Schedule Estimating Relationship (SER) Development Using Missile and Radar Datasets

Sara Jardine, Maxwell Moseley, Justin Moul, and Donald Trapp

Naval Center for Cost Analysis

Abstract

The Naval Center for Cost Analysis (NCCA) led research to create datasets of schedule events and technical parameters for missile and radar systems. Collecting historical schedule data for missiles and radars presents several challenges, such as inconsistent or evolving definitions of acquisition milestones. This paper discusses the team's solutions to data collection challenges, as well as efforts to derive parametric schedule estimating relationships (SERs) from technical characteristics. This NCCA endeavor produced curated datasets for missiles and radars, to which all service cost analysis organizations contributed – forming a consistent basis upon which analysts can draw analogies between realized program schedules and plans depicted for future programs.

1 Introduction

According to the GAO, 86 major defense programs in the U.S. Department of Defense's (DoD) 2017 portfolio averaged a 38% delay in delivering operational capability.[1] A number of cost analysis organizations, including NCCA, have embraced a goal to understand the implications of schedule dependencies (and/or optimism) to arrive at more realistic program cost estimates. To pursue this goal, NCCA and the other services' cost agencies have embarked on a path to evaluate projected program schedules in light of data-driven methods based upon realized prior program histories. If successful, this cost analysis goal will avoid, or at least mitigate, the optimism bias noted by the GAO, which has historically contributed to cost growth and key capability delays.[2]

[1] U.S. Government Accountability Office. (2018, April). *Weapon systems annual assessment: Knowledge gaps pose risks to sustaining recent positive trends*. Washington, D.C.: GAO-18-360SP.

[2] U.S. Government Accountability Office. (1994, July). *Future years defense program: Optimistic estimates lead to billions in overprogramming*. Washington, D.C.: GAO/NSIAD-94-210.

NCCA, in initiating this research, hypothesized that challenges associated with developing and producing systems might correlate with specific technical characteristics. If so, the length of historical program schedules which are affected by such challenges might correlate with specific technical characteristics. To test this hypothesis, parametric schedule estimating relationships (SERs) needed to be developed for predicting the length of time between major acquisition milestones based on technical parameters for future programs. At commencement of the research, however, no consistent nor generally-accepted datasets of missile and radar technical and program date parameters existed.

NCCA embarked on this path in two phases. The first phase placed emphasis on creating, collaboratively, datasets of prior missile and radar program information.[3] While seemingly simple, framing the necessary date (dependent[4]) and technical (independent) parameters proved challenging. As discussed later, changes in DoD's acquisition directive over time, precluded easy equivalency of milestone definitions. On the technical side, for both missiles and radars, interviewing engineers and investigation of development and manufacturing techniques and processes yielded a plethora of potential characteristics – of which only a few could realistically be recorded for various reasons. Assuming a relatively successful first phase, phase two would statistically evaluate potential relationships between the independent and dependent variables.

The effort was a joint-service activity, where the following organizations and agencies, in addition to NCCA, participated in vetting and validating the datasets:

- Air Force Cost Analysis Agency (AFCAA)
- Deputy Assistant Secretary of the Army for Cost and Economics (DASA-CE)
- U.S. Navy Naval Air Systems Command (NAVAIR)
- U.S. Navy Naval Sea Systems Command (NAVSEA)
- Marine Corps Systems Command (MARCORSYSCOM)
- Missile Defense Agency (MDA)

As a result of the NCCA team's extensive research and data collection, the Government cost community will significantly benefit from these comprehensive, jointly vetted and approved datasets—the first of their kind—and descriptive statistics for 58 missile and smart munition programs and 53 radar programs. Using these datasets and individual program documentation, estimators will be able to assess the reasonableness of planned schedules and draw analogies when estimating future missile and radar programs. While the great variability in schedule durations for both missile and radar programs precluded the development of any statistically significant SERs, analysts can use the descriptive

[3] Note here we refer to not just "data," but rather "information." The distinction is important, as the research found, in particular cases, programmatic references which helped illuminate issues with reported data requiring further investigation and/or parameters adjustment.

[4] The dependent variable is the length of time or duration between schedule event dates.

statistics for approximations of future program schedules. The following sections highlight the team's data collection and analysis processes and capture the lessons learned from this extensive study.

2 Data Collection

2.1 Schedule Data

For this study, the research team consulted unclassified, open sources for schedule data. Where more than one data source was available, the team established the following hierarchy of preferred data sources:[5]

1. Official DoD sources
2. DoD-sponsored research
3. Contractor sources
4. Web-based sources

Detailed schedule events and dates for each program came from congressionally mandated Selected Acquisition Reports (SARs) when available. Because the potentially important relationship between planned versus actual dates was beyond the scope of this study, the research team only included actual dates in the dataset, not the originally planned dates also in the SARs. Table 1 shows the schedule events for which the research team recorded dates, sources, exact wordings of the schedule events from the sources, and other comments in the dataset.

