Proceedings of the ASME 2021 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC/CIE 2021
August 17-20, 2021, Online, Virtual

DETC2021/71517

A MIXED-METHOD ANALYSIS OF SCHEDULE AND COST GROWTH IN DEFENSE ACQUISITION PROGRAMS

Atharva Hans*
School of Mechanical Engineering
Purdue University
West Lafayette, Indiana

Ashish M. Chaudhari
Postdoctoral Researcher
Aerospace Engineering
Texas A&M University
College Station, Texas

Ilias Bilionis
Associate Professor
School of Mechanical Engineering
Purdue University
West Lafayette, Indiana

Jitesh H. Panchal
Professor
School of Mechanical Engineering
Purdue University
West Lafayette, Indiana

ABSTRACT

Cost and schedule overruns are common in the procurement of large-scale defense acquisition programs. Current work focuses on identifying the root causes of cost growth and schedule delays in the defense acquisition programs. There is need for a mix of quantitative and qualitative analysis of cost and schedule overruns which takes into account program factor such as, technology maturity, design maturity, initial acquisition time, and program complexity. Such analysis requires an easy to access database for program-specific data about how an acquisition programs' technical and financial characteristics vary over the time. To fulfill this need, the objective of this paper is twofold: (i) to develop a database of major US defense weapons programs which includes details of the technical and financial characteristics and how they vary over time, and (ii) to test various hypotheses about the interdependence of such characteristics using the collected data. To achieve the objective, we use a mixed-method analysis on schedule and cost growth data available in the U.S. Government Accountability Office's (GAO's) defense acquisitions annual assessments during the period 2003-2017. We extracted both analytical and textual data from original reports into Excel files and further created an easy to access database accessible from a Python environment. The analysis reveals that technology immaturity is the major driver of cost and schedule growth during the early stages of the acquisition programs while technical inefficiencies drive cost overruns and schedule delays during the later stages. Further, we find that the acquisition programs with longer initial length do not necessarily have higher greater cost growth. The dataset and the results provide a useful starting point for the research community for modeling cost and schedule overruns, and for practitioners to inform their systems acquisition processes.

1 Introduction

Cost overruns plague the majority of large-scale complex systems covering transportation [1], power [2], defense [3,4] and space [5]. According to the Government Accountability Office (GAO) 77% of highway projects in the US result in cost over-

*Corresponding author. Email: hans1@purdue.edu

runs [6]. Similarly, in a study spanning 250 large transportation projects covering 20 counties, it was found that 90% of transportation projects cost 30% more on average [1]. In large infrastructure projects, including dams and nuclear reactors, cost escalation is the norm [2]. NASA's cost overrun averages 30% over the past decade [5]. 80% of the Navy lead and follow-on ships costed more than initially budgeted, 100% were behind schedule (half of them by more than 2 years), and many ships were delivered with deficiencies and incomplete work [3]. According to the GAO, the estimated increase in the acquisition cost for the Department of Defense's (DOD's) major defense acquisition (MDAPs) programs, from 1992 to 2007, was nearly 120% [7]. In addition, as of 2008, on average, MDAP's experienced 21 months delays in providing initial operational capabilities [7]. Therefore, in this paper we focus our attention on DoD's MDAPs. Insights gained from studying these DoD's MDAPs can assist in better understanding the reasons for cost and schedule overruns in large-scale complex systems.

Past studies have often attributed cost overruns and schedule delays in defense acquisition to infeasible and unrealistic requirements [8, 9, 10], excessive technical and manufacturing risk [11, 12, 13], technology integration issues [11, 12], immature technology [11, 12], concurrency in acquisition programs [14], testing inefficiency [13], funding instabilities [15, 8] and overly optimistic cost and schedule estimates [11, 12, 8, 16]. Despite identifying the major reasons for cost overruns and schedule delays in the MDAPs, the performance of the acquisition programs remain poor. There is a need for a more qualitative analysis of cost and schedule overruns which takes into account program factor such as, technology maturity, design maturity, initial acquisition time, and program complexity. This qualitative analysis requires an easy to access database for program specific data about how an acquisition programs' technical and financial characteristics vary over the time.

Therefore, the objective of this paper is twofold: *(i) to develop a database of major US defense weapons programs which includes details of the technical and financial characteristics and how they vary over time, and (ii) to test various hypotheses about the interdependence of such characteristics using the collected data.* We analyze U.S. GAO's yearly review reports of major defense acquisition programs from year 2003 to 2017 [17]. The majority of programs covered in GAO's defense acquisitions annual assessment are considered "major" by the Department of Defense (DoD). According to GAO, "a program is defined as major if its estimated research and development costs exceed \$365 million or its procurement exceeds \$2.19 billion in fiscal year 2000 dollars" [17]. Further, the technical characteristics of a program includes technology maturity, design maturity and complexity. Additionally, a program's financial characteristics encompass research and development (R&D) cost, procurement cost, total program cost, acquisition cycle time, and total procurement quantity. We extracted this quantitative data from GAO's reports, originally available in PDF format, into Excel files and further created an easy to access database accessible from a Python environment. Both the original Excel files and the database are made available through Purdue University Research Repository (PURR)[1]. Based on the literature review, we describe five hypotheses (Section 3.1) and test them on the dataset.

The results show technology immaturity as the major reason for cost overruns and schedule growth in the early part of the program. Further, the overruns experienced in the later part of the program can mostly be attributed to testing issues experienced by the program. Counter-intuitively, however, technology maturity at the development start of a program is positively correlated with cost overruns. As such, there seems to be multiple factors other than the technology immaturity that seem to drive costs significantly higher. Another observation is that the programs with higher technical complexity tend to have higher cost overruns. The dataset and the results are a useful starting point for the research community and systems engineering practitioners for modeling cost and schedule overruns and better inform the systems engineering processes.

The organization of this paper is as follows. We begin with review of the relevant work for studying cost growth and schedule delays in large-scale acquisition programs in section 2. In section 3, we introduce the dataset and explain various terminologies used by GAO. Then based on our review of literature, for the causes of cost and schedule overruns, we present a list of hypotheses that we tested in section 3.1. In section 4 we present our results where we highlight insights from the dataset. Finally, in section 5, we discuss the implication of this dataset from a perspective of qualitative modelling of the systems engineering process.

2 Related Work

We review existing studies on understanding the causes of cost overruns and schedule delays in large-scale acquisition programs. Specifically, we focus on the primary analysis methodologies used, the identified causes of cost and schedule growth, and the use of agent-based modeling for understanding the system acquisition process.

