3-2006

Data Analysis and its Impact on Predicting Schedule & Cost Risk

Steven M. Cross

DATA ANALYSIS AND ITS IMPACT ON PREDICTING SCHEDULE & COST RISK

THESIS

Steven M. Cross, SMSgt, USAF

AFIT/GIR/ENC/06M-01

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GIR/ENC/06M-01

DATA ANALYSIS AND ITS IMPACT ON PREDICTING SCHEDULE & COST RISK

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 Information Resource Management

Steven M. Cross, BS

SMSgt, USAF

March 2006

AFIT/GIR/ENC/06M-01

DATA ANALYSIS AND ITS IMPACT ON PREDICTING SCHEDULE & COST RISK

Steven M. Cross, BS
SMSgt, USAF

Approved:

Edward D. White	16 Feb 06
Dr. Edward D. White (Chairman)	date
Kevin Lee Elder	16 FEB 06
Dr. Kevin Elder (Member)	date
	16 FEB 06
Marc T. Lewis, Captain, USAF (Member)	date

AFIT/GIR/ENC/06M-01

Abstract

Many databases rely on documents (research) of the past to input data to create a comprehensive database. The Selected Acquisition Report (SAR) is one such document. The SARs are pervasive documents that have undergone decades of scrutiny by Congress and watchdog organizations such as the Government Accountability Office. Since the SAR has undergone such massive evolutionary changes, creating an accurate acquisition database presents a daunting task for the analyst and researcher alike. This research concerns itself with one such database.

From this prior research database, we look to fill in missing data. We first conduct a literature review to determine why we have database discontinuity. Indeed, a large part of our review entails SARs. We find there are no other complete sources for acquisition reporting. Also, the reason for missing data stems from 30 years of change within the Department of Defense administration. In addition, we grow concerned about the number of rebaselines a program undergoes in its lifecycle. Investigating this, we add the variable *# of Rebaselines* to schedule and cost regression models from prior research for statistical evaluation. We find our new variable is highly predictive with past schedule modeling but not predictive with prior cost modeling.

Acknowledgements

I want to first thank the Air Force and more specifically AFIT for the opportunity for an enlisted person to learn and grow academically; it's been a humbling experience. I subsequently would like to thank the OSD (AT&L) for their support allowing access to the SARs and pertinent information pertaining to the SARs that are not written down. Indeed, Mr. Larry Axtell and Chris Knoche from OSD were exceptionally valuable to this research. I also am indebted to Capt Robert Cardin with his insight associated with acquisition and database design. Another pivotal Air Force member that has helped me wrap my arms around acquisition is Capt James Monaco. I next would like to thank Dr. Edward White III my thesis advisor who is a gifted professor with integrity and character that has been encouraging through this entire process. I also thank Dr. Kevin Elder and Capt Marc Lewis who provided great insight and guidance to this research as readers.

I would be remiss if I did not thank my family. Especially I thank my wife for her selfless support; I would not be able to dig into this research without her. I also want to extend a special thank you to my children. My heart ached every time I had to cut their one-on-one time short and focus on this research. Finally, I want to thank my Lord Jesus Christ who gave me the opportunity to have life abundantly and a second chance to excel.

Steven M. Cross

Table of Contents

Page

Abstract .... **iv**

Acknowledgements .... **v**

Table of Contents .... **vi**

List of Figures .... **viii**

List of Tables .... **ix**

I. Introduction **1**

General Issue 1
Specific Issue 2
Scope and Limitations of the Study 3
Research Objectives 3
Thesis Overview 3

II. Literature Review **5**

Introduction 5
Reasons for Missing Data 5
Chapter Summary 40

III. Methodology **46**

Introduction 46
Search for Missing Variables Associated with Schedule Slippage 46
Data Assessment 48
Exploratory Data Analysis 52
Response Variables 53
Search for Predictor Variables 54
Logistic Regression 56
Multiple Regression 58
Chapter Summary 59

IV. Results and Discussion **61**

Chapter Overview 61
Preemptive Data Analysis (Schedule) 61
Multiple Regression Results (Schedule) 65

Validation Results (Schedule).. 73
Preemptive Data Analysis (Total RDT&E Cost).. 75
Multiple Regression Results (RDT&E Total Cost) 78
Validation Results (RDT&E Total Cost)... 85
Chapter Summary 86

V. Conclusions...88

Chapter Overview 88
Explanation of the Problem 88
Review of Literature 89
Review of Methodology 90
Restatement of Results.. 92
Limitations 94
Recommendations... 97
Possible Follow-on Thesis 100

Appendix A 102

List of Acronyms 102

Appendix B 103

List of Predictor Variables (From Sipple 2002) 103

Appendix C 105

List of Predictor Variables (From Monaco 2005)... 105

Appendix D 110

Final Model (Schedule)... 110

Appendix E 111

Validation Table (Schedule) 111

Appendix F 112

Final Model (Cost).. 112

Appendix G...113

(Monaco 2005) Explanation of the Independent Variables in Final Models............. 113

Bibliography 115

List of Figures

Figure **Page**

1. Stem-and Leaf Plot of Y (Schedule Variance %) 63
2. Preliminary Logistic Distribution of Y (Schedule) 64
3. Distribution of New Predictor Variable (Schedule) 64
4. Shapiro-Wilk Test for Normality - EMD Schedule Variance % Model (Schedule) 68
5. Cook's Distance Overlay Plot for Influential Data Points - EMD Schedule Variance % (Schedule) 69
6. Normality and Shapiro-Wilk with Program 125 Removed (Schedule) 70
7. Cook's Distance Overlay Plot with Program 125 Removed (schedule) 71
8. Parameter Estimates - VIF Scores- (Schedule) 72
9. Prediction Formula - Arrows Correspond with Formula Values and Parameter Estimates (Schedule) 72
10. Standardized Parameter Estimates of Variables [#-of Rebaselines, Actual Phase I + Planned EMD Length] (Schedule) 73
11. Stem and Leaf Distribution of Y (Cost) 76
12. Logistic Response Distribution Comparison (RDT&E Total Cost?) 77
13. Distribution of Predictor Variable - # of Rebaselines (Cost) 78
14. Distribution of Y and Residual Plot of Untransformed Model (Cost) 79
15. Distribution of Y and Residual Plot of Transformed Model (Cost) 80
16. Studentized Residuals - Distribution (Cost) 83
17. Cook's Distance Influence (Cost) 84
18. Review of Final Model Parameters (Schedule) 93
19. Multiple Regression Parameter Comparison between Genest (2004) & Cross (2006) 97
20. Multiple Regression Final Model - (Schedule) 110
21. Multiple Regression Final Model (Cost) 112

List of Tables

Table **Page**

1. Summary Table of Letter to Chairman of the Committee on Armed Services (1973) 12
2. 2005 GAO Report Problem/Solution Summary Table 21
3. (RAND) Summary Table of Most Common Changes to SAR up to 1992 25
4. (RAND) Summary Table of the SAR's Most Notable Problems up to 1992 27
5. CARS Summary Table of Significant Changes to the SAR (Mainly Congressional) 37
6. Significant Historical Literature Summary of the SAR (Part I).. 43
7. Significant Historical Literature Summary of the SAR (Part II) 44
8. Summary Table of Historical System Changes to the SAR.. 45
9. Past Research Significant Predictor Variables for Linear Regression 55
10. Past Research Significant Predictor Variables for Multiple Regression........................... 55
11. Monaco (2005) Final Model Variables.. 56
12. Monaco (2005) Multiple Regression Model.. 66
13. Final Model with # of *Rebaselines* 67
14. Multiple Regression Parameters Post Temporary Removal of Program 125 69
15. Multiple Regression Model Variables (Cost) 81
16. Comparison Regression Results of Monaco (2005) and Cross (2006) 95
17. 2005 GAO Report Problem/Solution Summary Table 99
18. 100% - Validation of Multiple Regression Model - EMD Schedule Variance % (Schedule) 111

DATA ANALYSIS AND ITS IMPACT ON PREDICTING SCHEDULE & COST RISK

I. Introduction

General Issue

Many databases used in research have missing data and disparity amongst the data points. Hence, the more data missing from the database, the less useful it is to the researcher. Many documents of the past are inputted by the hand of the researcher to create a comprehensive database. The Selected Acquisition Report (SAR) is one such document that dates back to the late sixties. Indeed, the SAR is a pervasive document that has undergone decades of change and scrutiny by Congress and watchdog organizations such as the Government Accountability Office (GAO). Since the SAR has undergone such massive evolutionary changes for over 30 years, creating an accurate acquisition database presents a daunting task for the analyst and researcher alike. This research concerns itself with one such database created by Sipple (2002:124-125) and revised by Monaco (2005).

Monaco's research pertains to a database derived from SARs; Monaco noted some limitations in his thesis starting on page 109. One such limitation pertained to the predictive model. Monaco needed a complete set of data in order for the statistical models to accurately predict the probability and magnitude of schedule growth within the time frame of the EMD phase of acquisition (defined as the interval between MS II and MS III). Monaco (2005) found that approximately 27% of programs that otherwise met the researcher's criteria did not have a reported value for one of the four necessary schedule dates, e.g. planned and actual dates for Milestone II and Milestone III. Of the programs missing the appropriate schedule dates, Planned MS II, Actual MS II, Planned MS III, and Actual MS III did not have complete data 56%, 28%, 72%, and 56% of the time respectively.

In addition, Monaco (2005) observed the following missing schedule dates that showed promise as possible predictor variables: First Unit Equipped (FUE), Preliminary Design Review (PDR), Production Contract Award (PCA), Critical Design Review (CDR), EMD Contract Award, and Initial Operating Capability (IOC). Due to the fact that the SARs contained missing schedule information, Monaco could not decompose the interval between MS II to MS III in order to create predictive models within smaller time frames. In particular, the FUE schedule date also appeared to be very predictable but only present in 19.4% of the programs (Monaco, 2005).

Specific Issue

The missing data in Monaco's thesis is the springboard for this research. This research determines if there is actually missing data in the SAR. If the data is missing, this research resolves to determine why it is missing. Once we determine why the data is missing, then we look for other venues or databases available that can fill in the holes. The primary objective of this research contends to build up the current database that Sipple (2002) and Monaco (2005) have derived and possibly add new variables to the model to enhance it.

Scope and Limitations of the Study

The Selected Acquisition Reports (SARs) are a collection of reports from major defense acquisition programs (MDAPs) e.g. programs that are in acquisition category (ACAT) I). These reports include key cost, technical, schedule, program variance, and other program information. The primary information that is gathered for this research depends on the accuracy and quality of these reports.

Another limitation pertains to the area of limited research material associated with the Selected Acquisition Report. A preliminary research examination reveals very little research directed solely on the Selected Acquisition Report. This research is also limited to only unclassified programs that complete the EMD phase of acquisition.

Research Objectives

The objective of this thesis is threefold. First, this research explores the reasons for the missing data. Second, this research attempts to build up the current database and add potential enhancing variables to it. Third, this investigation reruns the Monaco (2005) model and possibly other models. We then compare the results and determine if the model(s) remain stable and consistent with the original research. This thesis attempts to use the same logistic regression and multiple regression model(s) used in the research of Monaco and others to validate and compare results.

Thesis Overview

Finding answers to this research first requires literature review. This portion of the thesis, Chapter II, provides answers as to why there is missing data and helps us devise a suitable methodology to answer our research objectives. This section of research also aids to the possibility of adding new variables into the model. The literature review in Chapter II also provides a solid foundation for this research to formulate a suitable strategy to answer our research objectives in Chapter III. Indeed, in Chapter III we gather the data and the insight we gain in Chapter II to catapult this research into making some preliminary assumptions and framing a suitable methodology to answer our thesis. In Chapter IV, we test theory and validate it for accuracy against the schedule model of Monaco (2005) and the total RDT&E cost model of Genest (2004). Finally, in Chapter V, we provide conclusions to our thesis, and possible venues for follow-on research.

II. Literature Review

Introduction

Designing/Creating a major weapons system is complex and full of uncertainty. Analyzing the paper trail of the weapon system can be a daunting task if people, policy, and procedures are not consistent and meticulously followed. We find this disparity in multiple GAO reports and changing policy for more than 30 years. To fully understand the complexity of the information and documentation associated with a major weapons system, this research looks to historic documentation for some of the answers. Based on misinformation and data acquired from the selected acquisition report (SAR) and newly researched historical information and data, this thesis develops a foundation to enhance the database. We attempt to do this by filling in missing data or enhancing the data with new predictive variables. Chapter III covers this aforementioned methodology.

Reasons for Missing Data

This research looks for data across DoD to enhance the Monaco (2005) study on schedule slippage in an attempt to find its missing holes. In addition to looking forward to new data sources, this thesis also looks at the most commonly used data source for the analysis of major weapon systems, the SAR. As we describe in our methodology in Chapter III, we perform an exhaustive search for databases associated with MDAPs (major defense acquisition programs) and it proves to be futile. However, our analysis of the literature associated with the SAR supports the theory that the SAR is the only database comprehensive enough to provide the information we need for analysis. The following documents address some of these issues:

Letter to Chairman of the Committee on Armed Services, 1973

The Comptroller General prepared this document in response to a request from the office of the Chairman of the Committee on Armed Services. The request asked for a brief history including past and present recommendations and Department of Defense actions taken in response to the recommendations of the comptroller's office in relation to the DoD selected acquisition report (SAR). It is noteworthy to mention that there are numerous reproduction artifacts that make many portions of this document illegible. There are also stamps across the document that state “BEST DOCUMENT AVAILABLE.”

The letter starts with the SAR’s genesis and history. The historical portion noted the first DoD instruction to associate with the SAR. DoD instruction 7000.3 of February 1968 was the first authoritative guidance that established the mandatory SAR requirement. Before the introduction of the SAR, no recurring reports existed on major acquisitions for a comparison with prior and later estimates that retained cost, schedule, and performance data. The document went on to say the SAR system’s initial purpose focused on keeping its sponsor, the Assistant Secretary of Defense (Comptroller), apprised of the progress of selected acquisitions. Another primary purpose of the SAR dealt with comparing MDAP progress with planned technical, schedule, and cost performance.

DoD used the SAR in an experimental stage during 1968 and only required eight programs to be reported via the SAR. In early 1969, the Secretary of Defense established an objective that he be advised regularly of the status of major acquisitions. Concurrently, the Chairman of the Senate Arms Services Committee concluded that Congress should also be regularly informed of the progress of DoD acquisitions and requested periodic reports on such programs. After deliberation by all parties, they decided that SARs would be used to advise top DoD management and Congress of the progress of major acquisitions. As a result of this decision, the SAR became and remains the key recurring report from project managers and the services to inform the Secretary of Defense and Congress on the progress of its major acquisition programs.

The document went on to mention the improvements in the SAR since its inception. DoD, the office of the comptroller general, and other congressional committees championed these system improvements. The document also includes quotes from the House Committee on Armed Services in its report to Congress on April 24, 1970 (91st Congress, second session). We include the following quotes to show the essence of the SAR of the 70's:

> "With valuable suggestions made by the comptroller general, the SARs are being improved to the point where they can become a significant aid to better program management. The manner in which the SARs are presented to the committee, however, leaves much to be desired."
>
> * * * * * * * * *
>
> "The Department of Defense has sometimes arbitrarily eliminated statistical information or otherwise altered the material submitted to the Committee."
>
> * * * * * * * * *

"The committee is, likewise, disturbed by the timeliness with which the SARs are submitted to the Committee by the Department of Defense. In many cases the Committee has not received the SARs…until as much as three months after the close of the reporting period. This greatly lessens their effectiveness to the committee, particularly during the period when the annual authorization is being considered."

*   *   *   *   *   *   *   *   *

"In its attempt to gain a more detailed portrait of military spending, the committee has become concerned about the inconsistency of various reporting and estimating methods in relation to weapons costs" (Staats, 1973). This document went on to mention other recommendations to include improvements by DoD.

The first DoD improvement enhanced their definition of *costs* to include special vocabulary of the following four terms that should be uniformly applied and understood: "flyaway costs," "weapon system costs," "procurement costs," and "program acquisition costs." DoD intended to avoid any future discrepancies by clarifying the aforementioned terms to prevent the ambiguity associated by poorly defined definitions in times past.

DoD's next improvement pertained to the application of inflation. A report written in April 24, 1970 in response to the Secretary of Defense report of April 24, 1970, pointed out that some practical measure of inflationary trends was necessary so DoD issued a memorandum on June 30, 1970 titled Weapon System Costing. The memorandum stated that cost estimates should reflect the best estimates of the amounts ultimately to be paid, specifically incorporating anticipated changes in future prices. The estimates should also be accomplished on the basis of specific data applicable to a given system considering such factors as contract provisions, labor agreements, productivity, quantity changes, and the extent to which material is on hand or under fixed-price contract. The memorandum also stated that these policies would be set forth in the September 30, 1970 SARs.

Another DoD initiative made changes in data presentation. In 1970 and 1971 SARs were rather large, some reaching 60 pages or more. Since DoD recognized that management could not have time to review and analyze such large documents, they revised DoD instruction 7000.3 on September 13, 1971 to provide that no SAR would have more than 13 pages unless the assistant Secretary of Defense (Comptroller) granted a special waiver; DoD preferred 10 pages or less. In addition to the changes by DoD, the comptroller general also made some recommendations (Staats, 1973).

The comptroller first published a number of recommendations in a report in February 6, 1970 entitled "Status of the Acquisition of Selected Major Weapon Systems." The report recognized that the system presented a meaningful management tool for measuring and tracking the progress of major acquisitions. The office of the comptroller general did notice some serious shortcomings; they were missing: (1) a comparison of demonstrated performance with that specified in the contract, (2) the status of key subsystems essential to mission accomplishment (3) costs incurred in relationship to the costs planned to be incurred, (4) significant pending decisions that may affect the program, and (5) a comparison of quantities delivered with those scheduled to be delivered at the same time.

The comptroller general made a second review of the SAR system in August 1970 and published a report entitled "Acquisition of Major Weapon Systems" in

March 18, 1971. The report confirmed that improvements had been made since DoD issued the first report, however the SAR needed additional improvements. They concluded that SARs still did not: (1) contain a summary regarding overall acceptability of the weapon for its mission, (2) recognize the relationship of other weapon systems complementary to the system, or (3) reflect the status of programs (Staats, 1973).

In August 1971, the comptroller general generated a third review of the SAR system, which focused on evaluating its value to management. While DoD continued to improve the system, they identified two principal problems related to changing baselines for measuring progress and credibility of cost estimates. The GAO concluded that static baselines should be reported and maintained in the SAR and that DoD programs need to make complete and realistic cost estimates. The committee believed that both recommendations were essential in evaluating the progress of major acquisitions in making decisions on the systems future progress. In addition, they noticed a recurring problem associated with the undue delay in submitting SARs to top management through DoD. Another recurring problem concerned the limiting criteria for designating weapon systems far SAR reporting; it should be reassessed to improve management visibility on additional major weapon systems. The GAO published these findings in a report entitled "Acquisition of Major Weapon Systems," July 17, 1972 (Staats, 1973).

The report went on to mention GAO's next significant report between February and March 1973. The 1973 GAO report reviewed 68 staff studies to the Congress evaluating the SAR on applicable systems. The document went on to outline the following five recommendations:

1. Precise criteria should be established for adding and deleting major acquisitions.
2. Planning and development estimates that may change should not be deleted for any reason. SARs should contain a record of all estimates so that there is total visibility and track-ability from the program's inception.
3. Submitting SARs to top management through DoD are unduly delayed. For several years, SARs have been submitted to Congress nearly three months after the "as of " dates.
4. All program costs should be included. A number of systems under development include only research and development costs. Procurement costs are excluded and costs for these systems are therefore understated in SARs. Also, other systems are kept below the dollar criteria for consideration for SARs.
5. SARs should show a comparison of costs incurred, schedule milestones attained, and technical performance accomplished with what was planned for the same period of time and costs budgeted (Staats, 1973).

We include a summary table outlining the major changes associated with this document in Table 1.

Table 1. Summary Table of Letter to Chairman of the Committee on Armed Services (1973)

Letter to Chairman of the Committee on Armed Services (1973)	
Background History	February 1968 - first authoritative guidance that established the mandatory SAR requirement (DoD instruction 7000.3)
	1968 - the DoD used the SAR in an experimental stage
	1969 - Secretary of Defense established an objective to be advised regularly on the status of major acquisitions
	1969 - SAR's will be used to advise top DoD management and the Congress of the progress of major acquisitions
House Committee on Armed Services	April 24, 1970 - Report to Congress listed multiple concerns associated with the SAR
DoD Improvements	Enhanced definition of costs (flyaway costs, weapon system costs, procurement costs, and program acquisition costs)
	June 30, 1970 - Issued Weapon System Costing memorandum
	September 13, 1971 - Revised DoD instruction 7000.3 to ensure that no SAR would have more than 13 pages
Comptroller General (makes three reviews)	February 6, 1970 - Issued report called "Status of the Acquisition of Selected Major Weapon Systems"
	March 18, 1971- Issued report called "Acquisition of Major Weapon Systems"
	July 17, 1972 - issued report titled "Acquisition of Major Weapon Systems" (concluded that static baselines should be reported and maintained in the SAR - Also Made 5 additional significant reccomendations)

1984 GAO Report to Congress

The next document we find in our research, is a GAO report titled, "DoD Reports to the Congress Need More Realism," that makes several recommendations concerning the SAR. One significant portion of this report was found in Chapter III. Indeed, GAO voiced multiple concerns and made recommendations which we summarize here:

- Some of the weapon system's cost data reported to the Congress is not accurate, complete, or timely. SARs did not always report the most current and accurate program costs, sometimes reports contain costs inconsistently, and did not report operation and support costs.

- Costs known to be part of systems requirements are not always included in the officially approved program for the system.
- Congress introduced the unit cost exception report in 1982 because of their concern for the lack of currency of the cost data in the SARs.
- An expanded requirement for DoD certification of the validity of its estimates may result in more complete cost reporting to Congress.

With concerns outlined, GAO made the following recommendations:

- SARs report all relevant program costs, use the most current data, and report costs in a consistent manner. In the event of an exceptional situation where costs are excluded from the estimates, those costs should be clearly identified in the rationale for their explained exclusion.
- Clear criteria should be established regarding the cost to be included in the officially approved program for a given weapon system.
- DoD should disclose the total number of units it is considering for a program by providing a SAR footnote when the number is different from the approved program reported in the body of the SAR.
- Unit cost exception reports should disclose any anticipated cost growth that has not been included in the latest officially approved estimate (Bowsher, 1984).

