3-23-2017

A Longitudinal Study and Color Rating System of Acquisition Cost Growth

Cory N. D'Amico

A Longitudinal Study and Color Rating System of Acquisition Cost Growth

THESIS

March 2017

Cory N. D'Amico

AFIT-ENC-MS-17-M-181

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENC-MS-17-M-181

A Longitudinal Study and Color Rating System of Acquisition Cost Growth

THESIS

Presented to the Faculty

Department of Mathematics and Statistics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Cost Analysis

Cory N. D’Amico

March 2017

AFIT-ENC-MS-17-M-181

A Longitudinal Study and Color Rating System of Acquisition Cost Growth

Cory N. D'Amico

Committee Membership:

Edward D. White, Ph.D
Chair

Lt Col Jonathan D. Ritschel, Ph.D
Member

Capt Scott J. Kozlak
Member

AFIT-ENC-MS-17-M-181

Abstract

For decades, cost growth studies have been plentiful surrounding Department of Defense acquisitions. Many different angles have been looked at to try and discover how to better estimate cost, what causes cost growth, and how to mitigate it. This research addresses this through examining cost growth from a longitudinal perspective, evaluating cost growth factors at major program reviews, and assessing the cost growth by applying color rating metrics. The results of this analysis show that breaking cost growth into longitudinal segments of a programs lifecycle allows the true behavior of cost growth to be seen, when it can often be masked in the traditional approach of evaluating lifecycle cost growth. Additionally, when applying the proposed color rating system to cost growth factors, significant variables are found to have dependencies with cost growth factor color ratings. Significant relationships shown in the results were most commonly like-color predictor and response variables. Additionally DT&E is shown as a flag for high cost growth issues during a program lifecycle.

Acknowledgments

I cannot start this with anything except a sincere thank you to my husband, who has been my rock, my reality-check, and the sane to my crazy. I know you said I was insane when I set out to do this, and it turns out you were right. Thanks for seeing me through it regardless. We didn't know throughout this process that we would welcome our third child, but as luck would have it we complicated an already intense journey! Also, to those three beautiful children, you will never know how much I love you. This is for you. Thank you for being my daily reminder of what life is all about.

The biggest of thanks goes out to my research advisor, Dr. Edward White, for his guidance, support, and above all else, patience. While I embraced all of the profound statistical knowledge he bestowed on me, it was in our random, mystery-of-life-answering discussions that I truly gained appreciation. Thank you for being my sounding board regardless of the topic at hand.

Additionally, I wish to express my gratitude towards Lt Col Jonathan Ritschel and Capt Scott Kozlak for their time on my committee, as well as my fellow program cohorts for their humor and true friendship. Thank you all for making this an experience that I will never forget.

Cory N. D'Amico

Table of Contents

Abstract iv

Acknowledgments v

Table of Contents vi

List of Figures ix

List of Tables x

I. Introduction 1

General Issue 1

Problem Statement 2

Research Objectives and Scope 4

Research Questions 5

Methodology 5

Assumptions/Limitations 6

Summary 7

II. Literature Review 8

Chapter Overview 8

Cost Growth 8

Previous Research Findings – Foundational Analysis 9

Previous Research Findings – Longitudinal Studies 13

Summary 16

III. Methodology 17

Chapter Overview 17

Data Collection 17

Data Summary 18

Limitations 19

Data Set and Predictor Variables 21

Estimating Cost Growth 24

Data Standardization 25

Cost Growth 26

Color Coding Cost Growth 28

Two-Way Contingency Tables 30

Fisher's Exact Test 32

Chapter Summary 34

IV. Results and Analysis 35

Chapter Overview 35

Color Rating Summary 35

Tier I Results 38

Tier II Results 47

Chapter Summary 55

V. Conclusions and Recommendations 56

Chapter Overview 56

Research Questions Answered 56

1 –How does cost growth behave differently between a segmented longitudinal perspective and a traditional lifecycle perspective? 56

2 – Are there differences in cost growth trends from a cross-sectional perspective, than that of the overall program? 57

Findings 57

Limitations 60

Recommendations for Future Research 61

Chapter Summary 62
Appendix A 63
Appendix B 68
Appendix C 82
Bibliography 87

List of Figures

Page

Figure 1: DoD Budget by Funding Account.. 2

Figure 2: Distinction of Cost Growth Program Size..23

Figure 3: Distinction of Cost Growth Program Size..29

Figure 4: Contingency Table Color Chart for Green3 Given Green2 39

Figure 5: Contingency Table Color Chart for Red5 Given Red3 41

Figure 6: Contingency Table Color Chart for Red5 Given Green3.................................42

Figure 7: Contingency Table Color Chart for Green5 Given Modification 44

Figure 8: Contingency Table Color Chart for Red3 Given Red2 45

Figure 9: Contingency Table Color Chart for Amber6 Given Amber.............................48

Figure 10: Contingency Table Color Chart for Green6 Given Green 49

Figure 11: Contingency Table Color Chart for Amber13 Given Green6.........................51

Figure 12: Contingency Table Color Chart for Red6 Given Red 52

Figure 13: Contingency Table Color Chart for Amber10 Given Small...........................53

Figure 14: Tier I Predictor Variable Breakout...58

Figure 15: Tier II Predictor Variable Breakout..59

Figure 16: Tier I Response Variable Breakout 59

Figure 17: Tier II Response Variable Breakout...60

List of Tables

Page

Table 1: Aircraft Programs Selected for Analysis 19

Table 2: Review Relationship Tiers 27

Table 3 Predictor and Response Variable Combinations 31

Table 4: Sample Contingency Table 32

Table 5: Tier I Color Rating Summary 36

Table 6: Tier II Color Rating Summary 37

Table 7: Fisher's Exact Test Results for Green3 Given Green2 39

Table 8: Contingency Table for Green3 Given Green2 40

Table 9: Fisher's Exact Test Results for Red5 Given Red3 41

Table 10: Contingency Table for Red5 Given Red3 42

Table 11: Fisher's Exact Test Results for Red5 Given Green3 43

Table 12: Contingency Table for Red5 Given Green3 43

Table 13: Fisher's Exact Test Results for Red5 Given Green3 44

Table 14: Contingency Table for Green5 Given Modification 44

Table 15: Fisher's Exact Test Results for Red3 Given Red2 45

Table 16: Contingency Table for Red3 Given Red2 46

Table 17: Other Significant Tier I Results 47

Table 18: Fisher's Exact Test Results for Amber6 Given Amber 48

Table 19: Contingency Table for Amber6 Given Amber 49

Table 20: Fisher's Exact Test Results for Green6 Given Green 50

Table 21: Contingency Table for Green6 Given Green 50

Table 22: Fisher's Exact Test Results for Amber13 Given Green6 51

Table 23: Contingency Table for Amber13 Given Green6 52

Table 24: Fisher's Exact Test Results for Red6 Given Red 53

Table 25: Contingency Table for Red6 Given Red 53

Table 26: Fisher's Exact Test Results for Amber10 Given Small 54

Table 27: Contingency Table for Red6 Given Red 54

Table 28: Other Significant Tier II Results 55

A LONGITUDINAL STUDY AND COLOR RATING SYSTEM OF ACQUISITION COST GROWTH

I. Introduction

General Issue

The Department of Defense (DoD) has been facing steady pressure to reduce spending and program waste, and the near future will likely maintain this downward pressure. Since 2010, the DoD budget has demonstrated a decreasing trend, as the administration attempts to regain control of government spending. Figure 1, taken from the 2016 Annual Report on Performance of the Defense Acquisition System, shows the DoD funding by budget accounts through the 2017 President's Budget (PB). This highlights the steep drop in funding the DoD is facing in today's acquisition environment, which means increasing pressure on current programs, forcing them to perform efficiently and effectively with the funds that they manage.

While cost growth has historically been a burden on DoD funds, the increased scrutiny and oversight on budget and performance increases the necessity of more reliable cost estimates and tighter management of program cost growth.

DoD budgets are based on program inputs, and resources are planned and allocated accordingly. When these costs become unreliable due to cost growth, future budgets must adjust for the increased cost, constraining the flexibility of decision makers. Ultimately, as programs struggle to compete for limited resources to cover growing costs, the warfighter suffers from lost capability that cannot be funded, or must delay until further resources are available. Therefore, enhancing the ability of programs to predict, identify, and manage cost growth can allow for a more accurate management of resources and, in turn, likely reduce the need for unidentified resources in the future. This presents the DoD with the task of grasping a firm understanding of how cost growth behaves throughout a programs lifecycle, how to mitigate it when possible, and how to manage it when present.

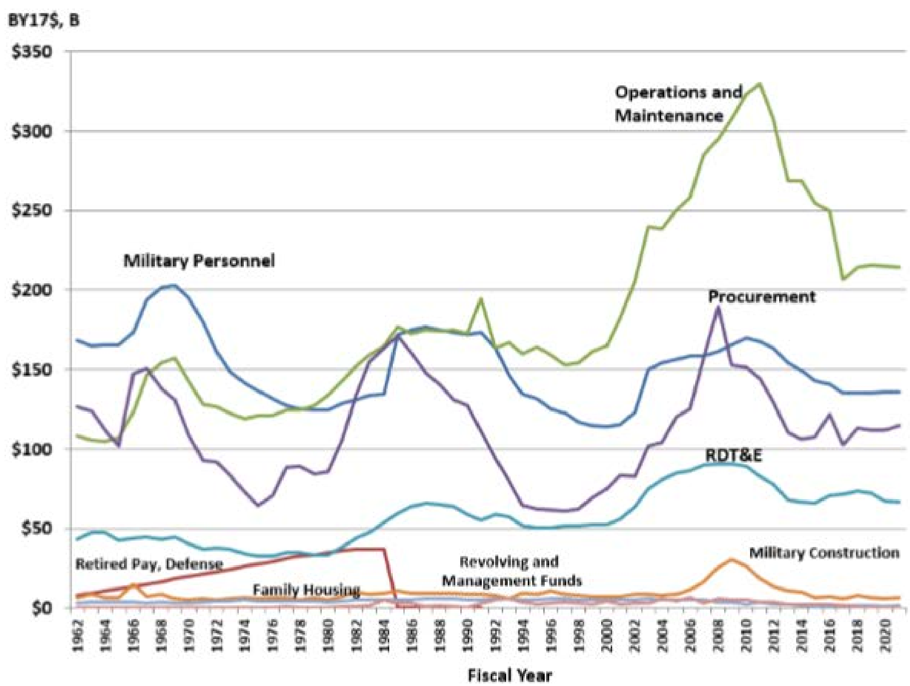


Figure 1: DoD Budget by Funding Account

Problem Statement

Historical performance of DoD programs shows that initial estimates are often over-optimistic compared to final program costs. In most circumstances, the initial estimate is constructed and baselined at Milestone B, when programs first become a program of record. The difficulty in estimating here is the multitude of changes a program will undergo over the course of its post-Milestone B lifecycle. It has been called to attention that even after accounting for uncertainty in their cost estimates, the official estimate is often looked at as a point estimate with little to no consideration for a change in trajectory, and no indication of uncertainty in the eventual budget process (DeNeve et al. 2015). This opens the flood gates for cost growth when comparing actuals back to this initial baselined estimate. So while the cost estimator is often burdened with the responsibility of accurate cost estimating, they are often an easy target to blame when cost growth goes rogue.

The GAO report in 2009 made waves when they reported the change to total acquisition cost from first estimate is 26%. This number is staggering, when you realize the planned commitments were at $1.6 trillion (GAO, 2009). However, cost growth studies have shown the figure to realistically be upwards of 50% (Arena et al., 2006). IDA and RAND both compiled studies in attempt to capture the full extent of cost growth over a program's lifecycle. These studies both showed more realistic cost growth estimates were 45 and 46% respectively (Cancian, 2010). While the cost estimating career field has expanded its role in recent years, the challenges that come with the nature of the unknown make accurate cost estimating a difficult task. If the strength of initial program estimating accuracy cannot necessarily be improved, monitoring the patterns of cost growth becomes increasingly important.

Common issues that typically lead to cost growth are unclear or undefined requirements, schedule delays, and technological complexity. Since it is unrealistic to assume all cost growth can be prevented and eliminated, helping management accurately forecast how cost growth affects the overall cost can be a valuable tool in early identification and risk reduction. Porter et al. (2009) describe one of the primary causes of cost growth to be weakness in management visibility, direction and oversight. Part of this remains due to the limited tenure of management personnel, and the nonstop rotation of decision makers. Often times decision makers are focused on the here and now, instead of the program completion that may be 10-15 years down the road. When problems are outside of the scope of their term, management can "kick the can" for a while. However, if decision makers have cost growth data that can be applied longitudinally throughout the entire lifecycle of a program, it becomes a more relevant part in the acquisition environment. Instead of accepting a "rule of thumb" figure for overall cost growth, using this research to help segment cost growth into short-term sections, makes the management oversight firsthand and something that decision makers can actively engage in real-time cost growth management.

As subsequent chapters explain, this research identifies the pattern of cost growth over the life of a program. In addition to identifying trends in cost growth, this research develops statistical guidelines of how cost growth may react given previous performance at specified milestones, enabling managers to more accurately predict future program cost growth levels.

Research Objectives and Scope

The objective of this research is to provide a tool for program managers and decision makers to determine where cost growth levels will project, given cost growth incurred during previous program milestones achieved. This research culminates in a color coding system to help identify cost growth risk, and identifies significant predictor variables for each color rating identifier.

For the scope of this analysis, 36 Aircraft programs across DoD are examined and their cost growth performance is evaluated across 4 critical evaluation points between the, initial Development Estimate (DE) at Milestone B and the final or most Current Estimate

(CE) available. Cost growth is measured by the cost growth factor (CGF), better defined as the ratio of actual cost to the estimated costs. When the CGF calculated is greater than 1.0, it indicates cost growth from the original estimate, an undesirable direction for the DoD program. Each program is evaluated at these specified program review points, and the CGF is documented accordingly. These trends are then mapped to a color rating system, which is defined in Chapter III.

Research Questions

This research is focused on addressing to major research questions.

1. How does cost growth behave differently between a segmented longitudinal perspective and a traditional lifecycle perspective?
2. What significant predictor variables forecast a given program review cost growth position?

To address these questions, an analysis of trends and traits of programs is conducted to find associated cost growth, to help identify causes associated with cost growth, and to project their impact on future performance. This research identifies targets, in which given a current state of cost growth, will provide a benchmark of ranges to predict the future state of a program’s cost growth. This tool ultimately helps leadership identify where program cost growth will lie at the next major evaluation point, given current program cost growth levels incurred to date.

Methodology

Selected Acquisition Reports (SARs) outline a program’s current state and report current funding estimates and actual expenses incurred for Major Defense Acquisition

Programs (MDAPs). This data is reported annually and provides insights into program performance at key milestones. For this study, SARs are used to reference program cost estimates and incurred expenses at selected dates. The SAR data used was collected directly from Defense Acquisition Management Information Retrieval (DAMIR), and also referenced from a database compiled by RAND, based on SAR data.

This research uses the data collected from the SARs to conduct an analysis on cost growth trends. First, a general analysis of the descriptive statistics is profiled on the data. This lays the groundwork for the analysis and helps to define the scope of the profiles. Next, the CGF is calculated, and a stoplight color rating system is developed to categorize cost growth factors in to easily identifiable risk categories that management can reference for top-level reviews. Once this color rating system is developed, it is applied to each CGF calculated. Finally, the color ratings are analyzed in relationship to each variable identified and tested for significance.

Assumptions/Limitations

Potential limitations include differing definitions of key milestones between programs assessed. In particular, IOC can be measured differently by each independent program. In this case, the data as reported in the SAR was used without altering the dates to standardize definitions. Other limitations include missing data within SARs such as missing program view dates, limited availability of SARS prior to the implementation of DAMIR in 1997, and a small sample size of aircraft programs meeting the appropriate criteria for this research.

Summary

Knowing cost growth is a certainty of Defense programs, the ability to capture the future cost growth into percentage brackets related to current performance levels may provide program leadership the opportunity to adequately prepare for future performance milestones. Instead of looking at long-term cost growth on a total program perspective, this research lays the foundation for new cost growth calculations, taking a more short-term decision making approach into consideration. Because management is not always concerned with long-term progress, but rather the near term and close range decisions at hand, having cost growth broken down into more manageable reviews should help aid decision makers.

The remaining chapters that follow discuss the Literature Review, Methodology, Analysis, Results, and Conclusion/Discussion. The literature review in Chapter II explores how cost growth is defined for this study, previous studies on cost growth relevant to this research, followed by a model on projected cost growth brackets. Chapter III walks through the creation of a cost database and the subsequent tests conducted in the analysis. Key analysis tests conducted include Contingency Table analysis and Fisher’s Exact Test for significance. These combine to make the foundation for Chapter IV, where the results of the analysis conducted are presented. Finally, Chapter V provides the conclusions of this research, as well as recommended future areas of study.

II. Literature Review

Chapter Overview

This chapter provides a foundation for this research. First, the method DoD uses to define cost growth is discussed and historical performance is evaluated. Previous studies on DoD cost growth are reviewed for foundational knowledge on cost growth. Additionally, the literature review looks at statistical techniques used to evaluate cost growth. Lastly, this review looks at studies with a longitudinal approach to cost growth in order to identify how cost growth can be segmented, and identify potential areas of information in need of further study.

Cost Growth

With budgets being stressed by increasing requirements and demands, programs have been forced to evaluate funding stressors. Therefore, there is increasing pressure to identify and reduce cost growth both at the contract and program level. Oftentimes, programs may not be able to reduce the cost growth. However, it can still be useful for a program to be able to accurately project where the cost growth is headed in the future and allow for decision makers to adequately prepare in advance of the next milestone. This research evaluates the probability of a program to fall into a certain cost growth bracket given its current performance cost growth level.

Cost growth can be interpreted or perceived in a variety of ways, particularly depending on the intent of the analysis. The simple answer is that cost growth, in its essence, is when something costs more than expected. However, cost growth can be a combination of a multitude of factors. The Select Acquisitions Report (SAR) is the foundational document used to assess cost growth on DoD programs, and is the official reporting document used by the General Accounting Office (GAO). The SAR divides cost growth in to seven main factors; Economic, Quantity, Schedule, Engineering, Estimating, Other, and Support (Cancian, 2010).