2.1 Reasons for Cost and Schedule Growth

Existing research contributes to identifying the major reasons for cost and schedule growths. Most of theses studies rely on a mix of qualitative and quantitative analysis of the selected acquisition reports (SARs) and interview data. Meier [18] focuses on analyzing the data from DoD programs acquisition reports and interviews with project executives. Their analysis identifies multiple causes. They also illustrate the sequential flow from inferred cause to systematic effects and reveal key interrelationships between different causes. Hofbauer et al. [19] investigate 92 major programs active in 2010 and 12 cancelled pro-

[1] Request access at: https://purr.purdue.edu/projects/dodoverruns

grams. Their analysis of respective SARs suggest that inaccurate cost estimates have the strongest correlation with net cost growth changes. Dwyer et al. [20] propose a methodology to perform fine-grained and longitudinal analysis of the program specific data. They propose a methodology based on design structure matrices, to model organizational structure and explain program-specific reasons of cost overruns. Riposo et al. [21] summarizes a selection of acquisition literature on potential sources of schedule and cost growth, along with the recommendations for improving schedules. Drawing on this literature, Table 1 summarizes the potential reasons into a few important categories.

Building on the insights of these studies, we attempt to organize the longitudinal (yearly) data of financial and technical characteristics for multiple defense programs. Such fine-grained data allow us to generate quantitative insights about the magnitudes of cost and schedule growths, and to analyze the causal relationships between financial and technical characteristics.

2.2 Knowledge-based Acquisition Process

The GAO assessments are the major data sources for acquisition analysis in most research. Therefore, it is important to review GAO's recommendations on best acquisition practices. At critical junctures in development, such as product development start, critical design review, and production start, the knowledge-based acquisition approach, as described by GAO, focuses on achieving high levels of knowledge in three aspects of a new product or weapon (technology, design, and production). GAO defines it's best practices for acquisition programs as follow: (i) at the start of product development, acquisition programs should have fully mature critical technologies, (ii) by the system-level critical design review, acquisition programs should have achieved design stability, and (iii) by the start of production, acquisition programs should have achieved production maturity. According to GAO's assessments, if a program does not achieve the above-mentioned level of knowledge in technology, design, and production, it faces an increased risk of technical issues, as well as cost and schedule increases.

2.3 Computational Models of Cost and Schedule Growth

Within the systems engineering research community, there is consensus on requiring a new methodology for acquisition process analysis. In order to generate quantitative insights about the magnitude of potential cost overruns or schedule delays, recent research has focused on using agent-based modeling of the system acquisition process. Vermillion and Malak [22] formulate delegation in system engineering as a principal-agent problem based on the game theory. Their analysis investigates modeling assumptions and propose extensions to improve the model. Similarly, Safarkhani et al. [23] utilize a principal-agent model to represent strategic behaviors of systems engineer and subsystem engineers. Their model demonstrates how the requirement setting, effort of subsystem engineers, and their preferences influence the system-level quality. Safarkahni et al. [24] model the system acquisition process between a sponsor agency and contractors. Their model captures the effects of requirements, contract types, number of bidders and problem complexity on bidders' behaviors and the utility to the sponsor agency. Meluso and Austin-Brennan [25] propose an agent-based model to represent cognitive biases in setting requirements and subsystem estimation strategies. At the level of individual subsystem designers, researchers propose behavior decision models for capturing the relationships between effort and the evolution of design component quality [26]. These model require analytical data for behavioral validation. In that context, the program-specific time series data such as the dataset presented in this paper is useful.

3 Dataset

We analyze U.S. GAO's yearly review reports of major defense programs from year 2003 to 2017. These review reports provide quantitative data as well as detailed descriptions of acquisitions—cost increases, schedule delays, and performance shortfalls for these programs. The acquisition programs included in the GAO assessments are based on the following three factors; high program total cost, phase in acquisition, and interest from the congressional committee.

We selected 19 weapons programs from GAO's assessments of these defense programs (originally available as PDF files). We chose the programs that have maximum possible quantitative data (with minimum missing data entries) for multiple consecutive years. For most weapons programs, GAO's assessment of a program summarizes results in a 2-page format. Figure 9 and Figure 10 in Appendix A show an example of GAO's assessment for the Advanced Extremely High Frequency Satellite (AEHF) program for the year 2004. The first page in the GAO's assessment (Figure 9) includes an outline of the acquisition program which contains a program description (annotated in Figure 9 as 1); a timeline of development (annotated in Figure 9 as 2); a baseline comparison of changes made to program cost, schedule and quantity. (annotated in Figure 9 as 3). The second page in GAO's assessment (Figure 10), not only includes a narrative evaluation of *technology maturity*, *design maturity* and *production maturity*, but also includes other program concerns and feedback from the program office.

Throughout it's assessments, GAO utilizes a consistent acquisition terminology for major program milestones; *program start date*, *development start date* and *production decision date*, as shown in the program's development timeline (annotated as 2 in Figure 9). In Figure 9, the *program start date* alludes to the initiation of an acquisition program. The *development start date*, meanwhile, refers to the dedication to product development that initiates DOD's system development and demonstration process. Finally, the *production decision date* refers to the decision to begin production and deployment, usually with low-rate initial

Area	Reason	Code
Requirements development and management	Infeasible or unrealistic requirements; Unstable requirements (e.g., key performance requirements, readiness requirements, reliability and support requirements); Inefficiencies in requirements process (e.g., serial nature of process and requirements evolution); Restructuring efforts	req-ineff
Managing technical risk	Excessive technical, manufacturing, or integration risk (general) or program complexity; Unanticipated design, engineering, manufacturing, technical difficulty, or technology integration issues	tech-unanti
	Overly optimistic expectations (technical risks, performance goals, system requirements, or design maturity)	tech-optim
	Immature technology	tech-imm
	Immature design	tech-design
	Immature production capabilities; Prototyping; Developmental and operational test failures; Deficient test planning or testing inefficiencies	tech-test
	Concurrency in complicated programs	tech-concur
Resource allocation	Funding instability or budget cuts	ra-fundinsta
Defense acquisition management	Lack of focus on schedule or inadequate schedule management	acq-focus
	Overly optimistic assumptions, including insufficient contingency funds in program budgets; Overly optimistic assumptions in cost and schedule estimates	acq-optim
	Personnel issues	acq-person
	Setting up competitions to select contractors	acq-compt
	Contractor performance; Inadequate incentives; Delivery delay	acq-incent
	Inadequate tailoring of the acquisition process	acq-tailor
Other	Post delivery issues	oth-datadelay

TABLE 1: Coding the potential causes of cost overruns and schedule delays (adapted from [21, 27, 28, 29])

production.

Specifically, from the GAO's assessments for the 19 programs in our dataset, we extracted three different types of data. *First*, we extracted financial characteristics from the first page of the GAO's assessment. This includes baseline comparison of program cost (R&D cost, procurement cost and total program cost), acquisition cycle time, and total quantity. For example, the information included in the block annotated as 3 in Figure 9 represent financial characteristics for the AEHF program for year 2003. The total program costs not only includes the costs for the program's main elements such as acquisition cost—research and development, and procurement, but also, includes the military construction and acquisition operation and maintenance costs. In GAO's assessment of a program for a particular year, all the cost information presented is in that year's dollars. For the purpose of this paper, we normalize cost changes by defining *the percent cost change* across the program instead of specific dollar amount change. Next, the schedule assessment includes *the acquisition cycle time*, which is defined as the number of months between the program start and the achievement of initial operational capability or an equivalent fielding date.

