1-2018

Unmasking Cost Growth Behavior: A Longitudinal Study

Cory N. D'Amico

Edward D. White
Air Force Institute of Technology

Jonathan D. Ritschel
Air Force Institute of Technology

Scott R. Kozlak

Recommended Citation

D'Amico, C. N., White, E. D., Ritschel, J. D., & Kozlak, S. J. (2017). Unmasking cost growth behavior: A longitudinal study. Defense Acquisition Research Journal, 25(1), 30–51.

Unmasking Cost GROWTH BEHAVIOR: A LONGITUDINAL STUDY

Cory N. D'Amico, Edward D. White, Lt Col Jonathan D. Ritschel, USAF, and Capt Scott J. Kozlak, USAF

This article examines how cost growth factors (CGF) change over a program's acquisition life cycle for 36 Department of Defense aircraft programs. Starting from Milestone B, the authors examine CGFs at five gateways: Critical Design Review, First Flight (FF), the end of Developmental Test and Evaluation (DT&E), Initial Operational Capability, and Full Operational Capability. Each CGF is assigned a color rating based upon the program's cost growth: *Green* (low), *Amber* (moderate), or *Red* (high). Significant findings include dependencies among similar CGF color ratings and cost growth occurring primarily between FF and the end of DT&E during a program's life cycle.

DOI: *https://doi.org/10.22594/dau.17-783.25.01*

Keywords: *Cost Growth, Longitudinal Analysis, Risk Assessment, Test and Evaluation, Developmental Test and Evaluation*

Department of Defense (DoD) acquisition programs historically have experienced cost overruns and schedule delays (Katz, Sarkani, Mazzuchi, & Conrow, 2015). This statement does not imply that all DoD acquisition programs will cost more and take longer than expected to complete. Rather, Katz et al. simply document a long line of reports and studies that suggest DoD acquisition programs are risky. Those individuals involved with planning and budgeting need to be cognizant of the reality that these programs often cost more and take longer than originally anticipated.

Many aspects such as the DoD acquisition system itself (the legislation, policies, and processes layered on top of practice), to possible incorrect incentives for program managers, or even the sheer complexity of today's DoD acquisition programs are potential contributing factors to this phenomena. O'Neil (2011) shows that controlling cost growth in defense acquisition has not improved in any material respect over at least the past four decades and has a variety of causes, including errors in the management or contracting process—but defects in the original concept are a very common cause. Other recent studies have shown that such information before program initiation may play a key role in mitigating future issues (Bolten, Leonard, Arena, Younossi, & Sollinger, 2008; Jimenez, White, Brown, Ritschel, Lucas, & Seibel, 2016).

Additional studies have examined specific platforms to analyze macro programmatic trends (Smirnoff & Hicks, 2008). One particular study by Kozlak, White, Ritschel, Lucas, and Seibel (2017) investigated cost growth with respect to DoD aircraft programs only. Their research examined Cost Growth Factors (CGF) at various program stages: Critical Design Review (CDR), First Flight (FF), end of Development Test and Evaluation (DT&E), Initial Operational Capability (IOC), and Full Operational Capability (FOC).

As a follow-on to that research, this article further scopes individual DoD aircraft programs (including modifications/upgrades) over time to identify potential common elements that suggest a high likelihood of experiencing cost growth at certain longitudinal points. In doing so, the aim is to highlight areas upon which to focus in the future to militate against cost growth.

Background

Cost growth is no newcomer to the acquisition hot-topics list. For decades, research has been rich in evaluating cost growth (Arena, Leonard, Murray, & Younossi, 2006; Bolten et al., 2008; Cancian, 2010; Christensen & Payne, 1992; Drezner, Jarvaise, Hess, Hough, & Norton, 1993). These studies have all been foundational in the development of cost research over the last three decades. Traditionally, most of these studies address cost growth over its complete life cycle and evaluate how growth escalates from program initiation—Milestone B (MS B)—to program completion. In contrast, the literature is scant with studies that evaluate cost growth from a longitudinal perspective, i.e., at particular time intervals or between these intervals.

On one end of the longitudinal spectrum, research has turned to Earned Value Management (EVM) with the intent of locating stabilization of cost growth at the individual contract level. Since the early 1990s, the EVM community has adopted what is known as the *stability rule* adopted from Christensen and Payne (1992), who studied 26 Contract Performance Reports across seven DoD aircraft contracts and evaluated the stability of Cost Performance Index (CPI) levels measured against percent complete. Christensen and Payne (1992) defined stability as a range of no more than 0.20, or an interval of +/- 10% (assumes symmetry), and demonstrated that the CPI stabilizes when the contract is at 20% completion for the interval definition, and at 50% completion for the range definition. It was this research, though based on very limited data, which established an industry standard of assuming CPI stability at 20% contract completion.