Economic cost growth is most easy to distinguish, as this is cost growth caused by inflation. Inflation is the general increase in prices of goods in the economy, making it external to the control of the program. Therefore, for most analyses of cost growth, it is removed. While it may appear that the cost of a program is getting more expensive throughout the life of the program, adjusting to base year dollars will remove the effect of inflationary increases, allowing a more accurate depiction of true program cost growth.

The cost growth caused by changes in quantity can also distort the perception of program costs rising. Although overall program costs may appear to increase, if quantities are increased, the true cost per item may not. While some may argue it is still cost growth, given the program is costing more than originally projected, for this research changes in quantity are accounted for to eliminate artificial cost growth.

While inflation and quantity are relatively easy to separate out for the purposes of analyses, the remaining factors are inherently more subjective. These remaining factors are all considered cost growth internal to the program and are the essence of data evaluated in this research. It is these decisions between Milestone B and the current estimate that create the pattern of longitudinal cost growth that is evaluated for inherent patterns in performance.

Previous Research Findings – Foundational Analysis

Once cost growth is defined, it can then be relatively measured. Many studies have looked at the cost growth issue within the DoD due to its pervasive nature. However, different approaches have been used to assess and predict cost growth trends. This section of the literature review highlights cost growth research that used as a foundation for this research.

Drezner et al. (1993) studied CGFs of weapon systems in order to quantify the magnitude of cost growth, and identify factors that affect it. Drezner et al. (1993) found that inflation and quantity have the greatest effect on reported cost growth outcomes, but could be accounted for. After removing the effects of both factors, program size and maturity are the two factors that have the greatest effect on total program cost growth; smaller programs tend to incur higher cost growth, while on average, maturity adds 2.2% above inflation per year (Drezner, 1993). Drezner et al. (1993) validates that historic trends shows cost growth averages about 20%, and has not seen significant fluctuations from that over the course of decades.

Drezner et al. (1993) assessed factors that affect cost growth, discovering that programs with prototyping incurred higher cost growth. This is validated in our research findings in Chapter IV. One explanation for this may be that prototyping tends to be present in programs with higher technical uncertainty, which are inherently subject to higher cost growth (Drezner, 1993). This research provided a foundation for the need to adjust for inflation and quantity changes, as well as account for prototyping as a variable.

Sipple et al. (2004) compiled a documentation of significant cost growth literature, to highlight the variety of statistical methods applied to cost growth in DoD acquisition. The analysis included a study on seven significant cost growth studies, compiled to summarize significant findings for cost professionals. This research served as a foundation of historical cost growth overviews.

Arena et al. (2006) presented a technical report that evaluated statistically significant drivers of cost growth for completed programs with similar complexities to those of the U.S. Air Force. Using data collected from the SAR, the CGFs for 68 completed programs are used in analysis. The data is segmented by funding category, milestone, and commodity type to accommodate different approaches in estimating.

Arena et al. (2006) observed there is consistent underestimation of cost, a point that is reiterated in later research. However, they continue on to find statistically significant drivers of cost growth. Major findings include that programs with longer duration have greater cost growth, that there are no statistically significant differences between the services, and little evidence to support that cost growth has improved over time (Arena, 2006). Additionally, significant cost growth occurs at MS II and MS III. Arena et al. (2006) is significant to this research because it presents key drivers of cost growth that help form the selected variables tested, as well as provides a rationale for using SARs in data collection.

Continuing on in research to identify cost growth drivers, Bolton et al. (2008) also produced a paper for the RAND Corporation exploring causally oriented variance categories. Bolton et al. (2008) defined four different categories of cost variance oriented towards the causes of cost growth: errors in estimation and planning, decisions by the government, financial matters, and miscellaneous sources, and reclassified the variances in the SARs into said categories. This research then evaluated cost growth for development, procurement, and total program, by identifying the significant cost drivers in each. According to this study, total cost growth is dominated by decisions, accounting for more than two-thirds of the growth. This includes quantity changes, requirements changes, and schedule changes.

Bolton et al. (2008) provided the foundation on the importance of quantity changes. Most cost growth studies standardize for quantity changes, which eliminates their impact. However, this study keeps quantity changes inclusive, stating it is still a factor of cost growth and should therefore be measured. Because of this inclusion, Bolton et al. (2008) were able to identify how truly significant the impact of quantity changes is in terms of growth. In fact, more than half of the average procurement cost growth is due to quantity changes (Bolton, 2008).

Because of these findings, this research heavily considered leaving the quantity change effect as part of the analysis; although the data was ultimately standardized remove those effects.

Deneve et al. (2015) focuses on the issues with DoD cost estimating instead of post-estimation drivers, posing the question that perhaps the problem is not the accuracy of the cost estimate, but that the estimates are estimating the wrong thing. Using an interesting analogy of cost estimating to hurricane forecasting, Deneve et al. (2015) describe how cost estimates often become the "sticker price," and lose the risk and uncertainty analysis in the translation to the budget process. They follow to create a macro-stochastic estimation model by identifying four categorical variables with strong relationships to CGFs: program type, iteration, funding years, and number of services. Deneve et al. (2015) conclude these groupings help predict the total cost from the baseline estimate. This research provided the foundation of issues with cost estimation in the DoD, an equal culprit to the cost growth problem.

Kozlak (2016) serves as a key foundation in this research. Multiple aspects of Kozlak's work were carried forward for additional study, including the identification of key aircraft program reviews and milestones, the objective to calculate CGFs at these reviews, and identifying the significant cost drivers that pertain to Aircraft programs.

Kozlak (2016) evaluated development, procurement, and total cost growth, defining median cost growth percentage against median program percent completion. For example, at IOC for procurement, the median percent of program completion is 48% and the median percent of total cost growth is 91%. In evaluating each review point as well as percent complete, Kozlak (2016) projected that at a median of 6.5 years after Milestone B, a program sustains approximately 91% of the total program cost growth. Aside from evaluating the performance of cost growth along a programs review cycle, Kozlak (2016) used logistic regression to identify individual (x) variables that are predictive of cost growth (y). The regression results showed Bomber, Prototype, and Weapon type to be the most predictive variables of cost growth.

The literature reviewed form the foundation of cost growth research, define significant cost growth drivers, and depict overall trends of cost growth within different segments of the DoD. This research combines the above efforts with a slightly varied approach of looking at cost growth from a longitudinal perspective. That is, how does cost growth behavior change as time changes? This research looks to dissect the life of a program into multiple evaluation points, each with its own evaluation factors. This approach can give leadership increased opportunity to evaluate, manage, and plan for cost growth on a more reliable timeline. While limited, some research has been conducted using this longitudinal approach, and is presented in the next section.

Previous Research Findings – Longitudinal Studies

This research takes the approach of evaluating cost growth on a longitudinal perspective, looking at each individual program and its cost growth over time from Milestone B to the Final or Current Estimate if still ongoing, which is defined in the SAR. While the research on cost growth within the DoD is vast, those that evaluate from a longitudinal perspective are limited. The remainder of this section summarizes the key literature on previous longitudinal cost growth studies.

On one end of the longitudinal approach spectrum, research has looked at Earned Value Management (EVM) with the intent of locating stabilization of cost growth. Since the early 1990's, the EVM community has adopted what is known as the "stability rule," from widely known Christensen and Payne research (1992). Christensen's findings show that the Cost Performance Index (CPI) commonly stabilizes when the contract is at 20% completion

(Christensen and Heise, 1993). This concept was explored more recently, applying Christensen's concepts of "stability" to more recent DoD data (Petter, 2014).

Petter (2014) addresses this "stability rule," indicating that recent research may provide contradictory evidence that the CPI does not in fact stabilize at 20% completion. Petter (2014) highlights the vague definition of "stability," and further summarizes it in to three broad categories: range definition, absolute interval definition, and relative interval definition. These three definitions become the basis for Petter's analysis. Using a variance analysis, Christensen's "stability rule" was both supported and contradicted, depending on the definition of stability used. With the range definition of stability, the 20% complete "stability rule" is supported, whereas either interval stability definitions did not (Petter, 2014). Petter (2014) then continues the research continues to break down a contact's life-cycle phase: Production or Development. Here the longitudinal aspect of the research is shows, though applied to EVM and a contract basis, making Petter (2014) both a foundational study, an updated approach to the "stability rule," and a take on the longitudinal approach to cost growth.

While EVM is a valuable tool, and understanding the CPI and its predictive measures on final cost, EVM can be restrictive in that MDAPS are often split into a variety of contracts throughout the life of the program. Therefore, the CPI limits measuring and predicting program cost growth to only a small portion of the overall cost growth. This can be helpful to short-term management of cost growth, but restricts long-term strategic thinking.

The other spectrum of the longitudinal approach encompasses how cost growth trends over the course of history. Davis et al. (2016) performed one of the most recent evaluations of cost growth within DoD MDAP programs. However, their approach to a longitudinal evaluation was looking at cost growth rates by year, historically, across the DoD over the last three decades.

Davis et al. (2016) used a 5-year moving average of the annual changes to assess patterns of cost growth in attempt to model growth.   David et al. (2016) reveal that the pattern of historical cost growth partially follows the defense budget, providing a new look at what may drive cost growth. Other results show structural changes and legislation contribute to shifts in growth. This approach of longitudinal evaluation shows how the acquisition process, policy reform, and the behavior of the defense acquisition system play a role in cost growth trends over time. While individual programs were assessed, they were not evaluated as cost growth independently across its own lifespan.

Our research has shown that the longitudinal approach has looked both from a focused, contract specific EVM perspective, as well as a timeline driven, DoD-wide overview. EVM is focused on a contract basis, as EVM only looks at and measures performance on a contract level, which can mask deeper issues on a total program level. Many DoD programs are comprised of multiple contracts during its lifecycle. While EVM is a valuable tool in tracking performance, the complexity of DoD programs and contract structure necessitate the need for research on how a program performs over the duration of all contracts. In contrast, the time-phased approach looks at how cost growth has changed within the DoD over time and not necessarily on an individual program basis.

While this "longitudinal" concept has been explored, it leaves gaps in looking at the concept specifically applied to an individual program. This facilitates the need for exploration of the longitudinal approach and evaluation of cost growth trends throughout the specified life of evaluation. For this research, the question of how cost growth behaves from a longitudinal perspective across the life of the program is examined; more specifically, how is current cost growth a predictor of where cost growth will perform at the next measured milestone and what is the relationship of cost growth levels between key performance milestones?

Summary

The literature review presented information on the sources of cost growth, standard methods for estimating that growth, research on when cost growth occurs in weapon systems, and potential variables to consider in our research. Chapter III, Methodology, discusses how we approach addressing the goals of our research.

III. Methodology

Chapter Overview

The purpose of Chapter 3 is to present the details of data used for analysis in this research, as well as the methods used for evaluation. First, the sources used for data collection are discussed. Next, the compilation and evaluation of the data is explained, followed by the method used for data standardization. Finally, how cost growth is identified and analyzed at each program review is discussed.

Data Collection

In order to analyze program cost growth factors, reliable data that contains program review information must be obtained. In most previous studies pertaining to cost growth, the SAR is used as a credible source of data, laying a foundation for the SAR as a reputable source. The SAR is one of the best resources of data for acquisition programs, particularly for the data points collected for evaluation within this research.

For this research, SAR data is utilized for program estimates, key dates, and program relevant information. The primary resource used for SARs is the Defense Acquisition Management Information Retrieval (DAMIR), which identifies various data sources that the Acquisition community uses to manage Major Defense Acquisition Programs (MDAP) and Major Automated Information Systems (MAIS) programs and is the authoritative source for Selected Acquisition Reports (SAR), SAR Baseline, Acquisition Program Baselines (APB), MAIS Annual Reports (MAR), MAIS Original Estimates (MAIS OE), and Assessments. (DAMIR webpage).

Additionally, RAND Corporation has worked extensively with historical SARs, compiling the annual reports into a SAR database. This database, which is built electronically using separate Microsoft® Excel sheets per program, proved valuable for this research, particularly for years that older SARs were not obtainable in DAMIR.

While SARs report key program dates, often gaps were left for the specific program deliverables tracked for the purposes of this research. The Air Force Cost Analysis Agency (AFCAA) previously compiled a database for the purpose of tracking DoD aircraft programs, including information on significant program review dates. The AFCAA database, RAND database, and DAMIR portal were combined to provide the information needed for analysis.

Data Summary

Cost growth is seen across programs of all types, in all services, and in many varying degrees. However, the complexities of DoD acquisition make normalizing program milestones across different services difficult, creating incompatibilities in evaluation. Therefore, this research focuses only on aircraft programs within the DoD. Furthermore, the analysis is limited to only Acquisition Category 1 (ACAT 1) programs, since the reporting requirements for SAR do not extend to lower value thresholds. ACAT 1 programs are Major Defense Acquisition Programs (MDAPs), estimated to require eventual expenditure for research, development, test and evaluation (RDT&E), including all planned increments, of more than $480 million in Fiscal Year (FY) 14 constant dollars, or procurement, including all planned increments, of more than $2.79 billion (FY 2014 constant dollars) (DAU, 2015). Once the given criteria are applied to the SAR database, 36 DoD Aircraft programs are captured for evaluation. Table 1 lists the programs used for analysis in this research.

Table 1: Aircraft Programs Selected 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-1C
C-5B	F-22	Reaper

Limitations

The collection of data within the DoD, particularly with respect to the acquisition environment, is in and of itself a proven challenge. While in theory the DoD has an acquisition process to guide key reporting dates and decisions, in practice many programs deviate from this standard process. It therefore becomes difficult to create a standard set of variables and tracking measures that are consistent enough across the population to perform a thorough analysis. While ideally this research method could be applied to a variety of programs, both aircraft and otherwise, the lack of common review points for measurement makes the application difficult. Therefore, this study is limited to aircraft-only programs.

This research also heavily relies on the data contained within the SAR for each program, which has limitations of its own. Many previous studies, including Kozlak (2016), Arena et al. (2006), and Hough (1992) have discussed limitations surrounding the SAR. The key limitations that are relevant to this study are summarized here.

1. High–Level Data. While useful for obtaining dates and estimate information, SAR does not provide in-depth explanations for said data. This means high-level assumptions must be made, and critical details may be overlooked.
2. Reporting Guidelines and Requirement Changes. Over the historical span of SAR reporting, many changes have evolved the process. This leaves room for inconsistent comparisons among programs. Since this study spans the years from the 1960's on, it is particularly susceptible to this limitation.
3. Lack of Risk and Confidence Levels. Without this information, the estimates contained in the SAR are taken at face value. If an estimate contains a significant amount of associated risk, it may not be a valid representation of true program cost.
4. Each program creates its SAR for reporting. Since each individual program is responsible for populating and submitting a SAR, each SAR will be unique. This makes it difficult for comparison to other programs and virtually impossible for an apples to apples analysis.

Also, DAMIR, while a valuable resource for collecting SAR data, only contains reports post-1997. Prior SAR copies have not been electronically converted into the system, making information hard to locate on older acquisition programs. While the RAND database was validated against information on DAMIR for post-1997 SARS, older programs were unable to be verified. It is therefore an assumption of this research that the RAND database is accurate in its information collection and conversions. Given there were no errors detected in programs that could be verified on DAMIR, it is a reasonable assumption that the RAND database is accurate for evaluation.

In spite of the limitations within the SAR, its historical use in cost growth studies and relatively stable capture of vital information make it a reasonable source of information for this analysis.

Another limitation is the assumption that this research and the results that follow may be subject to a Type I error; that is, the incorrect rejection of the null hypothesis. Even though this research is exploratory in nature and uses an alpha value of 0.10, the volume of variables tested (720 total combinations) presents a much narrower window of significance. Therefore, while it is likely the top variables noted play some important role, the level of true significance is unknown until further research can be conducted. However, even with the knowledge of a possible Type I error, the general purpose of this research is to find general trends in longitudinal cost growth and meets that goal accordingly.

Data Set and Predictor Variables

In order to evaluate cost growth patterns across a program lifespan, key program dates, all encompassed by the Milestone B estimate and the Last Reported SAR (LRS) were selected. Each key date represents a new bracket of evaluation. These dates, subsequent to the development estimate at Milestone B, are used to track how cost growth moves across the longitudinal span of the program, relating each to its predecessor evaluation point, as well as the DE. The research of Kozlak (2016) identified four key dates critical to the evaluation of DoD Aircraft cost growth between Milestone B and LRS. As such, the same program events are used for the purposes of this research, to include:

1. Development Estimate point (Milestone B)
2. Critical Design Review (CDR)
3. First Flight (FF)
4. Development Test and Evaluation (DT&E)

5. Initial Operating Capability (IOC)

6. Final Estimate or Last Reported SAR (LRS)

For each program review point, the Current Estimate is considered the estimate provided in the SAR for the year in which the review point occurred.

In addition to the date and estimate value at each program review, other program information is captured by for analysis. The following predictor variables considered for evaluation of program cost growth are listed and described as follows:

- *Air Force – Binary Variable*
    - This variable identifies if the aircraft is developed for the Air Force.

- *Navy – Binary Variable*
    - This variable identifies if the aircraft is developed for the Navy.

- *Pre-1997 BY – Binary Variable*
    - This dummy variable identifies if the base year evaluated occurs before 1997. The year 1997 identifies the point in which SAR information becomes available electronically within DAMIR.

- *Prototype – Binary Variable*
    - This variable represents programs that create a prototype, or prototypes, of a weapons system before production of that weapons system begins. More than one type of prototype for a weapons system can be created in a given program.

- *Modification – Binary Variable*
    - This variable is concerned with programs whose existence serves as a modification to a pre-existing weapons system.

- *Small Program – Binary Variable*
    - This variable identifies if the estimate at Contract Award (Milestone B) is considered "small." See Figure 2 for evaluation of program size.

- *Medium Program – Binary Variable*
    - This variable identifies if the estimate at Contract Award (Milestone B) is considered "medium." See Figure 2 for evaluation of program size.

- *Large Program – Binary Variable*
    - o This variable identifies if the estimate at Contract Award (Milestone B) is considered "large." See Figure 2 for evaluation of program size.

- *Duration <180 (Months) – Binary Variable*
    - o This variable indicates if the program length from CA-LRS was less than 180 months. The 180 month period was determined a significant break in contract length when all programs were evaluated on a distribution analysis.

Figure 2: Distinction of Cost Growth Program Size

Other variables are present in the database for program information and reference; however, they were not included in tested variables for the purposes of this study. These variables include:

- *Contractor – Binary Variable*
    - o This variable identifies which contractor is responsible for the program development. If a contractor only appeared once, it was identified in the "Other" variable.