<u>1988 GAO Report to Congress</u>

In the GAO report to Congress in 1988 titled "MAJOR ACQUISITIONS: Summary of Recurring Problems and Systemic Issues: 1960-1987." It outlined years of acquisition reform and concluded that significant change in acquisition progress since the 60's had not been realized. The report did not significantly address SAR reporting but did mention that the SAR is where GAO acquired its data for analysis. Appendix II.1 of the report mentions the number of problems associated with data reporting in Selected Acquisition Programs from 1980-1983 as 5 for 1980, 0 for 1981, 4 for 1982, and 3 for 1983. Although the report doesn't specifically mention the SAR as being the culprit, one can reasonably assume that the SAR is directly involved with this problem (Wath, 1986).

<u>1989 GAO Report to Congress</u>

This report shows how DoD could substantially improve the quality, timeliness, and presentation of data it provides to Congress on its major system acquisitions. The report focuses on three key areas:

- Present the SARs in the same manner as the internal acquisition status reports (Defense Acquisition Executive Summary, DAES).
- Enhance the presentation and usefulness of SAR data.
- Improve SAR presentation of schedule data.

With these three key areas identified, we further expound upon these starting with the first main point.

In reference to DoD's internal acquisition reporting, GAO found that the DAES presented significant improvements over the SAR, for example the DAES contains: quarterly reporting, the use of a standard microcomputer-based model to generate the DAES reports that reflect "as of" data current to within 60 days, and contractor data to improve quality reporting.

In reference to enhancing the presentation and usefulness of SAR data, GAO found that DoD could further enhance the utility and arrangement of SAR data through use of automation and graphs. GAO found that collectively SARs contain approximately 2000 to 3000 pages of data and are only provided to Congress as a typed copy. Ultimately they believed that DoD could establish a single SAR database to allow Congressional oversight committees access to SAR data through their own personal computers. DoD could format SAR data in such a way that congressional staff could query the system for key performance indicators relating to each acquisition program. The report also discussed other benefits of moving to computer technologies to aid in the timing and providing helpful graphics for data analysis.

In reference to the GAO comment on improving SAR presentation of schedule data, the report references an excerpt from the December 31, 1987 SAR report on the AMRAAM program schedule. As shown by Table 1.3 in their example, the program did not report the initial development estimates for low rate and full rate production phases for the AMRAAM, and did not calculate the actual schedule variances (Conahan, 1989). This report further emphasizes the recurring evolution of the SAR and the need for significant improvement in the vicinity of graphics and presentation. With the 1989 GAO report summarized, our next report further examines the SAR and prescribes noteworthy changes.

<u>2005 GAO Report to Congress</u>

This report is the most current evaluation that the GAO has performed in relation to the SAR. The report focused on the following key areas which we expound upon in the proceeding paragraphs:

- Information for Congress on unit cost performance could be more complete.
- Rebaselining can occur during each phase of acquisition, sometimes frequently on individual programs.
- Year-to-year unit cost changes are not measured and reported.
- In some cases, DoD reduces the magnitude of unit cost growth for Nunn - McCurdy determinations.
- Reporting of rebaselinings could be timelier and unclassified data is unnecessarily restricted.

In reference to information for Congress on "unit cost performance could be more complete," GAO believes that DoD could do better especially when it comes to program rebaselines. GAO mentions a new baseline serves an important management control purpose when program goals are no longer achievable. A new baseline also presents an important perspective on the program's current status and acquisition strategy. However, by comparing the latest unit cost estimate with the most recent approved baseline, DoD provides an incomplete perspective on program performance because a rebaseline shortens the period of performance reported and resets the measurement of cost growth to zero. The SAR also does not present how unit cost growth is adjusted for Nunn-McCurdy determinations when a rebaseline reduces the number of units to be procured or increases system capabilities. For example, the GAO report highlighted four programs that emphasize the highest number of rebaselinings in Table 1 of their report: F/A-22 (14 rebaselines), DDG 51 (11 rebaselines), SM-2 Block IV (11 rebaselines), SSN-21 (10 rebaselines). Of the 81 programs reported in 2003, 49 or (60%) had multiple rebaselinings over the life of the program.

Cumulative unit cost changes are also not measured and reported in constant dollars. DoD reports to Congress on the unit cost growth of programs in two different sections of the SAR. One section reports the current estimate against the latest approved baseline in both then-year and constant dollars. However, because of rebaselining, the latest approved baseline may be in place for only a short time and the measurement of unit cost growth is reset to zero. The other section reports the cumulative historical change only in then-year dollars, which includes the effects of inflation. For example, Table 2 of the GAO report highlights six programs that failed to show cumulative changes to unit cost in constant dollars as compared from the latest APB to the original APB. AMRAAM (87 to 254 months), AAWS-M (Javelin) (34 to 174 months), FMTV (7 to177 months), USMC H-1 Upgrades (20 to 87 months), V-22 Vertical Lift Aircraft (20 to 212 months), and F/A-22 (4 to143 months).

The report also mentions that year-to-year unit cost changes are not measured and reported. For example, DoD does not measure and report changes in unit cost between the latest budget request and the prior budget. Of the seven rebaselined programs between 2004 and 2005, DoD reset the cost growth for all of them to zero. For example, the Stryker armored vehicle program reported a little over 1% unit cost growth in the two months since it rebaselined, but DoD did not report the significant fact that the program experienced a 20% growth between annual budget requests. GAO highlighted four programs showing differences in the percentage in PAUC (program acquisition unit cost) from the calculation changes reported to Congress from the current estimate compared to the latest APB, which did get reported and the current estimate compared to the prior-year estimate that did not get reported. The report mentioned these following programs in Table 4 of the report: Advanced Extremely High Frequency Satellites (2.78% to 49.88%), F/A 22 Raptor (0% to 27.39%), Stryker (1.34% to 20.97%), and Force XXI Battle Command (0% to 72.79%).

The GAO report also describes that in some cases, DoD reduces the magnitude of unit cost growth for Nunn-McCurdy determinations. Congress established unit cost as the key measure of buying power compared to the average cost to buy each unit. Congress also established the Nunn-McCurdy cost increase thresholds of 15% and 25% for unit cost growth that require detailed reporting. If quantities decrease but cost is the same or did not decrease proportionately, unit cost would necessarily increase. When it comes to determining Nunn-McCurdy breaches, DoD allows for the exclusion of the unit cost increases associated with reductions in quantity or increases in capabilities on individual programs. These actions are referred to as programmatic adjustments and require the approval of a new acquisition program baseline. These adjustments allow DoD to reduce the number of the magnitude of unit cost increases reported to Congress that would otherwise exceed the Nunn-McCurdy thresholds. The effects of these adjustments are not explicitly visible in the reports to Congress. For example, the GAO report showed in Table 5 that the Comanche Helicopter Program projected a unit cost of $31,275 as of the baseline and a projected quantity of 1,213. The current estimate showed a unit cost of $50,621 and a projected quantity of 650 which is a 61.86 percentage change. The GAO report goes on to mention that if DoD had reported the adjustments for unit cost growth to account for reduced quantities or increased capabilities, the number of programs that would have had Nunn-McCurdy breaches reported to Congress (between fiscal years 2001 and 2003) would have increased by roughly 50% (from 17 to 25 programs).

The GAO report highlights another issue concerning the fact that reporting rebaselinings could be timelier and that DoD unnecessarily restricts unclassified data. DoD typically reports the latest program rebaselined information in its December SAR. Sometimes data can arrive for congressional consideration as much as 12 months after a program has rebaselined. Between the April 2003 and the April 2004 SARs to Congress, DoD rebaselined nine programs and on average failed to report these rebaselines for six months. Between 1996 and 2003, about two thirds of the rebaselines occurred between April and December. A new baseline approved after early April may not be reported to Congress before the enactment of authorization and appropriations legislation. The GAO report highlighted the DD (X) Destroyer program. The report mentions that the program established a new baseline on the April 23, 2002, but did not report this new baseline to Congress in the SAR until April 2003. As a result, between April 2002 and the passage of the fiscal year 2003 defense budget, the SAR provided to Congress did not reflect the approved baseline for the program. Table 7 of the GAO report highlighted seven programs that had reporting lag time from eight to 12 months.

In addition to problems with lag time, DoD unnecessarily classified SAR data. DoD classified about 50% of the SARs submitted to Congress in 2003, involving a total acquisition investment of $454 billion. Astonishingly, only a small amount of data contained in each classified SAR is actually classified. The classified material is generally only one of the 18 sections in a report which generally falls in the Performance Characteristics section. Performance characteristics include such items as speed, range, etc. Because the SARs are classified, special handling procedures must be used by those congressional staff with the proper clearances access to the unclassified cost and schedule data. This practice also completely blocks access for those staff without clearances to the unclassified cost and schedule data. As a result, congressional oversight of DoD's adherence to establish cost and schedule baselines is unnecessarily constrained.

With problems identified, the 2005 GAO report then made the following recommendations:

- Measure and report a full history of unit cost performance in constant dollars by comparing the latest cost in quantity estimates with: the first full estimate (typically the original acquisition program baseline established at milestone B), the current approved baseline, or if the program rebaselines, the prior approved program baseline, and the estimate established with the previous year's budget request.
- Fully disclose to Congress the nature and extent of programmatic adjustments affecting Nunn-McCurdy thresholds and terminations, pending any congressional direction on this issue.
- Notify Congress when rebaselining actions are approved.
- Separately report classified and unclassified SAR information (Levin, 2005).

We feel that the 2005 GAO report covers the most significant and current scrutiny that the SAR has undertaken to date. GAO made several points of concern, which we summarize in Table 2 and then shift to the 1992 RAND study.

Table 2. 2005 GAO Report Problem/Solution Summary Table

Problem	Solution
Information for Congress on unit cost performance could be more complete and Year-to-year unit cost changes are not measured and reported.	Measure and report a full history of unit cost performance in constant dollars by comparing the latest cost in quantity estimates with: the first full estimate (typically the original acquisition program baseline established at milestone B), the current approved baseline, or if the program rebaselines, the prior approved program baseline, and the estimate established with the previous year's budget request.
Rebaselining can occur during each phase of acquisition, sometimes frequently on individual programs.	N/A
Reporting of rebaselinings could be timelier	Notify Congress when rebaselining actions are approved.
In some cases, DoD reduces the magnitude of unit cost growth for Nunn -McCurdy determinations.	Fully disclose to Congress the nature and extent of programmatic adjustments affecting Nunn-McCurdy thresholds and terminations, pending any congressional direction on this issue.
Unclassified data is unnecessarily restricted.	Separately report classified and unclassified SAR information.

Pitfalls in Calculating Cost Growth from Selected Acquisition Reports, 1992

RAND (1992) begins with the preface that highlights two primary prerequisites for researching the patterns and causes of cost growth, a reliable database and a consistent methodology for insuring the comparability of the data. The study's main objective concerned identifying the weaknesses of the database and how those weaknesses influence calculations of program cost growth. This study provides an excellent background of the SAR and points out numerous difficulties with respect to the measurement of cost growth. RAND found the following most notable problems areas:

- Failure of some programs to use a consistent baseline cost estimate.
- Exclusion of some significant elements of cost.
- Exclusion of certain classes of major programs (e.g., special access programs).

- Constantly changing preparation guidelines.
- Inconsistent interpretation of preparation guidelines across programs.
- Cost sharing in joint programs.

RAND (1992) also looked at the baseline problem, more specifically which baseline to use. The study pointed out that the originally chosen baseline may not be as stable as first thought. This study cited two primary reasons. First, baseline estimates are occasionally updated subsequent to their creation. Second, once evolutionary model changes are introduced, the baseline may reflect a work scope far different from the current one (Hough, 1992). It appears that RAND shows some concern in the area of a revolving baseline and this could present a serious problem when trying to standardize schedule milestones across multiple programs by only looking at the most current SAR.

RAND goes on to talk about the changing development estimate (DE). RAND discusses the importance of planning, development, and production estimates for the same program and emphasizes that the development estimate is the most frequently used baseline in the SAR. RAND states that although the changing of the development estimate is not common, it sometimes happens and the analyst must be careful in selecting the correct baseline date. RAND mentions that as a general rule, once a baseline has been selected, the first estimate presented as the baseline should be used for calculations of cost growth. RAND (1992) cited multiple examples to support their view, one of which was the Bradley fighting vehicle system. The Mechanized Infantry Combat Vehicle (MICV) preceded the Bradley fighting vehicle. Subsequently, DoD combined the MICV with the Bradley fighting vehicle as two separate programs within one SAR, hence the development estimate's accuracy is questionable. The schedule date problem basically demonstrates how a service can take a canceled or restructured program, often for poor cost performance, and then breathe life into the program and take the opportunity to update the development estimate. RAND also believes that the development estimate occasionally is changed in a manner that puts the program in a more favorable light and sometimes changes are not made when they should.

Evolutionary model changes also concerned RAND. They emphasize that successful systems such as the F-15 or F-16 fighter aircraft realized quantities that have been greatly increased, and perhaps more importantly, the configuration has been modified so much that current models only remotely resemble originally planned estimates. For example, the study highlights the F-15E models currently in production at the time of this study are still reported in the SAR against the F-15A development estimate. RAND (1992) also further emphasizes that since improvements at current stages may only later be packaged as a new model. Therefore, the costs associated with evolutionary changes are difficult and often impossible to extract from SARs.

The RAND study also points out a related problem that occurs when baseline estimates and CEs (current estimates) are based upon an inconsistent program definition. The GAO has found cases over time where DoD program managers developed program estimates from differing work breakdown structures and therefore could not be comparable. RAND emphasizes that this is a great example of why research that places blind faith in the most current SAR, without examining program histories and making appropriate adjustments, can lead to bias results in flawed conclusions.

The report also reflects that contractor-borne expenses also pose an area for inconsistency. Because the SAR only includes the estimated cost to the government and not total investment costs, true cost growth is underestimated depending on how much the contractor must contribute. RAND (1992) believes that this can generally happen in one of two ways. One way is when cost overrun occurs on a firm fixed-price or fixed-price incentive contract. Costs above the ceiling price are borne entirely by the contractor and therefore, are not reflected in SAR estimates of total procurement cost. The second way is when contractors use their own funds to improve R&D efforts.

The RAND study mentions another potential pitfall which pertains to incomplete data and the evolution of the SAR database. RAND states that typically, studies are based on data contained in the active SARs as of December of a given year. The study emphasizes that exclusion of inactive SARs without systematic rationale may lead to biased results that may not reflect current procurement policy. RAND believes that cost growth should identify what portion of the total SAR population is included and why the sample is representative of the whole is satisfactory for meeting the study objectives. There are also numerous programs that are not even reported in the SAR database due to the sensitive and classified nature of these top-secret programs like the F-117A Stealth Fighter. RAND (1992) mentions of these black programs protect advanced technology in reference to our national security, but they also prevent the public eye at program costs.

Another area of apprehension to RAND concerns the historical changes in SAR preparation guidelines. RAND further states that there are too many changes to identify in this report but highlights the most common changes up to 1992. We include them in Table 3.

Table 3. (RAND) Summary Table of Most Common Changes to SAR up to 1992

Most Common Changes to SAR up to 1992 (RAND)
Reporting in base year and current dollars
Changes in threshold for SAR reporting
Reduced frequency of reporting (quarterly to annual)
Escalation of prior-year actuals to base-year dollars
Inclusion of production estimates for baseline
Change in the selection of base year from the first year of funding to the fiscal year of the development estimate (DE)
Addition of program cost-quantity data
Use of the baseline cost-quantity relationship to calculate quantity-cost variances
Reduction from nine to seven cost-variance categories

The RAND (1992) study further emphasizes that many of these changes are not significant in estimating cost growth and may even simplify the process; however, some of the changes have distinct and subtle effects. For example, RAND mentions that before 1984, the development estimates in base-year dollars often include actual dollars. In fact, one program's expenditure preceded the base-year of the development estimate. In 1984, administration changed the rules so that earlier *actuals*, if any, would be inflated to the base-year of the development estimate. Other changes became more critical, like the guidelines intended for quantity changes to be calculated based on the original cost-quantity curve. Many programs based their current unit prices on quantity variance calculations until December 1979 when the office of the Secretary of Defense increased manning for review of SARs. The study also finds it interesting that DoD had not fashioned a guide for the preparation and review of SARs until May 1980. The author of the RAND report believes, that although that many of changes appear to be minor, it reinforces the caution against temporal comparisons.

Along with changes in guidance is the problem with inconsistent preparation techniques. RAND (1992) mentions that inconsistencies may result from deliberate manipulation of the data as though it came from unintentional errors. Even though the RAND study does not cite specific examples in this particular analysis, RAND states that there is sufficient evidence available to suggest that certain programs under pressure may resort to liberal interpretations of the reporting instructions or, as RAND calls it, "creative financing" to reduce the apparent cost growth. The study cites two general possibilities for creative financing. One possibility is shifting costs between appropriations (R&D and procurement) and between fiscal years to take advantage of lower escalation rights. Another possibility is to make unrealistic cost improvement assumptions for out-year production lots to avoid a Nunn-McCurdy unit cost breech. RAND believes that this type of activity is not ubiquitous; but it would be youthful to deny that it exists.

Cost sharing in joint programs is another area of concern. The RAND (1992) states that when two or more services spend money in a joint program, SARs for the individual services can either be arbitrarily broken out or can be split in such a way that one service absorbs the fixed costs of production. The study mentions that the situation is most common in the case of air-to-air missiles used by both Air Force and Navy fighters. When individual service cost growth factors are calculated, the actual performance of the program may be distorted. The author further states that usually the lead service absorbs most of the cost growth in a program because it funds the majority of R&D.

RAND concludes that although there are numerous problems with the SAR, many of these concerns may be addressed through careful research and analysis. RAND believes that the SAR is still useful for capturing broad-based trends in temporal patterns. However, the conclusions drawn from a high-level analysis must be tempered by the weaknesses found in the data (Hough, 1992). Since we describe this report in great detail, we include a summary table (Table 4) for clarity of what RAND considers the SAR's most notable problems as of 1992 and then we review DOD guidance.

Table 4. (RAND) Summary Table of the SAR's Most Notable Problems up to 1992

(RAND) Summary Table of the SAR's Most Notable Problems up to 1992
Failure of some programs to use a consistent baseline cost estimate
Exclusion of some significant elements of cost
Exclusion of certain classes of major programs (e.g., special access programs)
Constantly changing preparation guidelines
Inconsistent interpretation of preparation guidelines across programs
Cost sharing in joint programs

<u>DoD Instruction 5000.2</u>

The first DoD instruction that we review lays the foundation for the acquisition model. It also outlines key players and responsibilities in the acquisition process. In fact, the Program Management enclosure takes up a primary portion of this instruction. This particular enclosure defines the key responsibilities of the Program Manager (PM) and Program Executive Officer (PEO). The instruction mentions that a Program Manager shall be designated for each acquisition program. DoD shall designate the PM and PEO no later than program initiation. The instruction goes on to state that a Program Manager shall serve in the position for four years in accordance with the Defense Acquisition Workforce Improvement Act. The PM shall be given budget and guidance in a written charter of his or her authority, responsibility, and accountability for accomplishing approved program objectives. Unless a waiver is granted, a PEO shall be assigned acquisition program responsibilities. The PEO shall be dedicated to executive management and shall not have other command responsibilities (DoD, 2003b). With our first instruction reviewed, we transition to DoD directive 5000.1.

DoD Directive 5000.1

This particular directive is the bedrock for the Defense Acquisition System. It outlines key leadership hierarchy and their responsibility; defines pertinent terms and methodology; and directs personnel oversight. The significant portion of this directive is again, the assignment of the program manager (PM) and the overarching duties of that assignment. In this directive, the "Total Systems Approach" is defined by emphasizing that the PM shall be the single point of accountability for accomplishing program objectives for total life-cycle systems management, including sustainment (DoD, 2003a).

DoD 5000.2-R

This particular regulation goes into great detail in relation to most aspects of the acquisition process. One area of significance is associated with the Acquisition Program Baseline (APB). The guidance states that every acquisition program shall establish an APB beginning at program initiation. The PM shall base the APB on user's performance requirements, schedule requirements, and estimate of total program cost. Performance shall include interoperability, supportability, and as applicable, environmental requirements. The APB is part of the Consolidated Acquisition Reporting System

(CARS). The PM is responsible for using the CARS to prepare the APB. In other words, the APB is where all the baseline schedule dates are estimated. A very interesting note is in the area of schedule dates in reference to the APB. The regulation states that schedule parameters should minimally include dates for program initiation, major decision points, and the attainment of initial operating capability (IOC). The PM may also propose for milestone decision authority approval, or other specific, critical system events as necessary. This particular area causes some concern because this research focuses on why all SAR reports in the Monaco (2005) database are missing multiple schedule dates. This guidance suggests that all programs do not need to report those dates but only need to report the previously described minimum schedule dates.

The Selected Acquisition Report (SAR) portion is another significant area of this regulation. In accordance with 10 U. S. C. 2432, the PM shall submit a SAR to Congress for all ACAT I programs. The SAR is required to report the status of total program cost, schedule, and performance; as well as program unit cost and unit cost breech information. The regulation goes on to state that each SAR shall include a full life cycle cost analysis for the reporting program, each of its evolutionary blocks, as available, and for its predecessor program, if applicable. The regulation also states that the SAR for the quarter ending December 31 shall be called the annual SAR. The PM is required to submit the annual SAR within 60 days after the president submits the following fiscal year’s budget to Congress. Annual SARs are required to reflect the president's budget in supporting documentation. Quarterly SARs are submitted on an exception basis for the following reasons: (1) the current estimate exceeds the Program Acquisition Unit Cost (PAUC) objective or the Average Procurement Unit Cost (APUC) objective of the currently approved APB, both in base-year dollars, by 15% or more; (2) the current estimate includes a six-month or greater delay, for any schedule parameter, that occurred since the current estimate reported in the previous SAR; or (3) milestone B or milestone C approval occurs within the reportable quarter.

The Secretary of Defense may also waive the requirement for submission of SARs for a program for a fiscal year if: (1) the program has not entered system development and demonstration; (2) a reasonable cost estimate has not been established for the program; (3) or the system configuration for the program is not well defined. This particular portion is of importance and explains why some programs have multiple SARs for a particular fiscal year and some programs only have one. It's noteworthy to mention that this piece of information explains why some programs are missing SAR reports for a given fiscal year.

Similar to the SAR is the Defense acquisition executive summary (DAES). The DAES is a multi-part document, reporting program information and assessments; PM, PEO (program executive officer), CAE (component acquisition executive) comments; and cost and funding data. The DAES is designed to be an early warning report to USD(AT&L) and ASD(C3I). The DAES describes actual program problems, warns of potential program problems, and describes mitigating actions taken. At a minimum, the DAES are for program assessments, unit costs, and current estimates. It's also reports the status of exit criteria and vulnerability assessments. The DAES shall also present total cost and quantities for all years, as projected, through the end of the current acquisition phase. The regulation also states USD shall designate which ACAT I programs shall be reported using the DAES and assign each program a quarterly reporting group. The PM uses the CARS to prepare the DAES estimate and must provide both hard and electronic copies to USD(AT&L) by the last working day of the program's designated quarterly reporting month. It's noteworthy to mention that the DAES is an internal document to the undersecretary of Defense and not directly reportable to Congress.