Petter, Ritschel, and White (2015) recently explored this concept of CPI stability through an analysis of modern DoD data. Their study highlighted the vague definition of stability, as denoted within their literature review, and further summarized the definition into three broad categories: range definition, absolute interval definition, and relative interval definition. [Note: For more granularity, see Table 3 in Petter et al. (2015, pp. 348–349).] Their findings suggest Christensen and Payne's stability rule was both supported and contradicted, depending on the definition of stability used. Petter et al.

findings suggest that for this article, a similar longitudinal pattern might appear. That is, if a DoD aircraft program did not display cost growth by the 20% complete point (which Kozlak et al. [2017] suggests is shortly before FF), then there is a high probability the program will remain on budget. In contrast, programs that displayed cost growth at 20% complete would be expected to remain permanently over budget.

While EVM is a valuable tool, and understanding the CPI and its predictive measures on final cost is valuable, EVM can be restrictive because DoD acquisition programs are often split into a variety of contracts. Therefore, the CPI limits measuring and predicting program cost growth to only a small portion of the overall cost growth since the metric is based on a contract and not an aggregate program level. However, EVM is useful in the sense that instead of looking at simply the start and finish of a program, this process allows for investigating longitudinal patterns over time that may provide insight into discovering factors that are associated with cost growth at various points in time.

> **“The CPI limits measuring and predicting program cost growth to only a small portion of the overall cost growth since the metric is based on a contract and not an aggregate program level.**

The other spectrum of the longitudinal approach encompasses how cost growth trends over the course of history. Davis and Anton (2016) evaluated a 5-year moving average of annual cost growth within the DoD over the last three decades to assess and analyze patterns. Their macro longitudinal approach showed how the acquisition process, policy reform, and the behavior of the defense acquisition system played a role in cost growth trends over time. While Davis and Anton examined individual programs in the collective, their analysis did not evaluate how factors of an individual program (for instance type of program, new start versus modification, branch of service, etc.) possibly correlated with cost growth during various stages of systems development.

Kozlak et al. (2017) combined the idea of assessing cost growth percentage at a portfolio level (DoD aircraft) with respect to percent complete rates (like that of EVM) to determine at what stages during an aircraft's acquisition life cycle does cost growth occur the most. Overall, Kozlak et al. determined that the first spike in percent of total (development and procurement) cost growth occurs at FF, and that by IOC (occurring approximately at 48% of program completion), an aircraft program realizes 91% of its total acquisition cost growth. Their analysis did not investigate what factors may contribute to the likelihood of an aircraft program experiencing cost growth as assessed by CGFs at five specific gateways (which is defined in the next section). This article aims to fill in this literature gap by providing that assessment.

Methods and Analysis

The CGF used in this article, which divides the cost variance (actual cost) by the development estimate at MS B, mirrors the CGFs used by both Arena et al. (2006) and Kozlak et al. (2017). The equation for the calculation of the CGF is as follows:

$$\frac{Actual}{Estimated} \qquad (1)$$

Consistent with the methodology of Arena et al. (2006), the final Selected Acquisition Report (SAR) is considered to be the “actual” cost despite SAR reporting stopping at the 90% program completion point. In describing their rationale for using the final SAR as “actual” cost, they state that “the cost for the final SAR should be very close to the final cost, as most of the funding has been spent at that point” (Arena et al., 2006). The denominator in equation (1) comes from the cost estimate at MS B as reported in the SAR.

A CGF of 1.0 indicates the program did not go over or under the cost estimate, and the actual cost matched the estimated cost. If the CGF is greater than 1.0, the program sustained growth, calculated as the CGF – 1.0 to determine the percent cost growth. Conversely, if the CGF is less than 1.0, the program did not sustain cost growth; rather, the program cost less than the estimate. To calculate the percent cost growth, subtract 1 from the CGF (Drezner et al., 1993). For program costs, this article considers only those associated with acquisition (development and procurement) and not those for operations and support.

This article explores how CGFs change between specific review points for DoD aircraft programs, each of which is designated a Major Defense Acquisition Program (MDAP). As defined in Department of Defense Instruction 5000.02 (DoD, 2017), such programs are estimated to require eventual expenditure for research, development, test and evaluation, including all planned increments, of more than $480 million in Fiscal Year (FY) 14 constant dollars. Likewise, procurement expenditures, including all planned increments, are estimated at more than $2.79 billion (FY 2014 constant dollars). Aircraft MDAPs customarily pass through program review points in a particular order, namely CDR, FF, the end of DT&E, IOC, and then FOC; however, some programs deviated from this traditional progression. Thus, for this analysis, these five review points will simply be termed *gateways* in the collective and denoted as Gateway 1 – Gateway 5 individually. [Note: For brevity, in lieu of stating end of DT&E, going forward, simply DT&E will be used.]

