model and results applicable for the other Services and allied nations too. The need has never been greater, as Figure 1 shows.

There is an urgent need to align contract parameters with contract risk to achieve better outcomes. More than ever, sound, data-driven metrics are needed to better illuminate contract risk and engender more informed pricing strategies in the face of these challenges:

- Limited capacity and competition in the defense industrial base (DIB) such as in the case of manufacturing urgently needed military systems for Ukraine, such as the M142 High Mobility Artillery Rocket System (HIMARS);
- Massive consolidation in the DIB, within additional details provided in Appendix 1, which, in turn, increases vendor pricing power; and
- Continued cost growth in major defense acquisition systems, as shown in Appendix 2.

On the first count, only Lockheed Martin manufactures HIMARS, with potential implications of any sole-source procurement highlighted in Figure 2. As noted by Dr. Bill LaPlante, USD (A&S), HIMARS is "… produced in Camden, Arkansas, in a big factory that used to be *literally a diaper factory*!" [3]. With a sole-source producer of "St. HIMARS," as bloggers dub it, the efficacy of contract parameters lessens – adding to the need to review their selection and methods to incentivize the vendor.

---

3. Transcripts from "Getting Weapons into Production," 2022 Conference hosted by George Mason University and the Defense Acquisition University.

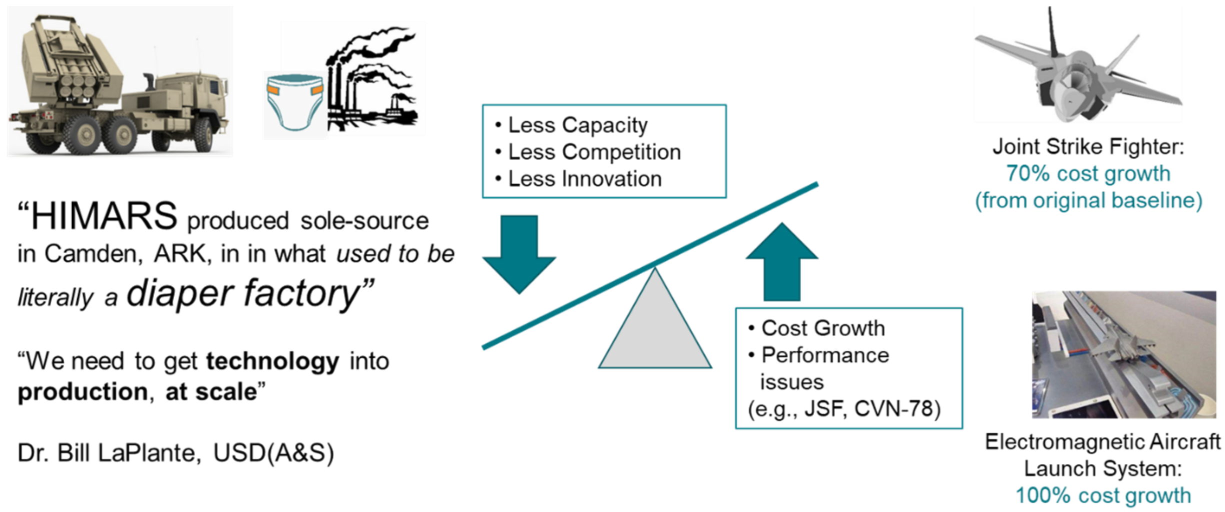


Figure 2: HIMARS Production for Ukraine

On the second count, the statistics are stark. Since the 1990s, the U.S. defense sector has consolidated substantially, transitioning from 51 to five aerospace prime contractors [4]. The number of suppliers for tactical missiles, fixed-wing aircraft, and satellites have all declined dramatically. More than 90% of missiles now come from just three sources [5].

In the shipyard industrial base, the situational is arguably even worse from a competition perspective. Two companies, General Dynamics (GD) and Huntington Ingalls Industries (HII), own the five largest U.S. shipyards. Even in cases of so-called competitive procurement, considerations of maintaining the industrial base heavily influence the allocation of contracts between the big three East Coast yards, GD/Bath Iron Works (BIW), GD/ Electric Boat (EB), and HII/Newport News Shipbuilding (NNS), and one Gulf Coast yard, HII/ Ingalls.

Further, only HII/NNS builds aircraft carriers; only GD/EB and HII/NNS build submarines; and only Ingalls builds the U.S. Navy's big amphibious attack vessels [6].

Finally, second-tier shipbuilders such as Eastern, Halter Marine, and Austal often struggle with limited workload.

Given the absence of robust competition in the DIB, such as exists in other sectors of the global economy (e.g., the automobile industry or even the commercial shipbuilding industry), it is crucial to counter-balance industry's pricing power with carefully constructed contract types and incentives – using the risk framework presented in this paper.

On the third count, cost growth remains an issue across the board in defense procurement. Notable recent examples include 100% cost growth for the Electromagnetic Aircraft Launch System (EMALS), and 20+% cost growth for both the CVN-78, USS *Ford*, and for the second ship in the class, CVN-79, the USS *Kennedy*.

This paper shows that the selections of contract types and incentive packages for some of these high-cost-growth procurements were of questionable efficacy – based on ex-post grading using the risk framework.

---

4. "State of Competition within the Defense Industrial Base," Undersecretary of Defense (Acquisition and Sustainment), 2022.

5. Ibid.

6. The amphibious vessels include the LHAs, LHDs, and LPDs. The difference between Landing Helicopter Dock (LHD) and Landing Helicopter Assault (LHA) is mainly a matter of emphasis. Both have a full-length flight decks and utilize helicopters. The LHD mainly utilizes landing crafts to bring troops and equipment ashore while a LHA uses all or mostly air assets for the same mission. The Landing Platform Dock (LPD) vessels carries Landing Craft Aircushion Cushion (LCACs).

2.0 Contract Risk Management Framework

2.1 Objective

Defense contracts have frequently been conceived with disconnects between incentives that are designed and employed by the government and motivational factors that drive the contractor. These fundamental disconnects result in financial motivations that too often encourage contractors to expend extra effort on performance goals that are not important enough to the user to justify their increased cost, and that result in less-than-desired system interoperability, reliability, and sustainability [7]. The end result is an unsatisfactory outcome for both parties instead of win-win. The contractor falls short of achieving its targeted return on investment (ROI) and the government its expected value for money.

The contract risk management framework seeks to fix some of the misalignments through illumination of the elements of risk in the establishment of contract type, methods of payment, and incentives. More specifically, the scoring framework seeks to quantify risk from an ex-post numerical evaluation of the many factors that make or break a program and its contracts, such as the experience of the contractor, stretch in technology, solidity of requirements, and presence or absence of robust competitive procurement.

The benchmarks derived from the assessment of several dozen contracts across multiple programs serve to inform ongoing and future development, design, and procurement efforts. This analysis, in turn, will help the government and contractor forge and maintain cooperative (win-win) relationships throughout the contracting process to ensure equitable returns for vendors while delivering systems on time, on budget, and that meet effectiveness, reliability, and sustainability threshold requirements [8].

2.2 Elements of Risk

The pricing parameters of a contract are highly dependent on the nature of the procurement. For example, the contract type and package of incentives for a design and development contract for a next-generation fighter aircraft differ fundamentally from those of a steady-state production contract. The decision calculus, then, requires careful consideration, measurement and assessment of the risk factors and uncertainties which evolve as a program proceeds through the acquisition lifecycle. Importantly, some programs achieve stability much later than others. Joint Strike Fighter (JSF), for example, is still using low rate initial production (LRIP) contracts because achievement of full-operational capability remains elusive, and the concurrent development contract was drawn out over the better part of two decades.

The contract scoring matrix specifies and assesses the elements of risk denoted in Figure 3 for each contract or CLIN in an acquisition program.

Sound pricing strategy requires illumination of the risks that influence results

- Stability of Requirements
  - Degree of firmness and completeness. Perfectly defined requirements are unique and unambiguous, complete and consistent, measurable, traceable, and verifiable (testable).
  - Risk results from as-yet-unseen changes in threat (i.e., “we didn’t fully understand or anticipate the problem”) or in design (i.e., “we didn’t fully understand the required solution”).

---

7. Examples of poor contract outcomes include Joint Strike Fighter with over 70% cost growth and Littoral Combat Ship (LCS), and its mine countermeasure module. The U.S. Navy’s Remote Minehunting System (RMS), which was to be deployed from LCS, was cancelled due to poor effectiveness, poor reliability, and cost growth.

8. Threshold requirements are must-achieve metrics in the U.S., validated by the Joint Chiefs of Staff (Joint Requirements Oversight Council, or JROC) for Major Defense Acquisition Programs. Failure of a system to meet its Key Performance Parameter (KPP) threshold/initial minimum rescinds the JROC validation, brings the military utility of the associated system into question, and may result in a reevaluation of the program or modification to production increments.

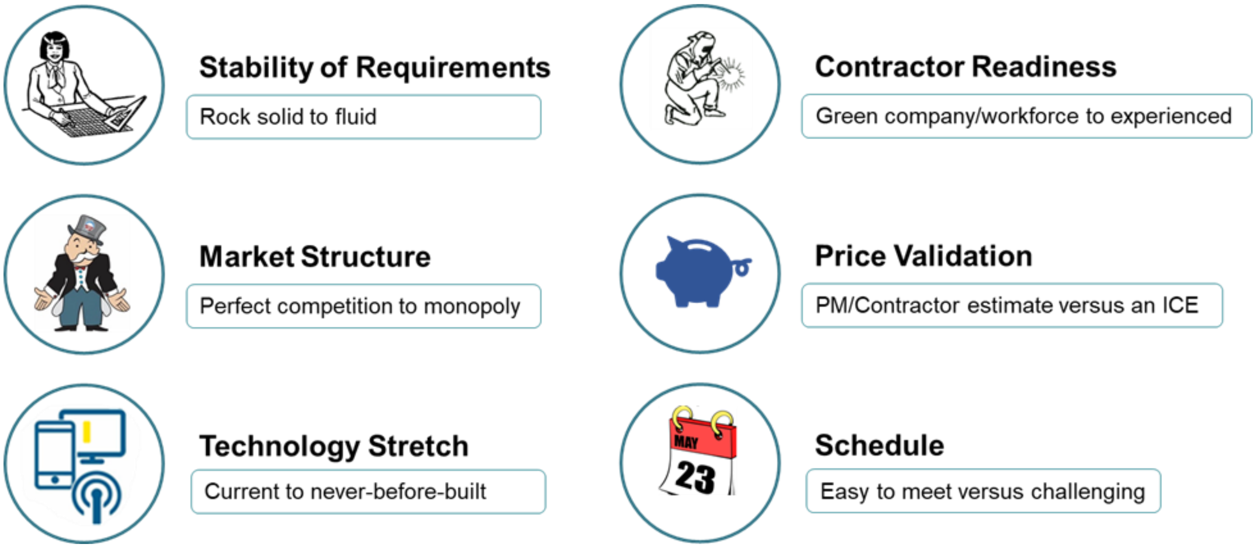


Figure 3: Elements of Risk

- Market Forces
  - Degree of competitive procurement.
- Maturity of Technology
  - Degree to which a platform and its systems are existing state-of-the-art and technically feasible, achievable, and obtainable.
- Contractor Readiness
  - Degree to which the company has experience with the design and build of the same or similar platform or systems.
- Price Validation
  - Extent to which a contract's target cost (both direct and indirect costs) and target price have been estimated by an independent authority outside the influence of the program office or the company.
- Schedule
  - Likelihood of failing to meet schedule plans and the effect of that failure.

3.0 Risk Scores

The Model uses a weighted average of scores for each of the elements of risk based on anchored, ratio scales. Each of these features (the weights and the scales) are discussed below.

3.1 The Weights

The contract risk factors are not equally important. There's no *a priori* reason they should be. Therein lies a problem in creating a valid scoring procedure using subject-matter experts. The opinions of practitioners vary according to their knowledge and experience. As Arrow's *Impossibility Theorem* indicates, all techniques to rank-order preferences, other than using a dictator, will violate at least one commonly accepted measure of fairness [9].

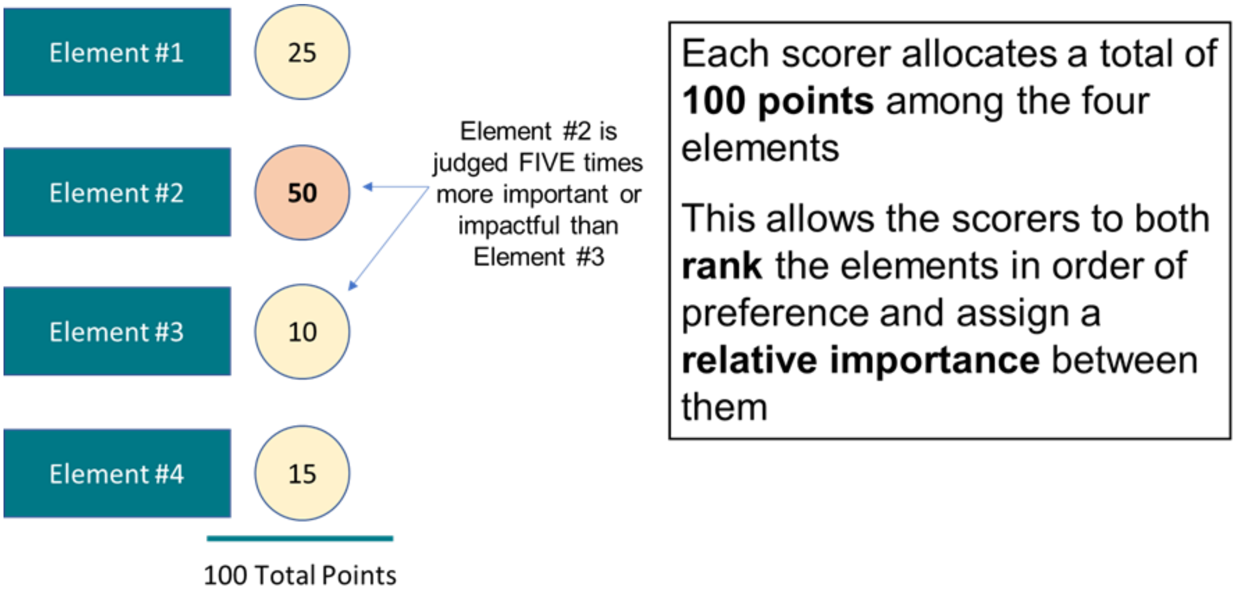
This research uses the highly regarded "Borda Count" technique to measure the rank order and relative importance of the risk factors, while recognizing that no flawless procedure exists for doing so [10]. Figure 4 presents a generic example.

Each scorer is given a total of 100 points to

---

9. Arrow, Kenneth; "A Difficulty in the Concept of Social Welfare," Ph.D. dissertation, 1951.

10. Major league baseball in the U.S. uses a modified Borda Count technique to choose its most valuable player.



Element #2 is judged FIVE times more important or impactful than Element #3

Each scorer allocates a total of **100 points** among the four elements

This allows the scorers to both **rank** the elements in order of preference and assign a **relative importance** between them

Figure 4: Borda Count Technique

distribute among the factors or elements [11]. A score of 20, for example, means that the factor is judged to be twice as important or impactful as a factor with a score of 10. Consensus is achieved on the scores.

The Borda Count procedure contrasts with ordinal ranking. Ordinal numbers signify order or position. It is common in Analyses of Alternatives, for example, to use ordinal rankings and to represent them by numbers or letters, which are merely shorthand for category labels:

1 represents Best; 2 represents Second Best; 3 represents Worst.

These rank orderings (1, 2, and 3) are ordinal not cardinal numbers, which express a quantity. Rank Order says nothing about the value of the score, only the order of the score. The numbers are merely shorthand for Best, Second Best, and Worst. The numbers could just as easily be letters such as *X, Y,* and *Z*, or *α, β,* and *γ*.

Unfortunately, an all-too-common occurrence in the U.S. DoD is to perform numerical computations (arithmetic) on ordinal ranking. The result is **totally meaningless**. It would be equivalent to saying that 11 Ensigns in the U.S. Navy (each with an O-Rank of 1) exceed the authority of the Chief of Naval Operations (with an O-Rank of 10).

3.2 The Scales

In a similar vein, ratio scales are used to assess the risk and uncertainty of individual contracts associated with the programs and contracts. The scales rate the best case as 1.0, the worst case as 2.0, with anchors provided in 0.25 increments.

In this scoring paradigm, a value of 1.50 represents 50% more impact or risk than a value of 1.00, and a value of 2.00 represents twice the impact. The use of ratio versus ordinal scales permits numerical manipulation of the scores using common arithmetical operations of addition, subtraction, multiplication, and division.

The guideposts of 1.00, 1.25, 1.50, 1.75, and 2.00 in the framework serve to anchor the scores across programs and contracts. For example, the category "Maturity of Technology" refers to level of

---

[11] This technique avoids the pitfall of cardinal ordering by measuring the amount by which one requirement or factor is judged more important than another. For more details, see "How to Use Rank Ordering for Comparison of Friendly COAs," Professors Downes-Martin and Volpe, 1 September, 2005, War Gaming Department, Unites States Naval War College.

technological sophistication or advancement required of the prime or vendor relative to the current state of the art. Any scores are allowed between the two “goalposts” of 1.00 and 2.00, with Technology Readiness Levels (TRLs) providing a useful gauge.

- Minimum Risk
  - The lowest score of 1.00 indicates that all or most technological requirements have been achieved on an identical item currently or previously in production by the prime contractor. In this case, few, if any, changes to the item (a system, component, or the platform itself) are required. No significant integration, weight, or size issues need to be addressed.
- Maximum Risk
  - The highest score of 2.00 represents new technology. That is, the item in question is significantly beyond the current state of the art. A new approach or concept is necessary to achieve the system requirement. In addition, the new concept has yet to be demonstrated, even in a laboratory environment. Unprecedented integration, weight, and size issues may have to be resolved before the system can meet operational requirements.

4.0 Model

4.1 Domain

The team assessed risk for 39 contracts within the domain of U.S. ship and ship-system design and construction programs. The choice of the programs within the broad domain was based on the team’s collective hands-on experience in generating cost estimates and analyses in support of both senior shipyard and U.S. Navy leadership. The experience included cost analysis support on surface combatants and amphibious vessels (for a private-sector company) and support to senior Pentagon officials on Remote Minehunting System, *Zumwalt* Class surface combatants, and *Ford* Class carriers.

The authors took pains to avoid selection bias when establishing the content of the sample. Contracts were selected from each of the major shipyards in the U.S. industrial base today.

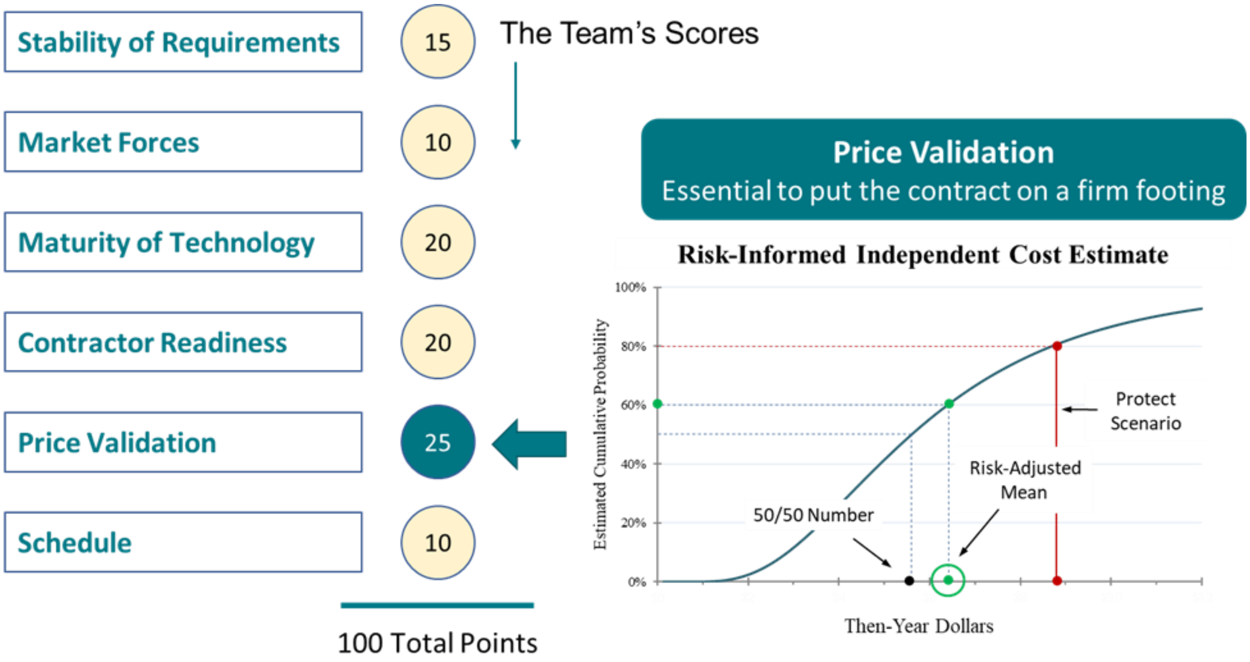


Figure 5: Risk Weightings

4.2 Weights

The team first established the risk-and-uncertainty weights shown in Figure 5, leveraging decades of experience in cost analysis, and using the Borda Count technique.

Interestingly, price validation was deemed the most important element of contract risk, perhaps due to the experience of team members with an entire spectrum of contracts and contract outcomes, where the quality and independence of the cost estimate proved essential in the establishment of a sound baseline.

4.3 Scoring Matrix

The scoring matrix uses anchored scales, with an example illustrated in Figure 6 for Market Forces, and with details presented in Table 1. The anchors are pre-defined benchmarks that are set at various points in the range of values (1.0 to 2.0) to increase the objectivity of the scoring.

A value of 1.00 is associated with robust competition, with four or more companies bidding for the work. A good example is the Navy's design and build contract for FFG-62 *Constellation* Class frigates, where four industry teams submitted different designs and prices.

A value of 1.50 is associated with a type of duopoly where there is vigorous competition between the two firms. It represents 50% more risk than the baseline value of 1.00, using the ratio scale. Duopoly in the defense market sometimes takes the form of allocation of work between the two firms to manage the industrial base (e.g., surface combatants for the Navy). The risk consequently higher (a value of 1.75).

Finally, a maximum value of 2.0 occurs in a sole-source environment, such as design and construction of aircraft carriers, where only Newport News Shipbuilding does the work.

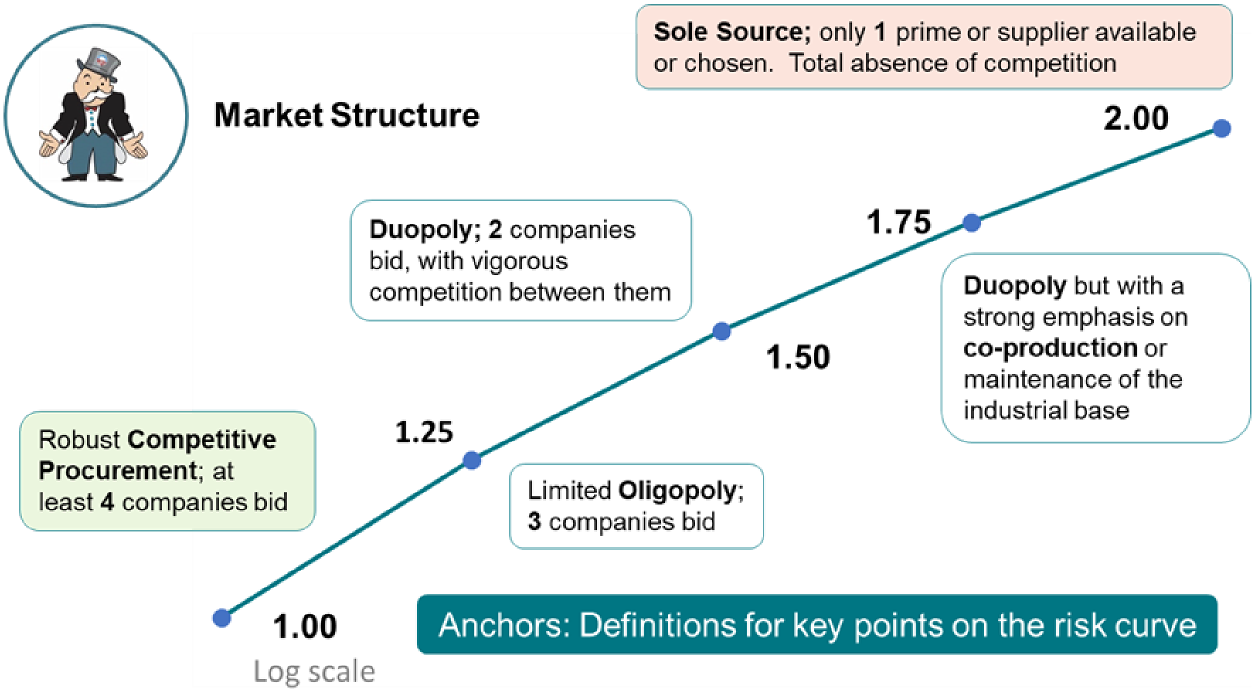
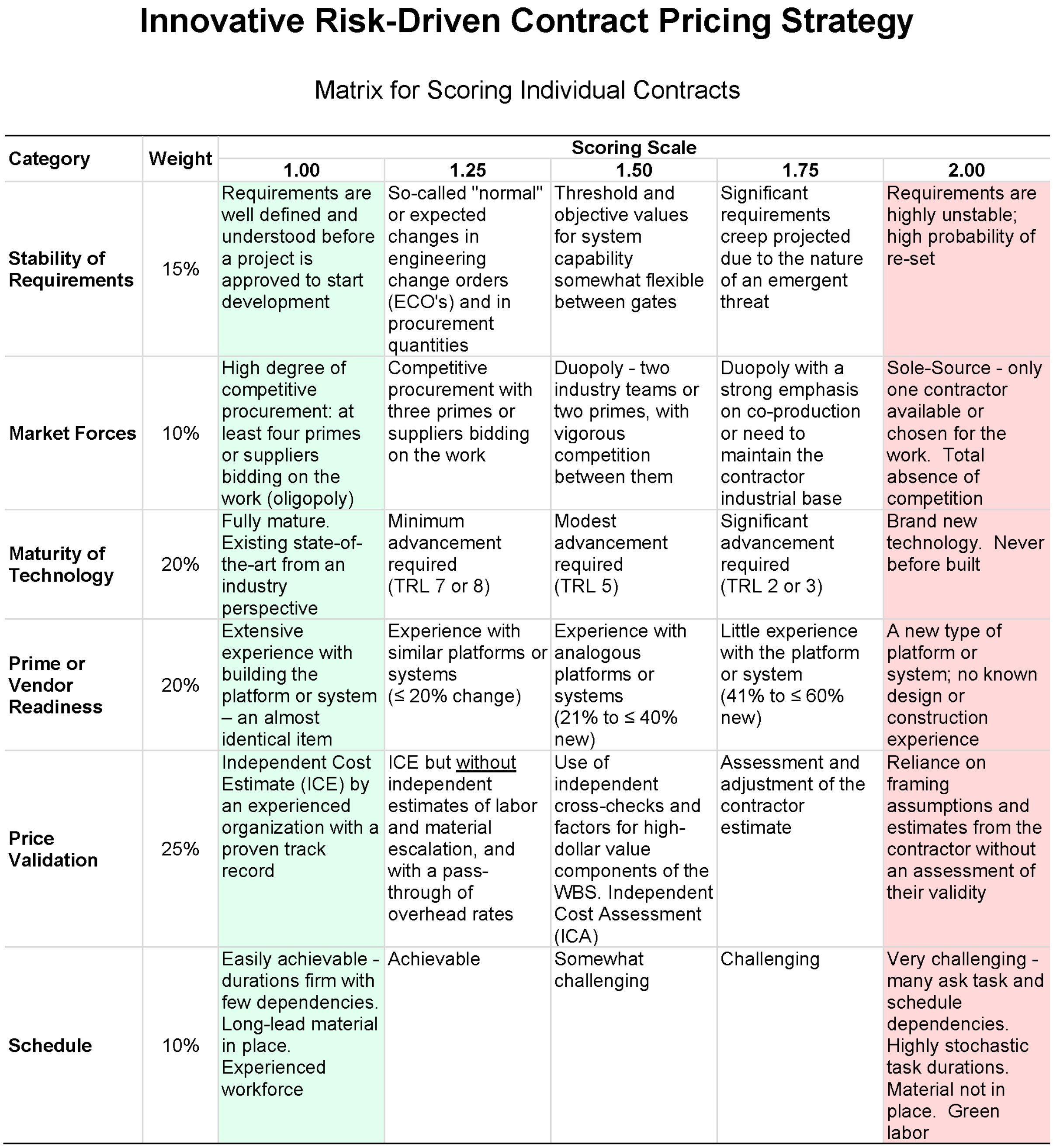


Figure 6: Anchored Scales



Innovative Risk-Driven Contract Pricing Strategy

Matrix for Scoring Individual Contracts

Category	Weight	Scoring Scale				
		1.00	1.25	1.50	1.75	2.00
Stability of Requirements	15%	Requirements are well defined and understood before a project is approved to start development	So-called "normal" or expected changes in engineering change orders (ECO's) and in procurement quantities	Threshold and objective values for system capability somewhat flexible between gates	Significant requirements creep projected due to the nature of an emergent threat	Requirements are highly unstable; high probability of re-set
Market Forces	10%	High degree of competitive procurement: at least four primes or suppliers bidding on the work (oligopoly)	Competitive procurement with three primes or suppliers bidding on the work	Duopoly - two industry teams or two primes, with vigorous competition between them	Duopoly with a strong emphasis on co-production or need to maintain the contractor industrial base	Sole-Source - only one contractor available or chosen for the work. Total absence of competition
Maturity of Technology	20%	Fully mature. Existing state-of-the-art from an industry perspective	Minimum advancement required (TRL 7 or 8)	Modest advancement required (TRL 5)	Significant advancement required (TRL 2 or 3)	Brand new technology. Never before built
Prime or Vendor Readiness	20%	Extensive experience with building the platform or system – an almost identical item	Experience with similar platforms or systems (≤ 20% change)	Experience with analogous platforms or systems (21% to ≤ 40% new)	Little experience with the platform or system (41% to ≤ 60% new)	A new type of platform or system; no known design or construction experience
Price Validation	25%	Independent Cost Estimate (ICE) by an experienced organization with a proven track record	ICE but without independent estimates of labor and material escalation, and with a pass-through of overhead rates	Use of independent cross-checks and factors for high-dollar value components of the WBS. Independent Cost Assessment (ICA)	Assessment and adjustment of the contractor estimate	Reliance on framing assumptions and estimates from the contractor without an assessment of their validity
Schedule	10%	Easily achievable - durations firm with few dependencies. Long-lead material in place. Experienced workforce	Achievable	Somewhat challenging	Challenging	Very challenging - many ask task and schedule dependencies. Highly stochastic task durations. Material not in place. Green labor

Table 1: Scoring Matrix for Individual Contracts

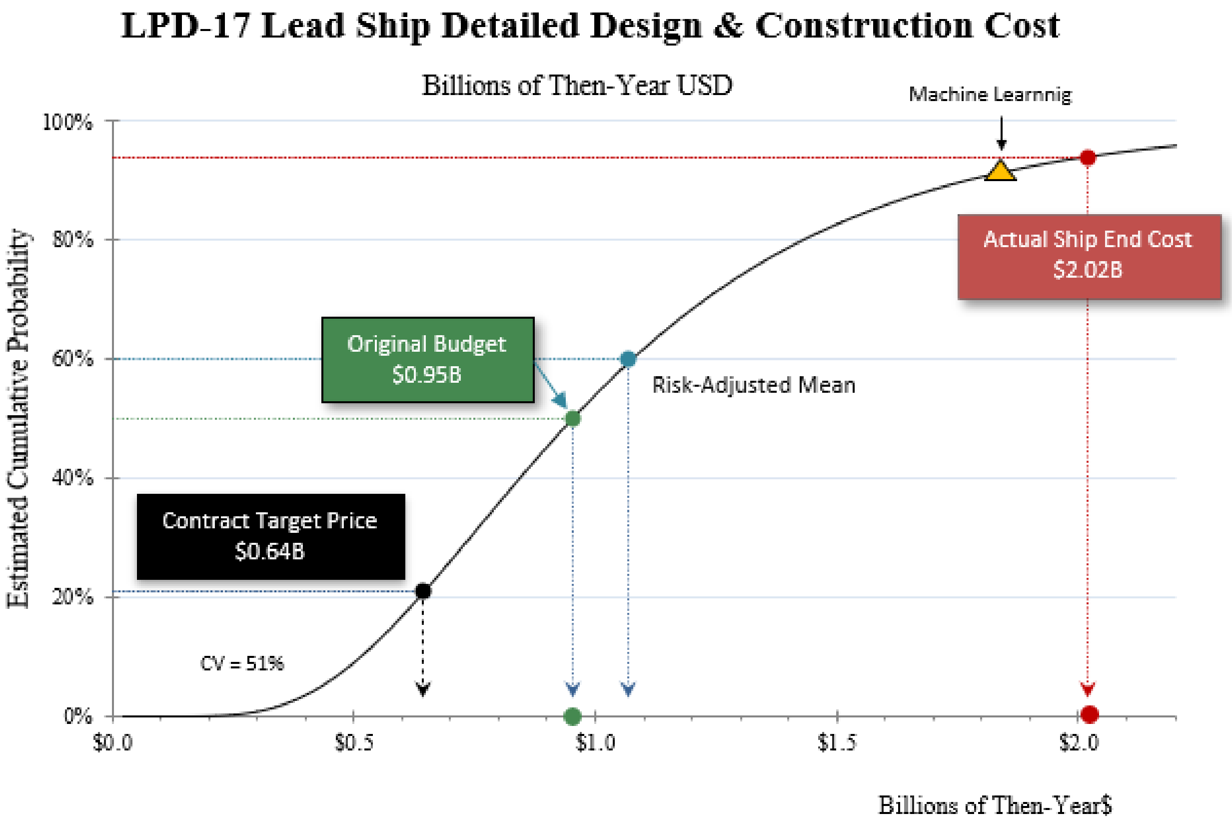


Figure 7: LPD-17 Cost Growth

Using the anchored scales and weights of the table, the team scored each of the 39 contracts (see Appendix 4). The LPD-17 *San Antonio* Class lead-ship Detailed Design and Construction (DD&C) contract provides a good example.

The contract type was originally cost plus award fee (CPAF). It was then then changed to cost plus incentive fee (CPIF) as technical problems emerged and cost growth became egregious – eventually reaching 100%, as Figure 7 shows [12].

In aggregate, the risk score for the lead-ship (USS *San Antonio*) contract was 1.71, as Figure 8 shows, or 70% higher than a no-risk baseline case. Note the maximum-risk score of 2.00 for Price Validation, as the program office bought into the shipyard's framing assumption that they would be at unit #4 on a learning curve – from the get-go, due to computer aided design (CAD) and collocation of the contractor and government management offices in New Orleans. The latter had no positive effect, and the CAD software bombed. Additional details of the scoring are presented in Appendix 3.

---

12. Problems persisted through the next five contracts. The Naval Sea Systems Command (NAVSEA) chose not to employ a fixed-price incentive (FPI) contract on the program until LPD-22.

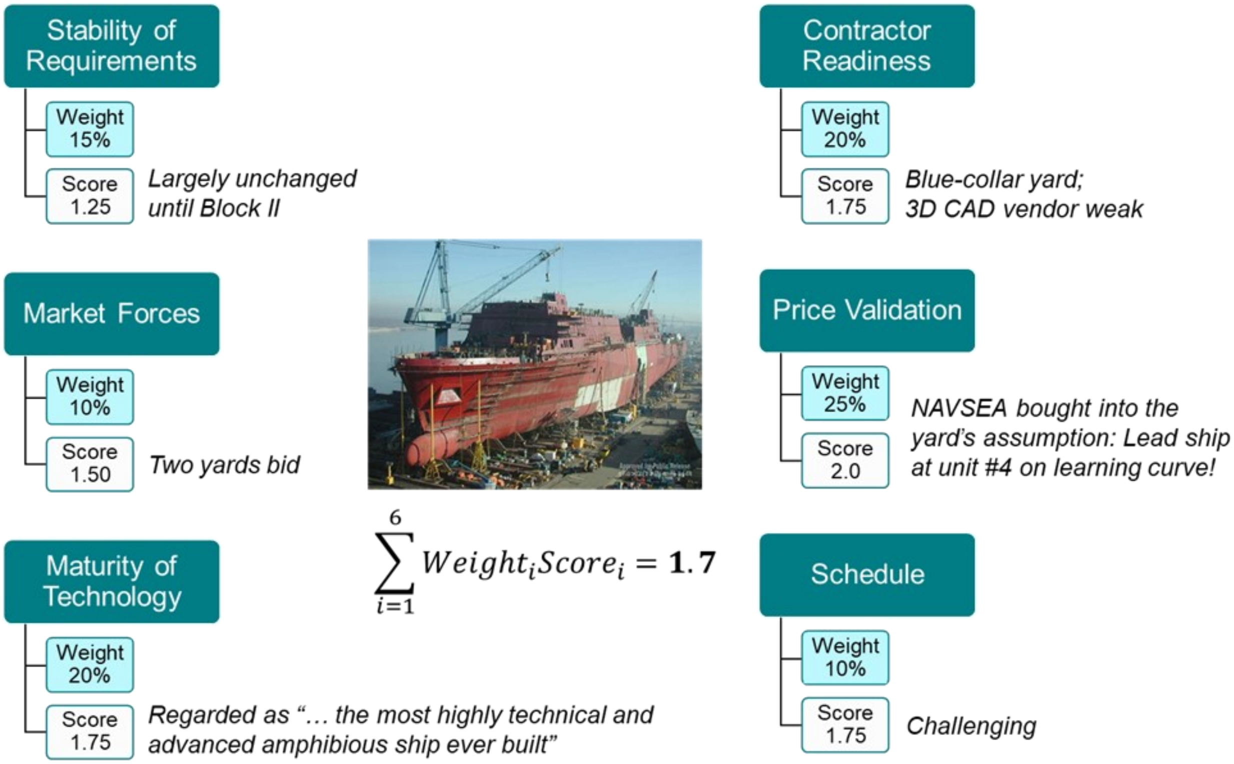


Figure 8: LPD-17 Lead-Ship Design and Construction Contract

4.4 Summary of Contract Risk Scores

Table 2 summarizes results of the scoring, with Appendix 4 presenting numerical details for each contract. The value of 1.70 for Market Forces stands out. This score is a reflection of a current concern of the Undersecretary of Defense for Acquisition and Sustainment: namely, diminished competition in the defense marketplace.

The scores represent useful benchmarks for judging contract risk for programs and contracts, both ongoing and future. Remarkably, the aggregate risk score is about 1.5, or the midpoint between no risk and maximum risk. The CVs show a marked consistency between the different categories of risk. Interestingly, the aggregate CV of 12% is lower than any of the individual category values, indicating that pluses and minuses tend to offset each other.

	Stability of Requirements	Market Forces	Maturity of Technology	Contractor Readiness	Price Validation	Schedule	Aggregate Weighted
Average (μ)	1.40	1.70	1.44	1.46	1.46	1.50	**1.47**
Std Dev (σ)	0.22	0.34	0.28	0.27	0.24	0.23	**0.18**
CV (σ/μ)	15.9%	20.0%	19.5%	18.3%	16.7%	15.2%	**12.5%**

Table 2: Summary of Contract Risk Scores for U.S. Naval Contracts

Key take-aways at a macro level include:

- Stability of Requirements: Capability-based planning is the gold standard in the U.S. and other NATO nations and alliance partners such as Australia. That said, requirements are never known with perfect certitude nor perfectly translated into ship technical and performance requirements. Requirements churn influenced the selection of contract types on programs such as DDG-1000, Littoral Combat Ship (LCS), and Remote Minehunting System (RMS). The values of 1.4 and 1.5 for these programs accord closely with moderate risk and uncertainty.
- Market Forces: The lack of robust competition has been and remains a problem for the Office of the Secretary of Defense and for the Services. The U.S. DoD generally but not always competes contracts for ship conceptual design and lead-ship design and construction. However, once a shipyard is selected for the work, they become a de-facto sole source. Hence the value of 1.7 for Market Forces. This phenomenon holds for other platform categories, too.
- Maturity of Technology: The aggregate score of 1.44 indicates moderate risk. Indeed, the Weapon Systems Acquisition Reform Act (WSARA) of 2009 strove to reduce risk in acquisition through, among other factors, stressing the development of prototypes early on. However, outliers occur, such as the previously mentioned LPD-17, with a risk score of 1.7 for technology immaturity. The practical reality is that the lead ship is generally a hybrid of a prototype and a proven operational model (cf. CVN-78). Even where engineering development models (EDMs) are undertaken to reduce risk with good intentions, the program can go overboard with too much simultaneous unproven technology, as in DDG-1000.
- Contractor and Vendor Readiness: Not surprisingly, the major U.S. shipyards with their long history of building the most complex military vessels in the world score well in terms of experience or fit, on average, with a value of 1.46. But, again, note the outliers. A classic example, as mentioned above, is Avondale Industries. The yard had never constructed a vessel of the complexity of LPD-17. Consequently, a good part of the work had to be transferred to Ingalls (a yard experienced with designing and building surface combatants) to complete the effort. (As a sad postscript, Avondale ceased construction of naval vessels after Hurricane Katrina.)
- Price Validation: For major weapon-system acquisition projects, the U.S. DoD produces Program Office Estimates (POEs), Component Cost Positions (CCPs), and Independent Cost Estimates (ICEs), with the latter usually performed by the Office of the Secretary of Defense (OSD) Cost Assessment and Program Evaluation (CAPE). This thorough and complete review process produces a risk score of 1.46, indicating good results, on average. But, yet again, there are outliers, such as LPD-17 for which Navy acquisition officials bought into an erroneous framing assumption from the shipyard. The assumption, which proved blatantly false, was that the lead ship would come in at a unit price normally found at the 4th unit on a learning curve – due to the efficiencies of computer-aided design (CAD) and collocation of the program office with the shipbuilder.
- Schedule Challenge: The challenge in meeting schedule is heavily influenced by the technology maturity and contractor readiness to perform the work. On average, the U.S. ship construction programs fare moderately well.

5.0 Program Insights

For each CLIN, the contract type and set of incentives should align with the risk profile of the framework, and with the project and contract scores representing invaluable benchmarks for future acquisitions. Take-aways gleaned from the individual project scores related to contract types, contract incentives, and methods of payment include:

- Effectiveness of Pricing Approach: The contract type is usually cost-plus (or a hybrid such as CPAF, CPFF, CPIF) for conceptual design, DD&C, and for the development of new technologies. The contract type then tends to become fixed-price for follow-on contracts. The effectiveness of the pricing approach in motivating the contractors is difficult to discern clearly, with these examples illuminating some of the issues:

- LSD-41 Class Ships: The contract for LSD-41 lead ship construction was originally CPAF. It was converted to CPFF (with ceiling) based on forecasts that the yard would significantly overrun target cost. The contract for LSD-42 was also CPAF. It was converted to FPI with a 50/50 share line and 123% ceiling due to poor contractor performance. In any event, moderate cost growth ensued for both contracts.
- LPD-17 Class Ships: The original contract type was CPAF based on controlling Total Ownership Cost (TOC) of the vessels, or, more specifically, future maintenance costs. In the face of cost growth, however, NAVSEA renegotiated the contract. It changed the contract type from CPAF to CPIF, with the incentive fee tied to controlling construction costs.
  - Nevertheless, the lead ship experienced 100% cost growth.
  - Further, because of egregious technical and performance issues, the government needed to use cost plus for five ships before changing to FPIF; that said, cost growth and schedule delays slowly but steadily decreased.
- CVN-78 Class Ships: The contract type for CVN-78 was composed of multiple cost reimbursable type contracts, including a massive Construction Preparation (CP) contract, circumventing full funding rules. This was advantageous for the U.S. Navy given immature technologies and poorly defined requirements, and it gave the yard (Newport News) the chance to reduce cycle times, maintain schedule, and maximize efficiency. The contract type for CVN-79 was fixed price incentive fee.
  - In general, FPIF is appropriate only when requirements are stable and technologies are mature. This was not the case with CVN-79, with the lead-ship (CVN-78) having been delivered at only 80% complete.
  - The contracts for the Electromagnetic Aircraft Launch System (EMALS) and Advanced Arresting Gear (AAG), both crucial for achieving planned aircraft sortie rates, were cost-plus. Cost growth reached 100% on the former. Production units were priced using fixed-price contracts. This pricing approach was effective only because cost growth was captured in the development contracts, absorbed fully by the government.
- Effectiveness of Performance Incentives: Incentives can be a beneficial tool in controlling contractor and vendor behavior. But the ability of fixed-price incentives to shape outcomes is a dubious proposition when new technologies are present. A good practice seems to be using FPIF or FFP contracts only after risks have been mitigated. Many of the projects in the sample do exactly this, but the following is a classic counterexample.
  - Remote Minehunting System: The contract was cost plus for development. Contracts were then awarded for Lot 1 and Lot 2 production using a fixed-price incentive strategy even though the RMS could not meet reliability thresholds. The incentives under production did not have their intended effects, and the project was eventually cancelled.
- Price Validation**:** Price validation has the highest weighting amongst the six factors that influence overall contract risk. The U.S. has performed poorly in estimating ship and ship-system costs, as Appendix 2 shows. No matter what the contract type and incentive structure, poor results are likely to result in the absence of a realistic, accurate, and complete cost baseline, with risk accounted for. Put another way, overruns relative to unrealistic contract target costs cannot be entirely blamed on poor performance.
  - Examples of poor estimates include LPD-17 lead and follow-on ships, EMALS, AAG, and RMS.