Given the very small sample size and the wide mix of contractors, this variable is likely a weak predictor for this analysis.

- *Quantity – Continuous Variable*
    - This variable reports the quantity predicted at the time of the program review.

Quantity is used to calculate the estimate/quantity factor to standardize the CGF. However, it is not tested as a variable.

- *Percent complete – Continuous Variable*
    - This variable measures the cost spent to date compared to the estimate value in the Last Reported SAR.

This is a common variable when evaluating cost growth. Percent complete is calculated based on the actual cost to date as a percentage of estimate at LRS. However, due to the inconsistencies of actual costs reported and the limitations of the SAR, this variable is limited in integrity, and is left out accordingly as a predictor variable for this analysis.

Estimating Cost Growth

To assess cost growth, this study evaluates what is referred to as a Cost Growth Factor (CGF). This is the calculation of current estimate as it relates to a previous, or original baseline, estimate. The CGF method divides the estimate plus the cost variance (actual) by the estimate (Drezner et al., 1993)

The neutral state of a CGF is where actuals equal the estimate, returning a value of 1.0. This indicates there has been no program cost growth. When the CGF is greater than 1.0, the program actuals are higher than the estimate, showing that there has been cost growth sustained by the program. Conversely, a CGF less than 1.0 indicates that the current cost of the program is less than the estimate, or a projected cost-underrun. To calculate the percent cost growth, subtract 1 from the cost growth factor (Drezner et al., 1993)

A program carries multiple estimates over its lifecycle. First, a Planning Estimate (PE) is the estimate made during the Concept Exploration and Definition stage. Because this estimate is prior to Milestone B, it is not referenced for analysis in this research. The Development Estimate (DE), often referred to as a baseline estimate, occurs at Milestone B and is used as the foundational estimate in this research. Finally, the Current Estimate (CE) is the most up to date estimate. If a program is complete, the CE is the actual cost of the program (Calcutt, 1993).

Estimators calculate cost growth from a baseline estimate, the PE, DE, or CE. Typically, the DE at MS B is the baseline estimate for cost growth. MS B is the point in the schedule where a program enters full-scale development and officially becomes a "program of record." Once a program of record is established, the program is required to file official cost reports with Congress (Porter et al., 2009). As formal cost reports materialize, cost growth becomes easier to track, and for this reason, the estimator measures cost growth from the DE when possible.

Data Standardization

The quantities each aircraft program produces typically shift both upwards and downwards throughout the stages of a program's lifecycle. Accounting for this change is a contested topic in terms of cost growth. The argument presents two ways: Cost growth associated with change in quantity is still cost growth that needs to be measured, and cost growth associated with change in quantity needs to be standardized to view cost growth inherent to the system. For this research, the cost growth factor is standardized to account for any change in quantity. SARs list the quantities estimated and produced for each aircraft program. In order to standardize the cost growth factor for quantity, a cost per aircraft is calculated at each program review point. This amount is then used in the calculation of the cost growth factor.

In addition to quantity, the database is standardized for consistent Base Year (BY) reporting. While not all 36 programs are evaluated in the same BY, each individual program maintains a common Base Year across each review point, making comparison of cost growth consistent. Since the comparison is from within each program and not across all programs, it is not necessary to convert all programs to the same BY. 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 review points.

Cost Growth

Once the cost per aircraft at each program review is calculated, the cost growth factor for a variety of combinations is found in order to evaluate the cost growth from a longitudinal perspective. This combination of interlocking cost growth evaluation is what distinguishes this research from cost growth research in the past. Each combination of cost growth factors is evaluated in order to present a tool of identification for cost growth management. Table 2 shows the combinations of cost growth evaluated. Each relationship is assigned a number for analysis naming purposes and is identified to the right of the relationship title.

In Tier I, relationships 1-5, the overall cost growth incurred from Contract Award to each subsequent program review is evaluated. Tier I measures cost growth while holding the first review point of Contract Award constant. More specifically, the first evaluation point measured is Contract Award to Critical Design review, while the last cost growth factor in this tier represents total growth from Contract Award to Last Reported SAR.

Table 2: Review Relationship Tiers

Tier I
Contract Award - Critical Design Review (1)
Contract Award - First Flight (2)
Contract Award - Development Test & Evaluation (3)
Contract Award - Initial Operating Capability (4)
Contract Award - Last Reported SAR (5)
Tier II
Critical Design Review - First Flight (6)
First Flight - Development Test & Evaluation (10)
Development Test & Evaluation - Initial Operating Capability (13)
Initial Operating Capability - Last Reported SAR (15)
Tier III
Critical Design Review - Development Test & Evaluation (7)
Critical Design Review - Initial Operating Capability (8)
Critical Design Review - Last Reported SAR (9)
First Flight - Initial Operating Capability (11)
First Flight - Last Reported SAR (12)
Development Test & Evaluation - Last Reported SAR (14)

Secondary evaluation includes program review combinations noted in Tier II, relationship numbers 6, 10, 13, and 15. Relationship 1 is also re-considered as a part of Tier II analysis. This tier is depicting the cost growth at each longitudinal cross section, as the program moves through the lifecycle. Instead of the overall cost growth measured in Tier I, dissecting the program down to each intermittent phase of the program lifecycle can help to show a more definitive image of how cost growth behaves at subsections of the program. Tier II is the focus of management decision making tools provided in this research.

Finally, Tier III combinations are evaluated showing remaining combinations of phases. While this research focuses primarily on results from Tier I and Tier II, Tier III is used for further analysis and confirmation of suspected indicators.

Color Coding Cost Growth

A stoplight chart is a color coded decision tool that helps quickly identify project status and risk levels using visual aids. Typically, Green, Amber, and Red are used to indicate the status of a program. However, subjectivity is involved in defining the conditions as to what constitutes red, green, or amber status, and must be rationally defined. For the purposes of this research, a distribution summary of cost growth factors is evaluated for color break points. A review of the distribution of CGFs for each of the 15 program review relationships, as well as the overall distribution, is analyzed for logical divisions. Figure 3 shows the distribution analysis for Contract Award-Critical Design Review as an example of the distribution breakout. The other distribution summaries can be found in Appendix A.

Green indicates low to no risk and identifies programs that incur a CGF of 1.0 or less. As previously indicated, this means that the actuals are at or below the estimated value. Amber signifies the program has encountered some risk, but is not yet in jeopardy of serious trouble. Programs that fall in this category are highlighted for extra attention and scrutiny, but are still within the range of correction. This research shows that programs stay moderately stable up until the 10% cost growth range; therefore, Amber coded programs are defined as those which encounter cost growth ranging from 1.0 to 1.10. After 10%, cost growth starts to vary widely with large inconsistencies and predictabilities. Therefore, a CGF that lies above 1.10 is identified as Red.

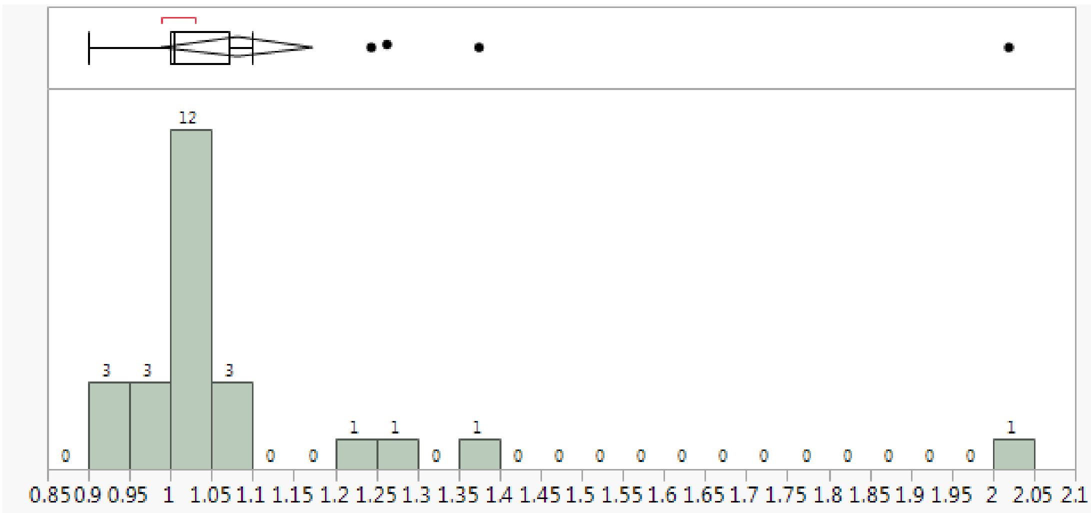


Figure 3: Distribution Used for Color Rating Assignment

Following the categorized levels identified for each color, the CGF at each Tier and Tier sub-section is allocated to its respective color rating in order to assign a risk level color for that section's cost growth. This adds the response variables to the database. The following variables are now incorporated for analysis:

- *Green – Binary Variable*
    - This variable identifies if the analyzed CGF falls at 1.0 or below.
- *Amber – Binary Variable*
    - This variable indicates if the CGF falls between the range 1.0 < CGF ≤ 1.10.
- *Red– Continuous Variable*
    - This variable indicates if the CGF is > 1.10.

This step is captured for each of the 15 program review relationships identified above in Table 2. The first relationship is identified as Green/Amber/Red, while the second relationship is identified by Green2/Amber2/Red2. This identification helps subsequent analysis between different program review points. In addition to the evaluation of the CGF between each of the six program review points (totaling the 15 relationships), the assigned color at each review point for each relationship is compared to look for trends in programs across their lifecycle. For example, if a program is Green at the first program review, is it more likely to be Green at evaluation point 2? This analysis reviews each combination of factors to help provide insight to longitudinal cost growth behavior patterns.

Two-Way Contingency Tables

This research focuses on multinomial experiments looking for significant relationships with respect to two qualitative factors. Each combination of factors present in the database are compared in order to distinguish the most statistically significant relationships in a program lifecycle that helps to predict the color status of a program at any given review point. Among the 15 program review relationships, 720 predictor/response combinations are tested for independence. Table 3 shows the predictor variables and response variable combinations that are tested in this research. Only valid relationships are denoted by a "1". Relationships not indicated by a "1" are those in which the predictor variable is not applicable to the response variable. For instance, Green cannot predict a response of Green; therefore, no relationship is present and no "1" is denoted.

Table 3: Predictor and Response Variable Combinations

	Predictor																					
Response	AF	Navy	Pre97	Post97	Prototype	Mod	Small	Meduim	Large	<180 mo	G	A	R	G6	A6	R6	G10	A10	R10	G13	A13	R13
G	1	1	1	1	1	1	1	1	1	1												
A	1	1	1	1	1	1	1	1	1	1												
R	1	1	1	1	1	1	1	1	1	1												
G2	1	1	1	1	1	1	1	1	1	1	1	1	1									
A2	1	1	1	1	1	1	1	1	1	1	1	1	1									
R2	1	1	1	1	1	1	1	1	1	1	1	1	1									
G3	1	1	1	1	1	1	1	1	1	1	1	1	1									
A3	1	1	1	1	1	1	1	1	1	1	1	1	1									
R3	1	1	1	1	1	1	1	1	1	1	1	1	1									
G4	1	1	1	1	1	1	1	1	1	1	1	1	1									
A4	1	1	1	1	1	1	1	1	1	1	1	1	1									
R4	1	1	1	1	1	1	1	1	1	1	1	1	1									
G5	1	1	1	1	1	1	1	1	1	1	1	1	1									
A5	1	1	1	1	1	1	1	1	1	1	1	1	1									
R5	1	1	1	1	1	1	1	1	1	1	1	1	1									
G6	1	1	1	1	1	1	1	1	1	1	1	1	1									
A6	1	1	1	1	1	1	1	1	1	1	1	1	1									
R6	1	1	1	1	1	1	1	1	1	1	1	1	1									
G7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1						
A7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1						
R7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1						
G8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
A8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
R8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
G9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
A9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
R9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
G10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1						
A10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1						
R10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1						
G11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1						
A11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1						
R11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1						
G12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1						
A12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1						
R12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1						
G13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
A13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
R13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
G14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
A14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
R14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
G15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
A15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
R15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

The method of analysis used is a contingency table (also known as a cross tabulation or crosstab) to look for independence among variables. A contingency table displays the frequency distribution of the variables. They are heavily used in survey research because they provide a basic picture of the interrelation between two variables. Table 4 shows a basic summary of a contingency table between two variables, A and B. For each variable there are two responses. A response of “no” indicates the variable does not meet the criteria, and a response of “yes” indicates the variable meets the criteria. For each combination of variables A and B, four combinations are possible:

1. No, No
2. No, Yes
3. Yes, No
4. Yes, Yes

Each combination is counted, and the count is represented in the table by n_{11}, n_{12}, n_{22}, n_{22}, respectively. The counts are then totaled in each column (C_1 and C_2) and Row (R_1 and R_2), and a total count of observations, n, is determined. Once each row and column is populated, the counts are then compared to expected values and measured for independence.

Table 4: Sample Contingency Table

		Variable B		
		No	**Yes**	**Totals**
	No	n_{11}	n_{12}	R_1
Variable A	**Yes**	n_{21}	n_{22}	R_2
	Totals	C_1	C_2	n

Fisher’s Exact Test

To assess the significance of the contingency table, a P-value is calculated. A variety of methods can be used in calculating a P-value. Due to the small sample size of this research, Fisher’s Exact Test is used. Fisher’s Exact Test is a statistical test used to determine if there are nonrandom associations between two categorical variables drawn from calculated contingency tables. Apart from other significance tests available, a benefit of using Fisher’s Exact Test is that it does not estimate the probability of a value;

rather the test calculates the exact probability of receiving the observed data (Kozlak, 2016). While the test is available for any sample size, it is particularly useful when sample sizes are small, as other applicable tests are limited. Although the total sample size used in this study is 36 programs, each combination of variables results in many cases where there is an even smaller sample set. This makes Fisher's Exact Test an acceptable measure for independence between a program having defined color rating and the explanatory variables considered in this thesis.

To understand Fisher's Exact Test, two assumptions are necessary. The first assumption is that all observations are independent. Second, the test operates under fixed, or conditioned, row and column totals. This second assumption is what distinguishes Fisher's Exact Test from other statistical independence tests with unconditioned rows and columns (McDonald, 2009).

The results of Fisher's Exact Test produce both a 1-tailed and a 2-tailed hypothesis test. The 1-tailed test uses a single tail of the probability distribution (either left tail, or right), and examines changes in a single direction. The hypothesis for a 1-tailed test is:

- H_o: the factors are the same
- H_a: the probability the color rating is greater for the factor = '1' than '0' (right tailed)
- H_a: the probability the color rating is greater for the factor = '0' than '1' (left tailed)

The 2-tailed test uses both tails of a probability distribution (both right and left) and examines changes in both directions. The hypothesis for a 2-tailed test is:

- H_o: the factors are the same
- H_a: the factors are not the same

This research uses the 1-tailed hypothesis tests to identify if a categorical factor increases the likelihood a program will have the tested color rating. The null hypothesis states the categorical variable does not affect the color rating. If the p-value for a right sided test is significant, the predictor variable increases the likelihood of the tested response color rating. If the results were significant for the left sided test, the predictor variable is less likely to produce the tested response color rating.

Fisher's Exact Test is used to look at each categorical predictor variable, identified above, against each color rating. In turn, as the color ratings progress from Green to Green2 and forward, the previous color ratings also become tested predictor variables.

Chapter Summary

This chapter details the research methodology used. The collection of data from the SAR and DAMIR is described. Predictor variables that provide a link to the response variable were identified and explained. The reasoning for the methodology used is provided along with a detailed explanation of the Fisher's Exact Test for analysis. The next chapter introduces the results to the testing involved in this research, and the analysis of findings.

IV. Results and Analysis

Chapter Overview

This chapter provides the results from the methodology outlined in Chapter III. First, a color code is distinguished for each of the 15 program review relationships. Second, an analysis of the specified techniques is conducted on the 720 predictor/response relationships in the database. The findings are then measured and compared. Finally, the details of the significant variables are discussed and summarized.

Color Rating Summary

The overview of our database revolves around how the risk of cost growth is depicted across the program lifecycle. As discussed in Chapter III, risk levels are defined as Green, Amber, or Red, and are applied on a basis of CGF at the given program review point. Each program receives a color rating at each of the 15 relationship points, although this research focuses on the results from Tier I and Tier II relationships. Tier III findings did not produce results that enhance the longitudinal evaluation, and therefore were excluded from analysis. While the color rating does not tell the complete story, which is supported by our further analysis, it does provide a management level overview of how cost growth has historically behaved throughout the longitudinal aspect of a program. Table 5 shows the color rating summary for Tier I relationships.

Table 5 shows how the color rating of a program changes as a program moves from Contract Award to each program review point. It is important to know that these relationships depict overall cost growth as they all start at CA and move forward. While this is a good indicator of program health, it is also a commonly studied and traditional view of looking at cost growth. Therefore, this research breaks down the program lifecycle further, reviewing each individual subsection of a program for a more distinct picture on the behavior of cost growth, indicated by Tier II relationships.

Table 5: Tier I Color Rating Summary

Program	CA-CDR	CA-FF	CA-DTE	CA-IOC	CA-LRS
A-10	Yellow	Red	Red	Red	Red
AV-8B	Green	Yellow	Yellow	Green	Yellow
B-1A	Green	Red	Red		Red
B-1B	Green	Green	Green	Green	Yellow
B-1B CMUP Computer Upgrade	Yellow	Yellow	Red	Yellow	Red
B-1B CMUP JDAM	Green	Green	Green	Green	Green
B-2 RMP		Green	Red	Yellow	Red
C-130 AMP	Red	Red			Red
C-17A		Red	Red	Red	Red
C-5 RERP	Yellow	Yellow	Red	Green	Red
C-5B		Green	Green		Green
E-2C		Red	Red	Red	Green
E-2D	Yellow	Yellow	Red	Red	Red
E-6A	Yellow	Green	Green	Green	Green
EA-18G	Yellow	Yellow	Yellow	Red	Green
EF-111	Yellow	Red	Red	Red	Red
FA-18EF	Yellow	Yellow	Red	Green	Red
F-14A	Green	Green	Red	Green	Red
F-15	Yellow	Yellow	Red	Red	Red
F-16	Green	Green	Green	Red	Yellow
F-22	Red	Red	Red	Green	Red
F-35	Red	Red			Red
RQ-4 (GLOBAL HAWK)		Red	Red	Red	Red
MQ-4C	Yellow	Green	Green		Green
P-8A	Yellow	Green	Green	Yellow	Red
S-3A		Yellow	Yellow	Green	Red
T-6 (JPATS)	Green	Green	Red	Red	Red
T-45TS	Yellow	Red	Red	Red	Red
V-22 FSD	Red	Red	Red	Red	Red
F-22 Inc 3.2B Mod	Green	Green			Green
KC-46	Green	Green			Green

Tier II relationships take a more in depth view of cost growth, by looking at each CGF independently of historical behavior. Instead of moving from CA forward, Tier II only looks at the relationship of a program review cost growth as it relates to the cost growth of the preceding review point. This can help management predict a more short-term planning strategy. While the program may have an overall CGF of red, the near term in real-time may have a level or decreasing CGF. While long-term mitigation may be in need, management can strategically plan for real-time responses. Table 6 shows the color ratings for Tier II evaluation.

