NAVAIR Cost Growth Study
A Cohorted Study of The Effects of Era, Size, Acquisition Phase, Phase Correlation and Cost Drivers

The goal of this study was to further analyze Navy programs to determine the cause of cost growth (technical or cost estimating) and to develop a concept for estimating future cost growth. This understanding will allow program offices to plan for cost growth (adjust cost estimates and budget with cost risk dollars) and/or mitigate the high risks in the program to minimize cost growth.

A significant advance presented in this paper is the introduction of cohort tracking. Programs that did have a Dem Val phase are compared to programs that did not. This distinction will be shown to have a significant effect on predicted cost growth in both RDT&E and Procurement appropriations. The mixing of these two types of programs causes significant errors. It will be shown, for example that the growth during E&MD for programs that *did* have a Dem Val is quite different than the growth during E&MD for programs that *did not* have a Dem Val. This effect has been completely overlooked in previous studies, and is it is critical to include it.

This paper presents analysis and conclusions on the following topics:

- Does program size affect cost growth? There is a clear trend showing a higher tolerance for cost growth in smaller programs.
- Does the era of the program effect cost growth? There is convincing evidence that cost growth has decreased over time.
- Does correlation exist? Analysis shows that correlation exists between phases and between appropriations in the EMD phase.
- Does the structure of a program's life cycle affect cost growth? Analysis of cohorted groups determined that the inclusion of a PDRR phase as well as the maturity of a program impacts its cost growth.
- Is cost growth affected by commodity type? Descriptive results from this study suggest that cost growth does differ for various commodities.

Authors:
Jessica R. Summerville
Megan E. Dameron
Cari L. Pullen
Richard L. Coleman
Donna M. Snead

NAVAIR Cost Growth Study

R.L. Coleman, M.E. Dameron, C.L. Pullen,
J.R. Summerville, D.M. Snead

NORTHROP GRUMMAN

LITTON TASC
A Subsidiary of Northrop Grumman

Table of Contents

Introduction

Purpose:
The goal of this study is to further analyze Navy programs to determine the cause of cost growth (technical or cost estimating) and to develop a concept for estimating future cost growth. This understanding will allow program offices to plan for cost growth (adjust cost estimates and budget with cost risk dollars) and/or mitigate the high risks in the program to minimize cost growth.

The paper will:

- Investigate and understand previous cost growth studies of DoD and Navy aircraft, helicopters, missiles, and NAVAIR electronics and use these as a starting point for this research.
- Review SAR, CCDR, and any other sources to collect the appropriate data needed to determine cost growth of Navy ACAT I, II, and III programs.
- Analyze/process data and incorporate into a NAVAIR Cost Growth Database.

The attached Appendix provides a summary of "thumb rules" derived from the research presented here. This section has been and will continue to be updated as new data is added to the database.

Background:
Risk is an important part of cost estimation, and can be used to adjust budgets for historical cost growth. Cost growth is the increase in the cost of a system from inception to completion, often expressed as a percent of the initial estimate. Cost risk refers to the funds set aside to cover predicted cost growth. There are several types of cost growth and risk: Cost Estimating (CE), Schedule / Technical (S/T), Requirements, and Threat. CE risk is the risk due to cost estimating errors, and the statistical uncertainty in the estimate. S/T risk is risk due to inability to conquer problems posed by the intended design in the current CARD. Requirements Risk occurs due to as-yet-unseen design shift from the current CARD arising due to CARD shortfalls due to the inability of the intended design to perform the (unchanged) intended mission. In other words, we didn't understand the solution. Threat Risk is risk due to as-yet-unrevealed threat shift from the current STAR, i.e. we didn't understand the problem. CE and S/T growth are often broken out separately. Requirements and threat growth are generally implicitly included in S/T growth.

Data:
Three main data sets were used for this analysis, each with different information and shortcomings.

RAND 93: The RAND 93 data set is derived from Selected Acquisition Reports (SARs). SARs divide estimates into three groups:

1. Planning Estimate (PE) refers to an estimate made around Milestone I (PDRR) and adjusted during the phase.

2. Development Estimate (DE) refers to an estimate made at Milestone II, just prior to EMD and adjusted during the phase.
3. Production Estimate (PdE) is an estimate made at Milestone III, just before the start of production and adjusted during the phase.

The RAND 93 data set contains information from the SARs for 244 DoD programs, 36 of which are NAVAIR. Total Acquisition Cost Growth Factors (CGFs) are given for PE, DE, and PdE for both RDT&E and Procurement. Some programs have PE, DE, and PdE, while others have only a subset of the three estimates. In addition, some have both RDT&E and Procurement, but others only have information for one appropriation. The RAND 93 data is provided only at the total cost growth level – breakouts of CE and S/T growth are not available.

SAR 00: The SAR 00 data set is also derived from SARs and includes 16 NAVAIR programs. All information available is for the DE phase. For each program, CGFs are provided for each of the growth categories above as well as for Total, CE Growth, and Total Less CE (S/T). These CGFs are provided for Total Acquisition, RDT&E, and Procurement. However, program cost is included only at the Acquisition level, not broken out into RDT&E and Procurement funds. Therefore, CGFs cannot be dollar weighted by appropriation.

Contract Data: Contract data for small NAVAIR programs was collected from a variety of sources, including CPSs, CPRs, C/SSRs, CFSRs, and CCDRs. Program names were coded due to issues with proprietary Contract data. The data set includes estimates for 25 contracts – 21 RDT&E and 4 Procurement (Note 6 additional data points were provided very close to completion of this work. They are included in the database, but not in the analysis presented here). In a few cases, the data set contains multiple contracts on the same program. For this analysis, each contract is treated as separate data point. The data is provided in $TY, so the midpoint in time was used for escalation to $FY00. The data set does not indicate the PDRR/EMD blend of the estimates (i.e., it is not known if the estimates correspond to PE, DE, or PdE, or which programs had which phases). In addition, the Contract Data does not separate CE and S/T growth.

Analysis:

The remainder of the paper will present the analysis performed. Descriptive analysis (average, standard deviation, and the like) will be provided for each data source. The SAR 00 and Contract Data will be compared to RAND 93. "Cohort Tracking" will be performed for RAND 93 and SAR 00. NAVAIR programs will be compared to DoD and Ships. In addition, the effects of program size and era on cost growth will be explored. Finally, phase-to-phase correlation will be examined in the RAND 93 data set.

