10-2017

Estimating an Acquisition Program's Likelihood of Staying within Cost and Schedule Bounds

Ryan Trudelle

Edward D. White
Air Force Institute of Technology

Clay M. Koschnick

Jonathan D. Ritschel
Air Force Institute of Technology

Brandon M. Lucas

Recommended Citation

Trudelle, R., White, E., Koschnick, C., Ritschel, J., & Lucas, B. (2017). Estimating an acquisition program's likelihood of staying within cost and schedule bounds. Defense Acquisition Research Journal, 24(4), 600–625.

ESTIMATING AN ACQUISITION PROGRAM'S Likelihood of Staying Within COST and SCHEDULE BOUNDS

Capt Ryan Trudelle, USAF, Edward D. White, Lt Col Clay Koschnick, USAF, Lt Col Jonathan D. Ritschel, USAF, and Lt Col Brandon Lucas, USAF

Program managers use prior experience to spot potential programmatic areas of concern. Augmenting this experience, the authors present an empirical procedure to estimate the likelihood of a program not exceeding two schedule and cost thresholds: (a) 15 percent of the initial total acquisition cost estimate from Milestone (MS) B to Initial Operating Capability (IOC); and (b) 15 percent of the estimated length (in months) between MS B and IOC—the second bound being 25 percent of the cost and schedule estimate. Using logistic regression and odds ratios, the authors analyze 49 Department of Defense programs and generally find that electronic system programs, extremely large programs (exceeding $17.5 billion in Base Year 2017 dollars), programs procuring smaller quantities of units, and programs with shorter schedules (less time from MS A to MS B and projected time from MS B to IOC) experience smaller percentages of cost growth and schedule slippage.

DOI: *https://doi.org/10.22594/dau.17-775.24.04*
Keywords: *Cost Growth, Schedule Slippage, Schedule and Cost Thresholds, Logistic Regression Model, Odds Ratios*


KB93342655I
FEDERAL RESERVE NOTE
100
82
ONE HUNDRED DOLLARS
UNITED STATES OF AMERICA

As members of the professional acquisition workforce, we should consider any tool at our disposal to mitigate cost growth and schedule slippage in today's fiscally constrained environment. It is our duty as good stewards of the taxpayer's money to ensure that Department of Defense (DoD) programs are fielded on time and on budget. This includes being aware of program characteristics that may lead to future cost growth and schedule slippage. To investigate this, we employ a statistical technique that is often adopted in the biostatistical community—logistic regression, a technique that predicts the likelihood of an event occurring or not. Using this method, we identify possible cost and schedule variables that may indicate a program will experience significant cost growth and schedule slippage.

Specifically, we consider cost and schedule growth percentages of 15 percent or 25 percent as proxies for significant and critical overrun thresholds, where cost is defined as the total acquisition cost (development, production, military construction, but not operating and support). We selected these percentages based upon leadership's recognition that these levels are typically considered as being above and beyond an acceptable level (Schwartz & O'Connor, 2016). With this in mind, we categorize defense acquisition programs based on their cost and schedule performance at the time they meet Initial Operating Capability (IOC) versus what they estimated at Milestone (MS) B. To this end, we consider a program to be *Green* if it is within a specified percentage of its estimated cost and schedule, and *Red* if it is not. The intent of our research is to ascertain what factors may be statistically significant in predicting the probability at MS B that a DoD acquisition program will fall into either category by IOC.

Background and Database

The National Aeronautics and Space Administration (NASA) has been implementing Joint Cost and Schedule Confidence Level (JCL) analysis since 2009. JCL policy, as written in NASA Procedural Requirement 7120.5E, states that projects are required to perform a JCL with the intent that they demonstrate a 70 percent probability that cost will be equal to or less than the targeted cost, and schedule will be equal to or less than the targeted schedule date. With respect to DoD acquisition programs and joint assessment of cost growth and schedule slippage, such literature is scant with studies, though many analyze program cost and schedule performance individually (Arena, Leonard, Murray, & Younossi, 2006; Cancian, 2010; Monaco & White, 2005). In contrast, NASA has circulated numerous reports regarding the joint risk of cost growth and schedule slippage.

> ***A program manager might only see the tip of the iceberg in terms of the final cost of a program over its entire lifetime at MS B; however, decisions made by MS B actually determine over 70 percent of a system's final total life-cycle costs.***

In particular, one study by Burgess and Krause (2014) examined the interaction between the phasing estimating relationship (PER), the cost estimating relationship (CER), and the schedule estimating relationship (SER). The CER is the total program cost from the System Requirements Review (SRR), a date that occurs prior to MS B, through launch; while the SER is the time, in months, from SRR to launch. Given these cost and schedule estimates, the PER relays the annual funding profile for the program and serves as the starting point for analyzing cost and schedule ramifications. Burgess and Krause used historical data from 37 NASA programs for their study and developed multiple regression models to analyze these relationships.