Second, we extracted technical characteristics for the programs. This includes programs' technology maturity and design maturity as percentages from textual description. GAO defines *technology maturity* in terms of the critical technologies identified by a program to be successful. To assess technology maturity, GAO utilizes a measurement system, referred to as technology readiness levels [30]. In GAO's evaluations, a critical technology that has achieved level 7 on the technology readiness scale; a prototype that has been tested in a operational setting, is said to be mature. If a GAO report mentions that two out of four critical technologies are mature, we label the technology maturity percentage for that program as $2/4 = 50\%$. GAO defines the *design maturity* as "the percentage of engineering drawings completed or projected for completion by the design review" [17]. GAO considers a design as mature when a program office has released 90% of the engineering drawings for manufacturing. To assess *production maturity*, GAO uses the number of critical manufacturing processes and extent of statistical control achieved for those processes. To measure the statistical control achieved for the production processes, GAO uses a statistical measure called the process capability index [31]. Generally, the second page of the GAO's assessment for the program describes the state of technology and design maturity (see Figure 10). Note that the dataset does not include a program's production maturity data. This is because most programs included in the GAO reports do not report the process control data necessary to assess production maturity.

Third, the dataset includes narrative descriptions of major reasons for cost and schedule overruns, which are provided on the second page of GAO's assessment. Based on these descriptions, we further perform text analysis to code the reported reasons into various categories provided in Table 1. Section 4.1 provides more details on the text analysis process and its results.

The dataset includes these three pieces of information for each of the 19 selected programs and for every year such program was included in GAO's assessments between 2003-2017. For example, the data for the AEHF program is included in the overall dataset for each year during its reporting period of 2003-2011, except in the case of missing data (see program#1 in Table 2).

3.1 Hypotheses

Based on the review of literature [21] and the GAO reports [17], we extract the following hypotheses that can be tested on the dataset.

H1: Cost overruns and schedule delays experienced by a program early in the acquisition cycle are primarily due to immaturity in the programs' critical technologies.

H2: Cost overruns and schedule delays experienced by the program in the later part of the acquisition cycle can mostly be attributed to testing issues experienced by the program.

H3: Programs with lengthy initial acquisition times have higher total program cost growth [15,32].

H4: Programs with lower technology maturity at development start of the acquisition program tend to have higher future cost overruns [17].

H5: Programs with higher technical complexity (e.g. space programs) tend to have higher cost overruns and schedule delays as compared to programs with lower complexity (e.g. radio communication programs) [33,34].

4 Results

4.1 Text Analysis of Written Explanations

To identify recurring reasons behind schedule and cost growth, we conducted a text analysis on the GAO reports across 19 DoD weapons acquisition programs in our dataset. Based on the review of various possible reasons from Refs. [21,27,28,29], we formulated a set of codes, provided in Table 1. Using these codes as the baseline, we coded the narrative description of the product development knowledge of individual programs based on the GAO best practices. This analysis resulted in the program-specific, yearly database of text analysis codes for every program in the database.

Figure 1 represents the frequency with which different reasons are cited as responsible for cost and schedule overruns across the selected programs. We observe that the unanticipated reasons stemming from testing ('tech-unanti,' 'tech-test'), delays in achieving full technology maturity ('tech-imm') and design maturity ('tech-design'), and inefficiencies in setting requirements ('req-ineff') are the top reasons. Further, Figure 2 compares the frequency of different reasons cited for the first half and the second half of the available program data.

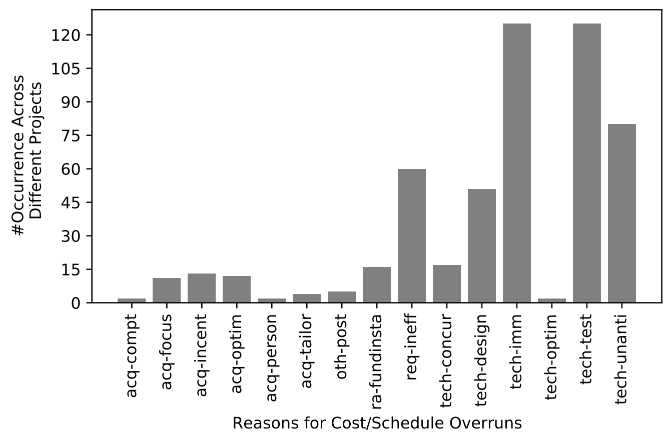


FIGURE 1: Frequency with which different reasons are cited in the GAO reports for cost and schedule overruns across all the 19 weapons acquisition program.

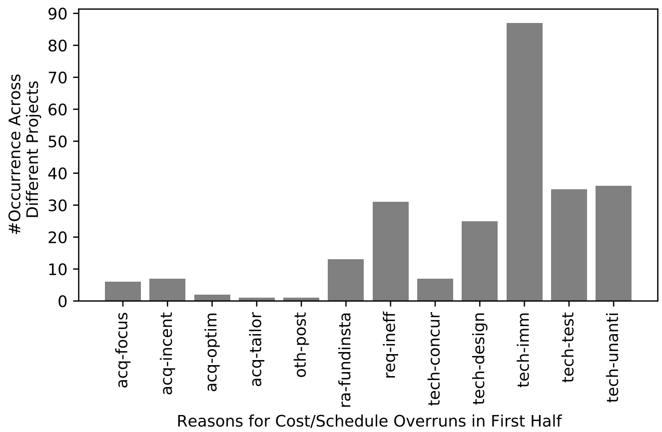
Technology immaturity tends to be a significant reason for cost and schedule overruns in the early stages of weapons acquisition program cycle (hypothesis H1 supported). When a program enters the development phase with immature technology, it is harder and takes longer to attain knowledge in succeeding elements-design and production [35]. For example, in the case of the Airborne Laser (ABL) program, the program had released 96% of the total expected design drawings, but the program still had to demonstrate full technology maturity. This mismatch increased the potential for design changes. However, in later stages, deficient test planning or testing inefficiencies ('tech-test') appears to be causing significant overruns (hypothesis H2 supported). This result is expected because the early stage of a weapons acquisition programs primarily includes the development phase where the program needs to demonstrate it's critical technologies, while the later stage of program deals with issues related to testing.