Table 6: Tier II Color Rating Summary

Tier II	CA-CDR	CDR-FF	FF-DTE	DTE-IOC	IOC-LRS
A-10	Yellow	Red	Yellow	Green	Yellow
AV-8B	Green	Yellow	Green	Green	Red
B-1A	Green	Red	Red		
B-1B	Green	Green	Yellow	Green	Yellow
B-1B CMUP Computer Upgrade	Yellow	Yellow	Red	Yellow	Green
B-1B CMUP JDAM	Green	Green	Green	Green	Green
B-2 RMP			Red	Green	Green
C-130 AMP	Red	Red			
C-17A			Red	Green	Green
C-5 AMP				Green	Red
C-5 RERP	Yellow	Yellow	Red	Green	Green
C-5B			Green		
E-2C			Green	Green	Green
E-2D	Yellow	Yellow	Red	Red	Yellow
E-6A	Yellow	Green	Yellow	Yellow	Yellow
EA-18G	Yellow	Yellow	Green	Green	Green
EF-111	Yellow	Red	Red	Green	Yellow
FA-18EF	Yellow	Yellow	Red	Yellow	Yellow
F-14A	Green	Green	Red	Yellow	Green
F-15	Yellow	Yellow	Red	Green	Yellow
F-16	Green	Green	Yellow	Yellow	Yellow
F-22	Red	Red	Red	Green	Yellow
F-35	Red	Green			
RQ-4 (GLOBAL HAWK)			Yellow	Yellow	Red
MQ-4C	Yellow	Green	Green		
P-8A	Yellow	Green	Green	Yellow	Red
S-3A			Yellow	Green	Yellow
T-6 (JPATS)	Green	Green	Red	Yellow	Yellow
T-45TS	Yellow	Yellow	Red	Green	Red
V-22 FSD	Red	Red	Green	Green	Yellow
F-22 Inc 3.2B Mod	Green	Green			
KC-46	Green	Green			
MQ-1C			Green	Green	
Reaper (block 1)			Green	Yellow	

The results in Tier II color rating appear to be more incomplete than Tier I. This is due to the fact that the most common missing date among programs is for CDR. Therefore, the first two relationships in Tier II are affected, where in Tier I, only one relationship would be affected. However, even with incomplete data, the results of color ratings for Tier II show an overall summary that provides significant insight. Table 6 shows that cost growth factors rate red significantly less than in Tier I (29 vs. 64). Also, of the 29 red ratings, 13 (45%) occur during FF-DTE (relationship 10). Even before running contingency tables and Fisher's Exact Test, this shows strong support that not only is the CGF significantly different when looking at a program from a longitudinal perspective versus a lifecycle perspective, the primary driver of high CGF is the Development Test and Evaluation phase. This insight can help decision makers plan best for mid-program cost increases, while avoiding over-anticipation of program end turmoil.

The color rating assignment and comparison paves the way for further analysis. Now that an overall picture is painted, we dive further by running contingency tables between the 720 combinations of variables, in order to best determine what are strong predictors of the defined color rating at each given review point. While we evaluate 720 relationships across the 15 stated relationships, this research focuses on those within Tier I and Tier II.

Tier I Results

Tier I contingency tables test every nominal variable against each of the 15 color ratings within Tier I relationships (Green through Red5). As previously mentioned, Tier I shows overall cost growth as a program moved from Contract Award to each review point of the lifecycle. For each contingency table evaluated, a Fisher's Exact Test value is compared to an alpha value of α = 0.10. This section walks through the significant findings within Tier I, focusing on the five most significant dependencies, and summarizing the remaining significant findings. A general analysis summarizes our Tier I findings before the discussion of Tier II examination.

Green3 given Green2

The most significant dependency in Tier I evaluations is Green3 (CA-DTE) given Green2 (CA-FF). Figure 4 shows the contingency table results for this analysis. While it seems fairly logical that if a program is Green at one evaluation point, it is more likely to be Green at the second evaluation point, this relationship is extremely significant, with the P-value for Fisher's Exact Test at 0.0001 (Table 7).

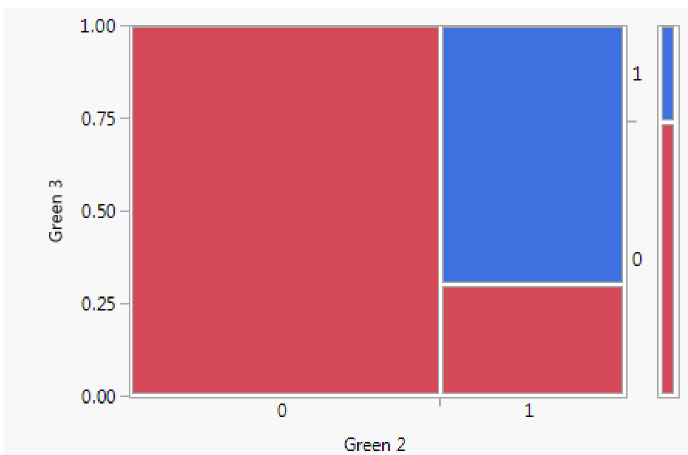


Figure 4: Contingency Table Color Chart for Green3 Given Green2

Table 7: Fisher's Exact Test Results for Green3 Given Green2

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	1.0000	Prob(Green 3=1) is greater for Green 2=0 than 1
Right	0.0001*	Prob(Green 3=1) is greater for Green 2=1 than 0
2-Tail	0.0001*	Prob(Green 3=1) is different across Green 2

The contingency table in Table 8 shows the expected count of programs being green at both points is roughly 2.5 whereas this analysis encountered 7 programs. It is also important to note that Green3 encompasses DTE, which is the area of most significant cost growth in a program. 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 DTE with minimal cost growth.

Table 8: Contingency Table for Green3 Given Green2

Contingency Table

Green 2 \ Green 3 (Count / Expected)	0	1	Total
0	17 12.5926	0 4.40741	17
1	3 7.40741	7 2.59259	10
Total	20	7	27

Maintaining Green at point 3 could be a critical factor in a program's decision structure and funding profile. This relationship shows management that efforts to maintain a Green rating at FF can truly impact its overall CGF. This may aid pushing management to make early investments to mitigate cost growth, as it provides the most return on investment.

Red5 given Red3

In contrast to the results provided above in the first analysis, the second most significant dependency in Tier I evaluations is Red5 (CA-LRS) given Red3 (CA-DTE). Figure 5 shows the contingency table results for this dependency. 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 (Table 9).

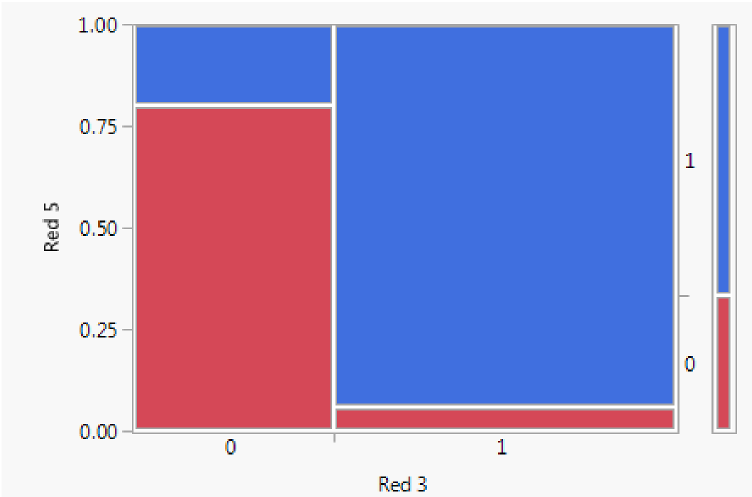


Figure 5: Contingency Table Color Chart for Red5 Given Red3

Table 9: Fisher's Exact Test Results for Red5 Given Red3

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	1.0000	Prob(Red 5=1) is greater for Red 3=0 than 1
Right	0.0002*	Prob(Red 5=1) is greater for Red 3=1 than 0
2-Tail	0.0002*	Prob(Red 5=1) is different across Red 3

The contingency table in Table 10 shows the expected count of programs being Red at both points is roughly 11.3 whereas this analysis encountered 16 programs that completed LRS in a Red rating. It is also important to highlight, that of the 17 programs that were rated Red at DTE, 16 of them continued to rate Red at LRS. This is a significant indicator to management as to where the program is headed if cost growth is out of control by DTE.

The first two dependencies analyzed highlight the importance of program status as of DTE, providing further encouragement for decision makers to invest early in mitigating cost growth drivers.

Table 10: Contingency Table for Red5 Given Red3

Contingency Table

Red 3	Count / Expected	Red 5: 0	Red 5: 1	Total
	0	8 3.33333	2 6.66667	10
	1	1 5.66667	16 11.3333	17
	Total	9	18	27

Red5 given Green3

Further supporting the first two analyses, the relationship between Red5 (CA-LRS) and Green3 (CA-DTE) is also statistically significant. While the analysis of Red5 given Red3 indicated a high likelihood that a program is Red at DTE will stay Red at LRS, the Red5 given Green3s relationship shows that Green at DTE makes it unlikely a program will be Red at LRS. Figure 6 shows the contingency table results for this analysis. The Fisher's Exact Test shows a P-Value of 0.0017 (Table 11); however this is a left sided result, showing that it is less likely to return a "1" at Red 5 than 0.

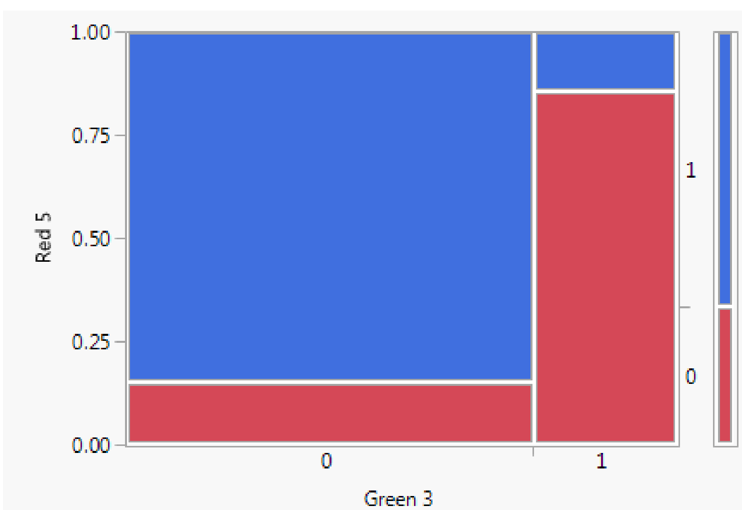


Figure 6: Contingency Table Color Chart for Red5 Given Green3

Table 11: Fisher's Exact Test Results for Red5 Given Green3

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.0017*	Prob(Red 5=1) is greater for Green 3=0 than 1
Right	1.0000	Prob(Red 5=1) is greater for Green 3=1 than 0
2-Tail	0.0017*	Prob(Red 5=1) is different across Green 3

The contingency table in Table 12 shows the expected count of programs being Red at LRS given Green at DTE to be approximately 4.7. However, of the 7 programs that had a rating of Green at DTE, only 1 encountered a rating of Red at LRS. This is even further validation that strong control of cost growth early can minimize the risk of extensive cost growth throughout the entire program lifecycle.

Table 12: Contingency Table for Red5 Given Green3

Contingency Table

Green 3 \ Red 5 (Count / Expected)	0	1	Total
0	3 6.66667	17 13.3333	20
1	6 2.33333	1 4.66667	7
Total	9	18	27

The top three significant relationships all return with a commonality; DTE indicating a very relevant target for programs in terms of cost growth management.

Green5 given Modification

The fourth result, when ranked by lowest P-value, shows Modification as a predictor of Green at review point 5 (CA-LRS). Figure 7 shows the contingency table results for this analysis. The Fisher's Exact Test has a P-Value of 0.0017 (Table 13); indicating that for Green5, it is more likely for Modification=1 to return a value than 0.

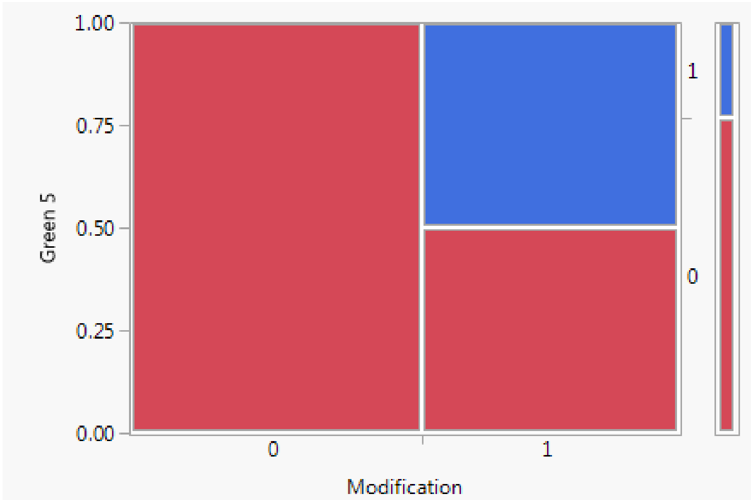


Figure 7: Contingency Table Color Chart for Green5 Given Modification

Table 13: Fisher's Exact Test Results for Red5 Given Green3

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	1.0000	Prob(Green 5=1) is greater for Modification=0 than 1
Right	0.0017*	Prob(Green 5=1) is greater for Modification=1 than 0
2-Tail	0.0017*	Prob(Green 5=1) is different across Modification

The contingency table in Table 14 shows that the expected count of programs being Green at LRS given the program is a Modification is approximately 3.2. However, 7 programs met these criteria.

Table 14: Contingency Table for Green5 Given Modification

Contingency Table

Modification \ Green 5 (Count / Expected)	0	1	Total
0	16 12.2667	0 3.73333	16
1	7 10.7333	7 3.26667	14
Total	23	7	30

It is also important to note here that a solid color bar appears in the contingency table. This shows that of all programs not coded "Modification," zero returned a value of "1" at LRS. While not the tested variable, this is extremely important to note for management planning purposes, as it indicates there is a very low change of a non-mod program ending in a Green rating.

Red3 given Red2

Once again DTE returns to the forefront. The final significant finding in the top 5 of Tier I, shows the dependency between Red3 (CA-DTE) and Red2 (CA-FF). Figure 8 shows the contingency table results for this analysis. The Fisher's Exact Test shows a right sided P-Value of 0.0052 (Table 15) depicting that the probability for a Red3 rating is greater for Red2=1 than 0.

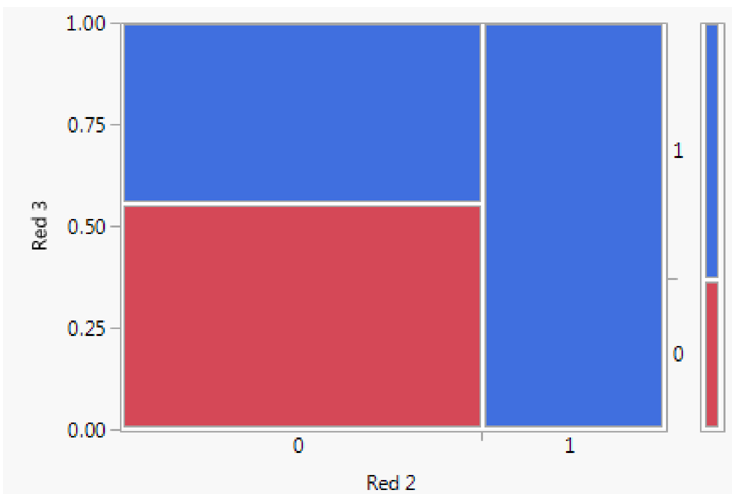


Figure 8: Contingency Table Color Chart for Red3 Given Red2

Table 15: Fisher's Exact Test Results for Red3 Given Red2

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	1.0000	Prob(Red 3=1) is greater for Red 2=0 than 1
Right	0.0052*	Prob(Red 3=1) is greater for Red 2=1 than 0
2-Tail	0.0088*	Prob(Red 3=1) is different across Red 2

The contingency table in Table 16 shows the expected count of programs being Red at DTE given the program is Red at FF is approximately 5.7, whereas the analysis returned a value of 9. Also, once again a solid colored bar appears in the contingency table. This is represented by the circumstance that every program that produced a Red rating at FF returned a red rating at DTE. This is perhaps another look at how early cost growth performance dictates long-run outcomes.

Table 16: Contingency Table for Red3 Given Red2

Contingency Table

	Red 3			
	Count Expected	0	1	Total
Red 2	0	10 6.66667	8 11.3333	18
	1	0 3.33333	9 5.66667	9
	Total	10	17	27

Other Significant Tier I Results

Of the Tier I relationships tested, 32 flagged as significant dependencies at the confidence level $\alpha = 0.10$. Table 17 shows the remaining 27 significant findings. Each of these findings is sorted by their level of significance, not by the variables impacted. Also noted is the direction of the tail on the test, to define how the relationship is related.