From their analysis, they developed a set of tools to give decision makers the ability to quantify trade-offs between cost, schedule, and phasing in their program. In terms of crosswalking to the DoD, one can look at this comparison in terms of Earned Value Management (EVM). Within EVM, Schedule Performance Index (SPI) and Cost Performance Index (CPI) allow one to assess the health of a project. Specifically, SPI and CPI help a program manager analyze the efficiency of schedule performance and cost performance of any project. Additionally, these metrics assist in planning trade-offs that pinpoint where to leverage funds to best right a program that may have gotten off track, while simultaneously keeping cost growth and schedule slippage coupled.

The Burgess and Krause (2014) analysis allowed the program manager to conduct a programmatic "health assessment" in which the estimating relationships are analyzed to determine if they fall within a standard deviation of the mean observed historical value. Our research deviates from theirs in that we analyze categorical indicators for programs that fall under or over the 15 percent and 25 percent baseline thresholds. We aim to describe what indicators may correlate to a program being designated Green or Red in the future based upon characteristics at MS B.

We glean two important facts from the findings of Burgess and Krause (2014): (a) longer duration from SRR to Preliminary Design Review (PDR) suggests increased likelihood of program schedule lengthening; and (b) a higher percentage of new designs appears to increase the likelihood of increased cost in acquisition programs. Jimenez, White, Brown, Ritschel, Lucas, and Seibel (2016) concluded similar findings. Longer time between MS A and MS B correlated to programs having generally longer schedule duration. Their research also deemed the following variables statistically significant for predicting increased schedule duration: (a) whether a program is a new effort or modification to an existing program, (b) the amount of raw funding (adjusted for inflation) prior to MS B for a program, and (c) the percentage of Research, Development, Test and Evaluation (RDT&E) total funding profile allocated at MS B.

These were not the only studies that suggested the usefulness of information obtained prior to MS B. A study conducted by Deitz, Eveleigh, Holzer, and Sarkani (2013) examined the importance of developing a robust Analysis of Alternatives prior to MS B and the effects it may have on program success. The most important finding of their research is that while only 10 percent of a program's total life-cycle costs are realized prior to MS B, 70 percent of a program's total life-cycle costs are determined by decisions made by MS B (Deitz et al., 2013). That is, a program manager might only see the tip of the iceberg in terms of the final cost of a program over its entire lifetime at MS B; however, decisions made by MS B actually determine over 70 percent of a system's final total life-cycle costs. The cautionary note here is that short-term savings might result in long-term costs eventually.

Similar to Jimenez et al. (2016), this suggests pre-MS B data are very important to predicting program outcomes. Unfortunately, not many programs have pre-MS B data since a DoD acquisition program does not officially begin until MS B; however, these recent studies seem to suggest that problems relating to cost growth and schedule slippage might be detected early. To investigate this possibility, we focused on this earlier information because as other references have discussed, it's a window into technology maturity. As the Government Accountability Office (GAO) has repeatedly pointed out, it has consistently been found that the vast majority of programs began system development without mature technologies and moved into system demonstration without design stability (GAO, 2009).

We used information collected in the Selected Acquisition Report (SAR) data, as retrieved by the Defense Acquisition Management Information Retrieval (DAMIR) system, to investigate this pre-MS B information. The programs selected were unclassified and designated at MS B as either a Major Defense Acquisition Program (MDAP) or not. [Note: As defined in DoDI 5000.02 (2015), an MDAP is a program estimated to have research and development costs greater than $480 million or procurement costs greater than $2.79 billion (in FY2014 constant dollars).] The program SAR must contain an MS A date or funding in the funding profile at least 1 year prior to MS B (indicates the year in which MS A may have occurred). Correspondingly, the program SAR must contain an MS B date and an IOC date that occurred prior to the last reported SAR. Without the MS B date and funding information, we are unable to ascertain the duration of MS A, the funding spent up to MS B, the projected funding needed to reach IOC, or the projected duration of MS B to IOC. With an actual reported IOC date, this indicates that the program is complete up to IOC and ensures we are not using projected values as actual costs in our model.

The requirement that the program must contain a SAR within 1 year of reaching MS B allows us to ascertain what the program's cost and schedule estimate was at MS B, and if the actual cost and schedule from MS B to IOC are within 15 percent or 25 percent of this estimate. We are assuming that both cost and schedule estimates are realistic and attainable. We allow 1 year from the time MS B occurs because the program may not have been required to report a SAR at the time MS B occurred. We agree that this might seem like an unnecessary limitation; however, we wish to obtain the 'truest' estimates set at the beginning of a program's initiation and militate against numbers still changing. That is, we don't want to treat actual values as estimates, when in fact the estimates are actual costs. This might inadvertently have programs underreporting cost growth and schedule slippage. Table 1 summarizes the inclusion and exclusion criteria. Based on these criteria, we use 49 programs in our analysis—the specific programs are listed in the Appendix.

TABLE 1. PROGRAM INCLUSION TABLE

Inclusion/Exclusion Criteria	Programs Included	Programs Removed	Program Count
Jimenez et al. (2016) Database	56		56
DAMIR Query (MDAP/ Pre-MDAP)	187		243
Double counted from Jimenez et al. (2016) Database		29	214
IOC Occurs after Last SAR		61	153
Missing Milestones A or B		74	79
No SAR within 1 year of MS B		24	55
Missing IOC		4	51
Classified		2	49
Final Number			49

It is important to stress that we are not using these selected 49 programs to infer to the entire collection of all SAR programs—simply because most, as documented, do not contain any pre-MS B data. The purpose of this study is not only to explore whether such data appear to be predictive of cost growth and schedule slippage, but also to purposely examine the information with a different statistical mindset. If such analysis does confirm that pre MS-B data appear to be again statistically significant, then perhaps more emphasis should be made to retain all programmatic data prior to MS B for all DoD programs.