Program #	Program Name	Type	Reporting Period	Initial Total Program Cost (in Million USD)	Final Total Program Cost (in Million USD)	Initial Acquisition Time	Final Acquisition Time	Initial Quantity	Final Quantity
1	Advanced Extremely High Frequency Satellites (AEHF)	Satellite	2003-2011	6,276.5	12,919.6	111	170	5	6
2	LHA6	Ship	2007-2017	3,449.3	9,989.3	146	176	1	3
3	C-130 Avionics Modernization Program (C130AMP)	Aircraft	2004-2012	4,132.3	6,204.3	–	–	519	221
4	DD(X) Destroyer (DDG)	Ship	2004-2017	37,751.5	23,645.7	128	264	32	3
5	E-2 Advanced Hawkeye (E-2AHE)	Aircraft	2004-2013	15,180.7	19,938.4	95	136	75	75
6	Excalibur Artillery Round (EAR)	Missile	2003-2015	5,059.1	1,934.1	–	206	76677	7474
7	Joint Strike Fighter (JSF)	Aircraft	2003-2017	231,806.3	336,152.4	175	237	2866	2457
8	Global Hawk Unmanned Aircraft System (GHUAV)	Aircraft	2004-2014	5,671.1	9,874.5	55	125	63	45
9	Joint Air-to-Surface Standoff Missile (JASSM)	Missile	2003-2011	2,281.6	7,201	75	87	2469	5018
10	Joint Land Attack Cruise Missile (JLENS)	Missile	2006-2012	6,665.9	7,857.8	97	103	16	16
11	Joint Tactical Radio System Handheld, Manpack, Small Form Fit (JTRSHMS)	Radio Communication	2005-2017	10,888.2	9,563.7	85	124	328674	271202
12	JTRS Airborne, Maritime, Fixed-Station (JTRSAMF)	Radio Communication	2006-2017	8,844.7	3,314.1	80	186	27102	14222
13	JTRS Ground Mobile Radio (JTRSGMR)	Radio Communication	2004-2011	17,164.7	15,867.7	55	127	108388	87079
14	Mobile User Objective System (MUOS)	Satellite	2004-2015	1,186.5	1,267.7	90	–	6	6
15	Multi-mission Maritime Aircraft (MMA)	Aircraft	2005-2014	32,640.1	34,336	160	164	115	122
16	National Polar-orbiting Operational Environmental Satellite System (NPOESS)	Satellite	2003-2010	6,519.1	13,161.5	113	193	6	4
17	Small Diameter Bomb (SDB)	Missile	2004-2006	1,648.7	1,629	62	61	24070	24070
18	SDB Increment II (SDBII)	Missile	2007-2017	5,100.3	4,286.8	72	96	17163	17163
19	Space Based Infrared System (SBIRS)	Satellite	2003-2013	4,730.6	18,777.6	86	–	5	6

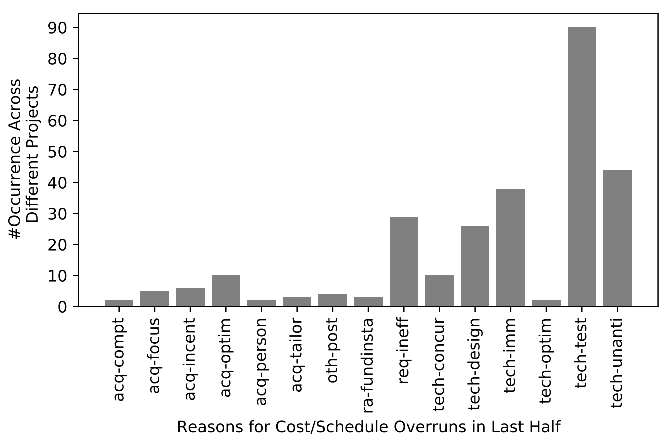
TABLE 2: Programs recorded in the dataset with budget, schedule and acquisition quantity information at the start year and the end year of the reporting period. Initial budget is inflation adjusted for the end year of reporting period. Note that our dataset includes budget, schedule and acquisition quantity information for all the years in the 'Reporting Period' column; not just the start year and the end year of reporting period.

4.2 Program Essentials Over Time

To better understand the reasons for the cost and schedule overruns, we analyze how program essentials such as the total program cost change, acquisition cycle time change, technology maturity, and design maturity change over the duration of the acquisition cycle. Figures 3a and 3b show the total program cost change and the acquisition cycle time change for 19 project over the years respectively. The percent changes for a given year are obtained by comparing that year's cost(or schedule) estimate to the available program baseline cost(or schedule) estimates. For instance, in the case of the LHA6 program, the end point of the dotted line in Figure 3a represents a 200% total cost change between the start year (as of 01/2006) and the end year of the reporting period (as of 09/2016). Similarly, the total program cost changes over 9 years of reporting for the AEHF program is highlighted using a dashed line in Figure 3a). Note that the cost growth and schedule growth may not be simultaneously re-



(a) First half of the available program data.



(b) Seconds half of the available program data.

FIGURE 2: Frequency of different reasons are cited in the GAO reports for cost and schedule overruns across the selected weapons acquisition programs (supports hypothesis H1).

ported for each year. That is why the lengths of reporting years for LHA6 program in Figures 3a and 3b are different.

These figures show trends of significant percentage change in the acquisition cycle time for some programs such as Program 4, Program 12, and Program 13 (see Table 2 for program details). These programs either incurred significant cost overruns in the early design stage or experienced increased cost risk because of which they underwent restructuring. The restructuring further led to a significant increase in the acquisition cycle time and decrease in the number of quantities being procured.

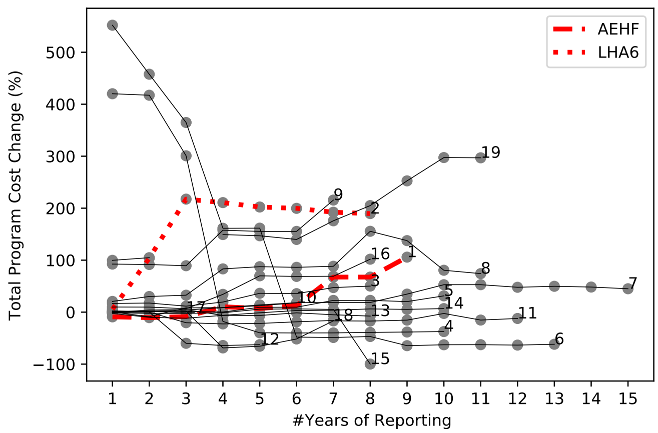
We also observe a decreasing trend in the percentage cost changes over time for programs such as Program 4 and Program 6. Program 4 and program 6 had initial total cost overruns of approximately 420% and 552% respectively as given in Figure 3a. According to GAO's first report (year 2004), Program 4's initial cost increases was due to changes in cost estimates, technology development programs, and program restructuring. For example, the total quantities to be produced were unclear in Program 4 between years 1998 to 2008, and eventually reported to be 32 for report year 2009. Over the years, there were conflicts between how GAO reports cost and schedule growths and how the Navy perceived the acquisition program. The GAO's 2006 year report indicates only one of the twelve critical technologies as mature for Program 4. The Navy's comment in the GAO 2006 report notes that comparing the Program 4's technology readiness levels to the GAO-developed best practices may not be valid. This was because the the detail design and construction were spread over 5 years (longer than GAO's yearly reporting cycle) given the complex nature of shipbuilding.

Similarly, since its development began in 1997, Program 6 encountered four major restructures, reduced production quantities, and increased unit costs. Our text analysis found major reasons such as initial technology and design immaturity during the early part of the program, limited funding (which led to program restructuring in 2001), and testing issues in the later part of the programs. Apart from restructuring, the percentage cost reductions may be attributed to the changes in the baseline report year used for comparison. Starting year 2008, GAO compared the yearly estimates with those year 02/2003, instead of year 1997. The GAO reports did not specify a reason for the change.