Table 17: Other Significant Tier I Results

Response	Predictor	Fishers Exact Test	Direction
Red2	Red	0.0055	Right
Green2	Green	0.0065	Right
Green5	Red3	0.0078	Left
Red3	Green2	0.0104	Left
Red5	Modification	0.0131	Left
Red4	Red2	0.0136	Right
Red	Large	0.0162	Right
Red2	Prototype	0.0179	Right
Green5	Green2	0.022	Right
Green5	Green3	0.0239	Right
Red5	Green	0.0252	Left
Red2	Medium	0.026	Left
Red5	Red2	0.026	Right
Amber2	Amber	0.0266	Right
Red3	<180 months	0.0311	Left
Green3	Red2	0.0358	Left
Green2	Prototype	0.0396	Left
Green3	Amber2	0.0567	Left
Amber	Pre97	0.0749	Left
Amber2	Medium	0.0787	Right
Amber	Navy	0.082	Right
Green4	Red2	0.0875	Left
Green4	Red3	0.0907	Left
Red4	Green2	0.0965	Left
Red4	Red3	0.0965	Right
Green	Pre97	0.0982	Right
Red2	Large	0.0983	Right

Tier II Results

Tier II contingency tables test every nominal variable against each of the 15 color ratings within Tier II relationships (relationships 1, 6, 10, 13, and 15). Unlike Tier I, Tier II shows the individual cross sections of the program lifecycle which provides better insight into the true behavior of cost growth at each longitudinal section, as opposed to the overarching view traditionally taken. Just as tested in Tier I, for each contingency table evaluated, a Fisher's Exact Test value is compared to an alpha value of α=0.10. This section discusses the significant findings within Tier II, focusing on the five most significant dependencies, and summarizing the remaining significant findings. A general analysis summarizes the Tier II findings before the discussion of the overall analysis.

Amber6 given Amber

The most significant dependency in Tier I evaluations is Amber6 (CDR-FF) given Amber (CA-CDR). Figure 9 shows the contingency table results for this analysis. This is the first appearance of Amber in the analysis, and it consists of not one side, but both sides of the dependency. In Tier I analysis, programs depicted Green or Red much more than Amber. Dissecting the programs on a small scale shows a lot more variability to the behavior of the cost growth. Table 18 shows the P-value for Fisher's Exact Test at 0.0100, depicting that the probability for a Amber6 rating is greater for Amber=1 than 0.

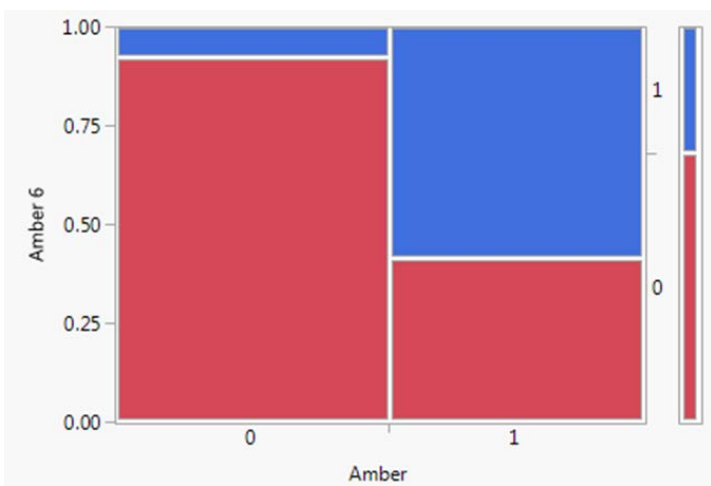


Figure 9: Contingency Table Color Chart for Amber6 Given Amber

Table 18: Fisher's Exact Test Results for Amber6 Given Amber

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9995	Prob(Amber 6=1) is greater for Amber=0 than 1
Right	0.0100*	Prob(Amber 6=1) is greater for Amber=1 than 0
2-Tail	0.0112*	Prob(Amber 6=1) is different across Amber

The contingency table in Table 19 shows the expected count of programs being Amber from CDR-FF given the program is Amber at CA-CDR, is approximately 3.84. The analysis results, however, were nearly double at 7 occurrences.

Table 19: Contingency Table for Amber6 Given Amber

Contingency Table

Amber \ Amber 6 (Count / Expected)	0	1	Total
0	12 8.84	1 4.16	13
1	5 8.16	7 3.84	12
Total	17	8	25

Green6 given Green

The second significant finding relates to the dependency of Green6 (CDR-FF) to Green (CA-CDR). This follows similar logic to Tier I results, in that early program cost growth mitigation, can aid in future increments. Figure 10 shows the contingency table results for this analysis. Table 20 shows the P-value for Fisher's Exact Test at 0.0159 revealing that the probability of Green6=1 is higher for Green=1 than 0.

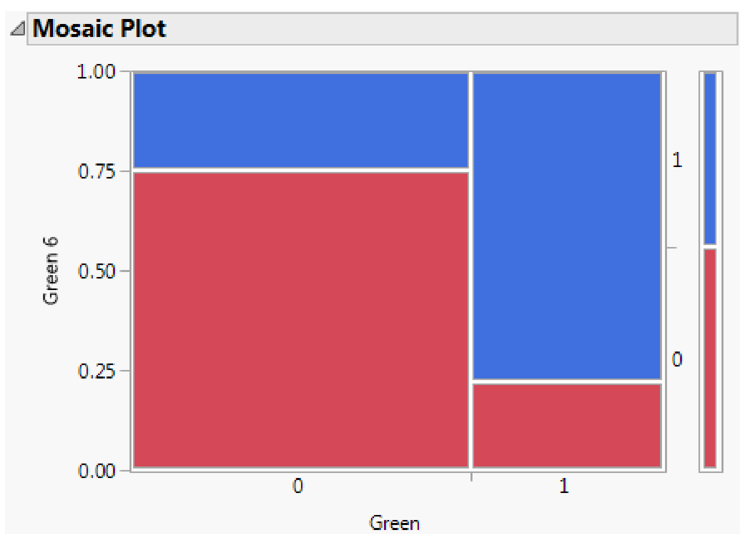


Figure 10: Contingency Table Color Chart for Green6 Given Green

Table 20: Fisher's Exact Test Results for Green6 Given Green

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9988	Prob(Green 6=1) is greater for Green=0 than 1
Right	0.0159*	Prob(Green 6=1) is greater for Green=1 than 0
2-Tail	0.0168*	Prob(Green 6=1) is different across Green

The contingency table in Table 21 shows the expected count of programs being Green during FF-DTE given the program is Green during CA-FF, is approximately 3.96, whereas the analysis conducted returned a value of 7.

Table 21: Contingency Table for Green6 Given Green

Contingency Table

		Green 6		
Green	Count / Expected	0	1	Total
	0	12 / 8.96	4 / 7.04	16
	1	2 / 5.04	7 / 3.96	9
	Total	14	11	25

Amber13 given Green6

Result 3 is perhaps one of the most unique seen thus far. The analysis now breaks from traditional Green-Green relationships. This result highlights the dependency between Amber at DTE-IOC and Green at CDR-FF. Figure 11 shows the contingency table results for this analysis. Table 22 shows the P-value for Fisher's Exact Test at 0.0292, revealing that the probability of Amber13=1 is greater for Green=1 than 0.

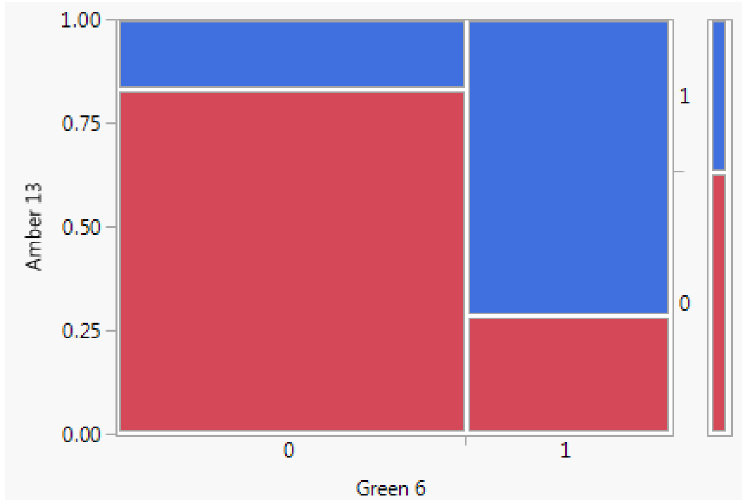


Figure 11: Contingency Table Color Chart for Amber13 Given Green6

Table 22: Fisher's Exact Test Results for Amber13 Given Green6

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9983	Prob(Amber 13=1) is greater for Green 6=0 than 1
Right	0.0292*	Prob(Amber 13=1) is greater for Green 6=1 than 0
2-Tail	0.0449*	Prob(Amber 13=1) is different across Green 6

This seems counter-intuitive to previous results in that it indicates programs is more likely to experience cost growth rather than sustain lower levels previously demonstrated early in the program. The contingency table (Table 23) shows the expected count of a program being Amber during DTE-IOC given the program is Green during CDR-FF, is approximately 2.6 whereas the analysis conducted returned a value of 5.

This appears to be an anomaly to previously discovered trends and is worth noting for further research. Underlying program commonalities need to be investigated to understand this relationship. However, it is a striking discovery, further validating the need for longitudinal review of data. Patterns like this aren't apparent when looking at cost growth from a whole-program perspective.

Table 23: Contingency Table for Amber13 Given Green6

Contingency Table

Green 6	Amber 13: Count / Expected	0	1	Total
	0	10 7.57895	2 4.42105	12
	1	2 4.42105	5 2.57895	7
	Total	12	7	19

Red6 given Red

The next significant relationship reverts back to traditional patterns revealing a dependency between Red6 (CDR-FF) and Red (CA-CDR). Figure 12 shows the contingency table results for this analysis. Table 24 shows the P-value for Fisher's Exact Test at 0.0312, revealing that the probability of Red6=1 is greater for Red=1 than 0.

The contingency table in Table 25 shows the expected count of a program being Red during CDR-FF given the program is Red during CA-CDR is 0.96, or almost 1. However, of the 4 programs that are rated Red at CA-CDR, 3 remain red at CDR-FF.

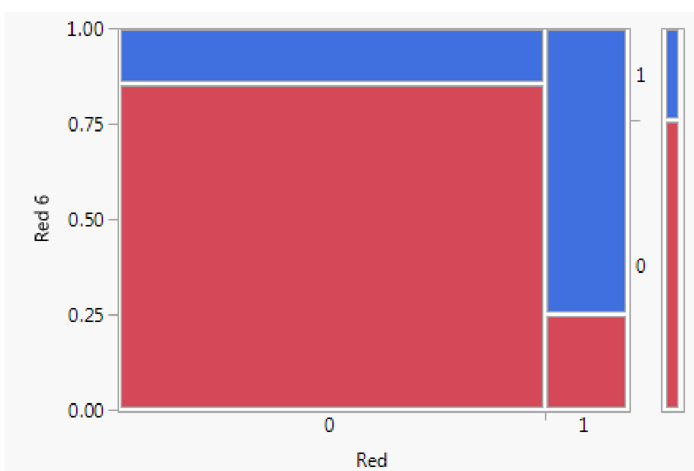


Figure 12: Contingency Table Color Chart for Red6 Given Red

Table 24: Fisher's Exact Test Results for Red6 Given Red

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9988	Prob(Red 6=1) is greater for Red=0 than 1
Right	0.0312*	Prob(Red 6=1) is greater for Red=1 than 0
2-Tail	0.0312*	Prob(Red 6=1) is different across Red

Table 25: Contingency Table for Red6 Given Red

Contingency Table

	Red 6			
Red: Count / Expected	0	1	Total
0	18 15.96	3 5.04	21
1	1 3.04	3 0.96	4
Total	19	6	25

Amber10 given Small

To round out the strongest five relationships in Tier II, program size finally makes an impact on color rating. This result shows the significant dependency between Amber at FF-DTE and a program size of Small. Figure 13 shows the contingency table results for this analysis. Table 26 shows the P-value for Fisher's Exact Test at 0.0433, revealing that the probability of Amber10=1 is greater for Small=1 than 0.

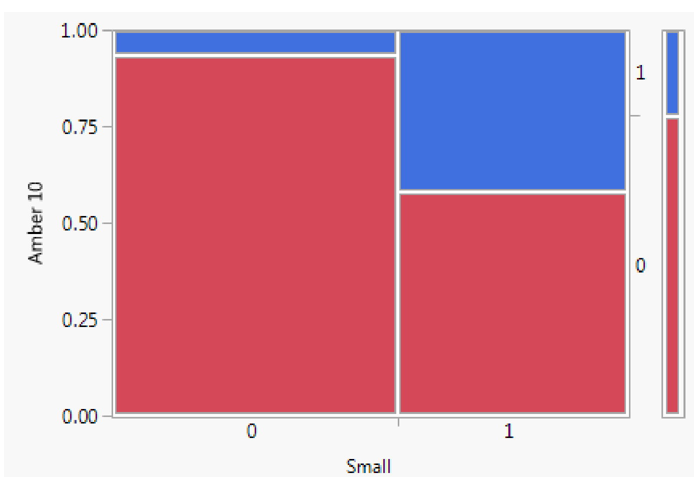


Figure 13: Contingency Table Color Chart for Amber10 Given Small

Table 26: Fisher's Exact Test Results for Amber10 Given Small

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9969	Prob(Amber 10=1) is greater for Small=0 than 1
Right	0.0433*	Prob(Amber 10=1) is greater for Small=1 than 0
2-Tail	0.0602	Prob(Amber 10=1) is different across Small

The contingency table in Table 27 shows the expected count of programs being Amber during FF-DTE given the program is Small is 2.7, whereas this analysis observed 5 programs.

Table 27: Contingency Table for Red6 Given Red

Contingency Table

		Amber 10		
	Count Expected	0	1	Total
Small	0	14 11.6667	1 3.33333	15
	1	7 9.33333	5 2.66667	12
	Total	21	6	27

Other Significant Findings

Of the Tier II relationships tested, 15 flagged as significant dependencies at the confidence level α = 0.10; much less than Tier I. Table 28 shows the remaining 10 significant findings.

Table 28: Other Significant Tier II Results

Response	Predictor	Fishers Exact Test	Direction
Green10	AF	0.0547	Left
Green10	Navy	0.0547	Right
Red10	Green6	0.0635	Left
Green15	Modification	0.0642	Right
Green13	Green6	0.0674	Left
Green6	Amber	0.0749	Left
Green13	Red6	0.0851	Right
Amber15	Red6	0.0851	Right
Amber15	Green10	0.0775	Left
Amber15	Amber10	0.0775	Right

Chapter Summary

In this chapter, the overall color rating trends of both Tiers I and Tier II are explained to present the foundation of the research. The contingency table analysis of significant variables is shown, and a separate in-depth analysis for each of the top five predictor variables for both Tiers I and II is presented. Statistical testing and patterns found within the data were discussed as they relate to recommended program management response on future programs. Lastly, significant variables outside of the top ten in-depth analyses are summarized. In the next chapter, the research is concluded and broad discussions and meaning to the analysis is presented.

V. Conclusions and Recommendations

Chapter Overview

The major finding in this research is the identification of statistically significant predictor variables on cost growth risk ratings. This research also evaluates the traditional approach to long-term cost growth review compared to a new longitudinal perspective of sectional cost growth. This chapter reviews the initial research questions to validate that the research accomplished the intended goal. Additionally, the limitations of findings are reviewed, and areas for future research are identified.

Research Questions Answered

1 –*How does cost growth behave differently between a segmented longitudinal perspective and a traditional lifecycle perspective?*

The answer to this question lies in two parts. First, when you map the color ratings from Tier I (the traditional lifecycle perspective), with Tier II (the longitudinal perspective), you can tell a distinct difference in how cost growth behaves at each review measured. The color schemes are dramatically different, with Red dominating the traditional approach and significantly more Green and Amber in the longitudinal approach. There is no denying that having a more refined view of an issue can provide insight that cannot be seen within the larger, lifecycle trends. It is in this fact that this research provides a fresh perspective on cost growth trends. Though limited in scope, this research paves the way for future research, in looking at cost growth as a near-term trend than a long-term burden.

The second part to this question is a bit more blurry than the first. While there still remain distinctions between both perspectives, both hold a pretty common trend of current performance dictating future performance. Very little significance was found between characteristics of programs (i.e. size, length) for both tiers. This makes it harder to distinguish the differences between the two. Overall, the predictor variables seem to be consistent between both perspectives, although the significance varies among them.

2 – *Are there differences in cost growth trends from a cross-sectional perspective, than that of the overall program?*

With respect to both Tier I and Tier II evaluations, the color ratings that were defined, and the available data, the answer is yes. 47 total relationships were found to be significant at $\alpha=0.10$. 65% of the significant predictor variables were color ratings of previous performance. This highlights the fact that, while there are some program traits that lend to a given cost growth pattern, the overwhelming response of cost growth is previous performance. This reiterates previous studies in that the importance lies with making maximum effort to minimize cost growth early in a programs lifecycle, to ensure minimal cost growth in the future.

Findings

The biggest finding was the undisputed influence DTE has on cost growth relationships that are found to be significant. This is perhaps the big-picture take away from this research, in that it is a key indicator for management as to where cost growth performance has been, and where it will continue to go. DTE really appears as the crux of the program, solidifying the cost growth fate, as you may have it. These findings echo the previous work of Rosado (2011) which uses regression analysis to show DT&E as a level 3 WBS element is a significant driver for overall program EAC growth.

However, when you look at the overall results from a bigger picture perspective, both the predictor and response variables are fairly evenly distributed over the phase of the lifecycle impacted. This held true for both Tier I and Tier II findings. Figures 14 through 17 show the breakout of occurrences each program review appeared in either the predictor or response variable respectively. Note that the "Other" category for predictor responses accounts for any descriptive variable not defined as a program review point.

The ambiguity of predictor and response variables within significant findings, make this research hard to project a summary result for management use. The interweb of relationships makes for a very complex environment, making it unclear as to which factors can really help management project their color rating risk.