Table 1. Schedule Events Collected for Missile and Radar Programs

Major Schedule Events
Milestone (MS) A (Decision and Award)
MS B (Decision and Award)
MS C/Low-Rate Initial Production (LRIP) (Decision and Award)
Full-Rate Production (FRP) (Decision and Award)
Initial Operational Capability (IOC)
Other Schedule Events
Preliminary Design Review (PDR)
Critical Design Review (CDR)
Development Test and Evaluation (DT&E) (Start and End)
Initial Operational Test and Evaluation (IOT&E) (Start and End)

[5] There has become more concern, in developing realistic and reproducible cost estimates, to underscore the necessity to reflect the "pedigree" of data upon which estimates have been created. This collaboratively built hierarchy is intended to reflect the assessed pedigree of the collected data.

Because this study spans decades of missile and radar programs, evolutions in DoD Instruction (DoDI) 5000.02's definitions of major milestones required the team to make an important distinction between MS III and MS C.[6] In 2000 DoDI 5000.02 changed its major milestone designations from I, II, and III to A, B, and C, where MS I corresponds to MS A and MS II corresponds to MS B. MS III and MS C are not aligned: the old DoDI 5000.02 defined MS III as the FRP decision; while the new DoDI 5000.02 defines MS C as LRIP.[7] The schedule dataset in this study shows an average of 27 and 29 months, for missiles and radars respectively, between a program's MS C award and FRP decision, a noticeable difference.

The research team made another important distinction between MS decision and MS decision award dates. For the major milestones, the team identified both the MS decision dates and the corresponding contract award dates. For some programs, both the decision and award dates were available; but more often, the team found only the decision date or the award date. If only one of the dates was available, the team estimated the missing date using the average time between MS decision and award dates. Missile programs experienced an average delay of two months between MS decision and award, while radar programs experienced an average delay of three months between MS decision and award. These averages exclude significant outliers, such as programs that faced extensive contract award delays from vendor protests or contract negotiations.

2.2 Technical Data

For each of the 58 missiles and 53 radars, the research team compiled values for readily available, public domain technical parameters according to the hierarchy of data sources in Section 2.1. Table 2 lists the key missile and radar technical parameters for which the team recorded values. The team chose "high-level" technical parameters to describe missile and radar programs because the schedule data similarly provides high-level overviews of the complete development programs.

Particularly for the missile programs, the need for the cross-service collaborative teaming proved invaluable for adjudicating the hierarchy of data sources – both schedule and technical data. Regardless, the inability to find legitimate parameters consistently across missile and radar programs exacerbated the testing of hypothesized correlations, even for those relationships prophesied by engineering experts and literature.

[6] Note: this study did not attempt to incorporate programs, though allowed per DoD's acquisition directives through accelerated procedures (e.g., JUONs). In the future, the recent "NDAA-18 Section 804" language may further complicate the picture.

[7] In this paper, MS C always means LRIP and MS III always means FRP. Other missile schedule studies and compilations the team consulted for this project have often, erroneously, equated MS III to MS C.

Table 2. Key Technical Parameters

Missile Technical Parameters
Weight (lbs)
Maximum Range (nautical miles [nmi])
Velocity (ft/s)
Diameter (in)
Length (in)
Altitude (ft)
Volume (in^3)
Total Impulse (pounds of force per second)
Year of Technology (1962-2010)

Radar Technical Parameters
Frequency Band (GHz)
Weight (lbs)
Antenna Aperture (ft^2)
Number of Elements
Number of Transmit/Receive Modules
Average Power (kW)
Peak Power (kW)
Minimum Range (nmi)
Maximum Range (nmi)

Additionally, Table 3 and Table 4 list the key classifications for missiles and radars, respectively, for which the research team recorded data. The team used these qualitative parameters to group programs, stratify the datasets, and to suggest possible dummy variables during SER analysis.

Table 3. Key Missile Classifications

Category	Count
Missile – Rocket Propelled	40
Missile – Cruise	9
Smart Munition	9
Total	**58**

Type	Count
Air-to-Air	8
Air-to-Ground	22
Ground-to-Ground	14
Ground-to-Air	14
Total	**58**

Lead Service	Count
Air Force	12
Army	21
Navy	16
Joint	9
Total	**58**

Table 4. Key Radar Classifications

Category	Count
Airborne	31
Non-Airborne	17
Strategic	5
Total	**53**

Array Type	Count
AESA	27
PESA	7
Mechanically Scanned	10
Other	5
Total	**53**

Lead Service	Count
Air Force	25
Army	8
Navy	17
MDA	3
Total	**53**

2.3 Programmatic Information

The team recorded programmatic history for each missile and radar program. The team then evaluated the programmatic histories to better understand anomalies in schedule durations and to explain significant outliers in regressions. For some radar programs, enough information was available to make schedule adjustments to durations by a specific number of months for activities unrelated to actual program development or production.[8] Table 5 provides an excerpt from the team's radar programmatic history research.