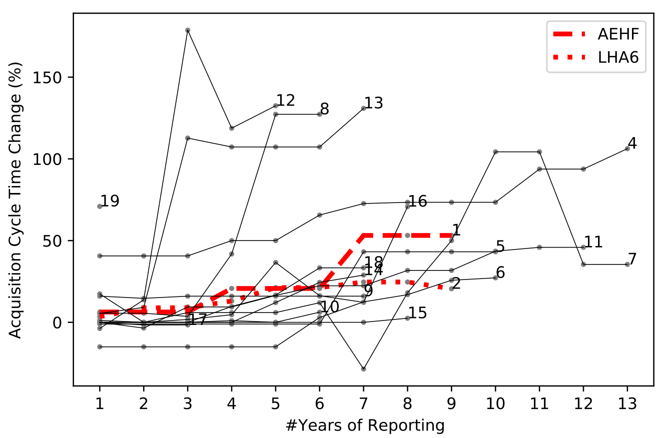
Figures 4a and 4b present how the technology maturity and design maturity change over time for the selected programs. The complex and sequential nature of the systems engineering process is prominent as observed from maturity trends in the AEHF and the LHA6 programs. Both these programs attain high level of technology and design maturity but still incur cost and schedule overruns (see Figure 3a and Figure 3b). For the AEHF program, the major cost overruns occurred in the later stage due to testing issues. While for LHA6 program, the initial design immaturity (which resulted in high levels of rework during the ships' construction) seems to be the reason for cost and schedule overruns. The LHA6 program began development with mature critical technologies but only 65% design drawings completed. These examples highlight how issues early in the program cascade to the later stages of the program cycle.

4.3 Correlations Between the Program Essentials

Figure 5 presents the correlation between the initial acquisition cycle times and cost growth for the programs in our dataset. Note that, in Figure 5, some programs (such as Programs 2 and 3) are missing because the initial acquisition time for these programs was unavailable in the GAO's assessments. The results indicate that the said effect is not statistically significant. Prior studies denote longer acquisition cycle times as one reason behind greater cost growth. Longer programs are generally at the risk of changing requirements and other time-related



(a) Total program cost change (%) over the years.

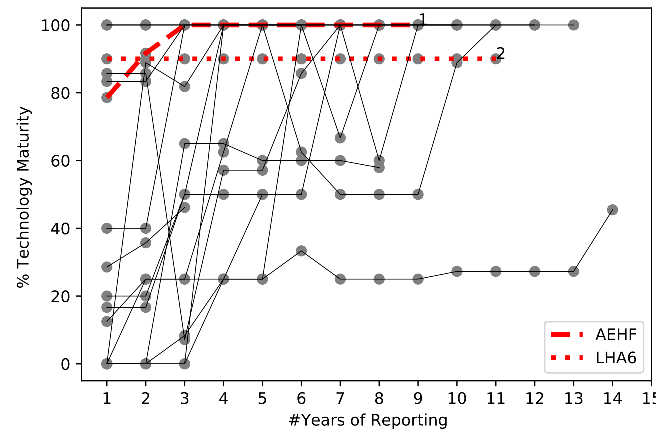


(b) Acquisition cycle time change (%) over the years.

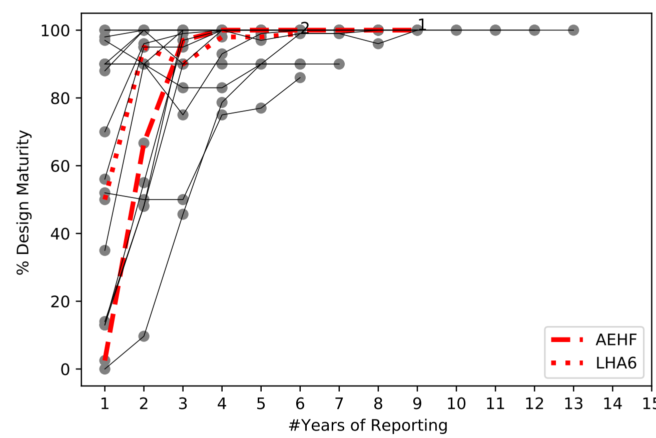
FIGURE 3: Program cost change and acquisition cycle time change data for all 19 weapons acquisition programs in our dataset. The numbers (1,2,...,19) at the end of the lines represent different programs. Program names corresponding to these numbers can be found in Table 2. Further, the dashed and the dotted lines represent AEHF Satellites program and LHA 6 Amphibious Assault Ship Replacement Program respectively. (supports hypothesis *H2*).

costs [15, 36]. However, this hypothesis (hypothesis H3) is not supported by the data.

Further, Figure 6 presents how a program's total cost varies with the technology maturity at the development start. Programs 7 and 12 for which the technology maturity at the development start was not available in the GAO reports are excluded from the analysis. We observe that the programs with low technology maturity at the development start do not necessarily cause higher cost overruns. In fact, the opposite trend is more likely. Hypoth-



(a) Technology Maturity (%) over the years.



(b) Design Maturity (%) over the years.

FIGURE 4: Technology maturity change and design maturity change data for all 19 weapons acquisition programs in our dataset. Further, the dashed and the dotted lines represent Advanced Extremely High Frequency (AEHF) Satellites program and LHA 6 Amphibious Assault Ship Replacement Program respectively.

esis H4 is not supported by the given data. While this may seem counter-intuitive, but here we have to keep in mind that systems engineering is a complex process where multiple factor determines the programs performance. For example, even though the AEHF Satellites program had all 14 critical technologies mature, the program incurred significant cost and schedule overruns due to testing and integration issues.

4.4 Effects of Program Type on Program Essentials

To better understand how the type of a program affects the cost and schedule overruns, we assigned the selected weapons

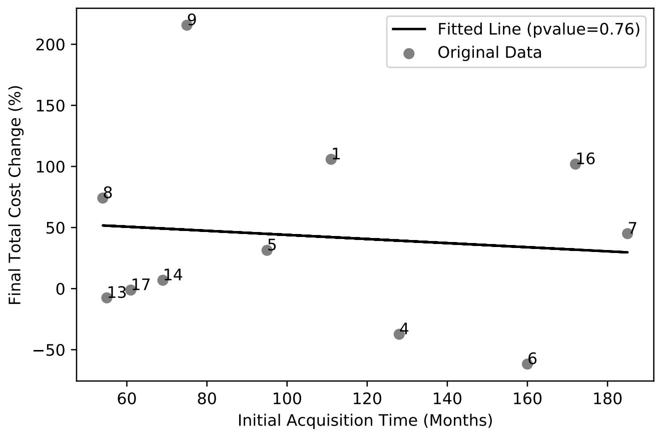


FIGURE 5: Analyzing the correlation between initial acquisition cycle time and total cost change for different programs.