Another significant portion of this regulation is associated with the Consolidated Acquisition Reporting System (CARS). CARS is the personal computer-based data entry and reporting software package. It maintains and reports information on Defense programs. CARS is mandatory requirement for all MDAPs and MAIS (Major automated information systems) programs, but non-MDAP and non-MAIS program may still use the system. CARS has three reporting modules: the APB, the SAR, and the DAES. CARS includes analysis routines, such as the Computational Module that supports the SAR cost change calculations, and SAR and DAES data checks. The Director, Acquisition Resources and Analysis, maintains a CARS helpline for user support. A unique program number (PNO) identification system controls the use of CARS. The office of USD(AT&L) is the focal point that assigns a PNO to each using ACAT I program.

The CARS software specifies the format of the APB, SAR, and DAES, except for narrative or memo type information. The three reporting modules share some, but not all, of the CARS data. For example, the DAES and the SAR report the APB. The modules also share some contract information. The CARS software also includes instructions for preparing the APB (acquisition program baseline), SAR, DAES, and UCR (unit cost reporting), including administrative procedures. This last piece of information concerning SAR, DAES, and CARS has a significant impact on this research because the SAR and DAES get their data directly from the CARS which is inputted by the PM.

In summary to this regulation, we find some significant nuggets of information. First, DoD only requires PMs to minimally include schedule dates for program initiation, major decision points, and the attainment of initial operating capability (IOC). This subjective statement answers one of our research objectives e.g. *Why are programs missing data in their respective SAR?* Hence, they are really not missing all the schedule dates mentioned in Chapter I, like FUE for example, because the PMs are not required to provide them. Second, the Secretary of Defense may also waive the requirement for submission of SARs for a program for a fiscal year which explains why some programs are missing SARs for a given date. And Lastly, SAR and DAES get their data directly from the CARS which is inputted by the PM, which leads us to believe that we cannot extract more schedule data from the systems other than what we already have in the SAR (DoD, 2001). The next form of guidance we review is the Defense Acquisition Guidebook.

Defense Acquisition Guidebook, 2005

The Defense Acquisition Guidebook is a 520 page document that the Defense Acquisition University maintains. This large document's purpose is to assist the acquisition workforce in using and understanding the DoD acquisition policies and the guidance associated with those policies. Its general goal is to provide members of the acquisition community and their industry partners with an interactive, on-line reference to policy and discretionary best practices. This document may be redundant in that it duplicates information from 5000.1 and 5000.2 but it also expands upon the regulations and clarifies common discrepancies. Since it is an online document, the Defense Acquisition Guidebook adapts to change faster than conventional policy (DAU, 2004).

With the guidebook highlighted, we progress to other pertinent sources of information associated with the SAR, such as the CARS web site.

Consolidated Acquisition Reporting System (CARS) web site

Due to the rapid change the acquisition system has undergone, there is no centrally published information on CARS that is thorough enough to stand on its own. Therefore, this research turns to the CARS DoD sponsored web page to enhance our literature review. CARS, as described before, is a personal computer-based data entry and reporting System combining common and unique Defense Acquisition Executive Summary (DAES) and Selected Acquisition Report (SAR), and Acquisition Program Baseline (APB) components into a unified repository from which DAES and SAR reports and APB documents can be printed. CARS leadership designed the system to do the following:

- Reduce field workload for preparation of Baseline, DAES, and SAR reports.
- Generate standardized, automated project status information for use by the Program Manager, Component Acquisition Executive (CAE), and Defense Acquisition Executive (DAE) staffs.
- Provide more timely and accurate analytical tools for Office of the Under Secretary of Defense (Acquisition, Technology, and Logistics) and the military services.
- Improve acquisition data management.
- Establish acquisition-wide software, data, and documentation standards (Flaharty, 2005). From the CARS web site, we find some very useful information in the CARS User Guide.

CARS User Guide, Oct 2005

Although the user Guide is not doctrine, it is published guidance from the CARS web site that directly describes how each component in the SAR is derived. Since it is the most complete document pertaining to how a SAR is created, this research deems it noteworthy to appraise. The first interesting feature concerns the "points of contact." The manual provides information that is not just for technical support. There is also contact information for MDAP, MAIS, DAES, and SAR functional support POCs. The next significant portion of the guide focuses on the components of the program. The program is set up in a modular structure that consists of four modules: Baseline, DAES, SAR, and Budget which we describe in the proceeding paragraphs.

The baseline module describes in detail how baseline dates are presented in the SAR, which is why it is of interest to this thesis. For example, all milestones come from approved schedule milestones included on the baseline disk as provided by the Defense Acquisition Executive (DAE) or the Component Acquisition Executive (CAE).

The SAR module contains the most comprehensive information on the SAR found thus far. The SAR background section also presented a very comprehensive summary of the most comprehensive policy changes that have occurred since the inception of the SAR. After the SARs inception in 1969, in 1975 the FY 1976/7T Authorization Act established the SAR as a legislative reporting requirement to be submitted to Congress. Over the past three decades, the SAR has changed several times. In 1983, a SAR Improvement Task Force significantly reduced the contents of the SAR. In 1985, the FY 86 Authorization Act restored the information that had been removed and added production to aid in operating and support cost information. The FY 87

Authorization Act provided for limited reporting for Pre-milestone B (formally milestone II) programs and relaxed some reporting criteria. The FY 1990 Authorization Act merged the Unit Cost Exception Report (UCR) with the SAR and eliminated separate unit cost exception reporting.

In continuing with regulatory actions to the SAR, The FY 92 Authorization Act gave the Secretary of Defense the authority to waive selected acquisition reporting for certain programs. The act also allowed the Secretary of Defense to change the content of the SAR as long the SECDEF notified the appropriate House and Senate committees in advance. The Federal Acquisition Streamlining Act of 1994 changed the baseline for unit cost reporting purposes from the prior President's budget to the approved acquisition program baseline (APB) and substituted the “procurement unit cost" for the current “year” procurement unit cost. In 1996, DoD made additional changes in SAR format and content which reduced the volume of the SAR by about 20 to 30%. The development of CARS proved to be a very significant change in the preparation of the SAR. Since CARS infancy stage in 1990, it provided a standardized, automated system for generating the SAR.

The SAR baseline segment is another comprehensive area of the SAR module in CARS. This section discusses the types of baseline estimates and also explicitly covers SAR baseline changes. SAR baselining (SAR baseline, baseline performance characteristics, schedule milestones, and cost estimates) are established for the initial SAR. Depending on the phase of the acquisition cycle and the time the initial SAR is submitted, the baseline values are represented by a Planning Estimate (PE), a Development Estimate (DE), or a Production Estimate (PdE).

The document goes on to mention that baselines are changed (e.g. from PE to DE or DE to PdE) at major milestone decision points after review and approval by the USD(AT&L). An APB is reflected in the SAR up to and including the first time a DE is approved as the SAR baseline at milestone II (now Milestone B). A DE is reflected in the SAR up to and including the first time a PdE is approved as the SAR baseline at milestone III (now Milestone C). SAR baselining is a two-step process. The first step is the submission of a SAR which shows the old SAR baseline with the new SAR baseline being reported in the approved program (APB); this SAR is usually called the "transition" SAR. The second step is the submission of a SAR that shows the new SAR baseline in the SAR baseline column (OSD, 2005b). This section seems to suggest that the SAR baseline can change, depending on when each estimate is submitted. With a comprehensive look at CARS, we feel it necessary to include a summary table of the major changes noted in the SAR since its inception (See Table 5) Following the summary table, we move our review to the Ostrich/English Dictionary.

Table 5. CARS Summary Table of Significant Changes to the SAR (Mainly Congressional)

CARS Summary Table of Significant Changes to the SAR	
1969	SARs Inception
1975	1975 the FY 1976/7T Authorization Act established the SAR as a legislative reporting requirement to be submitted to Congress
1983	SAR Improvement Task Force significantly reduced the contents of the SAR
1985	FY 86 Authorization Act restored the information that had been removed and added production to aid in operating and support cost information
1986	The FY 87 Authorization Act provided for limited reporting for Pre-milestone B (formally milestone II) programs and relaxed some reporting criteria
1989	The FY 1990 Authorization Act merged the Unit Cost Exception Report (UCR) with the SAR and eliminated separate unit cost exception reporting
1991	The FY 92 Authorization Act gave the Secretary of Defense the authority to waive selected acquisition reporting for certain programs; also allowed the Secretary of Defense to change the content of the SAR as long he notified the appropriate House and
1994	The Federal Acquisition Streamlining Act of 1994 changed the baseline for unit cost reporting purposes from the prior President's budget to the approved acquisition program baseline (APB) and substituted the "procurement unit cost" for the current "year" procurement unit cost
1996	The Department made additional changes in SAR format and content which reduced the volume of the SAR by about 20 to 30%
1996	CARS system is created and is mandatory for all SAR preparation

Ostrich/English Dictionary, 2005

The Ostrich/English Dictionary is a document that describes the Ostrich database. The Ostrich database is the database that the CARS system partially feeds into. It is also the database that the Defense Acquisition Management System (DAMIR) uses. Oracle software created the database but it is designed so that it works with the CARS software which is a legacy program. Therefore, the files contained in this database are not normalized and cannot be used in the typical robust fashion that a relational database is designed to perform. Mata-Toledo and Cushman (2000) state that normalization (of a database) is an objective criteria based upon the analysis of the relations, their schemes, their primary keys and how their functional dependencies are standardized. This evolutionary process is described as levels of normalization from first normal form (1NF) to fifth normal form (5NF). The objectives for a normalization process are to: (1) make it feasible to represent any relation in the database, (2) obtain powerful retrieval algorithms based on a collection of primitive relational operators, (3) free relations from undesirable insertion, update, and deletion anomalies, and (4) reduce the need for restructuring the relations as new data types are introduced (Mata-Toledo et al., 2000). It is noteworthy to mention that if the programmer normalized the database in at least 3NF, the data could be used in an almost unlimited fashion to perform queries that would be highly useful to the analyst. Currently, the data must be programmed so that it may be functional with web-based applications. This type of programming takes a great deal of time and cannot be reproduced from the raw database (Rosenberger, 2005).

Defense Acquisition Management Information Retrieval (DAMIR), 2005

DoD officially released DAMIR in March 2005 and provided right of entry to Congress in April 2005. Congress finally has limited electronic access to Selected Acquisition Report (SAR) data. DAMIR is a DoD initiative to provide enterprise accessibility to acquisition program information. The primary goal of DAMIR is to streamline acquisition management and oversight by taking advantage of the capabilities of a web based environment. DoD describes DAMIR's purpose to identify the various data sources the acquisition community uses to manage Major Defense Acquisition Programs (MDAP) and Major Automated Information Systems (MAIS) programs and provide a unified web-based interface through which to present that information. DAMIR allows for visibility of the aforementioned programs by OSD, Military Services, Congress and other participating communities. Now they can have access to information relevant to their missions regardless of the agency or where the data is located. DAMIR is slated to evolve in such a way that its components are to replace the need for the legacy Consolidated Acquisition Reporting System (CARS). The current DAMIR capability consists of two major web-based components: Purview and the Virtual Library.

Purview is an executive information system that displays program information such as mission and description, cost, funding, and schedule data. DoD developed the DAMIR initiative to provide a comprehensive view of the current state of all MDAP and MAIS programs. Purview is the presentation layer for structured data currently collected in CARS. It continues to be the solution for structured acquisition data presentation as the DAMIR initiative moves forward, and web services begin pulling this information directly from the service acquisition databases (USD, 2005).

The establishment of purview has dramatically improved information access but there are also some limitations. The major benefit of Purview to the researcher is that the data is centrally located via the web and can be reviewed separate from a secure area which is a significant step up from the current way of extracting the data. The current method entails acquiring a secret security clearance and authorization to a DoD secured area. Once access to a secured area, the researcher may review the Adobe Acrobat files on CD-ROM and then print the pertinent files (assuming the secure area possesses a printer). Then the security manager reviews the documents for potential classified data and finally the data is keyed by hand into a database.

The ability to print the unclassified portions of SAR reports from one's current personal computer is another benefit with Purview. Although access improvements are realized with this system, there are some limitations. One limitation is that all the data viewed on the web can only be reviewed or printed. There is not an easy way to capture the data and convert it to a flat file for analysis. The inability to directly download the data can be understood because of the data structure that the web component uses. The Ostrich database is the database that the web component uses which has not been normalized into third normal form due to the limitations of the CARS legacy software. Another significant limitation is that the SAR reports in the system date back only to 1997. Any SAR reports that are older than 1997 still need to be accessed via the classified SAR CD-ROMs.

The second component of DAMIR is the Virtual Library. The Virtual Library is a search tool for unstructured data discovery. The programming team, realizing that a typical search engine cannot find all the necessary information the acquisition community needs, created this search discovery tool to find information stored in various formats. Some examples of such formats are MS Word documents, or ".pdf" files, and also other incongruent sources, like Oracle databases, file servers and web servers (USD, 2005).

Chapter Summary

In this chapter, we review a plethora of historical documents that dissect acquisition reporting from many different angles. It is from this information where we derive the answers to many of our research objectives associated with the multiple missing schedule dates from the Monaco database (derived from the SAR for its information).

First, DoD only requires PMs to minimally include schedule dates for program initiation, major decision points, and the attainment of initial operating capability (IOC). This subjective statement answers one of our research objectives e.g. *Why are programs missing data in their respective SAR?* Hence, they are not missing all the schedule dates mentioned in Chapter I, like FUE for example, because the PMs are not required to provide them.

Second, the Secretary of Defense may also waive the requirement for submission of SARs for a program for a fiscal year which explains why some programs are missing SARs for a given date. And Lastly, SAR and DAES get their data directly from the CARS which is inputted by the PM, which leads us to believe that we cannot extract more schedule data from the systems other than what we already have in the SAR.

Third, there appears to be no other central repository for reporting the performance of MDAP programs. Indeed, there is only one vector for historical analysis of MDAP reporting based on this literature review, the SAR. This exhaustive search tentatively answers another one of our research objectives *Is there any other comprehensive data sources for acquisition reporting other than the SAR?* We expound upon this research objective in the next chapter.

Fourth, each MDAP is not required to use identical baseline schedule variables. It is very difficult to use schedule milestone data for statistical analysis with so much flux and variance between each MDAP, particularly in the area of changing baselines which leads us to focus a large portion of our effort on this in Chapter III. Indeed, we take a hard look in the area of baselines amongst various MDAPs to see if this affects the Monaco (2005) model. One way we accomplish this is by adding the variable called *"# of Rebaselines*" which we describe in the next chapter.

Finally, the SAR has undergone rapid evolutionary changes from its infancy in the late sixties. As mentioned in various reports, these changes directly affect the data when analyzed across 30 years. Indeed, to fully realize the historical flux of the SAR, we close this chapter with two abridged summary tables to recap three decades of change to the SAR (See Table 6 & Table 7). We also include a table that summarizes the system evolution of SAR based acquisition in Table 8.

Table 6. Significant Historical Literature Summary of the SAR (Part I)

Significant Historical Literature Summary of the SAR (Part I)	
1973 Letter to Chairman of the Committee on Armed Services	
(Background History)	February 1968 - first authoritative guidance that established the mandatory SAR requirement (DOD instruction 7000.3)
	1968 - the DoD used the SAR in an experimental stage
	1969 - Secretary of Defense established an objective to be advised regularly on the status of major acquisitions
	1969 - SAR's will be used to advise top DOD management and the Congress of the progress of major acquisitions
(House Committee on Armed Services)	April 24, 1970 - Report to Congress listed multiple concerns associated with the SAR
	Enhanced definition of costs (flyaway costs, weapon system costs, procurement costs, and program acquisition costs)
	June 30, 1970 - Issued Weapon System Costing memorandum
	September 13, 1971 - Revised DOD instruction 7000.3 to ensure that no SAR would have more than 13 pages
(Comptroller General makes three reviews)	February 6, 1970 - Issued report called "Status of the Acquisition of Selected Major Weapon Systems"
	March 18, 1971- Issued report called "Acquisition of Major Weapon Systems"
	July 17, 1972 - issued report titled "Acquisition of Major Weapon Systems" (concluded that static baselines should be reported and maintained in the SAR - Also Made 5 additional significant reccomendations)
1984 GAO Report - Most Current Problems	Some of the weapon system's cost data reported to the Congress is not accurate, complete, or timely. SARs did not always report the most currently anticipated program costs, sometimes reports contain costs inconsistently, and did not report operation and support costs
	Costs known to be part of systems requirements are not always included in the officially approved program for the system
	Congress introduced the unit cost exception report in 1982 because of their concern for the lack of currency of the Cost data in the SARs
	An expanded requirement for DoD certification of the validity of its estimates may result in more complete cost reporting to the Congress
1989 GAO - Most Current Problems	Present the SARs in the same manner as the internal acquisition status reports (Defense Acquisition Executive Summary, DAES)
	Enhance the presentation and usefulness of SAR data.
	Improve SAR presentation of schedule data.

Table 7. Significant Historical Literature Summary of the SAR (Part II)

Significant Historical Literature Summary (Part II)	
RAND Identifies Evolutionary Changes up to 1992	Reporting in base year and current dollars
	Changes in threshold for SAR reporting
	Reduced frequency of reporting (quarterly to annual)
	Escalation of prior-year actuals to base-year dollars
	Inclusion of production estimates for baseline
	Change in the selection of base year from the first year of funding to the fiscal year of the development estimate (DE)
	Addition of program cost-quantity data
	Use of the baseline cost-quantity relationship to calculate quantity-cost variances
	Reduction from nine to seven cost-variance categories
RAND Identifies Common Problems up to 1992	Failure of some programs to use a consistent baseline cost estimate
	Exclusion of some significant elements of cost
	Exclusion of certain classes of major programs (e.g., special access programs)
	Constantly changing preparation guidelines
	Inconsistent interpretation of preparation guidelines across programs
	Cost sharing in joint programs
2005 GAO Report - Most Current Problems	Information for Congress on unit cost performance could be more complete and Year-to-year unit cost changes are not measured and reported.
	Rebaselining can occur during each phase of acquisition, sometimes frequently on individual programs
	Reporting of rebaselinings could be timelier
	In some cases, DoD reduces the magnitude of unit cost growth for Nunn -McCurdy determinations
	Unclassified data is unnecessarily restricted

Table 8. Summary Table of Historical System Changes to the SAR

Summary Table of System Changes to the SAR	
1968-69	SAR's Inception (Paper Copies)
1989	The Inception of DAES
1990	CARS is created/required as a reporting tool
1990	Data associated with CARS is segregated into three database (SAR, DAES, APB)
1996	The Department made additional changes in SAR format and content which reduced the volume of the SAR by about 20 to 30%
March 2005	DoD officially released DAMIR (Available Programs = 1997 to Present)
April 2005	DoD allows congress access to selected data in DAMIR

(Knoche, 2006)

III. Methodology

Introduction

This chapter outlines the procedures that we use to conduct this research. The research in this chapter begins with a search for missing schedule variables from the Monaco (2005) database using the literature results as a template. From our search for missing variables, we incorporate a new variable that was realized from the literature review. We then review the data source that has been used by prior researchers and how we incorporate the new variable into our model. This research also reviews the methodology that has been used by prior research in this vein of study. Lastly, we explain the use of response and predictor variables along with preliminary data analysis.

Search for Missing Variables Associated with Schedule Slippage

This research from the literature review realized that filling in missing schedule dates for MDAPs using the SAR is futile because each program had varied schedule dates and each MDAP is not required to report the missing dates that are mentioned in the Monaco (2005) thesis. Dates like First Unit Equipped (FUE), Preliminary Design Review (PDR), Production Contract Award (PCA), Critical Design Review (CDR), and EMD Contract Award are not mandatory requirements (DoD, 2001). Also, as this research combed over hundreds of SARs, we find that the schedule date called FUE, is primarily an Army term and is not present in Air Force programs other than joint ventures (DAU, 1999). This point is significant because the Monaco (2005) and Moore (2003) models showed much promise in the area of predictability using FUE. Also, it is noteworthy to mention that some of the researched historic SARs contained a small amount of the missing dates, however, the dates are classified and therefore unable to be reported. The DAES also proved to be unfruitful to fill in the holes because the program schedule dates for the SAR and the DAES for each program shows an identical match, hence if the SAR does not have a schedule date, neither does the DAES. The DAMIR system also does not prove useful to fill in the missing schedule dates because the DAMIR pulls its data from the same place that the SAR and DAES get their data from, the CARS. Also, as mentioned previously in Chapter II, the database for SAR and DAES only date back to 1997. We can only find the data in paper copies of the SAR prior to 1997.

Since the SAR, DAES, and DAMIR do not provide any venues to fill in the gaps in the Monaco (2005) database, this research looks for other avenues to fill in the missing dates which are highlighted in the proceeding paragraphs. However, this research does not find any other plausible databases or systems that could alleviate the missing information dilemma.

This research continues by taking a closer look at the SAR to see if the GAO is correct in their assessment of the SAR. This research evaluated hundreds of SARs and we find the results consistent with GAO's findings. In particular, the GAO mentions an area of concern in their 2005 report associated with the number of programmatic rebaselines. Again, this research also came to similar conclusions. However, this unique variable had not been used in any previous research so we decide to add a variable to each program called "*# of Rebaselines*" and see if this variable proves to capture more of the uncertainty associated with the prior research in this venue. It is also noteworthy to mention that the development estimate (DE) changes across numerous SARs, multiple times due to rebaselines, therefore the accuracy associated with the amount of schedule slip remains to be a key area of concern. This research realizes the need to study this effect, however, this is beyond the scope of this current research.

Data Assessment

The database of this research employs a modified version of a database originally built by Sipple (2002) and modified by Bielecki (2003), Moore (2003), Genest (2004), Lucas (2004), McDaniel (2004) , Rossetti (2004), and Monaco (2005). The original use of the database predicted cost growth within the cost categories of the Engineering and Manufacturing Development (EMD) phase using both logistic regression to predict if cost growth occurs and multiple regression to predict the amount of cost growth and lastly used by Monaco (2005) to predict schedule risk. The database consists of information from the 1990 to 2003 SARs, the 1996 RAND report, and the Defense System Cost Performance Database: Cost Growth Analysis Using Selected Acquisition Reports.