- Efficacy of the FPI Strategy: There's no evidence to suggest that one type of contract or set of contract performance incentives uniformly and consistently produces better outcomes than any others. The U.S. DoD advocates the use of fixed-price incentive contracts early-on, with 50/50 share lines [13]. However, each project is a non-repeatable experiment. Each is unique. Upfront flexibility and realism are critical in trying to influence the contractor to better manage the costs, schedule, and quality of the project. The government should be realistic when choosing a contract type. (In particular, FPIF contracts with a 50/50 share line and 120% ceiling price are patently unrealistic for most Development contracts, as will be demonstrated in Appendix 5.)
- Value of Flexibility:
  - If risk is high, then an economic analysis or analysis of alternatives (cost and capability) should reflect that risk and laser-focus on the importance of any new technology.
  - When using incentives, the contractors should reap their benefit only if the original goals were met, and not prematurely, as in the case of the Remote Multi-Mission Vehicle (RMMV) for the U.S. Navy's Remote Minehunting System. Incentives should not be awarded to cover additional costs incurred by the contractor that were not in the original estimate. Again, this is an important lesson from RMS, where the contractor was paid several hundred million dollars in additional expenditures to correct faulty work. Once technologies are mature enough that the risk profile supports FPIF or FFP contract types, then the fixed-price award should be used after the completion of a development contract.
    - In this fashion, without the presence of an active, on-going development contract, the risk of cost overruns becomes the burden of the contractor or vendor, and not the government.
    - When the government finally decides to move to a FPIF or FFP contract type, then the development contracts should be complete or near complete. These fixed-price contract types should not be available to the contractors as a source to cover any subsequent cost growth, if and when it occurs. That would only serve to incentivize undesirable behavior on the part of the vendor.
- No Guarantees: The use of any particular contract type or set of incentives is no guarantee of success. The use of an FPIF contract, in theory, shares overrun risk between the government and the contractor. But, in reality, risk does not decrease in many cases because of an immature design. While the contractor may be incentivized to control cost, the technology/ design issues can overwhelm even the best intentions of program managers.
  - A best practice is to use cost-plus contracts for targeted new technologies or in cases where the contractor is inexperienced.
- Uniqueness of a Program: A best practice is to treat each contract within a program as an individual action, based on its specific risk profile and not broad guidance. That is, influence the outcome by making decisions based on a particular situation or set of circumstances. In the U.S. DoD, and particularly the Department of the Navy (DoN), there is an enormous amount of institutional or cultural bias to overcome in awarding certain contract types for certain technologies. The Navy has historically used FPIF contracts for shipbuilding. The Navy's FFG-62 DD&C contract is a good, recent example. If substantial risk is present, the choice is questionable. Focus acute attention on the use of concurrent development and production contracts.
  - CVN-79 and RMS contracts are good examples of concurrence that resulted in unintended consequences – program

---

13. Better Buying Power 2.0; Secretary Frank Kendall, USD(AT&L).

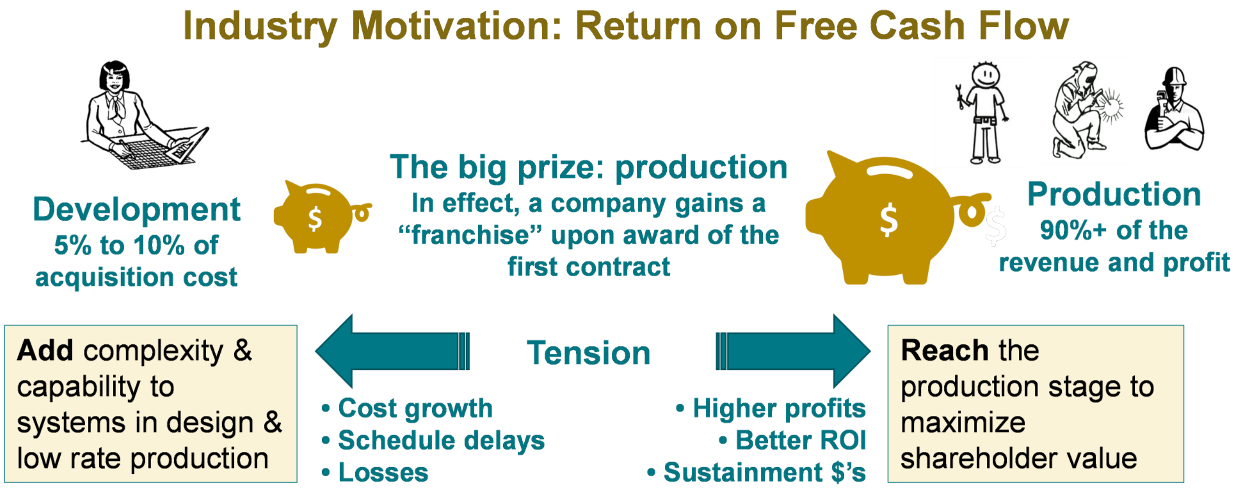


Figure 9: Cross-Phase Contract Tensions

  perturbation and the failure of incentives to work.

- Alignment with Industry Best Practices: Shipyards such as Fincantieri build both commercial and military vessels. On the commercial side, in their construction of cruise vessels at their yards on the Adriatic, they take great pains to eliminate as much risk as possible. They focus on perhaps one innovation at a time, and test the effectiveness of the new system (in concert with the buyer) before proceeding with ship construction.
  - Implementing best practices for defense projects, however, is very difficult due to the changing nature of the threat and the exacting requirements that follow.
- Contractor Motivation: There's an inherent tension between incentive provisions in development and low-rate production contracts versus cross-contract incentives downrange, as Figure 9 shows.

Given the cost of designing and developing complex weapon systems, coupled with limited competitive procurement and the cost of bringing onboard a second source, a company's winning bid early on (say at Milestone B) often implies the award of a "franchise" for the entire acquisition phase, and even into sustainment. This dynamic can and does impact a firm's strategic pricing perspective. If the firm adds complexity and capability to early design, it likely achieves higher unit price downrange. Examples are Joint Strike Fighter and DDG-1000 *Zumwalt* Class ships, or the "eight-billion-dollar boat," as it is sometimes dubbed. On the other hand, the additional complexity (especially if price validation is poor) increases the likelihood of cost growth, schedule delays, and contract losses during development.

A contractor's prime motivation is arguably to maximize the free-cash-flow return on invested capital for all contracts across all projects in the portfolio. This profit motive might induce the firm to trade short-term losses for future gains and could easily swamp the incentives of development contracts.

- Best Approach: Approaches, strategies and practices for future procurements include:
  - Assess the risks of the project and contract using the framework presented above.
  - Specifically, adjust the content of the contract, and set financial parameters, accordingly:
    - If Technology Maturity is high-end risky, consider moving the tasks into block upgrades rather than inclusion in the baseline
    - If Market Forces is too high, focus attention on more competition at the Tier 1 vendor level, if feasible

- Perform government Independent Cost Estimates (ICEs) early-on to validate costs
  - Make risk analysis and cost/capability tradeoffs (the knee in the curve) part of the analysis.

6.0 Operational Construct

The risk-scoring framework provides an analytical basis to support internal government and government-contractor deliberations on upcoming ship design and construction contracts. Application of the framework will help engender better-informed decisions related to choices of contract type and incentives – with the ultimate goal of increasing the effectiveness of the pricing approach at acceptable cost and risk to all parties.

The first step in making the framework operational is to establish a team to score the upcoming project/ contract(s), with representation from the requirements, engineering, and contracting communities. Participation by the contractor might be beneficial, too, per the discretion of government acquisition authorities. A cross-discipline approach helps ensure that all sources of risk are assessed thoroughly from a 360-degree project management and execution perspective.

Conduct a formal scoring session according to the following steps.

6.1 Collect Intelligence

Prepare for the scoring session by collecting information pertinent to choosing the contract types and incentives of the upcoming contract [14]. Data includes requirements documents, programmatic information, metrics on past contractor performance (including cost growth, schedule slippages, and cash flow), and benchmark risk scores as presented in this paper.

Obtain the details of risk scores for best-fit analogies – to include not only the raw numbers but the rationale behind them.

6.2 Evaluate Evidence

Share and explain details of the upcoming contract to the group of scorers to help ensure a minimum-common-denominator degree of understanding. Vigorous open discussion of prospective values of category weights and risk scores will strengthen the integrity of the exercise.

6.3 Establish Weights

Establish the weights of each of the six risk categories in a formal scoring session, where: there are k participants who make individual choices: w_1 = the weight for Stability of Requirements; w_2 = the weight for Market Forces; …; and w_6 = the weight for Schedule Challenge.

Expanding this notation, the second subscript in the term w1i represents the input from the i^{th} scorer for the first weight. That is, w_{11} is Scorer #1's input for weight w_1, w_{21} is Scorer #1's input for weight w_2, and so on.

Compute first and second moments (mean and variance) of the probability distribution for scoring the weights accordingly.

---

14. This step is akin to the military function of "Intelligence Preparation of the Battlespace."

Means:

$\mu_{w1} = \sum_{i=1}^{k} w_{1i}/k$ or the average value of the first weight, w_1, across the k scorers

$\mu_{w2} = \sum_{i=1}^{k} w_{2i}/k$ or the average value of the second weight, w_2, across the k scorers

$\vdots$

$\mu_{w6} = \sum_{i=1}^{k} w_{6i}/k$ or the average value of the sixth weight, w_6, across the k scorers

Variability:

σ_1 = standard deviation of the k scores for the first weight, or the observations $w_{11}, w_{12}, \ldots, w_{1k}$

σ_2 = standard deviation of $w_{21}, w_{22}, \ldots, w_{2k}$, and so on.

Coefficients of Variation:

$CV_1 = \sigma_1/w_1,\ CV_2 = \sigma_2/w_2,\ \ldots,\ CV_6 = \sigma_6/w_6$ [15]

The mean estimate of each weight is a measure of its relative importance or influence within the set of six categories of risk (e.g., Stability of Requirements versus Schedule versus Price Validation). A CV, on the other hand, is a measure of degree of consensus in the assessment of influence. The lower the CV, the stronger the consensus, and with a value of zero indicating unanimity.

For example, the mean estimated weight μ_{w1} for Stability of Requirements might be 20%. But the uncertainty of this estimate might be relatively high, with a CV of say 50%, compared to the other five CVs ranging from, say, 15% to 25%. The 50% CV represents a significant difference of opinion amongst the scorers. This might be due to factors such as scorers' unique perspectives or varying degrees of knowledge and experience. Additional group discussion in such cases will pay dividends in terms of a richer understanding of the risks that influence contract outcomes.

6.4 Score the Contract

With category weights established using mean values (μ_{w1}, μ_{w2},…,μ_{w6}), the next step is to generate a score for each of the risk categories for the upcoming contact, using ratio scales from 1.0 to 2.0, with anchors provided in 0.25 increments, as proposed in this study [16]. The mathematical procedure is the same as scoring the weights; i.e., compute means, standard deviations, and CVs.

6.5 Actionable Intelligence

The scoring results represent data-borne information or actionable insight that helps improve the effectiveness of pricing strategy by illuminating elements of risk that influence outcomes. Better understanding of risk, in turn, leads to better selections of contract types, incentives, and methods of payment.

The scoring results inform choices for the upcoming contract based on the following:

- The aggregate or total-contract risk score
- The Impact Factors of each of the risk categories
- A comparison to benchmarks such as averages and analogies.

---

15. A *CV* is a probability distribution's standard deviation divided by its mean. CVs can be thought of as the reciprocal of a signal-to-noise ratio. They are independent of unit of measurement, allowing for comparisons across probability distributions, which in this case are those for the six risk categories.

16. It is left as an exercise to the reader to show that the six mean values will sum to 100%. (Even if they didn't, they could be normalized to do so.)

An Impact Factor is similar to a Beta coefficient in regression analysis – it allows comparisons of the effect of risk scores, across the six categories, on a contract outcome. The higher the Impact Factor, the more influential is the element of risk.

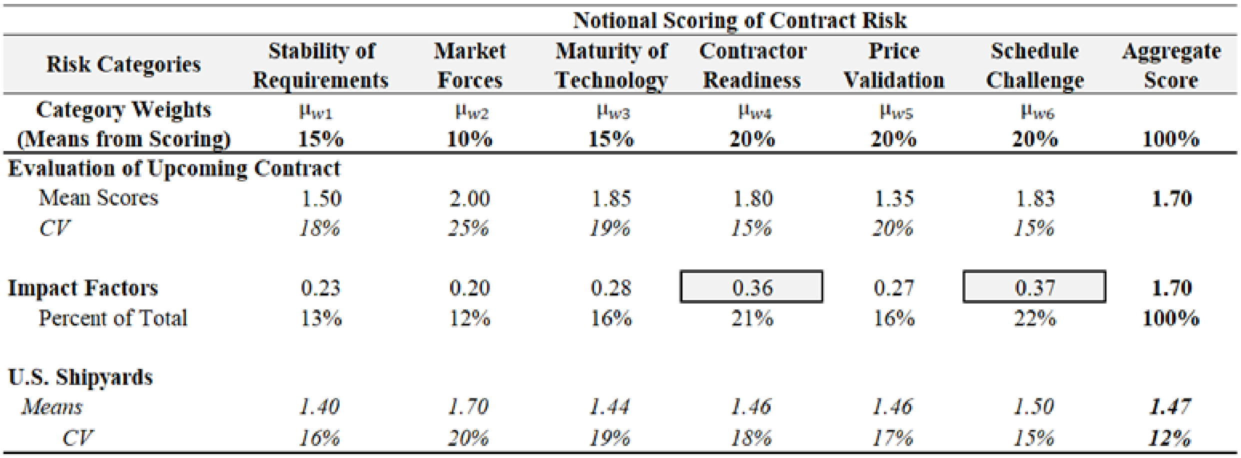
$$\text{Impact Factor} = \text{Category Weight} \times \text{Risk Score}$$

In the notional example of Table 3, two Impact Factors stand out, Contractor Readiness and Schedule Challenge. They influence the degree of overall contract risk more than any of the other elements, or 43% in total.

This is actionable intelligence which addresses where to apply incentives to diminish project risk using contract types such as cost-plus award fee (CPAF), cost-plus incentive fee (CPIF), and fixed priced incentive (FPI). For maximum leverage, it is better to apply incentives for the "big-ticket," more impactful elements of Schedule and Readiness rather than the "lower-hanging fruit," less impactful elements of Requirements and Market Forces, in the notional example.

- Schedule: Impact factor of 0.37. Reduce risk by rewarding the shipyard to meet schedule using metrics such as the following:
  - Threshold and objective calendar dates for each incentivized milestone
  - Design milestones such as a preliminary design review and critical design review
  - Construction milestones such as percent complete.
- Contractor Readiness: Impact factor of 0.36. Reduce risk by rewarding the shipyard to improve readiness using metrics such as the following:
  - Percent complete for (detail) design prior to commencement of construction
  - Percent vacant jobs filled for hard-to-fill professions and trades such as naval engineers and electricians
  - Demonstrated improvements to manufacturing processes.

The risk elements of Maturity of Technology and Price Validation might be addressed, too, as secondary considerations, since each represents 16% percent of total impact.



Notional Scoring of Contract Risk

Risk Categories	**Stability of Requirements**	**Market Forces**	**Maturity of Technology**	**Contractor Readiness**	**Price Validation**	**Schedule Challenge**	**Aggregate Score**
Category Weights (Means from Scoring)	μ_{w1} **15%**	μ_{w2} **10%**	μ_{w3} **15%**	μ_{w4} **20%**	μ_{w5} **20%**	μ_{w6} **20%**	**100%**
Evaluation of Upcoming Contract							
Mean Scores	1.50	2.00	1.85	1.80	1.35	1.83	**1.70**
CV	*18%*	*25%*	*19%*	*15%*	*20%*	*15%*	
Impact Factors	0.23	0.20	0.28	0.36	0.27	0.37	**1.70**
Percent of Total	13%	12%	16%	21%	16%	22%	**100%**
U.S. Shipyards							
Means	*1.40*	*1.70*	*1.44*	*1.46*	*1.46*	*1.50*	***1.47***
CV	*16%*	*20%*	*19%*	*18%*	*17%*	*15%*	***12%***

Table 3: Notional Scoring of Contract Risk

- Maturity of Technology: Reduce risk by incentivizing the shipyard to do the following:
  - Achieve incremental improvements to Technology Readiness Levels (TRLs) and Manufacturing Readiness Levels (MRLs) according to plan
  - Invest in test-beds during the Engineering and Manufacturing Development (EMD), and certainly before construction
  - Experiment with more than one technology as a contingency measure [17].
- Price Validation: Reduce risk by incentivizing the shipyard to help validate price by doing the following:
  - Presenting forward-pricing labor rates and overhead rates up to six years out, based on alternative outyear workload scenarios, **and** including justification.
  - Presenting fully documented shipyard cost estimates that meet U.S. and NATO standards [18].

The results of the scoring session also support the fundamental consideration of selection of the contract type [19]. The total risk score applies here, which is the sum of the six Impact Factors:

$$Total\ Risk = \mu_{w1} \cdot Score_{w1} + \mu_{w2} \cdot Score_{w2} + \cdots + \mu_{w6} \cdot Score_{w6}$$

Risk decreases as a program progresses through the Adaptive Acquisition Framework and into sustainment, as Figure 10 shows [20].

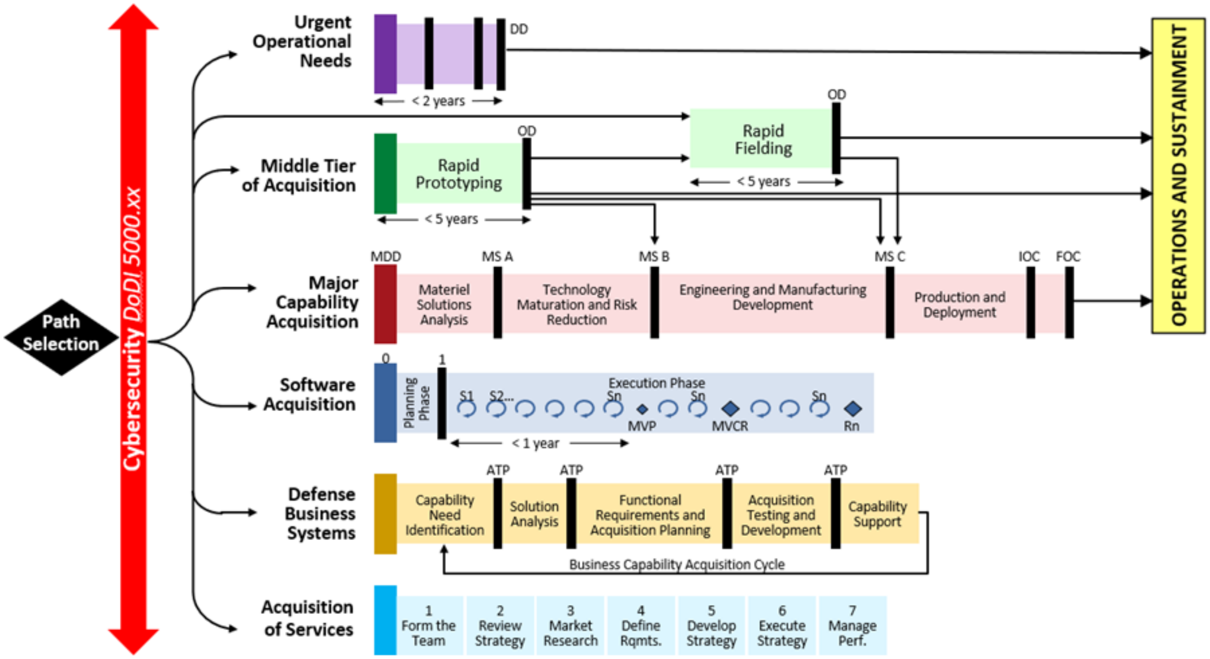


Figure 10: Adaptive Acquisition Framework

---

17. See case studies on *Ford* Class carriers and Remote Minehunting System.

18. NATO Research and Technology Organization (RTO) Technical Report "Methods and Models for Life Cycle Costing (Méthodes et modèles d'évaluation du coût de possession)," June 2007.

19. Or more technically, types (plural) at the Contract Line Item Number (CLIN) level within the contract.

20. https://aaf.dau.edu/

The diminution of risk is corroborated by cost growth studies in the U.S. [21]. For U.S. contracts (and amongst NATO partners and Australia), cost-reimbursable is the common contract type early-on; that is, up until full-scale production of ships. However, this is not always the case (e.g., FFG-62 *Constellation* Class frigates). Further, in some cases, cost-reimbursable was used for several ships in the class, and a block upgrade (LPD-17 *San Antonio* Class). To complicate matters, a cost-reimbursable contract should have been used, in retrospect, for contracts such as the USS *Kennedy*, the second ship in the CVN-78 *Ford* Class carrier project.

Is there a numerical value that represents a tipping point in choosing between cost reimbursable and fixed-price incentive?

Based on a wide range of contracts and many ship acquisition programs, 1.5 seems a reasonable value. It represents a middle ground between low and high risk in the ratio scales of the scoring framework. That is, aggregate scores above 1.5 suggest the use of a cost-reimbursable contract while those below suggests the use of a fixed-price incentive. Contract ceilings (maximum expenditures) might be invoked in cases of high risk where the scores approach 1.7 or above.

The aggregate risk score influences the choice of the contract type, and the impact factors influence the application of incentives.

Importantly, however, this value is a rough-order-of-magnitude metric. Future research will focus on additional ex-post scoring of contracts and offer contract metrics for each of the phases of acquisition. In addition, subsequent research will present metrics probabilistically by conflating distributions of the scoring process.

As a corollary, the program should strive to drive down risk based on details of the scoring. For example, if Technology Maturity is greater than 1.7, then consider ways to mitigate the risk such as the use of test beds or an evolutionary approach to acquisition.

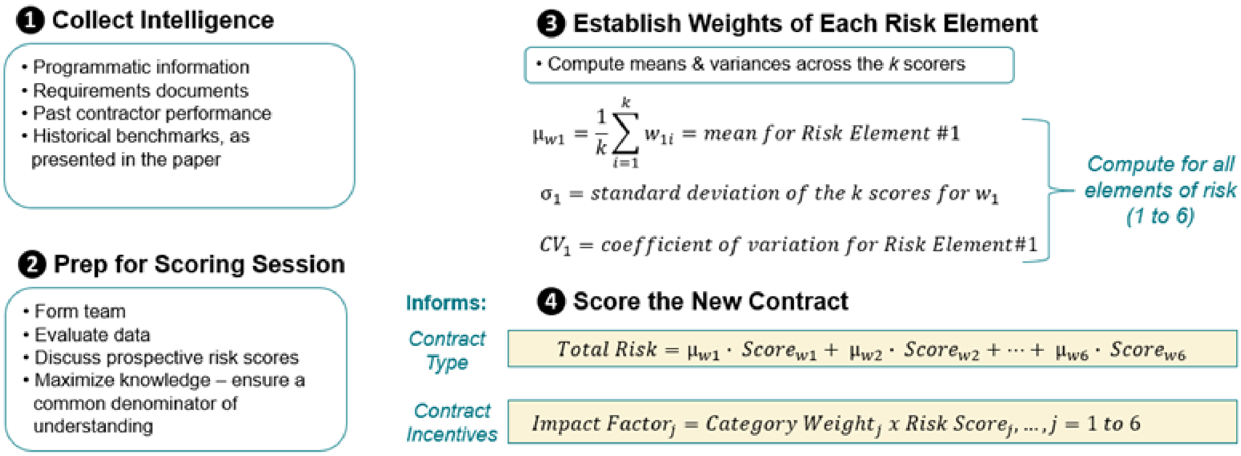


Figure 11: Contract Risk-Assessment Model

---

21. Weapon Systems Acquisition Reform Act (WSARA), 2009; "Enhanced Scenario-Based Method (eSBM) for Cost Risk Analysis," Journal of Cost Analysis and Parametrics, 2012, Garvey, Braxton, Flynn, Lee.

7.0 Summary

It must be cautioned that there is no "silver bullet" contracting solution guaranteed to produce optimal outcomes in all situations, even when restricted to the fairly uniform case of sole-source programs or programs in oligopolistic markets.

There is a common aphorism in the golf world, "You can't win a Major on Thursday, but you can lose a Major on Thursday" [22]. The import is that those who shoot well in the opening round are often overtaken over the course of the remaining three rounds by better (and steadier) players only a few shots back, but that a poor opening round can doom a player, even the best, by digging too deep of a hole to climb out of. The analogy is that a poor choice of contract type, incentives, and methods of payment may doom a project to failure, but even an optimal choice will not guarantee success.

As an industry executive once opined during an ICEAA conference, "You can't manage your way out of a bad deal." To minimize this possibility, the contract risk-assessment model, summarized in Figure 11, increases the odds of a win-win outcome for government and industry through a sound, statistical selection of contract type and incentives.

This research tilts the odds in the golfer's favor (carrying the analogy one step further) and represents a significant advance in the application of sound, data-driven metrics to better illuminate contract risk and engender more informed pricing strategies.

8.0 Next Steps

Despite the important advance that the research described in this paper represents, there is additional research and analysis to be accomplished in the interest of better contracting and acquisition decisions.

A more direct linkage between the risk scores and specific incentive contracting parameters (particularly, share line and ceiling price) is achievable if the risk scores could be translated into CGF and CV for the estimate itself. This has been done once before in the so-called BMDO Risk Model, cited in CEBoK Module 9 as the Historical Outputs-Based Model [23]. An updated application for Ship programs would involve the melding of the CGFs in Appendix 2 and the Risk Scores in Appendix 4. One immediate challenge is the former, being derived from SARs, are at the program level, whereas the latter are at the contract level. The authors are considering this research for presentation at a future ICEAA conference.

Regardless of whether a CGF and a CV are derived from a historical model as just described or are a result of the independent estimate for a new contract, the linkage between that stochastic cost estimate and proposed contract parameters can be achieved using the previously published Risk-adjusted Contract Price Methodology (RCPM). Rather than repeat that material here, an illustrative example of the unfortunate consequences of a mismatch between program risk and contract parameters is included in Appendix 5.

---

22. The annual quartet of Major tournaments comprises the Masters, the PGA Championship, the Open Championship (aka the British Open), and the U.S. Open. The opening round traditionally occurs on Thursday, with the remaining three rounds continuing throughout the weekend.

23. The Ballistic Missile Defense Organization (BMDO) was the successor to the Strategic Defense Initiative Organization (SDIO) – Reagan's so-called "Star Wars" program – and predecessor of the Missile Defense Agency (MDA).

***Brian Flynn** supports national security efforts at Technomics' Arlington, Virginia, headquarters. He has a deep understanding of and strong working relationships with many of the cost and operations research organizations in the U.S. DoD and the NATO Alliance. His areas of expertise include capability portfolio analysis and AoAs, weapon system and software cost estimating – including the cloud, EVM, econometric modeling, and corporate financial health. Dr. Flynn has undying pride in his service as a Marine.*

***Peter Braxton** is a Subject Matter Expert and Employee Owner at Technomics, Inc. He has over 20 years of experience performing cost and risk analysis and delivering associated training for a broad spectrum of federal government clients. The inaugural VP for Professional Development and a multiple ICEAA Educator of the Year winner, he has shown a long-standing commitment to knowledge sharing within the community. He holds an AB in Mathematics from Princeton and an MS in Operations Research from the College of William and Mary. He is a semi-retired game show contestant and avid cruciverbalist.*

***Robert. Nehring** is a project manager for Technomics, where he has developed many tools, data visualizations, databases, and cost estimates for programs throughout DoD and DHS. He is an ICEAA-Certified Cost Estimator/Analyst who has won many awards, including the ICEAA 2013 Cost Estimator/ Analyst of the Year for Technical Achievement and Best Conference Paper at the 2012 SCEA/ISPA Workshop. He also developed the innovative Galaxy Chart concept and many other models and tools throughout his career.*

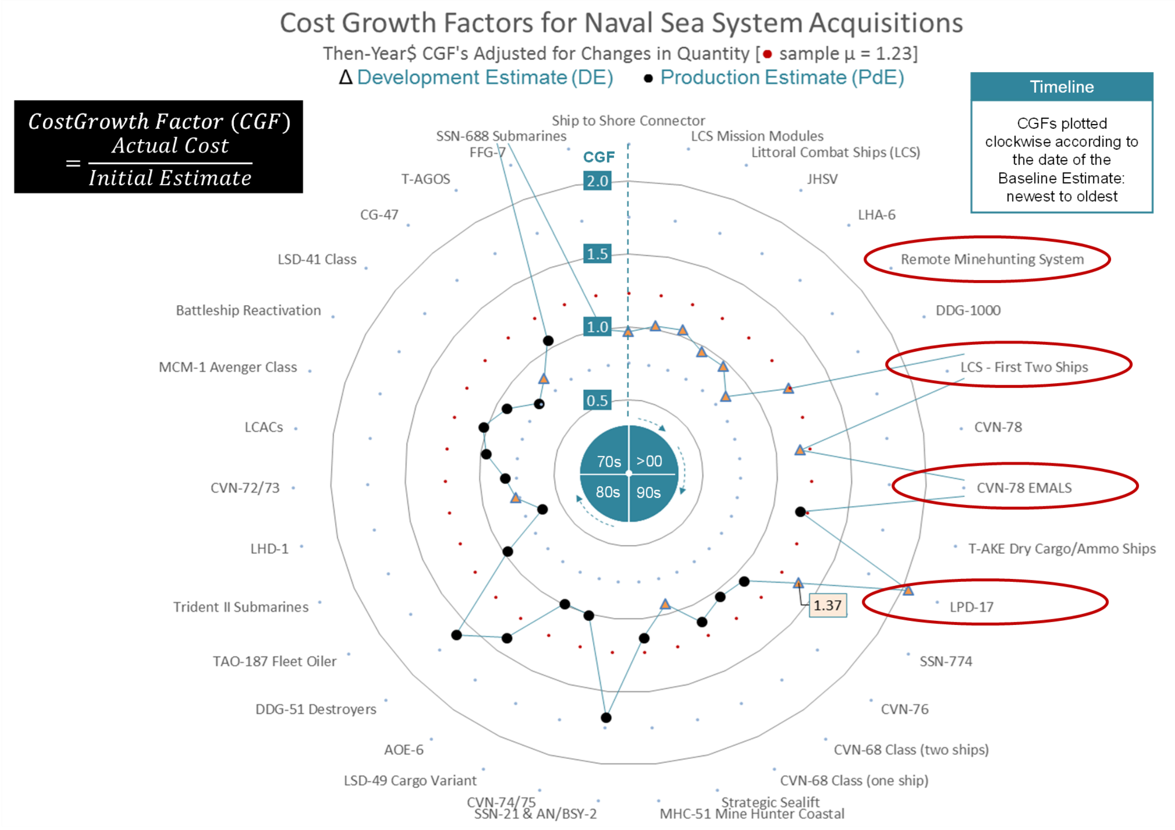
Appendix 1: Consolidation of the Defense Industrial Base [24]

Weapons Category	Total U.S. Contractors			Current U.S.-Based Prime Contractors
	1990	1998	2023	
Tactical Missiles	13	3	3	Boeing, Raytheon Technologies, Lockheed Martin
Fixed-Wing Aircraft	8	3	3	Boeing, Northrop Grumman, Lockheed Martin
Expendable Launch Vehicles	6	2	2	Boeing, Lockheed Martin
Satellites	8	5	4	Boeing, Lockheed Martin, Hughes, Northrop Grumman
Ships and Submarines	8	5	4	General Dynamics, Fincantieri Marinette, Huntington Ingalls, Austal
Tactical Wheeled Vehicles	6	4	3	AM General, Oshkosh, General Motors
Tracked Combat Vehicles	3	2	1	General Dynamics
Strategic Missiles	3	2	2	Boeing, Lockheed Martin
Torpedoes	3	2	2	Lockheed Martin, Raytheon Technologies
Rotary Wing Aircraft	4	3	3	Bell Textron, Lockheed Martin (Sikorsky), Boeing

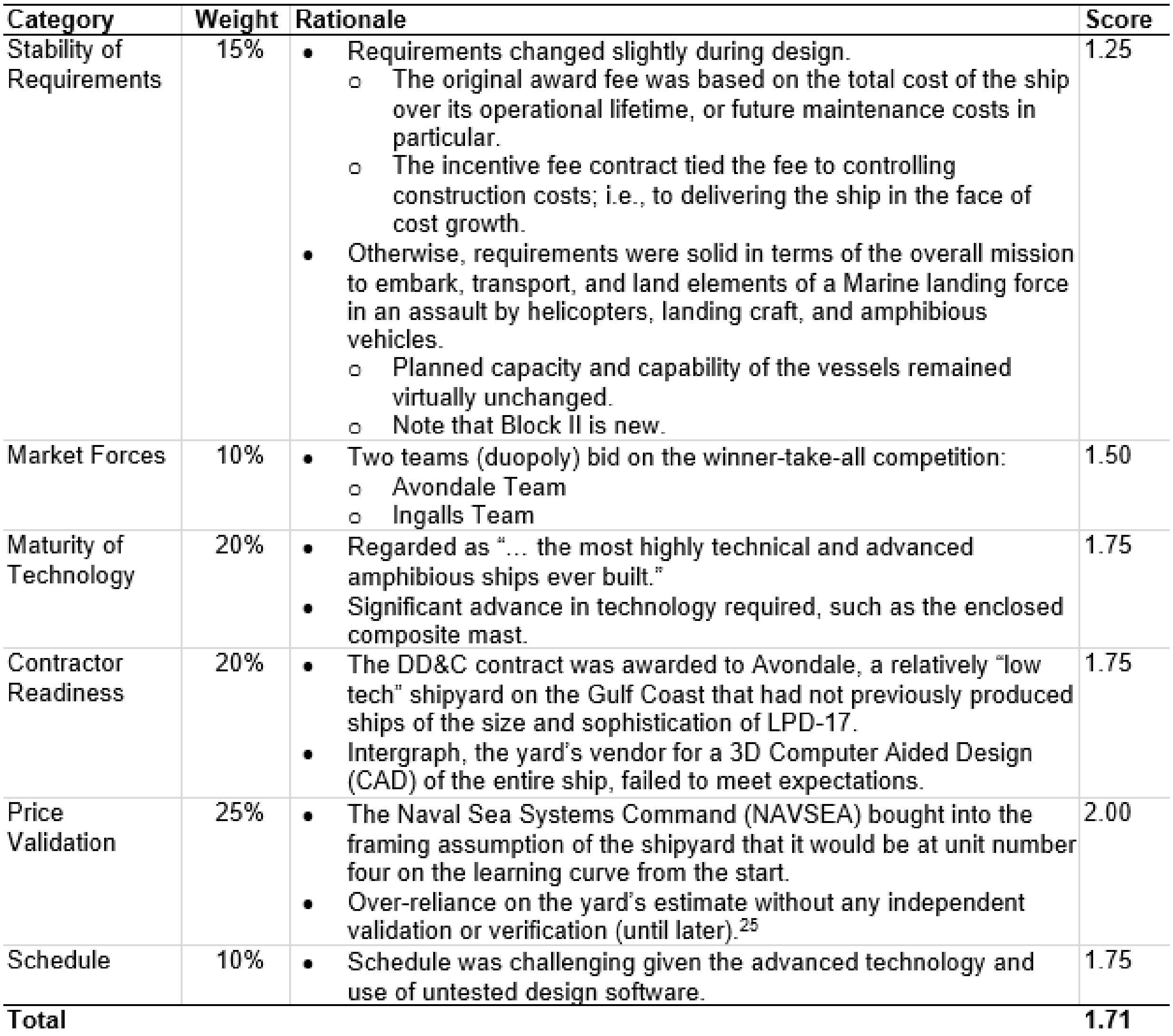
---

24. Source: Undersecretary of Defense (Acquisition and Sustainment).

Appendix 2: Cost Growth on U.S. Ship Contracts



Appendix 3: LPD-17 Scoring Details



Category	Weight	Rationale	Score
Stability of Requirements	15%	• Requirements changed slightly during design. o The original award fee was based on the total cost of the ship over its operational lifetime, or future maintenance costs in particular. o The incentive fee contract tied the fee to controlling construction costs; i.e., to delivering the ship in the face of cost growth. • Otherwise, requirements were solid in terms of the overall mission to embark, transport, and land elements of a Marine landing force in an assault by helicopters, landing craft, and amphibious vehicles. o Planned capacity and capability of the vessels remained virtually unchanged. o Note that Block II is new.	1.25
Market Forces	10%	• Two teams (duopoly) bid on the winner-take-all competition: o Avondale Team o Ingalls Team	1.50
Maturity of Technology	20%	• Regarded as "… the most highly technical and advanced amphibious ships ever built." • Significant advance in technology required, such as the enclosed composite mast.	1.75
Contractor Readiness	20%	• The DD&C contract was awarded to Avondale, a relatively "low tech" shipyard on the Gulf Coast that had not previously produced ships of the size and sophistication of LPD-17. • Intergraph, the yard's vendor for a 3D Computer Aided Design (CAD) of the entire ship, failed to meet expectations.	1.75
Price Validation	25%	• The Naval Sea Systems Command (NAVSEA) bought into the framing assumption of the shipyard that it would be at unit number four on the learning curve from the start. • Over-reliance on the yard's estimate without any independent validation or verification (until later).[25]	2.00
Schedule	10%	• Schedule was challenging given the advanced technology and use of untested design software.	1.75
Total			**1.71**

---

[25] Dr. Flynn assisted in the development of an Independent Cost Estimate (ICE) of the acquisition program after problems surfaced.

Appendix 4: Risk Scores for U.S. Naval Programs and Contracts

Project Parameters		Contract Risk Scores Based on Ratio Scales						
Ship Class and Contract	Contract Type	Stability of Requirements	Market Forces	Maturity of Technology	Contractor Readiness	Price Validation	Schedule Challenge	Aggregate Score
LPD-17 *San Antonio* Class (Amphibious Transport)								
Lead-Ship DD&C	CPAF → CPIF	1.25	1.50	1.75	1.75	2.00	1.75	**1.71**
LPD-18	CPIF	1.25	1.50	1.75	1.50	2.00	1.75	**1.66**
Steady-State LPD-22	FPIF	1.10	2.00	1.00	1.00	1.25	1.25	**1.20**
1st Block II: LPD-30	CPFF	1.25	2.00	1.25	1.25	1.25	1.25	**1.33**
2nd Block II: LPD-31	FPI	1.10	2.00	1.00	1.10	1.25	1.25	**1.22**
FFG-62 *Constellation* Class (Surface Combatant)								
Lead-Ship DD&C	FPI Firm Tgt	1.50	1.00	1.25	1.70	1.75	1.50	**1.50**
FFG-63	Option	1.25	1.00	1.10	1.30	1.75	1.25	**1.33**
Steady-State FFG-64	Option	1.10	1.00	1.00	1.20	1.75	1.20	**1.26**
DDG-1000 *Zumwalt* Class (Surface Combatant)								
DD(X) EMD	CPAF	2.00	1.50	1.75	1.75	1.50	1.50	**1.68**
System Design & Int'n	CPAF	1.75	1.50	1.75	1.75	1.50	1.50	**1.64**
Lead-Ship DD&C: BIW	CPAF	1.75	1.75	1.75	1.75	1.25	1.50	**1.60**
DD&C: Ingalls	CPAF	1.75	1.75	1.75	1.50	1.25	1.50	**1.55**
DDG-1001 Construction	FPI	1.50	1.75	1.50	1.25	1.25	1.25	**1.39**
DDG-1002 Construction	FPI	1.25	1.75	1.25	1.20	1.25	1.25	**1.29**
Advanced Gun System (AGS)								
Initial Design and Build	CPAF	1.75	1.75	1.75	1.75	1.50	1.75	**1.69**
AGS for DDG-1002	FPI	1.25	1.75	1.25	1.10	1.25	1.25	**1.27**
SSN-774 *Virginia* Class (Fast Attack Submarine)								
1st Boat in Block I	CPFF	1.40	1.75	1.35	1.40	1.25	1.60	**1.41**
1st Boat in Block II	FPIF	1.30	1.75	1.30	1.30	1.20	1.50	**1.34**
1st Boat in Block III	FPI Firm Tgt	1.25	1.75	1.25	1.35	1.10	1.50	**1.31**
1st Boat in Block IV	FPI Firm Tgt	1.25	1.75	1.15	1.25	1.00	1.40	**1.23**
1st Boat in Block V	FPIF	1.30	1.75	1.25	1.50	1.00	1.60	**1.33**
RMS Multi-Mission Vehicle (Mine Reconnaissance)								
Engineering Dev Model	CPFF	1.60	2.00	1.80	1.80	1.50	2.00	**1.74**
Initial Production (LRIP)								
Hardware	FFP	1.60	2.00	1.80	1.80	1.50	2.00	**1.74**
Engineering Svcs	CPFF	1.60	2.00	1.80	1.80	1.50	2.00	**1.74**
CVN-78 *Ford* Class (Nuclear Aircraft Carrier)								
"CVN-21" Constr Prep	CPIF, CPAF, CPFF	1.50	2.00	1.75	1.50	1.25	1.75	**1.56**
EMALS SDD	CPAF	1.50	2.00	2.00	2.00	1.75	1.75	**1.84**
Advanced Arresting	CPAF	1.50	2.00	1.75	1.75	1.75	1.75	**1.74**
Lead-Ship DD&C	CPIF, CPAF, CPFF	1.50	2.00	1.60	1.60	1.50	1.50	**1.59**
CVN-78 Prod'n	FFP	1.75	2.00	1.50	1.50	1.50	1.50	**1.59**
CVN-79 Constr Prep	CPFF/CPIF	1.30	2.00	1.50	1.50	1.50	1.60	**1.53**
CVN-79 DD&C	FPIF	1.20	2.00	1.40	1.30	1.60	1.40	**1.46**
CVN-79 & 80 Prod'n	FFP	1.25	2.00	1.40	1.30	1.30	1.40	**1.39**
T-AGS-66 *Pathfinder* Class (Oceanographic Survey)								
T-AGS-60	FFP	1.50	1.25	1.25	1.25	1.50	1.50	**1.38**
T-AGS-61 to T-AGS-65	FFP	1.20	2.00	1.25	1.00	1.50	1.25	**1.33**
T-AGS-66 and 67	FFP	1.20	2.00	1.00	1.00	1.50	1.25	**1.28**
LSD-41 *Whidbey Island* (Landing Ship Dock)								
LSD-41 DD&C	CPFF	1.50	1.25	1.50	1.60	1.50	1.50	**1.50**
LSD-42	CPAF → FPI	1.25	1.25	1.30	1.50	1.50	1.25	**1.37**
LSD-43	FPI	1.25	1.25	1.25	1.25	1.50	1.25	**1.31**
New Award (Avondale)								
LSD-44 to -48	FPI	1.20	1.00	1.25	1.75	1.80	1.50	**1.48**

Appendix 5: Calibration of Geometry for Incentive Contracts

An effective contract incentive structure relies on aligning Government and Contractor interests during execution, as illustrated in the bottom half of Figure 12. If the initial negotiations, wherein the parties' interest are naturally in opposition (as illustrated in the top half), fail to establish a reasonable Target Cost and other key parameters, then the program risks getting "stuck" in that Negotiation phase, with a continual parade of contract changes, instead of working together under the incentive mechanism in Execution.

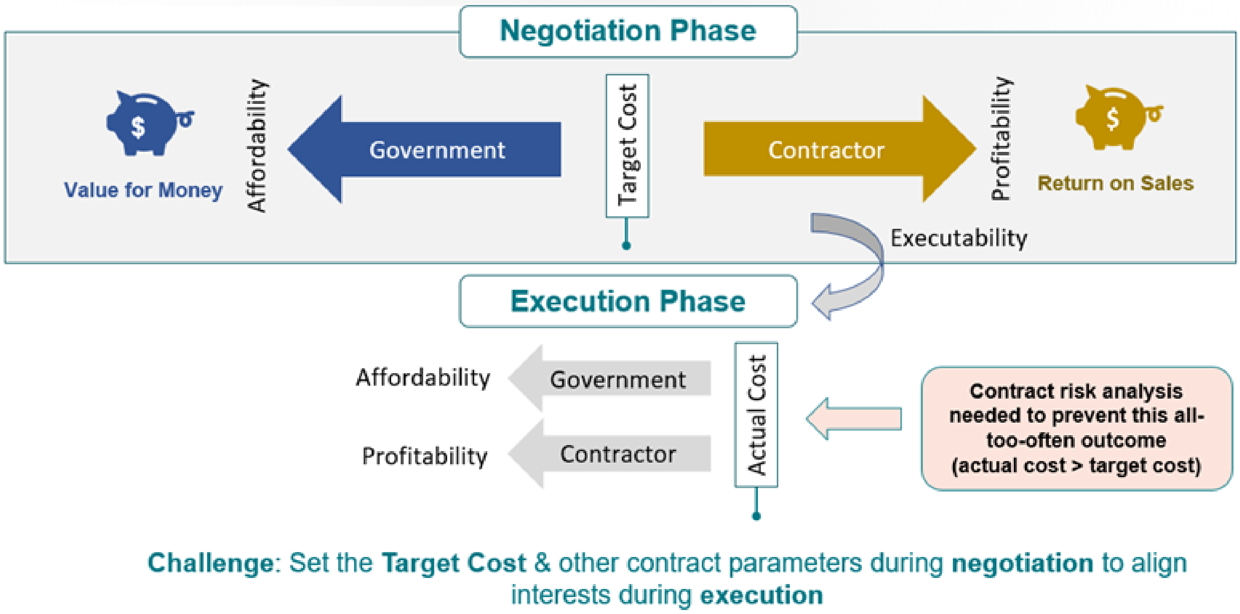


Figure 12: Contract Type as an Effective Contract Management Mechanism

This appendix provides a cautionary tale as to how not to set up an incentive arrangement and practical advice on how to avoid those potentially disastrous outcomes. In particular, it seeks to debunk the DPC default of a 50/50 shareline with 120% Ceiling Price for FPIF contracts.

The following metaphor is meant to illustrate the need for adequate mechanisms above target cost to encourage cost control while acknowledging that there is significant risk and uncertainty, especially for development and lead ship contracts. As shown in Figure 13, imagine a runaway truck barreling down the steep ROS curve from the favorable (underrun) outcomes on the left to the unfavorable (overrun) outcomes on the right. The illustration uses an FPI example with a target cost of $100M, target profit of 12%, ceiling price of 140%, and 80/20 and 70/30 sharelines over and under, respectively. (Graphics are generated using the Technomics Contract Incentive Impact Tool (CIIT), which is available upon request).

Appendix 5: Calibration of Geometry for Incentive Contracts

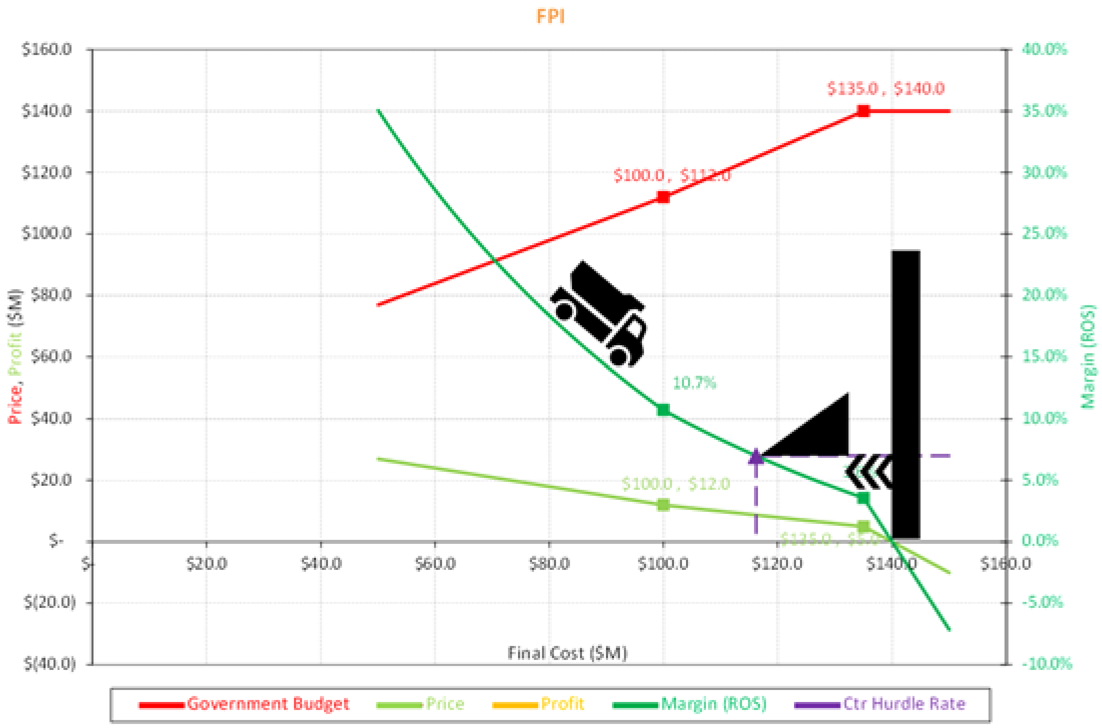


Figure 13: Runaway Truck with Safeguards

The first warning sign comes as the estimate at complete (EAC) passes target cost, forecasting an ROS of less than the 10.7% target. The truck starts applying its brakes, losing speed … and profit at 20 cents on the dollar of overrun. It's trying to stop short of the corporate hurdle rate of 7.0% ROS, which occurs at a final cost of $116.25M, or a 16.25% overrun, but the truck is too heavy and moving too fast. The driver steers off the motorway and onto a runaway truck ramp, which continues to slow the truck until the point of total assumption (PTA). In this case, that occurs at a final cost of $135M (35.0% overrun), and profit has eroded to $5M or an ROS of 3.57%, well below the hurdle rate but at least still positive. The truck has slowed significantly but is still in danger of crashing. The ground crew deploys caltrops (tire spikes) to shred the tires of the truck, and the truck lurches forward, now losing dollar for dollar of profit. It rumbles to a stop just short of the wall, the point at which profit disappears, a final cost of $140M (40% overrun).