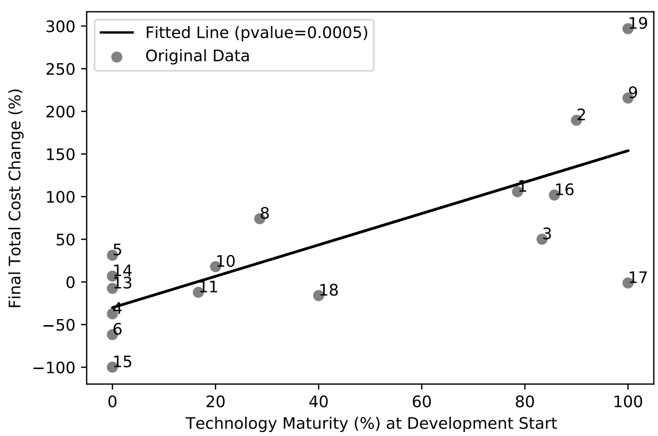


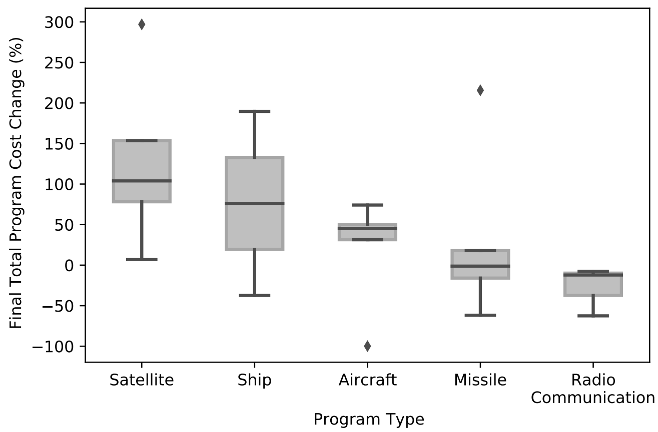
FIGURE 6: Final total program cost change (%) as a function of technology maturity at the development start of the program.

acquisition programs into 5 different types based their weapons category: satellite, ship, aircraft, missile and radio communication. Assuming the size of program budget, testing difficulty, and problem size as proxies for problem complexity [37], these program types are in a order of decreasing complexity. Figure 7 compares the differences in the final total program cost change (%) and final acquisition cycle time change (%) between the program types. The final quantities refer to the last year these projects were included in the GAO reports. In Figure 7a, we observe that the satellite programs has the highest final total program cost change (%) while radio communication programs has the lowest. Testing satellite's critical technologies is much harder than that of the radio devices because testing satellites in the space environment is infeasible. Moreover, the issues discovered during testing (later stage of the acquisition process) have a higher impact on cost and schedule of a program since they lead to costly technology or design changes [38]. This is further supported by the text analysis results (4.1) for Space-Based Infrared System (SBIRS) High program where testing inefficiency was one of the major reason for cost overrun in this project. Inversely, in Figure 7b, we observe that the radio communication programs happened to have the highest increase in the acquisition cycle time. Radio communication programs such as JTRS GMR and JTRS AMF struggled to mature and integrate multiple key technologies. They had been forced to make design changes which caused the programs to incur significant schedule problems. The schedule delays required these program to be restructured. Therefore, for the given dataset, hypothesis *H5* is partially supported; even though complex programs (e.g. space programs) tend to have higher cost overruns as compared to less complex programs (e.g. radio communication programs), the same is not true schedule delays.

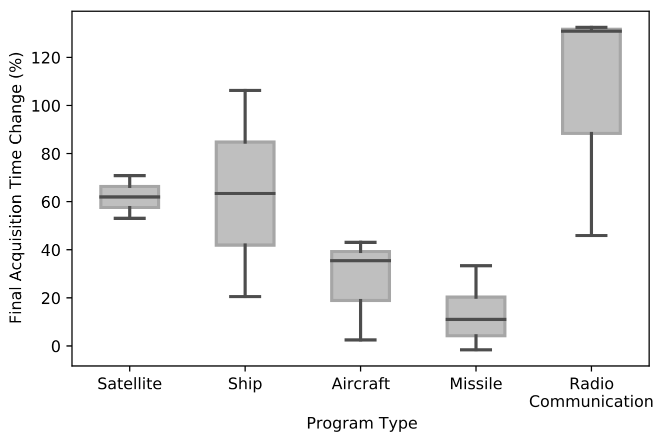
Further, Figure 8 highlights the differences in the critical technology maturity (%) at the development start across different program types . The satellite programs happen to be on the higher end of the technology maturity. Even with the higher percentage of the critical technologies mature, satellite programs have relatively larger final total program cost change (%) as compared to the other program types as seen from Figure 7a. For example, the AEHF Satellites program even with all 14 critical technologies mature, encountered problems with integration and testing of these technologies which led to additional cost and schedule overruns.

5 Conclusions

This paper presents a dataset for 19 DoD weapons acquisition program. The dataset contains information about how following program essentials change over time: (i) cost (R&D cost, procurement cost and total cost), (ii) schedule, (iii) procurement quantity, (iv) technology maturity, and (v) design maturity. We extracted the qualitative and quantitative data, originally available as PDF files, from GAO's reports (between years 2003-2017) into Excel files and further created an easy to access Pandas library DataFrame in a Python environment. Both, the raw dataset (Excel files) and the Pandas library DataFrame are made available through Purdue University Research Repository (PURR). The results highlight the complex (and serial) nature of the systems engineering process; where multiple factors contribute to the success of a program. One might expect that low technology maturity at the development start of a program should cause greater cost overruns, but that is not necessarily the case. Other issues, such as design maturity, program complexity and testing issues needs to be taken in conjunction with the technology maturity while reasoning about the cost and schedule overruns. Similarly, some programs, even with high technology and design maturity at the beginning of the acquisition cycle incur



(a) Final total program cost change (%) as a function of different program type.



(b) Final acquisition cycle time change (%) as a function of different program type.

FIGURE 7: Comparison of final total program cost change (%) and final acquisition cycle time change (%) as a function of program type.

significant cost and schedule overruns due to testing issue later in the later part of the project; this further leads to more costly design changes [38]. Future work could consider determining the causal factors for cost and schedule overruns in large systems engineering programs, e.g., through creating a game-theoretic model of the acquisition process that is capable of predicting cost overruns and schedule delays given program information knowable at a given point in time. This game-theoretical model can be calibrated on the data from major US DoD weapons acquisition programs (presented in this paper) using a Bayesian approach. The causality analysis can be carried out using Pearl's causal inference framework [39] for structural equations.

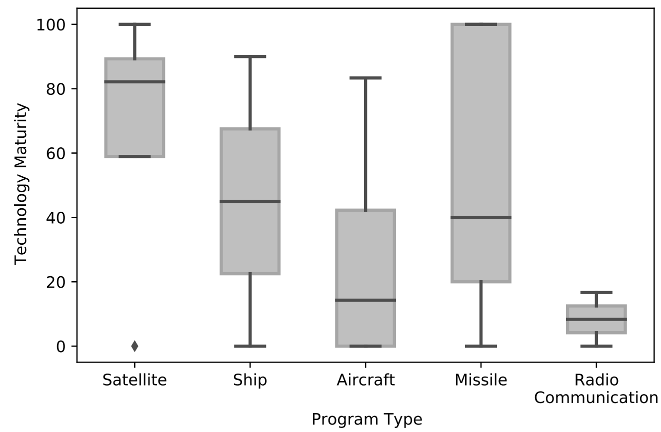


FIGURE 8: Technology maturity at the development start for different program types.