For each of the 49 DoD acquisition programs in our database, we use two SARs. For the response variables, we use the last reported SAR for each program to gather the actual cost (development and procurement, not operating and support costs, and beyond) and schedule duration for each program from

MS B to IOC. [Note: Future research could investigate how many programs are put into operations without operating and support funds identified and sufficiently appropriated.] For the candidate explanatory variables, we use the SAR from the year in which MS B occurred or, if this is unavailable, the SAR within 1 year of MS B. The cost and schedule estimate from MS B to IOC in this SAR becomes the current estimate with respect to measuring cost growth and schedule slippage. The cost growth percentage is calculated as *Current Cost Estimate at MS B – True Cost from MS B to IOC / Current Cost Estimate at MS B.* A similar calculation is computed for schedule.

Methodology

As noted earlier, the purpose of this article is to identify predictor variables that may determine the likelihood (probability) that a DoD acquisition program will experience cost growth and schedule slippage above certain thresholds for those programs with pre-MS B data. Similar to NASA's approach, we are looking jointly at cost growth and schedule slippage. We recognize that programs change, possibly due to forces outside of the program manager's control. Thus, we employ two separate threshold values. The first is the 15 percent threshold above the current estimate (both cost and schedule) from MS B to IOC established at MS B. The second threshold is set at 25 percent.

All costs in our models are in Base Year 2017 (BY17) dollars using the 2016 Office of the Secretary of Defense inflation indices, which prevents inflation from influencing our model. For the 15 percent and 25 percent response categories, we initially assign each of our 49 programs in the database to one of four mutually exclusive categories: *Green/Green, Green/Red, Red/Green*, and *Red/Red*. A program is considered Green if the final cost growth (or schedule slippage) from MS B to IOC is less than the chosen overrun threshold; a program is considered Red if it equals or exceeds the overrun threshold. For example, a Green/Red rating with respect to a 15 percent threshold indicates the program was within cost, but over schedule.

Initially, we aimed to identify variables that may predict which of the four categories a DoD program might fall into at MS B; however, the limited sample sizes for Green/Red and Red/Green prevented this. Combining these groups only resulted in nine programs with a significant overrun and six programs with a critical overrun. Although this combined category lacked the requisite statistical power to conduct any logistic regression analysis, it is also reflected that a program tended to stay either Green for both cost and schedule or Red for both. Therefore, we only focus on the Green/Green and Red/Red categories for both the 15 percent and 25 percent thresholds. These designations are listed in the Appendix for each program in our database.

To build our initial logistic regression model, we use a mixed stepwise approach to identify the most predictive variables; a 0.1 level of significance was selected for the entry and exit criteria due to the exploratory nature of our work. For the finalized model, the resultant predictor variables from the stepwise procedure must meet the overall model Type I error of 0.1 and require each variable to be significant according to the Holm-Bonferroni criteria (Holm, 1979). [Note: A Type I error is the probability of determining an explanatory variable is predictive when in fact it isn't.] We use JMP® Pro 12 for all statistical analysis performed in this article.

A logistic regression model predicts the probability of a program identifying with a particular group by way of the following equation:

$$y = \frac{e^{f(x)}}{e^{f(x)} + 1} \tag{1}$$

where y is a binary variable indicating a program's group, e is the natural exponent function, and $f(x)$ is considered the logit or log-odds function (Gaudard, Ramsey, & Stephens, 2006) and can be written in the form:

$$f(x) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p \tag{2}$$

Equation (1) represents an *s*-shaped curve (White, Sipple, & Greiner, 2004) whose values range from 0 to 1 (probability).

The X variables in (2) typify the standard explanatory variables used in linear regression; however, the β coefficients do not represent the mean change in the response. Instead, e^{β_i} represents the odds ratio (OR) of a particular program in our database belonging to either Green/Green or Red/Red when the X variables are dichotomous (i.e., $X_i = 1$ when a characteristic is present or $X_i = 0$ when a characteristic is not present). Continuous explanatory variables do not possess this easy interpretation of ORs because there is no natural baseline group to compare. Therefore, all explanatory variables have been converted to this dichotomous setting. [Note: For completeness, we did investigate the continuous settings of these variables, but none of the findings appeared to contradict what we state in this article.]

For categorical variables, this transformation is straightforward. For example, a dummy variable might be coded a 1 if the program is an Air Force acquisition program, and 0 if otherwise (i.e., an Army, Navy, or Marine program). For continuous data, we discretize (i.e., create categorical groupings) by utilizing histograms to determine potential break points in the data. These break points often coincide with quartiles (25th percentage, 50th percentage, or 75th percentage) of the histograms, which is a common practice in the biomedical community.