Table 5. Example of Radar Programmatic History

Short Name	Program Overview	Radar Affiliation	Schedule Impacts
APG-66	Fire control radar, coherent, multi-mode, digital fire control sensor designed to provide all-weather air-to-air and air-to-surface modes with advanced dogfight and weapon delivery capabilities.	Upgrade to APQ-120	Schedule durations were shorter since development started through IR&D that resulted in developing a prototype of the system before a requirement was stated.
APQ-164	1970s-era first-ever PESA radar with a beam-steering controller and real-time processor that allowed simultaneous SAR and Terrain Following (TF) imagery.	Technology from APG-66 and common LRUs with APG-68	None Identified
AWACS RSIP	Supported air defense and tactical operations by providing extended, all-altitude radar surveillance over land and water. Improved target identification and resistance to radar jamming.	None Identified	Development schedule delays resulted from software development problems and integration slippage. Delays in IOT&E were due to a mishap that damaged radar components.

The team also collected radar genealogy information. Figure 1 shows an example of the genealogy for airborne radars. An understanding of radar genealogies will help analysts determine the most suitable radar analogies to use for a new radar program.

[8] For example, in one radar case, an "above-program" decision to change planned technology, and a significant contract protest, resulted in extensive program delays totally unrelated to the planned program.

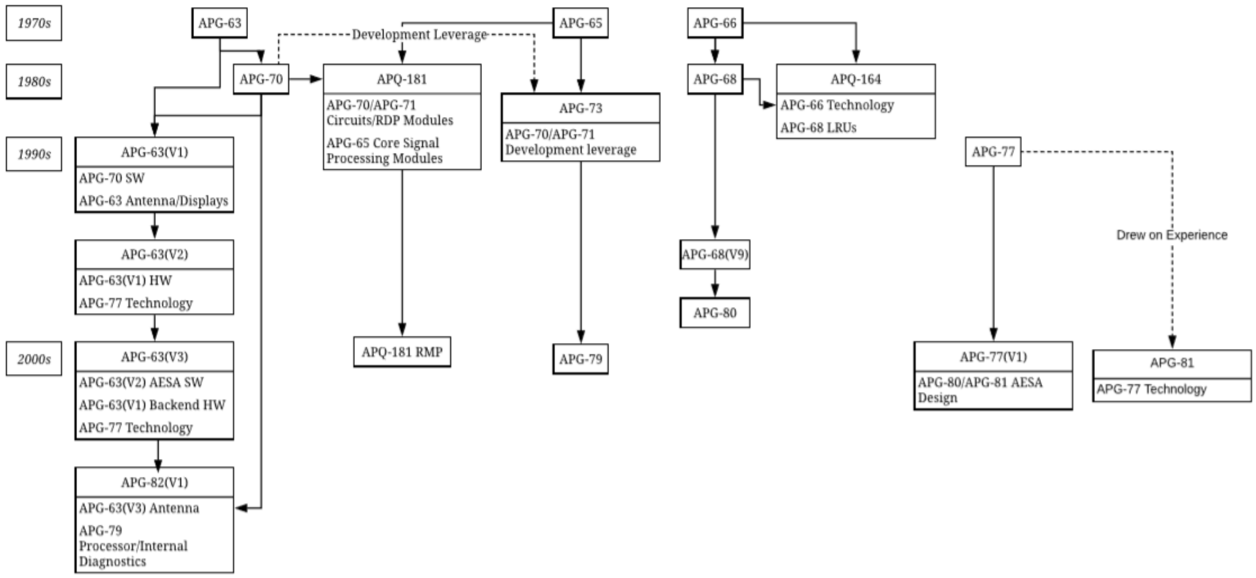

Figure 1. Airborne Radar Genealogy Example

2.4 Datasets

The team meticulously documented the schedule, technical, and programmatic information in a dataset for missiles and a separate dataset for radars. Within each dataset, for the schedule data collected, each date has the source, the exact wording of the milestone or event in the source, and any comments. This level of documentation was also done for technical values collected and for programmatic information.

For the missile dataset, a sample of four missiles is shown in Figure 2 illustrating schedule and technical data for which values were found. SAR schedule data was first extracted from the DoD DAMIR database and supplemented from other sources as illustrated for the AMRAAM missile at the top of Figure 2. Based on the detailed description of each event in the source, each date was assigned to one of the schedule events in Table 1; these assignments were not always obvious and sometimes required adjudication by the team and collaborating Service analysts. The dates for the major events were then summarized in the "Missile Schedule and Tech Data Master" file as illustrated in the middle of Figure 2. This process was repeated for every missile. Missile technical parameters are illustrated at the bottom of Figure 2. Eleven different official and unofficial sources were used. Often the values differed significantly for the same missile program. To resolve this problem, the team determined a primary source for each technical characteristic. The primary source was determined based on the lead Service for each missile. If the primary source did not contain data for a specific technical parameter, the team used one of the supplemental sources based on a comparison and evaluation of each of the supplemental sources. To indicate confidence in the preferred values, the team assigned a high, medium, or low pedigree to each preferred value.