Snapshot Comparison – RAND 93 and SAR 00

Principal Analysts – Cari L. Pullen and Megan E. Dameron

Purpose:

To determine the comparability of the RAND 93 and SAR 00 data sets. If the two data sets are comparable, then analysis performed using the larger, but older, RAND 93 data can be applied to the smaller, but newer, SAR 00 data set.

Approach:

"Snapshot Comparison" is the comparison of data on the same programs at two or more different points in time. Programs common to both RAND 93 and SAR 00 were extracted. These programs were:

Program	Platform
AV-8B (Harrier)	Aircraft
F/A-18A/B/C/D	Aircraft
F-14D (Tomcat)	Aircraft
CH/MH-53D/E	Helo
AIM-54C (Phoenix)	Missile
HARM (AGM-88)	Missile
JSOW Baseline and BLU-108	Missile
E-6A TACAMO	Electronic
SH-60R	Electronic
Lamps MK III	Electronic
AIM-9M (Sidewinder)	Missile
F/A-18E/F	Aircraft

Analysis was conducted on the two subsets of common programs. Note that these subsets contained only NAVAIR programs. There were limitations to both data sets. For example, Cost Estimating (CE) risk was not broken out in RAND 93. In SAR 00, averages could not be dollar weighted for RDT&E and Procurement, as dollar values were available only at the Total Acquisition level.

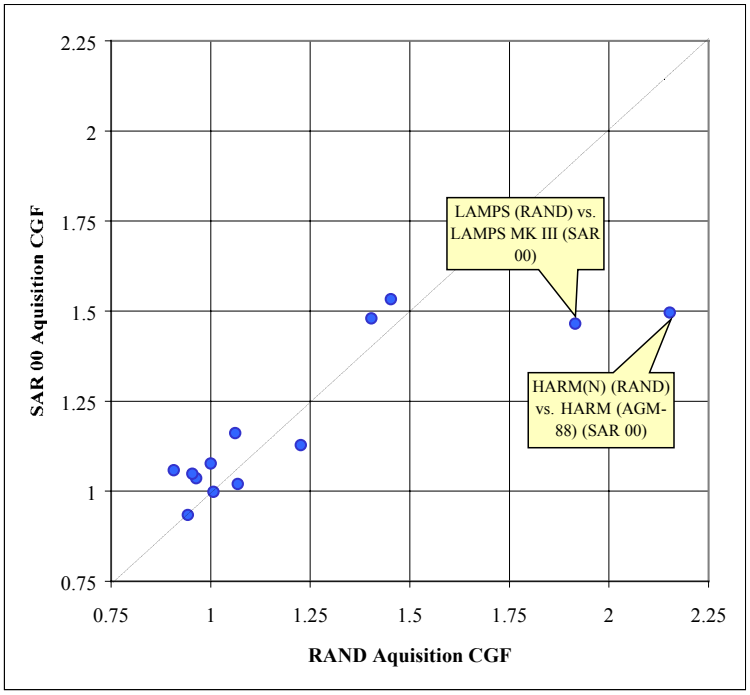
Results:

Acquisition Cost Growth Factors (CGFs) were calculated for both RAND 93 and SAR 00. No major difference was found between the data sets. The Acquisition CGF was 1.11 for RAND 93 and 1.12 for SAR 00.

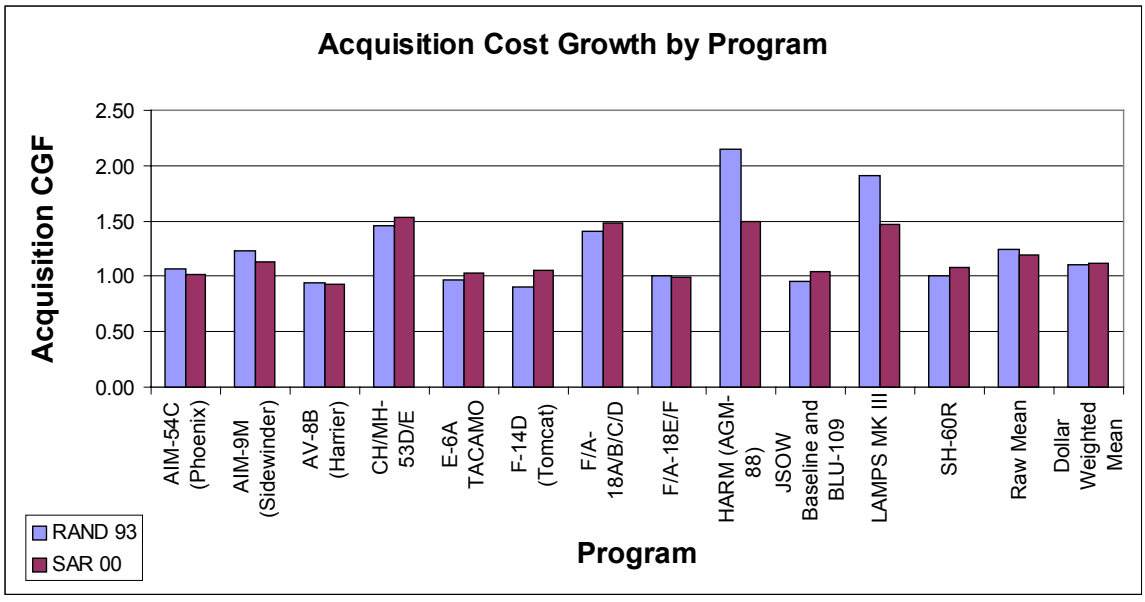
RAND 93		
	Total $	Dollar Weighted Mean
Total	136,954	1.11
R&D	16,855	1.25
Proc	120,099	1.09

SAR 00		
	Total $	Dollar Weighted Mean
Total	159,094	1.12
CE		1.01
Total less CE		1.11

The following scatter diagram plots the Acquisition CGFs from RAND 93 and SAR 00 for each program. If a point falls on the 45° line, then the program's CGF is identical in both data sets. So, from the graph below, it appears that the RAND 93 and SAR 00 data sets are comparable with the exception of HARM and LAMPS.