ACKNOWLEDGMENT

The authors gratefully acknowledge financial support from the National Science Foundation through NSF CMMI Grants No. 1662230 and No. 1728165.

REFERENCES

[1] Cantarelli, C. C., Flybjerg, B., Molin, E. J., and Van Wee, B., 2013. "Cost overruns in large-scale transportation infrastructure projects: Explanations and their theoretical embeddedness". *arXiv preprint arXiv:1307.2176*.

[2] Locatelli, G., 2018. "Why are megaprojects, including nuclear power plants, delivered overbudget and late? reasons and remedies". *arXiv preprint arXiv:1802.07312*.

[3] GAO, 2017. "Navy shipbuilding: Past performance provides valuable lessons for future investments". *https://www.gao.gov/assets/700/692331.pdf*.

[4] Maddox, I., Collopy, P., and Farrington, P. A., 2013. "Value-based assessment of dod acquisitions programs". *Procedia Computer Science*, ***16***, pp. 1161–1169.

[5] GAO, 2018. "Nasa: Assessments of major projects". *https://www.gao.gov/products/GAO-18-280SP*.

[6] Cost, F.-A. H., 2003. Oversight of major highway and bridge project–issues and options. Tech. rep., GAO-03-764T, Washington, DC.

[7] Schinasi, K. V., 2009. *Defense Acquisitions: Better Weapon Program Outcomes Require Discipline, Accountability, and Fundamental Changes in the Acquisition Environment: Testimony*. DIANE Publishing.

[8] Mayer, K. R., Thomas, K., Bodilly, S. J., Camm, F., and Webb, T. J., 1993. Barriers to managing risk in large scale weapons systems development programs. Tech. rep., RAND CORP SANTA MONICA CA.

[9] Pernin, C. G., Axelband, E., Drezner, J. A., Dille, B. B., Gordon, I., Held, B. J., McMahon, K. S., Perry, W. L., Rizzi, C., Shah, A. R., et al., 2012. Lessons from the army's future combat systems program. Tech. rep., RAND ARROYO CENTER SANTA MONICA CA.

[10] Decker, G., and Wagner Jr, L., 2011. Army strong: Equipped, trained and ready: Final report of the 2010 army acquisition review.

[11] Blickstein, I., Boito, M., Drezner, J. A., Dryden, J., Horn, K., Kallimani, J. G., Libicki, M. C., McKernan, M., Molander, R. C., and Nemfakos, C., 2011. Root cause analyses of nunn-mccurdy breaches, volume 1: Zumwalt-class destroyer, joint strike fighter, longbow apache and wideband global satellite. Tech. rep., RAND NATIONAL DEFENSE RESEARCH INST SANTA MONICA CA.

[12] Blickstein, I., Drezner, J. A., Libicki, M. C., McInnis, B., McKernan, M., Nemfakos, C., Sollinger, J. M., and Wong, C., 2012. Root cause analyses of nunn-mccurdy breaches. volume 2: Excalibur artillery projectile and the navy enterprise resource planning program, with an approach to analyzing program complexity and risk. Tech. rep., RAND NATIONAL DEFENSE RESEARCH INST SANTA MONICA CA.

[13] Fox, B., Boito, M., Graser, J. C., and Younossi, O., 2004. Test and evaluation trends and costs for aircraft and guided weapons. Tech. rep., RAND CORP SANTA MONICA CA.

[14] Younossi, O., Stem, D. E., Lorell, M. A., and Lussier, F. M., 2005. Lessons learned from the f/a-22 and f/a-18e/f development programs. Tech. rep., Deputy Chief of Staff, Plans and Operation (Air Force), Washington DC.

[15] Drezner, J. A., and Smith, G. K., 1990. An analysis of weapon system acquisition schedules. Tech. rep., RAND CORP SANTA MONICA CA.

[16] Younossi, O., Lorell, M. A., Brancato, K., Cook, C. R., Eisman, M., Fox, B., Graser, J. C., Kim, Y., Leonard, R. S., Pfleeger, S. L., et al., 2008. Improving the cost estimation of space systems. past lessons and future recommendations. Tech. rep., RAND Corp. Santa Monica.

[17] GAO. "Defense acquisitions: Assessments of selected major weapon programs".

[18] Meier, S. R., 2010. "Causal inferences on the cost overruns and schedule delays of large-scale us federal defense and intelligence acquisition programs". *Project Management Journal,* ***41***(1), pp. 28–39.

[19] Hofbauer, J., Sanders, G., Ellman, J., Morrow, D., Berteau, D., Ben-Ari, G., and Lombardo, N., 2011. Cost and time overruns for major defense acquisition programs. Tech. rep., Acquisition Research Program.

[20] Dwyer, M., Cameron, B., and Szajnfarber, Z., 2015. "A framework for studying cost growth on complex acquisition programs". *Systems Engineering,* ***18***(6), pp. 568–583.

[21] Riposo, J., McKernan, M., and Duran, C. K., 2014. "Prolonged cycle times and schedule growth in defense acquisition: A literature review".

[22] Vermillion, S. D., and Malak, R. J., 2015. "Using a principal-agent model to investigate delegation in systems engineering". In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. 57052, American Society of Mechanical Engineers, p. V01BT02A046.

[23] Safarkhani, S., Bilionis, I., and Panchal, J. H., 2020. "Toward a theory of systems engineering processes: A principal–agent model of a one-shot, shallow process". *IEEE Systems Journal,* ***14***(3), pp. 3277–3288.

[24] Safarkhani, S., Bilionis, I., and Panchal, J. H., 2020. "Modeling the system acquisition using deep reinforcement learning". *IEEE Access,* ***8***, pp. 124894–124904.

[25] Meluso, J., and Austin-Breneman, J., 2018. "Gaming the system: An agent-based model of estimation strategies and their effects on system performance". *Journal of Mechanical Design,* ***140***(12).

[26] Chaudhari, A. M., Bilionis, I., and Panchal, J. H., 2020. "Descriptive models of sequential decisions in engineering design: An experimental study". *Journal of Mechanical Design,* ***142***(8).

[27] Bolten, J. G., Leonard, R. S., Arena, M. V., Sollinger, J. M., and Younossi, O., 2008. *Sources of weapon system cost growth: Analysis of 35 major defense acquisition programs*, Vol. 670. Rand Corporation.

[28] Katz, D. R., Sarkani, S., Mazzuchi, T., and Conrow, E. H., 2015. "The relationship of technology and design maturity to dod weapon system cost change and schedule change during engineering and manufacturing development". *Systems Engineering,* ***18***(1), pp. 1–15.

[29] Smirnoff, J. P., and Hicks, M. J., 2008. "The impact of economic factors and acquisition reforms on the cost of defense weapon systems". *Review of Financial Economics,* ***17***(1), pp. 3–13.

[30] Mankins, J. C., 1995. "Technology readiness levels". *White Paper, April,* ***6***(1995), p. 1995.

[31] Pearn, W., and Chen, K., 1999. "Making decisions in assessing process capability index cpk". *Quality and reliability engineering international,* ***15***(4), pp. 321–326.