NCCA DAMIR DATABASE - SAR SCHEDULE DATA EXTRACT

SAR Date	Program	SAR Schedule Event	Key Event	Actual Event Date
IDA P3014	AMRAAM	Milestone 0	MS 0 Decision	Oct-75
12/25/2015	AMRAAM	Milestone I (DSARC)	MS I Decision	Nov-78
IDA P3014	AMRAAM	EMD Start/EMD Contract Award	MS II Decision	Dec-81
12/25/2015	AMRAAM	Milestone IIIA (DAB)	LRIP Decision	Jun-87
12/25/2015	AMRAAM	Milestone IIIB (DAB) (Lot IV Full Go-Ahead Rate Production)	FRP Decision	Apr-91
12/25/2015	AMRAAM	IOC Air Force	IOC	Sep-91

MISSILE SCHEDULE & TECH DATA MASTER

50 programs	Actual Major Schedule Events					Major Milestone Durations (months)					Actual Other Schedule Events					
Missile/Ordnance Program	MS A Decision	MS B Decision	LRIP/ MS C	FRP/ MS III	IOC	MS A to B	MS B to LRIP	LRIP to FRP	MS B to FRP	MS B to IOC	PDR	CDR	DTE Start	DTE End	IOTE Start	IOTE End
ACM (AGM-129A)		Feb-83	Jul-86	Jul-91	Jan-93		41.0	60.0	101.0	119.1	Sep-84	Mar-85				
ALCM (AGM-86B)	Feb-74	Jan-77	Aug-78	Apr-80	Dec-82	35.0	19.0	20.0	39.0	71.0	Jun-77				Oct-79	Feb-80
AMRAAM (AIM-120A)	Nov-78	Dec-81	Jun-87	Apr-91	Sep-91	37.0	66.0	46.0	112.0	117.1	Sep-82	Mar-85			Oct-83	Jun-90
ATACMS-APAM (MGM-140A	Oct-82	Feb-86	Jan-89	Nov-90	Aug-90	40.1	35.0	22.0	57.0	54.0	Sep-86	Mar-87	Mar-89	Dec-89	Mar-90	Jun-90

Missile Technical Parameters

	Preferred Value	Source Used	Army 2016 Deskbook	AF Fact Sheet - Official	AF MSSRH 2011	Navy Fact File 2017-Official	Navy Weapons Handbook 2008-Official	Contractor	US Msl Data Book 2011/2001/1997	Designation-Systems.net	Fed Am Sc - Fas.org	Global Security.org	Jane's 2012
			1	2	3	4	5	6	7	8	9	10	11
	Weight (lbs) - Includes boosters												
	Preferred	Source	DB 16	AF FS	MSSRH	NFF	NWH 08	Contractor	DSA 11	Parsch	FAS	GS	Jane's
ACM (AGM-129A)	Value 2	AF FS	Value 1	Value 2	Value 3				Value 4	Value 5		Value 4	Value 5
ALCM (AGM-86B)	Value 2	AF FS	Value 1	Value 2	Value 3				Value 4	Value 1		Value 2	Value 1
AMRAAM (AIM-120A)	Value 2	AF FS	Value 1	Value 2	Value 3	Value 1	Value 4	Value 5	Value 4	Value 4		Value 2	Value 4
ATACMS-APAM (MGM-140A Block I)	Value 1	DB 16	Value 1		Value 2				Value 1	Value 2	Value 2		Value 2

Figure 2. Sample Missile Dataset

For the radar dataset, a sample of 15 radars is shown in Figure 3 illustrating schedule and technical data for which values were found. Because there were only 9 radars with SARs and because supplemental schedule data was less available than for missiles, dates were entered directly into a file like the missile "Missile Schedule and Tech Data Master". Radar technical parameters are illustrated at the bottom in Figure 3. The gray highlighted cells shown for schedule and technical data dramatically portray the amount of missing schedule and technical data. The missing data reduced the number of observations available for regressing schedule durations against technical parameters. There were far fewer missing data points in the case of missiles.