The chart below shows the RAND 93 and SAR 00 Acquisition CGFs for each program. The two data sets are very close for all programs except HARM and LAMPS.



Conclusions:

The RAND 93 and SAR 00 data sets are comparable; therefore, general conclusions from either data set are reasonable. SAR 00 is newer and should be used whenever possible. RAND 93 has more information and larger n, so some general and comparison study (NAVAIR vs. DoD vs. Ships, Size, Era, and Correlation) can be conducted there.

Contract Data Analysis

Principal Analyst – Jessica R. Summerville

Purpose:

To explore cost growth in small programs using Contract data from sources such as CPSs, CPRs, C/SSRs, CFSRs, and CCDRs.

Approach:

Descriptive statistics were computed for both RDT&E and Procurement appropriations, though it should be noted there were only 4 Procurement observations at the time the analysis was done. Descriptive and inferential statistical analysis comparing contract data with SAR data from small programs was performed only for the RDT&E data. There are three RDT&E programs with extremely high cost growth factors (over 20 times that of the original estimate) that may be considered outliers. However, these points cannot be removed without further information. Therefore, analysis is shown both with and without the three potential outliers.

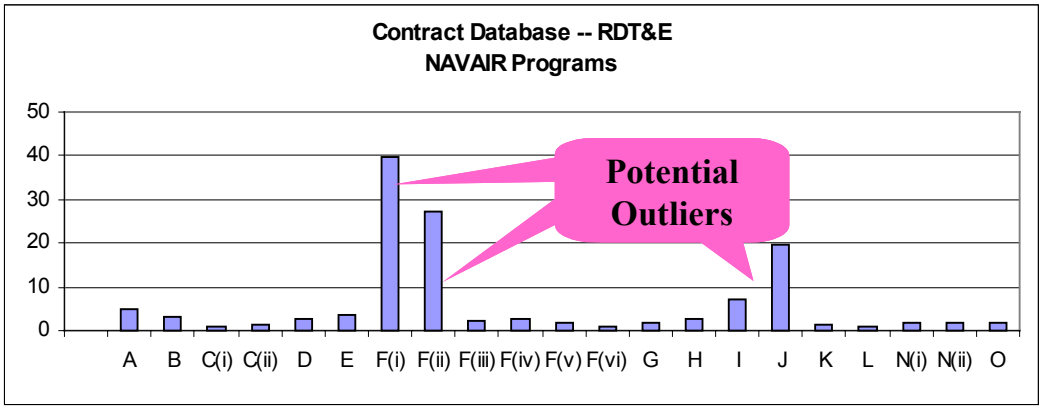
Note that the data used in this analysis is presented in a coded format (e.g. Contract A, Contract B, etc.) due to proprietary issues with contract data. The following caveats and concerns should also be noted:

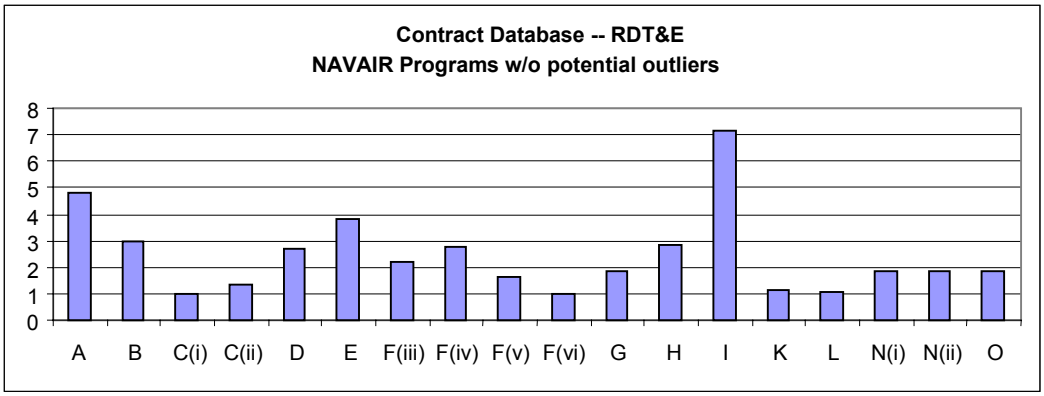
- In a few cases, there are multiple contracts on the same program. Each contract was treated as a separate data point.
- The data was provided in $TY, so the midpoint in time was used for escalation to $FY00.
- The PDRR/EMD blend of estimates within the data set is unknown.
- Variance in cost due to changes in quantity from the onset of each program were not available, and therefore the CGFs were not adjusted for this. This is more of a concern in Procurement.

Results:

RDT&E:

The following bar charts show the RDT&E cost growth by program, both with and without potential outliers. In general, the CGFs are much larger than typically seen in SAR databases.