This research begins its data collection by thoroughly reviewing the most recent databases modified by Monaco (2005) and Genest (2004) since Monaco looks at schedule growth and Genest evaluates total RDT&E cost growth. As mentioned in the previous paragraph, all past researchers in this venue used the same database. However, since we want to evaluate our new variable in relation to schedule and cost, we hone in on these two studies. The databases consist of programs listed in the SARs from 1990 to 2003. The Monaco (2005) research designed its database to predict schedule growth, while the Genest (2004) designed its research database to predict total research and development cost growth. It is noteworthy to mention, that there are other minor variations in these two databases, such as additional schedule related variables and excluded programs in the case of Monaco (2005). This research desires a big picture look at how this new variable interacts with the two aforementioned databases, therefore we use these two separate databases to evaluate schedule and cost growth. In keeping with the consistency of the Monaco (2005) research, we add a program to the database when it is classified as a mature program (minimum three years in EMD). Also for consistency, the previous researchers do not include programs that fall under the new milestone labeling scheme of A, B, and C compared to a I, II, and III labeling format. In like manner, this research also does not include programs using the new naming convention. To be consistent with the Monaco (2005) database, we follow the same selection criteria in relation to how the Monaco research differs from the previous versions of the database. For example, the Monaco (2005) research only includes programs that complete the EMD phase since U. S. Code: Title 10: Section 2432 states:

> "The requirements of this section with respect to a major defense acquisition program shall cease to apply after 90 percent of the items to be delivered to the United States under the program (shown as the total quantity of items to be purchased under the program in the most recent Selected Acquisition Report) have been delivered or 90 percent of planned expenditures under the program have been made." (US, 2005)

When a program meets the aforementioned criteria, one last SAR report based on the DE is submitted. This is the SAR that populates the Genest (2004) and Monaco (2005) databases. Monaco (2005) concludes that it is necessary to wait until a program completed the EMD phase to ensure the actual completion date for phase II is captured. In keeping with the pace of the Monaco (2005) research, we utilize the same sixty-eight programs, however, Genest (2004) is not as stringent in selection criteria for total research and development cost growth so there are 135 data points available. In addition, prior to the Monaco (2005) rendition, other researchers included programs that had been terminated, including Genest (2004). This research also does not include terminated programs in our schedule database to maintain parallelism with the Monaco (2005) database. The emanate aspiration of the Monaco (2005) thesis desired to create a predictive model for a cost estimator to use to build accurate schedule milestones for successful programs. If a program is terminated before completing EMD, then there would not be sufficient information to build a predictive model. Genest (2004), as previously mentioned, has more creative license to take a "big picture" look at total research and development cost growth. Genest (2004) therefore, has more data to work with.

The databases had uniformity; however, they are arranged into a flat file with no specific identifier that ties each program to their respective SAR. This made it very difficult to track each program across many years because one researcher may be using a different program than a preceding researcher and cause discontinuity between longitudinal research. Hence, we rearrange and reformat the data using a common naming identifier to allow for consistent trend analysis for years to come. Since each SAR in the DoD database uses a unique identification number called the PNO (program number), this research scrubs all the programs and realigns them in relation to the PNO. This research also changes the Excel file into a Microsoft Access format for collection purposes to allow for adding multiple SARs to the database. In fact, this initiative allows for meta-analysis of other databases which may add to the body of knowledge via future research.

In addition, we add new sources to our database in an attempt to find new predictive variables. The first data table that we add to the aforementioned database is a PNO table. It consists of a listing of MDAP programs and their respective PNO numbers. This research extracted the data from the CARS DoD website. This table allows for consistency and uniformity of the older Monaco (OSD, 2005a) and Genest (2004) databases and any new data that we be add to the primary database. The next source of data came from a series of three GAO reports called the DEFENSE ACQUISITIONS: Assessments of Selected Major Weapon Programs, which start in 2003 and end in 2005. These three reports choose a number of high risk weapon systems and assess the status using various indexes to include minimal schedule dates, and basic cost and quantity data. The reports do not include all the active programs so it is not an exhaustive database but it does add a different perspective to the primary research database but does not fill in any of the missing data gaps. This research teamed with another researcher, inputted the cost and schedule data by hand, and then validated the data for accuracy (Walker, 2003; Walker, 2004; Walker, 2005).

After an exhaustive search for missing variables, this research identifies the variable called "*# of Rebaselines*." This new variable identifies the number of rebaselines a program has undergone. We use this new variable to see if it may be an indicator of schedule and cost growth using the databases of Monaco (2005) and Genest (2004). For more information on the aforementioned databases, see the theses of Monaco (2005) and Genest (2005). It is important to understand the direction of this research in relation to using past research. This research was originally intended to fill in the missing variables from prior research, however, this research has taken a slightly different direction in an attempt to validate this new predictor variable against already established models that are empirically tested and have positive results. We have no intention of changing the direction of our thesis to create entirely new models because that is far beyond the scope of our research.

Exploratory Data Analysis

Since much research has already been accomplished in this area, we pick two "big picture" data models to use in order to narrow the scope of our research. We choose schedule growth and cost growth models with a significant focus on Monaco (2005) for schedule growth and Genest (2004) for RDT&E cost growth. A review of Sipple's research, indicates that cost growth during EMD is from a mixed distribution. Roughly half of the distribution is continuous, while the rest is massed around zero. In addition to a mixed distribution, Sipple finds a small portion of the programs have negative cost variance. Since Monaco and Genest found similar conclusions as Sipple, we duplicate the procedures established in the Sipple research. We provide an overview of these procedures next.

We first need to split the data into two sets, discrete and continuous. The theoretical probability of obtaining a specific value is zero with the continuous distribution. Such a probability does not accurately reflect the fact that many of the points of past research fall directly on zero. For the discrete distribution, we use logistic regression, and we use multiple regression analysis for the continuous distribution. First, we attempt to develop a logistic regression model to predict whether or not a program has cost or schedule growth from the full data set. Then we develop a multiple regression model from only the programs that actually demonstrate cost or schedule growth to predict the amount of cost or schedule growth we expect. Additionally, to ensure that we construct a robust model, we set approximately 20% of our data aside for validation before we begin any regression analysis. Finally, before performing regression, we must also choose the response and predictor variables. For more specific procedures, see Sipple's research (Sipple, 2002:59).

Response Variables

In keeping consistent with past research, we use the same modeling techniques. Sipple (2002) concerns itself with two different response variables, one that indicates if cost growth occurs and the other variable expresses the degree to which cost growth does occur. Monaco (2005) uses a similar model for schedule growth but is not concerned with cost growth. Since this research is focused on big picture predictability with our new variable, we address four response variables. Hence, this research and in like manner, uses two response variables for schedule growth and uses two response variables for cost (predicting if growth occurs, predicting the degree of growth). Since we evaluate two distinctly different families of models for our research (one for cost growth and one for schedule growth), we use great care to distinctly separate the two models.

To narrow the scope of this research, we use the response variables from Monaco (2005) and Genest (2004) for schedule and cost growth respectively. Monaco (2005) uses the variable "*EMD Schedule Slip?*" for its logistic regression model and "*EMD*

Schedule Variance %" as the response variable for the multiple regression model. Therefore, when we evaluate the predictability of our new X variable called *"# of Rebaselines*," we attempt to use the Monaco (2005) response variables for evaluating the predictability in reference to schedule growth (Monaco, 2005:124-125). In a similar fashion, this research attempts to use the Genest (2004) response variables to evaluate the predictability of our new variable against RDT&E total cost growth. Genest (2004) uses "*R&D (Total) Cost Growth?*" as the response variable for the logistic regression model and "*RDT&E Total Cost Growth %*" as the response variable for the multiple regression model (Genest, 2004:54-55).

Search for Predictor Variables

In an attempt to speed up our research, we look to past research of Sipple (2002) and other researchers that followed the same direction to identify proven predictor variables. Our technique remains the same barring the use of the FUE predictor variable due to its limited use outside of Army acquisition. The response variables, exploratory data analysis, and logistic regression are identical and hence the aforementioned research can be referred to for added insight. We include a list of the predictor variables used in the past research of Sipple (2002) in Appendix B. Sipple provides a brief description of the subcategories and includes elucidation of unclear elements when necessary (Sipple, 2002:61). Monaco (2005) adds additional variables and we include the unabridged list in Appendix C. We do not include the complete variable list from all past researchers because there is little variation from the two previously mentioned lists. We do however; include any significant predictor variables of past research in the proceeding paragraphs.

Since the crux of this research is not to create entirely new models to test our variable, we dial into already established and predictive variables of past research. Genest (2004) provides an excellent summary of significant past predictor variables of prior researcher's final models for logistic and multiple regression. We include these final model variables in Table 9 and Table 10. Genest's effort to flesh out these significant predictor variables of past theses greatly aids our research efforts. (Genest, 2004:51)

Table 9. Past Research Significant Predictor Variables for Linear Regression

Engineering - Sipple	Estimating - Bielecki	Schedule - Bielecki	RDT&E - Genest
Actual Length of EMD	Length of R&D in Funding Yrs	Maturity (Funding Yrs complete)	Svc > 3
MSIII-based Maturity of EMD %	Versions Previous to SAR	Army Involvement	Maturity (Funding Yrs complete)
Modification	Navy Involvement	Versions Previous to SAR	R&D Funding Yr Maturity %
Length of R&D in Funding Yrs	PE	Prototype	Risk Mitigation
Length of Prod in Funding Yrs	Lead Svc = DoD	Northrop Grumman	EMD Prototype
Actual Length of EMD (IOC-MSII)	Program have a MS I		Program have a MS I
Land Vehicle	Prototype		

Table 10. Past Research Significant Predictor Variables for Multiple Regression

Engineering - Sipple	Estimating - Bielecki	Schedule - Bielecki	RDT&E - Genest
Maturity from MS II	IOC-based Maturity of EMD %	Boeing	Northrop Grumman
No Major Def Contractor	Proc Funding Yr Maturity %	Land Vehicle	Funding Yrs of R&D Completed
Prog Acq Unit Cost	General Dynamics	Lead Svc = Navy	Maturity of EMD at IOC %
	Lead Svc = Navy	Program have a MS I	Prototype
	PE		Significant pre-EMD activity
			LRIP Planned

In addition to the aforementioned variables, we also include a summary table (See Table 11) from Monaco (2005) to add to our pool of significant predictor variables from prior research (Monaco, 2005:106).

Table 11. Monaco (2005) Final Model Variables

Model	Response Variable	Predictor Variable
Logistic	EMD Schedule Slip?	Aircraft?
		Svs > 1?
		Service = Navy Only?
		Planned EMD Length
		% Proc Funding (yrs) of Total Program
		Planned Unit Cost CY $M 2003 < 5 $M?
Multiple	EMD Schedule Variance %	Actual Phase I + Planned EMD Length
		Modification?
		MS III before IOC ?

Our new variable called *# of Rebaselines* comes from the original file that the 2005 GAO report used, which causes significant alarm pertaining to the number of MDAP programs experiencing rebaselines. The OSD provided this data, which we in turn validate against GAO's report (Levin, 2005). The database matches perfectly with the database of prior research in this vein, however, many of the older programs in our database are not populated because the data file contains programs that are active and only date back to the late 90's. Therefore, the number of programs in our model is significantly decreased due to the lack of rebaseline data. With the candidate predictor variables selected, this research turns to the model methodology of logistic regression.

Logistic Regression

Logistic regression is typically used to analyze possible predictive relationships when the response is either a nominal or ordinal. Logistic regression mainly predicts binary outcomes, usually coded '0' and '1 (Neter et al., 1996:567). Since we do not charter to fully develop a new model but build upon existing models, we do not fully follow procedures of the past. Past research utilizes logistic regression to develop a model that predicts whether or not a program either has schedule and/or cost growth or not. Therefore, in our database, we attempt to use the same coding as Monaco (2005) and Genest (2004) who incorporate a '1' for schedule or cost growth and a '0' for no schedule or cost growth. Our cost growth distribution and our schedule growth distribution contains only 0's and 1's, hence we characterize *whether or not a program has total cost growth or total schedule growth* as a Bernoulli random variable with probability p of success (success = one) (Neter et al., 1996:568). We, in fact, are not able to use logistic regression in our research, which we explain in Chapter IV. We include this information for continuity and clarity with past research. For example, we add the logistic regression technique of Sipple (2002) in the proceeding paragraph for this very reason.

Sipple established the following guidelines for utilizing logistic regression in cost growth analysis:

> We use JMP® 4 (SAS Institute, 2001) software to accomplish the logistic regression in order to help us identify the best model for estimating whether or not a program will have cost growth. JMP® uses maximum likelihood to estimate the coefficients of our model. Because JMP® has no automatic method, such as stepwise, for logistic regression, we manually compute thousands of individual regressions, recording our results on spreadsheets. We start with one-predictor models of all possible variables. Then we regress using all combinations of two predictor models and record the results. We continue this process, eventually whittling down the best combinations for use at the next level in order to cut down on the amount of regressions necessary. We stop when we reach a model her for which the gain of adding another variable does not warrant the additional complexity of the model that another variable adds. We intend to find several candidate models for each number of predictors and then narrow down to the best one for each number of predictors and validate the model using about 20 percent of the data that we set aside for validation. (Sipple, 2002:70)

Multiple Regression

Past research uses multiple regression to discover prediction models for the percent of total cost growth and schedule growth based on more than one predictor variable. As previously mentioned, we utilize two distinct models in our analysis; one for schedule growth and the other for total RDT&E cost growth. This research does not intend to re-create entirely new models however we investigate how the new predictor variable interacts with previously developed research. Since our goal is to test our new variable against already established models. We do not fully use the entire procedures for multiple regression. We do, however, include the following multiple regression guidelines established by Sipple (2002) for descriptive continuity with past research:

> We use JMP® for the multiple regression analysis. We use the stepwise method to identify those predictor variables that have a statistically significant impact on the ability of the model to predict our response variable, *Engineering %*. From our stepwise analysis, we build models using the standard least squares method, whereby JMP® estimates the form of the functional relationship between the predictors and the response variable that minimize the sum of squared deviations from the predicted values at each level of the predictors (Neter et al, 1996).
>
> Because of the large amount of candidate predictor variables, we exceed JMP®'s stepwise calculation abilities when we include all of our variables in a single run. In addition, we seek models with varying numbers of predictors. Thus, we must repeat the stepwise and standard least squares several times in order to achieve the desired results. As with logistic regression, we discover several candidate models for each number of predictors. Then we narrow our results to the best model for each number of predictors. We continue adding variables to the model until the number of variables equals about one tenth of the number of data points used in the model; this ensures we do not over-fit the model (Neter et al, 1996:437). We check the model's robustness using the same validation data as for the logistic regression. (Sipple, 2002:72)

Utilizing the methodology found in Sipple's study, we attempt to build four regression models. Two for schedule growth and two models for total RDT&E cost growth. However, since we could not take advantage of logistic regression with this research (See Chapter IV), we only build two multiple regression models. Since the variable *# of Rebaselines* is only populated in newer programs from the late 90's to present, this significantly reduces the number of programs that we can utilize in our models for schedule growth and total RDT&E cost growth. We start with 135 data points by way of the cost model. After we remove the data points that contain no rebaseline information and the 20% validation set, we are left with significantly less data points to use in our model. Likewise, our schedule model contains even less data due to the aforementioned reasons the stringent database guidelines utilized by Monaco (2005).

Both the schedule and cost models are highly similar. The methodological intent of the schedule growth database and models are to use logistic regression to predict whether or not there will be schedule growth. We then plan to use multiple regression to predict the percentage of schedule growth. Likewise, with the cost growth model, we attempt to use logistic regression to predict if total RDT&E cost growth occurs and then we make use of multiple regression to predict the percentage of total RDT&E cost growth. We also apply a log transformation to the response variable in our cost model to correct for heteroskedasticity in the residual plot (Sipple, 2002:72). It is noteworthy to mention that we do not need to transform the response variable in our schedule database. We discuss this and the final numbers of each database (schedule & cost) in Chapter IV.

Chapter Summary

This chapter is the genesis of our analytical process. We demonstrate the connection between our literature review and the analysis we perform. Our literature review is very important to this methodology. Interestingly, it was in our literature review that we find that there are no other places to fill in the missing data other then the SAR and that many of these MDAPS have multiple rebaselines. Furthermore, it is this specific finding that changes the course of our research. With no way to fill in the missing data, we must look to other ways to evaluate the data. One such way is to look at the data from a longitudinal perspective via the number of rebaselines each program has undergone. The analysis portion of this thesis therefore, takes a different direction by utilizing the new predictor variable *(# of Rebaselines*). Next, we review the SAR database and the various response and predictor variables that are associated with our new variable. Finally, we explain the reasoning for past research in relation to the use of logistic and multiple regression techniques and the process in which we incorporate many of these techniques in our research. In Chapter IV, we put into practice our methodology that we describe in Chapter III. We do this by testing the *# of Rebaselines* against formally established schedule and cost models of past research.

IV. Results and Discussion

Chapter Overview

This chapter details the results of both the attempted logistic analysis and the multiple regression analysis. In it, we describe the resulting models and their robustness. We also analyze the models for statistical legitimacy and practical expediency. We start by evaluating two families of models with the new predictor variable (*# of Rebaselines*) with logistic and multiple regression. We choose the two families of models because we want to test our new variable against the two main categories of past analysis in this venue. The first model family stems from the research by Monaco (2005) with respect to schedule growth. The second model family originates from the research by Genest (2004) with respect to total RDT&E cost growth. Finally, we validate each family of models to determine the accuracy, legitimacy, and reasonableness of each model. It is noteworthy to mention, that we evaluate and describe each of these families of models separately to maintain coherency with respect to our research.

Preemptive Data Analysis (Schedule)

As previously mentioned, our intentions are not to create entirely new research and disregard the work of Monaco (2005) with respect to schedule risk. On the contrary, we intend to build upon this research and test our new variable against Monaco's published work. With that said, we next look at the database of Monaco (2005) to ensure we have a foundational dataset to test our variable.

With the Monaco database as our first area of discussion, we first describe the database and any changes that we make to the date by Monaco (2005). We use the same database, however, many of the older programs are excluded because we do not have rebaseline data on programs from the early 90's. The original Monaco database contained 68 programs. With the removal of programs missing rebaseline information, our database decreases to 49 programs. Our subsequent task is to create a validation set of those 49 programs. To do this, we randomly select 80% of our 49 programs and use these programs as our working database. We then use the remaining 20% as our validation table. After we randomly create the 20% validation set and the 80% working database, our working database reflects a decrease from 49 to 39 programs.

With our working database formulated, we then look at the distribution of the multiple regression response variable used by Monaco called *EMD Schedule Variance %* in association with our modified database. The preliminary stem and leaf plot shows a smaller percentage of the programs massed at zero than the research of Monaco (2005). This is reasonable due to the decreased number of cases in our database, which lends further credence to omit logistic regression in this portion of our thesis (See Figure 1).

Stem and Leaf

Stem	Leaf	Count
15		
14	4	1
13		
12		
11	5	1
10	00	2
9	4	1
8	2	1
7	5	1
6	22239	5
5	2477	4
4	5	1
3	778	3
2	13469	5
1	48	2
0	1335578	7
-0	0000	4
-1	8	1

-1|8 represents -0.18

Figure 1. Stem-and Leaf Plot of Y (Schedule Variance %)

We next look at the distribution of both the logistic regression response variable used by Monaco and evaluate the distribution of our new predictor variable *# of Rebaselines* for its usefulness for logistic regression. Monaco's logistic response variable was *EMD Schedule Slip?* We analyze the response variable distribution of Monaco's full database and compare it to our new 80% working database. We find the distribution of the binary responses of our database significantly different than Monaco (2005) due to the removal of the programs that are missing rebaseline information and the randomization of that data set. For example, the percentage of programs that demonstrated schedule slips in the full Monaco database are 25% and our newly created working database is only 12.8% (See Figure 2). Not only is the percentage different, but the number of programs that experience schedule slip is also different. For example, the original Monaco database had 51 programs that demonstrate schedule slip and our current database only has 34 programs that exhibit schedule slip. This significant change concerns us in respect to validating the logistic model of Monaco and due to the dichotomy between programs that demonstrate schedule slip and those that do not.

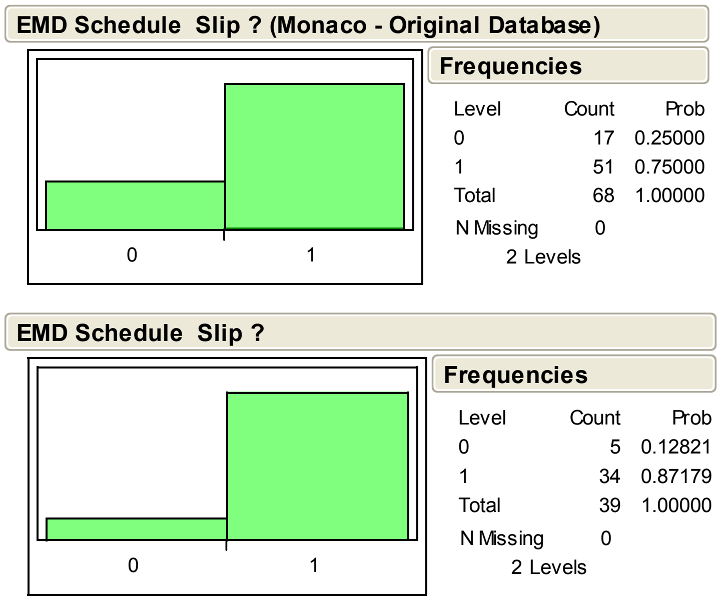


Figure 2. Preliminary Logistic Distribution of Y (Schedule)

Next, we look at the distribution of our new predictor variable *# of Rebaselines*. The predictor variable distribution also causes concern because of the very few programs that experience no rebaselines at all, which is only one (See Figure 3).

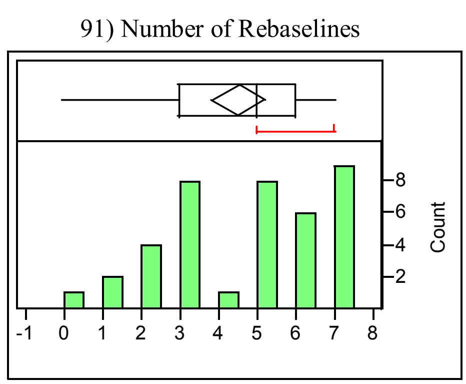


Figure 3. Distribution of New Predictor Variable (Schedule)

Since the distribution of the response variable in our model is different and our predictor variable only has only one demonstrated program with zero rebaselines, we feel it's superfluous to attempt logistic regression with our new predictor variable. Due to the disparity between the number of 1's and 0's in our logistic response variable, a logistic model would be highly unstable with such a disconnect. Therefore, we do not attempt logistic regression. For more information on the logistic regression model of past schedule research, see Monaco (2005) (Monaco, 2005:64-77).