We use two metrics to quantify the predictive capability of our logistic regression models. The first metric is the Area Under the Receiver Operating Characteristic Curve (AUC). The AUC indicates the sorting efficiency of a model with a value of 0.5 indicating merely random chance and a value of 1.0 indicating perfect prediction capabilities (Gaudard, Ramsey, & Stephens, 2006). The AUC is a single measure of the overall discrimination ability of a test. In general, "an AUC that is greater than 0.8 suggests that the diagnostics test has good discriminatory power" (McPherson & Pincus, 2016, p. 80). Since we have such a small subset of data for each group, it is infeasible to set aside a 20 percent validation pool. Given this limitation, we use a technique called bootstrapping (Efron & Tibshirani, 1994) to present a 90 percent confidence interval for the AUC value for each logistic regression model; these intervals provide the user predictive limitations of the model.

The second metric to demonstrate the utility of our logistic regression models is the OR for each explanatory variable and its corresponding confidence bound (either the lower or upper value in the confidence interval that is closest to the value of 1). An OR equal to 1 indicates the explanatory variable does not affect the odds of a program belonging to either the Green/Green or Red/Red category. An OR > 1 implies a higher odds of a program belonging to the Green/Green category, while an OR < 1 suggests lower odds of belonging to the Green/Green category (Szumilas, 2010). With respect to the confidence interval of an odds ratio, either the lower or upper confidence bound is used to estimate the precision of the OR. In practice, this bound is often used as a proxy for the presence of statistical significance if it does not overlap the null value (e.g., OR = 1) (Szumilas, 2010).

Lastly, to prevent model extrapolation, the ranges of the continuous independent variables over which the models are useful must be consistent with the bounds of the programs used in our analysis. Using the models outside these ranges may invalidate the results. Only three continuous explanatory variables proved statistically significant in our models. For projected duration from MS B to IOC, the range is 30 to 109 months. For projected percent complete at MS B, the range is 15 percent to 70 percent. For the duration from MS A to MS B, the range is 13 to 125 months.

Results

The following subsections illustrate the logistic models derived from the stepwise procedure along with an explanation of each significant explanatory variable. The first subsection highlights the results regarding the Green/Green and Red/Red groups for the 15 percent overrun threshold (*Significant*), while the second subsection highlights the results for the Green/Green and Red/Red groupings for the 25 percent overrun threshold (*Critical*).

Significant Overrun

For this analysis, 15 programs (approximately 31 percent of our database) fall in the Green/Green group and 25 (approximately 51 percent of our database) programs are in the Red/Red group. Table 2 summarizes the logistic model and associated predictive explanatory variables for determining the likelihood of a DoD acquisition program experiencing less than 15 percent cost and schedule growth from MS B to IOC. The model has an AUC of 0.88 suggesting good model discrimination. All of the estimated ORs and their associated confidence bounds are well above or below 1. Overall, these metrics suggest reasonable confidence in our findings.

TABLE 2. SIGNIFICANT PREDICTOR VARIABLES FOR DETERMINING LIKELIHOOD OF PROGRAM EXPERIENCING COST GROWTH/SCHEDULE SLIPPAGE LESS THAN 15 PERCENT

Variable	Estimate	Odds Ratio	Odds Ratio Bound	Chi-Square	P-Value	Relative Percent Effect
Intercept	-3.32	N/A	N/A	3.20	0.0735	N/A
Projected MS B to IOC <= 58 months	3.63	37.83	5.55	7.37	0.0066	25.9
Program Cost > $17.5B	3.37	29.13	4.58	7.01	0.0081	23.0
Electronic System Program	3.27	26.37	3.64	6.09	0.0136	21.3
Projected % Complete at MS B <= 35%	3.32	27.76	3.98	5.99	0.0144	18.2
MDAP	-3.34	0.036	0.49	3.75	0.0529	11.6

Note. Numbers rounded to two significant digits. AUC = 0.88 with a 90% bootstrapped confidence interval (1,000 samples) of (0.84, 0.98). The family-wise error rate for the independent variables is 0.10.

The electronic system program variable indicates if the DoD acquisition program is an electronic user interface system, avionics control system, radio network system, or similar electronic system. The OR suggests that such systems typically display cost growth and schedule slippage less than 15 percent. This appears to be in keeping with Bolten, Leonard, Arena, Younossi, and Sollinger (2008), who also concluded that electronic systems appear to be historically cheaper and thus less susceptible to cost growth and schedule slippage.

For the projected MS B to IOC duration <= 58 months explanatory variable, this finding suggests that acquisition programs whose projected MS B to IOC duration is equal to or less than 58 months (or approximately 5 years) typically display cost growth and schedule slippage less than 15 percent. We theorize this may be indicative of relatively shorter scoped programs whose technology may be relatively more mature. [Note: The 58-month timeframe was flagged from the numbers in our database. For practical purposes, we would suggest using a 5-year cut-off as the boundary point.]

The extra large program explanatory variable suggests that acquisition programs with a high cost (greater than $17.5 billion BY17 dollars in total project acquisition cost) typically experience cost growth and schedule slippage less than 15 percent. This is logically expected since larger programs do not have the flexibility of having sizeable overruns given the sheer amount of dollars involved before DoD oversight and/or Congressional reviews intervene and possibly cancel the program. Thus, we treat the extra large program explanatory variable as more of a covariate than a traditional explanatory variable.