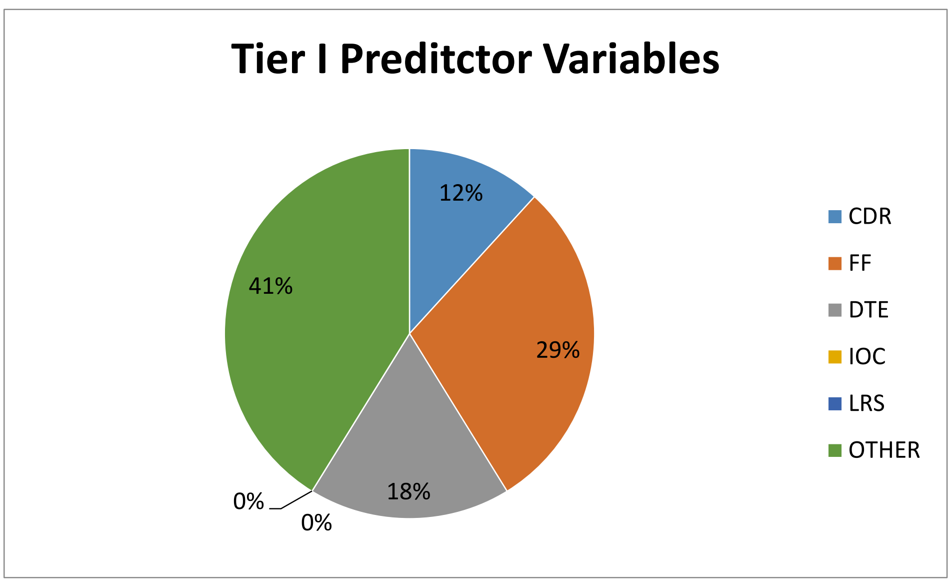


Figure 14: Tier I Predictor Variable Breakout

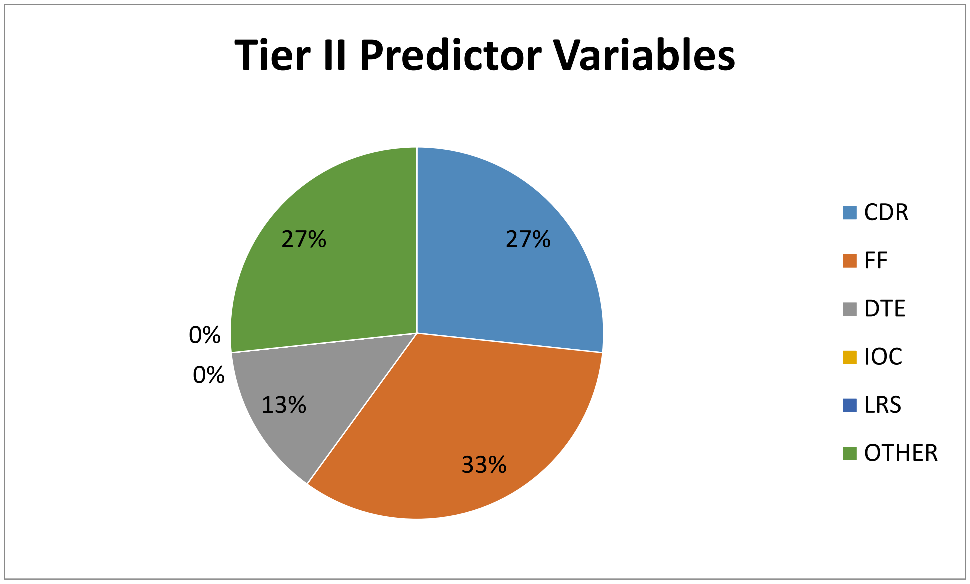


Figure 15: Tier II Predictor Variable Breakout

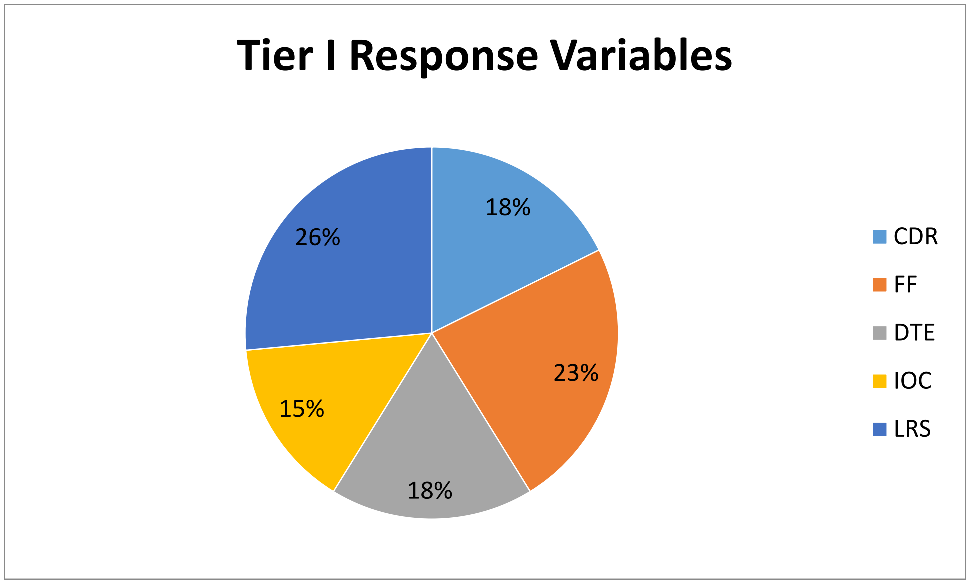


Figure 16: Tier I Response Variable Breakout

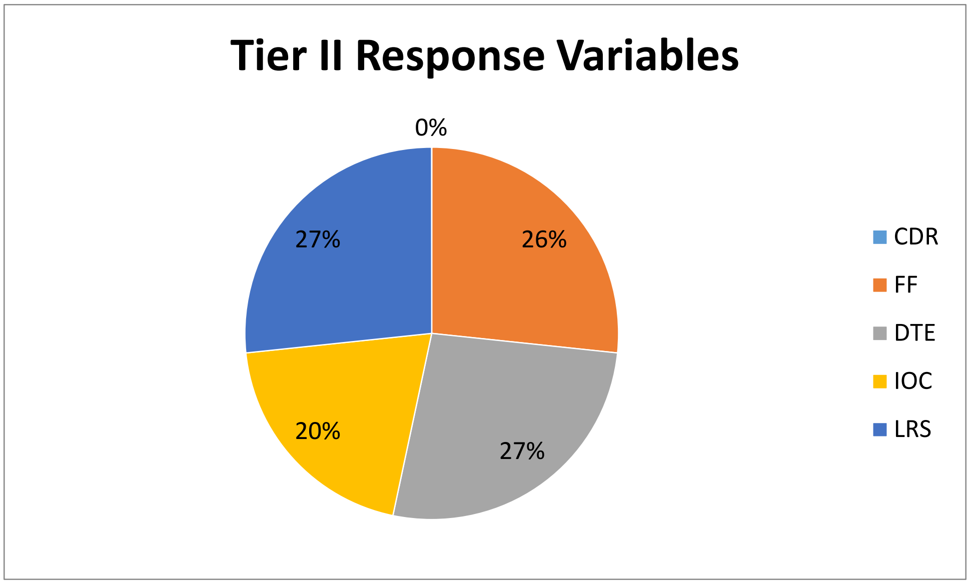


Figure 17: Tier II Response Variable Breakout

Limitations

We recognize several major limitations of this research, which may impact the applicability to real-world scenarios. First and foremost, the results are a product of the inputs. That being said, the limitations of our data collection filter through to weaker results. As mentioned earlier, the inherent issues with the SAR pass through to create issues within our database. Most notably, the magnitude of missing review dates really impacted the ability to analyze cost growth performance from a longitudinal perspective. Every missing date impacts at least two data points. With an already limited data set, this further reduces the sample size available at each testing interval.

For further research on this database, or for a continuation of its progress, a dive into additional sources to help complete and validate data would be essential in improving the results. It is recommended to visit sources of hardcopy SAR data to help with older program information. There were a variety of aircraft programs that were excluded due to the unavailability of data, which could be added in to the population given it can be located.

Another limitation to the analysis is the subjectivity of how the color ratings are applied. This research considers Green to represent cost growth of 1 or below. Many would argue that programs are still satisfied with risk if their cost growth falls just slightly above 1. The subjective nature can lead to varying results, potentially changing the analysis all together. Understanding this limitation is necessary in evaluating programs in question.

Recommendations for Future Research

Recommendations for future research encourage the exploration and use of the original SAR database, as well as our modified research database. Whereas our research is the first to explore predicting cost growth risk from a longitudinal perspective, it should be acknowledged that follow-on research and other methodologies used to predict cost growth should be encouraged. The cost estimating community can only estimate the known, but as this research has shown, there are many ways cost growth can be analyzed to help decision makers best prepare for known outcomes. Further exploration is highly encouraged to help refine the concept of short-term cost growth analysis. Ways in which this research can be carried forward include:

- Collect more SAR data to further populate our research database with more both more programs, and with population of missing data. This could help confirm the significance of predictor variables, and perhaps identify new ones as well.

- Perform analysis with a different cost rating scale. Perhaps a survey of program experts could help depict what real-world breakpoints would be useful for Green, Amber and Red distinguishers.

- Apply this research method to programs outside of Aircraft. While new program review markers would have to be identified (for instance, First Flight), the application of the longitudinal perspective to other programs would be interesting.

- Develop a statistical model for this analysis. While this research only carries as far as identifying the significant variables, it could be continued to develop a model for real-world application.

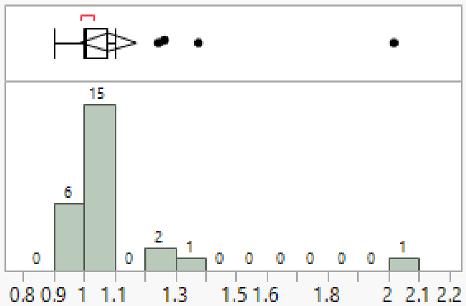
Chapter Summary

This chapter discusses the relevant findings, limitations, and future research opportunities available for continued analysis. This research serves as a stepping stone for future efforts. The findings contained within this research bridge a gap between previous longitudinal studies, particularly as it relates to a program. This research finds itself sandwiched between high-level timeline studies, and detailed EVM studies on individual contracts.

The application of color ratings to cost growth facilitates a new perspective and perhaps a more hands-on approach for management involvement. This research should serve as a tool to help bring cost growth analysis to the forefront of decision making, with a simple and easily identifiable system. While exploratory in nature, the analysis of variables against this color rating system has validated findings in previous cost growth studies, as well as highlighted new areas for further investigation.

Appendix A

CA-CDR CG



Quantiles

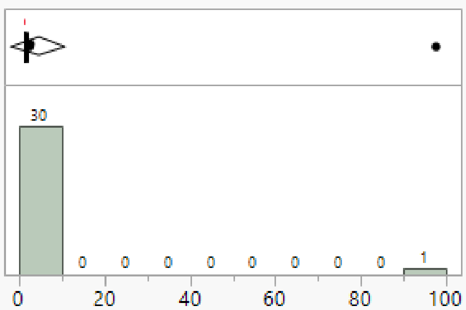
100.0%	maximum	2.0197
99.5%		2.0197
97.5%		2.0197
90.0%		1.30816
75.0%	quartile	1.07175
50.0%	median	1.0039
25.0%	quartile	0.99875
10.0%		0.93452
2.5%		0.9001
0.5%		0.9001
0.0%	minimum	0.9001

Summary Statistics

Mean	1.080792
Std Dev	0.222758
Std Err Mean	0.0445516
Upper 95% Mean	1.172742
Lower 95% Mean	0.988842
N	25

Figure A1: CGF Distribution Summary for CA-CDR

CG from CA-FF



Quantiles

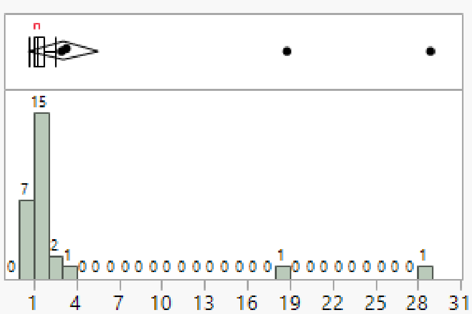
100.0%	maximum	97.8021
99.5%		97.8021
97.5%		97.8021
90.0%		2.15542
75.0%	quartile	1.2735
50.0%	median	1.0317
25.0%	quartile	0.9791
10.0%		0.88922
2.5%		0.8327
0.5%		0.8327
0.0%	minimum	0.8327

Summary Statistics

Mean	4.2873871
Std Dev	17.359525
Std Err Mean	3.1178627
Upper 95% Mean	10.654912
Lower 95% Mean	-2.080138
N	31

Figure A2: CGF Distribution Summary for CA-FF

CG from CA-DT



Quantiles

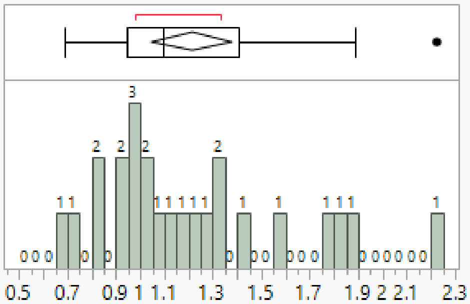
100.0%	maximum	28.9083
99.5%		28.9083
97.5%		28.9083
90.0%		6.38026
75.0%	quartile	1.6981
50.0%	median	1.2105
25.0%	quartile	0.9897
10.0%		0.87328
2.5%		0.6939
0.5%		0.6939
0.0%	minimum	0.6939

Summary Statistics

Mean	3.0532185
Std Dev	6.188453
Std Err Mean	1.1909683
Upper 95% Mean	5.501289
Lower 95% Mean	0.6051481
N	27

Figure A3: CGF Distribution Summary for CA-DT

CG from CA-IOC



Quantiles

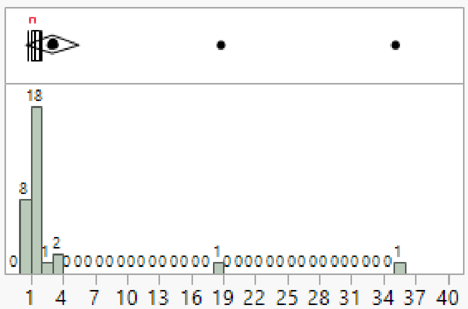
100.0%	maximum	2.2253
99.5%		2.2253
97.5%		2.2253
90.0%		1.85155
75.0%	quartile	1.406675
50.0%	median	1.09045
25.0%	quartile	0.946575
10.0%		0.7674
2.5%		0.6883
0.5%		0.6883
0.0%	minimum	0.6883

Summary Statistics

Mean	1.2135125
Std Dev	0.3973912
Std Err Mean	0.0811171
Upper 95% Mean	1.3813161
Lower 95% Mean	1.0457089
N	24

Figure A4: CGF Distribution Summary for CA-IOC

CG from CA-LRS



Quantiles

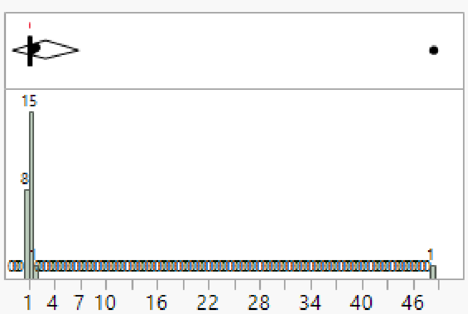
100.0%	maximum	35.1348
99.5%		35.1348
97.5%		35.1348
90.0%		3.15842
75.0%	quartile	1.7179
50.0%	median	1.2482
25.0%	quartile	0.9821
10.0%		0.84188
2.5%		0.7579
0.5%		0.7579
0.0%	minimum	0.7579

Summary Statistics

Mean	3.0606968
Std Dev	6.7508684
Std Err Mean	1.2124918
Upper 95% Mean	5.5369353
Lower 95% Mean	0.5844583
N	31

Figure A5: CGF Distribution Summary for CA-LRS

CG from CDR-FF



Quantiles

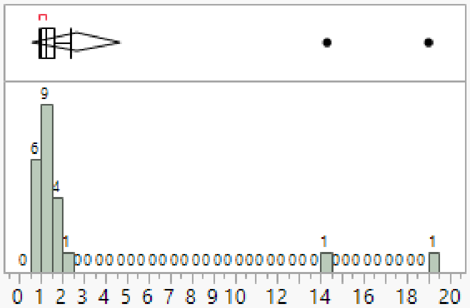
100.0%	maximum	48.4241
99.5%		48.4241
97.5%		48.4241
90.0%		1.60444
75.0%	quartile	1.10325
50.0%	median	1.0078
25.0%	quartile	0.99075
10.0%		0.94758
2.5%		0.8278
0.5%		0.8278
0.0%	minimum	0.8278

Summary Statistics

Mean	2.968064
Std Dev	9.4721322
Std Err Mean	1.8944264
Upper 95% Mean	6.877968
Lower 95% Mean	-0.94184
N	25

Figure A6: CGF Distribution Summary for CDR-FF

CG from CDR-DTE



Quantiles

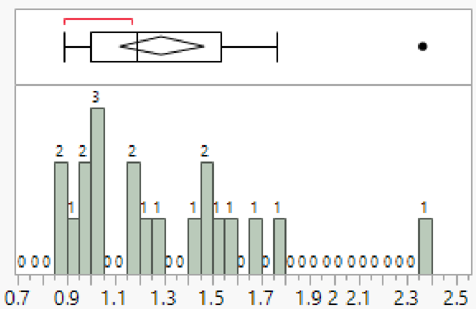
100.0%	maximum	19.0394
99.5%		19.0394
97.5%		19.0394
90.0%		10.72886
75.0%	quartile	1.570325
50.0%	median	1.19165
25.0%	quartile	0.98995
10.0%		0.89992
2.5%		0.863
0.5%		0.863
0.0%	minimum	0.863

Summary Statistics

Mean	2.6508318
Std Dev	4.6112193
Std Err Mean	0.9831153
Upper 95% Mean	4.6953319
Lower 95% Mean	0.6063317
N	22

Figure A7: CGF Distribution Summary for CDR-DTE

CG from CDR-IOC



Quantiles

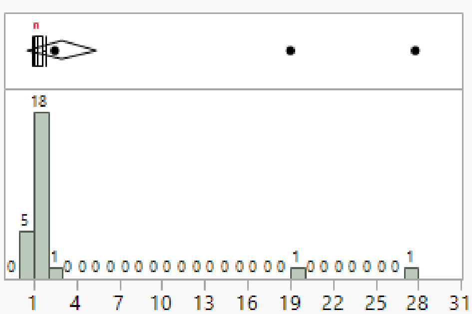
100.0%	maximum	2.3654
99.5%		2.3654
97.5%		2.3654
90.0%		1.75739
75.0%	quartile	1.53245
50.0%	median	1.1865
25.0%	quartile	0.99955
10.0%		0.89979
2.5%		0.8878
0.5%		0.8878
0.0%	minimum	0.8878

Summary Statistics

Mean	1.29155
Std Dev	0.3740955
Std Err Mean	0.0836503
Upper 95% Mean	1.4666321
Lower 95% Mean	1.1164679
N	20