By contrast, let's look at an FPI with only 10% target profit, 120% ceiling price, and 50/50 sharelines, as shown in Figure 14 below. This effectively makes the truck heavier and faster, the hill steeper and shorter, and the brakes less effective – a recipe for disaster! Barreling down the overrun hill, the truck is losing 50 cents on the dollar of profit and blows through the hurdle rate point at $104.3M (less than a 5% overrun). There is no time to deploy safeguards like a runaway ramp or caltrops. The driver desperately tries to apply the brakes, but they barely have an effect. Now the PTA and the point of zero profit are the same, at an actual cost of $120M (only a 20% overrun), and truck smashes into this wall and disintegrates into a fireball as the driver dives clear in a last-ditch attempt to avoid certain death.

Appendix 5: Calibration of Geometry for Incentive Contracts

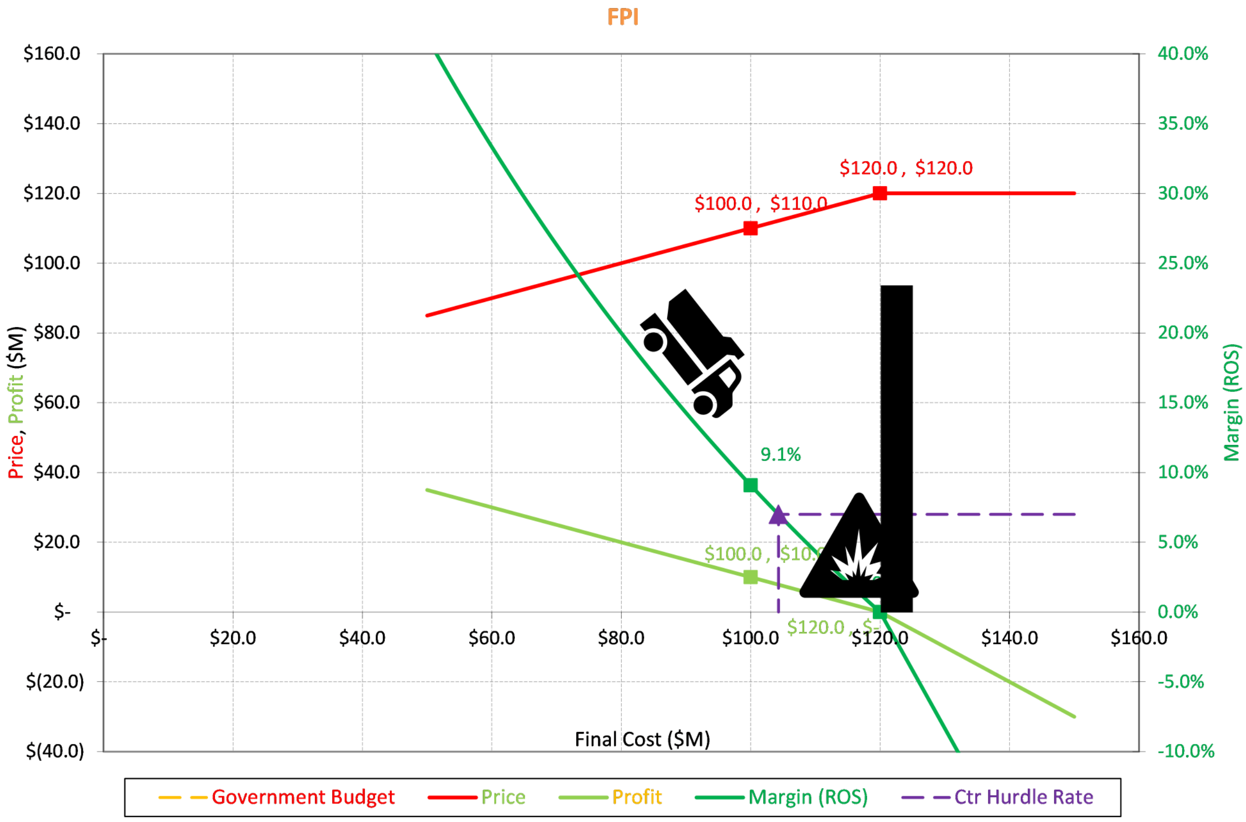


Figure 14: Runaway Truck without Safeguards

This narrative is a bit hyperbolic, but the scenario is still instructive. For a high-risk contract, steep sharelines (for the contractor), low target profit, and low ceiling price make for an unrealistically narrow range over which a cost control incentive is maintained for the contractor. If the government puts in place such punitive restrictions, it may prove to be self-defeating, as the contractor loses all motivation, and the project goes off the proverbial rails. The driver's bailing is symbolic of the project manager's literally quitting (or being fired) or mentally checking out. The government achieves a Pyrrhic victory, and any smugness at having negotiated such a parsimonious ceiling price evaporates upon the realization that both the delivery of the ship(s) and the health of the shipyard – a crucial component of the nation's Industrial Base – are at risk.

The 40% overrun in the previous scenario is not inevitable. Through a combination of prudent management and good fortune, we certainly hope to stop short of the first safeguard, or maybe even get the truck into reverse and end up to the left of target cost – a favorable underrun. Since the history of defense acquisition is littered with cost and schedule growth, however, it behooves us to put the safeguards in place.

Human Capital Impacts in Military Acquisitions

David K. Smith
Jonathan D. Ritschel
Brandon M. Lucas
Edward D. White
Shawn M. Valentine

The Department of Defense has historically struggled to control both cost and schedule growth within acquisitions programs. Many studies have investigated these issues, but very few have explored the impact of human capital in improving performance outcomes. This study performs contingency table analysis to evaluate the impact of personnel, base, and Acquisition Category (ACAT) on cost and schedule performance. The results of the study suggest that the program office estimating team composition has little to no impact on performance metrics. While it may be surprising to some that this research does not establish personnel manning as a significant driver of cost or schedule deviations, a closer examination of the findings reveals an encouraging narrative. The evidence suggests that personnel manning is currently done efficiently, representing a good news story. Given the lack of personnel significance, other factors should be investigated to control cost and schedule growth. The results also suggests that ACAT 3 programs are less likely to have performance issues of any kind, while base has no impact on performance. While the results are caveated by a small sample size, it is an important first examination of the issue especially as it relates to schedule.

The Department of Defense (DoD) has historically struggled to control both cost (Jones et al., 2023; Arena et al., 2006; Younossi et al., 2007) and schedule (Jones et al., 2023; Government Accountability Office, 2023; Riposo et al., 2014) growth within its' Major Defense Acquisition Programs (MDAPs). That stark reality has resulted in numerous research and government organizations such as RAND, the Institute for Defense Analysis (IDA), and the Government Accountability Office (GAO) among others to study the issue. The majority of these research efforts approach the cost issue by analyzing Defense Acquisition System (DAS) processes, overarching strategic decisions, and economic factors (Bolton et al., 2008; Lorrel et al., 2017). For schedule, the literature primarily focuses on risk, technological complexity, and cost estimates (Riposo et al., 2014; Monaco and White, 2005).

In contrast, the role of human capital in influencing cost and schedule outcomes has received little attention in previous studies. To the best of our knowledge, only two unpublished studies (Feuring, 2007; Gray, 2009) have directly researched cost and schedule growth along this dimension. This paper aims to bridge the gap by investigating the impact of human capital on cost and schedule outcomes. Without research on the impact of personnel on cost and schedule performance, it is difficult to draw conclusions about the effective allocation of our most valuable asset – people.

Cost and Schedule Performance in the DoD

Cost issues have plagued DoD Acquisitions for decades (Fox, 2011). Numerous studies have corroborated the claim. Arena et al. (2006) investigated 68 completed programs and found that the vast majority (70%) had a Cost Growth Factor (CGF) that was 1.25 or higher. Youssoni et al. (2007) expanded on the work of Arena et al. and found that the average CGF in DoD programs was 1.46. A decade later, Kozlak et al. (2017) reviewed 30 aircraft programs and discovered that the median development CGF for aircraft acquisition was 1.43. The results of all three studies suggests that cost growth is most prevalent in the development phase. More recently, D'Amico et al. (2018) found that cost performance has not changed from the previously poor performance, even in the modern era. Additionally, Jones et al. (2023) examined cost performance across four decades and found that while average CGF outcomes remained predominately the same over time, the standard deviations of CGFs have decreased, suggesting that cost estimators and/or the processes behind them may be improving over time. Additionally, Jones et al. (2023) examined cost performance across four decades and found that while average CGF outcomes remained predominately the same over time, the standard deviations of CGFs have decreased, suggesting that cost estimators and/or the processes behind them may be improving over time.

In investigating the sources of cost growth, Bolton et al. (2008) reviewed 35 mature programs and found that technological requirements, quantities, and production schedules were the most problematic sources of growth, all of which relate to decisions made by executive leadership. Lorell et al. (2017) found that extreme cost growth programs all exhibited the following traits: immature technology, unclear requirements, unrealistic estimates, and phase concurrency. Authors such as Cancian (2010), have investigated what could be done to improve outcomes. He highlighted early and accurate cost estimates, strategic restraint in program selection, and a focus on requirements as the most important principals in improving cost outcomes. More recently, McNicol (2022) sought to answer whether the literature has sufficiently identified the causes of cost growth. His answer was nuanced, suggesting that we know much about extreme (>100%) cost growth but do not have answers for those programs that fall in the 30-100% growth category.

Along with cost, schedule issues in the DoD have also been problematic (Jones et al., 2023). Hofbauer et al. (2011) found that for 98 MDAPs, programs were an average 22 months behind schedule from their initial estimate. The GAO (2023) found that of 26 MDAPs assessed, over half reported delays. Further, when these programs attempted rapid acquisition techniques, 37.5% of the programs were delayed by more than a year. Jones et al. (2023) assessed over 120 MDAPs and found that regardless of decade, schedules tend to have a mean Schedule Growth Factor (SGF) of approximately 1.20, varying about 0.2 standard deviations.

Searching for sources of schedule growth, Riposo et al. (2014) summarized over 100 sources to investigate schedule growth in the DoD and cited three major causes of schedule slippage: difficulty managing technical risk, poor initial assumptions regarding estimates or requirements, and funding instability. Monaco and White (2005) found the most prevalent variables in predicting schedule slippage were the instance of technical issues (negative impact), whether competition was utilized in contract sourcing (positive impact), and if the program utilized prototyping (positive impact).

Absent from all the aforementioned studies is an examination of human capital. Human capital is defined as “the skills, knowledge, and experience possessed by an individual or population, viewed in terms of their value or cost to an organization or country” (Oxford Languages Dictionary, 2023). Blundell et al. (1999) describes three main components of human capital: early ability, skills gained through formal education, and skills gained through work. These components are highly complementary as early ability tends to increase the likelihood for higher levels of formal education, which leads to higher levels of skills gained through work. Many established economic models which measure Gross Domestic Product (GDP) include components of human capital as explanatory variables (Mankiw et al., 1992: Krueger et al., 2001).

The power of specialized human capital is further harnessed through teams by utilizing each members’ specialized skills to generate synergies. Teams have greater combined previous experience, intellect, monitoring of mistakes, and information sharing (Kugler et al., 2012). Kugler et al. (2012) summarized 134 sources on the subject and found that teams consistently outperformed individuals. Teams outperformed individuals in a multitude of scenarios, including ultimatum games (Bornstein and Yaniv, 1998), signaling games (Cooper and Kagel, 2005), and mutual fund management (Bliss et al., 2008).

While teams tend to perform better than individuals, composition matters. Studies indicate team-training is important in producing improved performance (McEwan et al., 2017; Salas et al., 2008; Morey et al., 2002). The allocation of human capital is equally as important, and a widely used team role model constructed by Belbin (1993) integrates the impact of diverse personnel traits to improve performance outcomes. Even so, research on the impact of personnel allocation has been limited. In a review of the team effectiveness literature from 1997-2007, Mathieu et al. (2008) highlighted roles and allocation of team members as a field that needs more focus.

Even though human capital has shown to be impactful in improving performance outcomes in the economics literature (Mankiw et al., 1992; Kugler et al., 2012), there have only been two studies (Feuring, 2007; Gray, 2009) which have investigated the impact of human capital on DoD program performance. Feuring (2007) and Gray (2009) used Cobb-Douglas functions to analyze total portfolio yearly growth in conjunction with aggregated career-field demographic statistics. In both studies, teams with higher education levels tended to perform better. This finding is important (and expected), but it does not address the question of team composition.

Our paper seeks to fill the gap by examining how the composition of human capital in defense program office estimating teams affects cost and schedule performance. The development of a unique dataset that delineates individual program office estimating team personnel by functional category is the key new piece of information. This information allows us to examine, for instance, the cost growth of a specific program over time in conjunction with the human capital allocated to the estimating team. In contrast, the two previous studies (Feuring and Gray) did not have data broken down at the program level, they only had aggregates for the career-field writ large. It is important to emphasize that this paper is *only* looking at the composition of the *program office cost estimating team,* not the total program office personnel.

Data and Methods

Data

The data we used to conduct the analyses presented in this article originated from the Air Force Life Cycle Management Center in the form of Program Office Estimates (POEs) briefings. Within each brief, there are time-phased schedule milestone charts, budget outlays, Program Acquisition Unit Cost (PAUC) information, and team compositions. The POE data ranges from as early as 2000 and as late as 2022, with 2016 being the average. The number of POEs within each program ranges from one to eighteen, with an average of three. The POEs come from Wright-Patterson, Hanscom, and Eglin Air Force Base (AFB). Table 1 shows the initial number of POEs and Programs by base.

Table 1. POEs and Programs

Base	POEs	Programs
Wright-Patterson	702	250
Hanscom	268	105
Eglin	243	58
Total	1213	413

Of the 413 available programs, only 36 meet the criteria to be included in the analysis. The 166 programs that contain a single POE are not used because it is impossible to measure percent change across time with a single datapoint. Next, 80 programs that did not report personnel are removed. The next criteria remove 77 programs whose first POE occurs at 35% or greater into development, past where program performance stability has been established (Christensen and Heise, 1993; Kim et al., 2019). This allows an isolation of programs before other factors can drown out the impact of human capital on performance. The final two inclusion criteria are related to measuring schedule through actual production outlays. Forty-five programs are

Table 2. Inclusion Criteria

	Total	Wright-Patterson	Hanscom	Eglin
Starting Data Set	413	250	105	58
Single POEs	166	112	40	13
Missing Personnel Data	80	41	26	13
Started > 35% into Development	77	42	22	13
Incomplete Program	45	21	11	13
Phase Concurrency	10	6	3	1
Ending Data Set	**36**	**28**	**3**	**5**

removed because they do not move from development into production within the available POEs. Without entry into production, it is impossible to analyze when programs had planned production outlays in their first available POE and when those outlays actually occurred. This transition also highlights, for this study, the shift from development to production, defined by when production dollars are first spent. Lastly, ten programs, which are in the development and production phases simultaneously, are removed because entry into production has already occurred. Including these programs would invalidate the results because regardless of human capital impacts, these programs will always have perfect schedule performance by our measure. Table 2 shows the inclusion criteria.

Figure 1 shows the ACAT distribution by base. Programs at Wright Patterson (78%) are represented the most by far, with a total of 28. Hanscom (8%) and (14%) Eglin combine for a total of eight programs. ACAT 3 programs are most prevalent, with a total of 22. There are nine ACAT 2 programs and five ACAT 1 programs in the dataset.

Variables

Variables in this study are either directly reported or derived from information in the POEs. The directly reported variables include the personnel categories and associated number of people, the ACAT, and the base. *It is important to reiterate, that the number of people represents individuals who are identified by the program office as part of the program office cost estimate team. It is not the full count of personnel from the program office.*

The derived variables are average personnel percentage dichotomous variables, cost breach dichotomous variables, schedule breach dichotomous variables, and percent complete metrics. Program personnel information is discretized based on the first available POE, and performance metrics are discretized based on the difference between the first and last POE. The set of variables enables us to measure how personnel decisions in the early stages of programs impacts performance in the later stages of programs. Table 3 shows the variables used in this study, and Equation 1 shows how personnel information is discretized. [Note: If the Equation 1 statement is true, it is coded as 1; else it is coded as 0].

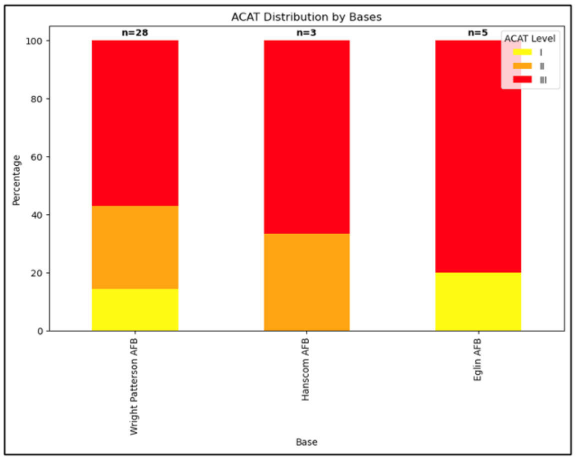


Figure 1. ACAT Distribution by Base

Table 3. Variables

Variable	Type	Values	Definition
Personnel Category	Given	- Cost Analyst - Budget - Program Management - Engineering - Contracting - Logistics	The categories outlined in the team composition slides within POEs
ACAT	Given	- ACAT 1 - ACAT 2 - ACAT 3	The ACAT category of the program as stated in the POEs
Base	Given	- Wright Patterson - Eglin - Hanscom	The base the POE originated from
Mean % Personnel Dichotomous Variable	Derived	1, 0	Indicates if the programs first POE had a personnel percentage higher than the total first POE personnel percentage average
Significant Cost Breach Dichotomous Variable	Derived	1, 0	Indicates if the PAUC grew more than 15% from the initial POE estimate
Critical Cost Breach Dichotomous Variable	Derived	1, 0	Indicates if the PAUC grew more than 30% from the initial POE estimate
Significant Schedule Breach Dichotomous Variable	Derived	1, 0	Indicates if the production milestone entry grew more than 15% from the initial POE estimate
Critical Schedule Breach Dichotomous Variable	Derived	1, 0	Indicates if the production milestone entry grew more than 30% from the initial POE estimate
% Complete	Derived	Ranges from 0%-100%	The percent of budget dollars executed divided by the final budget dollars obligated

$$IF(\,\%\ of\ Personnel\ Category_{First\ POE} > Mean\ \%\ of\ Personnel\ Category_{Total} \quad (1)$$

To arrive at both significant and critical cost breach dummy variables, PAUC percent growth metrics for significant (1.15 CGF) and critical breaches (1.3 CGF) are used. These thresholds mimic the Nunn-McCurdy current acquisition program baseline (APB) breach criteria. It is important to note the ultimate goal of the study is to measure cost and schedule performance, not instances of breaches. Nunn McCurdy threshold percentages are only used as an understandable ruler to measure performance. Equation 2 and 3 shows how cost breach information is discretized. In the case of schedule, projected years from entering production is compared to the actual years it took to enter production. This measure of schedule performance differs from most studies because it is based only on entry into production rather than the entire program. Actual production entry is known for all programs due to the inclusion criteria discussed above. For significant and critical schedule breaches, the same percent breakpoints (15% and 30%) are used. Equation 4 and 5 shows how schedule breach information is discretized. [Note: If the Equation statement is true, it is coded as 1; else it is coded as 0].

$$IF(\frac{PAUC_{Latest\ POE} - PAUC_{Earliest\ POE}}{PAUC_{Earliest\ POE}} + 1 > 1.15, 1, 0)$$

(2)

$$IF(\frac{PAUC_{Latest\ POE} - PAUC_{Earliest\ POE}}{PAUC_{Earliest\ POE}} + 1 > 1.3, 1, 0)$$

(3)

$$IF(\frac{Actual\ Production\ Entry - Production\ Entry\ Estimate_{\ Earliest\ POE}}{Production\ Entry\ Estimate_{Earliest\ POE}} + 1 > 1.15, 1, 0)$$

(4)

$$IF(\frac{Actual\ Production\ Entry - Production\ Entry\ Estimate_{\ Earliest\ POE}}{Production\ Entry\ Estimate_{Earliest\ POE}} + 1 > 1.3, 1, 0)$$

(5)

To illustrate the dichotomous variable schema in Equations 1-5, an example is provided. In one of the programs, the first POE indicated that 22% of the program office estimating team were Cost Analysts. Because this was less than the total dataset average of 23.4%, it received a 0, representing an underabundance of Cost Analysis personnel. The same logic was applied to the other five personnel categories. For performance metrics within the same program, the first POE (2018) estimated a PAUC of $4.30K (Constant Year dollars) and estimated that the first production dollars would be spent in 2020. The most recent POE (2022) highlighted a PAUC of $5.07K and that the first production dollars were spent in 2022. The percentage cost growth from $4.30K to $5.07K is 117.91%, and the project production entry estimate growth of two years to four years is 200%. Given this information, the program would receive a 1 for significant cost breaches and a 0 for critical cost breaches because 117.91% is greater than 115% but less than 130%. For schedule dichotomous variables, the program would receive a 1 in both significant and critical schedule breaches because 200% is greater than both 115% and 130%.

Lastly, a percentage complete metric is necessary to define the point where programs are when measuring changes. Percent complete is measured by dividing the percent of the budget executed by the total budget obligated. Due to the fact that budget obligations in programs that are experiencing cost growth typically grow across time, the amount of money spent in the first POE was divided by the latest budget obligation shown in the last POE rather than the perceived budget obligation in the first POE. Equation 6 highlights this nuance.

$$\frac{Budget\ Executed_{First\ POE}}{Budget\ Obligated_{Last\ POE}} \quad (6)$$

With the personnel categories, cost performance metrics, schedule performance metrics, base, and ACAT information put into categorical variables, contingency tables are then used to assess whether these categories are related to cost and schedule performance.

Method

Contingency tables are a tabular representation of categorical data that displays the frequency or count of occurrences corresponding to specific combinations of categories from the variables. From these tables, both the Pearson Chi-squared and Fisher's exact test can be used to assess if the outcomes in the table are associated with each other. The Fisher exact test is employed due to small expected counts (sample size) in some instances. These tests compare the observed frequencies of occurrences to the expected frequencies given independence, which tests if the variables are related or if the observed distribution is significantly different from what would be expected by chance (McClave et al., 2014). Additionally, variables compared in a contingency table has an odds-ratio, which quantifies the strength and direction of the association.

Both the Pearson Chi-squared test statistics and Fisher exact test statistics are analyzed to assess if the abundance or underabundance of specific personnel categories are related to cost and schedule performance. Both tests are run on all combinations of ACAT and base. Odds ratios are calculated for each comparison but are not interpreted unless the results are statistically significant. This study is exploratory, so a 10% level of significance is used to test the following hypotheses:

H_o: *There is no association between the two categorical variables.*

H_a: *There is an association between the two categorical variables.*

Analysis and Results

Table 4 shows the number of performance breaches within the final dataset for both cost and schedule performance metrics. Breach amounts represented as a percent of the dataset are shown in paratheses.

Table 4. Breach Counts

	Significant	Critical
Cost	14 (39%)	7 (19%)
Schedule	22 (61%)	15 (42%)

Schedule breaches are more common than cost breaches, with 61% of the programs incurring significant schedule breaches. Of those, nearly two thirds suffered critical schedule breaches. Cost breaches are less common, with 39% of the programs incurring significant cost breaches. Only 19% of the programs had critical cost breaches.

Moving into contingency table analysis, Chi-squared tests, Fisher tests, and odds-ratio calculations are conducted for comparisons of significant and critical cost/schedule performance metrics against all six personnel categories, all three ACATs, and all three bases. There are a total of 24 personnel comparisons, 12 ACAT comparisons, and 12 Base comparisons. Of the 48 tests run, half are related to cost performance, and half are related to schedule performance.

Table 5 shows the results of the personnel category dichotomous variables against both significant and critical cost breach dichotomous variables. In either case, there were no significant results. Odds ratios are not displayed for insignificant results, as the interpretation would be erroneous. The results suggest that there is not an association between program office estimating team composition and significant or critical cost performance.

Table 5. Personnel vs Cost Performance Metrics

	χ^2 p-value	Fisher p-value	Odds Ratio		χ^2 p-value	Fisher p-value	Odds Ratio
	Significant Cost Breaches				**Critical Cost Breaches**		
Cost Analysis	0.8484	0.7343	N/A		0.7417	0.675	N/A
Budget	0.5427	0.4955	N/A		0.8697	0.6843	N/A
Program Management	0.9393	0.7419	N/A		1	1	N/A
Engineering	0.3793	0.3074	N/A		1	1	N/A
Contracting	1	1	N/A		1	1	N/A
Logistics	0.3793	0.3074	N/A		0.2392	0.2036	N/A
* p-value < .1, ** p-value < .05, *** p-value < .01							

Next, the relationship between ACAT and cost performance is analyzed. For significant cost breaches, ACAT 3 was statistically significant at the 0.1 level for Fishers exact test. An odds ratio of 0.28 suggests that ACAT 3 programs are about 3.57 times less likely to incur a significant cost breach when compared to ACAT 1 and 2 programs. For critical cost breaches, ACAT 3 was statistically significant at the 0.1 level for Fishers exact test. An odds ratio of 0.18 suggests that ACAT 3 program are about 5.55 times less likely to incur a critical cost breach when compared to ACAT 1 and 2 programs. When Base was analyzed for both critical and significant cost breaches, there were no significant results. Table 6 shows the significant results for ACAT and Base comparisons against both significant and critical cost breaches.

Table 6. ACAT & Base vs Cost Performance Metrics

	χ^2 p-value	Fisher p-value	Odds Ratio		χ^2 p-value	Fisher p-value	Odds Ratio
	Significant Cost Breaches				**Critical Cost Breaches**		
ACAT 1	0.5829	0.3566	N/A		0.5204	0.244	N/A
ACAT 2	0.4298	0.2667	N/A		0.4658	0.3327	N/A
ACAT 3	0.1494	0.0924*	0.2812		0.1246	0.0842*	0.18
Eglin	0.6604	0.6283	N/A		1	1	N/A
Hanscom	1	1	N/A		1	0.4882	N/A
WPAFB	0.6153	0.4413	N/A		1	0.639	N/A
* p-value < .1, ** p-value < .05, *** p-value < .01							

Transitioning to schedule performance, the same process is followed except with schedule breaches as the variables of interest. Table 7 shows the results of personnel category dichotomous variables against significant schedule breach dichotomous variables. Only two statistically significant personnel categories are associated with significant schedule breaches. Cost Analyst was significant at the 0.05 level for both the Chi-squared and Fisher exact test. An odds ratio of 0.15 suggests that programs with more cost analysts are about 6.66 times *less* likely to incur a critical cost breach when compared to the baseline. Contracting was significant at the 0.10 level for both statistical tests. An odds ratio of 4.375 suggests that programs with more contracting personnel are about 4.38 more likely to incur a critical cost breach when compared to the baseline. When personnel are compared against critical schedule breaches, there were no significant results. These results suggest that the only significant associations between program office estimating team composition and schedule performance is Cost Analysis and Contracting, both for significant schedule breaches.

In comparing schedule breaches to ACAT and Base, ACAT 3 programs are once again significant for both significant and critical performance breaches. Like cost performance, base was not found to be significant for schedule performance. Table 8 shows the significant results for all ACAT and base schedule comparisons.

Table 7. Personnel vs Schedule Performance Metrics

	χ^2 p-value	Fisher p-value	Odds Ratio		χ^2 p-value	Fisher p-value	Odds Ratio
	Significant Schedule Breaches				**Critical Schedule Breaches**		
Cost Analysis	0.0241**	0.0159**	0.15		0.4274	0.3204	N/A
Budget	0.9393	0.7419	N/A		1	1	N/A
Program Management	0.4467	0.3217	N/A		0.7778	0.736	N/A
Engineering	1	1	N/A		0.2123	0.1755	N/A
Contracting	0.0874*	0.0858*	4.375		0.499	0.4998	N/A
Logistics	1	1	N/A		0.9097	0.7412	N/A
* p-value < .1, ** p-value < .05							

Table 8. ACAT & Base vs Schedule Performance Metrics

	χ^2 p-value	Fisher p-value	Odds Ratio		χ^2 p-value	Fisher p-value	Odds Ratio
	Significant Schedule Breaches				**Critical Schedule Breaches**		
ACAT 1	0.1533	0.1336	N/A		0.1661	0.138	N/A
ACAT 2	0.4298	0.4315	N/A		0.1719	0.1221	N/A
ACAT 3	0.0389**	0.0334**	0.1389		0.011**	0.0061***	0.1176
Eglin	0.6604	0.6283	N/A		0.5685	0.3761	N/A
Hanscom	1	1	N/A		0.7598	0.5588	N/A
WPAFB	0.6153	0.4413	N/A		1	1	N/A
** p-value < .05, *** p-value <.01							

A summary of the significant results from all 48 comparisons is shown in Table 9. Of all 24 comparisons between personnel categories and performance metrics, only two categories provided significant results, both in significant schedule breaches. Cost Analysts were significant at the 0.05 level, and Contracting was significant at the 0.10 level. The results suggest that a higher number of Cost Analysts reduce the chance of a significant schedule breach, while a higher number of Contacting personnel increases the chance of a significant schedule breach. In all 12 comparisons between base and performance metrics, there were no significant findings. In the 12 comparisons between ACAT and performance metrics, ACAT 3 vs ACAT 1 and 2 consistently showed significant results, while the other two categories did not demonstrate any statistical significance. For both significant and critical cost performance, ACAT 3 programs showed significance at the 0.10 level. For significant and critical schedule performance, ACAT 3 programs showed significance at the 0.05 level. In all four cases, programs that were ACAT 3 had a reduced chance to incur any sort of performance metric breach.

Table 9. Significant Results

Comparisons	χ^2 p-value	Fisher p-value	Odds Ratio	Interpretation
ACAT 3 vs Significant Cost Breaches	0.1494	0.0924*	0.2812	ACAT 3 Programs are 3.57 times less likely to incur a significant cost breach when compared to ACAT 1 and 2 Programs
ACAT 3 vs Critical Cost Breaches	0.1246	0.0842*	0.18	ACAT 3 Programs are 5.55 times less likely to incur a critical cost breach when compared to ACAT 1 and 2 Programs
Cost Analysis vs Significant Schedule Breaches	0.0241**	0.0159**	0.15	Programs with more Cost Analysts are 6.66 times less likely to incur a significant schedule breach when compared to the baseline
Contracting vs Significant Schedule Breaches	0.0874*	0.0858*	4.375	Programs with more Contracting personnel 4.38 more likely to incur a significant schedule breach when compared to the baseline
ACAT 3 vs Significant Schedule Breaches	0.0389**	0.0334**	0.1389	ACAT 3 Programs are 7.2 times less likely to incur a significant schedule breach when compared to ACAT 1 and 2 Programs
ACAT 3 vs Critical Schedule Breaches	0.011**	0.0061***	0.1176	ACAT 3 Programs are 8.5 times less likely to incur a critical schedule breach when compared to ACAT 1 and 2 Programs
* p-value < .1, ** p-value < .05, *** p-value < .01				

Additional Analyses

Given a surprising lack of personnel significance in the dataset, additional analyses were conducted where sample size was available. In the previous analysis, personnel comparisons were made for all 36 programs, regardless of base. Given the large amount of Wright-Patterson AFB and ACAT 3 data, we looked at personnel comparisons in those specific categories. An additional 24 personnel comparisons were conducted for programs that were only at Wright-Patterson AFB and another 24 personnel comparisons for programs that were ACAT 3. When isolating for personnel within Wright-Patterson AFB, the results mirrored those of the overall dataset, with Cost Analysts and Contracting compared against significant schedule breaches being the only statistically significant comparisons. Given that Wright-Patterson AFB represents 78% of the dataset, this is expected. When isolating for personnel in ACAT 3 programs, there were no statistically significant results.

A second analysis examined the possibility that “size,” in terms of budget, is really the key variable. To explore this, the impact of Cost Analysts personnel on budgetary quartiles of struggling programs was also investigated. The programs that experienced cost growth were split into quartiles based on how large their total budget was and assigned dummy variables for each quartile. Each quartile was compared against the Cost Analyst dichotomous variables, with the hypothesis being that the larger the budget, the more a prevalence of cost analysts would positively impact the program. There were no significant results.

Discussion and Conclusion

The DoD has consistently struggled to manage cost and schedule within its programs, a concern widely discussed in the literature (Arena et al., 2006; Younossi et al., 2007; Riposo et al., 2014; Jones et al., 2023; GAO, 2023). In response to these challenges, a significant emphasis on cost and schedule performance has emerged in the field of acquisition research. Existing research has approached the issue by investigating strategic decisions, DAS processes, economic factors, risk, technological complexity, and cost estimates (Bolton et al., 2008; Monaco and White, 2005; Lorrel et al., 2017; Riposo et al., 2014). While these studies provide valuable insights, they have not investigated the influence of human capital in shaping performance outcomes, an area that has received little attention to this point. The ultimate goal of this paper has been to fill that gap and investigate the impact of human capital in military acquisitions. This paper’s results represent an important first look at the human capital issue. Even so, it is unquestionably *not* the final word on the subject, and further investigation is warranted as more data becomes available.

This research finds the program office cost estimating team composition does not appear to have a significant impact on cost or schedule performance. Of the 24 personnel comparisons made, only two categories are found to be significant, both for significant schedule breaches. Personnel manning had no impact on cost performance. The two significant schedule results suggest that an abundance of Cost Analysts personnel reduces the chance of a significant schedule breach, and an abundance of contracting personnel increases the chance of a significant schedule breach. In addition to personnel, base does not appear to be an important factor in program performance, as there are no significant results. An investigation into ACAT affirms the established trend that ACAT 3 programs are less likely to incur cost or schedule breaches.

While it may be disconcerting to some that this research does not establish personnel manning as a significant driver of cost or schedule deviations, a closer examination of the findings reveals an encouraging narrative. The evidence suggests that personnel manning is currently done efficiently, representing a good news story. The consistent lack of significance in personnel comparisons indicates that controlling cost and schedule growth is not currently about the make-up of the program office estimating team. The Air Force appears to have gotten that right. Rather, researchers should focus on the other factors from the literature to hopefully improve performance outcomes.

The conclusions of this study must be tempered by the reality that the results were found with a very small sample size. We emphatically reiterate this point. This research is not the final word regarding human capital impacts, and it is highly recommended that future researchers duplicate the study once more data becomes available. The unique dataset used in this study is building each day, and a reexamination as the sample size increases in 3, 5, or 10 years mimicking the methods of this paper are highly encouraged. In the meantime, researchers should focus on the primary factors previously highlighted by the literature, which include DAS processes, technological maturity, and strategic decisions.

References

Arena, M. V., Leonard, R. S., Murray, S. E., & Younossi, O. (2006). Historical cost growth of completed weapon system programs (Report No. TR-343). Santa Monica, CA: RAND. https://www.rand.org/pubs/technical_reports/TR343.html

Belbin, R. M. (1993), Team Roles at Work, Butterworth-Heinemann, Oxford

Bliss, R. T., Potter, M. E., & Schwarz, C. (2008). Performance characteristics of individually-managed versus team -managed mutual funds. *Journal of Portfolio Management*, *34*(3), 110. https://doi.org/10.3905/jpm.2008.706248

Blundell, R., Dearden, L., Meghir, C., & Sianesi, B. (1999). Human capital investment: the returns from education and training to the individual, the firm and the economy. *Fiscal studies*, 20(1), 1-23. https://doi.org/10.1111/j.1475-5890.1999.tb00001.x

Bolton, J. G., Leonard, R. S., Arena, M. V., Sollinger, J. M., & Younossi, O. (2008). Sources of weapon system cost growth: Analysis of 35 major defense acquisition programs (Report No. MG-670). Rand Corporation. https://www.rand.org/pubs/monographs/MG670.html

Bornstein, G., & Yaniv, I. (1998). Individual and group behavior in the ultimatum game: are groups more “rational” players?. *Experimental economics*, *1*, 101-108. https://doi.org/10.1023/A:1009914001822

Cancian, M. F. (2010). Cost growth: Perception and reality. *Defense Acquisition Review Journal*, *17*(3), 389-403.

Christensen, D. S., & Heise, S. R. (1993). Cost performance index stability. *National Contract Management Journal*, *25*(1), 7-15.

Cooper, D. J., & Kagel, J. H. (2005). Are two heads better than one? Team versus individual play in signaling games. *American Economic Review*, *95*(3), 477-509. 10.1257/0002828054201431

D’Amico, C.N., White, E.D., Ritschel, J.D., and Kozlak, S.J. (2018) “Unmasking Cost Growth Behavior: A Longitudinal Study,” *Defense Acquisition Research Journal*, 25(1): 30-51. https://doi.org/10.22594/dau.17-783.25.01

Feuring, Jeffrey C., "The Impact of Human Capital on the Cost of Air Force Acquisition Programs" (2007). *Theses and Dissertations*. 3011. Air Force Institute of Technology, Wright Patterson AFB, OH

Fox, J. R., Allen, D., Moody, W., & Shiman, P. (2011). Defense acquisition reform 1960-2009: An elusive goal. U.S. Government Printing Office

Government Accountability Office. (2023). Weapon Systems Annual Assessment: Programs Are Not Consistently Implementing Practices That Can Help Accelerate Acquisitions (GAO-23-106059). U.S. Government Printing Office. https://www.gao.gov/products/gao-23-106059

Gray, Stephen E., "The Correlation of Human Capital on Costs of Air Force Acquisition Programs" (2009). *Theses and Dissertations*. 2579. Air Force Institute of Technology, Wright Patterson AFB, OH

Hofbauer, J., Sanders, G., Ellman, J., Morrow, D., Berteau, D., Ben-Ari, G., & Lombardo, N. (2011). Cost and time overruns for major defense acquisition programs. *Acquisition Research Program*.

Jones, S J., White, E. D., Ritschel J. D., & Valentine, S. M. (2023). Cost estimation trends for major defense acquisition programs. *Defense Acquisition Research Journal, 30*(2), 96-123. https://doi.org/10.22594/dau.22-894.30.0

Kim, D. B., White, E. D., Ritschel, J. D., & Millette, C. A. (2019). Revisiting reliability of estimates at completion for department of defense contracts. *Journal of Public Procurement*, *19*(3), 186-200. 10.1108/JOPP-02-2018-0006

Kozlak, S., White, E., Ritschel, J., Lucas, B., & Seibel, M. (2017). Analyzing cost growth at program stages for DoD aircraft. *Defense Acquisition Research Journal*, 24(3), 386–407. https://doi.org/10.22594/dau.16-763.24.03

Krueger, A. B., & Lindahl, M. (2001). Education for growth: Why and for whom? *Journal of economic literature*, *39*(4), 1101-1136. https://www.aeaweb.org/articles?id=10.1257/jel.39.4.1101

Kugler, T., Kausel, E. E., & Kocher, M. G. (2012). Are groups more rational than individuals? A review of interactive decision making in groups. *Wiley interdisciplinary reviews: Cognitive science*, *3*(4), 471-482. 10.1002/wcs.1184

Lorell, M. A., Payne, L. P., & Mehta, K. R. (2017). Program characteristics that contribute to cost growth (RAND Report No. RR-1761). RAND. https://www.rand.org/pubs/research_reports/RR1761.html

Mankiw, N. G., Romer, D. and Weil, D. N. (1992), A contribution to the empirics of economic growth, *Quarterly Journal of Economics*, vol. 107, pp. 407–37. 10.2307/2118477

Mathieu, J., Maynard, M. T., Rapp, T., & Gilson, L. (2008). Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. *Journal of management*, *34*(3), 410-476. 10.1177/0149206308316061

Mcclave, J. T., P George Benson, & Sincich, T. (2014). *Statistics for business and economics*. Pearson.

McEwan, D., Ruissen, G. R., Eys, M. A., Zumbo, B. D., & Beauchamp, M. R. (2017). The effectiveness of teamwork training on teamwork behaviors and team performance: a systematic review and meta-analysis of controlled interventions. *PloS one*, *12*(1), e0169604. https://doi.org/10.1371/journal.pone.0169604

McNicol, D.L. (2022). Can We Explain Cost Growth in Major Defense Acquisition Programs? Defense Acquisition Review Journal, 29(1), 2-20. https://doi.org/10.22594/dau.21-867.29.01

Monaco, J. V. & White, E. D. (2005). Investigating schedule slippage. *Defense Acquisition Review Journal*, *12*(2), 176–193. https://doi.org/10.21236/ada441770

Morey, J. C., Simon, R., Jay, G. D., Wears, R. L., Salisbury, M., Dukes, K. A., & Berns, S. D. (2002). Error reduction and performance improvement in the emergency department through formal teamwork training: evaluation results of the MedTeams project. *Health services research*, *37*(6), 1553-1581. 10.1111/1475-6773.01104

Oxford Languages. (2023). Human capital.

Riposo, J., McKernan, M., & Kaihoi, .Chelsea. (2014). Prolonged cycle times and schedule growth in defense acquisition: A literature review (RAND Report No: RR-455-OSD). Santa Monica, CA: RAND. https://www.rand.org/pubs/research_reports/RR455.html

Salas, E., Diaz, Granados, D., Klein, C., Burke, C. S., Stagl, K. C., Goodwin, G. F., & Halpin, S. M. (2008). Does team training improve team performance? A meta-analysis. *Human factors*, *50*(6), 903-933. 10.1518/001872008X375009

Younossi, O., Arena, M. V., Leonard, R. S., Roll, C. R., Jain, A., & Sollinger, J. M. (2007). Is weapon system cost growth increasing? A quantitative assessment of completed and ongoing programs (Report No. MG-588- AF). Santa Monica, CA: RAND. https://doi.org/10.7249/mg588

……

Lieutenant ***David K. Smith,*** *USAF, is a recent graduate of the Air Force Institute of Technology (AFIT) where he earned a Master's of Science in Cost Analysis. He is currently a cost analyst for the Air Force Lifecycle Management Center (AFLCMC), working under the digital directorate at Hanscom Air Force Base. (E-mail address: david.smith.464@us.af.mil)*

Dr. ***Jonathan D. Ritschel*** *is an Associate Professor of Cost Analysis in the Department of Systems Engineering and Management at AFIT. He received his BBA in Accountancy from the University of Notre Dame, MS in Cost Analysis from AFIT, and PhD in Economics from George Mason University. His research interests include public choice, the effects of acquisition reforms on cost growth in DoD weapon systems, and economic institutional analysis. (E-mail address: Jonathan.Ritschel@afit.edu)*

Dr. ***Brandon M. Lucas*** *is an Assistant Professor of Systems Integration & Cost Analysis in the Department of Systems Engineering and Management at AFIT. He holds a BA in History from the University of Texas at Austin, a MA in International Relations and ME in Teacher Education from the University of Oklahoma, a MS in Cost Analysis from AFIT, and a PhD in Economics from George Mason University. Dr. Lucas' research interests include profit analysis, cost & economic analyses, and incentive structures. (E-mail address: Brandon.Lucas@afit.edu)*

Dr. ***Edward D. White*** *is a Professor of Statistics in the Department of Mathematics and Statistics at AFIT. He received his BS in Mathematics from the University of Tampa, MAS from The Ohio State University, and PhD in Statistics from Texas A&M University. His primary research interests include statistical modeling, simulation, and data analytics. (E-mail address: Edward.White@afit.edu)*

Mr. ***Shawn M. Valentine*** *is an Operations Research Analyst at AFLCMC/FZCE at Wright Patterson AFB, Ohio. Mr. Valentine graduated from Ohio University with a Bachelor of Science (BS) in Actuarial Science and a Master of Science (MS) in Financial Economics. (Email address: shawn.valentine@us.af.mil)*

Trouble With the Curve:
Engineering Changes and Manufacturing Learning Curves

Brent M. Johnstone

Engineering changes pose a dilemma for estimators: If learning curves assume cost improvement due to repetitive build, what happens when that repetition is interrupted by a change of task? Design changes are common occurrences, but rarely addressed in learning curve literature. This paper addresses how to analyze an engineering change by breaking it into its pieces and outlines techniques to calculate the reversionary impact on the learning curve to derive the estimated cost of change.

Introduction

The learning curve demonstrates the cost benefit incurred by repetitively building a product over time. A cost curve "represents two facts: (1) that the time to do a job will decrease each time that job is repeated, and (2) that the amount of decrease will be less with each successive unit" (Fowlkes, 1963).

But what if the part design or the manufacturing process used to fabricate or assemble that part is altered? What if the job is no longer being repeated, but changes? What happens to everything that has been learned to date? A learning curve inherently assumes that the same task is being performed unit over unit – it is precisely that repetition of effort that is thought to create learning in the first place. So how do design changes impact the learning curve?

An engineering change will require the operator at a minimum to study and evaluate new planning to understand the change. He may need to review drawings and specifications for new engineering requirements. It may require him to learn new manufacturing methods or learn to use new or modified tooling. He may find himself in a new or altered work environment, accommodate new or changed production schedules, or submit to new or revised inspection criteria. (*Disruption*, n.d.) From an engineering or tooling perspective, changes may introduce errors in the design or tooling which may have to be subsequently fixed. The supply chain may have to start to produce new or revised parts which, if the engineering release is late, may create downstream part shortages on the assembly line. If we think of Anderlohr's five elements of learning (Anderlohr, 1969) – personnel, supervision, tooling, continuity of production, and methods -- we can see that any or all of these can be affected by an engineering change. All of this conspires to increase the number of hours to perform a changed task, at least initially.

The impact of an engineering change is best summarized by a General Dynamics training package from the early 1970s: Loss of learning "results from the re-introduction of problems associated with something new. These problems can vary widely and sometimes include tooling or engineering discrepancies. These discrepancies necessitate rework in fabrication or assembly areas. They contribute to the lost manhours or reduced efficiency. Therefore, they result in higher cost per completed end item from the point of reconfiguration." (*Learning Curves*, n.d.)

How then do we estimate this input? Surprisingly, most learning curve training packages make no reference to the manufacturing cost impacts of engineering changes. More commonly, they illustrate the cost impacts of production breaks. But engineering changes are far more common than a break in production. A long-running aircraft production program may experience as many as seven or eight major design changes before experiencing an actual production break if it ever does.

This paper reviews the ways we can assess and project these inputs.

Engineering Changes

In defense acquisition, engineering changes typically come in the form of an Engineering Change Proposal (ECP). We can think of engineering changes as doing one or more things:

1. A change may add tasks which did not previously exist.
2. A change may delete tasks which no longer must be performed.
3. A change may modify or reconfigure an existing task.

For discussion, Table 1. envisions a simple engineering change in a forward equipment bay of an aircraft which embodies all three cases.

Additions:
* Add two (2) new antennas
* Add coax cables
* Add provisions (brackets, fittings)
* New access door

Reconfigured:
* Relocate existing systems
* Relocate existing harnesses and tubes
* Move bulkhead penetrations to accommodate changed provisions

Deletions:
* Remove one (1) existing antenna
* Remove related provisions

Table 1. Example ECP with Notional Task Changes.

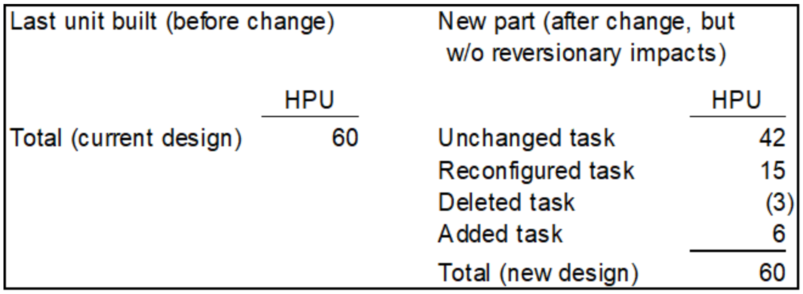
How would we evaluate the cost impacts of this change? We can separate the estimator's task into three categories:

1. Determine the baseline underlying learning curve prior to the change.
2. Isolate the portion of the total department task affected by the change. Identify the change to the affected cost centers and work breakdown structure. Relate this to the total department task.
3. Calculate the impact amount for the total.