Schedule Data

Key | SAR Program | Est Dec/Award (+-3 mos) | Uncertain | Missing Data

	Programmatic Information				Actual Major Schedule Events												Actual Other Schedule Events						
	Radar Name	Lead Service	Category	Platform	MS A Decision	MS A Award	Platform MS B Decision	MS B Decision	MS B Award	Platform MS C Decision	MS C/LRIP Decision	LRIP Award	Platform FRP Decision	FRP Decision	FRP Award	IOC (Platform for Airborne)	PDR	CDR	DTE Start	DTE End	IOTE Start	IOTE End	First Flight
1	3DELRR	Air Force	Ground	Ground	May-09	Aug-09		Sep-14	May-17								Sep-10	Jun-18					
2	APG-63	Air Force	Airborne	F-15	Aug-68	Nov-68	Jan-70	Jan-70	Sep-70	Oct-72	Oct-72	Jan-73	Oct-75	Oct-75	Jan-76	Sep-75		Apr-71		Jun-77		Nov-75	Nov-72
3	APG-63(V1)	Air Force	Airborne	F-15	Feb-94	May-94	N/A	Aug-94	Nov-94					Aug-99	Nov-99		May-95	May-96	Jul-97	Jun-98	Jun-98	Dec-00	Jul-97
4	APG-63(V2)	Air Force	Airborne	F-15	Jul-96	Oct-96	N/A	Feb-97	May-97		May-98	Aug-98		Dec-98	Mar-99			Dec-97					Feb-99
5	APG-63(V3)	Air Force	Airborne	F-15			N/A	Oct-02	Jan-03		Jul-07	Oct-07											May-06
6	APG-65	Navy	Airborne	F-18	Oct-75	Jan-76	Dec-75	Dec-75	Aug-76	Dec-77	Dec-77	Mar-79	Jun-81	Jun-81	Sep-81	Mar-83		Jun-77		Mar-82	Oct-80	Jan-81	Mar-78
7	APG-66	Air Force	Airborne	F-16	Sep-74	Dec-74	Apr-75	Apr-75	Nov-75	Dec-76	Dec-76	Mar-77	Oct-77	Oct-77	Jan-78	Jun-80	Apr-76	Jun-76		Jan-79	Dec-76		May-77
8	APG-68	Air Force	Airborne	F-16			N/A	Jan-81	May-81														
9	APG-68(V9)	Air Force	Airborne	F-16			N/A	Oct-99	Jan-00								May-00	Sep-00					Aug-01
10	APG-70	Air Force	Airborne	F-15E			N/A	Nov-82	Feb-83									Dec-84					Jan-85
11	APG-73	Navy	Airborne	F-18			Jan-76	Mar-89	Jun-90		Mar-91	Jun-91		Jul-95	Oct-95		Aug-90	Nov-91		Mar-94			Mar-92
12	APG-77	Air Force	Airborne	F-22			Jun-91	Jun-91	Aug-91	Aug-01	Aug-01	Sep-01	Mar-05	Mar-05	Nov-06	Dec-05	Apr-93	Jun-95					Nov-97
13	APG-77(V1)	Air Force	Airborne	F-22			N/A	Feb-02	Feb-02		Aug-06	Nov-06					Dec-02	Apr-03					Mar-04
14	APG-78	Army	Airborne	Apache			Dec-90	Dec-90	Dec-90	Sep-94	Sep-94	Dec-94	Oct-95	Oct-95	Dec-95	Nov-98					Jan-95	Mar-95	
15	APG-79	Navy	Airborne	F-18 E/F	Jan-00	Apr-00	N/A	Feb-01	Feb-01		Jul-03	Oct-03		Feb-07	Jan-07	Nov-06	Dec-00	Aug-01	Apr-02		Feb-03		Feb-03

Technical Data

	Programmatic Information				Technical Characteristics													
	Radar Name	Lead Service	Category	Platform	Array Type	Frequency Band	Freq. Range in GHz	Avg. Freq. in GHz	% New Design	Weight (Lbs)	Antenna Aperture (Ft^2)	No. of Elements	Transmit/ Receive Modules	Avg Power (KWs)	Peak Power (KWs)	Min Range (Nmi)	Max Range (Nmi)	Software Size (KLOC)
1	3DELRR	Air Force	Ground	Ground	AESA	C	Value 1	Value 1		Value 1	Value 1	Value 1	Value 1					
2	APG-63	Air Force	Airborne	F-15	Mech. Scan	I/J (includes X)	Value 2	Value 2		Value 2	Value 2				Value 1		Value 1	
3	APG-63(V1)	Air Force	Airborne	F-15	Mech. Scan	I/J (includes X)	Value 2	Value 2			Value 3						Value 2	
4	APG-63(V2)	Air Force	Airborne	F-15	AESA	X	Value 2	Value 2			Value 4						Value 3	
5	APG-63(V3)	Air Force	Airborne	F-15	AESA	X	Value 2	Value 2	Value 1	Value 3	Value 5	Value 2	Value 2				Value 4	
6	APG-65	Navy	Airborne	F-18	Mech. Scan	X	Value 2	Value 2		Value 4	Value 6						Value 5	
7	APG-66	Air Force	Airborne	F-16	Mech. Scan	X	Value 2	Value 2		Value 5	Value 7			Value 1	Value 2	Value 1	Value 6	
8	APG-68	Air Force	Airborne	F-16	Mech. Scan	X	Value 2	Value 2		Value 6	Value 8						Value 7	
9	APG-68(V9)	Air Force	Airborne	F-16	Mech. Scan	X	Value 2	Value 2		Value 7	Value 9						Value 8	Value 1
10	APG-70	Air Force	Airborne	F-15E	Mech. Scan	I/J (includes X)	Value 2	Value 2		Value 8	Value 10						Value 9	
11	APG-73	Navy	Airborne	F-18	Mech. Scan	X	Value 2	Value 2		Value 9	Value 11					Value 1	Value 10	
12	APG-77	Air Force	Airborne	F-22	AESA	X	Value 2	Value 2	Value 2	Value 10	Value 12		Value 3		Value 3		Value 11	Value 2
13	APG-77(V1)	Air Force	Airborne	F-22	AESA	X	Value 2	Value 2	Value 3	Value 11	Value 13	Value 3	Value 4		Value 4		Value 12	Value 3
14	APG-78	Army	Airborne	Apache	Mech. Scan	Ka	Value 4	Value 4			Value 14						Value 13	
15	APG-79	Navy	Airborne	F-18 E/F	AESA	X	Value 2	Value 2	Value 4	Value 12	Value 15	Value 4	Value 5				Value 14	Value 4

Figure 3. Sample Radar Dataset

3 Data Analysis

3.1 SER Process

After collecting and cleansing the schedule and technical data, the team used the dates for schedule events to calculate the duration (in months) between major acquisition milestones and events. The team began the evaluation of durations as a function of technical parameters by calculating the correlation coefficient *r* between each duration type (e.g., MS B to MS C/LRIP) and each major technical parameter. For ease of interpretation, the team squared the *r* values to calculate the coefficient of determination, R^2. The team then performed bivariate regressions for the technical parameter-schedule duration pairs with the highest R^2 values and a number of observations greater than or equal to five. Regressions used the Ordinary Least Squares (OLS) method in JMP®, a statistical analysis program from SAS. To evaluate the "goodness of fit" for each regression, the team performed a *t*-test. If the *p* values, calculated as part of the *t*-test, for each coefficient were less than 0.05, then the team accepted the regression coefficients as statistically significant.