The programs identified as MDAP in our database tend to suggest that this explanatory variable will lead to cost growth and schedule slippage greater than or equal to 15 percent. We believe this might have occurred for two reasons. One, this could be just an artifact of our database due to the large number of programs that identify as MDAP (45 of 49, or 92 percent) and the fact that all the programs in the Red/Red group are identified by this variable. It is also noteworthy that three of the four programs not identified as MDAP are in the Green/Green group. The second possible reason for our finding: MDAPs simply by their nature may be set up to fail. We theorize this isn't done on purpose per se, but we have seen too many instances whereby schedule timelines are simply too unrealistic. Additionally, in the rush to get the big ticket items initiated, cost estimates are lower than what they should be in order to "get the foot in the door."

Finally, the projected percent complete at MS B <= 35% variable (calculated as the actual time from MS A to MS B divided by the sum of the actual time from MS A to MS B and projected time from MS B to IOC) is statistically significant; this result suggests that programs that spend less time in the MS A to MS B phase relative to, and in comparison to, the MS A to IOC phase experience less cost growth and schedule slippage. This may be due to a high technology readiness level (TRL) early in the program's life or a lesser extent of new technology involved in the program. Such a conclusion is consistent with Dietz et al. (2013) who studied the pre-MS B process to identify cost-estimating relationships associated with identified TRLs. Their findings indicate that programs with a higher TRL entering MS B experience smaller levels of cost growth.

Regarding the Red/Red group, Table 3 displays the logistic model and associated predictor variables for determining the likelihood of a DoD acquisition program's actual MS B to IOC cost and schedule exceeding its MS B estimate by 15 percent or more. The model has an AUC of 0.85, suggesting good model discrimination. All of the estimated ORs and their confidence bounds are well above or below 1. Like before, these metrics suggest reasonable confidence in our findings.

TABLE 3. SIGNIFICANT PREDICTOR VARIABLES FOR DETERMINING LIKELIHOOD PROGRAM WILL EXPERIENCE COST GROWTH/SCHEDULE SLIPPAGE EQUAL TO OR EXCEEDING 15 PERCENT

Variable	Estimate	Odds Ratio	Odds Ratio Bound	Chi-Square	P-Value	Relative Percent Effect
Intercept	1.41	N/A	N/A	2.65	0.1038	N/A
Program Cost > $17.5B	-4.60	0.01	0.09	8.89	0.0029	33.1
Electronic System Program	-3.15	0.04	0.26	6.74	0.0094	22.1
Aircraft	3.29	26.86	3.19	5.00	0.0254	18.1
RDT&E at MS B Start >= $272M	1.87	6.48	1.64	4.47	0.0346	13.5
Qty Expected at MS B <= 305	-1.95	0.14	0.62	3.98	0.0461	13.2

Note. Numbers rounded to two significant digits. AUC = 0.85 with a 90% bootstrapped confidence interval (1,000 samples) of (0.79, 0.95). The family-wise error rate for the independent variables is 0.10.

Similar to the Green/Green model, both the explanatory variables of extra large programs and electronic system programs are statistically significant. However, both variables have negative parameter estimates (and thus ORs much smaller than 1), which indicates programs displaying these characteristics are much less likely to experience cost growth and schedule slippage equaling or exceeding 15 percent. This is consistent with our findings from the Green/Green group in Table 2.

> ***It is important to stress that we are not using these selected 49 programs to infer to the entire collection of all SAR programs—simply because most, as documented, do not contain any pre-MS B data.***

The explanatory variable identifying a program as a fixed wing aircraft is statistically significant in predicting whether a program is more likely to experience cost growth and schedule slippage equaling or exceeding 15 percent. We believe this is due to the large and complex nature of these programs, especially given the modern aircraft programs in our study, such as the F-22 and F-35. Or, as we discussed earlier, this could be a sign of rushing an aircraft into the inventory without realistically scrutinizing cost and schedule estimates at MS B.

Programs in our database that are expected to procure less than or equal to 305 units at MS B tend to indicate that they are less likely to experience cost and schedule growth equaling or exceeding 15 percent. In all honesty, we do not have a solid reason for why this is the case. We neither detected any dependencies between variables that would cause this result nor did we detect any issues in the analysis. We had suspected, given the other variables in play for just this model, that the F-22 and F-35 may be overly influencing the results. When we temporarily omitted these programs, the results, as presented in Table 3, didn't materially change. The p-value for aircraft increased to 0.0655; the others remained relatively constant. So, we present this finding as is and invite others to explore it further in the future.

The last predictor variable associated with programs experiencing cost growth and schedule slippage equaling or exceeding 15 percent is for programs that spend greater than $272 million in RDT&E by the start of MS B. We believe this may be indicative of programs with a low level of technological maturity, thus requiring larger and more complex development prior to MS B. It could also indicate that a program is integrating many highly sophisticated components and the final design is complex in nature. As mentioned for the Green/Green model, this is consistent with Dietz et al. (2013), who researched the pre-MS B process and found that a lack of maturity at MS B correlates with higher costs. [Note: As mentioned earlier, this $272 million amount was flagged from the numbers in our database. For practical purposes, we would suggest using $275 million or even $300 million; the point is that around this dollar amount, the probabilities statistically change.]