We want to determine the expected hours per unit (HPU) impact of added, deleted, and reconfigured tasks prior to any consideration of possible learning loss or setback. The current value of deleted tasks is relatively easy – presumably we have cost history on what it takes us to perform these tasks today. For reconfigured task, we likely have current cost history on the existing part number but do not have history on the modified part. We will not have current history on added tasks – by definition, we are not performing those tasks today. But an estimate can be developed through a variety of methods – industrial engineering standards analysis, expert judgment, analogy to other parts on this or other programs using complexity factors, etc. In addition, a cost assessment could be performed parametrically by using a weight in/weight out analysis to calculate an hours per pound delta.

After the expected cost of the added, deleted, and reconfigured tasks are calculated, we next calculate the percent contribution each of these categories makes to the total component or subcomponent cost. In our example change, this ECP affects only part of a larger subcomponent, so there are areas which are not impacted. Table 2 provides a breakdown of HPU by estimating category.

The left-hand side shows the current cost of the component (60 hours) before the change. The right-hand side shows the breakdown of the new part cost by category. Based on an industrial engineering analysis, we have concluded that 70% of the task, or 42 hours, will be untouched by the design change. Reconfigured tasks equate to 15 hours while the deleted task will remove 3 hours. Finally, the added task is estimated to be worth 6 hours.



Last unit built (before change)	HPU	New part (after change, but w/o reversionary impacts)	HPU
Total (current design)	60	Unchanged task	42
		Reconfigured task	15
		Deleted task	(3)
		Added task	6
		Total (new design)	60

Table 2. Sample ECP, Breakdown of Hours per Unit (HPU) by Category.

The manufacturing cost of the new design (60 hours) is equal to the cost of the current design (60 hours). Can we conclude there is no delta cost impact for the engineering change?

We cannot because our analysis is incomplete. At no point have we accounted for the impact of reversionary impacts on either the added or reconfigured tasks. Every engineering change requires a consideration of reversionary impact on the learning curve.

Reversionary Impact, Two Measurements

What do we mean by reversionary impact?

Reversionary impact is the unfavorable impact to cost that typically accompanies design changes. Reversionary impact, or loss of learning, is usually expressed in terms of setback on the learning curve. (Teplitz, 2014) To set back unit cost on a curve means to assume unit costs are based on cumulative unit positions earlier in the program. In other words, the program repeats a prior level of performance at a higher hours per unit cost.

Setback is typically calculated as follows:

Setback position = Break-in position x (1- Setback %)

Let us assume costs have been moving down an 85% unit learning slope. To date, 1,000 units have been built. To estimate the cost of the 1,001st unit, assuming a T-1 of 100 hours, we would calculate:

(100 * 1,001-0.32193) = (100 * 0.1979) = 20 hours

Now instead assume a 50% setback occurs at the same 1,001st unit. The setback position – the new cost position incorporating the loss of learning – would equal:

1,000 x (1 – 50%) = 500.

(Note that the setback calculation is typically based on the cumulative number of completed units up to, but not including, the break-in point, since we are measuring how much cumulative learning to date will be lost.)

Because of the reversionary impact, we have lost an estimated 500 units of learning. That means when we estimate the cost for unit number 1,001, for learning curve purposes we estimate it on the learning curve as if it were unit number 501 -- that is, 500 units back up the learning curve. Then for our same curve:

$100 * 501^{-0.32193} = 100 * 0.2329 = 24$ hours

This 50% setback on the cost curve has resulted in a 20% increase in hours (from 20 hours per unit to 24 hours per unit).

It's important to note that calculations of unit setback are *not* the same as calculations of learning loss as used in the Anderlohr production gap methodology (Anderlohr, 1968). A quick illustration will demonstrate the difference. Using the Anderlohr calculation of learning loss:

First unit cost	100 hours
Less T-1000 cost before setback	20 hours
Learned to date	80 hours
T-1001 cost before setback	20 hours
Less T-1001 cost after setback	24 hours
Hours of learning lost	(4) hours
Percent learning loss	5%

(4 hours learning lost / 80 hours learned before setback)

Reversionary impacts can be viewed in terms of unit setback (50% in our example) or Anderlohr learning loss (5% in the same example). The Anderlohr learning loss factor will typically be less than the percentage of unit setback because of the logarithmic nature of the learning curve where most of the learning occurs in the front end of the learning curve, usually in the first 50 units. For our purposes in this paper, reversionary impact will be measured by percent setback on the learning curve.

Figure 1 shows two different ways to visualize reversionary impacts (Asher, 1956). Assume an engineering change that breaks at unit 100. In one case, the first changed unit is plotted at unit 100 and the subsequent units are plotted as unit 101, unit 102, etc. Call this View A. In the second case, the first changed unit is plotted at unit 1 and the subsequent units are plotted at unit 2, unit 3, and so forth. Call this View B. Essentially View B treats the first impacted unit as if it were a completely new aircraft at T-1.

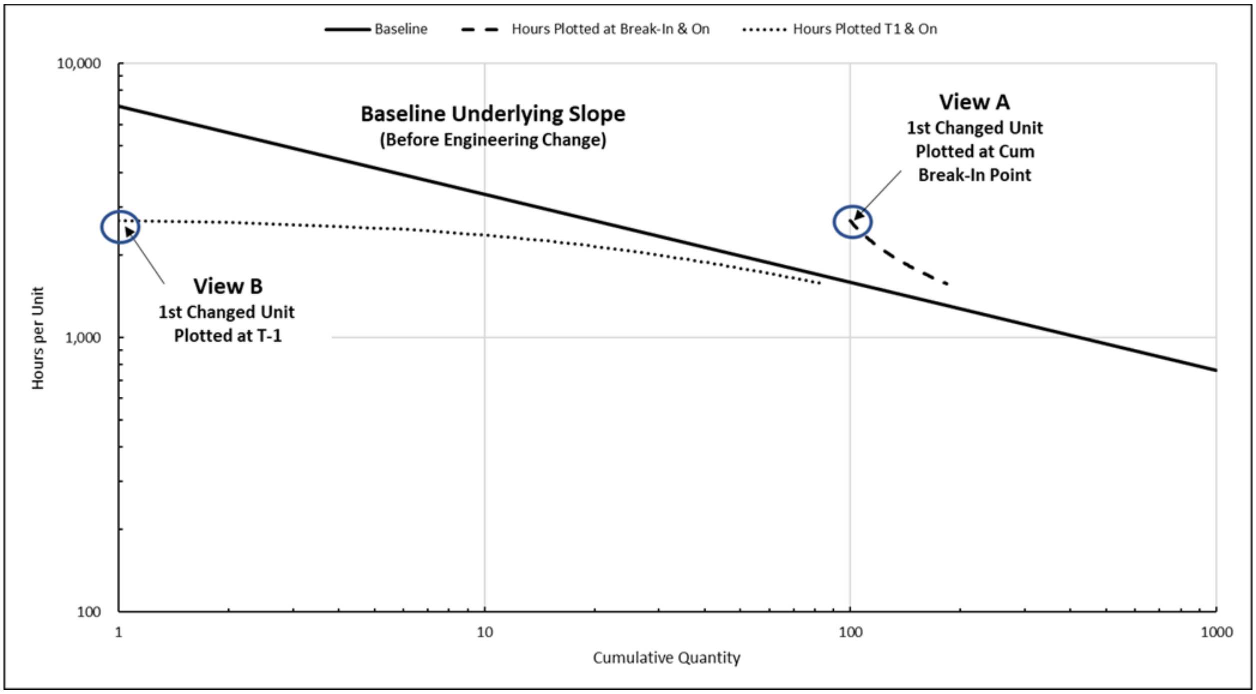


Figure 1. Two Views of Reversionary Impact

Both views portray the same hours per unit data. But View A shows more vividly the connection between the experienced baseline curve (before the change) and the delta impact created by the change. This is less obvious in view B. As changes accumulate over time, View A provides an historical continuity which demonstrates the impacts of design changes as well as other programmatic impacts such as schedule or manpower changes. That demonstration is not easily done using the View B plotting methodology.

Setback: How Much Is Enough?

The next question for the estimator is: How far should we set back the added and reconfigured tasks?

The obvious answer is that these tasks should be set back all the way to T-1 again. University of San Diego professor Charles Teplitz is one learning curve writer who supports this approach, writing "that portion of the task that has been altered has, in essence, suffered a setback all the way back to unit 1." (Teplitz, 1991)

However, there is reason to believe this is an overly conservative view. It tacitly assumes that learning is primarily operator driven, where a task change and the associated loss of muscle memory would create a significant impact. However, Jefferson suggests the operator learning only contributes 22% to overall cost improvement, while tooling improvements (34%) and engineering changes to assist production (23%) are bigger contributors. (Jefferson, 1981) We noted earlier that an engineering change can impact any of Anderlohr's five elements of learning – personnel, supervision, tooling, continuity of production, and methods. But many engineering changes may affect only one or two elements, not all of them. For example, a small change -- the movement of a harness from one location to another in a bay – may require an operator to learn a new location to install the harness and its associated bracketry. But he need not relearn how to route a harness through a hole in structure, install clamps and studs properly, make connections, or perform electrical bond. Moreover, it may not affect tooling, create any part shortages, or require learning new manufacturing processes or methods. It is hard then to imagine that such a design change would push the cost of implementation of a reconfigured or even added task all the way back to the first unit cost. In the author's experience, a careful breakdown of historical cost deltas associated with engineering changes rarely shows a setback all the way back to T -1 for adds or reconfigurations.

So, if we need not return all the way to T-1, how far do we set back?

A careful examination of prior experience with design changes and correlating the observed cost setback against the nature of the change to the configuration allows us to construct tables for the estimator to use when determining how much setback to apply. Such a table would say large setbacks for highly invasive design changes and work its way down to smaller setbacks for relatively benign changes. Such reconfiguration can come from ECPs initiated earlier in the program, from a prior program at the same facility or data from other facilities. (*Manufacturing Direct Labor Change Impacts: Setback/Learning Gain*, n.d.*)* An example of a notional setback chart for an aircraft is shown in Table 3.

Table 3 or something like it, while it relies on estimating judgment, is more defensible during a contract negotiation than relying on the analyst to simply pick a number out of the air. It will also force more consistent choices across the estimating team since each ECP will be analyzed using the same criteria.

Consistent choices are important because, as we will see, the cost of a change depends significantly on the choice of a setback value. Table 4 shows the sensitivity of various setback decisions to the total cost. From a baseline T-1000 cost, we can see the cost impacts if we choose a 50% or 80% setback. We can also see that for the same setback value, the cost impact varies substantially if the learning curve slope is 80% or 90%:

From the table, two things are apparent:

1. For a given learning curve slope, the higher the setback value, the larger the cost impact.
2. For a given setback percentage, the steeper the learning curve slope, the larger the cost impact.

X1% Setback (Highest)

- New weapons system or design concept

X2% Setback

- Current weapons system but major revision to design, e.g., outer mold line change, total subsystem affected by change

X3% Setback

- Relocation of aircraft systems components with associated rerouting of provisioning (harnesses, cables, tubes, ducts)
- Substantial wiring and tubing changes creating greater density and associated installation complexity
- Material substitution within established manufacturing techniques

X4% Setback

- Moderate change in structure in part design details
- Relocations of aircraft systems adjacent to original location
- Lesser number of wires, tubes, ducts added

X5% Setback

- Limited change in structure with changes confined to hole patterns and locations, revisions in tolerances, etc.
- Relatively small addition of wires, tubes, ducts

X6% Setback (Lowest)

- Minimal revisions in structural design
- Very limited added wires, tubes, ducts

Table 3. Notional Setback Criteria. .

Setback	Setback Unit	Slope = 80%		Slope = 90%	
		Unit Factor @ Setback	Increase from T1000	Unit Factor @ Setback	Increase from T1000
80%	200	0.1816	68%	0.4469	28%
50%	500	0.1352	25%	0.3888	11%
0%	1000	0.1082	0%	0.3499	0%

Table 4. Setback Sensitivity Table.

Recovery to the Baseline

Once we have established how far back up the learning curve our unit costs are expected to return, the next question is: what happens from that point forward in time? As it turns out, there are two different approaches in the industry regarding how to deal with this question.

The first method is the asymptotic recovery methodology illustrated by Figure 2. An engineering change breaks in at T-100 and cost returns to the equivalent point of T-20 on an 80% learning curve. We have lost eighty units of learning overall. The cost of the follow-on units (sequence numbers 101, 102, 103, 104, etc.) will be calculated using that same eighty-unit setback – that is, they will be calculated as if they were units 21, 22, 23, 24 and so on.

This method brings the unit cost down relatively quickly. After 50 units after the design change is initiated, the unit cost is reduced 33% from the setback point. But the unit cost for the redesigned configuration will never reach the learning curve for the configuration before the change. Even at T-1000, the unit cost will be calculated as if the program was eighty units higher on the learning curve – that is, at T-920. The delta difference will be small, but it will still exist. Therefore, we say that the recovery is "asymptotic," that the cost of the redesigned part will incrementally approach, but not actually equal, the baseline cost performance curve for the original design. The reversionary impact continues ad infinitum, although it eventually becomes so small that it cannot be distinguished on a normal logarithmic graphic.

The recovery is asymptotic because the redesigned part is assumed to come down the same learning curve slope as the original part design. How realistic is that assumption? Larry L. Smith comments: "[F]or most situations the items and units produced are similar and the work environment (company policy, management attitudes, etc.) is sufficiently stable that we expect the same rate of learning." However, Smith notes a learning curve slope change is appropriate if the ECP changes the manufacturing process from a manual process to a semi-automated or automated process. (Smith, 1976) Similarly, Teplitz writes: "Some changes affect the performance time or cost yet do not impact the slope of the learning curve. Others, on the other hand, could affect both requirement needs and learning curve slope." (Teplitz, 1991) Such a change in

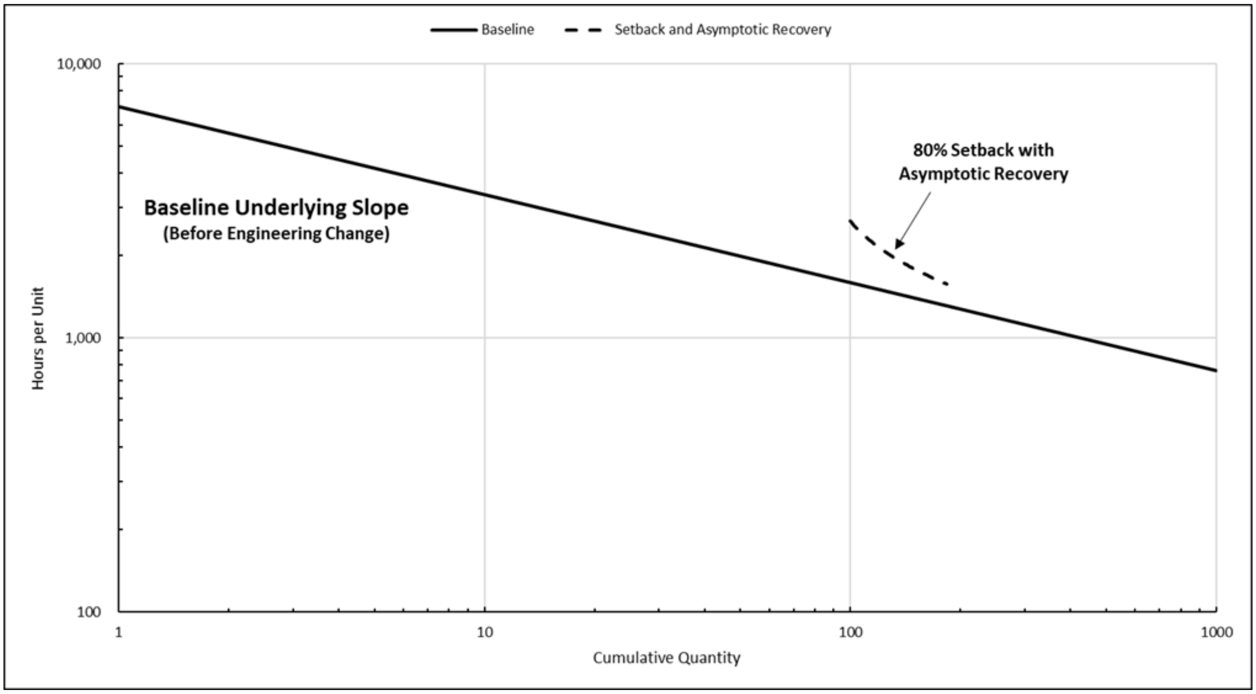


Figure 2. Setback with Asymptotic Recovery.

manufacturing process is not typical of most ECPs, however. Such changes usually require large tooling and facility non-recurring costs that most customers are unwilling to pay unless there is an immediate, near-term payback in cost savings.

A Different Approach

The method– which we will dub the "Variable Setback with Asymptotic Recovery" methodology -- is probably the most common industry approach for dealing with design changes. But it is by no means the only one. We'll contrast our first example with a second approach.

Instead of varying the unit setback depending on how extensive the design change, we could employ a universal rule of thumb applicable in all cases to how much setback is applied. One method is to employ a "one cycle" setback. Tracing back no doubt to the days of hand plotted charts on special paper pre-printed with logarithmic scales -- not so long ago in the author's career! -- this method moves the position on the cost curve back one logarithmic cycle from the break-in point. For example, if 700 units have been built at the time of the change, the setback is calculated as unit 70 (700 divided by 10). If 1,000 units have been built, the setback is calculated at unit 100 (1,000 divided by 10). A very extensive design change set back two cycles (from unit 700 to unit 7, or 700 divided by 100).

This one cycle setback methodology is paired with a different approach of recovery to the baseline. In this methodology, cost returns on a straight-line projection to the underlying baseline curve with intersection at some predetermined number of units. Unlike the asymptotic recovery, the manufacturing performance after the change returns to the pre-change cost curve at some point and continues as if the change had never occurred. The reversionary impact goes to zero. This can be seen in Figure 3.

This method also allows for the reversionary impact to be adjusted for the extent of the design change. However, instead of calibrating the amount of setback, the number of recovery units is calibrated to a higher number of units for significant design changes and a smaller number of units for more benign design changes. An extensive change may take 200 units to recover to the baseline, while a smaller change might only take 40 units.

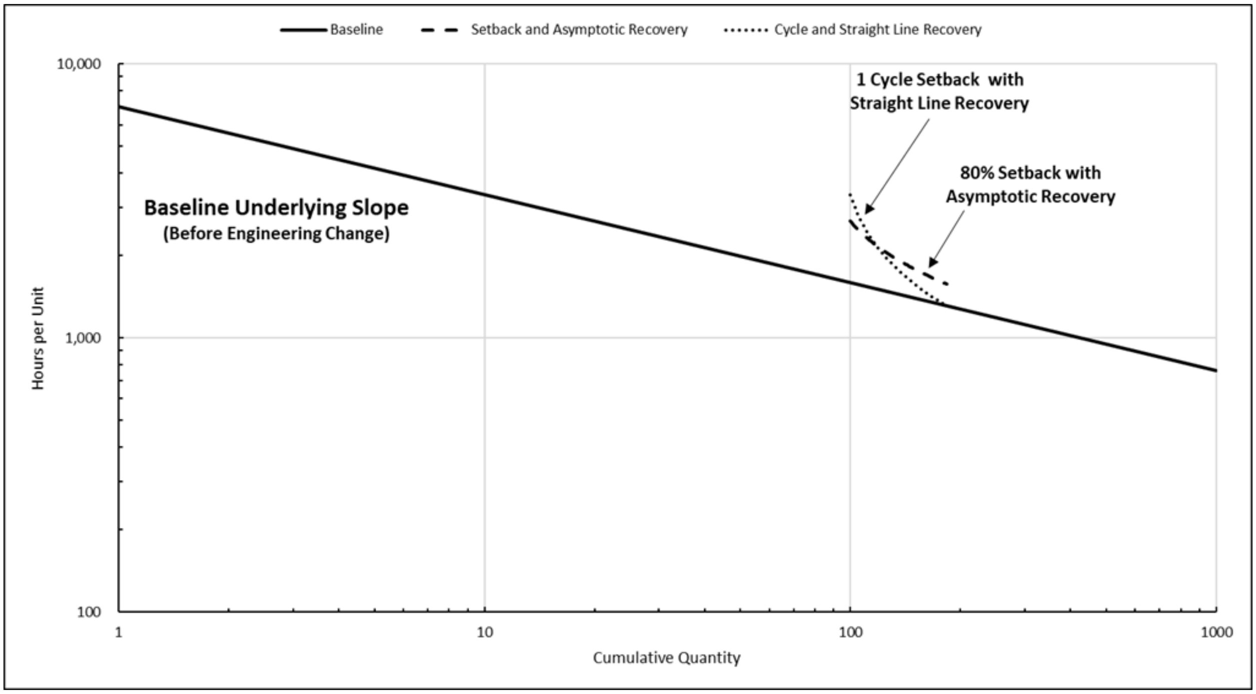
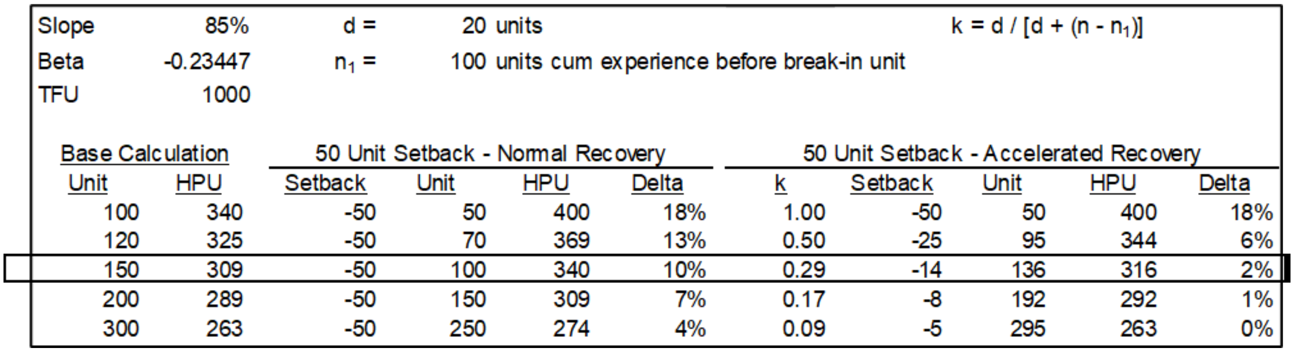


Figure 3. One Cycle Setback with Straight Line Recovery.



Slope 85% d = 20 units k = d / [d + (n - n_1)]
Beta -0.23447 n_1 = 100 units cum experience before break-in unit
TFU 1000

Base Calculation		50 Unit Setback - Normal Recovery				50 Unit Setback - Accelerated Recovery				
Unit	HPU	Setback	Unit	HPU	Delta	k	Setback	Unit	HPU	Delta
100	340	-50	50	400	18%	1.00	-50	50	400	18%
120	325	-50	70	369	13%	0.50	-25	95	344	6%
150	309	-50	100	340	10%	0.29	-14	136	316	2%
200	289	-50	150	309	7%	0.17	-8	192	292	1%
300	263	-50	250	274	4%	0.09	-5	295	263	0%

Table 5. Example of an Accelerated Recovery Curve

What are the advantages and disadvantages of these approaches? The one cycle setback rule – equivalent to a 90% unit setback applied in all cases – seems appropriate in some cases, less so for smaller design changes. That may be a more difficult “sell” during contract negotiations. A small design change would of course show a quick recovery over a short build run. Paired together, this may produce a shockingly steep effective learning curve that might be equally difficult to sell to the production managers responsible for carrying out the effort.

On the other hand, the asymptotic recovery approach is somewhat more difficult computationally than calculating a straight-line recovery. In addition, as the post-change hours per unit approach those on the baseline curve, there may come a point where the delta is so close it becomes immaterial. How long must we continue to carry it? The straight-line approach eliminates that concern.

More consequentially, the asymptotic recovery is usually underpinned by an assumption that the post -change learning curve slope will be the same as the slope before the change. But what if the learning curve slope is steeper? Then the post-change slope will intercept the pre-change cost curve, producing an answer like the straight-line approach.

There may be cases where such an accelerated recovery curve is desired. Cochran (1968) suggests a formula that is easily incorporated into the conventional recovery curve. He suggests use of a multiplier k_n for to be used for a n_0 amount of setback:

$$k_n = \frac{d}{d + (n - n_1)}$$

where n_1 represents cumulative units before break-in and n_0 or (n-n_1) represents the units of setback. The estimator chooses a value d as an accelerant for the recovery curve. Through experience, Cochran determined the optimal values of d are between 20 and 50. The smaller the value of d, the faster the recovery will be.

As an example, imagine an 85% baseline slope with a 50-unit setback for a design change implemented at unit 100 and an assumed d factor of 20. Table 5 shows the associated calculations. At unit 100, the units of setback are the same – 50 units – in the normal and accelerated recovery cases. However, at unit 150 the accelerated recovery curve only sets back 14 units, such that:

$$k_{150}\ x\ n_0 = \ 0.286\ x\ (150 - 100) = 14$$

We can use this to calculate the adjusted units of setback at T-150 as follows:

$$k_{150} = \frac{20}{20 + (150 - 100)} = 0.286$$

Consequently, instead of estimating an HPU of 340 hours at T-50 as the normal recovery curve would yield, an accelerated recovery curve would estimate an HPU of 316 hours (versus 309 HPU for the pre-change learning curve).

Figure 4 shows graphically the two different recovery slopes. In the first case, a normal asymptotic recovery is plotted. In the second, an accelerated recovery is plotted. Assuming a k factor of 20, the accelerated recovery HPU equals the baseline HPU by T-300 after 200 units have been built, where the normal asymptotic recovery curve is still 4% higher than the baseline at the same point.

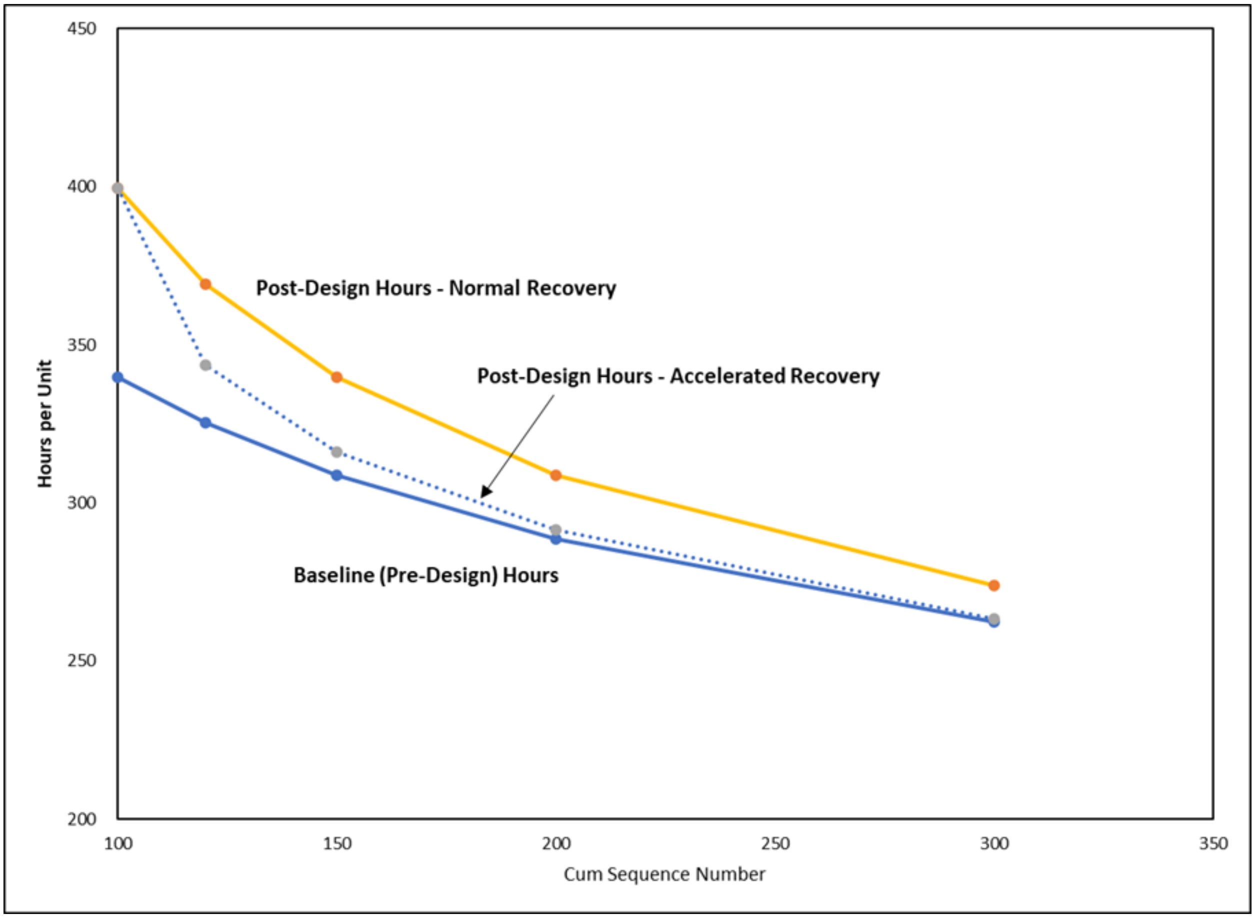


Figure 4. Accelerated Recovery Curve

Sample ECP – Calculations of the Cost Delta

Let us return to our original example – an ECP affecting the forward equipment bay of an aircraft. How might we apply these concepts to calculate the reversionary impact and the final cost impact?

In our example, we will apply the variable setback with asymptotic recovery methodology. We will apply a notional 75% setback since this change involves the relocation of aircraft systems components with rerouting of the associated provisioning.

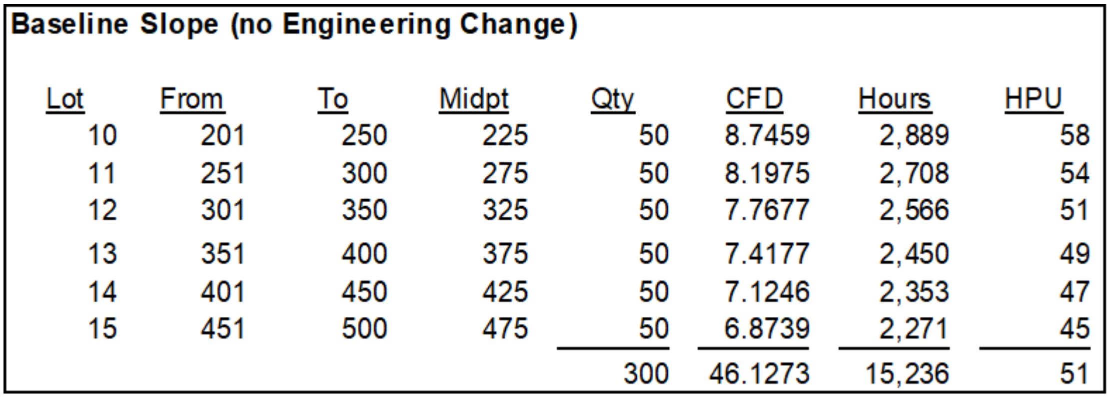
Table 6 shows the parameters associated with our notional ECP.

Table 7 calculates the baseline estimate for the task had no design change taken place. We will carry out the estimate for the next 300 units after the change, which breaks in at T-201. The cumulative factor difference (CFD) shown in the table is the sum of the individual learning curve unit factors over the range of aircraft in the lot, such that T-1 hours x CFD equals the total estimated lot hours, i.e., 330 T-1 hours x 8.7459 Lot 10 CFD = 2,889 Lot 10 hours. Table 8 shows the estimate for the added task. Notice we have applied a 75% setback, beginning our calculations at T-51 on the learning curve.

* Last unit built	200
* HPU at last unit built	60
* Assumed slope	80%
* Theoretical first unit (TFU)	330
* Setback	75%
* Equivalent last built	50

		HPU at Break
Added Task	10%	6.0
Reconfigured Task	25%	15.0
Deleted Task	-5%	(3.0)
No Change	70%	42.0
Total	100%	60.0

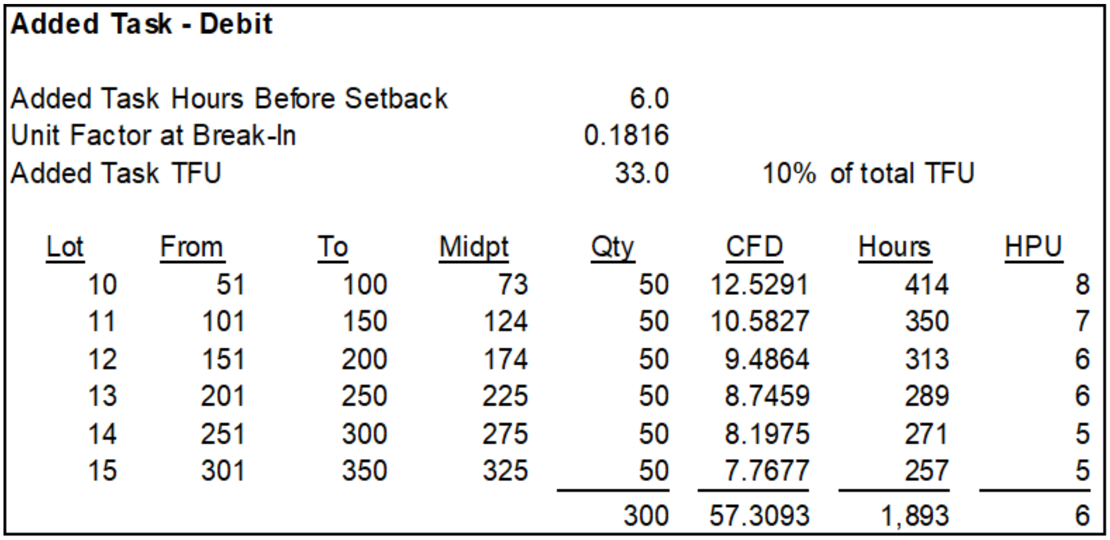
Table 6. Notional ECP Parameters



Baseline Slope (no Engineering Change)

Lot	From	To	Midpt	Qty	CFD	Hours	HPU
10	201	250	225	50	8.7459	2,889	58
11	251	300	275	50	8.1975	2,708	54
12	301	350	325	50	7.7677	2,566	51
13	351	400	375	50	7.4177	2,450	49
14	401	450	425	50	7.1246	2,353	47
15	451	500	475	50	6.8739	2,271	45
				300	46.1273	15,236	51

Table 7. Baseline Estimate.

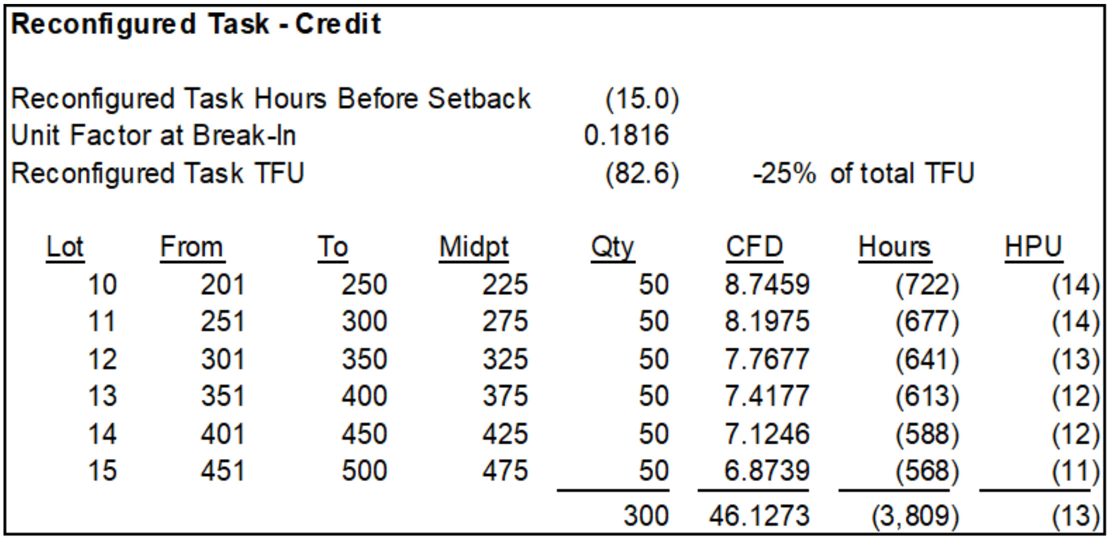


Added Task - Debit

Added Task Hours Before Setback	6.0	
Unit Factor at Break-In	0.1816	
Added Task TFU	33.0	10% of total TFU

Lot	From	To	Midpt	Qty	CFD	Hours	HPU
10	51	100	73	50	12.5291	414	8
11	101	150	124	50	10.5827	350	7
12	151	200	174	50	9.4864	313	6
13	201	250	225	50	8.7459	289	6
14	251	300	275	50	8.1975	271	5
15	301	350	325	50	7.7677	257	5
				300	57.3093	1,893	6

Table 8. Added Task – Debit Estimate.



Reconfigured Task - Credit

Reconfigured Task Hours Before Setback	(15.0)	
Unit Factor at Break-In	0.1816	
Reconfigured Task TFU	(82.6)	-25% of total TFU

Lot	From	To	Midpt	Qty	CFD	Hours	HPU
10	201	250	225	50	8.7459	(722)	(14)
11	251	300	275	50	8.1975	(677)	(14)
12	301	350	325	50	7.7677	(641)	(13)
13	351	400	375	50	7.4177	(613)	(12)
14	401	450	425	50	7.1246	(588)	(12)
15	451	500	475	50	6.8739	(568)	(11)
				300	46.1273	(3,809)	(13)

Table 9. Reconfigured Task – Credit Estimate.

Tables 9 and 10 show our estimate for the reconfigured task. We will calculate this task in two steps. First, we will credit the reconfigured task without setback as shown in Table 9. Second, we will debit the reconfigured task with setback as shown in Table 10. The delta, of course, is the reversionary impact.

Next, we calculate a credit for the deleted task as shown in Table 11. We have not applied any setback since the design change has eliminated this effort.

Finally, we take the totals of Tables 7 through 11 to show the sum of the added, deleted, and reconfigured tasks.

Our baseline is our method of manufacture and associated cost before the change is implemented. Mathematically, then:

- Debit = Hours for added and reconfigured tasks including reversionary impacts.
- Credit = Hours for any tasks eliminated by the change.
- Cost of change = Debit – Credit hours.

In total, the design change produces an additional 2,055 hours over the baseline. Over the course of the 300 units, the redesigned subcomponent will take 58 hours per unit (versus 51 hours per unit) or an increase of 13.5%.

A plot of the resulting hours is provided in Figure 5. Notice the "scallop" pattern seen in the earlier graphs of the post-change curve is not as pronounced. Less pronounced because, in this example, 70% of the subcomponent build is unaffected by the ECP, which has the effect of dampening the initial "bump" in the post-change hours per unit.

Reconfigured Task - Debit

Reconfigured Task Hours Before Setback	15.0	
Unit Factor at Break-In	0.1816	
Reconfigured Task TFU	82.6	25% of total TFU

Lot	From	To	Midpt	Qty	CFD	Hours	HPU
10	51	100	73	50	12.5291	1,035	21
11	101	150	124	50	10.5827	874	17
12	151	200	174	50	9.4864	783	16
13	201	250	225	50	8.7459	722	14
14	251	300	275	50	8.1975	677	14
15	301	350	325	50	7.7677	641	13
				300	57.3093	4,732	16
			Delta Hours for Reconfigured Tasks			923	3

Table 10. Reconfigured Task – Debit Estimate.

Deleted Task - Credit

Deleted Task Hours Before Setback	(3.0)	
Unit Factor at Break-In	0.1816	
Deleted Task TFU	(16.5)	-5% of total TFU

Lot	From	To	Midpt	Qty	CFD	Hours	HPU
10	201	250	225	50	8.7459	(144)	(3)
11	251	300	275	50	8.1975	(135)	(3)
12	301	350	325	50	7.7677	(128)	(3)
13	351	400	375	50	7.4177	(123)	(2)
14	401	450	425	50	7.1246	(118)	(2)
15	451	500	475	50	6.8739	(114)	(2)
				300	46.1273	(762)	(3)

Table 11. Deleted Task – Credit Estimate

Sum of the Totals

Lot	Baseline	Credit Reconfig	Debit Reconfig	Debit Added	Credit Deleted	Total Hours	HPU	% Delta
10	2,889	(722)	1,035	414	(144)	3,471	69	20.1%
11	2,708	(677)	874	350	(135)	3,119	62	15.2%
12	2,566	(641)	783	313	(128)	2,893	58	12.7%
13	2,450	(613)	722	289	(123)	2,726	55	11.3%
14	2,353	(588)	677	271	(118)	2,595	52	10.3%
15	2,271	(568)	641	257	(114)	2,487	50	9.6%
	15,236	(3,809)	4,732	1,893	(762)	17,291	58	13.5%

Baseline Hours (Lots 10-15)		15,236
Debits:		
Added Task	1,893	
Reconfigured Task Delta	923	
Credits:		
Deleted Task	(762)	
Total Cost of Change	2,055	
ECP Hours (Lots 10-15)		17,291

Table 12. Sum of Added, Deleted and Reconfigured Tasks.

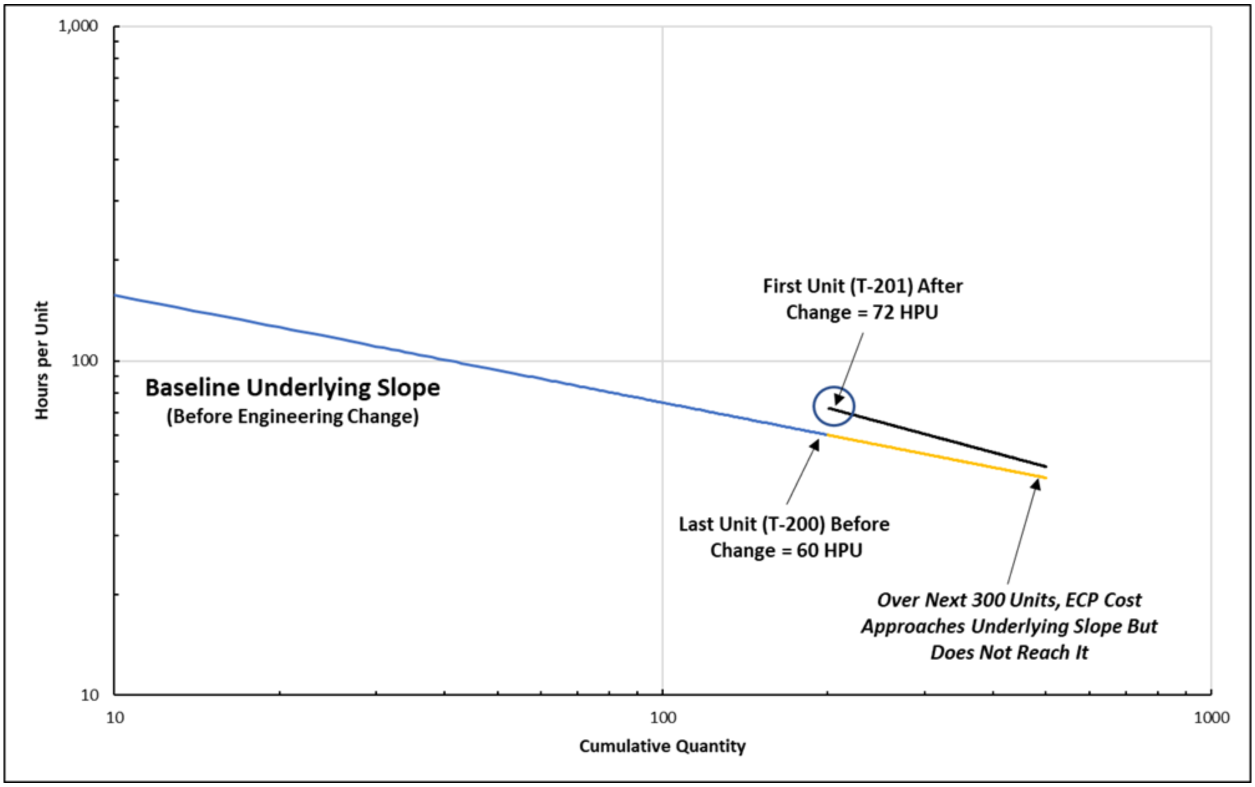


Figure 5. Notional ECP Hours per Unit, Before and After Change

Conclusions

The Greek philosopher Heraclitus said, “Everything changes, and nothing stands still.” Change on the manufacturing shop floor is a disrupter. It forces reevaluation of production schedules, tooling and plant layouts, crew tasks and responsibilities – and cost. Long cycle products like aircraft, missiles, and spacecraft – all of whom share lengthy production schedules, complex designs, and demanding customers – are particularly impacted by these engineering changes.

Estimating the impact of ECPs can be challenging for the cost estimator. It requires not only an understanding of the current shop conditions and the associated cost, but a careful breakdown of the tasks to be added or reconfigured by the ECP. Assistance from the design engineering, manufacturing engineering, and tooling functions can be invaluable in making these determinations. Lastly, the recognition that learning can be lost and regained over time, paired with a consistent and logical breakdown of the problem, can assist the analyst in making fair and reasonable cost estimates.

References

Anderlohr, George (September 1969). What Production Breaks Cost. *Industrial Engineering*, 34-36.

Asher, Harold (July 1956). *Cost-Quantity Relationships in Airframe Cost-Performance Studies*, Report No. R-291. Santa Monica, CA: Rand Corporation.

Cochran, Edward B. (1968). *Planning Production Costs: Using the Improvement Curve*. San Francisco, CA: Chandler Publishing.

Disruption. (n.d.). El Segundo, California: Northrop Aircraft.

Fowlkes, Tommie. (1963). *Aircraft Cost Curves*. Fort Worth, TX: General Dynamics Convair Division.

Jefferson, P. (May 1981). Productivity Comparison with the USA – Where Do We Differ. *Aeronautical Journal*, vol. 85 (844), 179-184.

Learning Curves. (n.d.). Fort Worth, TX: General Dynamics Fort Worth Division.

Manufacturing Direct Labor Change Impacts: Setback/Learning Gain. (n.d.). Fort Worth, TX: General Dynamics Fort Worth Division.

Smith, Larry L. (1986). *Cost Improvement Analysis*. Wright-Patterson AFB, Ohio: Air Force Institute of Technology.

Teplitz, Charles J. (1991). *The Learning Curve Deskbook*. New York, NY: Quorum Books.

Teplitz, Charles J. (2014). Learning Curve Setbacks: You Don’t Always Move Down a Learning Curve. *Journal of Applied Business and Economics*, vol. 16(6), 32-43.

Brent Johnstone *is a Lockheed Martin Fellow and production air vehicle cost estimator at Lockheed Martin Aeronautics Company in Fort Worth, Texas. He has 35 years’ experience in the military aircraft industry, including 32 years as a cost estimator. He has worked on the F-16 program and has been most recently the lead Production Operations cost estimator for the F-35 program. He has a Master of Science from Texas A&M University and a Bachelor of Arts from the University of Texas at Austin.*

CSI EU (Cost Scene Investigation – European Union)

Douglas K. Howarth

The commonly accepted story about why the Airbus A380 failed to make a return on its investment centers around its well-publicized development cost and schedule issues. Added to that was the movement of the airline industry from a traditional hub and spoke model to point-to-point routes; and the introduction of planes to the market, which had much more efficient engines and technology. Technology might have gone onto the A380, but, as Airbus CEO Tom Enders stated, "Airbus did not stumble into [building] the A380; we were very aware of the project's risks, but then technology and the market changed faster than anyone thought." [1]

Nico Buchholz, an ex-Airbus executive who later became Lufthansa's head of fleet strategy and ordered the plane for the airline, echoed Enders' sentiment, saying, "In 2000, you could not predict what crystallized in 2005—that the aircraft was technically outdated." [2]

That is exactly incorrect!