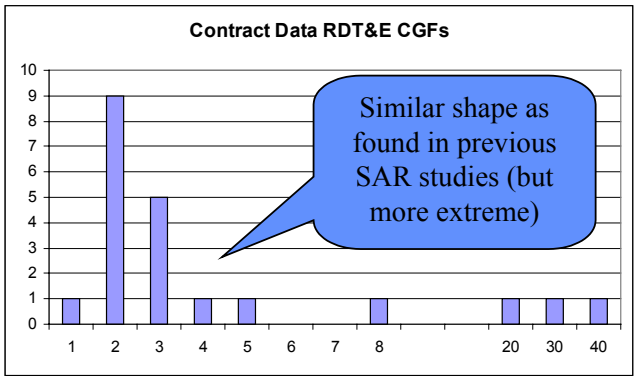




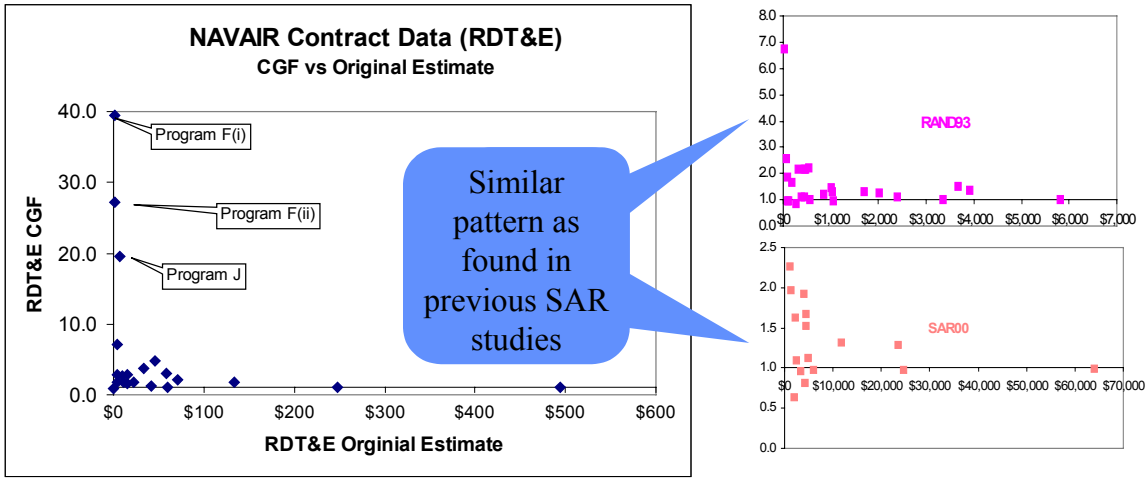
The RDT&E descriptive statistics are displayed below (n=21 for all data and n=18 with potential outliers removed). The dollar-weighted mean CGF of 1.61 is somewhat higher than dollar-weighted means observed in previous SAR studies.

	All Data	w/o Potential Outliers
Raw Mean	6.19	2.42
$-Wgt Mean	1.75	1.61
Raw Std Dev	10.07	1.56
CV	163%	64%

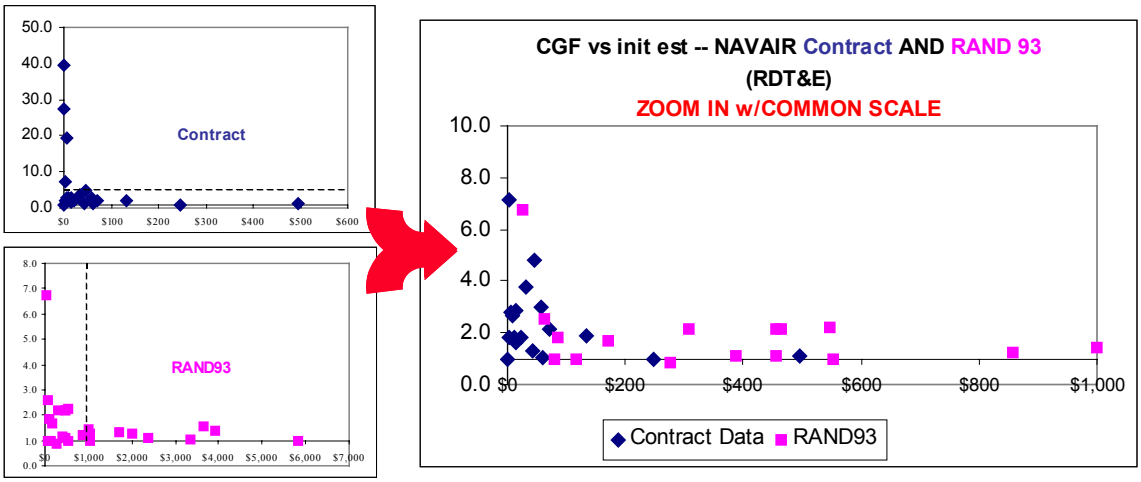
The histogram below shows the frequency of RDT&E CGFs. The shape of the histogram is similar to those observed in previous SAR studies, but is more extreme.



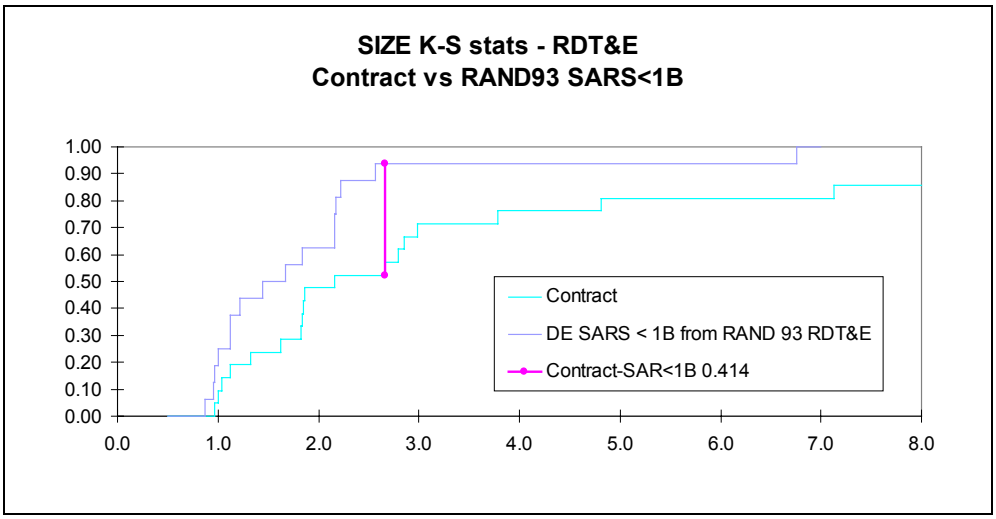
Next, a scatter plot of program size vs. CGF from the Contract Data was compared to plots from NAVAIR data in RAND 93 and SAR 00 (all DE data used to compare since the PDRR/EMD blend of Contract data is unknown). The following picture shows that the Contract Data has a similar pattern as that found in the two SAR based data sets. However, the Contract Data is much more extreme on the CGF axis, i.e. it has several programs with very high cost growth. This may be because the Contract data set consists of small programs and more growth is probably tolerated for small programs than for large programs.