Multiple Regression Results (Schedule)

Our subsequent step focuses on programs that have a positive schedule slip; the programs that have no slip or negative slip are excluded from the model in a similar fashion to Monaco's research. This resulted in our research excluding five programs from our model, which brings our total number of programs down to 34 in our working database. Our next step was to re-create Monaco's model and insert the same predictor variables to validate the robustness of that model against our modified database. Monaco's multiple regression model performs in a similar manner as described in that particular research, however, Monaco's adjusted R^2 is different from the original model, which is understandable because of the differing makeup of our database. The adjusted R^2 in the original model is 0.742518 and Monaco's model demonstrated an R^2 of 0.564831 against our modified database. It is noteworthy to mention, that all Monaco's predictor variables still perform in a highly prognostic manner. Since Monaco's predictor variables perform well, we use that model as a backdrop for testing our new variable *(# of Rebaselines*). For clarity, we include a table of Monaco's final model response variable and the proven predictor variables from that research in Table 12. This table is taken from Monaco's research and original database and reflects *p*-values in respect to that research.

Table 12. Monaco (2005) Multiple Regression Model

Monaco (2005) Multiple Regression Model		
Y Response Variable	**Predictor Variables**	**P Value**
EMD Schedule Variance % (positive only)	Actual Phase I + Planned EMD length	<.0001
	Modification?	<.0001
	MS III before IOC?	0.0406

With our backdrop solidified, we look to our new variable by itself in relation to Monaco's response variable. When we run the model in JMP® using multiple regression, we find no predictability in the variable *# of Rebaselines*. We achieve a *p-value* of 0.4316, which indicates no benefit to the model. However, when we rerun the model with Monaco's predictor variables in addition to the new variable *# of Rebaselines*, we find it highly predictive (See Table 13). Our preliminary run shows an adjusted R^2 value of 0.804748, which is significantly higher than Monaco's original model using our data, which achieved an R^2 of 0.564831. Due to identical missing data from predictor variables *Actual Phase I + Planned EMD Length, Modification?, and MS III before IOC?*, our database is limited to only 25 data points, which is similar to the experience of Monaco's research (Monaco, 2005:97-98).

Table 13. Final Model with # of *Rebaselines*

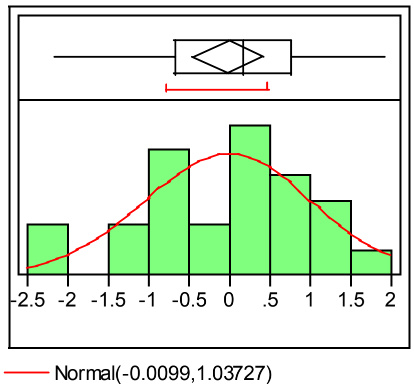
Final Model With New Variable		
Y Response Variable	**Predictor Variables**	**P Value**
EMD Schedule Variance % (positive only)	Actual Phase I + Planned EMD length	<.0001
	Modification?	<.0098
	MS III before IOC?	0.0011
	# of Rebaselines	<.0001

Before we move to the validation results, we need to first test our residuals of our multiple regression model to ensure the assumptions of the residuals are clearly met. The first assumption is independence. We follow the same assumptions as Monaco (2005) in that we assume independence is met due to the fact that we only use one SAR to obtain data for any one program except for the *# of Rebaselines* variable. The second assumption is normality of the residuals. We test this assumption by performing a Shapiro-Wilk goodness-of-fit test. Our last assumption is constant variance of the error term by performing a Breusch-Pagan test for constant variance (Neter et al., 1996:239).

In order to test the normality of the error term, we run a distribution of our studentized residuals in JMP®. We follow the same reasoning as Monaco (2005) for converting our residuals to a standard normal distribution with a mean of zero and a standard deviation of one to also look for outliers. Figure 4 displays the distribution of our model. Based on the fact that all data points are within three standard deviations of the mean, no outliers appear to exist in our model. Also, using an alpha of 0.05 as our threshold, we satisfy the assumption of normality with a *p*-value of 0.7754. The last assumption that we test is constant variance. Using Microsoft Excel®, we calculate a *p*-

value of 0.225678 for the Breusch-Pagan test. We again compare this *p*-value to an alpha of 0.05 indicating our assumption of constant variance holds.

Studentized Resid EMD Schedule Variance %



Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	Mu	-0.00991	-0.438073	0.418258
Dispersion	Sigma	1.03727	0.809932	1.443004

Goodness-of-Fit Test

Shapiro-Wilk W Test

W	Prob<W
0.975145	0.7754

Figure 4. Shapiro-Wilk Test for Normality - EMD Schedule Variance % Model (Schedule)

One last test we perform before being fully satisfied with our model for validation is the Cook's Distance test to identify influential data points. We use JMP® to graph an overlay plot of the Cook's Distance values. Figure 5 shows that we have an influential data point in our model. Program 125 is an influential data point with a Cook's Distance of 0.738. Neter establishes an influential data point as any point that exceeds the 50 percentile. Compared to Neter's established standard, Program 125 exceeds the 0.5 threshold (Neter et al., 1996:381). We therefore temporarily remove Program 125 from the model and rerun our regression model. The adjusted R^2 increases from 0.804748 to 0.836442 and the p-values lose some predictability in two of the predictor variables as indicated in Table 14. With an increase an adjusted R^2 and no dramatic change in the p-values, we feel confident in excluding Program 125, however, we need to recheck our assumptions since our model has been modified.

Table 14. Multiple Regression Parameters Post Temporary Removal of Program 125

Final Model With New Variable		
Y Response Variable	**Predictor Variables**	**P Value**
EMD Schedule Variance % (positive only)	Actual Phase I + Planned EMD length	<.0001
	Modification?	<.0013
	MS III before IOC?	0.0095
	# of Rebaselines	<.0033

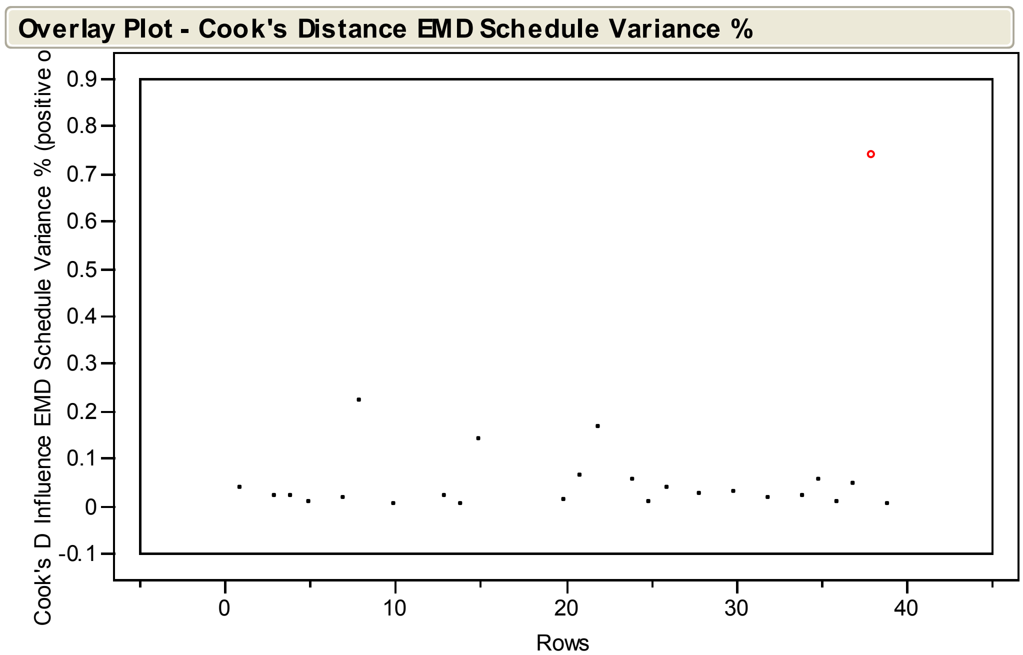
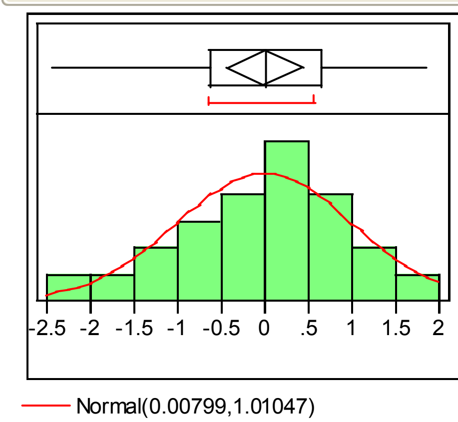


Figure 5. Cook's Distance Overlay Plot for Influential Data Points - EMD Schedule Variance % (Schedule)

We now double check our model for the assumptions of the error term since we removed Program 125. We still assume that independence is still valid as we previously mention the fact that the SARs are selected separately in different programs. Our next assumption is the assumption of normality of the error term. We run a distribution of our studentized residuals in JMP® and follow the same procedures as we mention previously. Figure 6 displays the distribution of our model. Based on the fact that all data points are within three standard deviations of the mean, no outliers exist in our model. Also, using an alpha of 0.05 as our threshold, we satisfy the assumption of normality with a *p*-value of 0.9679. The last assumption that we retest is constant variance. Using Microsoft Excel®, we calculate a *p*-value of 0.121211092 Breusch-Pagan test. We again compare this *p*-value to an alpha of 0.05 indicating our assumption of constant variance still holds even after we remove Program 125. Lastly, we need to double check for influential data points. The Cook's distance with Program 125 excluded does not significantly affect the model; all programs are less than 0.3 (Figure 7) which is less than Neter's 0.5 threshold (Neter et al., 1996:381).

Studentized Resid EMD Schedule Variance % (positive only)



Fitted Normal

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Location	Mu	0.007993	-0.418689	0.434676
Dispersion	Sigma	1.010467	0.785349	1.417444

Goodness-of-Fit Test

Shapiro-Wilk W Test

W	Prob<W
0.985040	0.9679

Figure 6. Normality and Shapiro-Wilk with Program 125 Removed (Schedule)

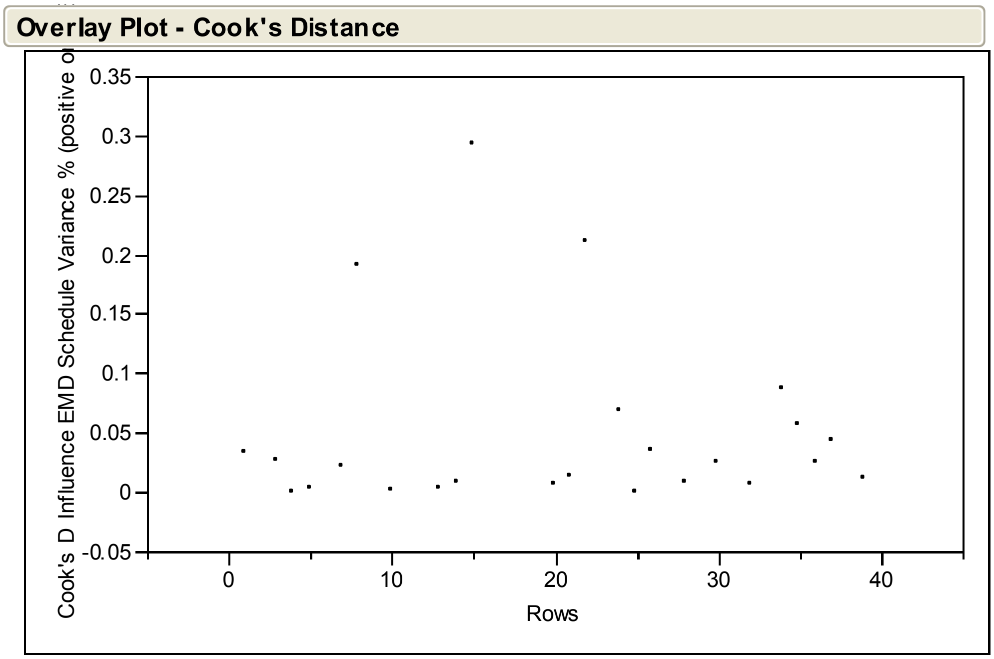


Figure 7. Cook's Distance Overlay Plot with Program 125 Removed (schedule)

Based on the performance measures, both individually and collectively, we believe our model provides us with a robust model with even greater predictive capability and statistical reliability than Monaco (2005). Also, with successful testing of the assumptions of the error term and the lack of outliers within our data set, we are certain in the ability of our model to correctly predict the degree of schedule variance in a program. Located in Appendix D we include the complete JMP® analysis of our final model *EMD Schedule Variance %.*

We now discuss the parameter estimates of our final model and the linear regression formula we submit for validation. The parameter estimates and linear regression formula are located in Figure 8 and Figure 9 respectively. The values of the predictor variables are input into the parenthesis adjacent to each of their respective parameters in Figure 9 for the linear regression formula. We mark each value in the parameter estimates table and draw an arrow to the respective portion of the linear regression formula in Figure 9 for clarity.

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	0.8758177	0.144169	6.07	<.0001	.
Actual Phase I + Planned EMD length	-0.009554	0.001079	-8.85	<.0001	1.1611273
MS III before IOC?	0.2125676	0.073656	2.89	0.0095	1.1766921
Modification?	-0.263873	0.069807	-3.78	0.0013	1.2700835
# of Rebaselines	0.0728925	0.021698	3.36	0.0033	1.4107758

Figure 8. Parameter Estimates - VIF Scores- (Schedule)

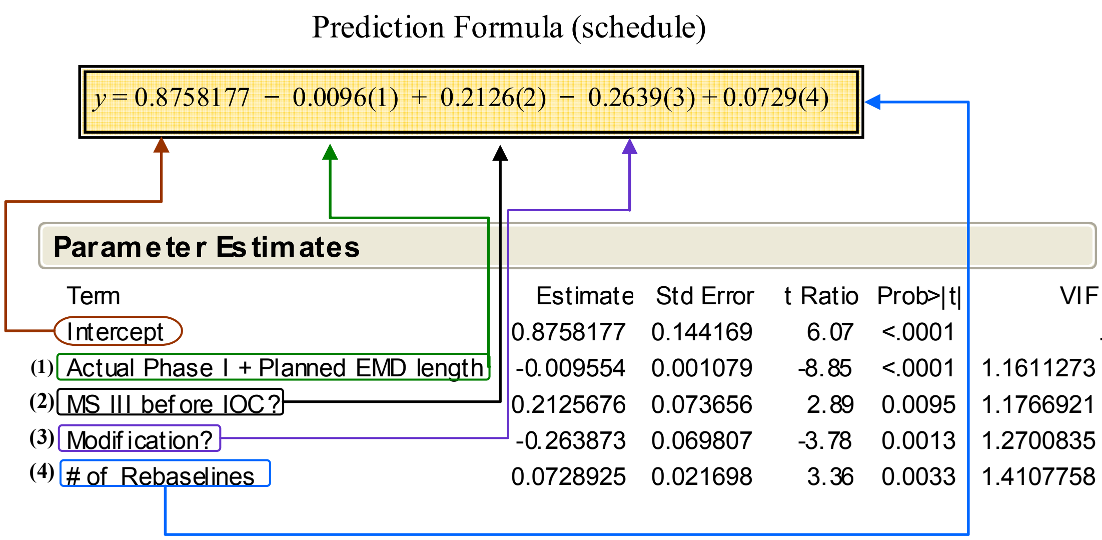


Figure 9. Prediction Formula - Arrows Correspond with Formula Values and Parameter Estimates (Schedule)

Furthermore, we standardize variable *Actual Phase I + Planned EMD Length* and variable *of Rebaselines* to get a proportional picture of the relationship between these two continuous variables. With these two variable standardized, we can evaluate the magnitude the variables have on each other in relation to the predictive formula. In Figure 10 we show that one unit change in *Actual PH I + Planned EMD Length* affects the predicted value roughly two times more than one unit change in *# of Rebaselines.* Furthermore, they have inverse effects from one another.

Parameter Est. - Standardized (# of Rebaselines & Actual PHI + Planned EMD Length)					
Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	0.4655704	0.076244	6.11	<.0001	.
MS III before IOC?	0.2125676	0.073656	2.89	0.0095	1.1766921
Modification?	-0.263873	0.069807	-3.78	0.0013	1.2700835
# of Rebaslines Standardized	0.1494284	0.044481	3.36	0.0033	1.4107758
Act PH I + Plnd EMD length Standardized	-0.29754	0.033617	-8.85	<.0001	1.1611273

Figure 10. Standardized Parameter Estimates of Variables [#-of Rebaselines, Actual Phase I + Planned EMD Length] (Schedule)

In addition to the parameter estimates, we also include the variance inflation factors (VIF) scores in Figures 8, 9, & 10. Various inflation is a consequence of multicollinearity and the VIF scores are a common way for identifying such a relationship. As a general rule, the VIF scores should not exceed 10 (Yu, 2004). All of our VIF scores are below two. Therefore, each of our independent variables appear to be explaining a different fraction of the variability in the model as noted in the R^2.

Validation Results (Schedule)

To validate our regression model, we added back the 20% of our database previously removed for model building. We begin our validation by computing the estimated confidence intervals of each program based on the parameter estimates and values of the independent variables within a 95% confidence interval. We are careful not to include the validation set in our *y* response. JMP® computes the upper and lower bounds based on the 95% confidence interval. We then look at the actual dependent responses of each program to see if they fall within the upper and lower bound computed by JMP®. The more programs that fall within the bounds of its confidence intervals, the confident we are in the predictability of our model. Of the 10 programs in our validation set, three programs demonstrate schedule slip; this equates to an 8% validation set. We proceed with validation even though we fall far below the established 20%. Of the three programs, two fell within the bounds of our test, which equates to a 67% prediction accuracy. This is an understandable finding due to the small number of validation programs and the diminutive number of programs that demonstrated schedule slip (3).

To compensate for the major limitation in our validation set, we use the entire database and perform the same validation procedures as previously described. When we compile 100% of the database, we have a total of 36 total programs that experienced schedule growth. Nine of these programs are missing data primarily from the variable *Planned IOC Date* similar to Monaco (2005) and therefore unusable (Monaco, 2005:97). Of the 27 usable programs for validation, only one program failed to fall within the upper and lower bounds with a 95% confidence interval. Therefore, 96.3% of our programs accurately predicted schedule growth. Furthermore, we also took an ultraconservative approach in validation and looked at the mean confidence intervals of each program and set the bounds to 95%. Of the 27 usable programs, 17 programs pass this stringent test for a 59% prediction rate. It is noteworthy to mention that five of the failed programs only fall short by a few percentage points, which theoretically would boost the prediction rate to 85%. We include the entire validation table in Appendix E. With the validated schedule model complete, we move our research to the cost family of model building.

Preemptive Data Analysis (Total RDT&E Cost)

As previously mentioned, our intentions are not to create entirely new research and disregard the work of Genest (2004) with respect to total RDT&E cost growth. On the contrary, we intend to build upon past research and test our new variable against Genest's published work. With that said, we next look at the database of Genest to ensure we have a foundational dataset to test our variable.

We first describe the changes in our cost database in contrast to Genest. We use the same database, however, many of the older programs are excluded in a similar manner as with our aforementioned research with schedule risk. The current cost database before modifications contains 135 programs. Our database reduces to 80 programs when we remove the programs missing rebaseline information. Our next task is to create a validation set of those 80 programs. After the appropriate rows are selected, we save the rows as a subset table in JMP®. From the remaining 20%, we then create another subset table, which is our validation table. After we randomly create the 20% validation set of those 80 programs, our 80% working database decreases to 64 programs.

Now that we have evaluated the differences between Genest's database and ours, we now look at the distribution of the multiple regression response variable used by Genest called *RDT&E % Total Changes*. The preliminary stem and leaf plot in Figure 11 shows a significant percentage of programs massed at zero, which indicates validity to use the two-step approach pioneered by Sipple (2002) and utilized by Genest (2004). Thus we attempt to apply that methodology to our research. As with the former research of Sipple (2002) and Genest (2004), we may need to transform the *y* variable in order to pass tests associated with our statistical assumptions.

Stem and Leaf Distribution of Y (Cost)

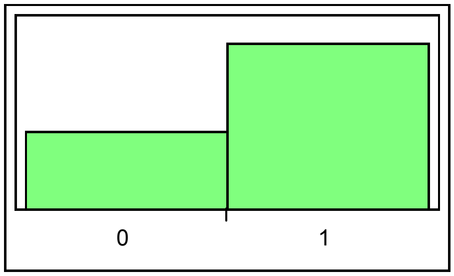
Stem	Leaf	Count
5		
4	5	1
4		
3		
3		
2	7	1
2	1	1
1	699	3
1	0001	4
0	555667778999	12
0	11111111222222333333444444	26
-0	4322111000000000	16

0|4 represents -0.4

Figure 11. Stem and Leaf Distribution of Y (Cost)

With a preliminary look at the multiple regression response variable, we turn our attention to the distribution of both the logistic regression response variable used by the research of Genest (2004) and evaluate the distribution of our new predictor variable *# of Rebaselines* against it's usefulness for logistic regression. Genest's logistic response variable was *Total R&D Cost Growth?* We analyze the response variable distribution of Genest's full database and compare it to our new 80% working database. We then find the distribution of the binary responses of our database significantly different than Genest (2005) due to the removal of the programs that are missing rebaseline information and the randomization of that data set. For example, the percentage of programs that demonstrate schedule slips in the full cost database are 32% (43 programs) and our newly created working database is only 20% (13 programs) in Figure 12. This significant change concerns us in respect to validating the logistic model of Genest (2004). This significant dichotomy between our 1's and 0's is similar to the obscurity in the previously described data set associated with schedule data.

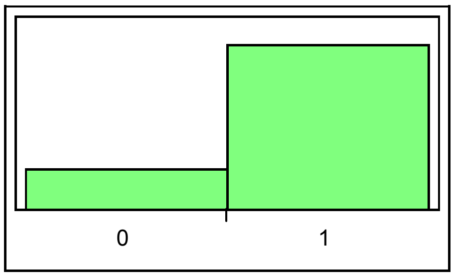
R&D Total Cost Growth? (Genest 2004)



Frequencies

Level	Count	Prob
0	43	0.31852
1	92	0.68148
Total	135	1.00000
N Missing	0	
2 Levels		

Total R&D Cost Growth? - 80% Database (Cost)



Frequencies

Level	Count	Prob
0	13	0.20313
1	51	0.79688
Total	64	1.00000
N Missing	0	
2 Levels		

Figure 12. Logistic Response Distribution Comparison (RDT&E Total Cost?)

Next, we look at the distribution of our new predictor variable *# of Rebaselines* (See Figure 13). The predictor variable distribution is within normal limits, however, we do notice that very few programs experience zero rebaselines (only four programs). Since the distribution of the logistic *y* response variable in our model is different from Genest and the disparity of our 1's and 0's are significantly skewed, we feel it is again unessential to attempt logistic regression with our new predictor variable. Indeed, due to such instability in our logistic response variable, we do not attempt logistic regression. For more information on the logistic regression model associated with total RDT&E schedule growth, see Genest (2004).