As we examine the A380, we'll find all the information needed to prove the aircraft's insufficient viability existed before its launch. Furthermore, its technology had nothing to do with its demise. It sealed its fate at the start.

This is the End:

In her Netflix series, Cunk on Earth, the fictional host of the mockumentary, Philomena Cunk (played by the actress Diane Morgan), asks a professor of Egyptology, "How did the Egyptians build the pyramids? Did they start at the top and work down, or start at the bottom and work up?' [3]

Buried in that silliness is a question with real import – how to come to a project's proper end? As laborers toiled to make the Great Pyramid at Giza reach ever higher, how could they engineer it to hit its tallest point in space within a centimeter or two? There are no drawings left for us to pore over to answer that question, but we know they spent much time pinching their starting points down. Glen Dash, an engineer who studies the Great Pyramid, found that "The builders of the Great Pyramid of Khufu aligned the great monument to the cardinal points with an accuracy of better than four minutes of arc, or one-fifteenth of one degree." [5] Having a well-conceived base foundation can lead to a proper ending.

But your results can vary, especially if you have not given the foundation sufficient thought. A trip through Italy will drive home this point.

Figure 1: The Great Pyramid at Giza has an exact base, which let it grow to previously unimaginable heights [4]

Figure 2 reminds us that the Pisa Tower started tilting because its builders did not realize the soil beneath differed from side to side. Starting in 1172 and completed in 1372, its construction team did not incorporate the lessons of the past.

Just over the Apennine Mountains, a little over 70 miles by air, another similar fiasco took place scant decades before.

Figure 3 shows that two towers lean precariously toward one another in downtown Bologna, Italy. The city has had to put straps around them to prevent them from crashing into one another. These towers, completed by 1119, might have given others some pause. Early on, intelligence about the lean of the Bologna Towers was insufficient for architects and

Figure 2: The Leaning Tower of Pisa went sideways because its builders didn't understand that the ground below one side was softer than the other [6]

Figure 3: The Bologna Leaning Towers [7]

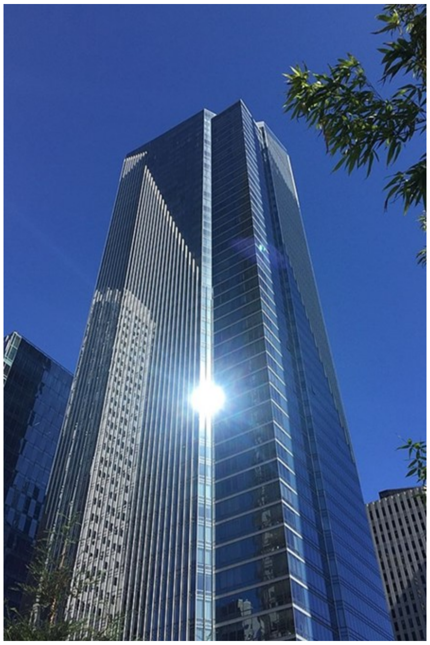


Figure 4: San Francisco's Millennium Tower [8]

engineers to take adequate measures to prevent Pisa's Tower from leaning.

Of course, such miscalculations are not solely a thing of the past. Figure 4's Millennium Tower lists 28 inches to the northwest as of 2022, as measured from the roof. [9]

In these instances of buildings going off-axis, builders aimed for a point in the sky and missed. To know how well or poorly a completed aerospace program did financially, we must first characterize where it aimed. Then we need to know the cost and associated revenues it would take to get to that point and how close it came to its target or by how much it missed.

The Targets:

If we think of the pinnacle of the Great Pyramid of Giza as a point in the sky, we can know where the structural architects wanted it to be and compare it to where it is. Such records of their intent are lost to history.

The aiming points of some more recent projects are easier to find. In Figure 5, we see that the Airbus people, as they launched their A380, thought they could sell 1250 units with a 2000 list price of $220M. At the same time, as Figure 6 reveals, in 1999-2000, the vehicle had a target weight of

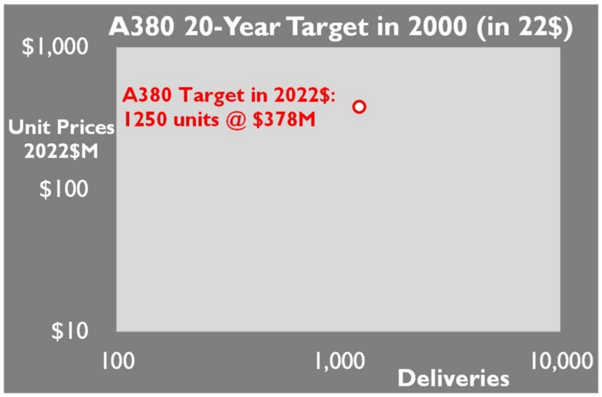


Figure 5: In 2000, Airbus projected 1250 A380 Units Sold [10] at a list price of $378M (in 2022$) [11]

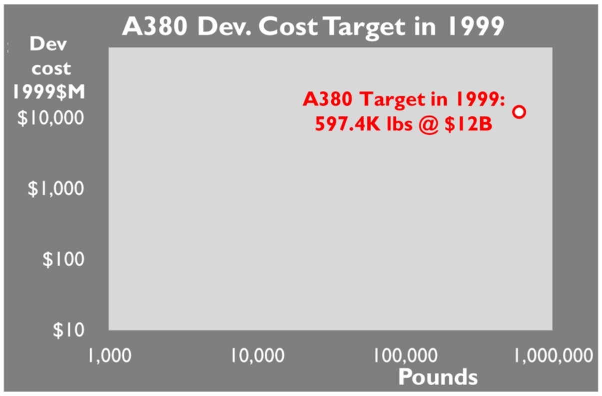


Figure 6: Airbus targeted $20.6B [12] (in 2022$) to develop an A380 weighing 597,400 pounds [13]

597,400 pounds and a development cost goal of \$12B. Those targets are without any reference points.

With these views, they are not unlike the builders of the Great Pyramid. While each aims for a point, neither offers enough perspective to gain insight. What happens when we take time to add some understanding to these issues?

Cost Target - Weight:

To see if the empty weight goal of the A380 was, at its start, a viable target, we might think of studying its weight. For context, we'll need to compare its weight history (as Manufacturer's Empty Weight, or MEW) to that of other programs. In Figure 7, we retrieve the weight data for 16 unnamed programs.

It turns out that these programs do not typically stay on target. Indeed, as Figure 8 shows, they often have a specific weight goal before they launch, one that drops upon the go-ahead. Then, over time that amount of mass grows.

When we normalize (to a starting value of "1" at program launch) and plot the empty weight data from each of the 16 models over time, we get Figure 8. Note there seems to be an upward trend over time. We can take advantage of that.

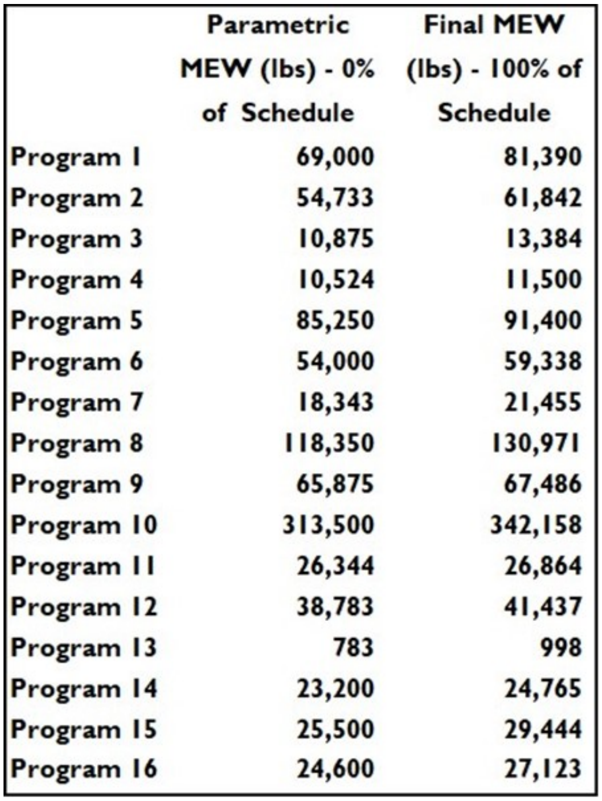
Using the data in Figure 7, we can try to predict the final Empty Weight from the like figure used at the program's start. When we do, we get Equation 1:

$$\text{Final MEW} = 1.48 * \text{Starting MEW}^{0.973} * \varepsilon \qquad (1)$$

Where:

Final MEW = Ultimate MEW, in pounds
Starting MEW = MEW at Go-Ahead, in pounds
ε = Error term for the equation

Equation 1 is well-correlated, with an Adjusted R^2 of 99.4%, a Mean Absolute Percentage Error (MAPE) of 4.3%, a Standard Error of 4317, and a P-Value of 8.93E-20. Any application of it outside its data range would be an extrapolation. If we were to apply it to smaller vehicles, say, one with a beginning empty weight of one pound, it suggests the aircraft would grow by nearly a half to 1.48 pounds.



	Parametric MEW (lbs) - 0% of Schedule	Final MEW (lbs) - 100% of Schedule
Program 1	69,000	81,390
Program 2	54,733	61,842
Program 3	10,875	13,384
Program 4	10,524	11,500
Program 5	85,250	91,400
Program 6	54,000	59,338
Program 7	18,343	21,455
Program 8	118,350	130,971
Program 9	65,875	67,486
Program 10	313,500	342,158
Program 11	26,344	26,864
Program 12	38,783	41,437
Program 13	783	998
Program 14	23,200	24,765
Program 15	25,500	29,444
Program 16	24,600	27,123

Figure 7: Target vs. Actual Empty Weights [14]

If we venture outside the database in the other direction for the exercise at hand, we could use Equation 1 to predict the A380 final weight from its starting condition. When we do that, as we discover in Figure 9, it forecasts a final A380 empty weight up 3.4%. In the end, though, the vehicle's MEW grew by 5.2% to 628,317 pounds. [15] While this increase is not trivial, it cannot explain the doubling of the program's projected Development Costs. Owing to the pervasive optimism in the industry, the weight prediction model projects it would take a vehicle with a starting MEW of 1,972,960 pounds to finish without any weight increases. Given the industry push to smaller and smaller unmanned aircraft, it would be beneficial to add such tiny planes to this mix, to see how the analysis of weight growth might be extended.

High Travel, Low Compatibility

To get a broad customer base for the A380, Airbus officials deliberately spread the work for its airplane about its several subsidiaries and critical suppliers. As we see in Figure 10A, this system, known as the Itinéraire à Grand Gabarit (in English, it roughly translates to "oversize convoy route"), is a water and road route in which the consortium invested

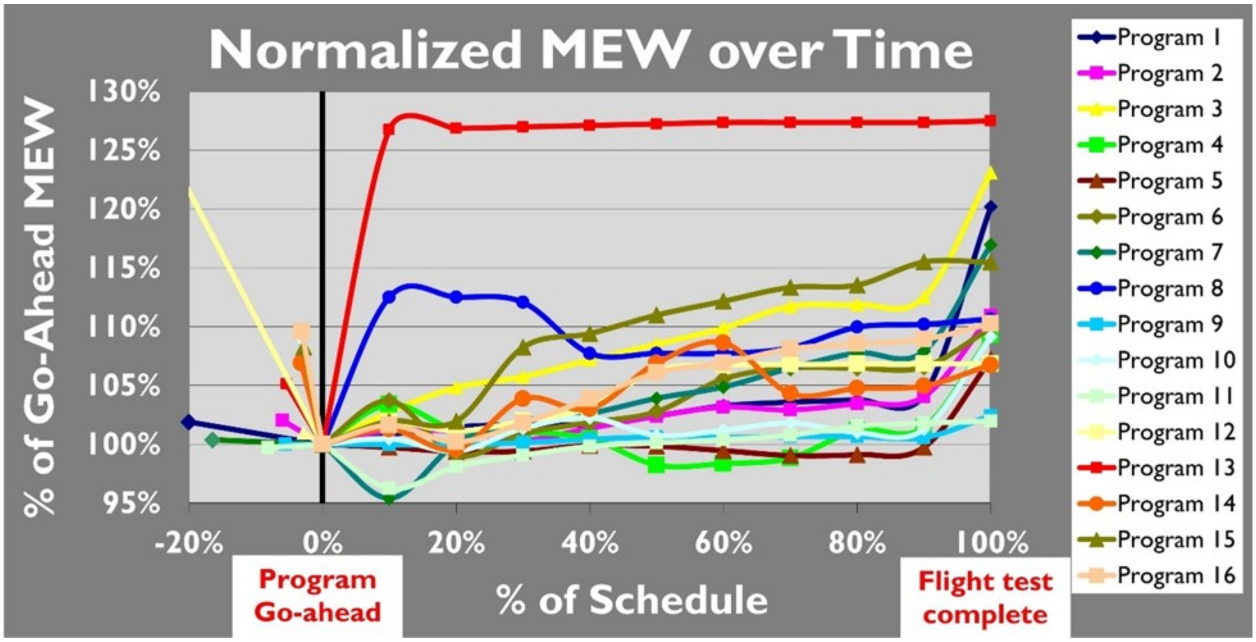


Figure 8: Aircraft Empty Weights typically fall as programs reach the Go-Ahead, then grow over time. The drop from the pre-launch figures to the ones at Go-Ahead represents systemic unfounded optimism.

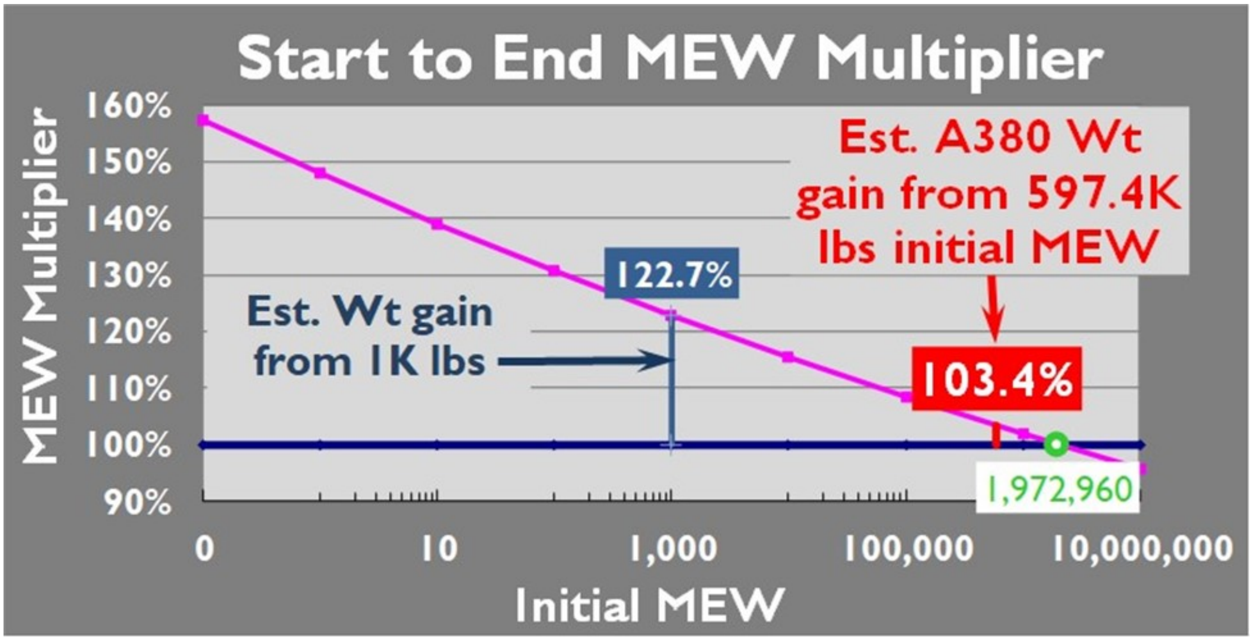


Figure 9: We find predictable weight growth when we take the Figure 7 estimated go-ahead Manufacturer Empty Weights and compare them to their ultimate weights. That growth, as a percentage, is higher for smaller vehicles than for larger ones. Here, a vehicle nominally weighing 1000 pounds at launch grows by 22.7% to its ultimate weight. Extrapolating the data, our weight increase models suggest that the A380's empty weight should have grown by 3.4% - available records show it grew by 5.1%.

hundreds of millions of dollars. The added time for pieces of the plane to get to final assembly in Toulouse added time and expense, but, more and more, getting several firms involved in a project seems to be the model for getting it launched in the first place.

More troubling than the parts moving long distances was the incompatibility of the software platforms across the consortium member countries. The CATIA (an acronym for computer-aided three-dimensional interactive application), invented by the French company Dassault Systems, offered added capability to engineers and designers. Initially released in 1982, it went through several revisions over the years. [18] By the time the program began, Spanish and German engineers were up to Version 4. Their French and British equivalents, however, were using Version 5.

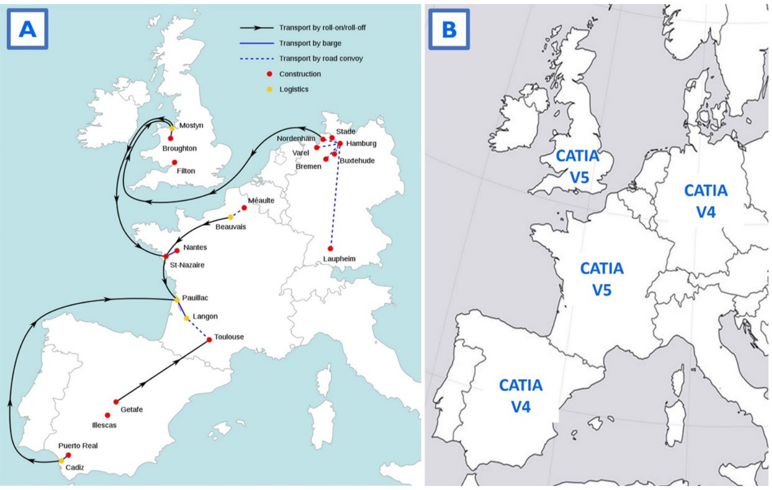


Figure 10: The building of the A380 was an international affair. The thinking was that if many countries participated in making the machine, lots of countries would buy the plane. Figure 10A shows the significant travel paths completed assemblies took to get to Toulouse, France, where the final assembly occurred. While that added cost to the plane, more significant was the difference in computer-aided design packages offered by CATIA. Spain and Germany worked from CATIA V4, while the United Kingdom and France used V5, a complete rewrite of the previous version. The lack of compatibility cost the program nearly two years in schedule and about $6.1B in then-year development cost dollars. [16]

The releases were not fully compatible, which created problems, most noticeably, in the electrical systems. Most wire harnesses came up short, forcing the program to endure a nearly two-year overall delay at $6.1B.

To get an idea of that issue's impact on Development costs, we should begin with an notion of what those costs should be. For that, we'll use the Figure 11 data (in 2022$), Development Cost information available to Airbus at their launch.

If we analyze the data in Figure 12, we'll have the information available to Airbus in 2000.

When we run a linear regression on the data, we get Equation 2:

$$\text{Dev Cost } \$2022M = 0.0486(MEW) - 1110 + \varepsilon \qquad (2)$$

Where:

Dev Cost 2022\$M = predicted development cost in 2022\$M
Final MEW = Ultimate MEW, in pounds
ε = Error term for the equation

Equation 2 has an adjusted R^2 of 92.6%, a MAPE of 46.0%, a Standard Error of $1,404,504,870, and a P-value of 8.73E-08.

Plotting the data in Figure 12, we find that if Airbus had used this equation as they launched, they would have predicted a 2022$ Development Cost of $27.9B using their projected starting weight (the blue point), or $29.4B (the green point) if they had allowed for weight growth to their final posted MEW. These figures compare to the Airbus estimate of $20.6B (their initial estimate of $12B inflated to 2022$). Richard Aboulafia put the final development cost between $31B and $37B. Using the midpoint of that range, or $34B, means the Airbus estimate was off by $34B - $20.6B = $13.4B, or 9.6 Standard Deviations from the estimate.

The company blamed its problems on schedule issues, but that doesn't seem to be the prime culprit. In Figure 13, we study the effect of size on schedule, as Operating Empty Weight (OEW) against the 1) Days from Launch to 1st Flight (Upper Right Chart), Days to 1st Certification (Lower Left Chart), and 3) Days to 1st Delivery (Lower Right Chart). Interestingly, we find the A380 actual schedule in keeping with the general trend, despite the company's position that their software compatibility issues cost them nearly two years of schedule. Surprisingly, the A380 took less time to develop than its smaller sister plane, the A350, which is less than half its size.

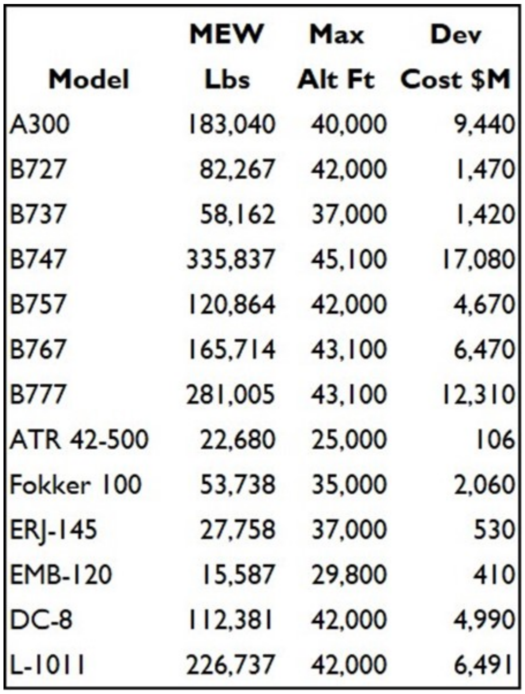
So, Airbus created a mess in their development phase, but they must have made up for it when they began delivering the planes.

Not really.

You'd want a production line to get its costs below its price as early as possible. However, well after their 200th delivery, "the [then] $445 million price tag of each aircraft was insufficient to cover the production cost. [That meant] …Airbus [was] losing money on each A380, and with orders evaporating, it made economic sense to shut down production."' [19] [20]It was probably much worse than they let on, as we have seen in Figure 14. [21]

The Starting Point

Both Boeing and Airbus considered a plane that would eventually become the A380 or something like it. For a brief time in the early 1990s, they even considered working on such a project together. [22]But Boeing decided not to pursue the new jumbo aircraft, while Airbus did.



Model	MEW Lbs	Max Alt Ft	Dev Cost $M
A300	183,040	40,000	9,440
B727	82,267	42,000	1,470
B737	58,162	37,000	1,420
B747	335,837	45,100	17,080
B757	120,864	42,000	4,670
B767	165,714	43,100	6,470
B777	281,005	43,100	12,310
ATR 42-500	22,680	25,000	106
Fokker 100	53,738	35,000	2,060
ERJ-145	27,758	37,000	530
EMB-120	15,587	29,800	410
DC-8	112,381	42,000	4,990
L-1011	226,737	42,000	6,491

Figure 11: Development Cost Database Before A380 Launch Date

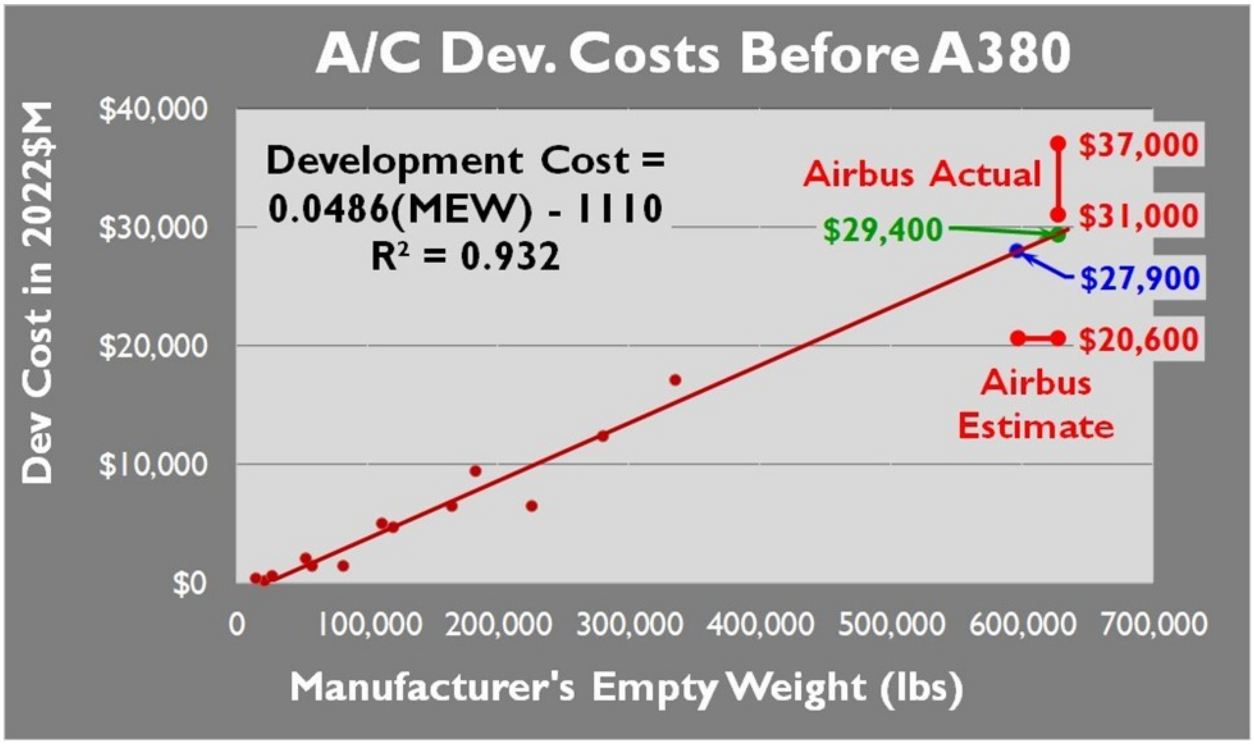
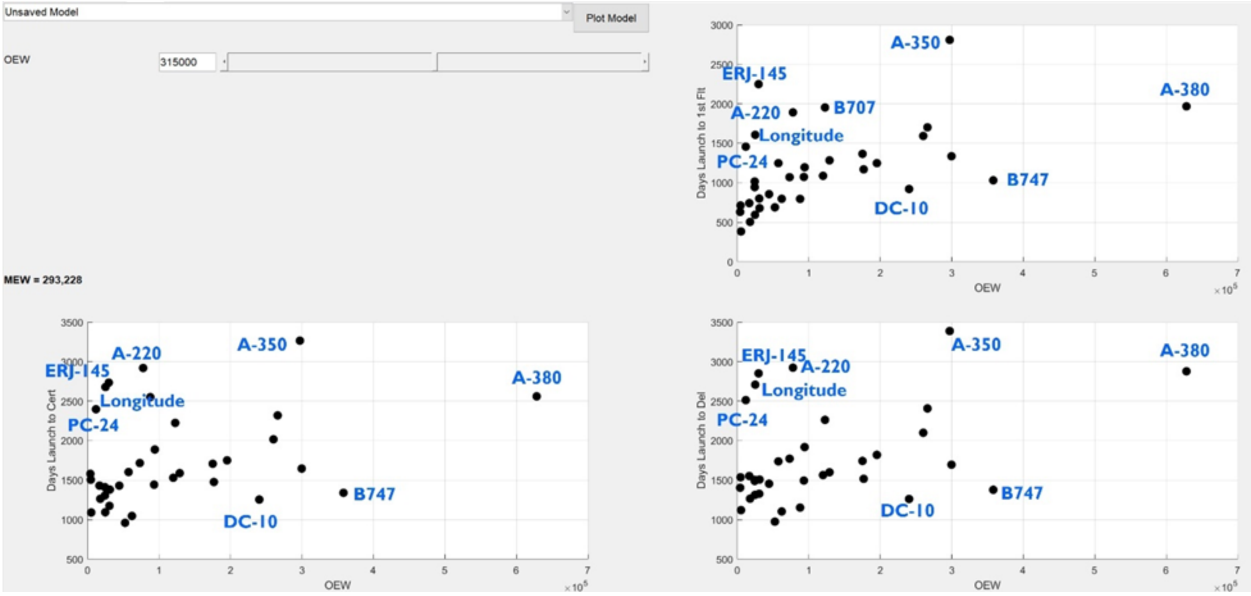


Figure 12: The original Airbus estimate for the A380 was 597,400 pounds, at $12B in 2000, which inflates to $20.6B in 2022. That's the leftmost of the points called the Airbus estimate – the rightmost one represents the exact cost of the A380's final weight. Had Airbus used this equation, their estimate would have been $27.9B. Its cost eventually rose to between $31B to $37B.


Unsaved Model

OEW 315000

MEW = 293,228

Figure 13: Airbus claimed a nearly two-year delay due to software incompatibility, leading to many of the A380's wire harnesses falling short. However, when we compare Operating Weight Empty (OEW) to 1) Days from Launch to 1st Flight (Upper Right Chart), Days to 1st Certification (Lower Left Chart), and 3) Days to 1st Delivery (Lower Right Chart), it does not appear to be the case. Instead, we find the A380 completed more quickly than its sister aircraft, the A350, at less than half its Operating Empty Weight. In this context, its Development Schedule looks reasonable. ***What About Production?***

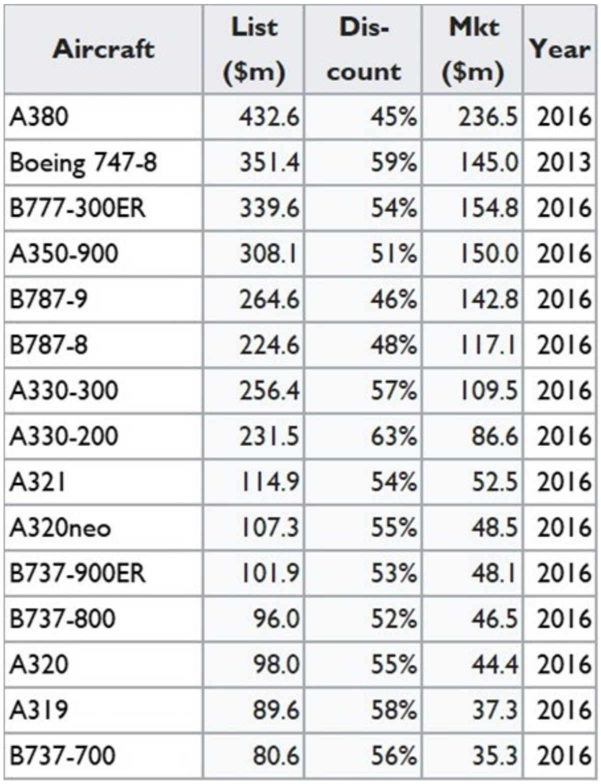
The break was due, in large part, to the differences in each firm's market projections, which we can observe in Figure 15. [23]

In 2000, as Airbus was about to launch the A380, they took little notice of how their market reacted to other products.

Had they examined their prime competitor, the Boeing 747, over the then past 20 years, they would have discovered the following in Figure 16:

The then-current models of the competitor's jumbo jet, the Boeing 747-200, -300, and -400, all had a much lower price tag than the Airbus entrant, selling for roughly 25% to 35% less than the A380. That fact is valuable information and is not to be ignored.

The primary observation we can make about Demand and Demand Curves is this: a higher price will tend to make fewer sales than competitors with lower price tags. Airbus had projected to sell over 1.5 times the number of A380s as Boeing sold B747s in the preceding two decades. They justified their numbers as being a function of seat cost. But eventually, sales will be limited by prices, no matter how hard program management tries to convince others otherwise.



Aircraft	List ($m)	Dis-count	Mkt ($m)	Year
A380	432.6	45%	236.5	2016
Boeing 747-8	351.4	59%	145.0	2013
B777-300ER	339.6	54%	154.8	2016
A350-900	308.1	51%	150.0	2016
B787-9	264.6	46%	142.8	2016
B787-8	224.6	48%	117.1	2016
A330-300	256.4	57%	109.5	2016
A330-200	231.5	63%	86.6	2016
A321	114.9	54%	52.5	2016
A320neo	107.3	55%	48.5	2016
B737-900ER	101.9	53%	48.1	2016
B737-800	96.0	52%	46.5	2016
A320	98.0	55%	44.4	2016
A319	89.6	58%	37.3	2016
B737-700	80.6	56%	35.3	2016

Figure 14: To attract enough customers, both Boeing and Airbus have to offer significant discounts

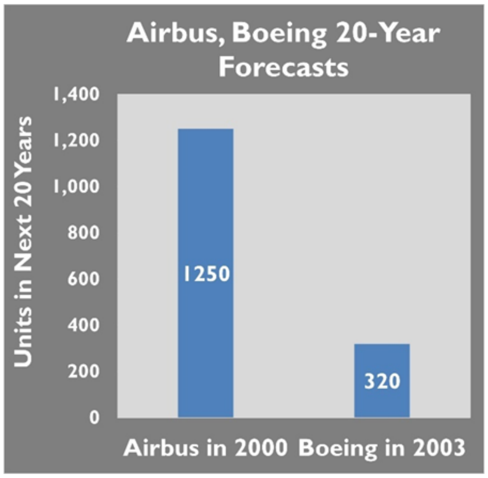


Figure 15: Airbus saw a market for its product that was nearly four times Boeing's projection

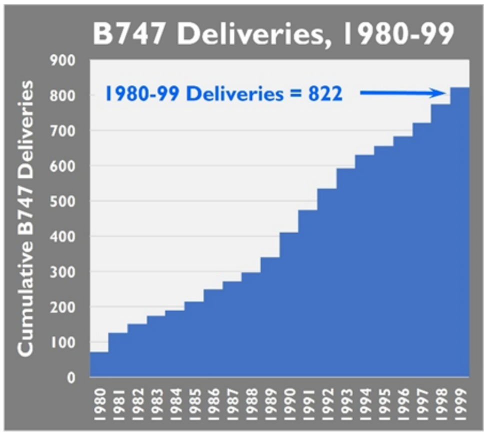


Figure 16: Boeing sold 822 B747s in the 20 years before the Airbus A380 launch; Airbus wanted to sell over 1.5X as many A380s in the following 20 years

In 2000, the world witnessed a multibillion-dollar example of this phenomenon. That year, Northrop Grumman completed the 21st and last example of their B-2 bomber. As revealed in Figure 17, the United States Air Force (USAF) originally wanted 132 vehicles, enough to form 11 squadrons of 12 aircraft each. However, the eventual recurring price of $1.2B was substantially more than the B-1B bomber, which the USAF purchased 100 units. While Northrop Grumman might have argued that they did not plan for the price to go so high, when it did, it came with consequences. The USAF Fighter/Bomber/Attack Aircraft Demand Frontier, which changed by about 2% from 1996 to 2021, was and remained a barrier to the number of units the service branch could absorb. While the variability about the curve allows for some margin of error, it did not allow for going over 6 times past (132 units (the original target) divided by 21 (the number delivered) = 6.29X) that limit.

Not surprisingly, we see the same behavior in commercial markets, specifically for airliners. In Figure 18, we plot 20-year quantities (from 1/1/1980 to 12/31/1999) and prices (which, in this industry, are much harder to find than like figures for USAF aircraft, as the United States Government (USG) must publish this data) for the then-current airliners for sale in 2000. We had to combine configurations (or "Dash Numbers," as they say in the aviation industry) to get the entire series produced over the period.

When we do, we find the 20-year Demand expressed by Equation 3:

$$1999\$M = 197Qty^{-0.188} * \varepsilon \qquad (3)$$

Where:

1999$M = Aircraft Model Price in 1999$M
Qty = Quantity sold from 1/1/1980 to 12/31/1999
ε = Error term for the equation

Equation 3, while not well-correlated with an R^2 of 45.8%, has a P-value of 3.5%, just below the 5% threshold typically used for this metric. The implication is that the opening position of the A380 sales target was about 9.7 standard deviations past its mean. Though widely off target, this miss approximates the error of 9.6 standard deviations calculated for their prediction of Development Cost. Note the eventual sales figure of 251 units is still vastly past the Demand Frontier.

Airbus A380 Summary

It is a too frequently appealing idea to find a product metric in which your firm excels and assume that it alone will draw in more customers than your competition. For Airbus, that measure was the cost per seat mile. While that is no doubt a crucial factor, there are always other market forces at work which suppliers must consider.

Supposing you will sell over 50% more units than your closest competitor with a product priced much higher is not borne out by Demand Frontier analysis. Without doing that work in advance, you might imagine you could exceed that limit by nearly 10 standard deviations. And that is what Airbus did. In the end, they lost tens of billions of Euros. But, for the lack of a detailed Demand Study before launch, the whole fiasco need not have happened.

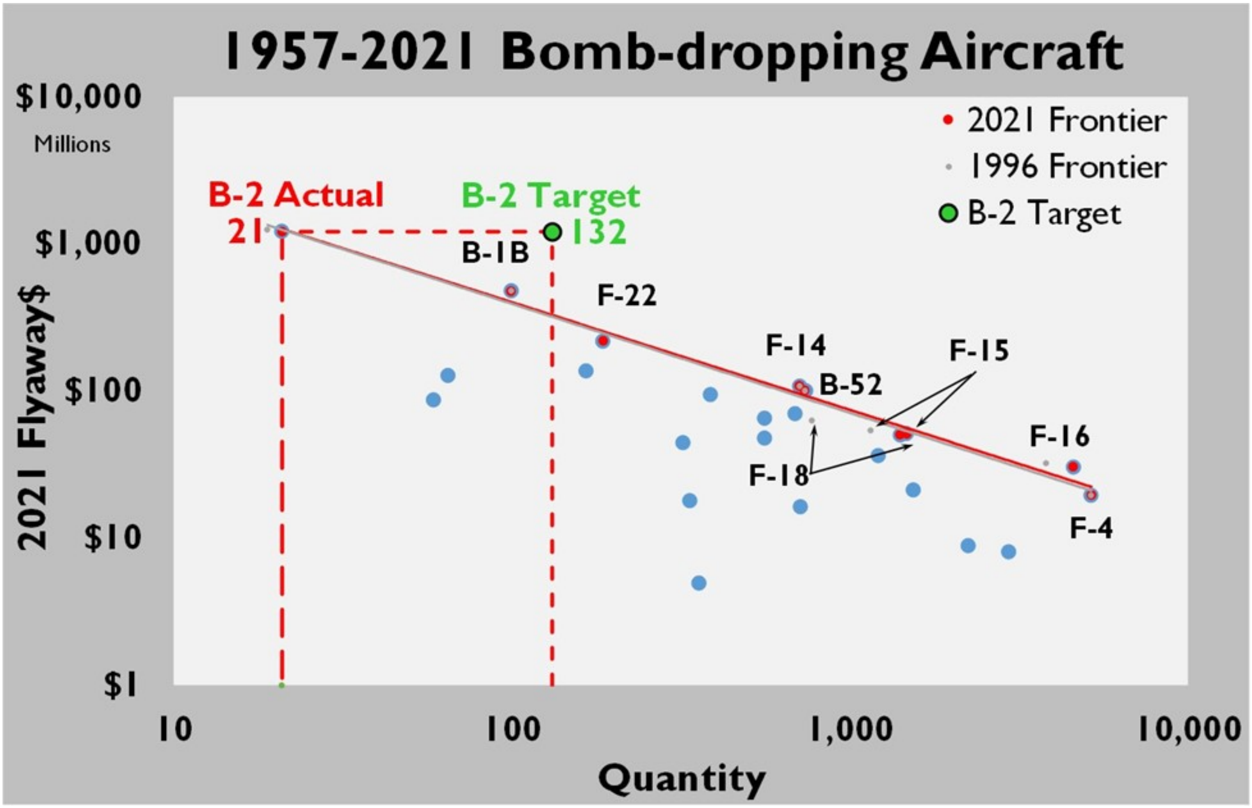


Figure 17: Northrop Grumman and the United States Air Force wanted to build 132 B-2 bombers. But, as the price rose, they found their sales limited to 21 units, almost precisely what the Demand Frontier limit was.

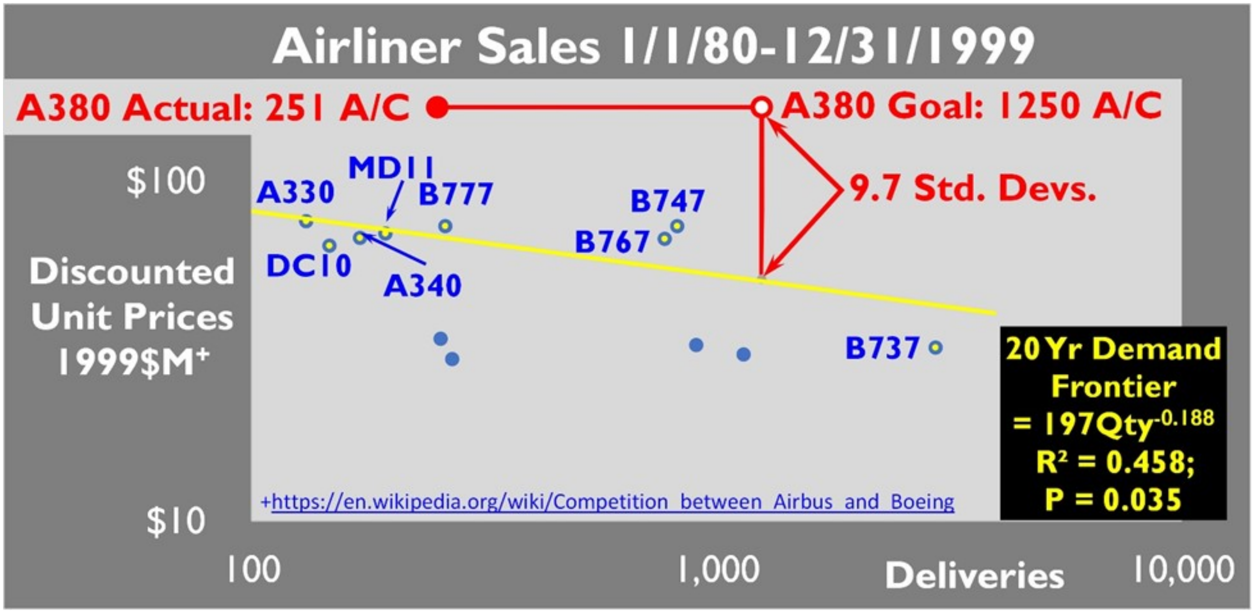


Figure 18: Airbus hoped to sell 1250 aircraft at their target price. But even their discounted price exceeded the Demand Frontier by 9.7 Standard Deviations at their goal quantity. Even the eventual sales figure of 251 units was vastly beyond the Frontier, implying many, if not most, or even all, sales at a loss.

Figure 19: The DeLorean DMC-12 [24]

Precursor: The DeLorean Debacle

It's not as if Europe had not previously seen how improperly constructed business analyses could lead to financial disaster. Less than two scant decades earlier, they endured the rapid rise and quicker demise of the DeLorean Motor Company.

The DeLorean DMC-12, pictured in Figure 19, with its gull-winged doors, mid-engine, and stainless steel body, was the brainchild of John DeLorean. The youngest person to become an executive at General Motors, he went to Northern Ireland to pursue his goal of building his innovative machine. He thought the gull doors, stainless steel body, and mid-engine design would attract sufficient buyers, similar to Airbus's thinking a lower cost per seat mile would attract customers.

To its credit, Airbus offered its customers an excellent value proposition with its A380, as the vehicle offered substantially more range and was slightly faster than its Boeing 747 counterparts.

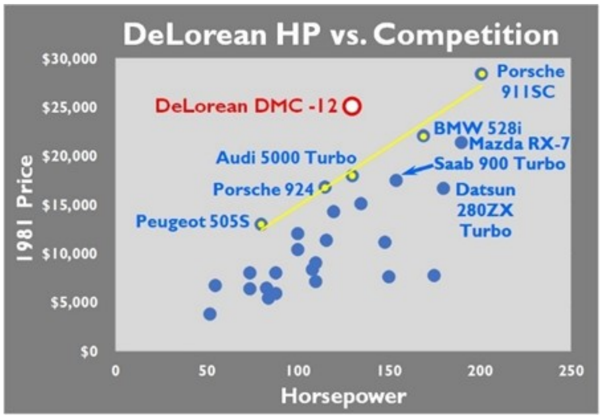


Figure 20: The DMC-12 wanted a lot more money per Horsepower than its competitors

Here is one critical area in which DeLorean failed to understand the business proposition from his customers' points of view. After many schedule delays and cost overruns, DMC-12 production began in late 1980. While good looks and innovation will always draw car buyers, those who buy sports cars want Horsepower—and the DMC-12 did not have nearly enough of it. Figure 20 shows the horsepower ratings and prices for the leading cars in 1981.

Usually, several features determine the Value or sustainable price of a product. For 1981 cars, that came down to Horsepower and the number of units sold, as depicted by Equation 4.

$$1981\ \text{Price} = 8546\text{HP}^{0.494} * 1981\ \text{Qty}^{-0.197} * \varepsilon \quad (4)$$

Where:

1981 Price = 1981 car model sales price
H/P = Installed Horsepower on each model
1981 Qty = Quantity sold in 1981
ε = Error term for the equation

In Figure 22A, we see that the set of features DeLorean put forth was worth, according to Equation 4, only $15,500, compared to its list price of $25,000. That puts it 1.43 Standard Deviations past its prediction (($25,000-$15,500)/$6,645 = 1.43). That proved to weaken Demand, as we'll see presently. Figure 22B shows us that to reach the desired target, without considering the Demand Frontier, DeLorean should have taken the vehicle up to 262 horsepower. Importantly, as we might have guessed, the phrase "without considering the Demand Frontier" hints that we ought to analyze Demand thoroughly.

Figure 23 depicts the interaction between the Demand Frontier, which applies to and limits the entire market, and Product Demand. This curve shows how the market-determined Value of a 1981 car falls as more units are produced. In that year, Product Demand fell according to its exponent (-0.1971), equating to an 87.2% Learning Curve if it were one of those. That means if a firm has found itself with a Learning Curve of, say, 90%, the Product Demand Curve and its associated Learning Curve might intersect. Figure 23 reminds us that Product Demand Curves are always flatter than the Demand Frontier they collectively comprise.