The SAR contains program estimates, key dates, and other relevant programmatic data. This study utilized the Defense Acquisition Management Information Retrieval (DAMIR) system to access relevant SARs in addition to using a database compiled by the RAND Corporation for the Air Force Cost Analysis Agency to access older SARs that were not obtainable in DAMIR. Cost growth was evaluated on a longitudinal perspective by investigating the change from the development cost estimate at MS B to the Final or Current Estimate, if still ongoing, for each individual program.

In total, the study's database consists of 36 DoD aircraft programs, including modifications/upgrades (Table 1). All cost figures are standardized to a common Base Year (BY) using the Office of the Secretary of Defense Comptroller Appropriation inflation rates to perform these conversions. While not all 36 programs are converted to the same BY, each individual program maintains a common BY, which allows for consistent comparison of costs longitudinally. In most cases, the BY is the original year in which the program was estimated. In some cases, however, the BY was updated at some point throughout the life of the program. In these instances, the most recent BY is applied across all gateways.

TABLE 1. 36 AIRCRAFT PROGRAMS CONSIDERED FOR ANALYSIS

A-10	C-27J	F-35
AV-8B	E-2C	RQ-4
B-1A	E-2D	MQ-4C
B-1B	E-6A	P-8A
B-1B CMUP	EA-18G	S-3A
B-1B JDAM	EF-111	T-6 (JPATS)
B-2 RMP	FA-18EF	T-45TS
C-130 AMP	F-14A	V-22 FSD
C-17A	F-14D	F-22 Inc 3.2B Mod
C-5 AMP	F-15	KC-46
C-5 RERP	F-16	MQ-IC
C-5B	F-22	Reaper

Note. AMP = Avionics Modernization Program; CMUP = Conventional Mission Upgrade Program; FSD = Full Scale Development; Inc 3.2B Mod = Increment 3.2B Modernization; JDAM = Joint Direct Attack Munition; JPATS = Joint Primary Aircraft Training System; RERP = Reliability Enhancement and Re-engining Program; RMP = Radar Modernization Program; TS = Training System.

In addition to standardizing to a BY, normalization for changes in quantity was also performed. Because estimated program quantities typically shift both upwards and downwards throughout the stages of a program's life cycle, these changes can produce illusions of cost growth. For this research, each CGF is standardized to account for any change in quantity. SARs list the quantities estimated and produced for each aircraft program. To standardize the CGF for quantity, a cost per aircraft is calculated at each program gateway. The method used is consistent with the methodology of Arena et al. (2006) that includes standardization through learning curves and first unit cost, which are derived from annual funding data provided in each program SAR. This is done prior to calculating the CGF at a particular gateway for a specific aircraft program.

Upon calculating the CGFs for all programs at each gateway, these responses (or dependent variable for this study) are then compared both longitudinally as well as cross-sectionally to assess for possible correlated factors (the independent variables). With respect to trends over time, this evaluation is accomplished in two ways. The first involves investigating how CGFs change from MS B (i.e., program initiation) to the various gateways. These evaluations are considered as Tier I and designated as relationships 1–5. As an example, relationship 1 compares the CGF at MS B (which would be 1 for all since the programs just started) to Gateway 1 (typically CDR), while relationship 5 assesses how the CGF has changed from MS B to Gateway 5 (typically FOC, or the last reported SAR for a completed acquisition program). Tier II evaluations then consider how CGFs change from gateway to gateway, incrementally, as a program progresses in time. Table 2 highlights the combinations of cost growth evaluated in this article. Each relationship is assigned a number for analysis naming purposes and is identified to the right of the relationship title.

TABLE 2. REVIEW RELATIONSHIP TIERS

Tier I
MS B - Gateway 1 (1)
MS B - Gateway 2 (2)
MS B - Gateway 3 (3)
MS B - Gateway 4 (4)
MS B - Gateway 5 (5)
Tier II
MS B - Gateway 1 (1)
Gateway 1 - Gateway 2 (6)
Gateway 2 - Gateway 3 (7)
Gateway 3 - Gateway 4 (8)
Gateway 4 - Gateway 5 (9)

Once cost growth is evaluated at each gateway point for both tiers, a color rating is assigned to the measured value to visually depict a program's cost growth going forward. This binning process puts all the numerical responses into categorical groups. These categorical groupings—*Green*, *Amber*, and *Red*—are used to indicate the status of a program. Subjectivity is involved in defining the conditions as to what constitutes Green, Amber, or Red status, and this subjectivity is a recognized limitation of the study. For this research, the following color ratings, based on the histogram of all the CGFs in the study's database, are associated with these respective CGFs:

Green:	If the analyzed CGF falls at 1.0 or below.
Amber:	If the CGF falls between the range $1.0 < \text{CGF} \le 1.10$.
Red:	If the CGF is > 1.10.