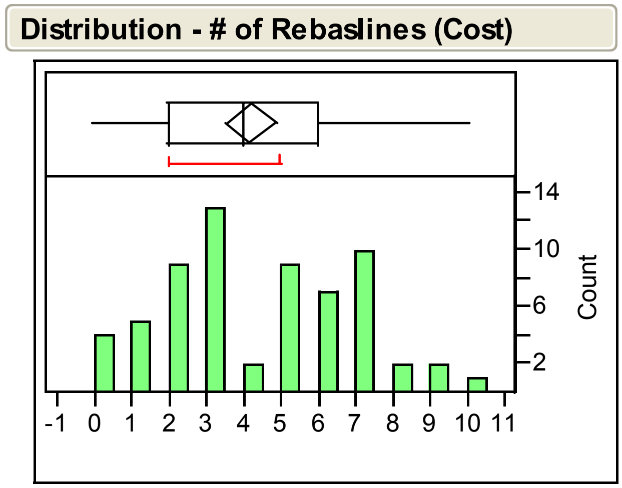
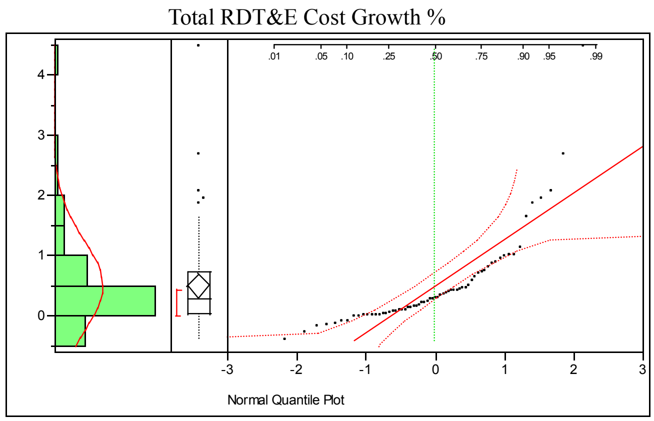


Figure 13. Distribution of Predictor Variable - # of Rebaselines (Cost)

Multiple Regression Results (RDT&E Total Cost)

Since we are unable to perform logistic regression, we evaluate programs that have positive schedule slip via multiple regression. The programs that have no cost growth or negative cost growth are excluded from the model in a similar fashion to Genest's research. This resulted in our research excluding 13 programs from our model, which brings our total number of programs down to 51 in our working database. We then to recreate Genest's model and insert the same predictor variables to validate the robustness of the model against our modified database. In order to reconstruct the Genest model, we need to transform the dependent variable, which mirrors Genest's findings due to the distribution of the residuals; they can not pass normality. First, we verify this with a Shapiro-Wilk test for normality. Indeed, we fail to pass normality and the residual overlay plot also takes on a clear pattern (See Figure 14). This flaring pattern does not indicate homogeneity of the residuals and is often the result of highly skewed data;

therefore, we take transformational measures. We then create a log transformation of the *y* dependent variable.



Shapiro-Wilk W Test

W	Prob<W
0.727410	<.0001

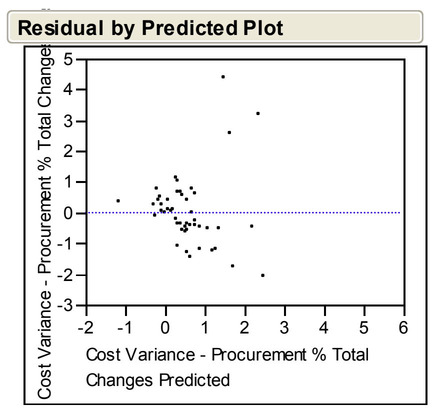
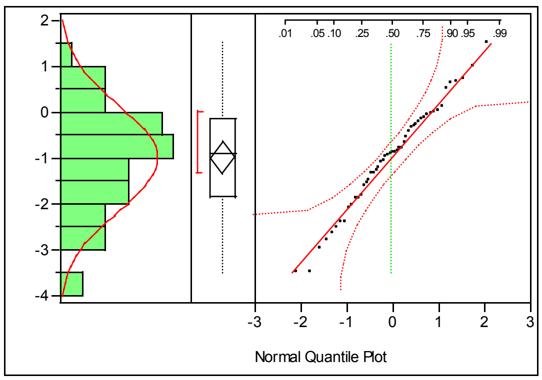


Figure 14. Distribution of Y and Residual Plot of Untransformed Model (Cost)

Once we transform the *y* response, we can see that the transformation now passes normality consistent with past research (see Figure 15). We can now evaluate Genest's multiple regression model. Genest's multiple regression model performed in a similar manner as described in their research, however, Genest's adjusted R^2 is significantly different from the original model, which we can account for because of the differing makeup of our database. The adjusted R^2 in Genest's original model was 0.362047 and the model demonstrated an R^2 of 0.2234 against our modified database. It is noteworthy to mention, that not all Genest's predictor variables still perform in a prognostic manner. The following predictor variables had *p*-values greater than 0.30, which is far higher than the standard of 0.05 outlined in Genest's model: (*LRIP Planned?, Northrop Grumman)*. Since some of Genest's predictor variables performed atypically, we are unable to fully utilize the model as a backdrop for testing our new variable *(# of Rebaselines*).

Total RDT&E Cost Growth %



Goodness-of-Fit Test

Shapiro-Wilk W Test

W	Prob<W
0.986996	0.8528

Residual by Predicted Plot

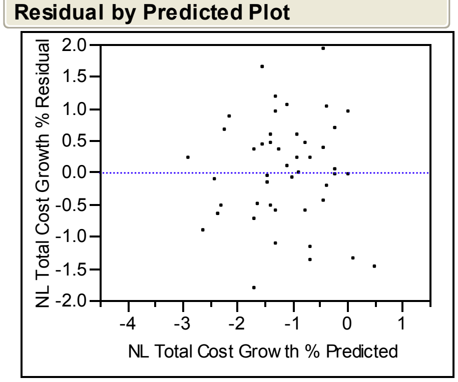


Figure 15. Distribution of Y and Residual Plot of Transformed Model (Cost)

Now that we have a suitably transformed y response variable, we now look to our new variable by itself in relation to Genest's response variable. When we run the model in JMP® using multiple regression, we find no predictability with the variable *# of Rebaselines* when it is used separately in the model. We also looked to prior research for additional predictor variables that could possibly interact with our new variable. We could not demonstrate *# of Rebaselines* to be predictive by itself or with other known predictor variables. However, we created a model by modifying the variable into a discrete format. This change produced a binary response called *Two Rebaselines?*. If a program had two rebaselines, it would be coded a 1 and if it did not have exactly two rebaselines, it was coded a 0. We include a table showing the variables in our final model below in (See Table 15).

Table 15. Multiple Regression Model Variables (Cost)

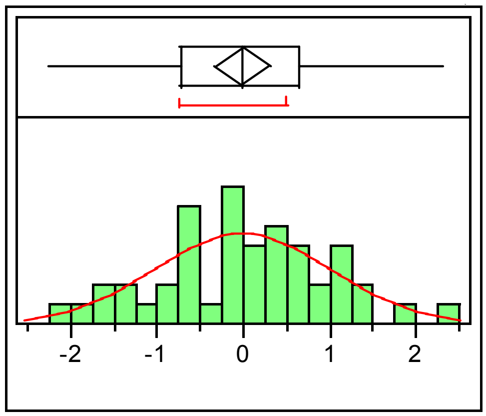
RDT&E Total Cost Multiple Regression Model (Cost)	
Y Response Variable	**Predictor Variables**
Total R&D Cost Growth %	Funding Yrs of R&D Completed
	Fixed-Price EMD Contract?
	Two Rebaselines?
	Sea

When we rerun the model with our new modified variable *Two Rebaselines?*, we find it considerably predictive. Our preliminary run shows an adjusted R^2 value of 0.40203, which is significantly higher than Genest's original model that achieved an R^2 of 0.36205. Due to missing data from predictor variables, our database is limited to only

46 data points, which is similar to the experience of Genest's research (Genest, 2004:43 & 49).

Before we move to the validation results, we first test the residuals of our multiple regression model to ensure the assumptions of the residuals are clearly met. The first assumption is independence. We follow the same assumptions as previously mentioned with our schedule research in that we assume independence is met. We again lend credence to this due to the fact that we use only one SAR to obtain data for only one program except for the *Two Rebaselines?* variable. The second assumption is normality of the residuals in which we test the assumption of normality by performing a Schapiro-Wilk goodness-of-fit test. The last assumption is constant variance of the error term. We perform a Breusch-Pagan test for constant variance (Neter et al., 1996:239).

We now test for normality and constant variance of the error term. In order to test the normality of the error term, we run a distribution of our studentized residuals in JMP® and convert our residuals to a standard normal distribution with a mean of zero and a standard deviation of one to also look for outliers. Figure 16 below displays the distribution of our model.



Goodness-of-Fit Test

Shapiro-Wilk W Test

W	Prob<W
0.989995	0.9589

Figure 16. Studentized Residuals - Distribution (Cost)

Based on the fact that all data points are within three standard deviations of the mean, no outliers appear to exist in our model. Also, using an alpha of 0.05 as our threshold, we satisfy the assumption of normality with a *p*-value of 0.9589. The last assumption that we test is constant variance. Using Microsoft Excel®, we calculate a *p*-value of 0.606381 for the Breusch-Pagan test. We again compare this *p*-value to an alpha of 0.05 indicating our error term is within normal limits.

One last test we perform before we are fully satisfied with our model for validation is the Cook's Distance test to identify influential data points. We use JMP® to graph an overlay plot of the Cook's Distance values. Figure 17 shows that we do not have any influential data points in our model. All values are less than 0.5 with no one program outside the norm (Neter et al., 1996:381).

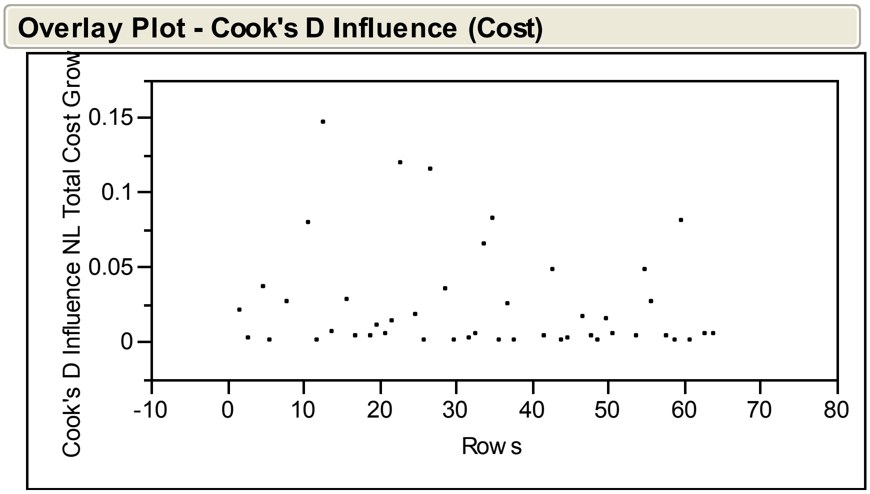


Figure 17. Cook's Distance Influence (Cost)

Based on the performance measures, both individually and collectively, we believe our model provides us with a robust model. Also, with successful testing of the assumptions of the error term and the lack of outliers within our data set, we are certain in the ability of our model to correctly predict the degree of RDT&E total cost growth variance in a program. However, given the extreme change in our new variable, we are cautious to recommend the model for its parsimoniousness. Especially since our research interests are to test the variable *# of Rebaselines* against the Genest (2004) database where the variable falls short in predictability. We cover more concerning this shortcoming in Chapter V. With that said, we can now subject our model to validation.

Located in Appendix F, we include the complete JMP® analysis of our final model *Total RDT&E Cost Growth %.* The final model also includes the parameter estimates with the variance inflation factors (VIF) scores. All of our VIF scores are below two, therefore, we have not over fit our model and each of our independent variables is explaining a different fraction of the variability in the model as noted in the R^2 (Yu, 2004).

Validation Results (RDT&E Total Cost)

To validate our regression model, we add back the 20% of our database previously removed for model building. We begin our validation by computing the estimated prediction intervals of each program based on the parameter estimates and values of the independent variables within a 95% prediction interval. JMP® computes the upper and lower bounds based on the 95% prediction interval. We then look at the actual dependent responses of each program to see if they fall within the upper and lower bound computed by JMP®. The more programs that fall within the bounds of its prediction intervals, the greater success we have with our model. Of the 16 programs in the 20% validation set, 5 programs demonstrate no schedule growth and two programs are missing data which brings the number of usable programs to 9. This brings our validation set to 11.3 % which is far below our goal of 20%. Nevertheless, we proceed cautiously with our validation set. Two programs of our validation set of 9 programs fell outside the bounds of our test which equates to 78% prediction accuracy. This is an understandable finding due to the small number of validation programs and the diminutive number of programs that demonstrated cost growth (9).

To compensate for the major limitation in our validation set, we use the entire database and perform the same validation procedures as previously described in our schedule growth section. When we compile 100% of the database, we have a total of 63 programs that experience cost growth. Eight of these programs are missing data which gives our research 55 usable programs for validation. Only four programs fail to fall within the upper and lower bounds with a 95% prediction interval. Therefore, 93% of our programs accurately predict schedule growth. Furthermore, we also took the same conservative approach as we did with the validation of our schedule model. We look at the mean confidence intervals of each program and set the bounds to a 95% prediction interval. Of the 55 usable programs, 21 programs pass this stringent test for a 38% prediction rate. With the validated schedule model complete, we review Chapter IV.

Chapter Summary

This chapter demonstrates our methodology in action. In Chapter III, we make obvious the need to research our new variable and compare it against established research of others in this area. Subsequently, we decide to observe our new variable in relation to cost growth and schedule growth using the databases from prior research in this vein.

Furthermore, we use the research of Monaco (2005) as a foundational model to evaluate our variable in association with schedule growth. Early in our preemptive analysis of our schedule database, we find our database variable is not suited for logistic regression due to the skewed dichotomy between the 1's and 0's of the responses in our database which leads us to reject logistic regression. However, we find Monaco's model to be very predictive with our modified schedule database for multiple regression. We then evaluate our new variable *# of Rebaselines* against the already established variables by Monaco and find Monaco's model significantly more predictive with our new variable. We also find that are schedule model is highly predictive and parsimonious from our validation results.

Finally, we attempt to use the research of Genest (2004) in a similar fashion as we did with Monaco (2005) in association with total RDT&E cost growth. Again, we find that our modified database is not well suited for logistic regression for the same reasons from our research with schedule growth. We look at the Genest's model and database and find that the model is not as predictive with our modified cost database. In fact, two of Genest's predictor variables are not enhancing the model so they are removed. In addition, we can not demonstrate any predictive ability of our new variable *# of Rebaselines* with the cost database. However, when we modify the variable into a discrete form called *Two rebaselines?*, we derive at a predictive model. This hybrid cost model passes all common statistical assumptions and validates as a predictive model. We are not secure concerning the utility of the cost model and discuss this matter later in Chapter V. In fact, we have generated a number of interesting discoveries in addition to our cost model in which we extrapolate in Chapter V.

V. Conclusions

Chapter Overview

In this chapter, we draw all our research together into a cohesive package. We first look at the original problems that we face in Chapter I and gauge our success on the basis of our research objectives. We then review our literature results and summarize our findings. Subsequently, we restate our methodology and evaluate our results. Finally, we address our limitations, provide our recommendations, and suggest possible follow-on theses for future research.

Explanation of the Problem

Data analysis with genuine data is complex and full of uncertainty. One may compound the complexity even more when their database is riddled with missing data. One such database is that of Monaco (2005). Monaco expounds upon this by highlighting the fact that missing data is the primary limitation. Indeed, Monaco (2005) finds that approximately 27% of programs that otherwise meets the researcher's criteria does not have a reported value for one of the four necessary schedule dates, e.g. planned and actual dates for Milestone II and Milestone III. Of the programs missing the appropriate schedule dates, Planned MS II, Actual MS II, Planned MS III, and Actual MS III did not have complete data 56%, 28%, 72%, and 56% of the time respectively. Monaco further states that one of the most promising predictor variables, FUE, is only present 19.4% of the time (Monaco, 2005:110). From our identified problem, we then formulate objectives for our research.

The objectives of this thesis as we describe in Chapter I are threefold. First, we explore the reasons for the missing data via our literature review in Chapter II. Second, our research attempts to build up the current database and add potential enhancing variables to it which we explain in Chapter III. Third, this investigation reruns the Monaco (2005) model and possibly other models and compares the results and determines if the model(s) remain stable and consistent with the original research. This thesis attempts to use the same logistic regression and multiple regression model(s) used in the research of Monaco and others to validate and compare results. We find the answer to our first research objective in our literature review in Chapter II, which we summarize in the next paragraph.

Review of Literature

We find the primary reasons for missing data in our literature review. DoD Regulation 5000.2-R proved very helpful to our research; it states that DoD requires PMs to minimally include schedule dates for program initiation, major decision points, and the attainment of initial operating capability (IOC). This vagueness allows PMs creative license to include a potpourri of schedule dates that suit their specific needs, however, this is not standardized across MDAPS (DoD, 2001). This finding is significant because Monaco (2005) and Moore (2003) found FUE as a very powerful predictor variable, however, since it is primarily an Army schedule parameter, it is missing across the majority of Major Defense Acquisition Programs (MDAPs). Another reason for missing data is that some of the data is classified and therefore we can not report the data which is emphasized by the 2005 GAO report as a primary problem with the SAR (Levin, 2005).

One final reason is found in the aforementioned RAND study and GAO reports, which highlight the fact that the SAR reporting requirements have changed significantly over the past three decades, which explains the large variation in schedule data. Indeed, we refer you back to the end of Chapter II to Table 6 and Table 7 to further emphasize the SAR's astonishing historical evolution. With one objective answered, we find the response to our second research objective in Chapter III (Methodology).

Review of Methodology

Before we move our research to statistical analysis, we must first answer our second research objective; we attempt to build up the current database and add potential enhancing variables to it. We start with a literature review as a template to acquire direction for our research. We then scour DoD for possible data sources, which results in no other possible avenues for new data that pertains to MDAPs other than the SAR. With a portion of our second objective unachievable, we need to approach our schedule variable problem from a different perspective. Indeed, we search for new variables that look at MDAPs from a longitudinal perspective. In fact, the 2005 GAO report criticizing programmatic rebaselining is the catalyst that we need for the creation of a new variable, *# of Rebaselines* (Levin, 2005). With a spark of hope towards achieving a portion of our second objective, we now move our research towards a methodology for statistical analysis to frame the answer to our second and third research objectives.

With our second and third research objectives, we combine them to structure our statistical methodology; we need to test our new variable against established research in this vein of study. We choose to evaluate the work of Monaco (2005) and Genest (2004) because Monaco takes a broad look at schedule and Genest evaluates RDT&E cost growth with a big picture approach. We feel using these models will best answer our last two research objectives. Our intentions are not to recreate new models because that causes redundancy and our outcome would be similar to the past and not add to the body of knowledge. We start our methodological approach by reviewing the past research of Sipple (2002) who initiated using the two step regression model approach which Monaco (2005) and Genest (2004) followed.

Then we prepare our database for analysis. We randomly create an 80% working database and reserve 20% of the database for validation. We need to do this procedure for two databases, one for schedule and one for total RDT&E cost. We also try to use the exact same methods of selecting programs for each database that Monaco and Genest use for analysis. We then adopt the two step model. The first step starts with logistic regression using a binary response that predicts weather or not schedule growth or cost growth will occur. The second step uses multiple regression to predict the degree or percentage that each program will experience schedule or cost growth. It is noteworthy to mention that since we evaluate our new variable against two separate models, we believe it is paramount to report each model family separately (schedule and cost) to avoid confusion. Now that we have a basic plan of attack, we review the variables of both past schedule and cost models from Monaco and Genest and then attempt to reduplicate their results and then analyze the models with our new variable which we perform in Chapter IV. We restate our results in the next paragraph.

Restatement of Results

We use the research of Monaco (2005) as a foundational model to evaluate our variable in association with schedule growth. Early in our preemptive analysis of our schedule database, we find our database variable is not suited for logistic regression due to the skewed dichotomy between the 1's and 0's of the responses in our database which leads us to reject logistic regression. However, we find that Monaco's model is very predictive with our modified schedule database for multiple regression. We then evaluate our new variable *# of Rebaselines* against the already established variables by Monaco and find Monaco's model significantly more predictive with our new variable. We also find that our schedule model is highly predictive and parsimonious from our validation results. One of the most striking results to our research is that the research of Monaco (2005) is very robust even with a modified database and fewer programs. Furthermore, our new variable, *# of Rebaselines,* boosted the adjusted R^2 of our model (.836442) by approximately 70% compared to Monaco's original model pitted against our database. We can theorize why this is true because each time a program rebaselines, it runs the risk of delaying program development, which leads to schedule growth.

We are pleased that this variable greatly increases the predictability of Monaco's research. We include the final model parameter results for review in Figure 18. By glancing at the predictor variables, one can theorize the weight of each predictor variable in our model. You can see that one unit change in *# of Rebaselines* increases the expected percentage of schedule slip by roughly 7.3%. To further see how the variables from Monaco (2005) interact, we encourage reviewing Monaco's *Explanation of the Independent Variables in Final Models* which we include in Appendix G. In addition, our model passes all the assumptions and the validation results are substantially significant. In fact, of the 27 usable programs for validation, only one program failed to fall within the upper and lower bounds with a 95% confidence interval. Therefore, 96.3% of our programs accurately predict the percentage of schedule growth.

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	0.8758177	0.144169	6.07	<.0001	.
Actual Phase I + Planned EMD length	-0.009554	0.001079	-8.85	<.0001	1.1611273
MS III before IOC?	0.2125676	0.073656	2.89	0.0095	1.1766921
Modification?	-0.263873	0.069807	-3.78	0.0013	1.2700835
# of Rebaselines	0.0728925	0.021698	3.36	0.0033	1.4107758

Figure 18. Review of Final Model Parameters (Schedule)

Finally, we attempt to use the research of Genest (2004) in a similar fashion as we did with Monaco (2005) in association with total RDT&E cost growth. Again, we find that our modified database is not well suited for logistic regression for the same reasons from our research with schedule growth. We look at the Genest's model and database and find that the model is not as predictive with our modified cost database. In fact, two of Genest's predictor variables are not enhancing the model so they are removed. In addition, we can not demonstrate any predictive ability of our new variable *# of Rebaselines* with the cost database.

However, when we modify the variable into a discrete form called *Two rebaselines?*, we derive at a predictive model. This hybrid cost model passes all common statistical assumptions and validates as a predictive model. We are not secure concerning the utility of the cost model since we made a number of changes to the predictor variables and modified our own variable. In fact, we make so many changes to the original RDT&E cost model that we suggest that a future researcher evaluate this model for usefulness. Although we have accomplished great strides in our research, we have a number of limitations, which we discuss next.