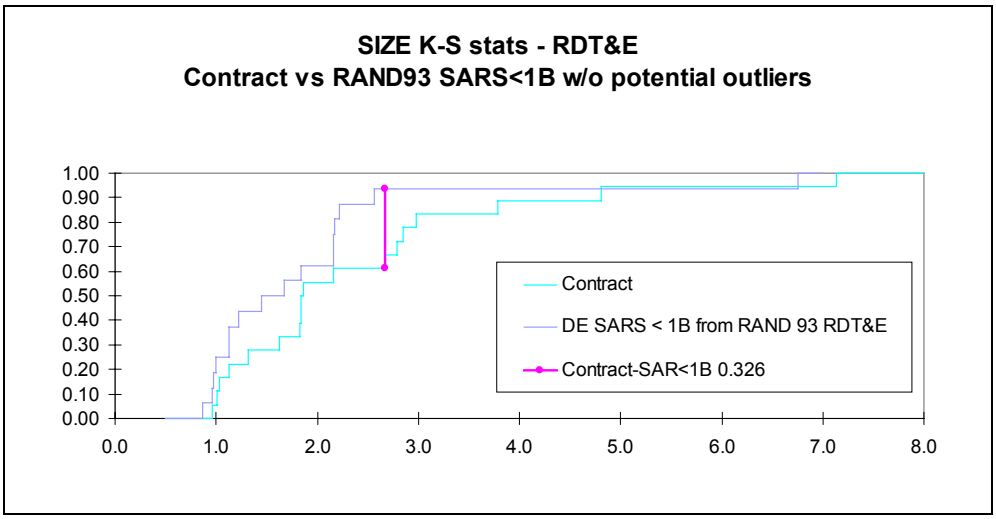


The next graph zooms in on the Contract and RAND 93 data using a common scale. Note that only programs less than $1B and programs with CGFs below 8.0 are shown, in order to make the graph scale useful. The Contract Data blends well with RAND 93 and continues the trend of smaller programs having larger cost growth.



Next, the distributions of the Contract and RAND 93 data sets were tested for identical distributions using a Kolmogorov-Smirnov (K-S) type test. The graphs below show the empirical cumulative distribution functions for both the Contract data (with and without potential outliers) and the RAND 93 data for programs less than $1B. The K-S test statistic is the maximum distance between the two curves, shown by the vertical line in the graphs.



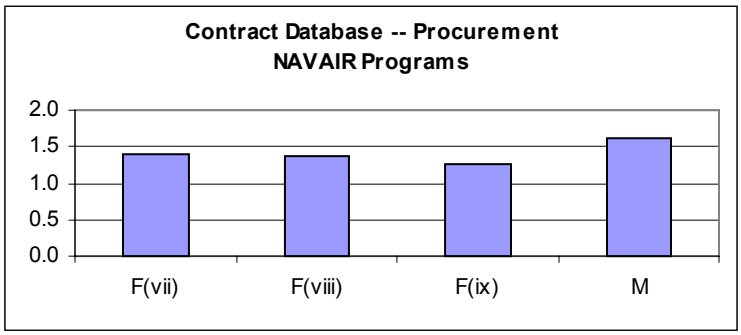


The test statistics are summarized in the table below. The K-S test supports identical distributions, with or without the potential outliers.

	K-S Critical Value	K-S Test Statistic	Identical Distribution?
w/All Contract	0.451	0.414	Yes
w/o potential outliers	0.467	0.326	Yes

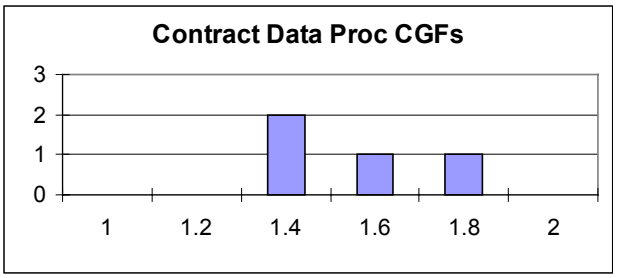
Procurement:

The following chart shows the Procurement cost growth by program. The CGFs here are not drastically different from typical SAR results, but the sample size is very small.



The Procurement descriptive statistics and histogram are displayed below (n=4). The dollar weighted mean CGF of 1.37 is more consistent with SAR results, but the sample size is too small to draw meaningful conclusions.

Raw Mean	1.42
$-Wgt Mean	1.37
Raw Std Dev	0.14
CV	10%



There was not enough Procurement data to perform any additional analysis.

Conclusions:

The Contract data mimics patterns observed in the SAR based data sets. The shapes of the distributions are very similar. In addition, the Contract data supports the idea that higher growth is tolerated more in small programs than in large programs. Descriptive statistics indicate more extreme growth and dispersion in the Contract data than in RAND 93 and SAR 00. Inferential statistics suggest no statistically significant difference between Contract data and RAND 93 data for programs less than $1 billion in RDT&E.

It is recommended at this point to continue using SAR data for predictive modeling, as the analysis of Contract data is still new territory with many caveats and lessons yet to be learned.

Cohort Tracking

Principal Analyst – Megan Dameron

RAND 93 Data:

Purpose:

Previous risk studies were performed without regard to which phases a particular program passed through. For example, programs with PDRR were grouped together with programs that did not have PDRR. A problem arises if cost growth in programs with PDRR is inherently different than that in programs without a planning phase. The purpose of this study is to determine how cost growth occurs throughout the phases of a program.

Approach:

"Cohort tracking" is the comparison of data on the same programs as they progress through time, usually in a single data set. Analysis was conducted using the RAND 93 data set. Programs were separated into the following categories for study:

- RDT&E
  - Programs with PE and DE for DoD, NAVAIR, and Ship
  - Programs with DE for DoD, NAVAIR, and Ship
- Procurement
  - Programs with PE, DE, and PdE for DoD, NAVAIR, and Ship
  - Programs with DE and PdE for DoD, NAVAIR, and Ship
  - Programs with DE for DoD, NAVAIR, and Ship

RDT&E Results:

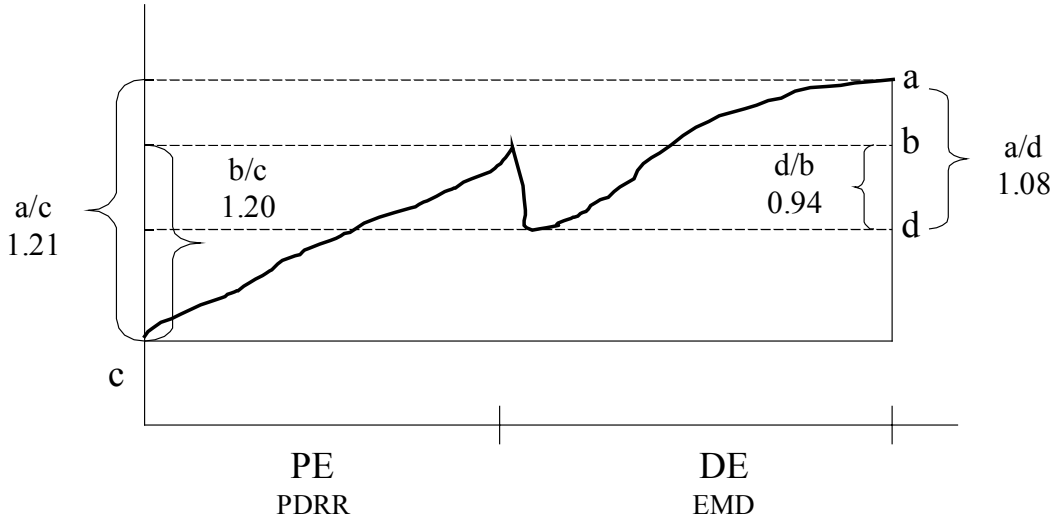
DoD Programs with PE and DE:

There were 25 DoD programs with both PE and DE estimates (i.e., programs with PDRR and EMD). Raw and dollar weighted means were calculated for both PE and DE. In addition, growth from the final PE estimate to the DE baseline estimate[1] was examined. Total growth was calculated by compounding, i.e. Total RDT&E = PE CGF * PE * DE CGF * DE CGF. Results are summarized in the table below.

RDT&E RAND 93 Cohorts - DoD

	Number	Total $	CGF Raw Mean	Raw Std. Dev	CV	CGF Dollar Weighted Mean
PE	25	62015	1.53	0.95	62%	1.20
PE to DE			1.000	0.49	49%	0.94
DE	25	69724	1.31	0.65	49%	1.08
Total RDT&E (compound)						1.21
Total RDT&E without discontinuity (compound)						1.30

A graphical depiction of the growth is shown below. The RDT&E estimate grows significantly during PE, then drops between the end of PE and the start of DE, and then continues to grow at a slower rate during DE. The discontinuity between the end of PE and the DE baseline is an artifact of quantity corrections, which is corrected with compounding.



NAVAIR Programs with PE and DE:
The same analysis was conducted for the five NAVAIR programs with PE and DE. The results are presented in the table below. The NAVAIR subset had a higher Total CGF (both with and without the discontinuity) than DoD, but the NAVAIR sample size is quite small.

RDT&E RAND 93 Cohorts - NAVAIR						
	Number	Total $	Raw Mean	Raw Std. Dev	CV	Dollar Weighted Mean
PE	5	7,999	1.75	0.63	36%	1.70
PE to DE	5	13582	0.86	0.49	57%	0.80
DE	5	10,897	1.32	0.49	37%	1.07
Total (compound)						1.46
Total RDT&E without discontinuity (compound)						1.81

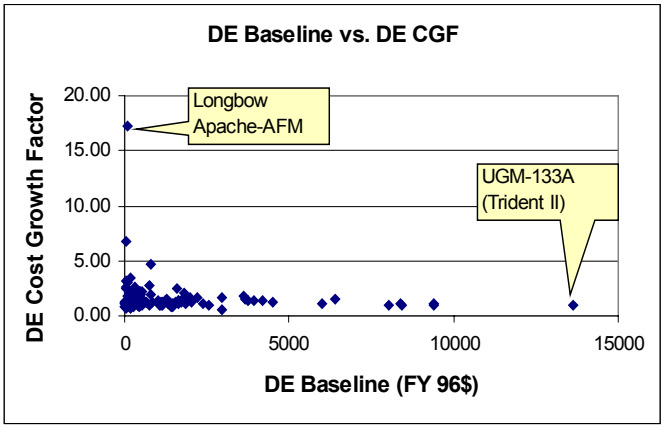
Ship Programs with PE and DE:
There were only 2 Ship programs with PE and DE (DDG 51 and SEALIFT). The results are presented in the table below. The Ship subset had a much higher Total CGF (both with and without the discontinuity) than DoD, but the Ship sample size is too small to draw conclusions.

[1] This shift of phases will be called a discontinuity in this paper.

RDT&E RAND 93 Cohorts - Ship						
	Number	Total $	CGF Raw Mean	Raw Std. Dev	CV	CGF Dollar Weighted Mean
PE	2	1287	1.58	0.89	56%	2.17
PE to DE			0.743	0.36	49%	0.49
DE	2	1376	1.51	0.77	51%	2.03
Total RDT&E (compound)						2.17
Total RDT&E without discontinuity (compound)						4.40

DoD Programs with DE only:

There were 140 DoD programs with DE only (programs that did not have a PDRR). Longbow Apache-AFM was removed from the analysis as an outlier as the data is believed to be a combination of a SAR on the airframe and one on the electronics, preventing a valid comparison. The scatter plot below shows the DE baseline and the DE CGF. Longbow Apache-AFM had a very high CGF of 17.30.



The results for RDT&E DoD programs with DE only are displayed below. Note that the DE dollar weighted average here (1.25) was higher than the dollar weighed average for DoD programs that did have a PDRR (1.08). However, the totals of the two groups are about the same (1.30 for programs with both PE and DE versus 1.25 for programs with DE only). So, it appears that programs occur most of their total RDT&E growth in the first phase of the program

RDT&E RAND 93 DE Only - DoD						
	Number	Total $	Raw Mean	Raw Std. Dev	CV	Dollar Weighted Mean
PE						
DE	140	172771	1.43	0.74	51.8%	1.25