Figure A8: CGF Distribution Summary for CDR-IOC

CG from CDR-LRS



Quantiles

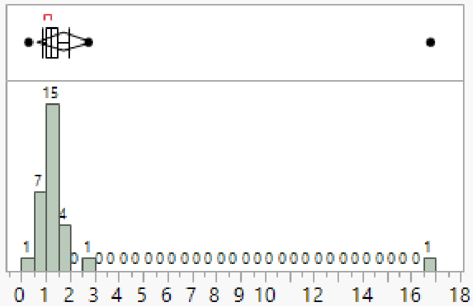
100.0%	maximum	27.8098
99.5%		27.8098
97.5%		27.8098
90.0%		7.4404
75.0%	quartile	1.528025
50.0%	median	1.2375
25.0%	quartile	1.0186
10.0%		0.93606
2.5%		0.8777
0.5%		0.8777
0.0%	minimum	0.8777

Summary Statistics

Mean	2.97675
Std Dev	6.1557097
Std Err Mean	1.207234
Upper 95% Mean	5.463095
Lower 95% Mean	0.490405
N	26

Figure A9: CGF Distribution Summary for CDR-LRS

CG from FF-DTE



Quantiles

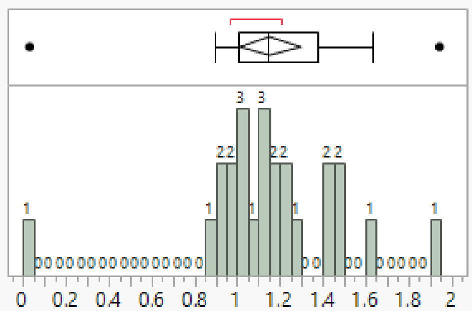
100.0%	maximum	16.8044
99.5%		16.8044
97.5%		16.8044
90.0%		1.9458
75.0%	quartile	1.4366
50.0%	median	1.1769
25.0%	quartile	0.99535
10.0%		0.9186
2.5%		0.2956
0.5%		0.2956
0.0%	minimum	0.2956

Summary Statistics

Mean	1.7576621
Std Dev	2.9261817
Std Err Mean	0.5433783
Upper 95% Mean	2.870722
Lower 95% Mean	0.6446021
N	29

Figure A10: CGF Distribution Summary for FF-DTE

CG from FF-IOC



Quantiles

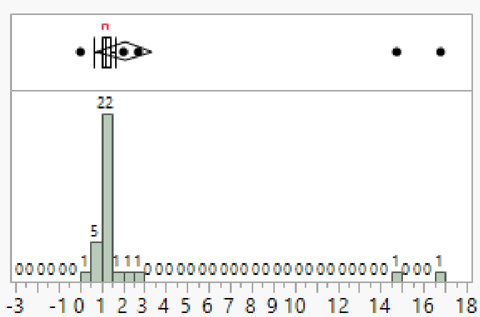
100.0%	maximum	1.9458
99.5%		1.9458
97.5%		1.9458
90.0%		1.5602
75.0%	quartile	1.376925
50.0%	median	1.14125
25.0%	quartile	1.00125
10.0%		0.9011
2.5%		0.0294
0.5%		0.0294
0.0%	minimum	0.0294

Summary Statistics

Mean	1.1548958
Std Dev	0.3480099
Std Err Mean	0.0710372
Upper 95% Mean	1.3018475
Lower 95% Mean	1.0079441
N	24

Figure A11: CGF Distribution Summary for FF-IOC

CG from FF-LRS



Quantiles

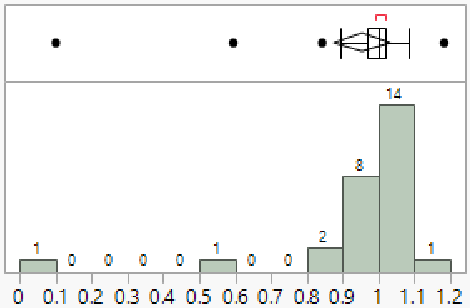
100.0%	maximum	16.8044
99.5%		16.8044
97.5%		16.8044
90.0%		2.44222
75.0%	quartile	1.39865
50.0%	median	1.1929
25.0%	quartile	1.0047
10.0%		0.90468
2.5%		0.0297
0.5%		0.0297
0.0%	minimum	0.0297

Summary Statistics

Mean	2.079697
Std Dev	3.5678946
Std Err Mean	0.6210907
Upper 95% Mean	3.3448174
Lower 95% Mean	0.8145766
N	33

Figure A12: CGF Distribution Summary for FF-LRS

CG from DTE-IOC



Quantiles

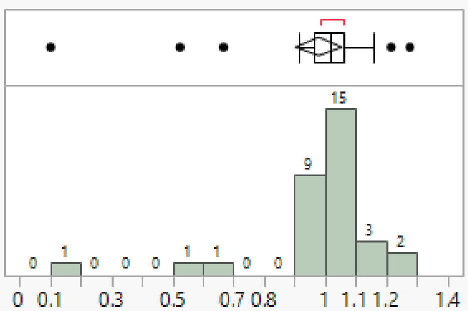
100.0%	maximum	1.1854
99.5%		1.1854
97.5%		1.1854
90.0%		1.07286
75.0%	quartile	1.0228
50.0%	median	1
25.0%	quartile	0.9692
10.0%		0.79448
2.5%		0.0994
0.5%		0.0994
0.0%	minimum	0.0994

Summary Statistics

Mean	0.9548222
Std Dev	0.19723
Std Err Mean	0.0379569
Upper 95% Mean	1.0328438
Lower 95% Mean	0.8768006
N	27

Figure A13: CGF Distribution Summary for DTE-IOC

CG from DTE-LRS



Quantiles

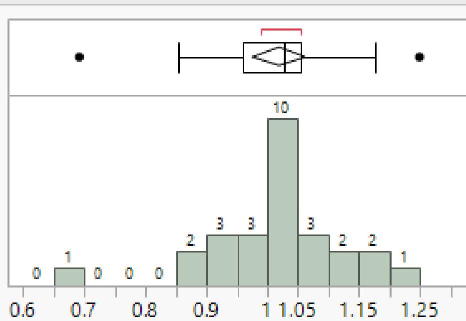
100.0%	maximum	1.2779
99.5%		1.2779
97.5%		1.2779
90.0%		1.15545
75.0%	quartile	1.059225
50.0%	median	1.01985
25.0%	quartile	0.9605
10.0%		0.74107
2.5%		0.1006
0.5%		0.1006
0.0%	minimum	0.1006

Summary Statistics

Mean	0.9813781
Std Dev	0.2111634
Std Err Mean	0.0373288
Upper 95% Mean	1.0575107
Lower 95% Mean	0.9052456
N	32

Figure A14: CGF Distribution Summary for DTE-LRS

CG from IOC-LRS



Quantiles

100.0%	maximum	1.2495
99.5%		1.2495
97.5%		1.2495
90.0%		1.15856
75.0%	quartile	1.0551
50.0%	median	1.027
25.0%	quartile	0.9601
10.0%		0.87602
2.5%		0.6917
0.5%		0.6917
0.0%	minimum	0.6917

Summary Statistics

Mean	1.0193593
Std Dev	0.1100117
Std Err Mean	0.0211718
Upper 95% Mean	1.0628784
Lower 95% Mean	0.9758401
N	27

Figure A15: CGF Distribution Summary for IOC-LRS

Appendix B

Table B1: Contingency Table Results for Red2 given Red

Contingency Table

Red 2 \ Red (Count / Expected)	0	1	Total
0	17 14.28	0 2.72	17
1	4 6.72	4 1.28	8
Total	21	4	25

Tests

N	DF	-LogLike	RSquare (U)
25	1	5.4465695	0.4955

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	10.893	0.0010*
Pearson	10.119	0.0015*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	1.0000	Prob(Red=1) is greater for Red 2=0 than 1
Right	0.0055*	Prob(Red=1) is greater for Red 2=1 than 0
2-Tail	0.0055*	Prob(Red=1) is different across Red 2

Table B2: Contingency Table Results for Green2 given Green

Contingency Table

Green 2 \ Green (Count / Expected)	0	1	Total
0	13 9.6	2 5.4	15
1	3 6.4	7 3.6	10
Total	16	9	25

Tests

N	DF	-LogLike	RSquare (U)
25	1	4.3366948	0.2655

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	8.673	0.0032*
Pearson	8.362	0.0038*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9997	Prob(Green=1) is greater for Green 2=0 than 1
Right	0.0065*	Prob(Green=1) is greater for Green 2=1 than 0
2-Tail	0.0090*	Prob(Green=1) is different across Green 2

Table B3: Contingency Table Results for Green5 Given Red3

Green 5

Contingency Table

Red 3

Green 5	Count Expected	0	1	Total
	0	5 7.77778	16 13.2222	21
	1	5 2.22222	1 3.77778	6
	Total	10	17	27

Tests

N	DF	-LogLike	RSquare (U)
27	1	3.5673883	0.2004

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	7.135	0.0076*
Pearson	7.090	0.0078*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.0152*	Prob(Red 3=1) is greater for Green 5=0 than 1
Right	0.9993	Prob(Red 3=1) is greater for Green 5=1 than 0
2-Tail	0.0152*	Prob(Red 3=1) is different across Green 5

Table B4: Contingency Table Results for Red3 given Green2

Contingency Table

Green 2

Red 3	Count Expected	0	1	Total
	0	3 6.2963	7 3.7037	10
	1	14 10.7037	3 6.2963	17
	Total	17	10	27

Tests

N	DF	-LogLike	RSquare (U)
27	1	3.7664872	0.2116

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	7.533	0.0061*
Pearson	7.400	0.0065*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.0104*	Prob(Green 2=1) is greater for Red 3=0 than 1
Right	0.9993	Prob(Green 2=1) is greater for Red 3=1 than 0
2-Tail	0.0127*	Prob(Green 2=1) is different across Red 3

Table B5: Contingency Table Results for Red5 given Modification

Red 5

Contingency Table

Modification

Red 5: Count / Expected	0	1	Total
0	2 5.33333	8 4.66667	10
1	14 10.6667	6 9.33333	20
Total	16	14	30

Tests

N	DF	-LogLike	RSquare (U)
30	1	3.5063890	0.1692

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	7.013	0.0081*
Pearson	6.696	0.0097*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.0131*	Prob(Modification=1) is greater for Red 5=0 than 1
Right	0.9989	Prob(Modification=1) is greater for Red 5=1 than 0
2-Tail	0.0187*	Prob(Modification=1) is different across Red 5

Table B6: Contingency Table Results for Red4 given Red2

Red 4

Contingency Table

Red 2

Red 4: Count / Expected	0	1	Total
0	11 8	1 4	12
1	5 8	7 4	12
Total	16	8	24

Tests

N	DF	-LogLike	RSquare (U)
24	1	3.6839891	0.2412

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	7.368	0.0066*
Pearson	6.750	0.0094*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9993	Prob(Red 2=1) is greater for Red 4=0 than 1
Right	0.0136*	Prob(Red 2=1) is greater for Red 4=1 than 0
2-Tail	0.0272*	Prob(Red 2=1) is different across Red 4

Table B7: Contingency Table Results for Red given Large

Contingency Table

Red \ Large: Count / Expected	0	1	Total
0	19 16.8	2 4.2	21
1	1 3.2	3 0.8	4
Total	20	5	25

Tests

N	DF	-LogLike	RSquare (U)
25	1	3.6563838	0.2923

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	7.313	0.0068*
Pearson	9.003	0.0027*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9996	Prob(Large=1) is greater for Red=0 than 1
Right	0.0162*	Prob(Large=1) is greater for Red=1 than 0
2-Tail	0.0162*	Prob(Large=1) is different across Red

Table B8: Contingency Table Results for Red2 given Prototype

Contingency Table

Red 2 \ Prototype: Count / Expected	0	1	Total
0	14 10.7667	5 8.23333	19
1	3 6.23333	8 4.76667	11
Total	17	13	30

Tests

N	DF	-LogLike	RSquare (U)
30	1	3.1311257	0.1525

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	6.262	0.0123*
Pearson	6.111	0.0134*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9981	Prob(Prototype=1) is greater for Red 2=0 than 1
Right	0.0179*	Prob(Prototype=1) is greater for Red 2=1 than 0
2-Tail	0.0227*	Prob(Prototype=1) is different across Red 2

Table B9: Contingency Table Results for Green5 given Green2

Green 5

Contingency Table

Green 2

Green 5: Count / Expected	0	1	Total
0	17 14.0968	6 8.90323	23
1	2 4.90323	6 3.09677	8
Total	19	12	31

Tests

N	DF	-LogLike	RSquare (U)
31	1	2.9905185	0.1445

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	5.981	0.0145*
Pearson	5.985	0.0144*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9980	Prob(Green 2=1) is greater for Green 5=0 than 1
Right	0.0220*	Prob(Green 2=1) is greater for Green 5=1 than 0
2-Tail	0.0316*	Prob(Green 2=1) is different across Green 5

Table B10: Contingency Table Results for Green5 given Green3

Green 5

Contingency Table

Green 3

Green 5: Count / Expected	0	1	Total
0	18 15.5556	3 5.44444	21
1	2 4.44444	4 1.55556	6
Total	20	7	27

Tests

N	DF	-LogLike	RSquare (U)
27	1	3.0200512	0.1955

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	6.040	0.0140*
Pearson	6.667	0.0098*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9986	Prob(Green 3=1) is greater for Green 5=0 than 1
Right	0.0239*	Prob(Green 3=1) is greater for Green 5=1 than 0
2-Tail	0.0239*	Prob(Green 3=1) is different across Green 5

Table B11: Contingency Table Results for Red5 given Green

Red 5

Contingency Table

Green

Count / Expected		0	1	Total
Red 5	0	3	6	9
		5.76	3.24	
	1	13	3	16
		10.24	5.76	
	Total	16	9	25

Tests

N	DF	-LogLike	RSquare (U)
25	1	2.8855863	0.1766

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	5.771	0.0163*
Pearson	5.740	0.0166*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.0252*	Prob(Green=1) is greater for Red 5=0 than 1
Right	0.9978	Prob(Green=1) is greater for Red 5=1 than 0
2-Tail	0.0308*	Prob(Green=1) is different across Red 5

Table B12: Contingency Table Results for Red2 given Medium

Contingency Table

Med

Count / Expected		0	1	Total
Red 2	0	10	10	20
		12.9032	7.09677	
	1	10	1	11
		7.09677	3.90323	
	Total	20	11	31

Tests

N	DF	-LogLike	RSquare (U)
31	1	2.9481692	0.1462

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	5.896	0.0152*
Pearson	5.188	0.0227*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.0260*	Prob(Med=1) is greater for Red 2=0 than 1
Right	0.9980	Prob(Med=1) is greater for Red 2=1 than 0
2-Tail	0.0472*	Prob(Med=1) is different across Red 2

Table B13: Contingency Table Results for Red5 given Red2

Red 5

Contingency Table

Red 2

Red 5: Count / Expected	0	1	Total
0	10 7.09677	1 3.90323	11
1	10 12.9032	10 7.09677	20
Total	20	11	31

Tests

N	DF	-LogLike	RSquare (U)
31	1	2.9481692	0.1462

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	5.896	0.0152*
Pearson	5.188	0.0227*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9980	Prob(Red 2=1) is greater for Red 5=0 than 1
Right	0.0260*	Prob(Red 2=1) is greater for Red 5=1 than 0
2-Tail	0.0472*	Prob(Red 2=1) is different across Red 5

Table B14: Contingency Table Results for Amber2 given Amber

Contingency Table

Amber

Amber 2: Count / Expected	0	1	Total
0	12 9.36	6 8.64	18
1	1 3.64	6 3.36	7
Total	13	12	25

Tests

N	DF	-LogLike	RSquare (U)
25	1	2.9806049	0.1722

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	5.961	0.0146*
Pearson	5.540	0.0186*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9984	Prob(Amber=1) is greater for Amber 2=0 than 1
Right	0.0266*	Prob(Amber=1) is greater for Amber 2=1 than 0
2-Tail	0.0302*	Prob(Amber=1) is different across Amber 2

Table B15: Contingency Table Results for Red3 given <180 Months

Contingency Table

Red 3 (columns) by Months <180 (rows)

Count / Expected	0	1	Total
0	1 3.7037	9 6.2963	10
1	9 6.2963	8 10.7037	17
Total	10	17	27

Tests

N	DF	-LogLike	RSquare (U)
27	1	2.7922145	0.1569

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	5.584	0.0181*
Pearson	4.979	0.0257*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.0311*	Prob(Red 3=1) is greater for Months <180=0 than 1
Right	0.9977	Prob(Red 3=1) is greater for Months <180=1 than 0
2-Tail	0.0415*	Prob(Red 3=1) is different across Months <180

Table B16: Contingency Table Results for Green3 given Red2

Contingency Table

Red 2 (columns) by Green 3 (rows)

Count / Expected	0	1	Total
0	11 13.3333	9 6.66667	20
1	7 4.66667	0 2.33333	7
Total	18	9	27

Tests

N	DF	-LogLike	RSquare (U)
27	1	3.4231063	0.1992

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	6.846	0.0089*
Pearson	4.725	0.0297*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.0358*	Prob(Red 2=1) is greater for Green 3=0 than 1
Right	1.0000	Prob(Red 2=1) is greater for Green 3=1 than 0
2-Tail	0.0593	Prob(Red 2=1) is different across Green 3

Table B17: Contingency Table Results for Green2 given Prototype

Contingency Table

Green 2 \ Prototype	Count / Expected	0	1	Total
	0	8	11	19
		10.7667	8.23333	
	1	9	2	11
		6.23333	4.76667	
	Total	17	13	30

Tests

N	DF	-LogLike	RSquare (U)
30	1	2.3794602	0.1159

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	4.759	0.0291*
Pearson	4.474	0.0344*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.0396*	Prob(Prototype=1) is greater for Green 2=0 than 1
Right	0.9951	Prob(Prototype=1) is greater for Green 2=1 than 0
2-Tail	0.0575	Prob(Prototype=1) is different across Green 2

Table B18: Contingency Table Results for Green3 given Amber2

Contingency Table

Green 3 \ Amber 2	Count / Expected	0	1	Total
	0	12	8	20
		14.0741	5.92593	
	1	7	0	7
		4.92593	2.07407	
	Total	19	8	27