As we discover in Figure 24, in 1981, DeLorean built past the Demand Frontier. The company made 7,500 DMC-12s, but the market's self-imposed unit sales limit of $25,000 that year was 6,000. As

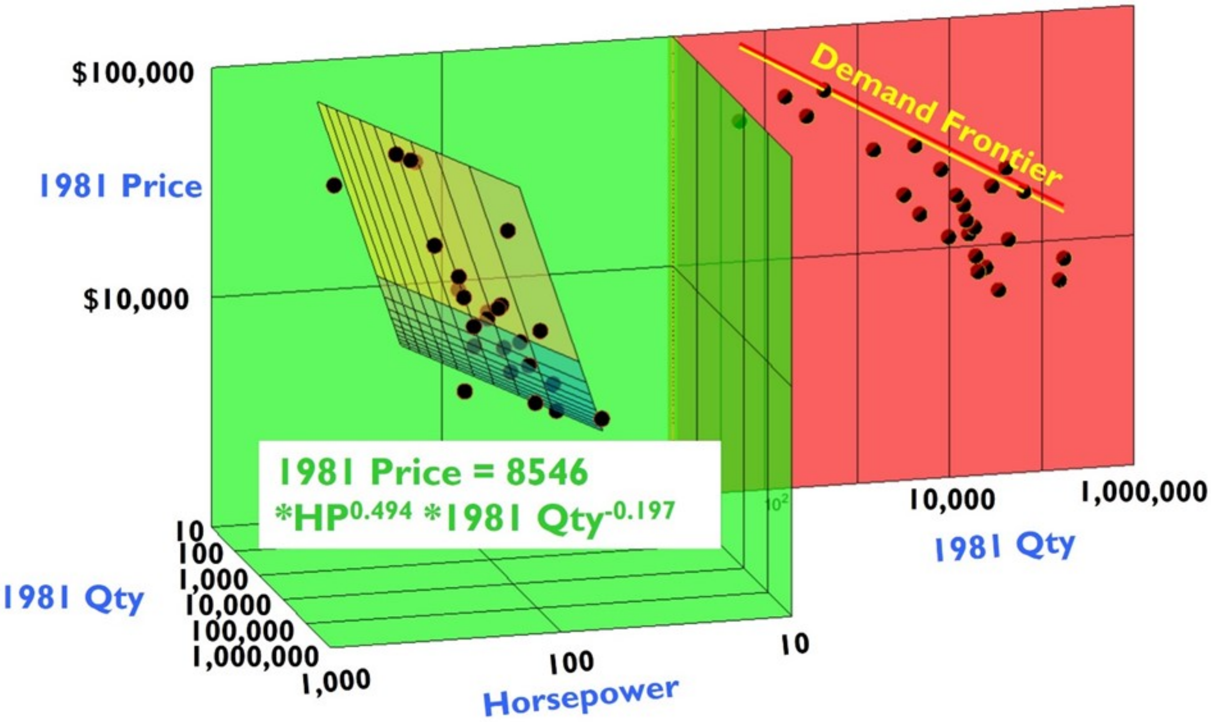


Figure 21: The prices for cars in 1981 went up with added Horsepower and down as quantities increased

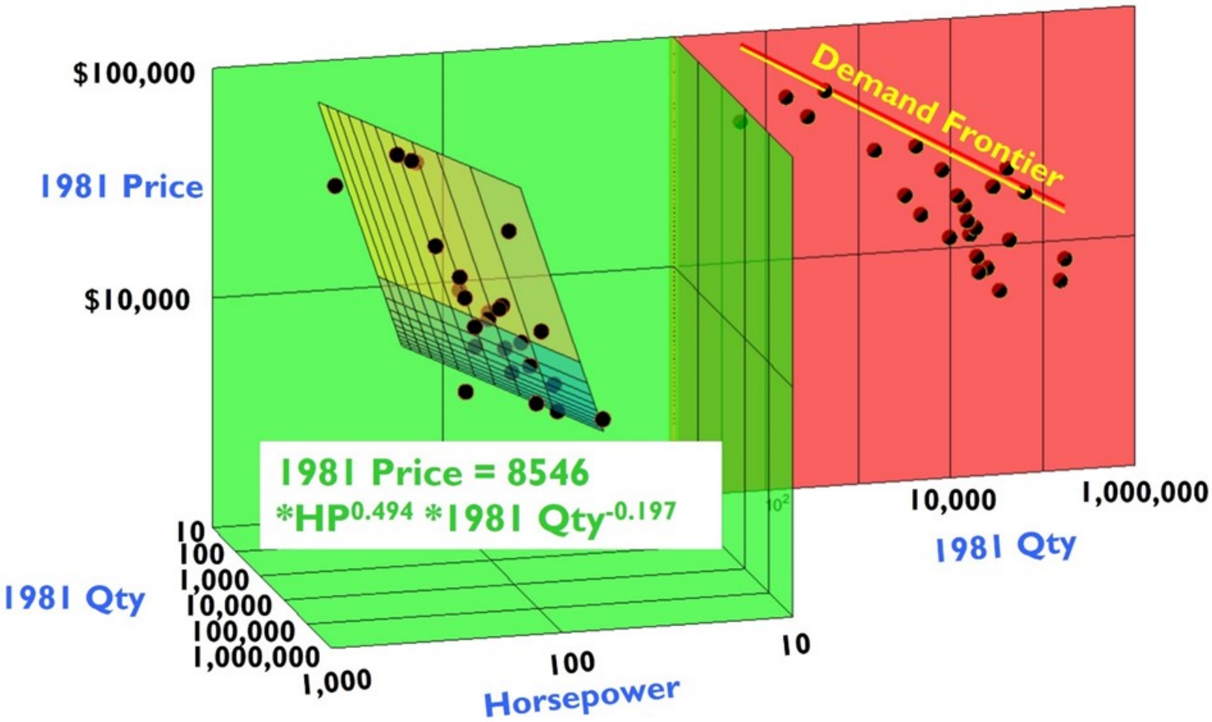


Figure 22: The DeLorean DMC-12 had insufficient Horsepower to sustain the price they wanted for the 7,500 units they hoped to sell that year. Their 1981 price was $25,000, but, as shown in A, that combination was only good for $15,500. In B, we discover that to make the $25,000 price, DeLorean would have had to install an engine with 262 horsepower, not the one with 130 they used equation 4's Adjusted R2 is 76.3%, a MAPE of 24.9%, a Standard Error of 6625, and a P-Value for the entire equation of 5.08E-08, and P-Values of 0.94% and 2.0E-05 for Horsepower and Quantity, respectively. This equation states that sustainable prices go up with Horsepower and down with Quantity, as shown in Figure 21.

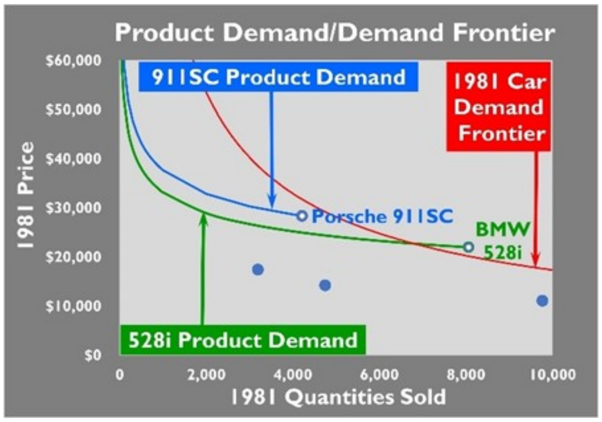


Figure 23: Product Demand Curves (as the Porsche 911SC and BMW 528i) are always flatter than their associated Demand Frontiers.

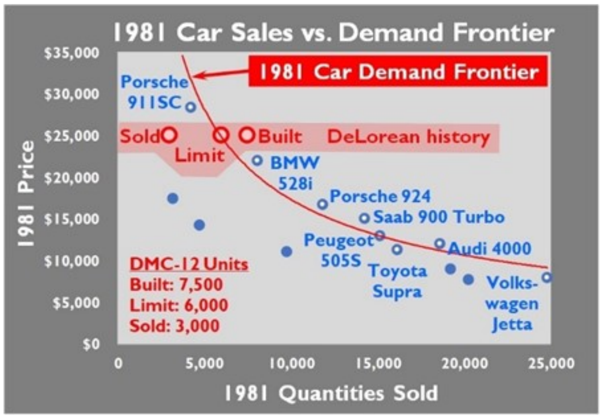


Figure 24: DeLorean built more DMC-12s (7,500) than the 1981 Demand Frontier would sustain (6,000) and ended up selling 3,000 units in 1981

always, not all firms can make it to the Demand Frontier, and in that year, DeLorean fell far short, only selling 3,000 models. They were left holding thousands of DMC-12s in inventory.

DeLorean DMC-12 Summary

There are many ways to sink a program, and DeLorean found most of them. Cost overruns can be fatal to a program, and the DMC-12 had them from the start and managed to get into production despite those setbacks. But, crucially, miscalculations regarding Value and Demand can be, and here were, equally devastating.

DeLorean bet its sleek design would be enticing, and, to an extent, it was. In the end, making a sports car requires a sporty engine. Value Analysis reveals they offered a little less than half of what they would need to sustain their price.

Disappointing for the firm was their miscalculation of Demand. Having not studied the applicable Demand Frontier, they were only too happy to attempt to exceed it dramatically. That approach seldom works, and it didn't work for DeLorean.

Conclusion

In market analytics, as in construction and rocketry, knowing where you are aiming is essential. Steps need to be taken to ensure that one's foundation is sound. The A380 and DMC-12 did not take the time to do that, just like the towers in Italy and San Francisco.

To date, market analysis has focused chiefly on cost and schedule. Both are crucial. Missing either of those targets by a large margin can make a program, or even a firm, go bankrupt.

The European firms Airbus and DeLorean Motor Company missed cost and schedule goals, impacting both greatly.

For the Airbus A380, well-researched analytics at the beginning of the program might have convinced the firm not to proceed. All the data needed to make that decision existed and was retrievable before the program launch. Firms in this pre-launch mode often rely on customer surveys to gauge market interest. They will take polls, sum up the results, discount them by some method, and then suppose they have a clear market picture. Ultimately, we should rely on buyers' actions, not their words. Observing past and present market reactions is the best way to predict future behavior.

DeLorean supposed, without analysis, that the beauty and innovation in his DMC-12 would more than make up for its lack of horsepower. It did not. Combining that error with guessing about Demand rather than analyzing it, the company created a recipe for financial ruin.

Between a firm's cost and schedule data, and the information its buyers reveal through their purchases, there is ample knowledge to refine new business cases compared to the ones done before.

Doug Howarth pioneered Hypernomics, the study of market actions across four or more dimensions. Hypernomics reveals the linked, opposing, non-physical forces known as the Law of Value and Demand, replacing the Law of Supply and Demand. The new field discovers buyers collectively form discernable patterns that may be statistically quantified and move over time. Analysis of those movements permits users to predict market positions with greater accuracy. Mr. Howarth has issued 15 peer-reviewed papers across four continents. His company, Hypernomics Inc., founded on his ideas, has worked for NASA, Virgin Galactic, Lockheed Martin, Raytheon Technologies, and Northrop Grumman. The US Patent Office issued US Patent 10,402,838 to him and two others in his company for the world's first 4D market analytic software. He has addressed the Royal Aeronautical Society and NASA three times each. Wiley published his book, "Hypernomics: Using Hidden Dimensions to Solve Unseen Problems," in 2024.

References

1. https://aviationweek.com/shownews/dubai-airshow/what-went-wrong-airbus-a380, retrieved Jan 31, 2023
2. Ibid.
3. https://www.netflix.com/watch/81594700?trackId=14170286, Episode 1, Retrieved February 5, 2023
4. https://commons.wikimedia.org/wiki/File:Kheops-Pyramid.jpg
5. Dash, Glen, "Occam's Egyptian razor: the equinox and the alignment of the pyramids," The Journal of Ancient Egyptian Architecture, Vol. 2, 2017, p. 1
6. Leaning Tower of Pisa. (2022, December 15). In https://en.wikipedia.org/wiki/Leaning_Tower_of_Pisa
7. Photo by Author, June 9, 2022
8. Millennium Tower (San Francisco). (2023, January 24). In Wikipedia. https://en.wikipedia.org/wiki/Millennium_Tower_(San_Francisco)
9. Van Derbeken, Jason (May 12, 2022). "Underground Wall Could Be Hitch for SF's Millennium Tower Fix Plan." NBC Bay Area.
10. https://www.globalsecurity.org/military/world/europe/a380.htm
11. Etsy, Benjamin C., and Ghemawat, Ghemawat, "Airbus vs. Boeing in Superjumbos: Credibility and Preemption," Harvard Business School Working Paper 02-061, August 3, 2001
12. Ibid
13. www.flug-revue.rotor.com/FRTypen/FRA3xxx.htm
14. Howarth, Douglas K., "The Checkmark Function," given to the 2008 ISPA Conference
15. https://en.wikipedia.org/wiki/Airbus_A380
16. https://calleam.com/WTPF/?p=4700
17. https://calleam.com/WTPF/?p=4700
18. Aboulafia, Richard, Airbus A380: The Death Watch Begins, Forbes, Jun 6, 2016
19. Goldstein, Michael, "Is Airbus Finally Ready To Shut Down A380 Production? Forbes December 27, 2017
20. https://en.wikipedia.org/wiki/Airbus_A380
21. https://en.wikipedia.org/wiki/Competition_between_Airbus_and_Boeing
22. DeLorean Motor Company. (2023, January 10). In Wikipedia. https://en.wikipedia.org/wiki/DeLorean_Motor_Company

What is the U.S. DoD Cost Estimation Community Saying About Agile?

Anandi Hira, CCEA®
Ben Kwok

The commercial industry experienced significant improvements in productivity, cost savings, and customer satisfaction by adopting Agile development. Some commercial organizations moved away from needing cost estimates for Agile projects, since they continue developing and improving their products until budget or schedule constraints are reached. However, the estimation and budget processes are imperative for the U.S. Government, particularly systems acquired through the Department of Defense (DoD) even if the contractors building the systems adopt Agile development. Agile adoption has posed big questions for the DoD cost estimation community: Are new cost estimation methods needed? How can we move away from existing source lines of code (SLOC) based models? Can DoD systems realize the benefits reported in the commercial industry? This literature review looks to answer these questions by reviewing papers and presentations from venues that primarily target the U.S. Government and DoD cost estimators. A total of 63 papers and presentations were found relevant to this study. The major categories addressed in this paper include: the benefits and challenges in adopting the methodologies within the U.S. Government acquisition process; empirical studies that compare performances between Agile and Waterfall projects; and cost estimation challenges and methods. The DoD is still fairly inexperienced at adopting the Agile development compared to the commercial sector, but development teams have made efforts to adopt Agile within the U.S. Government's existing Waterfall processes. This nested use creates challenges in identifying and predicting performance improvements. While there have been constant discussions on how to estimate Agile systems since 2010, very few estimation models have been published.

1. Introduction

The Agile Manifesto dramatically changed the way software is developed with both the promises and demonstrated proof of efficient software delivery and satisfied customers. Agile teams accomplish performance improvements by following methodologies and practices that support the Agile Manifesto and its Principles, such as: evaluating and reducing scope to fit budget and schedule needs and demonstrating progress and incorporating changes and feedback through delivered software [1]. However, the way the U.S. Government acquires and manages its systems, particularly through the Department of Defense (DoD), is very different from the commercial industry in several key ways:

- In most cases, the requirements are more important than the budget or schedule. In other words, even lower priority requirements cannot be excluded from the final product.
- DoD technical requirements and delivery schedules are driven by adversary threats, not the commercial marketplace.
- The estimation process is an essential part of the Government's acquisition and procurement process. The Government uses estimates to determine which programs to initiate and distribute annual funds.

These differences directly conflict with some of the ways commercial organizations execute Agile development. By continuously evaluating and limiting the scope, users and developers determine which features are most needed and do not expend resources unnecessarily. However, this might not be possible with projects concerning the U.S. Government and DoD as many of the requirements cannot be dropped over the lifecycle of the project. This distinction explains why the Government and DoD did not start talking about Agile development until 2010 (9 years after the Agile Manifesto was written in 2001) and has relatively few completed programs and systems that followed the Agile Principles.

This literature review looks through all papers and presentations from venues that target the U.S. Government and DoD cost estimation community to comprehensively understand and conglomerate how the DoD can apply Agile methodologies and practices to supply products and systems to the Government, whether it noticed performance improvements from adopting Agile development, and what estimation methods and models it uses. One goal of this literature review is to share the methods and challenges faced by the U.S. Government and DoD cost estimation community to encourage future research that addresses the challenges of Agile adoption and execution, and finds new estimation methods.

2. Background

2.1 Agile Software Development Methodology

In order to react to the increasing changes in technology and users' needs, a group of software developers came up with a way to speed software development and deploy more quickly to market/field. The group developed a manifesto and 12 principles to define the goal and main tenants to build software successfully [1]. The main tenants are to shorten the time it takes to get working software to users, and quickly and continuously get feedback from users. Several practices or processes (such as Scrum, eXtreme Programming, Kanban, etc.) were developed that adhere to the Agile Principles.

Two of the Agile Principles are to deliver working software between a couple weeks to a couple months and to allow requirements to change [1]. The Agile lifecycle model (see Figure 1) provides a process to follow both principles. The requirements as well as architecture and design are continually re-evaluated during a sprint, guiding what is developed, which is shared with the customers and/or users in preparation for the next sprint. This process continues until the customers/users are satisfied with the system or until schedule and budget constraints are met. In either scenario, the customers/users are happy with the outcome because they should receive working software with the highest prioritized functionality without exceeding schedule and budget constraints.

It does not necessarily make sense to try to estimate the effort and cost of the entire system upfront because all the requirements are not known, and the requirements can change throughout the lifecycle. However, Agile teams calculate their velocity (how much work they complete in a sprint) to plan the work for a sprint or the short-term.

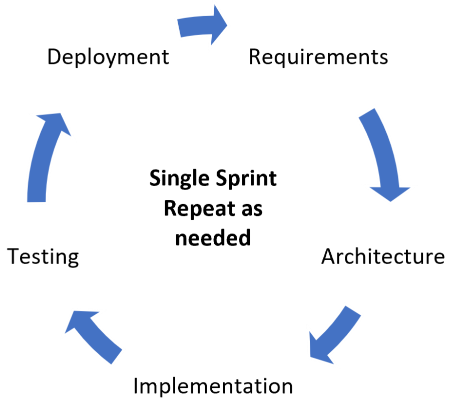


Figure 1.
Agile software development lifecycle model

2.2 U.S. Government's Budget and Cost Estimates Process

The Executive Branch formulates the Budget of the United States each fiscal year and submits it to the Congress. House and Senate Committees authorize and appropriate funding for the various Departments and Agencies including the Department of Defense (DoD). After the President signs the annual authorization and appropriation bills, DoD's appropriated funding is allocated to the various acquisition programs across the department.

The DoD's process to estimate and manage the resources it requires is called the Planning, Programming, Budgeting, and Execution (PPBE) Process [2]. The DoD must balance limited funding between resources needed for existing programs and systems, as well as resources to start new programs to meet future requirements. To obtain funding, a program must submit cost estimates for both near-term budgets and the overall life cycle of the program, which requires estimates at the total and annual levels. Additionally, large and high-interest programs and systems are required to submit updated estimates at specific milestones, subject to approval, to move on to the next phase of the program's lifecycle.

DoD programs have a Government Program Management Office (PMO) that oversees the cost, schedule, and technical execution (i.e., the meeting of requirements) by one or more contractors. Contractors must provide estimates when they respond to Request for Proposals (RFPs) and provide details on how they come to that estimate. The Government PMO usually develops an independent estimate using a variety of means including contractor provided technical and cost actuals. Additionally, independent cost estimates are completed at various other levels of the DoD to ensure the estimates are realistic. Having data and estimation methods at lower levels of detail assists in reconciling differences across multiple estimates to ensure they are independently evaluated and deemed realistic.

3. Related Work

This literature review covers the benefits and challenges of Agile adoption, as well as estimation in the U.S. DoD. An important note to make is that the benefits and challenges are from the viewpoint of the cost estimation community. There are several academic systematic literature reviews addressing challenges and estimation methods, but few studies which address the benefits or positive effects of adopting the Agile methodology. We will compare our results with these previously completed literature reviews later in the Literature Review Results section.

Only one systematic literature review addressed Agile adoption in the "public sector," which includes governments [3]. The authors of [3] list the benefits, challenges, and Agile practices from 17 papers [3]. In this effort, we look at grey literature (presentations and papers not peer-reviewed) which are shared at venues that typically target U.S. Government and DoD cost estimators for a more comprehensive, accurate, and updated view of their perspective on the benefits, challenges, and cost estimation methods being used.

4. Research Methodology

Literature reviews identify and evaluate research addressing a specific research question, topic, interest, or phenomenon. The results of the literature review provide a basis of the current foundational knowledge on the topic to provide relevant background, and evaluate for gaps for new research efforts [4]. This literature review explains the challenges the U.S. DoD and cost estimation community face in trying to apply Agile development while following the U.S. Government's acquisition and procurement processes. This review creates the basis of the existing knowledgebase to encourage future research to explore new ideas and methods that counter existing challenges.

4.1 Research Questions

Cost estimators work closely with program managers to develop a program's estimate, analyze how alternative solutions affect the program's estimate, analyze metrics to update the program's estimate and estimate future programs, and must understand how new processes and methodologies might in the future or have in the past affected performance to apply appropriate risk on estimates. Therefore, the questions addressed in this paper address all aspects of a cost estimator's interests and needs:

1. What benefits has the U.S. DoD experienced from adopting Agile development?
2. What are the challenges the U.S. DoD experiences trying to adopt Agile development?
3. In the commercial industry, the need for cost estimates reduced. Is the need for cost estimates reducing within the U.S. DoD?

4. Which size metrics and estimation methods is the U.S. DoD exploring or using? Which estimation models are shared with the community that organizations can use as a crosscheck or when estimating initial programs (when the organizations do not have their own historical data to leverage, yet)?

4.2 Research Process

Table 1 lists out the steps taken to complete this literature review.

Step 1: Set up	**Step 2: Manage references**	**Step 3: Notes Prep**	**Step 4: Findings**
• Identify topic and questions • Identify venues/ sources for papers and presentations • Identify search keywords • Find studies/ references	• Skim references for useful info • Copy/note useful info • Capture info and URL for bibliography • Identify major categories/ themes	• Create sections based on questions • Assign information to sections • Identify and assign to subsections, as needed	• Identify guidance, heuristics, findings, and knowledge base

To answer the questions, we identified the venues that specifically target the U.S. Government and DoD cost estimation community, listed in Table 2.

Venue	Website
International Cost Estimating & Analysis Association (ICEAA) Professional Development and Training Workshop	https://www.iceaaonline.com/
Joint IT and Software Cost Forum (formerly named, Software and IT-CAST Symposium)	https://www.dhs.gov/joint-it-and-software-cost-forum-2021
Practical Software and Systems Measurement Users' Group Workshop	https://psmsc.com/Default.asp
NASA Cost and Schedule Symposium	https://www.nasa.gov/offices/ocfo/cost_symposium
Agile SRDR (Software Resources Data Report) Subgroup	N/A

Table 2. List of venues and their websites

The last listed venue has semi-regularly scheduled meetings but does not have a publicly published website to keep an archive of presentations. Therefore, we were only able to utilize the presentations from the meetings we were able to participate in (via affiliation with the Space Systems Command (SSC)'s Financial Management Cost Research (FMCR) department).

Four out of five of these venues (all but the first one listed) are presentations-only type of venues. The first venue optionally accepts papers, but accepted papers is not a precondition to presenting findings or results of a research study, unlike academic conferences. The U.S. Government or DoD does not have an extensive database where the papers and/or presentations are stored, and nor do each of the venues listed above. In most cases, we had to peruse the conference proceedings or schedule for all past years' events and find sources that seem related to the topic of this study. Since the topic of this study is broad, the search terms are very simple. The search terms and the types of the 85 relevant studies found are in Figure 2. While 85 references were generally relevant, 63 of them answered the questions listed in the section 4.1. Research Questions.

Figure 3 shows the number of references found at each venue, including the duplicates. Most of the references came from both the ICEAA Professional Development and Training Workshop and Joint IT and Software Cost Forum.

The main categories and themes of the references, identified through Step 2 of the literature review (Table 1), are:

1. Cost Estimation
2. General/Experience (explaining what Agile is and/or experiences of how a team applied Agile development in their environment)
3. Metrics
4. Earned Value
5. Management
6. Data Collection

While most references covered multiple themes (e.g., General/Experience and Cost Estimation), one of the topics is usually the main objective of the paper or presentation (e.g., Cost Estimation). The results of Steps 3 and 4 are demonstrated through answering the research questions listed in the 4.1. Research Questions section in the following section.

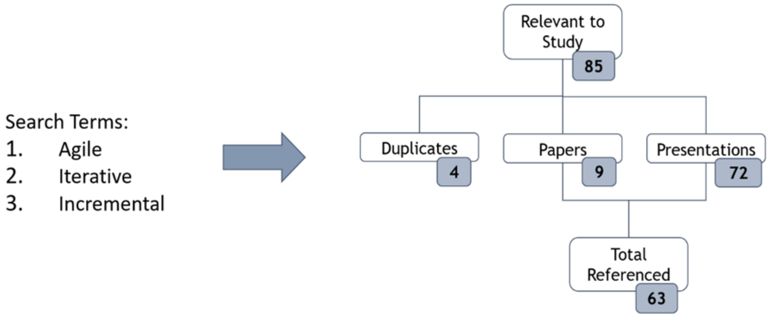


Figure 2. Search terms and demonstration of relevant studies found

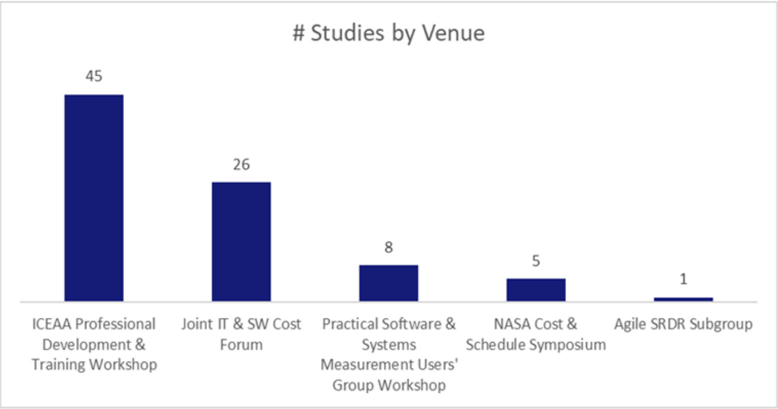


Figure 3. Number of references found by venue (including duplicates)

5. Literature Review Results

5.1 Benefits of Adopting Agile

While the commercial industry experienced several benefits as a result of Agile adoption, the U.S. DoD community questioned whether Agile can be applied to the large, complex systems the DoD needs to build, if the DoD community can provide "enough, up-front cost, schedule, and risk analysis to satisfy DoD regulatory and statutory requirements" and "support the persistent oversight and management requirements of DoD acquisitions," how to identify and have feedback loops with stakeholders and users, and so much more [5]. Essentially, the U.S. DoD was concerned whether it could meet the U.S. Government's acquisition and procurement requirements while adopting Agile development. Several references identified through this literature review analyzed the aspects and characteristics of Agile that could be applied in the DoD, while confirming that adopting Agile does not necessarily require the DoD to forego processes and steps that ensure product quality. Table 3 and Table 4 list out qualities and characteristics the references identified as part of adopting Agile development and those that would not align with the Agile Principles, respectively.

What Agile Is	References
Mindset, philosophy	[6] [7] [8]
Highly collaborative, self-organizing, cross-functional teams	[5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16]
Regular increments of functional software capability	[5] [6] [7] [17] [9] [10] [12] [13] [14] [15] [16] [18] [19] [20] [21] [22]
Flexibility and rapid response to change in requirements	[5] [7] [8] [9] [10] [11] [13] [15] [16] [18] [19] [20] [23] [24] [25]
Finding appropriate balance between anticipation and adaptation	[26]
Value/customer-driven or business/mission value Deliveries	[6] [7] [8] [17] [9] [11] [14] [16] [18] [19] [22] [27] [28] [29] [30]
Relies on traditional software development fundamentals	[27]
Reduced risk and uncertainty	[13] [14] [17] [18] [20] [28]
Reduced cycle time	[11] [28]
Increased transparency/visibility	[18] [25] [28]
Business value increases quickly	[14] [18]
Focus on eliminating waste and delays, “Just enough” upfront design to minimize rework	[5] [13] [14] [18] [19]
Involve users/customer/product owner throughout development lifecycle	[5] [8] [9] [11] [13] [16] [17] [18] [19] [21] [22] [24] [31]
Frequently tested software	[5] [6] [11] [29]
Focus on quality and security	[11] [14] [17] [29] [32]
Cross-section of project types, Cross-domain planning at program increments	[19] [29]
Continuous integration	[11]
Prioritize customer satisfaction	[7] [8] [11] [14] [19] [31]
Documentation is created as-built	[6]
Working product demonstrated to stakeholders at each iteration, demonstrate requirements met	[6] [13]
Break work into smaller, manageable segments	[6] [13] [17]
Better user experience	[6] [13] [17]
Measure progress on working software product	[6] [7] [14] [17] [19] [20]
Time-boxed/short iterations	[5] [7] [8] [13] [14] [21] [22]
Adaptive planning	[7] [13] [15] [17]
Face-to-face conversations	[9] [12] [14]
Delayed decisions until most info is available	[13]
Produces less documentation	[13] [17]
Establish Trust	[13]
Promoting user ownership	[13]
Motivation of knowledge worker	[19]
Incremental cycle	[5] [14] [19]
Decentralized decision makers	[19]
Scope is variable (cost and schedule are fixed)	[17] [22] [24]
Improved productivity	[12] [17] [33] [34]
Faster deployment to the field	[12] [17] [24] [34]
Produced fewer defects	[12]
Prioritized backlog (requirements)	[5] [14] [21]
Small teams	[5] [14] [21]
Prototyping	[5]
Co-located teams	[5] [14] [21]
Reduce integration cost due to continuous testing	[5] [25]
Reduce sustainment cost as defect rate is decreased	[5] [25]
Reduce code rework as customers are integrated into teams	[5]
Reduce incidents of massing cost growth as working code	[5]
Development at a sustainable pace (no overtime)	[8] [14] [21]

Table 3 List of characteristics that agree with the Agile Principles

The first column, under “What Agile Is,” in Table 3 contains a mix of practices, suggestions, and benefits. Of course, practices and suggestions are changes that can cause improvements or benefits. We took the information in Table 3 to create a causal graph that explains the practices and suggestions as causes of the benefits. The practices and suggestions listed in Table 3 fall into a few categories, which coincide with the Agile Manifesto. The Agile Manifesto states that the founders found higher value in 4 items over their counterparts [1]:

1. Individuals and iterations
2. Working software
3. Customer collaboration
4. Responding to change

Items 1 and 3 can be grouped together as high collaboration and items 2 and 4 can be grouped as incremental development. Besides these 2 categories for the practices and suggestions, we identified process improvement and quality from Table 3. The benefits can also be further categorized into short-term or immediate effects and long-term effects. Many of the benefits listed in Table 3 were direct effects of the practices and suggestions, while a few benefits may be side effects of applying the practices and suggestions or benefits that are not immediately realized. The resulting causal graph is in Figure 4. None of the references had listed lower costs as a benefit in writing. Though, participants would discuss whether the DoD could realize reductions in costs due to the nature of the systems (for example, requirements not necessarily being optional or optimal).

What Agile is Not	References
A Method	[6] [8]
“Magic bullet” that will make all issues disappear	[9] [16] [17] [35]
Unlimited or uncontrolled scope	[6] [12]
Unplanned	[6] [9] [12] [16] [20] [36]
Undocumented	[6] [9] [12] [16] [20] [36]
Unverified	[6] [12]
Mini Waterfall	[6] [12]
Trial and error	[6] [12] [36]
Synonym for flexible	[6] [12]
Synonym for fast	[6] [12]
Lack of formal requirements	[29]
Key benefit is cost savings	[17]
Necessarily more productive, faster, cheaper	[17] [31] [33]
Lack of discipline and process	[9] [36]
Just Scrum	[36]
Unarchitected	[9] [16]
Only good for small projects	[9] [16]

Table 4 List of characteristics that do not agree with the Agile Principles

Several empirical studies compared various performance measures (such as, productivity and schedule) between Waterfall and Agile DoD programs to determine whether the DoD can experience performance improvements claimed by the commercial industry (summarized in Table 5). While there are several studies that found Agile programs are more efficient, there are an almost equal number of studies that suggest that DoD programs do not experience significant improvements due to Agile adoption. Figure 5 quantitatively compares the number of studies that conclude the performance measures are better on

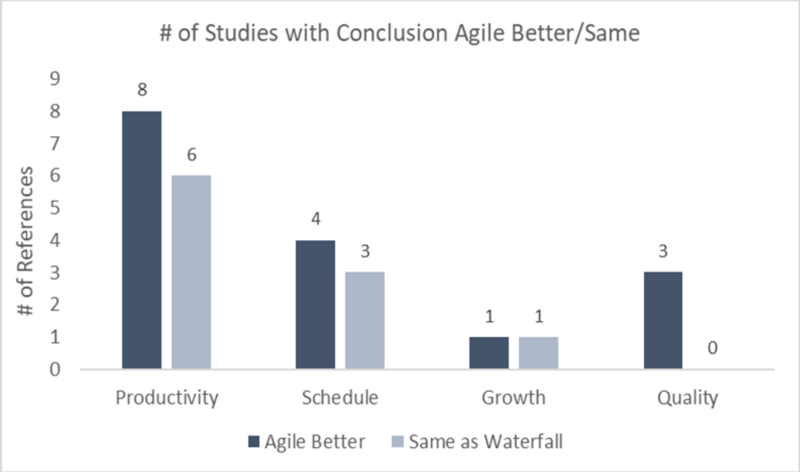


Figure 5 Comparison of number of studies that found Agile performed better than or similarly to Waterfall programs

Agile programs to the studies that conclude the performance measures are similar to Waterfall programs. Both Table 5 and Figure 5 demonstrate that Agile programs have the potential to perform better than Waterfall programs, but that it is not necessary or consistently experienced in the DoD.

The research question addressed in this section is: What benefits has the U.S. DoD experienced from adopting Agile development? The references identified for this literature review list several benefits the DoD can experience from Agile adoption, such as reduced risk, fewer defects, reduced sustainment costs, better user experience, and several more listed in Figure 4. Empirical analyses based on DoD programs/projects concluded that adopting Agile definitely led to improvements in product quality; and improvements in productivity and schedule can be attained, but not necessarily (Table 5 and Figure 5). Unfortunately, the empirical studies were not able to further analyze characteristics of the programs to identify why some benefited from performance improvements while others did not.

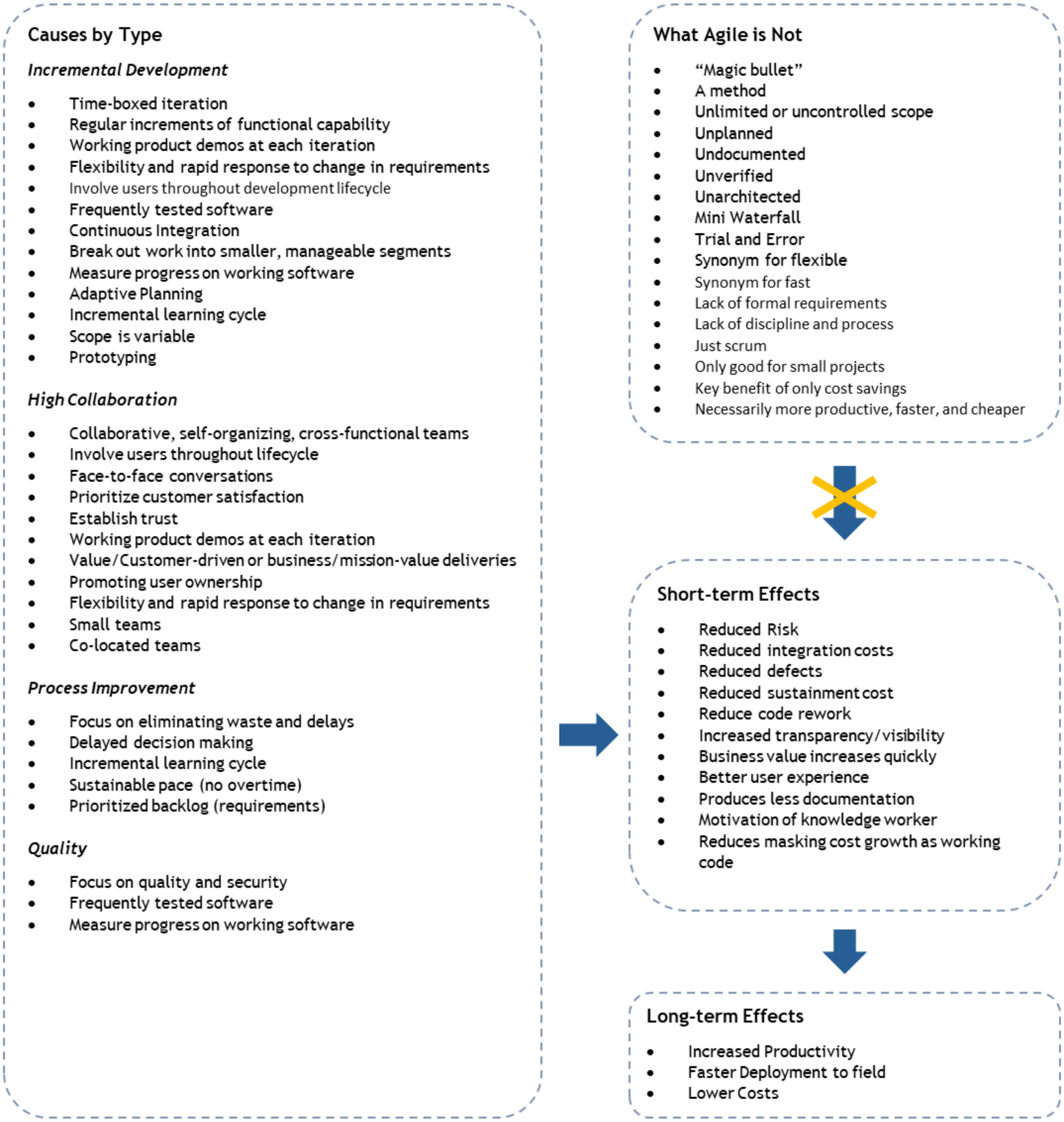


Figure 4 Causal graph and categorization of the Agile practices and benefits

Category	Improvements with Agile	Similar to Waterfall
Productivity/ Velocity/Cost	1. Productivity indices for Agile projects were significantly higher than the business average [12]. 2. Stable Agile teams are twice as productive [5]. 3. Agile programs more productive at component level [33]. 4. Agile very slightly outperforms Waterfall in productivity at high level. When grouping by application domains, Agile outperforms by about 15%. Agile team's velocity (requirements developed per month) is higher than Waterfall [34]. 5. Waterfall projects median productivity is 1.2 Function Points (FPs)/Staff-Month (SM). Agile is 2.8 FPs/SM. There is a 35% gain in productivity using Agile [37]. 6. Smaller projects (< $10M) tend to spend less resources per requirement – their requirements were generally less complex and defined at more granular level. Smaller projects have ~70-80% developers (less overhead) [38]. 7. U.S. average development productivity is 10 Function Points (FPs)/Staff Month (SM). This includes all activities from requirements through delivery. Average productivity for Agile teams is 36 FPs/SM. One developer can maintain approximately 750 FPs in 1 year. Developers can maintain about 1500 FPs on Agile projects [13]. 8. Development costs may be reduced by 10-20%, and productivity improved by 25% [16].	1. Agile projects use slightly more staff than non-Agile, though trend is very similar. Agile and non-Agile projects use nearly the same amount of project effort for projects with similar amounts of delivered functionality [12]. 2. Agile programs are not more productive at program level [33]. 3. Productivity not very different – slightly better for Agile, but this difference is removed for Government [31]. 4. Looked at SRDR (Software Resources Data Report) to compare Agile and non-Agile programs with Good or Good-Allocation Data Quality tags. Looked for analogous non-Agile data points for Agile and selected randomly from "Good" data points. Mann-Whitney U test and everything failed to reject null hypothesis (that there's a significant difference). No reason to believe (at this time) separate methodologies needed to estimate Agile programs. Differences in Initial and Final SLOC (source lines of code) per hour and hours per requirement between Agile and non-Agile data points not statistically significant [39]. 5. More cost and effort per Point (assuming Function Points because not clear in presentation slides) due to overhead and Points taking more effort than expected to finish. About 60/40 split between development/integration and overhead (Program Management, business support, functional Subject Matter Experts (SMEs), etc.) [38]. 6. Assuming program scope has been specified in a sizing metric, the cost risk of Agile software development is the same as any other program [24].
Schedule	1. Agile projects complete much more rapidly, schedule-wise [12]. 2. Agile process cut time to market in half [5]. 3. Agile average overrun is 16% while Waterfall is 25% [34]. 4. Exponent on Function Points (FPs) for schedule on Waterfall projects is 0.4. The exponent for Agile projects is 0.33 [13].	1. Agile is not faster at component or Program level [33]. 2. Looked at SRDR (Software Resources Data Report) to compare Agile and non-Agile programs with Good or Good-Allocation Data Quality tags. Looked for analogous non-Agile data points for Agile and selected randomly from "Good" data points. Schedule slip between Agile and non-Agile data points not statistically significantly different [39]. 3. Cost/Schedule overruns of 20-40% seem typical for Agile projects (rule of thumb is 20-30%) [38].
Growth	1. Hours growth is 50% across all projects, and 16% for Agile (based on only 14 data points, though) [24].	1. Cost, hours, software, and requirements growth between Agile and non-Agile data points not statistically significantly different [39].
Quality	1. Agile projects produced fewer defects [12]. 2. Teams using Scrum have 250% better quality [5]. 3. Average number of defects per Function Point (FP) is 5. The average for Agile projects is 3.5. When fixing a bug, on average, there is a 7% change the development team introduces a new bug. On Agile projects, there is a 2% chance. On average, 85% of bugs are removed during development or before the software becomes operational. On Agile projects, it's 92% [13].	

Table 5. Summary of studies comparing performance measures between Waterfall and Agile programs

5.1.1. Benefits Comparison to Related Work

Comparing the comprehensive list of benefits of Agile adoption from several systematic literature reviews [3] [40] [41] [42] [43] [44], the following benefits were listed in this study but not in the other literature reviews:

- Reduced risk
- Reduced sustainment cost
- Reduced code rework
- Produces less documentation, and
- Reduces masking cost growth as working code

Benefits listed from academic systematic literature reviews [3] [40] [41] [42] [43] [44], not mentioned in any of the presentations or papers reviewed in this study are:

- Acceleration of the development cycle
- Steady expenditure of budget
- Increased trust from customer
- Improvement in learning new technologies, and
- Knowledge-sharing

Some studies mentioned items that we considered practices or suggestions (e.g., manage changing priorities and better collaboration among stakeholders), and hence are omitted from the above list. One study found that 45.5% of Automated Information Systems (AIS) projects evaluated did not show increased efficiency, performance, and productivity or reduction in costs [45]. The authors of [45] also stated that the improvements in performance and reduction in costs are claimed among software developers and practitioners, but sufficient empirical evidence of these claims does not exist. Overall, the results in this study are similar to those from academic systematic literature reviews, with some changes in benefits due to the differences between Government and industry environments.

5.2 Challenges of Adopting Agile

The previous subsection looked at the benefits the DoD could and did experience from Agile adoption. This section addresses the challenges DoD programs and systems face in trying to adopt Agile, while following and fulfilling the U.S. Government's processes and requirements (the second research question). The challenges mentioned across the references can be grouped by role or type of challenge, with some overlap across these groups (as seen in Table 6).

The major causes of the challenges listed in Table 6 are:

1. The flexibility and delay in having an overall architecture/design and/or requirements
2. Trying to balance practices that ensure product quality (documentation and systems engineering) with Agile principle to do just enough
3. Uncertainty of which metrics to use to track projects or estimate costs, and inability to use existing data and models

Execution	Management	Cost Estimation
Lack of overall project design [5]	Lack of overall project design [5]	Lack of overall project design [5]
Product owner availability and access [35]	No industry benchmark data [35]	No industry benchmark data [35]
Replacing paper specifications with face-to-face customer meetings is not always suitable [13]	Key “value-added” metrics may not be identified or collected [17]	Key “value-added” metrics may not be identified or collected [17]
Upfront costs of setting up Agile, automated testing, infrastructure, and training [5] [25] [35]	Difficulty in determining velocity [35]	Predicting cost and schedule [7] [9]
Learning curve/optimal team performance requires time [35] [27]	Agile does not fit within budget cycles or the Government’s Fiscal Years well [35]	Lack of well-defined requirements and size metrics early in lifecycle [9] [23]
Issues due to not following systems engineering process [13] [30]	Establishing and monitoring metrics [9]	Do not know the efficiency of the workforce/development personnel [23]
Task seems to be on-time, until it is not and is pushed to the next increment [5] [30] [46]	Managing scope creep while maintaining flexibility [5] [7]	
Continually deferring functionality can cause failure to deliver on established cadence [13]	Evaluating and reporting progress [7] [9]	
Usual constraints still apply [6]	Capability Maturity Model Integration (CMMI) compliance [13]	
Different meanings or ways of implementing Agile development [13]	Prioritizing remaining work [7] [17]	
Poor management of technical debt can lead to complex and brittle systems [17]	Individuals may not be comfortable with having to provide detailed insight into daily activities [35]	
Just documenting at user stories level is not enough. Need cohesive and holistic documentation at system/sub-system levels. Need to determine right level of documentation [30]	Definition of done must be defined and “must have” and “nice to have” features identified; or the program may expend efforts on improvements with diminishing return [17]	
Organizational resistance, fear of change, individual skepticism, culture not conducive to Agile adoption [13] [35] [17]		

Table 6. Challenges of adopting Agile by role or type

The commercial industry may not face several of these challenges because the U.S. Government requires cost estimates at various times throughout a program’s lifecycle and requires the development teams to track and demonstrate their progress in meeting requirements. Additionally, most DoD systems require high levels of reliability and quality, as failures or errors in the software can cause significant damage, mission degradation, and/ or risk to human life.

5.2.1. Challenges Comparison to Related Work

Though this literature review primarily comes from the cost estimation community, many of the mentioned challenges apply to the general execution of Agile methodologies as well as program managers. Most of the challenges listed in Table 6 are focused on metrics. The challenges listed in academic systematic literature reviews generally come from the execution and program management perspective, and hence, their focus primarily is on the people (e.g., lack of senior management support and lack of communication) and development aspects (e.g., lack of automated tests and concerns with being able to maintain continuous testing and integration) [3] [47] [48] [49] [50] [51] [52] [53] [54]. Instead of identifying the challenges that do not overlap between this study and the academic systematic literature reviews, these are the challenges that were common between the 2:

- Product owner availability and access
- Upfront costs of setting up Agile, automated testing, infrastructure, and training
- Different meanings or ways of implementing Agile development
- Organizational resistance, fear of change, individual skepticism, and culture not conducive to Agile adoption
- Managing scope creep while maintaining flexibility
- Capability Maturity Model Integration (CMMI) compliance
- Definition of done must be defined
- Predicting cost and schedule (estimation)

The challenges identified in this study are mostly different from challenges identified in academic systematic literature reviews due to the difference in perspectives.

5.3 Reduced Need of Cost Estimates?

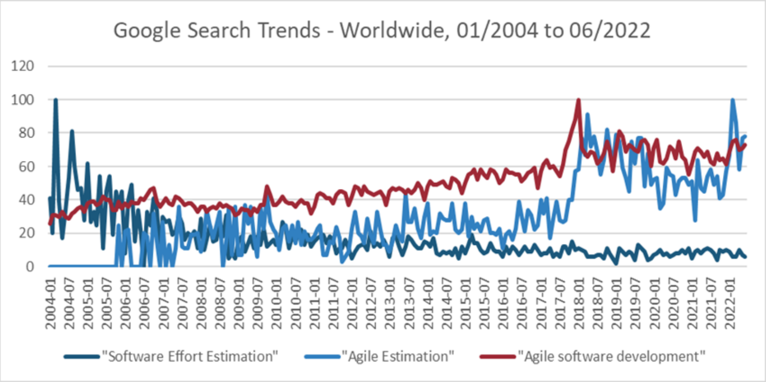


Figure 6. Google Trends on the topics "Agile software development", "Cost Estimation", and "Effort Estimation", normalized to the magnitude of the search term/topic – 100 represents the highest number of times the term/topic was searched, 50 represents half the occurrence at 100.

As the commercial industry adopted Agile development, the need for cost estimates reduced for some organizations as development teams work on tasks in order of priority and as long as resources are available. To demonstrate this trend, we looked at Google Trends for "Software Effort Estimation," "Agile Estimation," and "Agile software development." The results, in Figure 6, demonstrate the inversely proportional relationship of how many times "Agile software development" and software effort estimation is searched over time. On the other hand, the term "Agile Estimation" has been more widely searched since 2018. The data points represent the relative number of times the keywords or topics are searched with respect to itself. A data point with a value of 100 means that is the most times the term was searched, and 50 means it was half as many times as compared to the 100. The numbers do not indicate the actual magnitude of how many times the terms were searched compared to other terms.