The team identified outliers when evaluating bivariate regression scatter plots. To understand why a program was an outlier, the team used programmatic history. The team adjusted and/or excluded outlier durations if programmatic history showed reasonable justification. The research specifically looked for delays outside of normal development program delays, such as unspecified Congressional holds, extended contract disputes or negotiations, or a high degree of corporate development leverage.[9] After making outlier schedule adjustments or exclusions, the team re-ran the regressions.

If the regression coefficients were statistically significant and made sense technically, the team evaluated transformations (e.g., power and logarithmic) and triad nonlinear models. The team used CO$TAT, a statistical program from Tecolote Research, to evaluate these models using the Minimum Unbiased Percentage Error (MUPE) and OLS methods. Transformations and nonlinear models yielded worse statistics than the simple linear regression models except for one radar case, which had insignificant improvement.

Next, the team evaluated multivariate regressions. To evaluate multivariate regressions, the team used stepwise regression in JMP® to regress schedule durations using more than one technical parameter, where each parameter is systematically introduced and removed from the model to identify which combination of parameters yields the highest R^2 value. During this process, JMP® excluded missiles or radars missing values for the schedule duration and parameters, reducing the analysis sample size. The team applied stepwise regression to all schedule durations using all linear combinations of key technical parameters. Multivariate regression models evaluated for both missiles and radars also yielded worse statistics.

Finally, for each schedule duration, the team calculated descriptive statistics, generated a histogram, and identified the best fit continuous probability distribution. These statistics and distributions can help to make informed inferences about future missile and radar program durations for each schedule duration.

3.2 SER Results

During the regression process, the team ran and evaluated hundreds of different regression models for each dataset. The team considered a regression to be a statistically significant SER if it met the following criteria:

(a) $p < 0.05$ to reject the null hypothesis
(b) coefficient of determination $R^2 > 0.60$
(c) number of observations[10] $N > 10$

The missile and radar SER results are summarized below including scatter plots which show the great variability in the durations.

[9] See previous footnote

[10] According to the "Joint Agency Cost Estimating Relationship (CER) Development Handbook" (9 February 2018), CERs should have at least five observations. For CERs, five data points may be appropriate where the dataset used to develop the CER is small. However, in the context of the missile and radar datasets with 50 missiles and 53 radars, the number of observations should be relatively more.

3.2.1 Missile SER Results

Based on the above criteria, none of the missile SERs met the threshold for statistical significance. The highest R^2 for a missile SER was 0.08, meaning the best SER from the missile dataset could only explain about 8% of the schedule data variation. Figure 4 shows a scatter plot of schedule duration against a technical parameter with its bivariate fit and provides a visual example of the scatter in the missile data. The shaded region surrounding the fitted line represents the 95% confidence interval.