Green indicates low to no growth and identifies aircraft programs whose actual costs are at or below the estimated value at a respective gateway. Amber signifies the program has encountered some cost growth at or below 10%, while Red indicates a program has experienced more than 10% cost growth at a particular gateway.

In addition to tracking CGFs longitudinally, this study also investigates possible factors (independent variables) that might correlate to an aircraft program incurring a certain color rating at a respective gateway. These variables included the following: Air Force program, Navy program, modification (in lieu of a new start), prototype utilized, contractor, program size, percent of appropriated funds spent at each gateway, and percent program completion at each gateway. Because color ratings (the dependent variable) and some independent variables are dichotomous in nature, contingency table analysis was performed to determine association. The null hypothesis assumes these factors are independent; the alternative hypothesis assumes they are dependent. For continuous explanatory variables, customary ANOVA (Analysis of Variance) was conducted. The null hypothesis in this case is that the means of the continuous variables are equivalent across the color ratings, whereas the alternative hypothesis is that they vary.

To assess the significance of the contingency table, a p-value is calculated. Due to the small sample size of this research, Fisher's Exact Test was adopted. This particular statistical test is a permutation test in the sense that one first calculates all the possible combinations (tables) for the two factors being tested for association, given the observed row and column totals are the same. One then calculates the probability of each table. The total probabilities of the other tables are then summed, whose values are more extreme in the sense that these probabilities are less than or equal to the given table. If the total probability of such extreme tables is less than a given significance level (usually denoted α), then the data suggest the tested factors are statistically dependent.

No statistical assumptions are needed for a Fisher's Exact Test except for computational ability and time. [Note: The reader is directed to McDonald (2014, pp. 77–85) for more details regarding the use of Fisher's Exact Test or to Mehta and Patel (1983).] Due to the exploratory nature of this study, the analysis adopted a significance level of 0.10 for all tests to initially screen for potential predictors, and then lowered this to 0.01 to highlight those variables that do appear to be statistically significant. This approach is designed to mitigate spurious associations and is able to account for the high number of pairwise tests conducted (approximately 300).

To determine whether there is an increased or decreased likelihood that a particular categorical factor is correlated with a particular tested color rating (Green, Amber, or Red), a one-sided hypothesis test is applied. The null hypothesis assumes that the categorical variable does not affect the color rating (color level and factor are independent). If the p-value for a right-sided test is significant, then the predictor variable is associated with increasing the likelihood of the tested response color rating occurring. If the results were significant for the left-sided test, then the predictor variable is associated with decreasing the likelihood of the tested response color rating occurring.

Results

Of the 36 MDAP DoD aircraft programs in the database, 32 (or approximately 89%) had their four intermediate gateways (CDR, FF, DT&E, and IOC) occur in this customary progression. With respect to color rating of cost growth at each gateway in comparison to the estimate at MS B, which is denoted as Contract Award (CA), Table 3 highlights the color rating summary for Tier I relationships. [Note: A blank/white color reflects a program missing—either an estimate or quantity either at CA or at a particular gateway; five programs also were missing CA completely.] This table highlights the commonly studied and traditional view of looking at cost growth for programs and changing acquisition costs of each aircraft program from initiation to each gateway. As reflected both in the studies mentioned earlier in this article as well as Table 3, most of the programs experienced a Red rating of cost growth (exceeding 10%). Of the 26 programs that reported a SAR estimate by MS B in addition to actual costs at program completion, 18 (or approximately 69%) experienced cost growth exceeding 10% since program initiation. The mean and median cost growth percentages for these programs were 72% and 51%, respectively.