The third research question is: Is the need for cost estimates reducing within the U.S. DoD? This question looks to see if the U.S. Government and DoD's need for cost and effort estimates is reducing as the use of the Agile development becomes widespread and common. As mentioned in the subsection 4.2. Research Process, we found 5 major categories and themes across the references. Figure 7 displays the number of references that fall into each of the categories/themes, excluding the duplicates. One of the presentations equally covered 3 topics: Cost Estimation, General/Experience, and Metrics. Hence, the references add up to 83, instead of 81.

Cost Estimation naturally has the greatest number of references, since the venues target the U.S. Government and DoD cost estimation community. In order to understand if the need or importance of cost estimates reduced as the use of Agile development increases across the U.S. Government and DoD, we combine the Metrics, Earned Value,

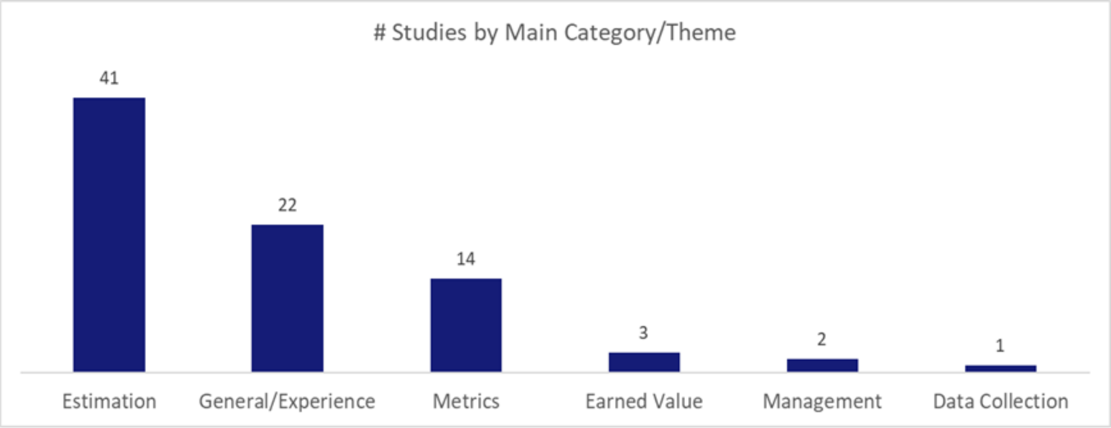


Figure 7. Number of references by major category/theme, demonstrating Estimation is the most popular category/topic

Management, and Data Collection categories/themes with General/Experience and compared to Estimation, shown in Figure 8. Both topics (Estimation and General/Experience) have grown proportionately over time.

The U.S. Government's acquisition, procurement, and budgeting processes have not changed as a result of Agile adoption. These processes require initial estimates in order to determine whether a program/system should be developed and then annual estimates to update the annual budget for programs and systems that have been approved. These estimates must be provided at the total and annual levels. No references implied a potential change in these processes since the Government must have a plan to distribute funds. As Richey stated, "While Agile approach is different from traditional software development methods (Waterfall), the needs for a high-quality, reliable cost estimate is still applicable for Government programs" [20].

Development teams and cost estimators still find estimates useful for several reasons, other than meeting the Government's requirements for estimates:

- Understand the capacity (the amount of work a team can accomplish in a certain amount of time) to prioritize work [20].
- Plan and commit to develop features in upcoming iterations and releases, which allow the team to plan to meet the customer's business objectives [8] [20].
- Make investment decisions [55].
- Compare alternative solutions or options [55].
- Verify or challenge vendor and contractor estimates [55].
- Control the program/project's costs [36] [55].
- Create benchmarks from completed programs' data and evaluate current programs' performance [55].

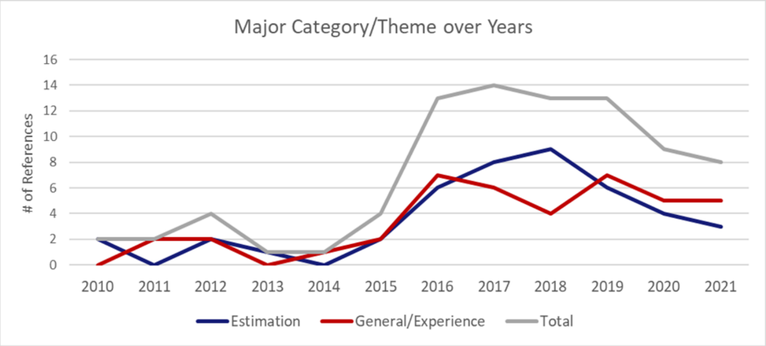


Figure 8. Number of references across major theme/topic by year, demonstrating that Estimation and General/Experience grew proportionately over time

- Prepare the market for the new features being developed [8].
- Manage and optimize resource allocations across multiple projects [8] [36].
- Provide status and progress, as well as show where the current status deviates from the plan [36].
- Provide early warnings to risks [36].

Therefore, in response to the third research question (Is the need for cost estimates reducing within the U.S. DoD?), the trend of the references indicate the answer is no.

5.4 Size Metrics and Estimation Methods Explored/Used

As explored in the previous subsection, cost estimates are required and needed to track progress, manage resources and programs, and to get necessary funds. While the DoD primarily used Source Lines of Code (SLOC) based cost estimation models for Waterfall programs, the DoD also recognizes that SLOC "does not fit the Agile framework well" [56] and has several drawbacks [21] [32] [56] [57]. Additionally, if Agile adoption leads to higher productivity, existing cost estimation models need to be appropriately adjusted to account for the changes. The references include a mix of presentations that make suggestions on which size metrics or methods seem the most promising and organizations sharing the size metrics and methods they are using. Figure 9 lists all the size metrics suggested or used in the U.S. DoD in order of granularity levels. The granularity levels of software size metrics depend on the details needed to calculate the size metric [58]. The size metrics are also grouped or color-coded by the major categories of size metrics – Agile metrics (blue), Function Points (dark blue), and Other (light blue with dark blue border).

Two size metrics in Figure 9 were created to provide possible estimation solutions for Agile projects or teams:

1. Messages: any communication between 2 elements/components of a system. These can be identified from the architecture/design [56].
2. Agilons: IFPUG Function Points where the complexity ratings are determined by the software development team in a Planning Poker fashion [65].

Figure 9 demonstrates 2 things: 1. The U.S. DoD is trying out several size metrics to find effective and suitable ways to estimate Agile programs and 2. The single-most suggested size metric is IFPUG Function Points. When

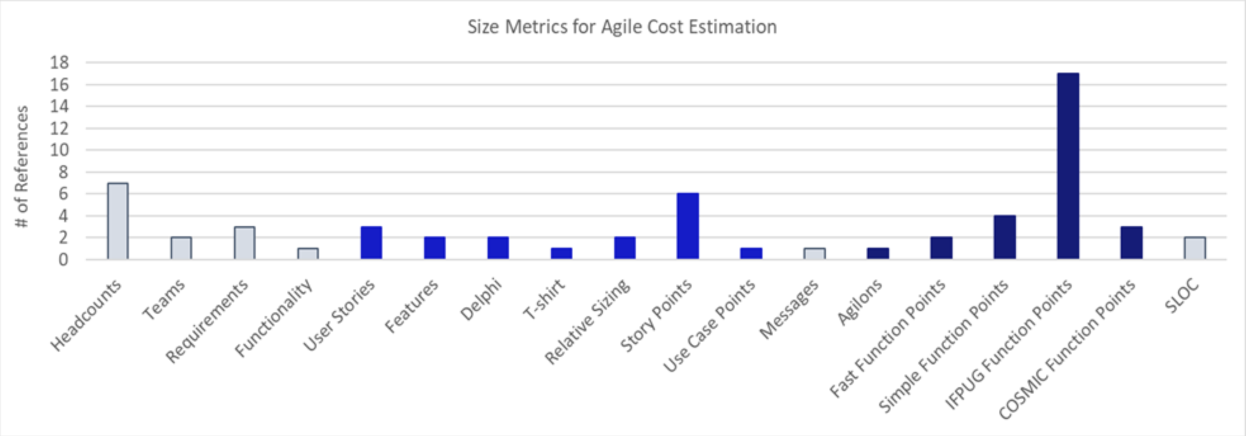


Figure 9 List of size metrics and number of references that suggested or used them within the U.S. DoD for Agile software cost estimation, in order of granularity level and color-coded by granularity level of size metric [58]. References used to create the graph: [6] [8] [9] [12] [13] [14] [15] [16] [19] [21] [23] [24] [26] [27] [28] [34] [36] [37] [46] [56] [57] [59] [60] [61] [62] [63] [64] [65] [66] [67] [68] [69] [70] [71] [72] [73] [74] [75] [76] [77] [78] [79] [80] [81]

Size Metric Category	**Agile**	**Function Points**
Size Metrics	User Stories Features Delphi (similar to Planning Poker but not necessarily Story Points) T-shirt Relative Sizing Story Points Use Case Points	Fast Function Points Single Function Points IFPUG Function Points COSMIC Function Points
Advantages	1. Convenient way to merge software size and complexity [8] 2. Easily map to requirements [56] 3. Easily map to build plan/schedule [56] 4. Suitable for Agile development environments [56] 5. Simplifies risk calculation [56] 6. Risk more manageable by Program Managers [56] 7. Tracking user stories provides an accurate inventory of delivered capability [70] 8. Excellent to facilitate communications with less technical stakeholders so team can get handle on the scope of requirements [8]	1. Uniform sizing [14] [21] [69] 2. Objective, not subjective, sizing [14] [21] [69] 3. Consistent regardless of team composition and experience [14] [21] [69] 4. Better measure of and predictor of velocity [14] [21] [76] 5. Independent of language, platform, and technology [14] [21] [69] 6. Can use to size, manage, and prioritize backlog [14] [21] 7. Helps users/clients quantify the number of requirements amended in software [76] 8. Size based on requirements [21] 9. Results vary by +/- 35% based on various factors including project's complexity and analysts' skill level [56] 10. Calculation error (difference between final and initial estimates) on average 14% (SLOC estimates error are 85% on average) [24] 11. Higher accuracy compared to Story Points for estimation [67]
Disadvantages	1. Subjective and cannot be replicated [14] [21] [72] [77] 2. Variation among teams [14] [21] [71] [77] 3. No rules on how to size [14] [21] 4. Inconsistent and unpredictable [14] [21] [77] · Story Points and T-shirt sizes most inconsistent across teams [69] 5. Not a size metric - accounts for effort, complexity, risk, and experience of estimators [71] [72] 6. Cannot be used to develop productivity metrics [14] [21] 7. Optimistic bias [14] [21] 8. Difficult to determine velocity [14] [21] 9. Does not support accountability in project management [77] 10. Does not provide a sound foundation for estimation [77] 11. Cannot be used to evaluate against industry data [14] 12. Frequently unknown beyond next few iterations [73] 13. Does not provide info regarding necessary team size [72] 14. Does not provide info regarding progress against target [61] [72] 15. Does not provide info regarding risk mitigation in outsourcing [72] 16. Story Points not available at beginning of lifecycle or at contract award because provided by development team [34] 17. Estimates based on these metrics tend to have the most risk [8]	1. Requires long counting time [56] [65] 2. Requirements need to be fully defined [65] 3. No flexibility to estimate in Agile [65] 4. To fully realize benefits, need to improve mapping of requirements to Function Points [56] 5. Not granular enough to apply (hinders exploring trade space or solutions space) [57] [65] 6. Does not assess/measure non-functional requirements [57] [69] 7. Not applicable for sizing bugs/defects [69] 8. Need to group data into categories (either application domains or development type) for better accuracy [46] [58] [82] 9. Non-intuitive to engineers, difficult to explain how estimates are derived, hard to diagnose why an estimate is wrong [57]

Table 7 Advantages and Disadvantages of Agile Metrics and Function Points for Agile cost estimation

combining the size metrics by group (Figure 10), both Agile metrics and Function Points are almost equally being considered. Table 7 goes over the advantages and disadvantages the references note for these 2 categories of size metrics for estimation.

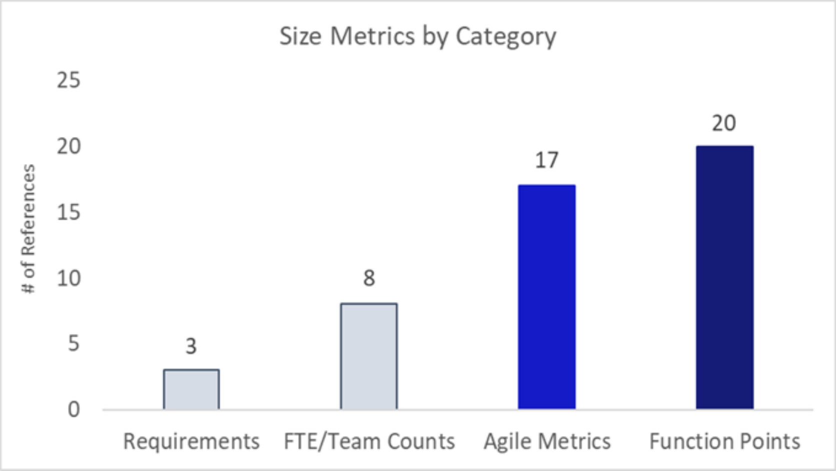


Figure 10. List of size metric categories and number of references that suggested or used metrics from the categories

Software development teams use Agile metrics as part of their development processes, therefore making it easier to collect the data through the lifecycle. However, Agile metrics are only available starting after the program has been initiated and handed off to the development team and cannot be used across teams or organizations. On the other hand, Function Points (regardless of the variant or type) can use descriptions of the system's functionality along with rules to ensure consistency and objectivity in the sizing. This information is available before the program has been initiated, when initial estimates are needed to get contract award from the U.S. Government. For these reasons, the U.S. DoD is primarily exploring both Agile metrics and Function Points for cost estimates.

While most of the references explore or use a single metric to perform all their estimation needs, a few references suggest using multiple size metrics but by different groups or at different parts of the lifecycle. These suggestions are listed, but generally suggest using Function Points for initial estimates and then either Agile Metrics or other metrics for the remainder of the estimates:

1. Use Function Points for fixed price contracts or for contract award estimates, and Story Points to size requirements and plan sprints at the team-level [26] [36] [72].
2. Use Function Points (can use Natural Language Processing to automatically count from requirements) for initial estimates needed to request contract award. Use Functionality while the requirements and architecture/ design are being negotiated. When the architecture/design has enough detail to identify the messages that are passed between elements or components of the system, can use messages to SLOC conversion ratios to use existing SLOC-based estimation models for estimates [56].

Along with size metrics, the U.S. DoD explored and uses tools or methods either along with or independently of size metrics. The 4 categories of these tools/methods are:

1. Proprietary cost estimation models/tools. The 2 mentioned are PRICE and SEER for Software, both of which allow Agile metrics and Function Points as size inputs [22] [68] [71].
2. Complexity and additional cost/effort drivers to build more accurate cost estimation models, regardless of the size metric being used [6] [7] [8] [12] [16] [68].
3. Natural Language Processing to convert requirements to Function Points [46] [63] [66] [67] and Machine Learning to build estimation models using available data [19].
4. Estimating using analogy, which is generally done by looking at similar programs to estimate a current program's costs [15] [36] [75] [83].

Size Metric	Model/Rule of Thumb	Usage/Method	Usage Needs
Headcounts/ Teams		• Team size × # sprints × team burn-rate [23] • Team size × sprint length × # sprints [6] [22] • # teams × team size × duration × cost per head [61] • Burn rate × duration [24]	• Need to know how many sprints needed [9] [24]. • Assume established/ consistent velocity [6].
Requirements	• Effort – 200 × Req0.512 × Staff 1.001 × D1.15 where D = 1 for mission support, 2 for Auto Info Sys, 3 for Engineering, and 4 for Real-time [78] • (Updated study to above): • Effort = 173 × Req0.539 × Staff0.463 × 2.3D1 × 3.7D2 × 3.9D3 • Schedule = 1.7 × Req0.34 × Staff0.19 × 2.3D1 * 3D2 × 4.5D3 • where D1=Auto Info Sys, D2 = Engineering, D3 = Real-time [34]		• Req = Requirements • Total Requirements = Functional + Interface requirements • Staff = Initial peak staff estimate • D#'s are for Application Domains
Functionality		• Size (unit) × productivity ($/unit) [24]	• Use size metric of choice

User Stories	• Can convert stories to SLOC and then use a SLOC-based estimation method (used by Quantitative Software Management (QSM)). Conversion ratios based on complexity [12]:
	• Low: 370 SLOC/Story
	• Average: 610 SLOC/Story
	• High: 915 SLOC/Story
		• If ongoing project, use user story from project/ historical staffing to complete estimate. If new project, use organizational averages to estimate how much functionality can be implemented based on headcount estimates [70].	• SLOC conversion method:
			  • Complexity of each User Story.
			  • SLOC-based estimating methodology.
			• Direct estimates:
			• Historical headcounts/ effort.

Delphi		• Identify requirements source. Employ disciplined Delphi method with participation from software subject matter experts (SMEs) and key Program Management Office (PMO) SMEs. They will provide low, expected, and high estimates to inform uncertainty analysis. Can use Monte Carlo to estimate schedule [57] [81].	• Analogous data (which is usually limited) [57] [81]. • New type of cost model required [57] [81]. • Sizable effort/ coordination needed to run scoring/ estimation session [57] [81].
Story Points	• Air Force's sizing guidelines [60]: • 0.5 Story Points: < 1 day • 1: 1-3 days • 5: 3-5 days • 8: 5-8 days	• Use at team-level for sizing requirements, planning sprints, calculate velocity [26]. • Express requirements in Story Points, rank/size Story Points, estimate and/or document velocity, spread sprints over time to develop time phase [22].	• Need to know how many sprints needed [9] [24]. • Assume established/ consistent velocity [6].
Use Case Points (UCPs)		• Express requirements in UCPs, rank/size UCPs, estimate and/or document velocity, spread sprints over time to develop time phase [9].	• Need to know how many sprints needed [9] [24]. • Assume established/ consistent velocity [6]. • Estimation model and/or historical data based on UCPs.

Table 8.1 List of Size Metrics, estimation models or rules of thumbs if available, explanations of how to use them, and needs for the suggested usage

Size Metric	Model/Rule of Thumb	Usage/Method	Usage Needs
Messages	•	• Messages are any communication between 2 elements of a system. Part of the system architecture and likely well understood. Found high correlation with SLOC. Requirements are mapped to the required number of messages, which are used to convert to SLOC to get effort estimates. Can classify new, modified, and reused messages [56].	• Messages to SLOC conversion ratios.
Agilons	•	• Development teams estimate the low/ average/ high ratings for the transactions of IFPUG Function Points in a planning poker fashion [65].	• Estimation model and/or historical data based on Function Points (FPs).
Simple Function Points	•	• Use AI/NLP to map action verbs to Simple Function Points components – automated calculation [59] [63] [66]. • Use historical and industry data to determine throughput rate and estimate effort and schedule [59] [63].	• Estimation model and/or historical data based on Simple Function Points.
IFPUG Function Points (FPs)	• Rules of thumb from [13]: • Productivity: 36 FPs/Staff Month (includes all activities from requirements through delivery) • 1 developer can maintain about 1500 FPs per year • Schedule (months) is approximately FPs$^{0.33}$	• Northrop uses a tool called MARINE (Machine Assessed Requirements Inspection and Evaluation) developed by Logapps to count Function Points from user stories [71]. • Tool called ScopeMaster developed to automatically count IFPUG and COSMIC Function Points from requirements. Also able to detect incomplete requirements, users and objects, find problems and suggest changes, and propose functional test cases [67].	• Estimation model and/or historical data based on FPs.
COSMIC Function Points	•	• Individual user stories can be measured in COSMIC Function Points. Sizes can be added up for roll ups. Use in place of Story Points [77].	• Estimation model and/or historical data based on COSMIC Function Points.

Table 8.2 Continued List of Size Metrics, estimation models or rules of thumbs if available, explanations of how to use them, and needs for the suggested usage

Table 8 lists out most of the size metrics the U.S. DoD's cost community has explored (from Figure 9), and when appropriate, how a method (such as Natural Language Processing) could be used with the size metric. Table 8 puts everything together in one place – the size metrics the U.S. DoD's cost community has explored or is using, cost estimation models or rules of thumbs that have been shared for use, general guidance on how to use the size metric, and what an organization or cost estimator would need to be able to use the provided guidance.

The fourth research question addressed in this subsection is: Which size metrics and estimation methods is the U.S. DoD exploring or using? Do estimation models exist that organizations can use as a crosscheck or when estimating initial programs (when organizations do not have their own historical data to leverage, yet)? The response to this is: the U.S. DoD is exploring Agile metrics and various variants of Function Points to find methods that can be used consistently at different phases of the lifecycle and potentially across organizations. Despite the number of references that addressed Agile cost estimation, very few estimation models or rules of thumbs are provided in the references (see Table 8). The lack of estimation models being shared in the U.S. DoD cost estimation community poses a challenge for organizations. While most organizations have their existing SLOC-based estimation methods and models, they do *not* have data to develop new models based on either Agile metrics or Function Points. They would need to wait until enough programs complete before having enough data points to develop estimation models.

5.4.1. Cost Estimation Methods Comparison to Related Work

Nine literature reviews looked at estimation methods, size metrics, and cost drivers used for Agile estimation reported in previously published papers. The most popular estimation methods used are:

1. Regressions or models using some kind of size metric (such as Use Case Points, variants of Function Points, and Source Lines of Code (SLOC)) [84] [85] [86] [87] [88] [89]
2. Expert judgment [84] [85] [86] [87] [88] [89] [90]
3. Planning Poker [85] [86] [87] [88] [91], and
4. Neural networks [84] [87] [91].

The 3 most looked at size metrics are:

1. Story Points [84] [85] [86] [88] [90] [91] [92]
2. Function Points variants [84] [85] [86] [88] [89] [91] [92], and
3. Use Case Points [85] [87].

Finally, the most considered cost drivers – ignoring size – are the team's experience [85] [86] [91] [93] and algorithmic complexity [91] [93].

The results from this study are very similar in terms of most looked at size metrics and cost drivers, but slightly different in estimation methods. The academic papers focus on subjective estimation methods (expert judgment and Planning Poker) and neural networks, which do not provide explicit explanations or reasoning. However, the U.S. Government and DoD need to be able to replicate their cost estimates, explain how they reach an estimate, as well as explain the data and assumptions underlying the models used.

6. Threats to Validity

This systematic literature review attempts to understand the current knowledge base and status of Agile adoption and Agile estimation within the U.S. DoD. Therefore, this study has identified venues that specifically target the U.S. Government and DoD cost community. The results presented in this paper can be influenced and biased by a few factors:

1. Not all organizations share their methods and findings at these venues;
2. Some organizations consistently share their methods, findings, and suggestions on a regular basis (with some changes and updates);
3. Government-only and/or government-sponsored cost estimation working groups do not have websites or a public archive of past meetings; and
4. There may be additional government-only and/or government-sponsored cost estimation working groups (similar to the Agile SRDR Subgroup) of which we are not members.

Therefore, the conclusions made in this literature review are limited to methods and analytics that have been publicly shared and archived. This limitation does not necessarily detract from the study's results. Informal discussions and observations led to similar conclusions: organizations either use their existing SLOC-based estimation methods or analogy and expert judgment due to the lack of sufficient data or openly available models to initially use.

7. Conclusions

The U.S. DoD and Government can realize benefits by applying the Agile development methodology. However, improvements in productivity and reduction in costs and schedule is not guaranteed or as a default result. Reduction in risk, uncertainty, and sustainment costs as well as improved customer satisfaction and other direct effects of applying Agile development have been experienced by the U.S. DoD; thus, leading to continued Agile adoption in the U.S. DoD. Further research may indicate why some Agile projects experience productivity, costs, and schedule improvements while others do not. This inconsistent behavior leads to challenges when building cost estimates, as all stakeholders may expect to see reduced costs and schedules in the proposals; but there is no basis to determine whether such benefits can be experienced by a particular team or program.

While Agile practices embrace the fact that little is known upfront, Government processes require early understanding of the baseline for required estimates, budgeting, and progress tracking. Hence, developers, managers, and estimators have faced several challenges in Agile adoption within the U.S. DoD, particularly in trying to balance between Agile and structure, determine which metrics to use, and how to build cost estimates.

Hence, the U.S. DoD cost estimation community is still looking for estimation methods and models that allow them to build estimates using information available to them and using metrics that can be available when estimates are needed. However, the biggest challenge facing the community, is the two-fold: the lack of agreement or standard in the metrics being collected; and their availablityavailability to build and use cost estimation models.

Both Agile and Function Points are being considered for estimation purposes. A couple problems challenge the widespread use of Agile metrics:

1. These metrics are not consistent across teams, much less across organizations. The team compositions may change from program to program within an organization, making it difficult to collect and effectively use these metrics at a larger scale. Plus, this personnel turnover limits the ability to create generalizable cost estimation models, ones useable and applicable across different organizations, types of programs, and application domains.
2. These metrics would not be available until the development teams start working on the program. However, the DoD (or Government Program Management Office (PMO)) must submit estimates prior to the program being awarded to contractors. Hence, the Government program management office would not be able to use cost estimation models based on Agile metrics.

All variants of Function Points can be calculated using the primary functions of the system but can be tedious to calculate. Therefore, software development teams may not be inclined to adopt Function Points to manage and track their progress. As mentioned in section 2.2 (U.S. Government's Budget and Cost Estimates Process), the contractors and several levels of the Government perform independent cost estimates. Given the pros and cons of both Agile and Function Points as well as the needs of cost estimates, the solution will probably need to be a combination of methods. For instance, the DoD can use Function Points, requirements, or some metric based on the functionality (since the Government provides initial, high-level requirements) for its independent estimates, while contractors can use Agile metrics and past performance for their estimates.

The purpose of this literature review is to understand the benefits, challenges, and cost estimation needs the U.S. DoD has for Agile adoption through the four research questions posed. The goal of sharing these insights is to welcome and encourage research in how the Agile development methodology can be adjusted to meet the U.S. Government's needs and cost estimation methods and models for the U.S. DoD to use.

8. Acknowledgments

The authors thank Raj Palejwala, Adriana Contreras, and Ernest Rangel of the Space Systems Command (SSC) for supporting and reviewing this study. The authors also thanks Matt Murdough of Tecolote Research Inc. for reviewing, providing suggestions, and guiding this study from initiation to submission. This study was funded by the Space Systems Commands (SSC) under contract FA8802-21-F-0004.

9. References

[1] M. Beedle, A. van Benekum, A. Cockburn, W. Cunningham, M. Fowler, J. Highsmith, A. Hunt, R. Jeffries, J. Kern, B. Marick, R. C. Martin, K. Schwaber, J. Sutherland and D. Thomas, "Principles behind the Agile Manifesto," Agile Manifesto, 2001. [Online]. Available: https://agilemanifesto.org/principles.html. [Accessed 5 November 2021].

[2] Defense Acquisition University, "Planning, Programming, Budgeting & Execution Process (PPBE)," [Online]. Available: https://www.dau.edu/acquipedia/pages/ArticleContent.aspx?itemid=154. [Accessed 24 June 2022].

[3] I. Vacari and R. Prikladnicki, "Adopting Agile Methods in the Public Sector: A Systematic Literature Review," in *Embrapa Informática Agropecuária-Artigo em anais de congresso (ALICE). In: INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING AND KNOWLEDGE ENGINEERING*, Pittsburgh, 2015.

[4] B. A. Kitchenham, D. Budgen and P. Brereton, Evidence-Based Software Engineering and Systematic Reviews, vol. 4, CRC Press, 2015.

[5] E. Hawkes and C. J. I. Kupec, "A Primer on Agile Software Development for Cost Analysts," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2016.

[6] B. Hunt and M. Thompson, "Applying Earned Value to Agile Development Programs," in *International Cost Estimating and Analysis Association 2016 International Training Symposium*, 2016.

[7] B. Smallwood and O. Mahmoud, "Agile Project Management Controls: Issues with EVM and a Viable Alternative," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2016.

[8] A. Minkiewicz, "Are Parametric Techniques Relevant for Agile Development Projects?," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2012.

[9] B. Hunt, J. Kilgore and J. Swartz, "Software Cost Estimating for Iterative and Incremental Development Programs," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2012.

[10] A. Minkiewicz, "CMMI and Agile Development: A Binary Choice?," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2017.

[11] K. Davis, E. Ashwood, A. Smith and W. Lane, "Case Study: Agile Execution Environment Analysis (AE2A) in AF DCGS," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2019.

[12] D. Beckett, "Beyond the Hype: Evaluating and Measuring Agile Development," in *Practical Software & Systems Measurement Users' Group Workshop*, 2011.

[13] H. Nayhouse and C. Donadi, "The Challenge of Agile Estimating," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2011.

[14] D. French and C. Dekkers, "Using Function Points to Manage Agile Product Backlog: Fact vs Fiction," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2018.

[15] T. Coonce and G. Alleman, "How Should We Estimate Agile Software Development Projects and What Data Do We Need?," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2017.

[16] B. Hunt, "Estimating Agile Software Development," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2015.

[17] J. Manring, "Maturing the Economic Aspects of Agile Development in the Federal Government," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2016.

[18] K. Rice, "Keynote Address (Kevin Rice)," in *Joint IT and Software Cost Forum*, 2020.

[19] L. Walker, B. Anderson and M. Melancon, "Grandma's Secret Hotdish Recipe for Software Planning: SAFe and Analytics," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2021.

[20] K. Richey, "Agile Development and GAO Cost Estimating Best Practices," in *Joint IT and Software Cost Forum*, 2017.
[21] D. French, "Agile and Function Points: A Winning Combination," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2016.
[22] B. Hunt, "Estimating Issues Associated with Agile Development," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2010.
[23] M. Kennedy, "Estimating and Tracking Agile Software Development Projects," in *Software and IT-CAST Symposium*, 2019.
[24] A. Letcher, "Agile Software Development Cost Risk for Information Technology Programs," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2018.
[25] M. Jones, "NRO CAAG Agile Software Development Practices," in *Software and IT-CAST Symposium*, 2016.
[26] M. Harris, "An Approach to Fixed Price Agile," in *Practical Software & Systems Measurement User's Group Workshop*, 2018.
[27] E. v. d. Vliet, "Agile Estimation Challenges When Starting a New Team for a New Product," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2019.
[28] W. Pratt, "State of Lean and Agile at DHS," in *Joint IT and Software Cost Forum*, 2020.
[29] C. Costello, "Five Agile Metrics for the Organization," in *Practical Software & Systems Measurement Users' Group Workshop*, 2019.
[30] R. Pulverenti, "Agile Release Planning and Integrated Scheduling," in *Joint IT and Software Cost Forum*, 2021.
[31] J. Myers and A. Hawkins, "Out with the Old, In with the New? Examination of Agile Dev Impacts on Traditional Software Cost Estimating Methods," in *Joint IT and Software Cost Forum*, 2020.
[32] J. Boleng, "Keynote: How do we Estimate Agile Development?," in *Software and IT-CAST Symposium*, 2018.
[33] D. Strickland, "SRDR Agile Analysis (2021 Update)," in *Agile SRDR Working Group*, 2021.
[34] W. Rosa, "Early Effort and Schedule Models for Agile Projects in the US Department of Defense (DoD)," in *NASA Cost & Schedule Symposium*, 2021.
[35] D. French, "Agile: You're Doing it Wrong (or how to know you're doing it right)," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2017.
[36] T. Cagley, "Agile Estimation using Functional Metrics," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2015.
[37] K. McKeel and S. Dennis, "Understanding Federal Sector Agile Productivity: A Benchmark Study," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2019.
[38] B. Smallwood, "Agile Software Cost Factors Case Study," in *Software and IT-CAST Symposium*, 2018.
[39] K. Barbre, "What's the Big Deal? Is Agile Software Development Really Different in the DoD Acquisition Environment?," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2019.
[40] J. T. Farlik, "Project Success in Agile Development Software Projects," *Dissertation,* 2016.
[41] M. T. Sletholt, J. Hannay, D. Pfahl, H. C. Benestad and H. P. Langtangen, "A Literature Review of Agile Practices and Their Effects in Scientific Software Development," in *Proceedings of the 4th International Workshop on Software Engineering for Computational Science and Engineering*, 2011.
[42] P. Sfetsos and I. Stamelos, "Empirical studies on Quality in Agile Practices: A Systematic Literature Review," in *2010 Seventh International Conference on the Qualty of Information and Communications Technology*, 2010.
[43] R. Sandsto and C. Reme-Ness, "Agile Practices and Impacts on Project Success," *Journal of Engineering, Project, and Production Management,* vol. 11, no. 3, pp. 255-262, 2021.
[44] M. Moloto, A. Harmse and Z. Tranos, "Impact of Agile Methodology Use on Project Success in Organizations - A Systematic Literature Review," in *Proceedings of the Computational Methods in Systems and Software*, 2020.

[45] A. S. Freire, A. Meireles, G. Guimaraes, M. B. Perkusich, R. M. da Silva, K. C. Gorgonio, A. Perkusich and H. O. Almeida, "Investigating Gaps on Agile Improvement Solutions and Their Successful Adoption in Industry Projects - A Systematic Literature Review," in *The 30th International Conference on Software Engineering and Knowledge Engineering*, 2018.

[46] B. Grissom, C. Robbins and L. Patashnick, "Automation in Practice: assessment of an Agile Database," in *Software and IT-CAST Symposium*, 2018.

[47] D. Venkatesh and M. Rakhra, "Agile Adoption Issues in Large Scale Organizations: A Review," in *Materials Today: Proceedings*, 2020.

[48] M. Faisal Abrar, M. Sohail, S. Ali, M. Faran Majeed, I. Ali Shah, N. Rashid and N. Ullah, "De-motivators for the Adoption of Agile Methodologies for Large-Scale Software Development Teams: An SLR from Management Perspective," *Journal of Sfotware: Evolution and Process,* vol. 32, no. 12, p. e2268, 2020.

[49] A. Fogarty, A. Edgeworth, O. Smith, M. Dowling, M. Yilmaz, S. T. MacMahon and P. Clarke, "Agile Software Development--Do We Really Calculate the Costs? A Multivocal Literature Review," in *European Conference on Software Process Improvement*, 2020.

[50] J. Lopez-Martinez, R. a. H. C. Juarez-Ramirez, S. Jimenez and C. Guerra-Garcia, "Problems in the Adoption of Agile-Scrum Methodologies: A Systematic Literature Review," in *2016 4th International Conference in Software Engineering Research and Innovation (COINSOFT)*, 2016.

[51] S. Tarwani and A. Chug, "Agile Methodologies in Software Maintenance: A Systematic Review," *Informatica,* vol. 40, no. 4, 2016.

[52] K. Dikert, M. Paasivaara and C. Lassenius, "Challenges and Success Factors for Large-Scale Agile Transformations: A Systematic Literature Review," *Journal of Systems and Software,* vol. 119, pp. 87-108, 2016.

[53] W. R. Fitriani, P. Rahayu and D. I. Sensuse, "Challenges in Agile Software Development: A Systematic Literature Review," in *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2016.

[54] T. Raharjo and B. Purwandari, "Agile Project Management Challenges and Mapping Solutions: A Systematic Literature Review," in *Proceedings of the 3rd International Conference on Software Engineering and Information Management*, 2020.

[55] E. v. d. Vliet, "Scaled Agile Deliveries: Do We Still Need Estimates?," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2018.

[56] J. Goucher and J. Brown, "Alternatives to SLOC in the Modern Environment," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2016.

[57] B. Smallwood, "New Approach for Sizing and Estimating Agile Software," in *Software and IT-CAST Symposium*, 2016.

[58] A. Hira, "Calibrating COCOMO(R) II for Functional Size Metrics [Doctoral dissertation]," University of Southern California (USC), 2020.

[59] K. Mann, R. Hoang, C. Dekkers and C. Lucas, "Let's SiSE Agile Program Requirements!," in *Joint IT and Software Cost Forum*, 2021.

[60] T. Michelson, "Air Force: Standing up Agile Under a Waterfall," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2021.

[61] A. Smith and B. Opaska, "Cyber Mission Platform (CMP) Program: Analyzing the Full Suite of Agile Metrics," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2021.

[62] E. v. d. Vliet and J. Bloo, "Forecasting of Agile Deliveries Based on Intelligent Data Analysis," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2021.

[63] K. Mann and R. Hoang, "But wait, there's more! Using SFPA for your Cost, Schedule & Performance Needs," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2020.

[64] C. Pullen and M. Stahr, "How Agile is Agile: Characterizing the Agile Process for Gov Acquisitions," in *Joint IT and Software Cost Forum*, 2020.

[65] I. Brown, "Agile Estimation - What's the Point?," in *Joint IT and Software Cost Forum*, 2020.

[66] K. McKeel, "The Evolving Role of the Software Cost Estimator," in *Joint IT and Software Cost Forum*, 2020.

[67] C. Hammond and L. Franks, "Why and How to Use COSMIC FP Effectively on Agile Development Contracts," in *Software and IT-CAST Symposium*, 2019.

[68] A. Minkiewicz, "Agile Estimation - What are we Really Looking For?," in *NASA Cost & Schedule Symposium*, 2019.

[69] R. Thapar and C. Dekkers, "Pitfalls to Avoid in Agile Fsm Productivity Measurements," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2019.

[70] D. Seaver, "Agile to DevOPS and its Impact on Estimation and Measurement," in *Joint IT and Software Cost Forum*, 2018.

[71] J. Sautter and S. Nichols, "Agile Estimation," in *Software and IT-CAST Symposium*, 2018.

[72] H. v. Heeringen, "The Concepts of Size and Productivity in an Agile World," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2018.

[73] K. McKeel, "Accurate Agile Estimation," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2018.

[74] J. Dean, "Estimation Bias and Mitigation with Agile Estimate Guidance," in *Practical Software & Systems Measurement Users' Group Workshop*, 2017.

[75] G. Alleman and T. Coonce, "Estimating and Reporting Agile Projects using the SRDR and Earned Value Management," in *Practical Software & Systems Measurement Users' Group Workshop*, 2017.

[76] T. Coonce and G. Alleman, "How Should We Estimate Agile Projects and Measure Progress to Plan," in *Software and IT-CAST Symposium*, 2017.

[77] A. Abran, "Software Projects Estimation & Control: The Versatility and Contributions of COSMIC Function Points," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2017.

[78] W. Rosa and C. Wallshein, "Software Effort Estimation Models for Contract Cost Proposal Evaluation," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2017.

[79] A. Abran, "Should You Care about How Good an Estimation Process?," in *International Cost Estimating and Analysis Association 2016 International Training Symposium*, 2016.

[80] B. Smallwood and O. Mahmoud, "Estimating Software Development Cost for Agile Projects," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2013.

[81] B. Smallwood, "Software Schedule & Resource Planning with Agile Delphi Estimation," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2017.

[82] A. Hira, K. Mann, P. Cymerman and I. Brown, "Get to the Point. What's the Deal with Different Function Points Methodologies?," in *International Cost Estimating and Analysis Association Professional Development and Training Workshop*, 2021.

[83] R. Yeman, "Keynote Address (Robin Yeman)," in *Joint IT and Software Cost Forum*, 2020.

[84] D. Nguyen-Cong and D. Tran-Cao, "A Review of Effort Estimation Studies in Agile, Iterative and Incremental Software Development," in *The 2013 International Conference on Computing & Communication Technologies - Research, Innovation, and Vision for Future (RIVF)*, 2013.

[85] M. Usman, E. Mendes, F. Weidt and R. Britto, "Effort Estimation in Agile Software Development: A Systematic Literature Review," in *The 10th International Conference on Predictive Models and Data Analytics in Software Engineering (PROMISE)*, Torino, Italy, 2014.

[86] T. Schweighofer, A. Kline, L. Pavlic and M. Hericko, "How is Effort Estimated in Agile Software Development Projects?," in *Fifth Workshop on Software Quality Analysis, Monitoring Improvement, and Applications*, 2016.

[87] S. Bilgaiyan, S. Sagnika, S. Mishra and M. Das, "A Systematic Review on Software Cost Estimation in Agile Software Development," *Jounral of Engineering Science and Technology Review,* vol. 10, no. 4, pp. 51-64, 2017.

[88] E. D. Canedo, D. P. Aranha, M. D. O. Cardoso, R. P. Da Costa and L. L. Leite, "Methods for Estimating Agile Software Projects: Systematic Literature Review," in *Proceedings of the International Conference on Software Engineering and Knowledge Engineering (SEKE)*, 2018.

[89] A. R. Altaleb and A. M. Gravell, "Effort Estimation across Mobile App Platforms using Agile Processes: A Systematic Literature Review," *Journal of Software,* vol. 13, no. 4, pp. 242-259, 2018.
[90] J. Goljan, Developing a Cost of Delay (CoD) Framework for the DoD & Analyzing the Current State of Air Force Agile Cost Estimation, vol. 4946, Air Force Institute of Technology Theses and Dissertations, 2021.
[91] M. Fernandez-Diego, E. R. Mendez, F. Gonzalez-Ladron-de-Guevara, S. Abrahao and E. Insfran, "An Update on Effort Estimation in Agile Software Development: A Systematic Literature Review," *IEEE Open Access Journal,* vol. 8, pp. 166768-166800, 2020.
[92] T. Hacaloglu and O. Demirors, "Challenges of Using Software Size in Agile Software Development: A Systematic Literature Review," in *The Joint Conference of the 27th International Workshop on Software Measurement (IWSM) and the 11th International Conference on Software PRocess and Product Measurement (MENSURA)*, 2018.
[93] B. Alsaadi and K. Saeedi, "Data-Driven Effort Estimation Techniques of Agile User Stories: A Systematic Literature Review," *Artificial Intelligence Review,* pp. 1-32, 2022.

Anandi Hira *leads research to advance the state of the art in software cost estimation and supports programs with the resulting estimation and measurement processes in her current role at Carnegie Mellon University's (CMU's) Software Engineering Institute (SEI). Her most recent research objective is to develop a new, innovative, capability-based software cost estimation model. Anandi received her PhD in software cost estimation under Dr. Barry Boehm at the University of Southern California (USC), where she collected software development data and calibrated the COCOMO® II software cost estimation model for functional size metrics.*

Benjamin Kwok *is a chief scientist at Tecolote Research and has supported cost research efforts at Space Systems Command (SSC) for over 20 years. He is mostly known for being "the USCM guy" and is a regular speaker at the SSC CIPTs. He graduated from UCSD with a BS in electrical engineering and worked at Boeing for a few years prior to Tecolote. Lastly, he is passionate about data visualization and product design.*

Investigating Shifts in Engineering Manufacturing Development (EMD) Factors for Department of Defense Assets through Decadal Analysis

Capt. Michael Smith
Brandon M. Lucas
Jonathan D. Ritschel
Edward D. White
Shawn M. Valentine

Analysts and estimators often use cost factors to estimate future costs throughout several phases of a weapon system's development. Factors can be used as a primary methodology early in development when few programmatic or technical details are known, or as a secondary crosscheck throughout later development to assess other estimating methods. This research focuses on factor development for the Engineering and Manufacturing Development phase and uniquely investigates Level 2 Work Breakdown Structure elements to assess any changes or trends over time. Using standardized cost reports (DD 1921) as a basis, we analyzed seven decades and eight WBS elements by commodity type, contract and contractor type, and service branch. After completing several statistical tests, we determined which factors were stable, increasing, decreasing, or unpredictable over time. Our research empirically substantiates the need to evaluate data points (i.e. programs) carefully prior to inclusion for factor development rather than relying on simpler rules of thumb, such as use the 'latest data' or include 'all available data.'

Introduction

> *"It is a capital mistake to theorize before you have all of the evidence. It biases the judgement."*
> *– Sherlock Holmes, A Study in Scarlet (Doyle, 1930)*

For the cost estimator, perhaps one could replace 'theorize' and 'evidence' with 'estimate' and 'data', and have an equally appropriate statement. Holmes' second line provides the reason and importance of the statement; indeed, perhaps some well-intentioned rules of thumb (e.g., 'use the latest data' or 'use all of the data') about what data to include can bias estimates as well. Our research provides an empirical analysis of data commonly used to develop a factor estimate and detects whether a different approach may be recommended.

Many methods exist for a cost analyst to create accurate and robust cost estimates. Four standard methods are analogy, parametric, engineering build-up, and expert opinion. As part of the second methodology, parametric, factors can often provide a relatively quick and trusted estimate for weapon systems. Most commonly, factors provide a relationship between the prime mission product (PMP) and other elements within the system, which allows estimating at differing levels of the system. For this analysis, we developed parametric factors for the Level II Common Element in the standard Work Breakdown Structure (WBS).

Cost organizations sometimes develop standard factors to use as primary or secondary estimating methodologies when completing estimates. These factors often require regular updates to the underlying data, which implies '*time period'* may play a role in factor accuracy. Our research uses decadal analysis to identify possible trends within categories of data and the most applicable data based on similarity. Understanding the time-driven aspect of these factors may help analysts create more defendable and accurate estimates.

Literature Review

Much literature exists about acquisitions, cost estimating, methodologies, and parametric factors. While not exhaustive, some of these studies certainly bear mentioning. At a high level, one impetus for defendable acquisitions programs derives from the DoD Directive 5000.01 "Defense Acquisitions System." This directive outlines the need for a disciplined approach to acquisitions to deliver "products and services that satisfy user needs … at a fair and reasonable price" (Department of Defense, 2020). Congress relies on the Government Accountability Office (GAO) to encourage these objectives by establishing the guidelines and best practices for cost estimates. The GAO publishes the Cost Estimating and Assessment Guide as a reference from which the DoD acquisitions community can obtain procedures to create trusted and verifiable estimates (Government Accountability Office, 2020). The DoD and each military service have several regulations and instructions governing the topic as well. However, it is up to the cost estimator to tailor an estimate, in a timely manner, utilizing the best data available. Just as every program is part new and part heritage, estimates must be unique to the program and built from the knowledge of previous programs.

One outcome of a cost estimate is the point estimate. The point estimate is the culmination of data results in a reasonable estimate of a program or asset's costs. The point estimate is the best prediction of future costs. An analyst establishes a point estimate by collecting data within each element of the program. Methodologies are the process of collecting and applying this data to a new program. A cost estimator may choose a combination of the following methodologies to construct an accurate point estimate: Expert Opinion, Engineering Build-Up, Analogy, and Parametric. Which methodologies are appropriate depends on data availability, which stage in the acquisitions process the program is in, and how much data the estimator can collect in the time available (Department of the Air Force, 2007).

The parametric method relies on historical costs from comparable systems and utilizes the statistical relationship between many systems to determine the historical relationship between elements of the systems. The analyst develops a parametric estimate by collecting data on several similar programs and analyzing the cost drivers to determine if a statistically significant relationship exists between them. This method assumes the same relationship that drove costs in the past will continue to drive costs in the future (Government Accountability Office, 2020).

Since the parametric methodology is based on statistically verified factors, it remains valid when the system characteristics change. However, the system characteristic must remain within the relevant range of the dataset. The parametric method may prove inadequate when a new system does not significantly match the program or parameters of the historical program(s), when not enough data exists to create a factor, or when complexity hinders understanding of the baseline relationships (Government Accountability Office, 2020).

More specifically, cost factors can be derived via parametric methods. Cost factors can be used to estimate a wide range of system costs and several prior investigations demonstrate their significance. Blair (1998) developed and published factors from avionics programs at the Engineering and Manufacturing Development (EMD) stage to estimate future programs. Wren expanded the Blair factors using the data available in 1998. By averaging data from the Blair study and the more recent programs, he developed composite factors for each support element (Wren, 1998). Further updates came in 2015 when Otte expanded the research from just Air Force aeronautical programs. His factors provided updates to existing factors and created factors for new WBS elements such as System Test & Evaluation and common support equipment, among others. It also included data not only from the Air Force but also from the Navy, Army, and Foreign Military Sales (FMS) systems (Otte, 2015). Over the years, many departments have maintained cost factor handbooks that track common factors within an organization. They include the Marine Corps Cost Factors Manual, the Army Cost Analysis Handbook, the Air Force Cost and Planning Factors, and the Historical Air Force Construction Cost Handbook. Analysts utilize these handbooks within and between branches, and often, departments will develop internal Cost Factor handbooks for specific program types (Mislick & Nussbaum, 2015). In 2019, Markman et al. generated 443 factors from 102 development phase programs in categories of commodity, development type, contract type, and service branch (Markman et al., 2021). Edwards expanded this effort by extending factors into the production phase and created 3,462 unique factors from 145 programs (Edwards et al., 2021).