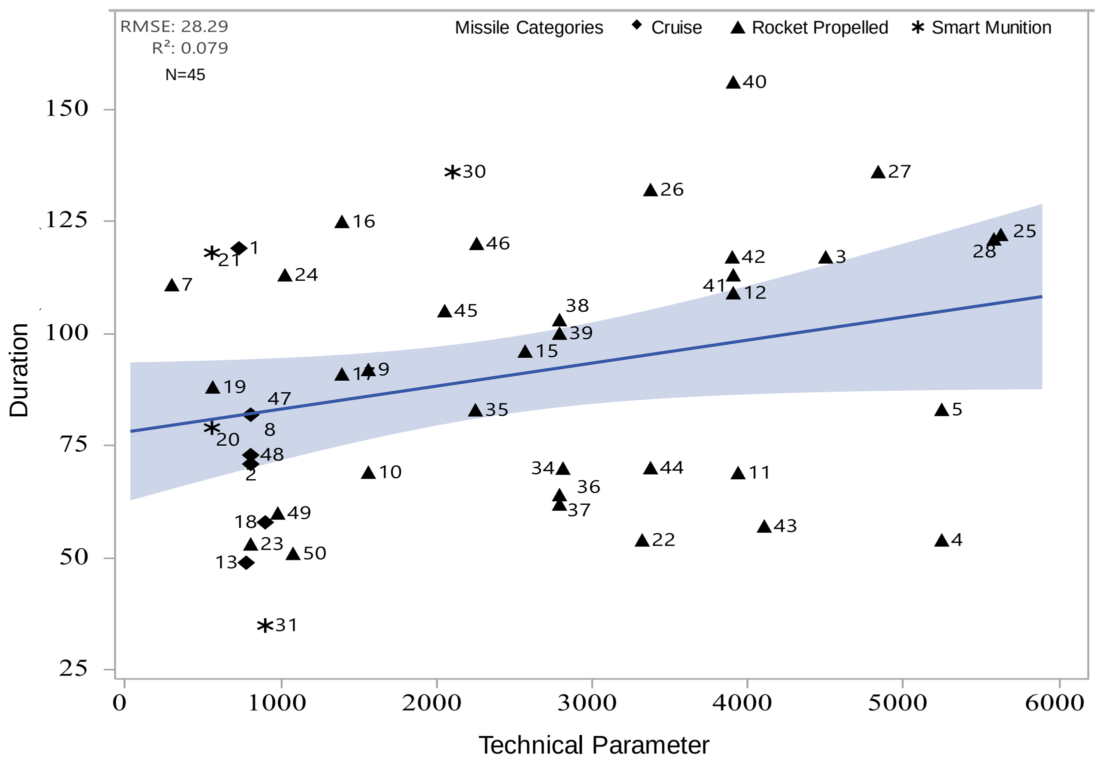

Figure 4. Bivariate Plot of Schedule Duration and Technical Parameter – Missile Example

The dispersion of the missile program durations in Figure 4 demonstrates the weak relationship between the schedule duration and technical parameter. The team also explored linear and nonlinear models such as power, exponential, and logarithmic models. The team then applied this process to the different missile stratifications and used dummy variables.[11] This comprehensive approach yielded no meaningful relationships. While stratification slightly improved R^2 values, the resulting models were still insufficient for predicting schedule durations. The scatter plots together with the

[11] In peer reviews, concern arose about establishing population inferences with, likely, vastly different types of systems in the dataset (e.g., issue of pooled regressions). Nevertheless, even when stratified (and compounded by missing data elements), correlative relationships failed to emerge.

programmatic histories and descriptive statistics can help in determining the best analogies for analysts to use when evaluating future missile program schedules

3.2.2 Radar SER Results

Based on the three criteria for statistically significant SERs discussed earlier, two radar SERs met two of the criteria; these regressions had coefficients with $p < 0.05$ and $R^2 > 0.60$. However, the SERs had only $N = 6$ and $N = 8$ observations, which is few observations relative to the 53 radars in the dataset. Since these SERs did not meet all three criteria, the research team did not consider these regressions to be acceptable SERs.

Figure 5 shows a scatter plot of schedule duration against a technical parameter against a schedule duration with its bivariate fit and provides a visual example of the scatter in the radar data. The shaded region surrounding the fitted line represents the 95% confidence interval.