TABLE 3. TIER I COLOR RATING SUMMARY

Tier I	Relationship				
Program	1	2	3	4	5
A-10	🟧	🟥	🟥	🟥	🟥
AV-8B	🟩	🟧	🟧	🟩	🟧
B-1A	🟩	🟥	🟥		🟥
B-1B	🟩	🟩	🟩	🟩	🟧
B-1B CMUP Computer Upgrade	🟧	🟧	🟥	🟧	🟥
B-1B JDAM	🟩	🟩	🟩	🟩	🟩
B-2 RMP		🟩	🟥	🟧	🟥
C-130 AMP	🟥	🟥			
C-17A		🟥	🟥	🟥	🟥
C-5 AMP					
C-5 RERP	🟧	🟧	🟥	🟩	🟥
C-5B		🟩	🟩		🟩
C-27J					
E-2C		🟥	🟥	🟥	🟩
E-2D	🟧	🟧	🟥	🟥	🟥
E-6A	🟧	🟩	🟩	🟩	🟩
EA-18G	🟧	🟧	🟧	🟥	🟩
EF-111	🟧	🟥	🟥	🟥	🟥
FA-18EF	🟧	🟧	🟥	🟩	🟥
F-14A	🟩	🟩	🟩	🟥	🟥
F-14D					
F-15	🟧	🟧	🟥	🟥	🟥
F-16	🟩	🟩	🟩	🟥	🟧
F-22	🟥	🟥	🟥	🟩	🟥
F-35	🟥	🟥			
RQ-4 (GLOBAL HAWK)		🟥	🟥	🟥	🟥
MQ-4C	🟧	🟩	🟩		
P-8A	🟧	🟩	🟩	🟧	🟥
S-3A		🟧	🟩	🟧	🟥
T-6 (JPATS)	🟩	🟩	🟥	🟥	🟥
T-45TS	🟧	🟥	🟥	🟥	🟥
V-22 FSD	🟥	🟥	🟥	🟥	🟥
F-22 Inc 3.2B Mod	🟩	🟩			
KC-46	🟩	🟩			
MQ-IC					
Reaper (block 1)					

Note. AMP = Avionics Modernization Program; CMUP = Conventional Mission Upgrade Program; FSD = Full Scale Development; Inc 3.2B Mod = Increment 3.2B Modification; JDAM = Joint Direct Attack Munition; JPATS = Joint Primary Aircraft Training System; RERP = Reliability Enhancement and Re-engining Program; RMP = Radar Modernization Program; TS = Training System.

In comparison, Table 4 highlights Tier II relationships and takes a more in-depth view of cost growth by looking at each CGF longitudinally from gateway to gateway. [Note: Like Table 3, a blank/white color indicates missing estimate and/or quantity data for one or both of the involved gateways.] Instead of moving from CA forward, Tier II only looks at the relationship of a program's cost growth as it relates to the preceding gateway. Table 4 looks much different than Table 3. Unlike a lot of cost growth starting earlier and continuing throughout, Table 3 primarily has the cost growth rating of Red appearing at the FF to the DT&E gateway (relationship 7 in Table 2). Of the 31 programs reporting this change in cost growth from FF to DT&E, 15 (or 48%) had an additional increase of over 10% from whatever cost growth the program had already experienced leading into FF. The mean and median cost growth percentages were 50% and 36%, respectively. The next highest occurrence of a Red rating happened at relationship 6, which had only six instances of cost growth exceeding 10%. The overall takeaway for Table 4 is that if an aircraft program experiences cost growth exceeding 10%, then such growth typically appears between FF and the end of DT&E.

TABLE 4. TIER II COLOR RATING SUMMARY

Tier II	Relationship				
Program	1	6	7	8	9
A-10	🟨	🟥	🟨	🟩	🟨
AV-8B	🟩	🟨	🟩	🟩	🟥
B-1A	🟩	🟥	🟥		
B-1B	🟩	🟩	🟩	🟨	🟨
B-1B CMUP Computer Upgrade	🟨	🟨	🟥	🟨	🟩
B-1B JDAM	🟩	🟩	🟩	🟩	🟩
B-2 RMP			🟥	🟩	🟩
C-130 AMP	🟥	🟥			
C-17A			🟥	🟩	🟩
C-5 AMP				🟩	🟥
C-5 RERP	🟨	🟨	🟥	🟩	🟩
C-5B			🟩		
C-27J			🟥		
E-2C			🟩	🟩	🟩
E-2D	🟨	🟨	🟥	🟥	🟨
E-6A	🟨	🟩	🟨	🟨	🟨
EA-18G	🟨	🟨	🟩	🟩	🟩
EF-111	🟨	🟥	🟥	🟩	🟨
FA-18EF	🟨	🟨	🟥	🟨	🟨
F-14A	🟩	🟩	🟥	🟨	🟩
F-14D			🟥		
F-15	🟨	🟨	🟥	🟩	🟨
F-16	🟩	🟩	🟨	🟨	🟨
F-22	🟥	🟥	🟥	🟩	🟨
F-35	🟥	🟩			
RQ-4 (GLOBAL HAWK)			🟨	🟨	🟥
MQ-4C	🟨	🟩	🟩		
P-8A	🟨	🟩	🟩	🟨	🟥
S-3A			🟩	🟨	🟨
T-6 (JPATS)	🟩	🟩	🟥	🟨	🟨
T-45TS	🟨	🟨	🟥	🟩	🟥
V-22 FSD	🟥	🟥	🟩	🟩	🟨
F-22 Inc 3.2B Mod	🟩	🟩			
KC-46	🟩	🟩			
MQ-IC			🟩	🟩	
Reaper (block 1)			🟩	🟨	

Note. AMP = Avionics Modernization Program; CMUP = Conventional Mission Upgrade Program; FSD = Full Scale Development; Inc 3.2B Mod = Increment 3.2B Modification; JDAM = Joint Direct Attack Munition; JPATS = Joint Primary Aircraft Training System; RERP = Reliability Enhancement and Re-engining Program; RMP = Radar Modernization Program; TS = Training System.