Critical Overrun

For this analysis, 20 programs (approximately 41 percent of our database) fall in the Green/Green group and 23 (approximately 47 percent of our database) programs are in the Red/Red group. Table 4 shows the logistic model and associated predictor variables for determining the likelihood that a DoD acquisition program's true MS B to IOC cost and schedule will be less than 25 percent larger than its MS B estimate. The model has an AUC of 0.84 suggesting good model discrimination. All of the estimated ORs and associated 90 percent confidence bounds are well above or below 1.

TABLE 4. SIGNIFICANT PREDICTOR VARIABLES FOR DETERMINING LIKELIHOOD A PROGRAM WILL EXPERIENCE COST GROWTH/SCHEDULE SLIPPAGE LESS THAN 25 PERCENT

Variable	Estimate	Odds Ratio	Odds Ratio Bound	Chi-Square	P-Value	Relative Percent Effect
Intercept	5.41	N/A	N/A	7.19	0.0073	N/A
Program Cost > $17.5B	3.34	28.25	5.01	7.58	0.0059	32.1
MDAP	-4.54	0.011	0.13	7.19	0.0073	27.8
MS A to MS B >= 28 months	-2.99	0.05	0.27	6.33	0.0119	26.0
1985 or Later for MS B Start	-1.69	0.19	0.69	4.08	0.0434	14.1

Note. Numbers rounded to two significant digits. AUC = 0.84 with a 90% bootstrapped confidence interval (1,000 samples) of (0.78, 0.93). The family-wise error rate for the independent variables is 0.10.

With respect to previous results regarding MDAP and extra large programs, we see similar results in this section. Extra large programs appear more likely to have cost growth and schedule slippage less than 25 percent, while MDAPs are less likely to have cost growth and schedule slippage under 25 percent.

For the MS A to MS B greater than or equal to 28 months' explanatory variable (or approximately 2.5 years), these programs appear less likely to experience cost growth and schedule slippage less than 25 percent. A possible explanation is that programs with relatively longer duration from MS A to MS B may indicate a program is relying upon complex technology that must be matured, which we believe is consistent with prior research conducted by Dietz et al. (2013).

The variable 1985 or later for MS B start indicates if a program is considered to be a part of the "modern" era of defense acquisition. This finding suggests modern programs are more likely to experience cost growth and schedule slippage over 25 percent. This could be due to the increasing complexity of modern programs, which include the Joint Strike Fighter (JSF) and other highly complex systems, and that increased complexity drives cost and schedule. This is consistent with the work conducted by Jimenez et al. (2016), who found that these modern programs tend to have a longer schedule.

Regarding the Red/Red group, Table 5 displays the logistic model and associated explanatory variables for determining the likelihood of a DoD acquisition program having its true cost and schedule from MS B to IOC exceeding its MS B estimate by 25 percent or more. The model has an AUC of 0.79, suggesting fair to good model discrimination. All of the estimated ORs and associated confidence bounds are well above or below 1.

TABLE 5. SIGNIFICANT PREDICTOR VARIABLES FOR DETERMINING LIKELIHOOD A PROGRAM WILL EXPERIENCE COST GROWTH/SCHEDULE SLIPPAGE EQUAL TO OR EXCEEDING 25 PERCENT

Variable	Estimate	Odds Ratio	Odds Ratio Bound	Chi-Square	P-Value	Relative Percent Effect
Intercept	0.54	N/A	N/A	1.47	0.2253	N/A
Electronic System Program	-2.74	0.06	0.33	5.72	0.0168	40.4
Program Cost > $17.5B	-2.51	0.08	0.40	4.98	0.0257	35.9
Aircraft	2.10	8.19	1.52	3.20	0.0737	23.7

Note. Numbers rounded to two significant digits. AUC = 0.79 with a 90% bootstrapped confidence interval (1000 samples) of (0.70, 0.89). The family-wise error rate for the independent variables is 0.10.

With respect to the explanatory variables of extra large programs, fixed wing aircraft, and electronic system programs, we see the same trends as we did in Tables 2–4; extra large programs and electronic system programs are less likely to experience cost and schedule growth greater than 25 percent, while fixed wing aircraft are more likely to experience cost growth and schedule slippage greater than or equal to 25 percent. There are no additional significant variables for this model.

With respect to any other findings, no other explanatory variables that we investigated appeared to flag as statistically significant for predicting the probability of a program being either Green/Green or Red/Red. Such variables included other weapon platforms (besides those mentioned in this article), Service (Air Force, Army, etc.), modification program, prototype developed, concurrency planned, or contractor. Additionally, we investigated the possibility of explanatory variables being "correlated." In the logistic regression analysis presented in this article, that translates to variables being dependent on one another. We detected no such issues with respect to the presented results.