Typical parametric factors are the averages of many factors across a vast time period. For example, if calculating the factors for a new bomber program, an analyst may draw from data on the B-52 Stratofortress and the B-1 Lancer, among others. There is a several decade gap between the developments of those aircraft. It may not be prudent to assume the relationships between their cost elements have not diverged, as many aspects of DoD acquisitions have evolved and developed over time. Even more recent reforms such as the Nunn-McCurdy Act of 1982, the Packard Commission of 1986, the Defense Acquisition Workforce Improvement Act (DAWIA) of 1990, the Federal Acquisitions Streamlining Act (FASA) of 1994, and the Weapon Systems Acquisition Reform Act (WSARA) of 2009 have led to changes not only in reporting, but also in the development of acquisitions programs. In addition, revisions to acquisition handbooks and standards, such as Military Standard Work Breakdown Structure for Defense Materiel Items (MIL-STD-881F), may create differences in the definitions of certain WBS elements. These changes ultimately can affect how managers develop and track systems, which could, in turn, effect the calculation of standard factors.

Database

Cost Data Summary Reports (CDSR), officially the DD Form 1921 or simply 1921s, contain the necessary cost data to develop and analyze factors. Contractors must submit these reports on contracts valued at $50 million or more on Acquisition Category I programs (DoD, 2007). DoD and contractors follow the guidance of MIL-STD 881F which provides a standardized reporting structure at Level 2 of the WBS.

The Office of Cost Assessment and Program Evaluation (CAPE) developed the Cost Assessment Data Enterprise (CADE) which stores a wide range of official cost data, including 1921s. The Air Force Life Cycle Management Center cost staff (AFLCMC/FZC) collects and summarizes these reports in an available database. Our research used a dataset that consists of 1921s spanning from 1951 until 2019, representing an extensive range of programs within the Engineering, Manufacturing, and Development (EMD) life-cycle phase.

Category	**Number**	**Remaining**
Development Programs in Database		620
Prototype/Experimental Programs	30	590
Unavailable Milestone B Date	11	579
Non-Final or Late Interim Data	168	411
No WBS Cost Values	2	409
Ground Vehicle	1	408
Final Dataset for Analysis		408

Table 1: Data Exclusion

The AFLCMC/FZC cost database contains 620 1921s in the EMD life-cycle, though we only included 408 1921s that fit the criteria for this research. We list the programs associated with these 408 1921s in Appendix A. The final dataset was selected following previous factor research criteria (e.g., Markman et al, 2021). Table 1 details the exclusion criteria and the number of programs utilized for this research.

While we excluded initial 1921 data, we included some interim 1921s based on input from AFLCMC/FZC. Any interim 1921s included in this dataset were practically complete from a cost perspective, but the 1921 does not administratively list them as such. We excluded prototype and experimental programs as their costs may not reflect traditional programs. There were 11 programs where Milestone B dates could not be determined and classified into decades. Finally, we excluded systems within CADE which had no EMD data or were in a non-readable format.

We classified the data into four categories: commodity, contract type, contractor type, and service. Within these categories, subcategories form the basis for our comparisons. For example, the category of contract type contains the subcategories Cost Plus and Fixed. Table 2 lists the subcategories and the number of 1921s associated with each. We then categorized the data by decade, which resulted in multiple factors for most programs. Categories and decades with fewer than five data points did not contain enough data to test. Table 2 gives an overview of the data categorizations for this research.

Table 3 shows the number of data points by decade and Level II WBS element. Amounts shown in very light gray were not analyzed due to lack of data points (i.e., <5). We will not include these decades in the Kruskal-Wallis or Wilcoxon tests, but we have included them for reference.

AFLCMC/FZC collected the 1921 data and extracted the relevant information into a central database, normalizing the data to fiscal year 2021 dollars using Office of the Secretary of Defense (OSD) inflation indexes and each report's "report as of" date. For decadal analysis, we have further categorized the data points by Milestone B date and rounded them down to the decade. For example, the 1990s decade includes all data with a Milestone B year from 1990 through 1999. We selected Milestone B as this is when a program leaves the Technology Maturation & Risk Reduction phase and enters the Engineering & Manufacturing Development phase.

WBS

The Work Breakdown Structure (WBS) is the framework for detailing the system requirements of a program. Organized in an hierarchical structure where each level specifies an aspect of the level above it, its purpose is to describe a system in enough detail to understand and manage the system. Creating standard WBS formats throughout DoD acquisitions ensures consistency between programs.

Category	**Total**	**% of Data**
1921s	408	100.00%
Commodity Type		
Aircraft	142	34.80%
Electronic/Automated Software	96	23.50%
Engine	14	3.40%
Missile	10	2.50%
Rotary Wing	87	21.30%
Space	24	5.90%
UAV	13	3.20%
Contract Type		
Cost Plus	212	52.00%
Fixed	110	27.00%
Contractor Type		
Prime	247	60.50%
Subcontractor	139	34.10%
Service		
Army	68	16.70%
Navy (including Marine Corps)	158	38.70%
Air Force (including Space Force)	150	36.80%
Joint	10	2.50%

Table 2: Data Categories

The design of the common WBS expresses which elements of a system are essential to understanding the cost and schedule of the program. The first level (Level 1) is the entire system or project. Every other element will eventually connect to the Level 1 WBS element, and so too will all costs. The next level (Level 2) consists of the major elements of the system. The Prime Mission Equipment (PME) is one of the Level 2 elements and consists of the direct deliverable of the system, such as the aircraft or software itself. Systems Engineering/Program Management (SE/PM) is the engineering, technical control, and business management of the system. System Test

Decade	**1950**	**1960**	**1970**	**1980**	**1990**	**2000**	**2010**	**Total**
WBS								
SEPM	1	5	24	46	51	220	38	385
ST&E	8	6	22	46	48	174	28	332
Training	6	4	13	29	30	90	8	180
Data	6	5	23	44	35	129	21	263
Site Activation			2	15	6	31	4	58
Support Equipment	7	4	12	29	27	61	10	150
Spares	7	3	2	11	11	36	6	76
G&A		8	21	35	46	216	39	365
Grand Total	35	35	119	255	254	957	154	1809

Table 3: Dataset Characteristics by Decade

and Evaluation (ST&E) consists of the design and production of models, prototypes, and hardware necessary to validate the system during the development stage. Training includes all deliverable training services, devices, equipment, and parts used to instruct personnel on the use and maintenance of the system. The Data element includes the production, acquisition, transformation, and storage of data used within the program. Site Activation includes real estate, construction, utilities, and equipment needed to house, service, and launch the PMP. Support Equipment includes Common Support Equipment and Peculiar Support Equipment. Common Support Equipment (CSE) is the design, development, and production of equipment necessary to support and maintain the system when not directly engaged in its mission. Peculiar Support Equipment (PSE) is the design, development, and production of equipment necessary to support and maintain the system when it is not directly engaged in its mission but is not CSE efforts. We have combined PSE and CSE into Support Equipment for our analysis as their definitions may overlap from program to program. Spares consist of spare components, assemblies, and subassemblies for the initial replacement of the PMP (Department of Defense, 2018). Figure 1 depicts the general structure of a DoD WBS.

Utilizing a standardized WBS makes estimation more consistent across programs and departments. For our purposes, it enables comparisons that make factor development and analysis possible. These elements are typical of most major acquisitions programs and will be of particular interest in our research.

Methods

Factor Calculation

The parametric WBS element factors used in this analysis are the mean (average) of the analogy factors for all programs contained within each WBS element, subcategory, and decade. The analogy factors are the ratio, expressed as a percentage, of the WBS Level II elements to the Prime Mission Equipment (PME) values. The PME values do not include contractor fees, miscellaneous expenses (general and administrative (G&A), undistributed budget, management reserve, or facilities capital cost of money (FCCM)). Table 4 depicts an example of an analogous factor calculation created by dividing System Engineering/Program Management (SEPM) by the program's PME value.

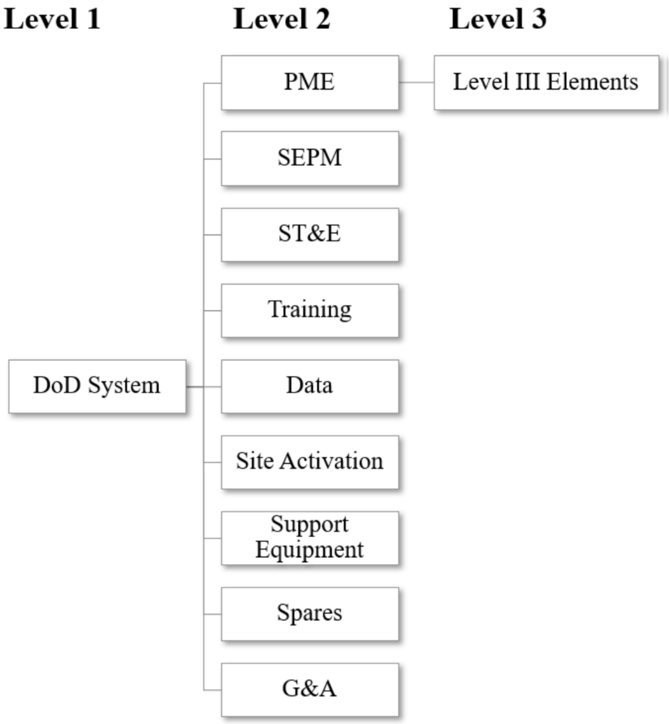


Figure 1: Example of WBS structure

We have categorized these Level II WBS factors and averaged them to develop parametric factors in each decade. Table 5 gives an example of how four programs in the same WBS element and decade calculate a single parametric factor.

Prime Mission Equipment	$ 718.3K
System Engineering/Program Management (SEPM)	$ 120.1K
Analogy Factor = 120.1/718.3 = 0.167 or 16.7%	

Table 4: Example of Analogy Factor Calculation

	PME	**SEPM**	**Ratio**
Program 1	$400K	$60K	0.150
Program 2	$280K	$20K	0.071
Program 3	$600K	$220K	0.367
Program 4	$180K	$52K	0.289
Total	$1,460K	$352K	0.877
Parametric Factor = 0.877 / 4 = 0.219 or 21.9%			

Table 5: Example of Parametric Factor Calculation

Statistical Tests

We utilize several statistical analyses to perform the hypothesis tests. These include descriptive statistics and the Kruskal-Wallis test. The Wilcoxon Rank Sum test serves as a multiple comparison test. We evaluated normality with Descriptive Statistics as any WBS categories found to be non-normally distributed needed non-parametric testing. This non-parametric testing indicates how each decade within a subcategory related to each other. A Kruskal-Wallis test this by comparing the distributions of the responses by decade (Kruskal & Wallis, 1952). The final test was the Wilcoxon test between each decade and all data points not within the decade to identify which decades were statistically different from the overall subcategory. This research compares decadal factors to traditionally calculated factors and illustrates any difference between the decade and the remaining data comprising a traditional factor.

Descriptive analysis determines if the data came from normal distributions. Normality is a necessary condition in parametric tests, as parametric tests assume the population from which we draw the samples is normally distributed. Data that fails the normality test must use non-parametric tests. Visually inspecting the data distribution and comparing the mean, median, and standard deviation determined that none of the data categories appeared normally distributed. To avoid the violation of normality in our testing, we used non-parametric tests for the remainder of our analysis.

We compare each category using the hypothesis test shown in Equation 1, where x represents the different decades in each subcategory for each comparison and y represents a parametric factor for the entire subcategory for comparison. For example, when comparing the decades for aircraft within SEPM, x is defined as the 1970s, and y would be the aircraft SEPM overall factor. Failure to reject the null hypothesis, $H0$, signifies that there is no difference between the medians of the WBS element and each decade. If we reject $H0$, then a difference does exist.

$$H0\text{: } \Delta x = \Delta y$$

$$Ha\text{: } \Delta x \neq \Delta y$$

Equation 1

Analysts use Kruskal-Wallis tests to determine whether several datasets come from the same distributions and have the same median values. We use it to determine if the data from one decade matches the distribution of the other decades' data.

We use the Wilcoxon Rank Sum test when comparing only two datasets. It is similar to a Student's t-test but without the assumption of normality. A Wilcoxon test compares the locations of the data points of two samples to determine if they are from the same distribution. To test each decade, we created a dummy variable to categorize the decade in question in one group and every other decade in the other. Given distributions of the same shape, the Null Hypothesis of a Wilcoxon test is that the medians of the two samples are equal and rejecting this Null Hypothesis would indicate which decades are not of the same distribution as the rest of the subcategory. Analyzing medians is less prone to the impacts of outliers; and as our data is skewed, we tend to see many possible outliers. Our tests compare the medians of the distributions, and the analysis identified trends and changes in the medians between decades.

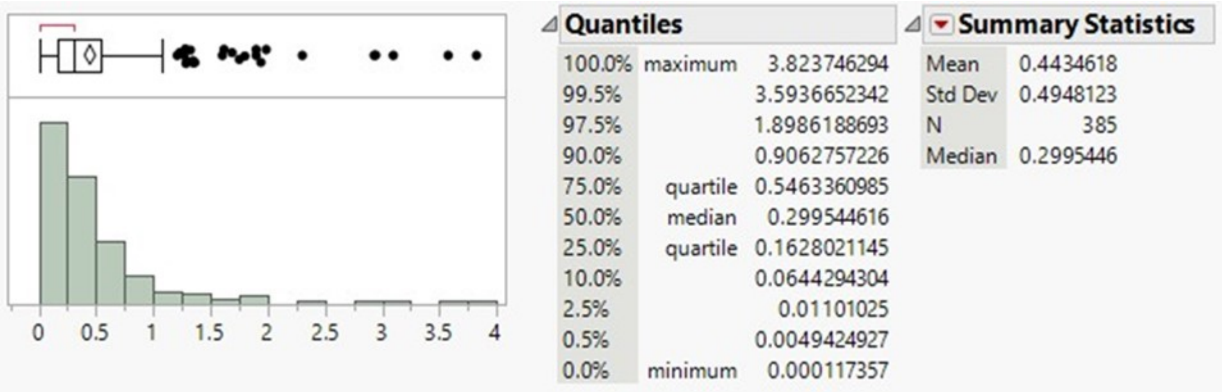
Results

"How often have I said to you that when you have eliminated the impossible, whatever remains, however improbable, must be the truth?"
— Sherlock Holmes, The Sign Of Four (1890)

We conducted the aforementioned statistical tests across all categories for each of the WBS elements to identify the distributional fit between decades, which allows us to recommend data for certain decades. However, we limit reporting and commenting to an example case and a few of the more interesting findings.

Example of Analysis using SEPM

The Systems Engineering and Program Management (SEPM) WBS element is a prominent factor in the analysis. It contains the most 1921s of the available programs with 386, representing 94.6% of the programs. SEPM factors ranged from 0.0117% to 911.4% of PME. At the high end of the range, the 911.4% factor may indicate reporting anomalies and/ or extreme issues in the upper value—as such, we trimmed it from the data set and kept the next maximum value of 382.3%. Figure 2 shows the distribution of SEPM values and provides descriptive statistics used in further analysis. The distribution is still skewed, but the standard deviation is much smaller at 49.5%, and the mean and median are closer to each other at 44.3% and 29.5%, respectively. The distribution is skewed right and far from normally distributed.



Quantiles

100.0%	maximum	3.823746294
99.5%		3.5936652342
97.5%		1.8986188693
90.0%		0.9062757226
75.0%	quartile	0.5463360985
50.0%	median	0.299544616
25.0%	quartile	0.1628021145
10.0%		0.0644294304
2.5%		0.01101025
0.5%		0.0049424927
0.0%	minimum	0.000117357

Summary Statistics

Mean	0.4434618
Std Dev	0.4948123
N	385
Median	0.2995446

Figure 2: SEPM Descriptive Statistics

Table 6 displays individual descriptive statistics for SEPM broken out by decade. As previously mentioned, excluded from the analysis are decades with less than 5 data points (e.g., 1950 in Table XX), but are included here for comparison to other decades (Howell, 2010).

SEPM Summary Table by Decade							
Decade	1950	1960	1970	1980	1990	2000	2010
Mean	0.26	0.11842	0.18893	0.34387	0.40447	0.47302	0.65357
Std Dev		0.08643	0.17072	0.29198	0.61088	0.49159	0.60133
N	1	5	24	46	51	220	38
Max	0.26	0.25175	0.65209	1.26801	3.82375	3.57635	3.09701
0.75	0.26	0.20532	0.26091	0.47241	0.44964	0.57181	0.83865
Median	0.26	0.0762	0.14311	0.27827	0.23298	0.34771	0.50689
0.25	0.26	0.05263	0.06816	0.11313	0.15362	0.19623	0.29008
Min	0.26	0.04866	0.0293	0.00597	0.00012	0.00531	0.01251

Table 6: SEPM Summary by Decade

The SEPM factor appears to be growing from 1960 through 2010. The mean, median, and quartiles have, for the most part, consistently increased. The mean has grown from 11.8% to 65.3%, and the median has grown from 7.6% to 50.7%. The mean is consistently within the third quartile, confirming the right skew of the distribution. Due to this type of skew, the median may be a better measure of the distribution as the median is less affected by outliers and highly skewed distributions.

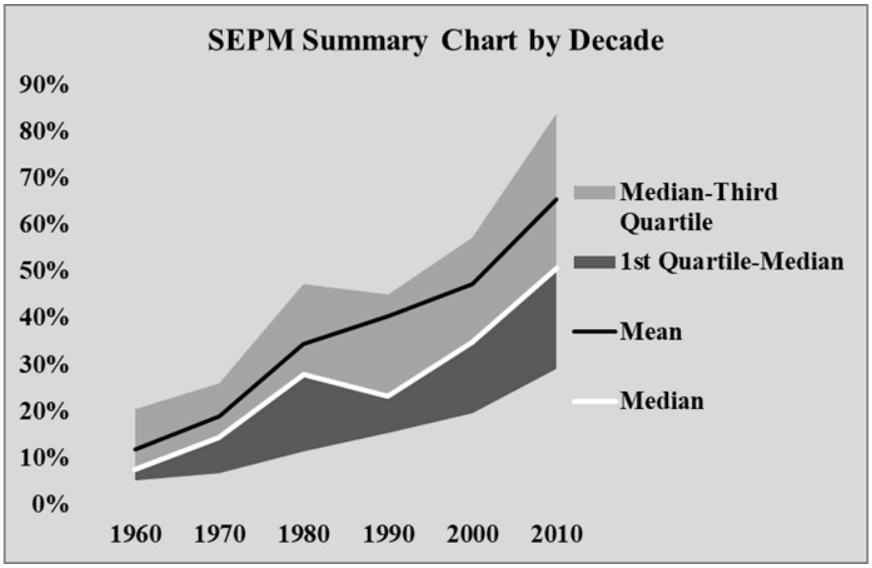


Figure 3: SEPM Summary by Decade

Figure 3 illustrates how the SEPM factor over time has changed quite dramatically, which indicates a rule of thumb such as ‘use all of the data’ is likely not the best method for a new weapon system. Additionally, note that SEPM appears to have an increasing trend over time, which indicates this cost category generally contains much more of a weapon system’s cost now than five decades earlier.

ST&E

The ST&E factors have been slightly decreasing, in contrast to the SEPM factors. There is also more variability in the ST&E factors through the decades, but overall, there has been a shift from higher to lower factors. The mean has changed from 33.3% to 21.9%, and the median has shifted from 32.0% to 10.2%. Again, the mean resides within the third quartile, except for the 1970s, where the mean and median are very close.

ST&E Summary Table by Decade

Decade	1950	1960	1970	1980	1990	2000	2010
Mean	0.33320	0.38307	0.42179	0.28501	0.17351	0.15181	0.21870
Std Dev	0.15924	0.26609	0.26592	0.27351	0.15695	0.18037	0.32311
N	8	6	22	46	48	174	28
Max	0.59372	0.83931	1.06772	1.07767	0.60513	1.05752	1.49831
0.75	0.47241	0.62385	0.60037	0.43902	0.26537	0.21052	0.32363
Median	0.32009	0.29513	0.40791	0.19222	0.12392	0.09505	0.10234
0.25	0.18537	0.16597	0.19156	0.07796	0.06399	0.03424	0.00345
Min	0.12973	0.14332	0.02533	0.00405	0.00050	0.00011	0.00000

Table 7: ST&E Summary by Decade

While not as consistent or large as the trend in SEPM, ST&E data contains a decreasing trend between the 1970 and 2000 decades. As a result, factors for ST&E tend to vary noticeably by decade, with recent decades tending to be quite a bit lower than earlier decades.

‘Categories’

Other WBS elements contain less significant or stable patterns, but often vary as well. As such, we include charts for completeness, but do not directly address each of them. These are available for review in Appendix B.

Based on the data, clear differences often exist between decades. Depending on the WBS element, cost factors may increase, decrease, or remain somewhat stable, but significant shifts and changes appear fairly routine.

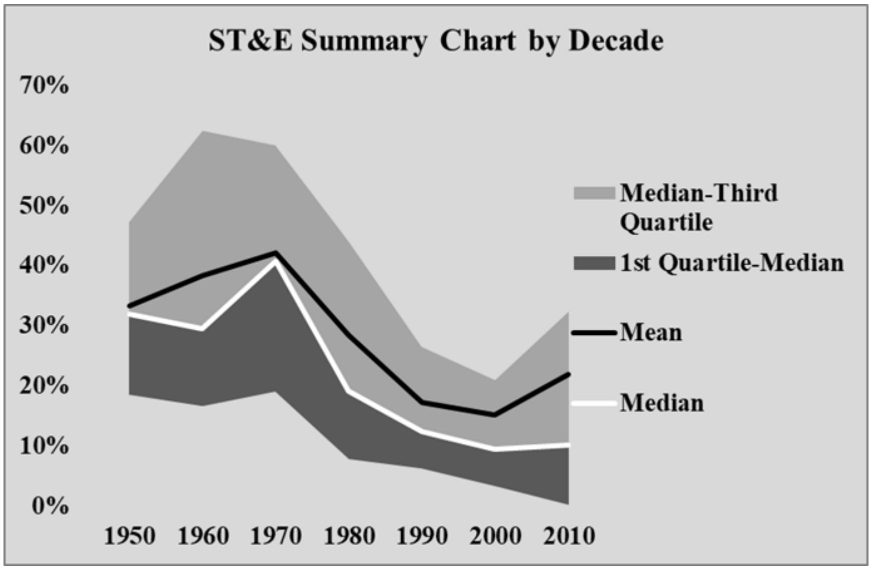


Figure 4: ST&E Summary by Decade

WBS	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Decade	1950		1960		1970		1980	
SEPM			11.8%*	7.6%*	18.9%***	14.3%***	**34.40%**	**27.80%**
ST&E	33.3%*	32.0%*	38.3%*	29.5%*	42.2%***	40.8%***	28.5%*	19.2%*
Training	**2.00%**	**2.00%**			**1.70%**	**0.40%**	**2.40%**	**0.40%**
Data	**2.10%**	**1.50%**	**2.50%**	**2.80%**	**5.80%**	**2.70%**	**4.60%**	**1.90%**
Site Activation							1.8%**	0.2%**
Support Equipment	5.40%	3.50%			16.3%*	5.9%*	**6.30%**	**3.20%**
Spares	8.9%***	10.1%***					**5.40%**	**1.50%**
G&A			10.3%***	10.0%***	22.30%	21.60%	17.0%**	17.9%**

WBS	Mean	Median	Mean	Median	Mean	Median
Decade	1990		2000		2010	
SEPM	**40.40%**	**23.30%**	**47.3%***	**34.8%***	**65.4%*****	**50.7%*****
ST&E	**17.40%**	**12.40%**	**15.2%*****	**9.5%*****	**21.90%**	**10.20%**
Training	**2.50%**	**0.30%**	**3.60%**	**0.60%**	**4.20%**	**0.30%**
Data	**2.90%**	**1.80%**	**3.10%**	**1.80%**	**2.20%**	**1.20%**
Site Activation	**4.10%**	**1.40%**	**6.3%****	**4.0%****		
Support Equipment	**2.7%***	**0.8%***	**10.30%**	**1.90%**	**1.4%****	**0.3%****
Spares	**6.30%**	**2.40%**	**6.30%**	**1.70%**	**1.50%**	**1.00%**
G&A	**25.70%**	**18.30%**	**26.10%**	**23.20%**	**35.90%**	**25.90%**

Table 8: WBS Summary Table and Significance Results

Statistical Tests & Results

Our research addresses the creation and statistical significance of factors between decades within several categories in each WBS element. We sought to examine the statistical differences between decades in Level II WBS factors for various DoD commodities, contract types, contractor types, and service branches. We found clear trends with statistically significant decadal differences in several subcategories within SEPM and ST&E. There are also several large spikes across decades.

For decades and subgroups with five or more data points, we display the mean and median. However, our analysis and discussion will focus on the medians of the distributions due to the manner of our tests and the skew of our distributions. In addition, we chose the median versus mean as a point of comparison as the median is less affected by extreme values. We exclude all subgroups with less than two decades, as we would not be able to compare them.

We indicate the results of the Wilcoxon tests on these numbers with asterisks. The p-values under 0.05 are moderately significant and have a single asterisk next to the mean and median values. P-values below 0.01 are significant and have two asterisks. P-values below 0.005 are highly significant have three asterisks. We have also marked those decades we would recommend using for future factor creation using bold format. Table 8 illustrates these significance levels. We will explore graphs of the medians by decade within some of the WBS sections to highlight changes to factor composition.

KEY	**Significance**	**Recommended**
Not Significant ($p>0.05$)	x.xx%	**x.xx%**
Moderately Significant ($0.05\geq p>0.01$)	x.xx%*	**x.xx%***
Significant ($0.01\geq p>0.005$)	x.xx%**	**x.xx%****
Highly Significant ($p\leq 0.005$)	x.xx%***	**x.xx%*****

Table 9: Wilcoxon Results Significance Levels Key

SEPM factors appear to increase from the 1960s to the 2010s, with highly significant factors in the 1970s and 2010s. Having significant factors early and late in the decades corresponds with an upward or downward trend. ST&E shows an opposite trend, decreasing in the median from a high in the 1970s to a low in the 2010s. ST&E has highly significant factors in the 1970s and 2000s. G&A factors increased from the 1960s to the 2010s but were only significant in the early decades. Table 9 illustrates the shifts in the median for these WBS elements. The relationship between an increase or decrease in the factors marked with significant decades will appear in these same WBS elements when discussing them at the subcategory level. Site Activation saw an increase between the 1980s and the 2000s, with the 2000s being significant. Support Equipment saw a decline with significant factors in the 2010s. Spares had highly significant factors in the 1960s, where the median was much higher than the other decades.

Since Training and Data had no significant decades, all the data appears to be from the same distribution. Future analysts could use all decades' data to develop factors as none of the decades contain data that differ greatly from the distribution.

For significant decades with visible trends such as SEPM, ST&E, and G&A, an analyst should consider using the most recent decades as the older decades might not reflect newer programs. In WBS elements like Spares, where one decade is significant, and the median is dissimilar from the remaining decades, we recommend excluding data from that decade and using the remaining decades to develop an estimate. Figures 5, 6, and 7 provide visualization of the median data from Table 9. Note that in some lines may not connect due to lack of data in intervening decades.

In general, our recommendations are to exclude early decades if they have significant Wilcoxon results, or later decades if they appear to represent an anomaly, such as with Spares. We recommend all decades when no decades are significant. These recommendations apply in the remaining analysis by subcategories unless we specifically outline decades in the analysis.

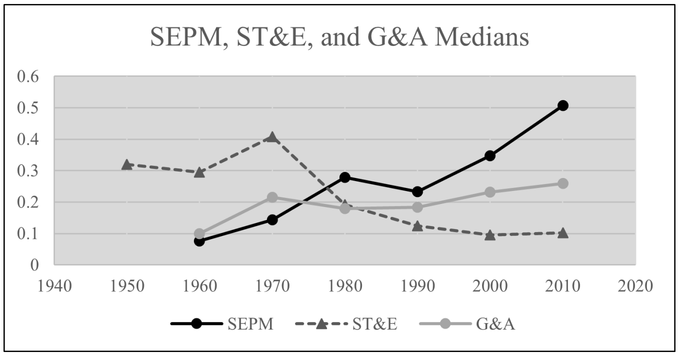


Figure 5: SEPM, ST&E, and G&A Medians by Decade

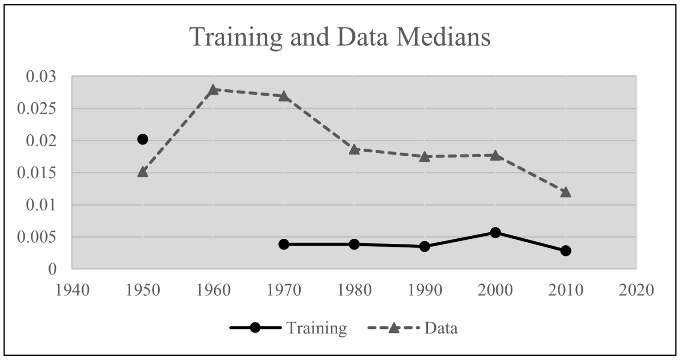


Figure 6: Training and Data Medians by Decade

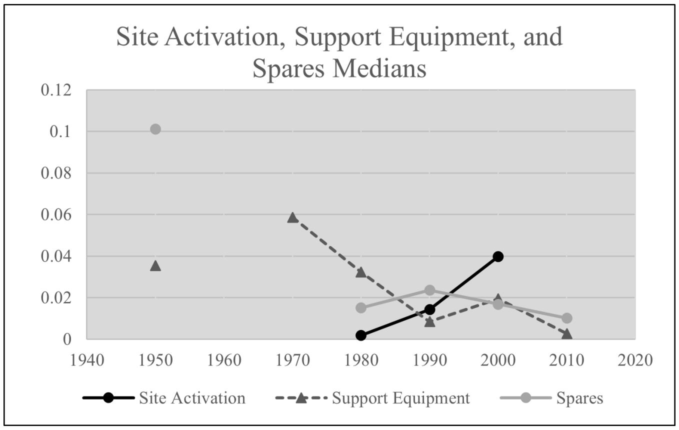


Figure 7: SA, SE, and Spares Medians by Decade

Significance of Research

We can interpret the findings of this research in many ways. First, analysts should avoid grouping the decades with significant differences with the other decades in that subcategory without justification. With several of the WBS elements, clear time trends were present in the data, and the significant results were indicative of the extreme ends of those trends. Where trends exist, using older data for a modern system may not be appropriate as the relationships those factors represented may no longer be most representative. Second, decades that have no significant differences are more likely to fully represent the subcategory in question, and a factor using all available decades may be appropriate.

We discovered an interesting relationship while analyzing these trends. The SEPM and ST&E decadal factors appear to change by decade in opposite directions. These two WBS elements are often the most significant cost drivers outside of PME. It is not apparent that these trends will continue, but it is interesting that as SEPM has become a larger cost driver, ST&E has become smaller. In such instances, analysts should likely use only the most recent or relevant programs to develop parametric factors. Not doing so could result in older or less relevant data points incorrectly influencing the factors when used for new programs.

For an estimator without the necessary time or tools to conduct a detailed analysis, we offer two tables with recommend decadal periods for data selection. Table 10 and Table 11 display an overview of the recommended decades for each WBS element and subcategory. An "X" in a decade indicates the decade within a subcategory that is most representative of future factors. Blank decades are either not recommended for use due to not being representative or were not evaluated due to lack of data points.

Decade	SEPM							ST&E							Training							Data						
	1950	1960	1970	1980	1990	2000	2010	1950	1960	1970	1980	1990	2000	2010	1950	1960	1970	1980	1990	2000	2010	1950	1960	1970	1980	1990	2000	2010
Commodity Type																												
Aircraft				X	X	X	X				X	X	X	X	X		X	X	X	X		X	X	X	X	X	X	X
Electronic/Automated Software					X	X	X					X	X	X						X	X						X	X
Rotary Wing				X	X	X						X	X						X	X						X	X	
Contract Type																												
Cost Plus				X	X	X	X				X	X	X	X			X		X	X				X	X	X	X	
Fixed						X	X				X		X	X						X							X	X
Contractor Type																												
Prime				X	X	X	X				X	X	X	X	X		X	X	X		X	X	X	X	X	X	X	X
Sub					X	X	X					X	X						X	X						X	X	
Service																												
Air Force				X	X	X	X					X	X	X	X		X	X	X	X		X		X	X	X	X	X
Navy				X	X	X	X				X	X	X	X				X	X	X				X	X	X	X	X
Army					X	X	X					X	X	X					X	X	X					X	X	X

Table 10: Recommended Decades Part 1

Decade	Site Activation							Support Equipment							Spares							G&A						
	1950	1960	1970	1980	1990	2000	2010	1950	1960	1970	1980	1990	2000	2010	1950	1960	1970	1980	1990	2000	2010	1950	1960	1970	1980	1990	2000	2010
Commodity Type																												
Aircraft				X		X					X	X	X		X			X	X	X				X	X	X		X
Electronic/Automated Software																										X	X	X
Rotary Wing													X												X	X	X	
Contract Type																												
Cost Plus					X	X						X	X						X	X				X	X	X	X	X
Fixed						X					X		X					X		X				X	X		X	X
Contractor Type																												
Prime				X	X	X						X	X	X				X	X	X						X	X	X
Sub												X	X													X	X	X
Service																												
Air Force												X	X	X					X	X				X	X	X	X	X
Navy						X					X	X	X							X	X					X	X	X
Army																										X	X	X

Table 11: Recommended Decades Part 2

Conclusion

This research utilized data from the AFLCMC/FZC database derived from CADE to develop and analyze decadal factors in eight WBS elements and across several commodities, contract types, contractor types, and service branches. The creation of robust factors requires utilizing the most extensive database available to the analyst. However, understanding changes in factor composition over time can assist in the creation of more defensible estimates in the future. The factors tested in this research highlight decadal differences between programs within categories. Analysts should consider these differences when creating parametric factors for cost estimates.

While we do not claim that some of these findings have not been observed by other estimators, the study does provide a wide-ranging historical and empirical analysis to test common rules of thumb. As we have demonstrated, some WBS element factors have increased, decreased, or had spikes, indicating that not all decades and time periods represent the overall WBS element or subcategory. Tables 10 and 11 can help estimators focus data collection and investigative efforts on the time periods most likely to benefit their specific analysis and estimate. Our research provides a framework for creating and improving parametric factors for cost estimates. Efficient and effective cost estimating relies on the most relevant and useful data. The importance of this research is informing cost estimators to take into consideration the decade, or time period in general terms, from which they calculate factors.

References

Department of the Air Force. (2007). Air Force Cost Analysis Handbook. Washington: U.S. Department of the Air Force.

Department of Defense (2007). Cost and Software Data Reporting (CSDR) Manual. DoD 5000.04–M–1. Washington: DoD.

Department of Defense (2022). Work Breakdown Structures for Defense Materiel Items. MIL-STD-881F. Washington: DoD.

Doyle, Arthur Conan, 1859-1930. (1930). The complete Sherlock Holmes. Garden City, N.Y. :Doubleday & Co.

Edwards, J. S., Ritschel, J. D., Plack, E. A., White, E. D., Koschnick, C. M., & Drylie, S.T. (2021). Improving Acquisitions in Science and Technology Programs: Creating Unique Cost Factors to Improve Resource Allocation Decisions. Acquisition Research Program.

Government Accountability Office. (2020). Cost estimating and assessment guide (Report No. GAO-20-195G). Washington, DC: U.S. Government Printing Office.

Howell, D. C. (2010). University of Vermont (7th ed.). Wadsworth, Cengage Learning, Belmont, CA.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-Criterion Variance Analysis. Journal of the American Statistical Association, 47(260), 583–621. https://doi.org/10.1080/01621459.1952.10483441

Markman, M., Ritschel, J., & White, E. (2021). Use of Factors in Development Estimates: Improving the Cost Analyst Toolkit. Defense Acquisition Research Journal, 28(96), 40–70. https://doi.org/10.22594/10.22594/dau.19-848.28.01

Mislick, G. K., & Nussbaum, D. A. (2015). Cost estimation: Methods and Tools. John Wiley & Sons.

Otte, Jim. (2015). Factor Study September 2015. Air Force Lifecycle Management Center Research Group, Wright Patterson Air Force Base, OH.

Wren, Don. (1998). Avionics Support Cost Element Factors. Aeronautical Systems Center, Wright Patterson Air Force Base, OH.

Appendix A

Aircraft

- A-10A
- A-6A
- A-6E
- A-6F
- AC-130U
- ASIP
- AV-8B
- B-1A
- B-1B
- B-2A
- B-52H
- B-58A
- C-130 AMP
- C-130J
- C-17A
- C-5A
- C-5A/B
- C-5M
- E-2C
- E-2D
- E-3 FMS
- E-3A
- E-6A
- E-7A
- E-8A
- E-8C
- EA-18G
- EA-6B
- F/A-18A
- F/A-18A/B/C
- F/A-18C/D/E/F
- F/A-18D
- F/A-18E/F
- F-101A
- F-102A
- F-104A
- F-105A
- F-106A
- F-14A
- F-14D
- F-14D, F/A18C/D
- F-15A
- F-15E
- F-16A
- F-16A/B
- F-16C/D
- F-22A
- F-35A/B/C
- F-4A
- F-4C
- F-5E
- F-5F
- KC-135A
- KC-135R
- KC-46A
- LVT MIDS
- P-8A
- RQ-4/E-10
- S-3A
- S-3B
- T-45TS
- T-46A
- VC-25A

Electronic/Automated Software

- 3DELRR
- AMDR
- AMF JTRS
- AN/TSC-154
- AN/TVQ-2
- B-1B
- CAC2S
- CANES
- CIRCM
- CNS/ATM
- Cobra Judy
- DCGS
- Distributed Battle Command System
- EA-18G
- F-15E
- F-16 Blk 30
- F-16 Blk30
- F-16 Blk40/50
- F-22A
- FAB-T
- FBCB2
- G/ATOR
- GCSS-A
- GSE
- IAMD
- ITEP
- JATAS
- JLENS
- JMPS
- JPALS
- JTRS
- JTRS GMR
- JTRS MIDS
- JTRS NED - MUOS Waveform
- LMP
- LVT MIDS
- MUOS
- N/A
- WIN-T

Engine

- A-10A
- A-4A
- A-6F
- A-7A
- B-1B
- CH-53K
- F/A-18A
- F/A-18E/F
- F-111F
- F-14A
- F-15A
- F-16A/B
- F-22A
- F-35A/B/C
- F-5E
- V-22

Missile

- AGM-129A
- AIM-9X
- ER
- JAGM
- MALD-J
- N/A

Rotary Wing

- AH-1Z & UH-1Y
- AH-64D
- AH-64E
- ARH-70A
- CH-47D
- CH-47F
- CH-53K
- CV-22
- H-1/AH-1Z
- H-1/UH-1Y
- HH-60A
- MH-60R
- MH-60S
- OH-58D
- RAH-66A
- SH-60B
- SH-60F
- UH-1N
- UH-60M
- V-22
- V-22/CV-22

Space

- AEHF
- EPS
- GPS - OCX
- GPS-IIIA
- GT - EPS
- NAVSTAR GPS - Block IIR
- NAVSTAR GPS - MUE
- NAVSTAR GPS - OCS
- NAVSTAR GPS Blk IIF
- NPOESS
- SBIRS HIGH
- TSAT

UAV

- MQ-1C
- MQ-4C
- MQ-9A
- RQ-1A
- RQ-4A/B
- RQ-5A
- RQ-7A
- RQ-8A

Appendix B

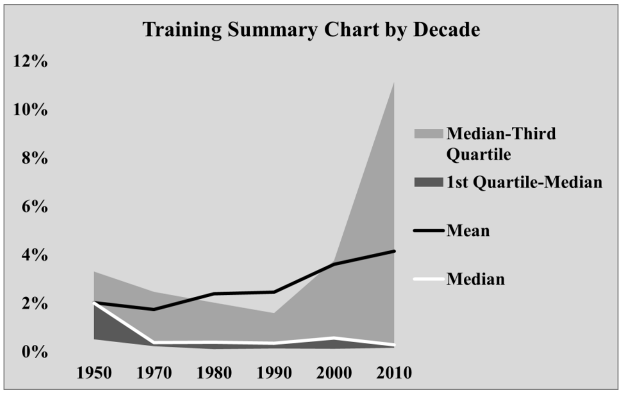


Figure 8: Training Summary by Decade

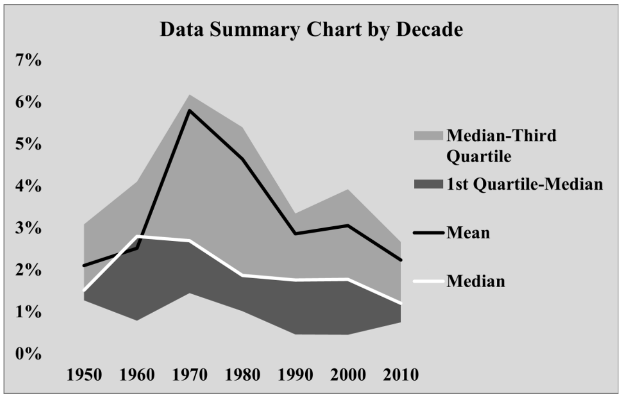


Figure 9: Data Summary by Decade

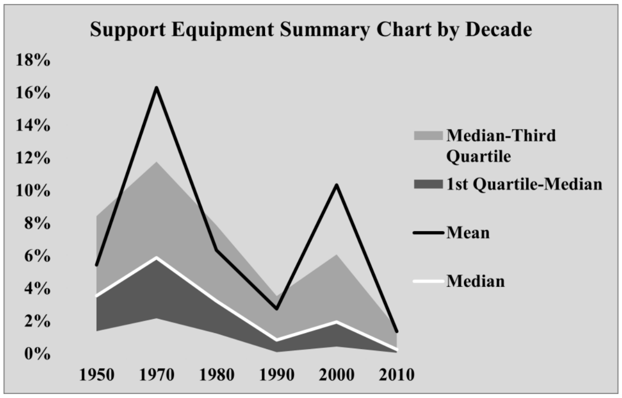


Figure 10: Support Equipment Summary by Decade

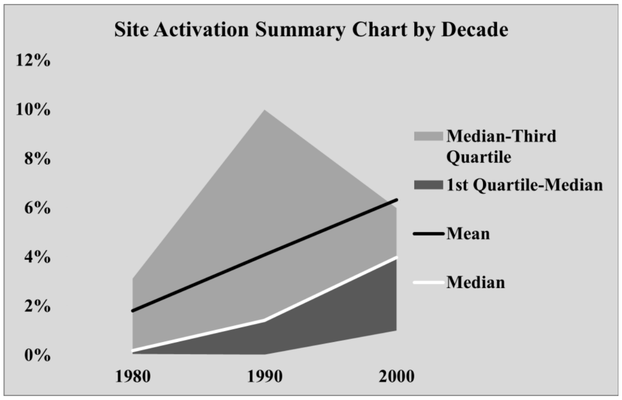


Figure 11: Site Activation Summary by Decade

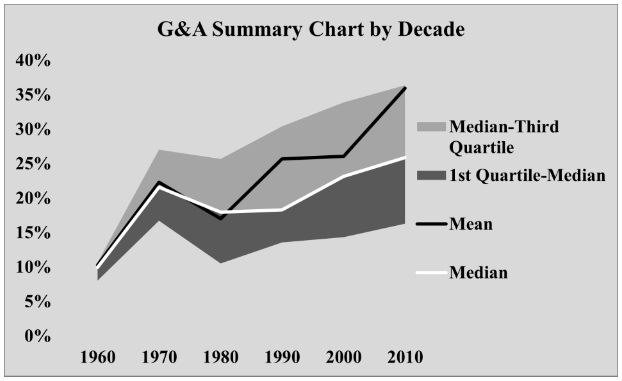


Figure 12: G&A Summary by Decade

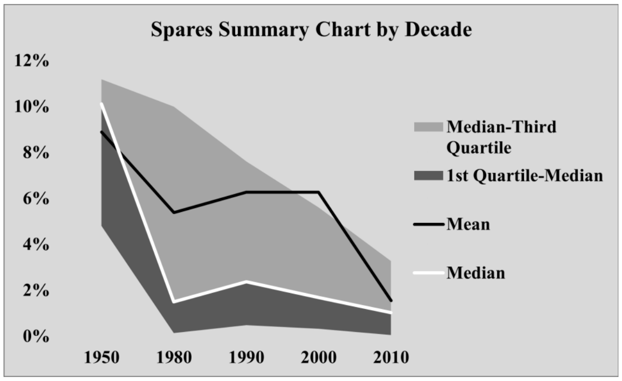


Figure 13: Spares Summary by Decade

Captain ***Michael J. Smith*** *is the Lead Cost Analyst, Missile Warning/Missile Tracking program, SNF, Los Angeles Air Force Base, California. Captain Smith enlisted in the Air Force following his graduation from Northwest University, WA. He then commissioned to Second Lieutenant through Officer Training School, Maxwell AFB, AL. He attended the Air Force Institute of Technology and received a Masters in Cost Analysis at Wright Patterson AFB, OH.*

Dr. ***Brandon M. Lucas*** *is an Assistant Professor of Systems Integration & Cost Analysis in the Department of Systems Engineering and Management at AFIT. He holds a BA in History from the University of Texas at Austin, a MA in International Relations and ME in Teacher Education from the University of Oklahoma, a MS in Cost Analysis from AFIT, and a PhD in Economics from George Mason University. Dr. Lucas' research interests include profit analysis, cost & economic analyses, and incentive structures. (E-mail address: brandon.lucas.8@us.af.mil)*

Dr. ***Jonathan D. Ritschel*** *is an Associate Professor of Cost Analysis in the Department of Systems Engineering and Management at AFIT. He received his BBA in Accountancy from the University of Notre Dame, his MS in Cost Analysis from AFIT, and his PhD in Economics from George Mason University. Dr. Ritschel's research interests include public choice, cost analysis, and economic institutional analysis. (E-mail address: jonathan.ritschel.1@us.af.mil)*

Dr. ***Edward D. White*** *is a Professor of Statistics in the Department of Mathematics and Statistics at AFIT. He received his BS in Mathematics from the University of Tampa, MAS from The Ohio State University, and PhD in Statistics from Texas A&M University. His primary research interests include statistical modeling, simulation, and data analytics. (E-mail address: edward.white@us.af.mil)*

Mr. ***Shawn M. Valentine*** *is an Operations Research Analyst at AFLCMC/FZCE at Wright Patterson AFB, Ohio. Mr. Valentine graduated from Ohio University with a Bachelor of Science (BS) in Actuarial Science and a Master of Science (MS) in Financial Economics. (Email address: shawn.valentine@us.af.mil)*

The International Cost Estimating and Analysis Association is a 501(c)(6) international non-profit organization dedicated to advancing, encouraging, promoting and enhancing the profession of cost estimating and analysis, through the use of parametrics and other data-driven techniques.

www.iceaaonline.com

Submissions:

Prior to writing or sending your manuscripts to us, please reference the JCAP submission guidelines found at

www.iceaaonline.com/publications/jcap-submission

Kindly send your submissions and/or any correspondence to:
jcapeditor@iceaaonline.org

International Cost Estimating & Analysis Association
4115 Annandale Road, Suite 306 | Annandale, VA 22003
703-642-3090 | iceaa@iceaaonline.org