To determine likely predictors of a color rating at a particular gateway, contingency table analysis as well as ANOVA, as described in the previous section, were used. With respect to Tier I (Table 2) contingency tables, these statistically test pairwise comparisons of various color ratings (Green, Amber, and Red) to the five Tier I relationships (1 through 5). For example, does Green1 associate with Green2? This answers the question: If an aircraft program going from CA to CDR (typically) experiences no cost growth, what is the likelihood the program will continue to experience no cost growth going from CDR to FF (again, typical gateway progression) such that CA to FF still shows no cost growth? Such questions allow a program manager to monitor cost growth throughout the program's acquisition cycle and its changing growth levels moving forward in time, but based on a comparison to the CA estimate.

The statistical analysis indicates that the most significant dependency in Tier I evaluations is Green3 (CA-DT&E) given Green2 (CA-FF). This statistical relationship is extremely significant, with the p-value for Fisher's Exact Test at 0.0001. It is also important to note that Green3 encompasses DT&E, which appears to be the area of most significant cost growth in a program as shown in Table 4. Therefore, this finding is valuable to management and decision makers in that if a program is still maintaining a Green color rating at FF, it is very likely that it will complete DT&E with no cost growth.

In a similar trend of matching color ratings, the second most significant dependency in Tier I evaluations is Red5 (commonly defined as CA-FOC) given Red3 (commonly CA-DT&E). Following the logic that a Green rating early on will aid long-term Green status, this relationship shows that hitting a Red rating early on will ultimately push the program into Red status for the long haul. This relationship is highly significant, with the p-value for Fisher's Exact Test at 0.0002. Again, there is that connection with DT&E. In terms of descriptive numbers, Table 3 highlights that of the 16 programs that were rated Red at DT&E, 15 of them (or approximately 94%) continued to rate Red at FOC. Other significant findings for Tier 1 include modification programs more likely to be Green5 (0.0017), Green3 less likely to be Red5 (0.0017), and Red2 likely to be Red3 (0.0052). No other findings were significant at the final 0.01 level of significance.

Moving to Tier II (Table 2, relationships 1, 6, 7, 8, and 9), similar ANOVA and contingency table analyses were conducted. Unlike Tier I, Tier II highlights the individual cross sections of the program life cycle, which provides different insights into cost growth. That is, Tier II investigates how the rate of cost growth is changing over time. Here, possible correlated variables are investigated to predict the rate of change. Utilizing the same methodology as in Tier I, each contingency table is evaluated using Fisher's Exact Test.

The first observation from the Tier II relationships was that none of the p-values from any of the hypotheses possessed the very low values found in the Tier I relationships. The lowest p-value corresponded to the test comparing Amber6 (Gateway 6, most commonly CDR-FF) given Amber1 (Gateway 1; most commonly CA-CDR) was 0.01 for the Fisher's Exact Test. The second observation found was that this was the first appearance of Amber in the analysis. In Tier I analysis, programs depicted Green or Red much more than Amber. For Tier II, this result suggests aircraft programs that display some cost growth going into CDR will more than likely continue to have some cost growth going from CDR to FF, but that this amount will not be more than 10%.

> **The presented results in this article suggest that aircraft programs that have no cost growth from CA to FF have a very good chance of reaching the end of DT&E without cost growth. In contrast, those programs that experience more than 10% cost growth by the end of DT&E have a very strong chance of remaining above 10% cost growth for the entirety of the program.**

Although it met the initial cut-off of 0.1, the statistical association of Green6 (relationship 6) to Green1 (relationship 1) did not meet the final p-value of 0.01, given its Fisher's Exact Test value of 0.0159. This possible finding is highlighted since it does support the association of similar colors like that of Amber6 to Amber1 (relationship 1). That is, an aircraft program that shows no cost growth from CA to CDR is more likely to still possess no cost growth when looking at just CDR to FF.