Limitations

Our primary limitation parallels the research of its predecessors; we suffer from missing data. In fact, all of our statistical limitations are caused by missing data. Indeed, we were not able to add to our database for the same reasons stated by Monaco (2005) and others. Ironically, one of our research objectives is to attempt to fill in the missing dates. We therefore could not oblige that objective due to the fact that there are not any other credible sources for that data. In addition, our data is even further restricted by reason that we can not gather rebaseline data on all our programs from both databases (schedule and cost). For example, the Monaco schedule database originally contained 68 programs, however, 19 of those programs did not have rebaseline data available from OSD. This leaves us with 49 programs minus the validation set of 10 programs, which results to 39 programs in our 80% working database. Of those 39 programs, they are missing the same data mentioned in Monaco (2005).

With the aforementioned missing data and the one program that we remove because it is an influential data point, we are left with 24 programs for our schedule database. Indeed, this leaves us with a 6:1 ratio of data points to each predictor variable which may limit the power of our test. However, since we are testing theory and using a proven model, we believe that our model is still very robust when we compare it to Monaco's research. We include a table showing the comparison of data/models between our database and the database of Monaco (2005) in Table 16. We also are missing data with the Genest cost database in a similar manner; however, we start out with more programs (135) when compared with the Monaco database. We also do not compare model descriptive statistics with the Genest (2004) model since our cost model is so dissimilar.

Table 16. Comparison Regression Results of Monaco (2005) and Cross (2006)

Multiple Regression Model	Monaco (2005)	Cross (2006)
R^2	0.7674	0.86489
Adjusted R^2	0.7425	0.8364
Data Point: Variable Ratio	10.7:1	6:01
P-Value Summation	0.0408	0.0143
# of Observations	32	24

The fact that we are missing data also affects our ability to perform logistic regression in both models (schedule & cost) since the dichotomy between our 1's and 0's responses are so different and the ratio between 1's and 0's are completely skewed. For example, the percentage of programs that demonstrated schedule slips in the full Monaco database are 25% and our newly created 80% working database is only 12.8%.

Additionally, since we have limited data, this also spills into our validation results. For example, in our schedule database, our schedule validation set starts with 10 programs. Three programs demonstrate schedule slip; this equates to an 8% validation set. We proceed with validation even though we fall far below the established 20%. Of the three programs, two fell within the bounds of our test, which equates to a 67% prediction accuracy. We understand this as a suitable finding due to the small number of validation programs and the diminutive number of programs that demonstrated schedule slip (3).

To compensate for the major limitation in our validation set, we use the entire database and perform the same validation procedures as previously described. When we compile 100% of the database, we have a total of 36 total programs that experience schedule growth. Nine of these programs are missing data primarily from the variable *Planned IOC Date* similar to Monaco (2005) and therefore unusable (Monaco, 2005:97).

Our subsequent limitation concerns our entire RDT&E cost database which is patterned from Genest (2004). First, we are not able to use logistic regression because of what we previously mention concerning the skewed dichotomous response variable. Second, we look at Genest's model and database and find that the model is not as predictive with our modified cost database. In fact, two of Genest's predictor variables are not enhancing the model so we remove them. In addition, we can not demonstrate any predictive ability with our new variable *# of Rebaselines* with the cost database.

However, when we modify the variable into a discrete form called *Two rebaselines?*, we derive at a predictive model. Nevertheless, the original model is comparatively different. We use additional predictive variables from past research, however, our model is far dissimilar. This hybrid cost model passes all common statistical assumptions and validates as a predictive model, however, we recommend the analyst to use caution with this model since it is grossly different than the original model. In fact, since our scope is not to build entirely new models, we believe this model requires further research before it is used in the field. We include Figure 19 to further emphasize the diversity between the two models.

Genest (2004) [R^2=0.3620]				
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-1.070473	0.784431	-1.36	0.1781
Northrop Grumman	1.3557629	0.664538	2.04	0.0463
Funding Yrs of R&D Completed	0.132762	0.025576	5.19	<.0001
Maturity of EMD at IOC%	-1.929685	0.813505	-2.37	0.0214
Prototype?	0.8669499	0.346592	2.5	0.0155
Significant pre-EMD activity	-0.968515	0.325376	-2.98	0.0044
LRIP Planned?	0.7522629	0.302415	2.49	0.0160
Cross (2006) [R^2=0.4203]				
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-1.907099	0.287564	-6.63	<.0001
Funding Yrs of R&D Completed	0.0783691	0.021204	3.7	0.0006
Fixed-Price EMD Contract?	-1.195237	0.395955	-3.02	0.0044
Two Rebaselines?	1.5694569	0.382439	4.1	0.0002
Sea	-1.152497	0.412615	-2.79	0.0079

Figure 19. Multiple Regression Parameter Comparison between Genest (2004) & Cross (2006)

Our final limitation concerns the lack of published literature on the SAR. Indeed, we could not find a peer reviewed article that is dedicated to the SAR. We had to resort to research from GAO reports, RAND studies, DoD regulations and even DoD internet sources, which causes some concern on our part. However, we feel that the information that we present in this research is credible and accurate. Now that we have presented our limitations, we make future recommendations in the proceeding paragraphs.

Recommendations

The two step regression technique created by Sipple (2002) continues to benefit the analysis community as a valuable tool based on the validation from other researchers like Monaco (2005) for example. We believe that the two step method is superior for detecting schedule and cost growth. In fact, we also recommend that the researcher and the analyst use the techniques of Monaco (2005) for predicting schedule risk. In addition, we find the Monaco (2005) multiple regression model even more predictive by at least 10% when we apply our new variable called *# of Rebaselines*. Although we highly suggest utilizing the schedule model, we do not recommend the analyst utilize our RDT&E cost model until it is further researched due to the drastic changes we make to the predictive model.

Another recommendation is based on our literature review and our search for missing data across DoD. We believe that the SAR has come a long way in 30 years, however, we are concerned that with so much change; programs become very difficult to compare across the board. In fact, we recommend that DoD explicitly identify and increase the number of the required standard schedule milestones to match the same schedule dates that are mentioned in the research of Monaco (2005). With a standard requirement for schedule parameters, the analyst should be able to predict even tighter increments of schedule and cost growth. Another area of concern is associated with spiral development. Since each spiral initiative is different, e.g. each program funds a different number of spirals and prototypes, how will the analyst of the future compare these programs? This is another reason why we think future research should focus longitudinally since we cannot find or recreate missing variables like FUE.

In addition, we would be remiss not to readdress the recommendations by GAO in their 2005 report. We also concur with the recommendations of the 2005 GAO report based on our research of the SAR. The report outlines many recommendations to include transformations in the rebaseline and reporting process. We include their major problems and recommendations in Table 17.

Table 17. 2005 GAO Report Problem/Solution Summary Table

Problem	Solution
Information for Congress on unit cost performance could be more complete and Year-to-year unit cost changes are not measured and reported.	Measure and report a full history of unit cost performance in constant dollars by comparing the latest cost in quantity estimates with: the first full estimate (typically the original acquisition program baseline established at milestone B), the current approved baseline, or if the program rebaselines, the prior approved program baseline, and the estimate established with the previous year's budget request.
Rebaselining can occur during each phase of acquisition, sometimes frequently on individual programs.	N/A
Reporting of rebaselinings could be timelier	Notify Congress when rebaselining actions are approved.
In some cases, DoD reduces the magnitude of unit cost growth for Nunn -McCurdy determinations.	Fully disclose to Congress the nature and extent of programmatic adjustments affecting Nunn-McCurdy thresholds and terminations, pending any congressional direction on this issue.
Unclassified data is unnecessarily restricted.	Separately report classified and unclassified SAR information.

Our last recommendation stems from the database that contains the SAR called Ostrich and the method for presenting the data called DAMIR. As we previously mention in Chapter II, the Ostrich database is not normalized in third normal form. We also previously mention that this is due primarily to the legacy system that feeds into the Ostrich database. In other words, the database is designed to support CARS and now the system is functioning partially in a relational format via DAMIR. Although we do not have all the facts of the genesis of the database, we do have the data dictionary and the shell of the database to examine. We recommend a complete transformation of this database. In our opinion, many systems are limited due to an improper design or have had small evolutionary changes to an old design which renders it unable to adapt to change. Additionally, we recommend using system analysis and design techniques in concert with lean principles to reengineer the process/system as outlined by James P. Womack and Daniel T. Jones, in their book *Lean Thinking* (Womack et al., 2003).

Possible Follow-on Thesis

We further support future researchers using this database and the additional metadata that we have created with this research. We encourage populating the database with newer programs and look for additional variables that take a longitudinal approach such as our variable call *# of Rebaselines.* In addition, we recommend researching the variability of rebaselined dates as mentioned in the 2005 GAO report. The report states further that the first rebaseline estimate date may not stay the same but change. Thus we recommend looking at each rebaseline, starting with the first development estimate from the early SAR as opposed to the development estimate date on the most recent SAR. We suggest this to see if the model demonstrates robustness for changing the development estimate baseline. In addition, we suggest that the researcher attempt to analyze rebaselined information incrementally between each baseline to possibly measure predictability in smaller units. Furthermore, we propose using our model that predicts schedule growth to see if an incremental change in the schedule model influences cost growth.

We also suggest further research with our modified cost model for usefulness. One may possibly combine that model with what we previously mention concerning a future schedule growth-cost growth model. It is quite possible that the researcher may be able to build a highly predictive model based on the Monaco (2005) model with our additional variable and use the results of that model to further define the relationship between schedule growth and cost growth.

One last recommendation for future research concerns the database that contains the SAR information. Earlier we recommended that the Ostrich database takes a business process reengineering transformation to further meet the needs of the DoD analyst. Since this process requires great care and planning, we recommend a joint partnership with the Air Force Institute of Technology (AFIT). There exists the possibility for group research that may save precious DoD money and time. Indeed, this also enhances the graduate-level education of the researchers at AFIT and may lead to other money saving collaborative efforts in the future.

Appendix A

List of Acronyms

ACAT – Acquisition Category
APB – Acquisition Program Baseline
AUC – Area Under the Receiver Operating Characteristic Curve
CAE – Component Acquisition Executive
CARS – Consolidated Acquisition Reporting System
CDR – Critical Design Review
DAE – Defense Acquisition Executive
DAES – Defense Acquisition Executive Summary
DAMIR – Defense Acquisition Management Information Retrieval
DE – Development Estimate
DoD – Department of Defense
EMD – Engineering and Manufacturing Development
FSD – Full Scale Development
FUE – First Unit Equipped
IOC – Initial Operating Capability
Ln – Natural Log
LR – Likelihood Ratio
LRIP – Low Rate Initial Production
MAIS – Major Automated Information Systems
MDAP – Major Defense Acquisition Program
MS – Milestone
PdE – Production Estimate.
PDR – Preliminary Design Review
PE – Planning Estimate
PEO – Program Executive Officer
PM – Program Manager
Proc – Procurement
Qty – Quantity
R&D – Research and Development
RAND – Research and Development Corporation
RDT&E – Research, Development, Testing, and Evaluation.
ROC – Receiver Operation Characteristic
SAR – Selected Acquisition Report
UCR – Unit Cost Reporting
USD – Under Secretary of Defense
USD (AT&L) – Under Secretary of Defense (Acquisition, Training & Logistics)
VIF – Variance Inflation Factor

Appendix B

List of Predictor Variables (From Sipple 2002)

Program Size Variables

- Total Cost CY $M 2002 (Continuous)
- Total Quantity (Continuous)
- Prog Acq Unit Cost (Continuous)
- Qty during PE (Continuous)
- Qty planned for R&D$ (Continuous)

Physical Type of Program

- Domain of Operation Variables (Binary)
    - Air, Land, Space, & Sea
- Function Variables (Binary)
    - Electronic, Helo, Missile, Aircraft, Munition, Land Vehicle, Ship, & Other

Management Characteristics

- Military Service Management (Binary)
    - Services (Svs) >1, Svs >2, Svs>3, Service = Navy Only, Service = Joint, Service = AF Only, Lead Svc = Army, Lead Svc = Navy,Lead Svc = DoD, Lead Svc = AF, AF Involvement, N Involvement, MC Involvement, & AR Involvement
- Contractor Characteristics (Binary)
    - Lockheed-Martin, Northrup Grumman, Boeing, Raytheon, Litton, General Dynamics, No Major Defense KTR, More than 1 Major Defense KTR, & Fixed-Price EMD Contract

Schedule Characteristics

- RDT&E and Procurement Maturity Measures (Continuous)
    - Maturity (Funding Yrs complete), Funding YR Total Program Length, Funding Yrs of R&D Completed, Funding Yrs of Prod Completed, Length of Prod in Funding Yrs, Length of R&D in Funding Yrs, R&D Funding Yr Maturity %, Proc Funding Yr Maturity %, & Total Funding Yr Maturity %
- EMD Maturity Measures (Continuous)
    - Maturity (Funding Yrs complete), Funding YR Total Program Length, Funding Yrs of R&D Completed, Funding Yrs of Prod Completed, Length of Prod in Funding Yrs, Length of R&D in Funding Yrs, R&D

Funding Yr Maturity %, Proc Funding Yr Maturity %, & Total Funding Yr Maturity, Maturity from MS II in mos, Actual Length of EMD, MS II-based Maturity of EMD%, IOC-based Maturity of EMD%, & FUE-based Maturity of EMD%

- Concurrency Indicators (Binary & Continuous)
    - MS III Complete, Proc Started based on Funding Yrs, & Proc Funding before MSIII (Binary)
    - Concurrency Measure Interval & Concurrency Measure % (Continuous)

Other Characteristics

- # Product Variants in this SAR (Continuous)
- Security Classification (Binary) – Class S, Class C, Class U, & Class at Least S
- Risk Mitigation (Binary)
- Versions Previous to SAR (Binary)
- Modification (Binary)
- Prototype (Binary)
- Dem/Val Prototype (Binary)
- EMD Prototype (Binary)
- Did it have a PE (Binary)
- Significant pre-EMD activity immediately prior to current version
- (Binary)
- Did it have a MS I (Binary)
- Terminated (Binary)

Appendix C

List of Predictor Variables (From Monaco 2005)

Program Size Variables

- *1 Planned Total Cost CY $M 2003* – continuous variable which indicates the planned total costs of the program at the time of the DE in CY $M 2003
- *2 Qty Planned for R&D* – continuous variable which indicates the R&D quantity at the time of the DE
- *3 Qty Planned for Proc* - continuous variable which indicates the procurement quantity at the time of the DE
- *4 Total Quantity planned* – continuous variable equal to *2 Qty Planned for R&D* plus *3 Qty Planned for Proc*
- *5 Planned Unit Cost CY $M 2003* – continuous variable which equals *1 Planned Total Cost CY $M 2003* divided by *4 Total Quantity planned*
- *6 ACAT 1*? – binary variable: 1 for yes and 0 for no

Physical Type of Program Variables

- Domain of Operation Variables

    - *7 Air* – binary variable: 1 for yes and 0 for no; includes programs that primarily operate in the air; includes air-launched tactical missiles and strategic ground-launched or ship-launched missiles
    - *8 Land* – binary variable: 1 for yes and 0 for no; includes tactical ground-launched missiles; does not include strategic ground-launched missiles
    - *9 Space* – binary variable: 1 for yes and 0 for no; includes satellite programs and launch vehicle programs
    - *10 Sea* – binary variable: 1 for yes and 0 for no; includes ships and ship-borne systems other than aircraft and strategic missiles

- Function Variables

    - *11 Electronic* – binary variable: 1 for yes and 0 for no; includes all computer programs, communication programs, electronic warfare programs that do not fit into the other categories
    - *12 Helo* – binary variable: 1 for yes and 0 for no; helicopters
    - *13 Missile* – binary variable: 1 for yes and 0 for no; includes all missiles
    - *14 Aircraft* – binary variable: 1 for yes and 0 for no; does not include helicopters
    - *15 Munition* – binary variable: 1 for yes and 0 for no
    - *16 Land Vehicle* – binary variable: 1 for yes and 0 for no

- *17 Space (Rand)* –binary variable: 1 for yes and 0 for no
- *18 Ship* – binary variable: 1 for yes and 0 for no; includes all watercraft
- *19 Other* – binary variable: 1 for yes and 0 for no; any program that does not fit into one of the other function variables

Management Characteristics

- Military Service Management

  - *20 # of Svs* - continuous variable to indicate the number of services involved in the program
  - *21 Svs > 1* – binary variable: 1 for yes and 0 for no; number of services involved at the date of the SAR
  - *22 Svs > 2* – binary variable: 1 for yes and 0 for no; number of services involved at the date of the SAR
  - *23 Svs > 3* – binary variable: 1 for yes and 0 for no; number of services involved at the date of the SAR
  - *24 Service = Navy Only* – binary variable: 1 for yes and 0 for no
  - *25 Service = Joint* – binary variable: 1 for yes and 0 for no
  - *26 Service = Army Only* – binary variable: 1 for yes and 0 for no
  - *27 Service = Marines Only* – binary variable: 1 for yes and 0 for no
  - *28 Service = AF Only* – binary variable: 1 for yes and 0 for no
  - *29 Lead Svc = Army* – binary variable: 1 for yes and 0 for no
  - *30 Lead Svc = Navy* – binary variable: 1 for yes and 0 for no
  - *31 Lead Svc = DoD* – binary variable: 1 for yes and 0 for no
  - *32 Lead Svc = AF* – binary variable: 1 for yes and 0 for no
  - *33 Air Force Involvement* – binary variable: 1 for yes and 0 for no
  - *34 Navy Involvement* – binary variable: 1 for yes and 0 for no
  - *35 Marine Corps Involvement* – binary variable: 1 for yes and 0 for no
  - *36 Army Involvement* – binary variable: 1 for yes and 0 for no

- Contractor Characteristics

  - *37 Lockheed-Martin* – binary variable: 1 for yes and 0 for no
  - *38 Northrop Grumman* – binary variable: 1 for yes and 0 for no
  - *39 Boeing* – binary variable: 1 for yes and 0 for no
  - *40 Raytheon* – binary variable: 1 for yes and 0 for no
  - *41 General Dynamics* – binary variable: 1 for yes and 0 for no
  - *42 No Major Defense Contractor* – binary variable: 1 for yes and 0 for no; a program that does not use one of the contractors mentioned immediately above = 1
  - *43 More than 1 Major Defense Contractor* – binary variable: 1 for yes and 0 for no; programs including more than one of the contractors listed above = 1
  - *44 Fixed-Price EMD Contract* – binary variable: 1 for yes and 0 for no

Schedule Characteristics

- *45 Program has a MS I?* – binary variable: 1 for yes and 0 for no
- *46 Planned Time from MS I to MS II* – continuous variable calculated by subtracting the MS I date from the MS II date based on the DE
- *47 Actual Time from MS I to MS II* – continuous variable calculated by subtracting the current estimate MS I date from the current estimate MS II date recorded in the last DE SAR
- *48 Phase I Slip?* - binary variable: 1 for yes and 0 for no
- *49 Length of Phase I Slip* – continuous variable equal to *47 Actual Time from MS I to MS II* minus *46 Planned Time from MS I to MS*
- *50 Planned EMD Length* – continuous variable calculated by subtracting the MS II date from the MS III date based on the DE
- *51 Actual Phase I + Planned EMD Length* – continuous variable calculated by adding *47 Actual Time from MS I to MS II* and *50 Planned EMD Length*
- *52 Planned Time from MS II to PDR* – continuous variable calculated by subtracting the MS II date from the preliminary design review date based on the DE
- *53 Planned Time from MS II to CDR* - continuous variable calculated by subtracting the MS II date from the critical design review date based on the DE
- *54 Planned Time from MS II to FUE* - continuous variable calculated by subtracting the MS II date from the first unit equipped date based on the DE
- *55 Planned Time from MS II to IOC* - continuous variable calculated by subtracting the MS II date from the initial operational capability date based on the DE
- *56 Planned Time from MS II to production contract award* - continuous variable calculated by subtracting the MS II date from the production contract award date based on the DE
- *57 Planned Time from EMD contract award to MS III* - continuous variable calculated by subtracting the EMD contract award date from the MS III date based on the DE
- *58 Planned Time from EMD contract award to IOC*- continuous variable calculated by subtracting the EMD contract award date from the initial operational capability date based on the DE
- *59 Planned Time from EMD contract award to production contract award*- continuous variable calculated by subtracting the EMD contract award date from the production contract award date based on the DE
- *60 Planned EMD start (proxy for tech complexity)* – continuous variable ascertained by calculating the number of months the planned EMD start date exceeds Jan 1, 1970.
- *61 Planned Time from MS II to LRIP* – continuous variable calculated by subtracting the MS II date from the LRIP date based on the DE

Other Characteristics

• *62 LRIP Planned ?* - binary variable: 1 for yes and 0 for no; indicates if the program had LRIP planned
• *63 Planned LRIP Quantity* – continuous variable which indicates the number of LRIP articles planned based on the DE
• *64 Funding YR Total Program Length* – continuous variable which indicates the total number of years for which the program has either RDT&E funding or procurement funding budgeted
• *65 Funding Yrs of R&D Completed* – continuous variable which indicates the number of years completed for which the program had RDT&E funding budgeted
• *66 Funding Yrs of Proc Completed* – continuous variable which indicates the number of years completed for which the program had procurement funding budgeted
• *67 Length of R&D in Funding Yrs* – continuous variable which indicates the number of years for which the program has RDT&E funding budgeted
• *68 Length of Proc in Funding Yrs* – continuous variable which indicates the number of years for which the program has procurement funding budgeted
• *69 R&D Funding Yr Maturity %* – continuous variable which equals *65 Funding Yrs of R&D Completed* divided by *67 Length of R&D in Funding Yrs*
• *70 Proc Funding Yr Maturity %* – continuous variable which equals *66 Funding Yrs of Proc Completed* divided by *68 Length of Proc in Funding Yrs*
• *71 % R&D funding of total program funding* – continuous variable which equals *67 Length of R&D in Funding Yrs* divided by *64 Funding YR Total Program Length*
• *72 % Proc funding of total program funding* - continuous variable which equals *68 Length of Proc in Funding* divided by *64 Funding YR Total Program Length*
• *73 Proc Started based on Funding Yrs* – binary variable: 1 for yes and 0 for no; if procurement funding is budgeted in the year of the SAR or before, then = 1
• *74 Proc Funding before MS III* – binary variable: 1 for yes and 0 for no
• *75 # Product Variants in this SAR* – continuous variable which indicates the number of versions included in the EMD effort that the current SAR addresses
• *76 Class – S* – binary variable: 1 for yes and 0 for no; security classification Secret
• *77 Class – C* – binary variable: 1 for yes and 0 for no; security classification Confidential
• *78 Class – U* – binary variable: 1 for yes and 0 for no; security classification Unclassified
• *79 Risk Mitigation* – binary variable: 1 for yes and 0 for no; indicates whether there was a version previous to SAR or significant pre-EMD activities