Discussion and Conclusions

In this article, we further delve into the association of pre-MS B data by predicting whether a DoD acquisition program will incur cost growth and schedule slippage. Metrics are analyzed jointly versus separating them. We first investigate possible explanatory variables that statistically predict the probability of a DoD acquisition program experiencing cost growth and schedule slippage less than 15 percent. We also model the likelihood that a program would experience cost and schedule growth in excess of (or equal to) 15 percent. These percentage increases are measured with respect to the MS B to IOC estimates at MS B, and the actual cost and schedule realized for MS B to IOC. We then replicate this process to determine which variables may be predictive if the threshold percentage increased from 15 percent to 25 percent.

Overall, we determined the following five variables appear to be predictive factors for determining if a DoD acquisition program will experience less cost and schedule growth:

1. Electronic system programs;
2. Programs having a projected MS B to IOC duration less than (or equal to) 58 months;
3. Extra large programs (exceeding $17.5 billion BY17);
4. Programs that expect to procure fewer than 305 units at the time of MS B; and
5. Programs with a projected percent complete at MS B less than (or equal to) 35 percent.

In contrast, MDAPs, fixed wing aircraft, programs where the duration between MS A to MS B is greater than (or equal to) 28 months, programs whose projected percent complete at MS B is greater than 38 percent appear, modern programs that enter MS B in 1985 or later, and programs that spend greater than (or equal to) $272 million (BY17) of RDT&E funding by the beginning of MS B appear to be predictive that programs are likely to experience more cost growth and schedule slippage. Table 6 captures those factor effects more generally controllable by a program manager.

TABLE 6. SIGNIFICANT PREDICTOR VARIABLES USUALLY CONTROLLABLE BY A PROGRAM MANAGER

Variable	Overall Program Effect	Relative Ranking
Projected MS B to IOC <= 58 months	Positive	1
Projected % Complete at MS B <= 35%	Positive	3
Qty Expected at MS B <= 305	Positive	5
MS A to MS B >= 28 months	Negative	2
RDT&E at MS B Start >= $272M (BY17)	Negative	4

Note. The Overall Program Effect column reflects the Variables' overall general effect on a program, while the Relative Ranking column compares only among the factors shown.

Our findings, with respect to variables that incorporate the time between MS A and MS B, are consistent with those of Dietz et al. (2013). These results suggest that programs with more technology uncertainty or immaturity at MS B have an increased likelihood of incurring higher cost growth and schedule slippage compared to more technologically mature programs. Additionally, our findings with respect to electronic systems programs are supported by Bolten et al. (2008), though we do acknowledge that most of those programs in our database were both small in nature (under $3 billion BY17) and consisted primarily of modifications.

As with any statistical model, there are limitations to our logistic regression models. First, the database was created from SARs that may contain incomplete information. The models built are only as good as the data used to create them. There were multiple constraints on the data collecting process that hampered the ability to create a more robust database; foremost, the lack of availability of pre-MS B data limited the programs that could be included. Additionally, the search parameters in DAMIR may have unintentionally excluded programs, which could have influenced the outcome of our analysis.

To gain insight on a program's potential for cost and schedule growth at such an early stage as MS B, we attempt to leverage the knowledge of the past to see where others have been. Our models may give program managers a glance at where they may be heading and highlight potential pitfalls. This set of logistic regression models is designed to provide a tool for the DoD acquisition community to make strategic program health assessments. Practically, these models offer the potential to help portfolio managers decide where to allocate risk dollars.

Our research differs from prior research in that our database is expanded beyond the NASA-only programs that were researched by Burgess and Krause (2014). Additionally, we utilize program characteristics across a large range of programs to develop logistic models that predict the probability of overrunning thresholds identified as being above acceptable levels of cost and schedule growth. No other DoD research, to our knowledge, relates cost or schedule growth probability to overrunning such important thresholds. The models may provide managers the ability to predict the possibility and severity of an overrun.

References

Arena, M. V., Leonard, R. S., Murray, S. E., & Younossi, O. (2006). *Historical cost growth of completed weapon system programs* (Report No. TR-343). Santa Monica, CA: RAND.

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

Burgess, E., & Krause, C. (2014, August). *Integrated estimating relationships.* Paper presented at NASA 2014 Cost Symposium, Hampton, VA. Retrieved from https://www.nasa.gov/sites/default/files/files/09_Burgess_NASA_Cost_Symposium_2014_IERs_Final_Tagged.pdf

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

Deitz, D., Eveleigh, T. J., Holzer, T. H., & Sarkani, S. (2013). Improving program success through systems engineering tools in the pre-milestone B acquisition phase. *Defense Acquisition Research Journal, 20*(3), 283–308.

Department of Defense. (2015). *Operation of the defense acquisition system* (DoDI 5000.02). Washington, DC: Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap.* Boca Raton, FL: Chapman and Hall/CRC Press.

Gaudard, M., Ramsey, P., & Stephens, M. (2006). *Interactive data mining and design of experiments: The JMP® partition and custom design platforms.* Cary, NC: North Haven Group. Retrieved from http://islab.soe.uoguelph.ca/sareibi/PROJECTS_dr/GRAD_FUTURE_dr/docs/Interactive_DataMining.pdf

Government Accountability Office. (2009). *Defense acquisitions: DoD must balance its needs with available resources and follow an incremental approach to acquiring weapon systems* (GAO Report No. 09-431T). Washington, DC: U.S. Government Printing Office.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70. Retrieved from https://www.ime.usp.br/~abe/lista/pdf4R8xPVzCnX.pdf

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

McPherson, R. A., & Pincus, M. R. (2016). *Henry's clinical diagnosis and management by laboratory methods* (23d ed.). Amsterdam, Netherlands: Elsevier.