While not the most significant in terms of p-value (and caution dictates future research is still required), the evaluation of percent complete (the amount of money expended up to and including a particular gateway in relation to the total acquisition cost for the aircraft) revealed some unique trends that should be further explored. In the evaluation of percent complete in terms of dollars expended, the analysis shows that at Gateway 3 (most commonly DT&E), 45% expended appears to be a possible significant mark for multiple gateways. In both Tier I and Tier II evaluations, programs that expended more than 45% at Gateway 3 (commonly DT&E) were less likely to be Red at relationship 5 (p-value 0.0120), as well as Red at relationship 7 (p-value 0.0311). Additionally, this variable also showed some statistical significance at Green3 (p-value 0.0496), Green5 (p-value 0.0135), and Green7 (p-value 0.0676). That is, expending more than 45% increased the likelihood of a program having no cost growth for these comparisons. Similar to earlier results, DT&E appears to play a role here. No other explanatory variables considered met the final 0.01 level of significance either regarding contingency tables' analysis or ANOVA.

Conclusions

Using a database of 36 DoD MDAP aircraft programs (including modifications/upgrades), this article investigated the longitudinal behavior of CGFs as programs progressed from MS B through five life-cycle gateways. A color rating system (Green, Amber, and Red) was presented to descriptively characterize the movement of CGFs from both a cumulative approach in Tier I, to the subsectional relationships in Tier II. Studying the macro-trend of CGFs such as in the Tier I evaluation, the patterns appear to suggest that programs turn Red early (experienced cost growth exceeding 10%), and commonly stayed there for the remaining effort. These results, even though they are limited to the DoD aircraft portfolio, mirror and support historical analysis of cost growth in general for DoD acquisition programs.

Delving deeper into where said cost growth appears, the Tier II analysis, which incorporates an incremental longitudinal perspective, suggests that cost growth mainly appears between FF and the end of DT&E, and that this gateway ultimately dictates an aircraft program's overall fiscal health. Such findings corroborate the EVM work of Rosado (2007). Rosado used regression analysis to show that DT&E at the level 3 Work Breakdown Structure element is a significant predictor for overall program Estimate at Completion growth. That research suggests issues discovered during DT&E drive overall contract cost growth.

Additionally, the presented results in this article suggest that aircraft programs that have no cost growth from CA to FF have a very good chance of reaching the end of DT&E without cost growth. In contrast, those programs that experience more than 10% cost growth by the end of DT&E have a very strong chance of remaining above 10% cost growth for the entirety of the program. This tangentially supports what Smith, White, Ritschel, and Thal (2016) mentioned in that having a solid and funded test plan often mitigates future cost growth. Conversely, lacking or possessing limited funds devoted to testing often has detrimental effects on the entire cost of the program as issues are discovered too late to fiscally rectify.

Lastly, this article's findings regarding statistical associations of Green/Green and Red/Red (with respect to cost growth percentages) further highlights that a good predictor of future cost growth is previous performance. This finding is similar to the work of Christensen and Payne (1992) in that poor performance early in the program is very difficult to offset later in the program with performance better than planned. Their claim is that the cost growth stabilizes at around 20% completion, indicating a program is unlikely to dramatically improve its cost overrun position in the remaining effort. While Christensen and Payne's work was centered on individual contracts, the concept seems to hold true even when evaluating initial estimates to actuals at the program level over the entire life-cycle performance. Management can therefore deduce that the importance lies with making maximum effort to minimize cost growth early in a program's life cycle to ensure minimal cost growth in the future.

With respect to possible limitations, any data issues present within a particular SAR would naturally affect this study's database. Most notably, missing review dates (which generated some of the blank cells in Tables 3 and 4) impacted the ability to analyze cost growth performance from a longitudinal perspective for every aircraft program in the database. Every missing date impacts at least two data points. For Tier I, five programs were missing

CA information, impacting everything else downstream from that. This effectively removed approximately 14% of the data for Tier I analysis. For Tier II, 45 out of 180 cells (or 25%) were missing, but some of these (at least nine cells) are due to programs not yet completed (e.g., F-35 or the KC-46).

Overall, the study's database attempted to capture all relatively modern DoD MDAP aircraft programs. Of the population of 36 programs, the data captured approximately 80% of the gateway information. Of that information analyzed, the statistical findings appear to be supported by other works cited in this article. Additionally, keeping the strict final criteria of a 0.01 level of significance militated against spurious relationships from appearing given the number of inferential tests conducted. Lastly, nonsignificant findings from any of the ANOVA tests were not unexpected given the large variance of CGFs from aircraft program to aircraft program as noted by Kozlak et al. (2017).