- *80 Versions Previous to SAR* – binary variable: 1 for yes and 0 for no; indicates whether there was a significant, relevant effort prior to the DE; a pre-EMD prototype or a previous version of the system would apply
- *81 Modification* – binary variable: 1 for yes and 0 for no; indicates whether the program is a modification of a previous program
- *82 Prototype* – binary variable: 1 for yes and 0 for no; indicates whether the program had a prototyping effort
- *83 Dem/Val Prototype* – binary variable: 1 for yes and 0 for no; indicates whether the prototyping effort occurred in the PDRR phase
- *84 EMD Prototype* – binary variable: 1 for yes and 0 for no; indicates whether the prototyping effort occurred in the EMD phase
- *85 PE?* – binary variable: 1 for yes and 0 for no; indicates whether the program had a Planning Estimate
- *86 Significant pre-EMD activity immediately prior to current version* – binary variable: 1 for yes and 0 for no; indicates whether the program had activities in the schedule at least six months prior to MSII decision
- *87 Funding Yrs of R&D Complete <= 10?* – binary variable which indicates if *65 Funding Yrs of R&D Completed* is less or equal to 10 years: 1 for yes 0 for no
- *88 Funding Yrs of Proc Complete <= 6?* – binary variable which indicates if *66 Funding Yrs of Proc Completed* is less than or equal to 6 years: 1 for yes 0 for no
- *89 Length of R&D Funding > 14 yrs?* – binary variable which indicates if *67 Length of R&D in Funding Yrs* exceeds 14 years: 1 for yes 0 for no
- *90 Length of Proc Funding >= 12 yrs?* – binary variable which indicates if *68 Length of Prod in Funding Yrs* exceeds or equals 12 years: 1 for yes 0 for no
- *91 R&D Funding Yr Maturity % > 80%?* – binary variable which indicates if *69 R&D Funding Yr Maturity %* exceeds 0.8: 1 for yes 0 for no
- *92 Proc Funding Yr Maturity % > 50%?* – binary variable which indicates if *70 Proc Funding Yr Maturity %* exceeds 0.5: 1 for yes 0 for no
- *93 LRIP Planned Quantity < 7?* – binary variable which indicates if *63 Planned LRIP Quantity* is less than 7: 1 for yes 0 for no
- *94 Planned Total Cost CY $M 2003 < 13,000?* – binary variable which indicates if *1 Planned Total Cost CY $M 2003* is less than 13,000 CY $M 2003: 1 for yes 0 for no
- *95 Planned Time from MS II to PDR < 9 months?* – binary variable which indicates if *52 Planned Time from MS II to PDR* is less than 9 months: 1 for yes 0 for no
- *96 Planned Unit Cost CY $M 2003 < 5?* – binary variable which indicates if *5 Planned Unit Cost CY $M 2003* is less than 5 CY $M 2003: 1 for yes 0 for no
- *97 MS III before IOC?* – binary variable which indicates if the planned MS III date occurs before the planned IOC date: 1 for yes 0 for no

Appendix D

Final Model (Schedule)

Final Model - Response EMD Schedule Variance % (positive only)

Summary of Fit

RSquare	0.864887
RSquare Adj	0.836442
Root Mean Square Error	0.151197
Mean of Response	0.482965
Observations (or Sum Wgts)	24

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	2.7803596	0.695090	30.4057
Error	19	0.4343502	0.022861	Prob > F
C. Total	23	3.2147098		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	0.8758177	0.144169	6.07	<.0001	.
Actual Phase I + Planned EMD length	-0.009554	0.001079	-8.85	<.0001	1.1611273
MS III before IOC?	0.2125676	0.073656	2.89	0.0095	1.1766921
Modification?	-0.263873	0.069807	-3.78	0.0013	1.2700835
# of Rebaselines	0.0728925	0.021698	3.36	0.0033	1.4107758

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Actual Phase I + Planned EMD length	1	1	1.7908816	78.3394	<.0001
MS III before IOC?	1	1	0.1904000	8.3288	0.0095
Modification?	1	1	0.3266500	14.2888	0.0013
# of Rebaselines	1	1	0.2579875	11.2853	0.0033

Residual by Predicted Plot

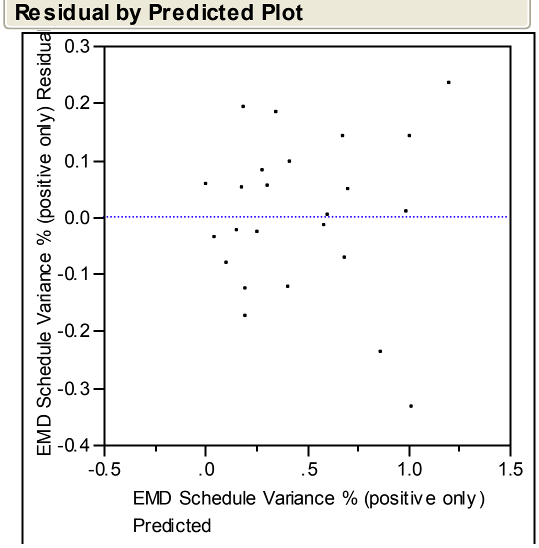


Figure 20. Multiple Regression Final Model - (Schedule)

Appendix E

Validation Table (Schedule) **Table 18. 100% - Validation of Multiple Regression Model - EMD Schedule Variance % (Schedule)**

Multiple - EMD Schedule Variance % (100% Validation) - [Schedule]							
Program #	Validation EMD%	Lower 95% Indiv Validation EMD% (+Only)	Upper 95% Indiv Validation EMD% (+Only)	Passed Indv Validation	Lower 95% Mean Validation EMD% (+Only)	Upper 95% Mean Validation EMD% (+Only)	Passed Mean Validation
17	0.029411765	-0.24729238	0.470713628	YES	-0.0578089	0.281230143	YES
26	0.621621622	M	M	M	M	M	M
28	0.287878788	0.074948129	0.750269789	YES	0.294844171	0.530373748	NO
32	0.615384615	0.257691497	0.968595998	YES	0.451279571	0.775007925	YES
35	0.237288136	-0.08382901	0.617671336	YES	0.115661811	0.418180518	YES
36	0.049180328	M	M	M	M	M	M
39	0.823529412	0.352587906	1.017344101	YES	0.583334763	0.786597245	NO
41	0.536585366	-0.00728624	0.71764389	YES	0.178445602	0.53191205	NO
44	0.258064516	M	M	M	M	M	M
57	0.573770492	0.256270484	0.922576397	YES	0.485286022	0.693560859	YES
68	0.137254902	-0.18653833	0.512794351	YES	0.014399203	0.31185682	YES
70	0.06779661	-0.33133091	0.358933752	YES	-0.12393033	0.15153317	YES
72	1.43902439	0.845742072	1.568075111	YES	1.032854105	1.380963078	NO
73	0.182926829	M	M	M	M	M	M
81	0.941176471	M	M	M	M	M	M
82	0.368421053	-0.02517241	0.656955444	YES	0.18869753	0.443085506	YES
83	0.233333333	-0.17746347	0.546412205	YES	0.008824987	0.360123753	YES
88	0.688888889	0.68292924	1.362883739	YES	0.898655777	1.147157202	NO
89	0.571428571	M	M	M	M	M	M
90	0.631578947	0.536860543	1.203216232	YES	0.765821377	0.974255399	NO
94	1	0.635517014	1.353164991	YES	0.82520093	1.163481074	YES
99	1.152173913	0.673908725	1.352795742	YES	0.89056948	1.136134987	NO
103	0.211111111	M	M	M	M	M	M
104	0.013157895	-0.30749398	0.411725434	YES	-0.11868488	0.222916341	YES
107	0.517241379	0.076276424	0.76824457	YES	0.282408117	0.562112877	YES
108	1	M	M	M	M	M	M
109	0.754098361	0.358499989	1.056926633	YES	0.560052708	0.855373914	YES
111	0.454545455	M	M	M	M	M	M
112	0.077586207	-0.15605002	0.565808487	YES	0.031317622	0.37844085	YES
113	0.384615385	-0.14381852	0.529540642	YES	0.077940011	0.307782115	NO
120	0.620689655	0.336116178	1.050331636	YES	0.527755965	0.858691848	YES
122	0.031578947	-0.12684946	0.543705589	YES	0.097681826	0.3191743	NO
133	0.369863014	-0.0480574	0.629807114	YES	0.169512656	0.412237058	YES
58	1.40625	0.567463314	1.289405102	NO	0.754786012	1.102082404	NO
60	0.369863014	0.072892797	0.80042707	YES	0.25727142	0.616048447	YES
86	0.157894737	-0.1837769	0.526339535	YES	0.010284281	0.332278358	YES

	Total Programs	Programs Missing Data	Total Usable Programs	% Pass			% Pass
	36	9	27	96.30%			59%

Appendix F

Final Model (Cost)

Final Multiple Regression Model (Total RDT&E Cost Growth %)

Summary of Fit

RSquare	0.47183
RSquare Adj	0.420301
Root Mean Square Error	0.862254
Mean of Response	-1.09931
Observations (or Sum Wgts)	46

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	27.231155	6.80779	9.1566
Error	41	30.482766	0.74348	Prob > F
C. Total	45	57.713921		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	21	15.806209	0.752677	1.0257
Pure Error	20	14.676557	0.733828	Prob > F
Total Error	41	30.482766		0.4788
				Max RSq
				0.7457

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	-1.907099	0.287564	-6.63	<.0001	.
49 Funding Yrs of R&D Completed	0.0783691	0.021204	3.70	0.0006	1.227627
46 Fixed-Price EMD Contract?	-1.195237	0.395955	-3.02	0.0044	1.1002052
2 Rebaselines?	1.5694569	0.382439	4.10	0.0002	1.1675066
11 Sea	-1.152497	0.412615	-2.79	0.0079	1.0205072

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
49 Funding Yrs of R&D Completed	1	1	10.155589	13.6595	0.0006
46 Fixed-Price EMD Contract?	1	1	6.774669	9.1121	0.0044
2 Rebaselines?	1	1	12.521152	16.8412	0.0002
11 Sea	1	1	5.800426	7.8017	0.0079

Residual by Predicted Plot

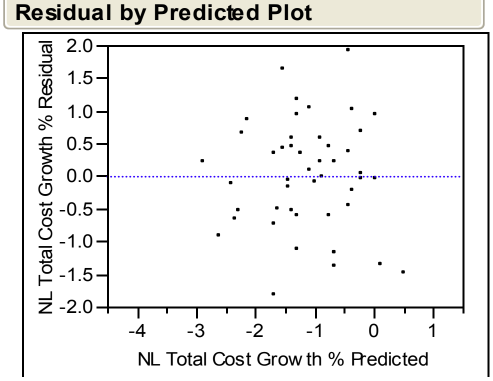


Figure 21. Multiple Regression Final Model (Cost)

Appendix G

(Monaco 2005) Explanation of the Independent Variables in Final Models

The logistic regression model uses planned EMD length as one variable to predict the probability of schedule growth for a future program and the multiple regression model uses actual time from MS I to MS II plus planned EMD length as one variable to predict the percent of schedule growth. That is, as planned EMD length increases the probability and magnitude of schedule growth decreases. It is logical to assume that a longer planned EMD length will decrease the probability of schedule growth. And if growth is expected, the longer duration should lesson the percentage of possible schedule growth. This intuition is validated by the effect this variable has on both response variables.

The predictor variable that has the greatest impact on determining the probability of schedule growth in the logistic regression model is variable *96 Planned Unit Cost CY $M 2003 < 5?*. This variable is calculated by dividing the total planned cost of a program by the total planned quantity of units. Variable 96 is in the top logistic regression model for each generation of model building from the one variable to final six variable model. Programs with per unit costs less than $5M have a greater probability of schedule slip. Just by the nature of the formula for variable 96, when the total quantity planned is high, or the total planned cost is low, planned unit cost will be low. In either case, variable 96 is usually coded as a '1'. During the model building process, we find that as planned total quantity increases, the probability of schedule growth increases and as planned total cost decreases, the probability of schedule growth increases as well. Perhaps the increased probability of schedule growth for programs with low per unit costs brings light to two underlying issues within weapon system developments. First, technical difficulty and unscheduled delays may affect schedule growth for programs with a large quantity of units more than programs fielding only a few units. Second, programs with lower total costs may fall under the radar scope for oversight.

We test this theory by looking further into the aircraft programs in our database. Variable 96 is coded '0' for all aircraft programs in our database meaning all aircraft programs within our database have per unit costs greater than $5M. Aircraft programs have high total costs and low quantity amounts compared to other weapon systems. Based on the parameter estimate of variable 96, we would expect to see aircraft programs remain within their allotted schedule more often than other weapon system types. We find that this is in fact the case in our database where 50% of aircraft programs don't slip compared to only 22% for all other program types. Also, for the programs that slip, aircraft slips by an average of 29.8% compared to 50.3% for all other program types. Previous research states that more management oversight is given to aircraft programs (Tyson, Harmon, and Utech, 1994:S-2). Also, we note that this directly corresponds to the parameter estimate of variable *14 Aircraft* in our final logistic regression model, which shows a decrease in the probability of schedule growth for aircraft programs.

Two of the variables in our final models which are cited in literature numerous times as predictors of schedule growth are: *21 Svs > 1?* and *81 Modification?*. It seems likely that having more than one service involved with a program would increase the likelihood of schedule slippage -- more risk is involved. Our model confirms this fact as the probability of schedule slippage increases with the existence of more than one branch of the armed forces. The modification variable also mimics the findings in literature where modification programs exhibit less schedule growth. This makes logical sense in that modification programs are more technologically mature.

Of the remaining predictor variables: *25 Service = Navy Only?, 72 % Proc Finding (yrs) of Total Program,* and *97 MS III before IOC?*, we do not find prior studies that directly confirm or refute the results of our models. Although we are unable to find a study that shows Navy only programs have a higher probability of a schedule slip, past literature shows that military service type is a possible predictor of determining schedule duration and the extent of schedule growth. For the variables 72 and 97, we follow with discussion on their behavior and possible meanings.

Variable *72 % Proc Funding (yrs) of Total Program* is calculated as the number of years of procurement funding divided by the total number of years of funding. As the percentage of funding increases, the probability of schedule slippage decreases. We believe this variable is acting as a proxy for prototype, technical maturity, and technical complexity. Programs with unproven technologies or technically complex developmental projects in many cases utilize a prototyping effort to increase the technical maturity of the weapon system thus reducing the risk of unscheduled delays. The binary variable for prototype in our database is highly correlated to variable 72. Data points in our database that are prototype programs show a lesser percentage of procurement funding than non-prototype programs. We believe this also relates to the technical maturity and complexity of programs as they begin the EMD phase. A program with low technical risk will have a higher percentage of procurement funding and a lesser percentage of research and development funding. The effect of this variable in our model shows that a higher percentage of procurement funding leads to a lower probability of schedule growth.

Our final multiple regression model shows that programs with scheduled MS III dates before the planned IOC dates incur a greater amount of schedule slip. Although past research has not looked into this area, we believe this makes logical sense. Programs that plan to start production prior to fielding an initial operating capability must hope for a smooth development phase. There is little room for unscheduled delays, technical problems, rework, etc. Furthermore, the data shows that programs with a planned MS III date before the planned IOC date have a higher amount of total quantity. Past research and our findings show that as the total quantity of units increases, schedule growth increases as well. We believe that variable *97 MS III before IOC?* may be acting as a proxy for total quantity planned (Monaco, 2005:106-109).

Bibliography

Bielecki, John V. *Estimating Engineering and Manufacturing Development Cost Risk Using Logistic and Multiple Regression*. MS Thesis, AFIT/GAQ/ENC/03-02. Air Force Institute of Technology, Wright Patterson AFB OH (AU), March 2003 (ADA413231).

Bowsher, Charles A. "DOD Needs To Provide More Credible Weapon Systems Cost Estimates to The Congress," 1984.

Conahan, Frank C. "WEAPON ACQUISITION: Improving DOD's Weapon Systems Acquisition Reporting," G.A. Office (ed.), 1989.

DAU. "AT&L Knowledge Sharing System", *Defense Acquisition University*, http://akss.dau.mil/askaprof-akss/normal/qdetail2.asp?cgiSubjectAreaID=9&cgiQuestionID=4171, 1 October 1999.

----. "Defense Acquisition Guidebook", http://akss.dau.mil/dag/, 14 November 2005.

DoD. *The Defense Acquisition System*. DoD Directive 5000.1. Washington: USD(AT&L), 12 May 2003a.

----. *Mandatory Procedures for Major Defense Acquisition Programs (MDAPS) and Major Automated Information System (MAIS) Acquisition Programs*. DoD 5000.2-R. Washington: USD(AT&L), 10 June 2001.

----. *Operation of the Defense Acquisition System*. DoD Instruction 5000.2. Washington: USD(AT&L), 12 May 2003b.

Flaharty, Elizabeth. "Consolidated Acquisition System (CARS)", *USD(AT&L)*, http://www.acq.osd.mil/cars/index.htm, 1 October 2005.

Genest, Daniel C. *Logistic and Multiple Regression: The Two-step Approach to Estimating Cost Growth*. MS Thesis, AFIT/GCA/ENC/04-01. Air Force Institute of Technology, Wright Patterson AFB OH (AU), March 2004 (ADA423097).

Hough, Paul G. "Pitfalls in Calculating Cost Growth from Selected Acquisition Reports," RAND (ed.), RAND, 1992.

Knoche, Christine. "Telephone interview," CARS Program Manager, OSD (AT&L), Washington DC, 13 February 2006.

Levin, Robert E. "DEFENSE ACQUISITIONS: Information for Congress on Performance of Major Programs Can Be More Complete, Timely, and Accessible," G.A. Office (ed.), 2005.

Mata-Toledo, Ramon A., and Pauline K. Cushman. *Fundamentals of Relational Databases*. New York: McGraw-Hill, 2000.

Monaco, James V. *Predicting Schedule Risk: A Regression Approach*. MS Thesis, AFIT/GCA/ENC/05-01. Air Force Institute of Technology, Wright Patterson AFB OH (AU), March 2005.

Moore, Gary W. *Estimating Procurement Cost Growth Using Logistic and Multiple Regression*. MS Thesis, AFIT/GAQ/ENC/03-02. Air Force Institute of Technology, Wright Patterson AFB OH (AU), March 2003 (ADA413830).

Neter, John, Michael H. Kutner, Christopher J. Nachtsheim, and William Wasserman. *Applied Linear Statistical Models*. Boston: McGraw-Hill, 1996.

OSD. "CARS Web Site", http://www.acq.osd.mil/cars/index.htm, 1 October 2005.

----. "(CARS) Consolidated Acquisition Reporting System User Guide", *OSD(AT&L)*, http://www.acq.osd.mil/cars/download.htm, 1 December 2005.

Rosenberger, Eric. "Ostrich/English Dictionary," CACI-CMS Information Systems, Inc., 2005.

Rossetti, Matthew B. *Logistic and Multiple Regression: A Two-pronged Approach to Accurately Estimate Cost Growth in Major DoD Weapon Systems*. MS Thesis, AFIT/GCA/ENC/04-04. Air Force Institute of Technology, Wright Patterson AFB OH (AU), March 2004 (ADA422947).

Sipple, Vincent P. *Estimating Engineering Cost Risk Using Logistic and Multiple Regression*. MS Thesis, AFIT/GAQ/ENC/02-02. Air Force Institute of Technology, Wright Patterson AFB OH (AU), March 2002 (ADA400576).

Staats, Elmer B. "Comptroller General Letter to the Chairman of the Committee on Armed Services House of Representatives," C. General (ed.), 1973.

US Code: Title 10: Section 2432. 12 Aug 2005 http://assembler.law.cornell.edu/uscode/html/uscode10/usc_sec_10_00002432----000-.html

USD. "Defense Acquisition Management Information Retrieval (DAMIR)", *USD(AT&L)*, http://www.acq.osd.mil/damir/, 2005.

Walker, David M. "DEFENSE ACQUISITIONS: Assessments of Major Weapon Programs," C. General (ed.), General Accountability Office, 2003.

----. "DEFENSE ACQUISITIONS: Assessments of Major Weapon Programs," C. General (ed.), General Accountability Office, 2004.

----. "DEFENSE ACQUISITIONS: Assessments of Major Weapon Programs," C. General (ed.), General Accountability Office, 2005.

Wath, Paul F. "MAJOR ACQUISITIONS: Summary of Recurring Problems and Systemic Issues: 1960-1987," U.G.A. Office (ed.), 1986.

Womack, James P., and Daniel T. Jones. *Lean Thinking* (First Free Press 2003 edition). New York: Free Press, 2003.

Yu, Alex Dr. "Multi-Collinerity, Variance Inflation and Orthogonalization in Regression", *Arizona State University College of Education*, http://seamonkey.ed.asu.edu/~alex/computer/sas/collinear_VIF.html, 12 November 2004.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
01-03-2006	**Master's Thesis**	Jun 2005 – Mar 2006

4. TITLE AND SUBTITLE

Data Analysis and its Impact on Predicting Schedule & Cost Risk

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Cross, Steven, M., Senior Master Sergeant, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/GIR/ENC/06M-01

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Marc T. Lewis, Captain, USAF
Aircraft and Weapon Systems Cost Analyst, Air Force Cost Analysis Agency
201 12th Street Suite 403
Arlington, VA 22202 DSN: 664-0403 Comm: 703-604-0403

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Many databases rely on documents (research) of the past to input data to create a comprehensive database. The Selected Acquisition Report (SAR) is one such document. The SARs are pervasive documents that have undergone decades of scrutiny by Congress and watchdog organizations such as the Government Accountability Office. Since the SAR has undergone such massive evolutionary changes, creating an accurate acquisition database presents a daunting task for the analyst and researcher alike. This research concerns itself with one such database.

From this prior research database, we look to fill in missing data. We first conduct a literature review to determine why we have database discontinuity. Indeed, a large part of our review entails SARs. We find there are no other complete sources for acquisition reporting. Also, the reason for missing data stems from 30 years of change within the Department of Defense administration. In addition, we grow concerned about the number of rebaselines a program undergoes in its lifecycle. Investigating this, we add the variable # of Rebaselines to schedule and cost regression models from prior research for statistical evaluation. We find our new variable is highly predictive with past schedule modeling but not predictive with prior cost modeling.

15. SUBJECT TERMS
Logistic Regression, Multiple Regression, Schedule Variance, Schedule Growth, Cost Growth, History of the SAR, Selected Acquisition Report, SAR, DoD Schedule Growth, Data Analysis, Cost Growth, Inferential Statistics, Schedule Growth in DoD Acquisition Programs, Cost Growth in DoD Acquisition Programs, Predicting Schedule Growth

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					Dr. Edward D. White III
REPORT	**ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*
U	U	U	UU	128	(937) 255-3636, ext 4540; e-mail: Edward.White@afit.edu

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