Monaco, J., & White, E. (2005, April–July). Investigating schedule slippage. *Defense Acquisition Review Journal, 12*(3), 177–190.

Schwartz, M., & O'Connor, C. V. (2016). *The Nunn-McCurdy Act: Background, analysis, and issues for Congress* (Report No. R41293). Washington, DC: Congressional Research Service. Retrieved from https://fas.org/sgp/crs/natsec/R41293.pdf

Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry, 19*(3), 227–229.

White, E. D., Sipple, V. P., & Greiner, M. A. (2004). Using logistic and multiple regression to estimate engineering cost risk. *Journal of Cost Analysis and Management, 6*(1), 67–79. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/15411656.2004.10462248

Appendix

List of Programs and Their Respective Designations

Program	15% Group	25% Group	Program	15% Group	25% Group
A-10 Thunderbolt II	3	3	COBRA Judy Replacement	3	2
C-17 Globemaster III	3	3	Harpoon Missile	3	3
F-22 Raptor	3	3	Navy Multiband Terminal (NMT)	2	2
AH-64 Apache	3	3	SH-60B Seahawk	3	3
B-1B Computer Upgrade	1	1	UGM-96A Trident I Missile	2	1
C-5 Reliability & Re-Engining Program (RERP)	3	3	SSN 774 (Virginia Class Submarine)	2	1
F-15 Eagle	1	1	UGM-109 Tomahawk	1	1
B-1B Joint Direct Attack Munition (JDAM)	1	1	SSBN 726 Submarine	3	3
FA-18 A/B Hornet	2	1	AGM-114A Hellfire Missile	3	3
AV-8B Harrier	1	1	OH-58D Helicopter	1	1
P-8 Poseidon	1	1	AAWS-M Javelin	3	2
V-22 Osprey	3	3	B-2 EHF Inc 1 Satellite Communications	1	1
F-35 Joint Strike Fighter (JSF)	3	3	AH-64E Apache Longbow Remanufacture	2	2
CH-47D Chinook	2	1	CH-47F Chinook	3	3
E-8A Joint Surveillance Target Attack Radar System (JSTARS)	3	3	UH-60M Blackhawk	3	3

List of Programs and Their Respective Designations Continued

Program	15% Group	25% Group	Program	15% Group	25% Group
Air-Launched Cruise Missile (ALCM)	3	3	Active Electronically Scanned Array (AESA)	1	1
Advanced Medium-Range Air-to-Air Missile (AMRAAM)	1	1	AGM-88E Advanced Anti-Radiation Guided Missile (AARGM)	3	3
Joint Air-to-Surface Standoff Missile (JASSM)	3	3	Cooperative Engagement Capability (CEC)	3	3
Joint Direct Attack Munition (JDAM)	1	1	E-2D Advanced Hawkeye (AHE)	3	3
Joint Primary Aircraft Training System (JPATS) T-6A	3	3	Littoral Combat Ship (LCS)	1	1
GBU-39 Small Diameter Bomb-I	1	1	MH-60S Knighthawk	3	3
National Aerospace System	2	2	Advanced Extremely High Frequency (AEHF)	3	3
AGM-88 High-speed Anti-Radiation Missile (HARM)	2	2	Evolved Expendable Launch Vehicle (EELV)	1	1
AIM-9X Block 1 Sidewinder Missile	2	1	Wideband Global SATCOM (WGS)	3	3
AN/BSY-1 Combat System	1	1			

Code 1 implies cost growth and schedule slippage less than 15% (or 25%).

Code 2 implies either cost growth or schedule slippage less than 15% (or 25%), but not both.

Code 3 implies cost growth and schedule slippage equal to or greater than 15% (25%).

Author Biographies

Capt Ryan Trudelle, USAF, is a cost analyst at the Space and Missile Systems Center at Los Angeles Air Force Base, California. Capt Trudelle holds a BS from the United States Air Force Academy and an MS in Cost Analysis from the Air Force Institute of Technology (AFIT), Wright-Patterson AFB, Ohio.

(E-mail address: Ryan.Trudelle@us.af.mil)

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

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

Lt Col Clay Koschnick, USAF, is an assistant professor of systems engineering in the Department of Systems Engineering and Management, AFIT. He received his MS and PhD in Operations Research from the Georgia Institute of Technology and the University of Florida, respectively. Lt Col Clay has served in acquisition management roles for a variety of space, aircraft, and information technology programs. His research interests include economic and decision analysis.

(E-mail address: Clay.Koschnick@afit.edu)

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

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

Lt Col Brandon Lucas, USAF, currently serves as an assistant professor and director of the graduate cost analysis program, AFIT. He holds an MS in Cost Analysis from AFIT and a PhD in Economics from George Mason University. Lt Col Lucas has served in the budget, cost, and finance communities at base, center, and Air Staff levels.

(E-mail address: Brandon.Lucas@afit.edu)