[32] Belay, A. M., and Torp, O., 2017. "Do longer projects have larger cost deviation than shorter construction projects?". *Procedia engineering,* ***196***, pp. 262–269.

[33] Ben-Ari, G., and Chao, P. A., 2009. *Organizing for a complex world: developing tomorrow's defense and net-centric systems*, Vol. 31. CSIS.

[34] Keller, S., Collopy, P., and Componation, P., 2014. "What is wrong with space system cost models? a survey and assessment of cost estimating approaches". *Acta Astronautica,* ***93***, pp. 345–351.

[35] Office, U. S. G. A., 2004. *Defense Acquisitions: Assess-*

ments of Major Weapon Programs: Report to Congressional Committees. US General Accounting Office.
[36] Arena, M. V., Leonard, R. S., Murray, S. E., and Younossi, O., 2006. "Historical cost growth of completed weapon system programs".
[37] Summers, J. D., and Shah, J. J., 2010. "Mechanical engineering design complexity metrics: size, coupling, and solvability". *Journal of Mechanical Design,* ***132***(2).
[38] Tranøy, E., and Muller, G., 2014. "7.1. 1 reduction of late design changes through early phase need analysis". In INCOSE International Symposium, Vol. 24, Wiley Online Library, pp. 570–582.
[39] Pearl, J., Glymour, M., and Jewell, N. P., 2016. *Causal inference in statistics: A primer*. John Wiley & Sons.

A Sample Program Assessment from GAO's Assessments of Major Weapon Pograms

Common Name: AEHF

Advanced Extremely High Frequency Satellite (AEHF)

1

The Air Force's AEHF satellite system is intended to replenish the existing Milstar system with higher capacity, survivable, jam-resistant, worldwide, secure communication capabilities for strategic and tactical warfighters. The system also includes a mission control segment. Terminals used to transmit and receive communications are acquired separately by each service. AEHF is an international partnership program that includes Canada, United Kingdom, and the Netherlands. First launch is scheduled for December 2006.

Source: AEHF Program Office.

2

Concept		System development			Production		
Program start (4/99)	Development start (9/01)	GAO review (1/04)	Design review (4/04)	Production decision (8/04)	First launch (4/07)	Initial capability (2/09)	

Program Essentials

Prime contractor: Lockheed Martin
Program office: El Segundo, Calif.
Funding needed to complete:
R&D: $2,452.0 million
Procurement: $473.6 million
Total funding: $2,925.6 million
Procurement quantity: 1

Program Performance (fiscal year 2004 dollars in millions)

	As of 10/2001	Latest 09/2003	Percent change
Research and development cost	$4,168.4	$4,356.1	4.5
Procurement cost	$1,232.9	$473.6	-61.6
Total program cost	$5,401.3	$4,829.8	-10.6
Program unit cost	**$1,080.254**	**$1,609.917**	**49.0**
Total quantities	5	3	-40.0
Acquisition cycle time (months)	111	118	6.3

3

The AEHF satellite program demonstrated most of its critical technologies at development start and expects to have all technologies demonstrated by the design review. The program has released two-thirds of its drawings and expects to complete 90 percent by the design review. In September 2003, the program office decided to delay the launch of the first two satellites by 4 months. The delay was necessary to accommodate changing security requirements and resolve fabrication issues relating to the critical cryptological equipment.

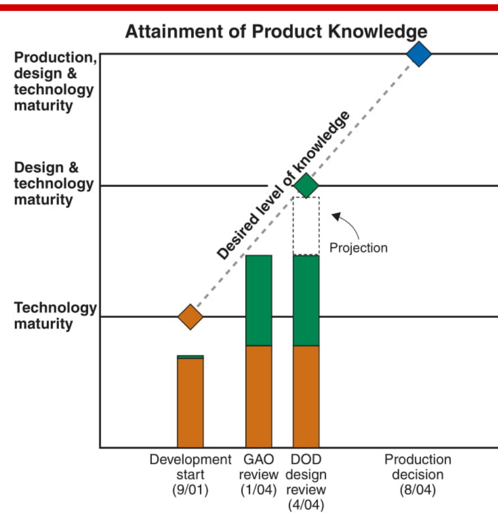


FIGURE 9: The first page of GAO's assessment for the AEHF program contains a profile of the product, a timeline of development, a baseline comparison of cost, schedule, etc., and quantity changes to the program. Source: GAO [17]

Common Name: AEHF

AEHF Program

Technology Maturity

Eleven of the critical technologies identified by the program office are mature. The remaining three technologies have engineering models and are undergoing testing in a relevant environment to simulate both launch and space atmosphere. Two of these three technologies have mature backup technologies. Only one technology, a component of the phased array antenna, does not have a backup technology that meets operational requirements. Program officials expect all technologies to be mature by the design review scheduled for April 2004.

Design Maturity

The program office has released over two-thirds of its expected drawings. The program office expects to release 90 percent of the expected drawings by the scheduled design review. In addition, preliminary design reviews are complete and the program office has initiated the subsystem design reviews. The program is also developing early software builds for the ground and space segments.

Production Maturity

The production maturity could not be assessed because the program office does not have statistical process control data. The Air Force currently plans to buy only three satellites. However, there have been some problems in producing a critical system component. The AEHF Comsec/Transec System (ACTS) is a suite of cryptological equipment installed in both the satellites and the terminals to limit access to authorized users. ACTS has already experienced significant cost growth and schedule delays due to changes in satellite architecture design, interface, and other requirements changes. ACTS consists of computer chips whose fabrication is more technically challenging than producing other computer chips. The challenge results from a security requirement to have separate foundries produce components of the chips that must be integrated together. During a major functional test in September 2003, a problem was discovered, and the program is evaluating ways to resolve the problem.

Concurrent development of ACTS and the AEHF satellite payload has resulted in a 4-month delay in the launch of the first two AEHF satellites, now scheduled for April 2007 and April 2008, respectively. ACTS is managed by the National Security Agency and is on the AEHF satellite payload critical path. The program office stated the launch delay was necessary to accommodate changes in ACTS security requirements and resolve ACTS production issues.

Other Program Issues

The current development contract includes the first two satellites and the mission control segment. A decision to buy a third satellite is planned after the design review. In December 2002, two satellites were deleted from the program because the newly developed Transformational Communications Architecture calls for the Transformational Satellite, assessed elsewhere in this report, to replace these AEHF satellites. Because the Transformational Satellites are early in development and may not progress in time to meet the military need, the Air Force has scheduled a progress review and decision point in early fiscal year 2005 to determine if additional AEHF satellites will be needed to meet operational requirements.

Program Office Comments

In commenting on a draft of this assessment, the program office noted that the AEHF program continues to progress through the system development and demonstration phase, meeting all scheduled milestones and is projected to meet all key performance parameters.

Page 18 GAO-04-248 Assessments of Major Weapon Programs

FIGURE 10: The second page of GAO's assessment for the AEHF program contains a narrative depiction of how the product development knowledge of an individual program compared to GAO's best practices. Source: GAO [17]