Overall, the application of a categorical color rating of cost growth in conjunction with a longitudinal perspective from gateway to gateway revealed and confirmed prior findings. Past performance does appear to be a predictor of future success, and DoD MDAP aircraft programs appear, on average, to have cost growth exceeding 10% of the development estimate from MS B. In these findings, DT&E appears to play a significant role in this occurrence. This doesn't necessarily imply that a DT&E problem exists with aircraft programs. Smith et al. (2016) suggest testing early and often to address issues such as flawed designs, inadequate incorporation of requirements, or architectural design so as to offset future cost growth. O'Neil (2011) echoes those concerns when noting that defects in the original concept are a common cause of cost growth. Ignoring these authors' concerns might result in cost growth between FF and the end of DT&E, as highlighted by this article's analyses with respect to DoD MDAP aircraft programs.

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-AF). Santa Monica, CA: RAND.

Bolten, J. G., Leonard, R. S., Arena, M. V., Younassi, O., & Sollinger, J. M. (2008). *Sources of weapon system cost growth: Analysis of 35 major defense acquisition programs* (Report No. MG-670-AF). Santa Monica, CA: RAND.

Cancian, M. F. (2010). Cost growth: perception and reality. *Defense Acquisition Review Journal, 17*(3), 389–404.

Christensen, D., & Payne, K. (1992). Cost performance index stability: Fact or fiction? *Journal of Parametrics, 10*, 27–40.

Davis, D., & Anton, P. S. (2016). *Annual growth of contract costs for major programs in development and early production* (Technical Report). Retrieved from Defense Technical Information Center. (Accession No. AD1006106)

Department of Defense. (2017). *Operation of the defense acquisition system* (DoDI 5000.02). Washington, DC: Office of the Under Secretary of Defense for Acquisition, Technology and Logistics.

Drezner, J. A., Jarvaise, J. M., Hess, R. W., Hough, P. G., & Norton, D. (1993). *An analysis of weapon system cost growth* (Report No. MR-291-AF). Santa Monica, CA: RAND.

Jimenez, C. A., White, E. D., Brown, G. E., Ritschel, J. D., Lucas, B. M., & Seibel, M. J. (2016). Using Pre-Milestone B data to predict schedule duration for defense acquisition programs. *Journal of Cost Analysis and Parametrics, 9*(2), 112–126. Retrieved from https://doi.org/10.1080/1941658X.2016.1201024

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

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

McDonald, J. H. (2014). *Handbook of biological statistics* (3rd ed.). Baltimore, MD: Sparky House Publishing.

Mehta, C. R., & Patel, N. R. (1983). A network algorithm for performing Fisher's Exact Test in r × c contingency tables. *Journal of the American Statistical Association, 78*(382), 427–434.

O'Neil, W. D. (2011). Cost growth in major defense acquisition: Is there a problem? Is there a solution? *Defense Acquisition Research Journal, 18*(3), 277–294.

Petter, J. L., Ritschel, J. D., & White, E. D. (2015). Stability properties in Department of Defense contracts: Answering the controversy. *Journal of Public Procurement, 15*(3), 341–364.

Rosado, W. R. (2007). *Comparison of development test and evaluation and overall program estimate at completion* (Master's thesis). Retrieved from Defense Technical Information Center. (Accession No. A549645)

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

Smith, N. C., White, E. D., Ritschel, J. D., & Thal, A. E., Jr. (2016). Counteracting harmful incentives in DoD acquisition through test and evaluation oversight. *The ITEA Journal of Test and Evaluation, 37*(3), 218–226.

Author Biographies

Ms. Cory N. D'Amico is a cost analyst at the Air Force Life Cycle Management Center in the Dayton, Ohio, area. She holds a BS in Accounting from the University of Kentucky, an MBA from the University of Dayton, and an MS in Cost Analysis from the Air Force Institute of Technology (AFIT).

(E-mail address: Cory.D'Amico@afit.edu)

Dr. Edward D. White is a professor of statistics in the Department of Mathematics and Statistics, AFIT. He holds a Master of Applied Statistics from Ohio State University and a PhD in Statistics from Texas A&M University. Dr. White's primary research interests include statistical modeling, simulation, and data analytics.

(E-mail address: Edward.White@afit.edu)

Lt Col Jonathan D. Ritschel, USAF, is an assistant professor of cost analysis in the Department of Systems Engineering and Management, AFIT. He holds a BBA in Accountancy from the University of Notre Dame, an MS in Cost Analysis from AFIT, and a PhD in Economics from George Mason University. Dr. Ritschel's research interests include public choice, the effects of acquisition reforms on cost growth in DoD weapon systems, research and development cost estimation, and economic institutional analysis.

(E-mail address: Jonathan.Ritschel@afit.edu)

Capt Scott J. Kozlak, USAF, is a cost analyst at the Air Force Cost Analysis Agency in the Washington, DC area. Capt Kozlak holds a BS in Management from the United States Air Force Academy and an MS in Cost Analysis from AFIT.

(E-mail address: Scott.J.Kozlak.mil@mail.mil)