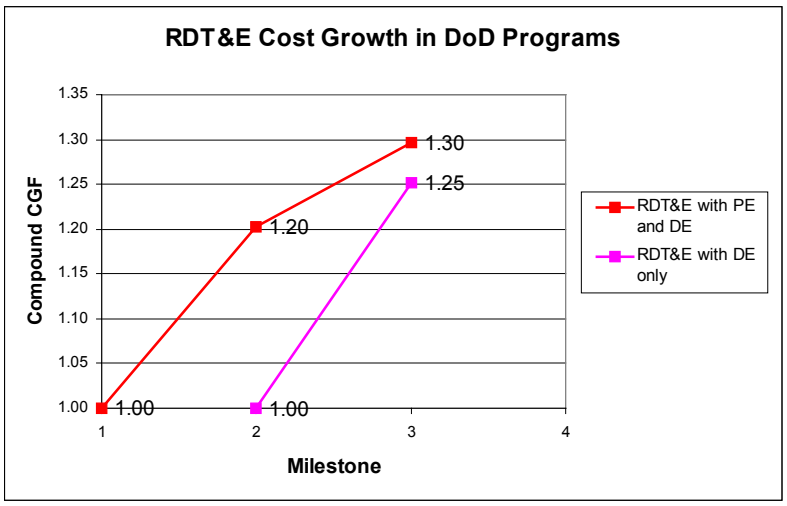
NAVAIR Programs with DE only:
There were 20 NAVAIR programs with DE only (no PDRR). The results are shown below. NAVAIR programs with DE only experienced lower total growth (1.37) than did NAVIAR programs with both PE and DE (1.81, but very small sample size). NAVAIR programs with DE only had a somewhat higher CGF (1.37) than DoD programs with DE only (1.25).

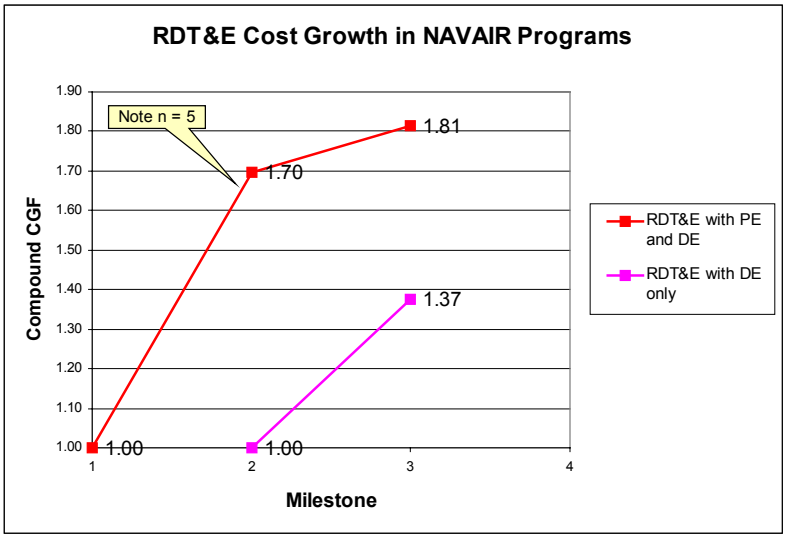
RDT&E RAND 93 DE Only - NAVAIR						
	Number	Total $	Raw Mean	Raw Std Dev	CV	Dollar Weighted Mean
PE						
DE	20	19877	1.72	1.28	74%	1.37

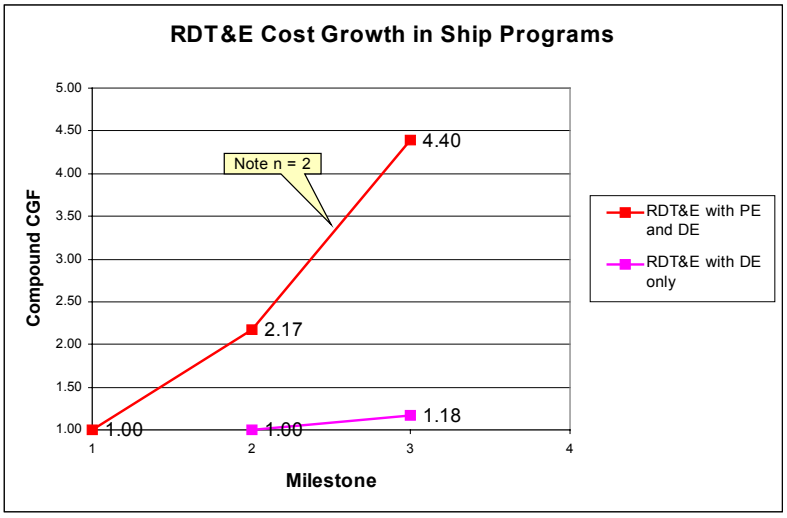
Ship Programs with DE only:
The RAND 93 data set contained 10 Ship programs with DE only. The results are summarized below. Ship programs with DE only had a somewhat lower dollar weighted mean (1.18) than did DoD programs with DE only (1.25).

RDT&E RAND 93 DE Only - Ship						
	Number	Total $	Raw Mean	Raw Std. Dev	CV	Dollar Weighted Mean
PE						
DE	10	6978	1.12	0.16	14.0%	1.18

RDT&E Graphs:
The following charts display RDT&E cost growth for the various commodities. Note that for all three groups (DoD, NAVAIR, and Ship), programs with a PDRR grew more in PE than in DE. For DoD programs, total RDT&E growth was approximately the same regardless of whether or not a PDRR occurred. This conclusion is not readily apparent for NAVAIR and Ship subsets, but both had very few data points for programs with PE and DE, limiting the ability to draw conclusions.


RDT&E Cost Growth in DoD Programs
Compound CGF
1.35
1.30
1.25
1.20
1.15
1.10
1.05
1.00
1.30
1.25
1.20
1.00
1.00
1
2
3
4
Milestone
RDT&E with PE and DE
RDT&E with DE only


RDT&E Cost Growth in NAVAIR Programs
Compound CGF
1.90
1.80
1.70
1.60
1.50
1.40
1.30
1.20
1.10
1.00
Note n = 5
1.81
1.70
1.37
1.00
1.00
1
2
3
4
Milestone
RDT&E with PE and DE
RDT&E with DE only


RDT&E Cost Growth in Ship Programs
Compound CGF
5.00
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
Note n = 2
4.40
2.17
1.18
1.00
1.00
1
2
3
4
Milestone
RDT&E with PE and DE
RDT&E with DE only