Tests

N	DF	-LogLike	RSquare (U)
27	1	2.9474891	0.1796

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	5.895	0.0152*
Pearson	3.979	0.0461*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.0567	Prob(Amber 2=1) is greater for Green 3=0 than 1
Right	1.0000	Prob(Amber 2=1) is greater for Green 3=1 than 0
2-Tail	0.0681	Prob(Amber 2=1) is different across Green 3

Table B19: Contingency Table Results for Amber given Pre-1997

Contingency Table

Pre 1997 (columns) by Amber (rows)

Count / Expected	0	1	Total
0	5 7.28	8 5.72	13
1	9 6.72	3 5.28	12
Total	14	11	25

Tests

N	DF	-LogLike	RSquare (U)
25	1	1.7386035	0.1014

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	3.477	0.0622
Pearson	3.381	0.0660

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.0749	Prob(Pre 1997=1) is greater for Amber=0 than 1
Right	0.9886	Prob(Pre 1997=1) is greater for Amber=1 than 0
2-Tail	0.1107	Prob(Pre 1997=1) is different across Amber

Table B20: Contingency Table Results for Amber2 given Medium

Contingency Table

Med (columns) by Amber 2 (rows)

Count / Expected	0	1	Total
0	17 14.8387	6 8.16129	23
1	3 5.16129	5 2.83871	8
Total	20	11	31

Tests

N	DF	-LogLike	RSquare (U)
31	1	1.6684207	0.0828

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	3.337	0.0677
Pearson	3.438	0.0637

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9880	Prob(Med=1) is greater for Amber 2=0 than 1
Right	0.0787	Prob(Med=1) is greater for Amber 2=1 than 0
2-Tail	0.0947	Prob(Med=1) is different across Amber 2

Table B21: Contingency Table Results for Amber given Navy

Contingency Table

NV

Amber	Count / Expected	0	1	Total
0	Count	10	3	13
	Expected	7.8	5.2	
1	Count	5	7	12
	Expected	7.2	4.8	
Total		15	10	25

Tests

N	DF	-LogLike	RSquare (U)
25	1	1.6523186	0.0982

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	3.305	0.0691
Pearson	3.232	0.0722

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9873	Prob(NV=1) is greater for Amber=0 than 1
Right	0.0820	Prob(NV=1) is greater for Amber=1 than 0
2-Tail	0.1107	Prob(NV=1) is different across Amber

Table B22: Contingency Table Results for Green4 given Red2

Contingency Table

Red 2

Green 4	Count / Total % / Col % / Row %	0	1	Total
0	Count	8	7	15
	Total %	33.33	29.17	62.50
	Col %	50.00	87.50	
	Row %	53.33	46.67	
1	Count	8	1	9
	Total %	33.33	4.17	37.50
	Col %	50.00	12.50	
	Row %	88.89	11.11	
Total		16	8	24
		66.67	33.33	

Tests

N	DF	-LogLike	RSquare (U)
24	1	1.7730015	0.1161

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	3.546	0.0597
Pearson	3.200	0.0736

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.0875	Prob(Red 2=1) is greater for Green 4=0 than 1
Right	0.9913	Prob(Red 2=1) is greater for Green 4=1 than 0
2-Tail	0.1782	Prob(Red 2=1) is different across Green 4

Table B23: Contingency Table Results for Green4 given Red3

Green 4

Contingency Table

Red 3

Green 4: Count / Expected	0	1	Total
0	3	12	15
	5	10	
1	5	4	9
	3	6	
Total	8	16	24

Tests

N	DF	-LogLike	RSquare (U)
24	1	1.5876495	0.1039

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	3.175	0.0748
Pearson	3.200	0.0736

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.0907	Prob(Red 3=1) is greater for Green 4=0 than 1
Right	0.9873	Prob(Red 3=1) is greater for Green 4=1 than 0
2-Tail	0.0994	Prob(Red 3=1) is different across Green 4

Table B24: Contingency Table Results for Red4 given Green2

Red 4

Contingency Table

Green 2

Red 4: Count / Expected	0	1	Total
0	6	6	12
	8	4	
1	10	2	12
	8	4	
Total	16	8	24

Tests

N	DF	-LogLike	RSquare (U)
24	1	1.5518394	0.1016

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	3.104	0.0781
Pearson	3.000	0.0833

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.0965	Prob(Green 2=1) is greater for Red 4=0 than 1
Right	0.9864	Prob(Green 2=1) is greater for Red 4=1 than 0
2-Tail	0.1930	Prob(Green 2=1) is different across Red 4

Table B25: Contingency Table Results for Red4 given Red3

Red 4

Contingency Table

Red 3

Red 4	Count / Expected	0	1	Total
0	Count	6	6	12
	Expected	4	8	
1	Count	2	10	12
	Expected	4	8	
Total		8	16	24

Tests

N	DF	-LogLike	RSquare (U)
24	1	1.5518394	0.1016

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	3.104	0.0781
Pearson	3.000	0.0833

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9864	Prob(Red 3=1) is greater for Red 4=0 than 1
Right	0.0965	Prob(Red 3=1) is greater for Red 4=1 than 0
2-Tail	0.1930	Prob(Red 3=1) is different across Red 4

Table B26: Contingency Table Results for Green given Pre-1997

Contingency Table

Pre 1997

Green	Count / Expected	0	1	Total
0	Count	11	5	16
	Expected	8.96	7.04	
1	Count	3	6	9
	Expected	5.04	3.96	
Total		14	11	25

Tests

N	DF	-LogLike	RSquare (U)
25	1	1.4822355	0.0864

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	2.964	0.0851
Pearson	2.932	0.0868

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9841	Prob(Pre 1997=1) is greater for Green=0 than 1
Right	0.0982	Prob(Pre 1997=1) is greater for Green=1 than 0
2-Tail	0.1153	Prob(Pre 1997=1) is different across Green

Table B27: Contingency Table Results for Red2 given Large

Contingency Table

Red 2	Count / Expected	Large 0	Large 1	Total
0		18 16.129	2 3.87097	20
1		7 8.87097	4 2.12903	11
Total		25	6	31

Tests

N	DF	-LogLike	RSquare (U)
31	1	1.5191919	0.0997

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	3.038	0.0813
Pearson	3.160	0.0755

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9868	Prob(Large=1) is greater for Red 2=0 than 1
Right	0.0983	Prob(Large=1) is greater for Red 2=1 than 0
2-Tail	0.1510	Prob(Large=1) is different across Red 2

Appendix C

Table C1: Contingency Table Results for Green10 given Air Force

Contingency Table

AF	Count / Expected	Green 10: 0	Green 10: 1	Total
	0	7 9.41379	6 3.58621	13
	1	14 11.5862	2 4.41379	16
	Total	21	8	29

Tests

N	DF	-LogLike	RSquare (U)
29	1	2.0803392	0.1218

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	4.161	0.0414*
Pearson	4.067	0.0437*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.0547	Prob(Green 10=1) is greater for AF=0 than 1
Right	0.9933	Prob(Green 10=1) is greater for AF=1 than 0
2-Tail	0.0923	Prob(Green 10=1) is different across AF

Table C2: Contingency Table Results for Green10 given Navy

Contingency Table

NV	Count / Expected	Green 10: 0	Green 10: 1	Total
	0	14 11.5862	2 4.41379	16
	1	7 9.41379	6 3.58621	13
	Total	21	8	29

Tests

N	DF	-LogLike	RSquare (U)
29	1	2.0803392	0.1218

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	4.161	0.0414*
Pearson	4.067	0.0437*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9933	Prob(Green 10=1) is greater for NV=0 than 1
Right	0.0547	Prob(Green 10=1) is greater for NV=1 than 0
2-Tail	0.0923	Prob(Green 10=1) is different across NV

Table C3: Contingency Table Results for Red10 given Green6

Contingency Table

Red 10

Green 6 — Count / Expected	0	1	Total
0	4 6.19048	9 6.80952	13
1	6 3.80952	2 4.19048	8
Total	10	11	21

Tests

N	DF	-LogLike	RSquare (U)
21	1	2.0094481	0.1383

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	4.019	0.0450*
Pearson	3.884	0.0487*

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.0635	Prob(Red 10=1) is greater for Green 6=0 than 1
Right	0.9933	Prob(Red 10=1) is greater for Green 6=1 than 0
2-Tail	0.0805	Prob(Red 10=1) is different across Green 6

Table C4: Contingency Table Results for Green15 given Modification

Contingency Table

Green 15

Modification — Count / Expected	0	1	Total
0	13 10.6667	3 5.33333	16
1	5 7.33333	6 3.66667	11
Total	18	9	27

Tests

N	DF	-LogLike	RSquare (U)
27	1	1.8855399	0.1097

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	3.771	0.0521
Pearson	3.759	0.0525

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9910	Prob(Green 15=1) is greater for Modification=0 than 1
Right	0.0642	Prob(Green 15=1) is greater for Modification=1 than 0
2-Tail	0.0969	Prob(Green 15=1) is different across Modification

Table C5: Contingency Table Results for Green13 given Green6

Green 6

Contingency Table

Green 13

Green 6	Count Expected	0	1	Total
	0	3 5.05263	9 6.94737	12
	1	5 2.94737	2 4.05263	7
	Total	8	11	19

Tests

N	DF	-LogLike	RSquare (U)
19	1	1.9960514	0.1544

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	3.992	0.0457*
Pearson	3.909	0.0480*

Warning: Average cell count less than 5, LR ChiSquare suspect.

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.0674	Prob(Green 13=1) is greater for Green 6=0 than 1
Right	0.9937	Prob(Green 13=1) is greater for Green 6=1 than 0
2-Tail	0.0739	Prob(Green 13=1) is different across Green 6

Table C6: Contingency Table Results for Green6 given Amber

Contingency Table

Amber

Green 6	Count Expected	0	1	Total
	0	5 7.28	9 6.72	14
	1	8 5.72	3 5.28	11
	Total	13	12	25

Tests

N	DF	-LogLike	RSquare (U)
25	1	1.7386035	0.1004

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	3.477	0.0622
Pearson	3.381	0.0660

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.0749	Prob(Amber=1) is greater for Green 6=0 than 1
Right	0.9886	Prob(Amber=1) is greater for Green 6=1 than 0
2-Tail	0.1107	Prob(Amber=1) is different across Green 6

Table C7: Contingency Table Results for Green13 given Red6

Contingency Table

Green 13

Red 6	Count Expected	0	1	Total
	0	8 6.31579	7 8.68421	15
	1	0 1.68421	4 2.31579	4
	Total	8	11	19

Tests

N	DF	-LogLike	RSquare (U)
19	1	2.5681106	0.1986

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	5.136	0.0234*
Pearson	3.685	0.0549

Warning: Average cell count less than 5, LR ChiSquare suspect.

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	1.0000	Prob(Green 13=1) is greater for Red 6=0 than 1
Right	0.0851	Prob(Green 13=1) is greater for Red 6=1 than 0
2-Tail	0.1032	Prob(Green 13=1) is different across Red 6

Table C8: Contingency Table Results for Amber15 given Red6

Contingency Table

Amber 15

Red 6	Count Expected	0	1	Total
	0	8 6.31579	7 8.68421	15
	1	0 1.68421	4 2.31579	4
	Total	8	11	19

Tests

N	DF	-LogLike	RSquare (U)
19	1	2.5681106	0.1986

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	5.136	0.0234*
Pearson	3.685	0.0549

Warning: Average cell count less than 5, LR ChiSquare suspect.

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	1.0000	Prob(Amber 15=1) is greater for Red 6=0 than 1
Right	0.0851	Prob(Amber 15=1) is greater for Red 6=1 than 0
2-Tail	0.1032	Prob(Amber 15=1) is different across Red 6

Table C9: Contingency Table Results for Amber15 given Green10

Green 10

Contingency Table

Green 10 \ Amber 15 (Count / Expected)	0	1	Total
0	7	11	18
	9	9	
1	5	1	6
	3	3	
Total	12	12	24

Tests

N	DF	-LogLike	RSquare (U)
24	1	1.9036925	0.1144

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	3.807	0.0510
Pearson	3.556	0.0593

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.0775	Prob(Amber 15=1) is greater for Green 10=0 than 1
Right	0.9931	Prob(Amber 15=1) is greater for Green 10=1 than 0
2-Tail	0.1550	Prob(Amber 15=1) is different across Green 10

Table C10: Contingency Table Results for Amber15 given Amber10

Amber 10

Contingency Table

Amber 10 \ Amber 15 (Count / Expected)	0	1	Total
0	11	7	18
	9	9	
1	1	5	6
	3	3	
Total	12	12	24

Tests

N	DF	-LogLike	RSquare (U)
24	1	1.9036925	0.1144

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	3.807	0.0510
Pearson	3.556	0.0593

Fisher's Exact Test	Prob	Alternative Hypothesis
Left	0.9931	Prob(Amber 15=1) is greater for Amber 10=0 than 1
Right	0.0775	Prob(Amber 15=1) is greater for Amber 10=1 than 0
2-Tail	0.1550	Prob(Amber 15=1) is different across Amber 10

Bibliography

Arena, Mark A., Leonard, Robert S., Murray, Shelia E., Younossi, Obaid, (2006). *Historical Cost Growth of Completed Weapon System Programs*, Santa Monica, CA: RAND Corporation.

Bolton, Joseph G., Robert S. Leonard, Mark B. Arena, Obaid Younossi, Jerry M. Sollinger (2008). *Sources of Weapon System Cost Growth: Analysis of 35 Major Defense Acquisition Programs*. Santa Monica, CA; RAND Corporation.

Calcutt, Harry M. (1993). *Cost Growth in DoD Major Programs: A Historical Perspective.* Executive Research Project. The Industrial College of the Armed Forces, National Defense University. Fort McNair Washington, DC.

Cancian, Col. Mark F. (2010). *Cost Growth: Perception and Reality*. Defense Acquisition University. Fort Belvoir, VA.

Christensen, David D. (1994). *Cost Overrun Optimism: Fact or Fiction*. Defense Acquisition University. Fort Belvoir, VA.

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

Christensen, D.S., & Payne, K. (1992). *Cost Performance Index Stability- Fact or Fiction?* Journal of Parametrics, 10, 27-40.

Davis, Dan & Philip S. Anton. (2016). *Annual Growth of Contract Costs for Major Programs in Development and Early Production.* Acquisition Policy Analysis Center. Washington, D.C.

Defense Acquisition University (DAU). (2015). *Glossary of Defense Acquisition Acronyms & Terms*. Retrieved from https://dap.dau.mil/glossary/pages/1382.aspx.

Deneve, Allen J., Erin T. Ryan, Jonathan D. Ritschel, Christine Schubert Kabban. (2015). *Taming the Hurricane of Acquisition Cost Growth – Or At Least Predicting It.* Defense ARJ, January 2015.

Drezner, Jeffrey A., Jeanne M. Jarvaise, Ronald Wayne Hess, Paul G. Hough, and D. Norton (1993). *An Analysis of Weapon System Cost Growth*, Santa Monica, CA: RAND Corporation, MR-291-AF.

Government Accountability Office. (2009). *Cost Estimating and Assessment Guide* (Report No. GAO-09-3SP). Washington, D.C.

Hough, Paul G. (1992). *Pitfalls in Calculating Cost Growth from Selected Acquisition Reports*, Santa Monica, CA: RAND Corporation, N-3136-AF.

Kozlak, S. (2016). *Predicting Cost Growth Using Program Reviews and Milestones for DOD Aircraft*. (Unpublished Master's Thesis).

McDonald, J.H. (2009). Handbook of Biological Statistics (2nd ed.) Sparky House Publishing, Baltimore, Maryland.

Petter, Jacob L., Jonathan D. Ritschel, Edward D. White (2015). *Stability Properties in Department of Defense Contracts: Answering the Controversy*. Journal of Public Procurement, 15(3), 341-364.

Porter, Gene, Brian Gladstone, C. Vance Gordon, Nicholas Karvonides, R. Royce Kneece, Jr., Jay Mandelbaum, William D. O'Neil (2009). *The Major Causes of Cost Growth in Defense Acquisitions.* Institute for Defense Analyses, Alexandria, VA.

Rosado, William R. (2011). *Comparison of Development Test and Evaluation and Overall Program Estimate at Completion.* MS thesis, AFIT/GCA/ENC/11-02. Air Force Institute of Technology (AU), Wright-Patterson AFB OH.

Rossetti, M.B., E.D. White (2004). *A Two Pronged Approach to Estimate Procurement Cost Growth in Major DoD Weapon Systems*, Journal of Cost Analysis and Management (Winter), 11-21.

Sipple, V.P., E.D. White, M.A. Greiner (2004). *Surveying Cost Growth*, Defense Acquisition Review Journal, 11(1), 78-91.

REPORT DOCUMENTATION PAGE			*Form Approved* *OMB No. 074-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
23-03-2017	Master's Thesis	Sep 2015 – March 2017

TITLE AND SUBTITLE	
4. A Longitudinal Study and Color Rating System of Acquisition Cost Growth	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
D'Amico, Cory N.	**5e. TASK NUMBER**
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/ENC) 2950 Hobson Way, Building 641 WPAFB OH 45433-8865	AFIT-ENC-MS-17-M-181
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)** **AFCAA**
Air Force Cost Analysis Agency 1500 West Perimeter Rd. Ste 3500 Joint Base Andrews NAF, MD 20762 Phone: (240) 612-5541; E-mail: scott.kozlak@us.af.mil ATTN: Capt Scott Kozlak	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION/AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

For decades, cost growth studies have been plentiful surrounding Department of Defense acquisitions. Many different angles have been looked at to try and discover how to better estimate cost, what causes cost growth, and how to mitigate it. This research addresses this through examining cost growth from a longitudinal perspective, evaluating cost growth factors at major program reviews, and assessing the cost growth by applying color rating metrics. The results of this analysis show that breaking cost growth into longitudinal segments of a programs lifecycle allows the true behavior of cost growth to be seen, when it can often be masked in the traditional approach of evaluating lifecycle cost growth. Additionally, when applying the proposed color rating system to cost growth factors, significant variables are found to have dependencies with cost growth factor color ratings. Significant relationships shown in the results were most commonly like-color predictor and response variables. Additionally DT&E is shown as a flag for high cost growth issues during a program lifecycle.

15. SUBJECT TERMS

Cost Analysis, Statistics, Regression Analysis, Defense Acquisitions, Schedule, Finance, Program Management

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** White, Edward D., Ph.D., AFIT/ENC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*
U	U	U	UU	101	(937) 255-3636, ext 4540 (NOT DSN) (edward.white@afit.edu)

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18