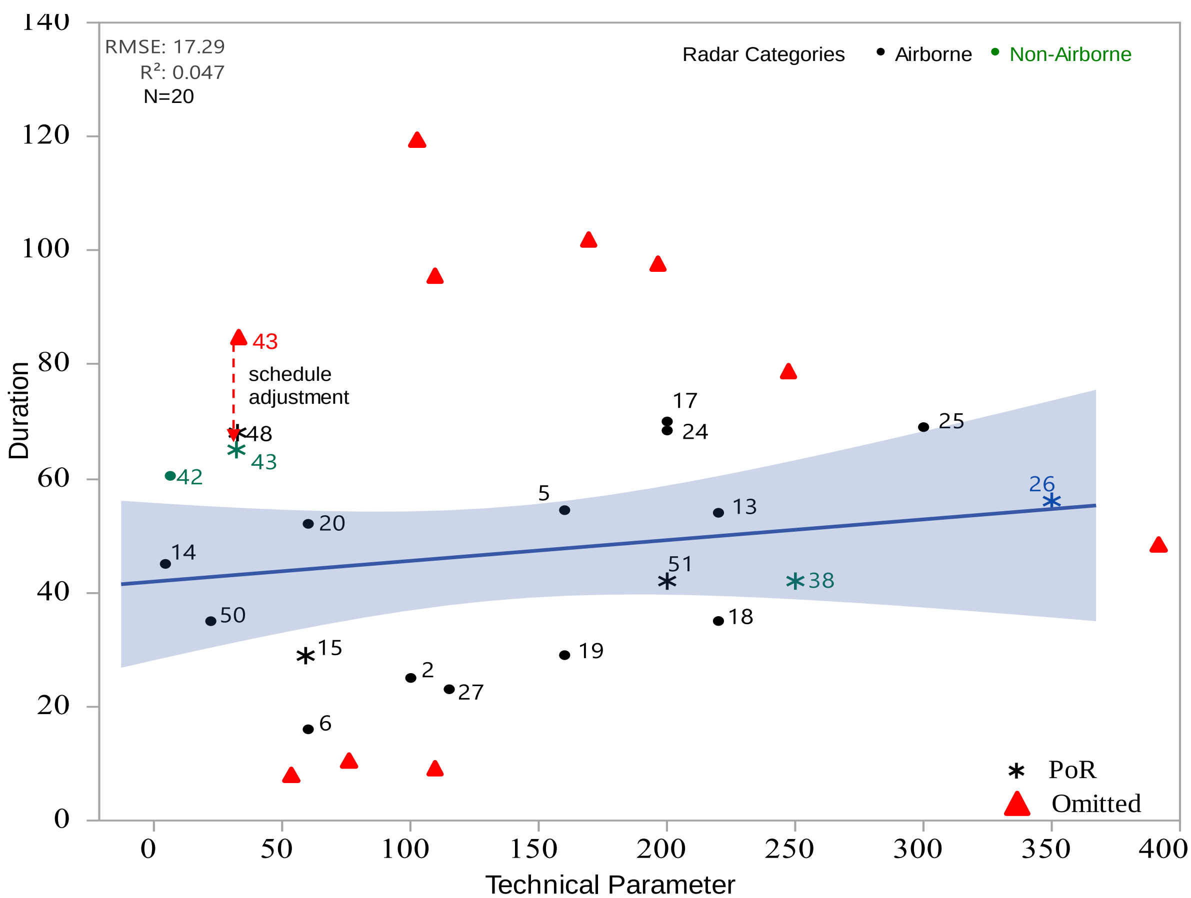

Figure 5. Bivariate Plot of Schedule Duration and Technical Parameter–Radar Example

The dispersion of the radar program durations in Figure 5 demonstrates the weak relationship between the schedule duration and technical parameter. The team also explored linear and nonlinear models such as power, logarithmic, triad models. The team then applied this process to the different radar stratifications and used dummy variables.[12] This comprehensive approach yielded no meaningful relationships. The scatter plots show the great variation in schedule durations. The scatter plots together with the programmatic histories and descriptive statistics can help in determining the best analogies for analysts to use when evaluating future radar program schedules

3.3 Descriptive Statistics

Table 6 and Table 7 provide descriptive statistics for major schedule durations of the full missile dataset and the full (non-strategic) radar dataset, respectively.

Table 6. Descriptive Statistics for Full Missile Dataset

	MS A to MS B	MS B to MS C	MS C to FRP	MS B to FRP	MS B to IOC
N	26	45	38	43	45
Mean	35	56	29	82	91
Standard Deviation	14	22	20	30	29
Coefficient of Variation (CV)	40%	39%	67%	36%	32%
Maximum	64	105	111	147	156
Median	35	56	24	81	88
Minimum	11	13	3.9	24	35

[12] See previous footnote.

Table 7. Descriptive Statistics for Full Radar (Non-Strategic) Dataset

	MS A to MS B	MS B to PDR	MS B to CDR	MS B to MS C	MS B to FRP	MS B to IOC
N	12	19	27	21	16	16
Mean	24	12	20	47	67	80
Standard Deviation	22	7	11	17	22	22
CV	88%	54%	57%	36%	33%	28%
Maximum	69	26	46	70	111	121
Median	16	13	18	51	60	78
Minimum	3	2	2	16	36	52

4 Lessons Learned

As a result of the missile and radar research, many lessons were learned. First, multiple reliable sources, had different schedule and technical information for the same program. The data conflict required a data hierarchy to be established to determine a preferred source.

Second, when collecting milestone schedule dates, there was value in collecting both the milestone decision and milestone decision contract award date. On average, there was a two- month delay for missiles and a three-month delay for radars between milestone decision and contract award; this excludes outliers where protests, contract negotiations and other circumstances caused long delays. Where only a Decision or Award date was known, the other date is estimated by adding or subtracting two or three months. Another reason for having both milestone decision and contract award dates is to be able to calculate durations starting at an award date which is when development engineering and production work generally begins.

A third lesson learned was the importance of collecting the programmatic history. This helps in understanding duration outliers, commonality with other missiles or radars, and whether a historical program provides a good analogy to a new program. Programmatic history research led to understanding if there were extenuating circumstances impacting the program schedules including technical problems causing major redesign, extended testing, and funding cuts.

A fourth lesson learned, which applies only to radars, is that a platform subsystem (i.e., aircraft or ship) may drive the schedule for the subsystem.

5 Conclusions

The research team produced a comprehensive, jointly vetted dataset of schedule events and technical parameters for 58 missiles and 53 radars. The research team used JMP® and CO$TAT to identify potential schedule estimating relationships (SERs) within the dataset. Bivariate linear and multivariate regressions were evaluated for the full dataset and for stratifications of the dataset using OLS and MUPE methods. The team also evaluated other linear models such as power and logarithmic transformations in addition to the nonlinear model.

During the regression process, the team ran and evaluated hundreds of different regression models. A regression was considered a statistically significant SER if (a) the regression had statistically significant coefficients, (b) the R^2 was greater than 0.60, and (c) there were a reasonable number of observations relative to the dataset. Although no SERs were acceptable, the resultant comprehensive missile and radar datasets developed in this study and the descriptive statistics are useful in evaluating future missile and radar program schedules. Also, the bivariate scatter plots, programmatic information, genealogies and technical information are useful in selecting analogies for estimating future missile and radar program schedules.

<u>Bottom Lines</u>:

- Bivariate plots exhibit too much "scatter" [see Figure 4 and Figure 5] to accept any derived SER.
- The collaboratively-collected data is incredibly useful as a foundation for analogy-based estimating.