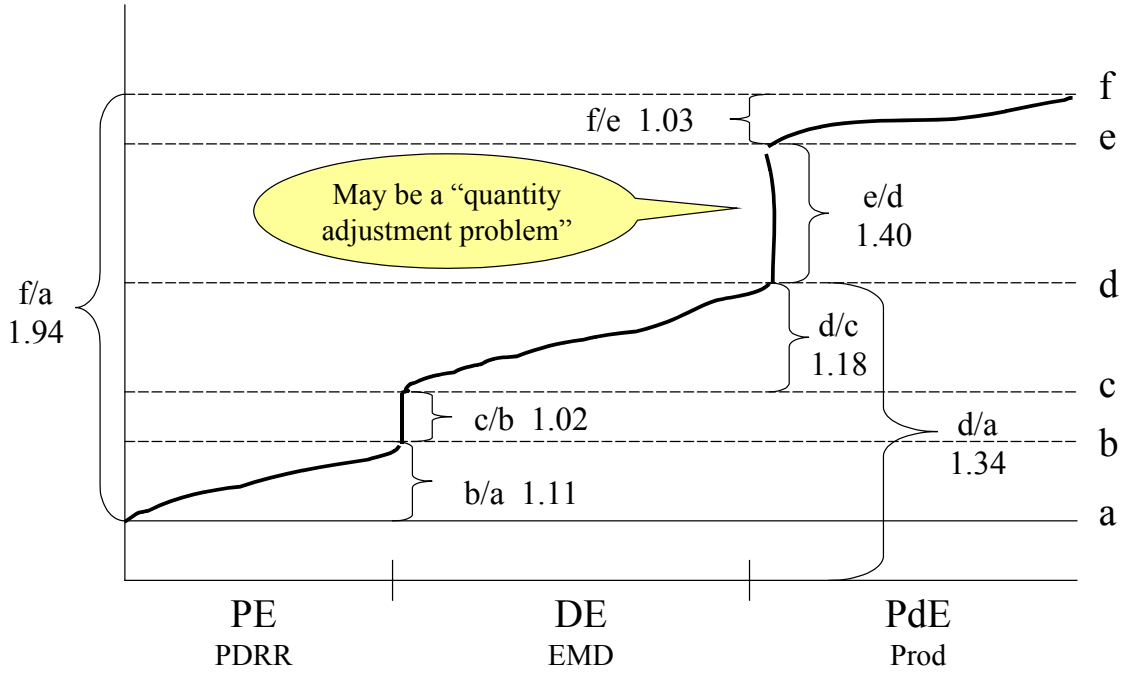
Procurement Results:

DoD Programs with PE, DE, and PdE

There were only 6 DoD programs with all three phases. Mean CGFs were calculated for PE, DE, and PdE as well as for the discontinuities from PE to DE and from DE to PdE. Totals were again calculated by compounding for both with and without discontinuities. For Procurement, totals were found for Procurement growth during RDT&E (PE and DE), during Procurement (PdE), and during Acquisition (PE, DE, and PdE).

Proc RAND 93 Cohorts - DoD						
	Total $	Number	Raw Mean	Raw Std. Dev	CV	Dollar Weighted Mean
PE	27995	6	1.11	0.49	44%	1.11
PE to DE	31054	6	0.91	0.44	48%	1.02
DE	31769	6	1.36	0.40	29%	1.18
DE to PdE	37416	6	1.45	0.70	48%	1.40
PdE	52547	6	1.06	0.23	22%	1.03
Total during R&D (compound)						1.34
Total during R&D without discontinuity (compound)						1.31
Total during Proc						1.03
Total Proc			1.74	0.97	56%	1.94
Total Proc without discontinuities (compound)						1.35

The graph below shows the Procurement cost growth over the three phases. It is believed that the discontinuity growth may be a problem of quantity adjustment. DoD programs with PE, DE, and PdE had a dollar weighted average CGF of 1.35 for Procurement during Acquisition without discontinuities.



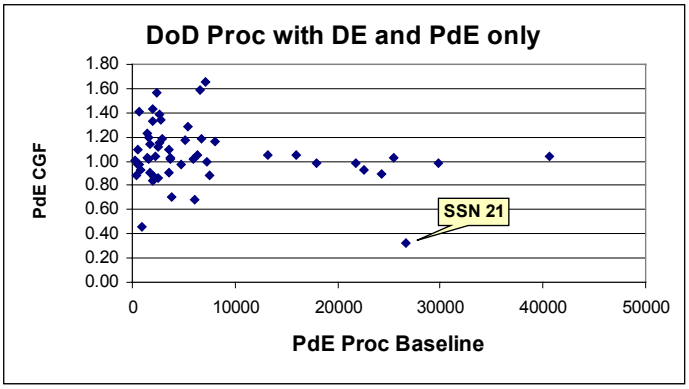
NAVAIR Programs with PE, DE, and PdE:
There was only one NAVAIR program with all three phases (C/MH-53). This program had a Procurement CGF during Acquisition without discontinuities of 1.64, as compared to the DoD average of 1.35.

Proc RAND 93 Cohorts - NAVAIR						
	Total $	Number	Raw Mean	Raw Std. Dev	CV	Dollar Weighted Mean
PE	1257	1	1.28			1.28
PE to DE	1605	1	0.78			0.78
DE	1257	1	1.28			1.28
DE to PdE	1605	1	0.75			0.75
PdE	1196	1	1.01			1.01
Total during R&D (compound)						1.28
Total during R&D without discontinuity (compound)						1.63
Total during Proc						1.01
Total Proc			0.96			0.96
Total Proc without discontinuities (compound)						1.64

Ship Programs with PE, DE, and PdE:
There was also only one Ship program with PE, DE, and PdE (DDG-51). This program had a Procurement CGF during Acquisition without discontinuities of 0.89, as compared to the DoD average of 1.35.

Proc RAND 93 Cohorts - Ship						
	Total $	Number	Raw Mean	Raw Std. Dev	CV	Dollar Weighted Mean
PE	9878	1	0.97			0.97
PE to DE	9552	1	1.74			1.74
DE	16622	1	0.98			0.98
DE to PdE	16224	1	1.33			1.33
PdE	21625	1	0.95			0.95
Total during R&D (compound)						1.64
Total during R&D without discontinuity (compound)						0.94
Total during Proc						0.95
Total Proc			2.07			2.07
Total Proc without discontinuities (compound)						0.89

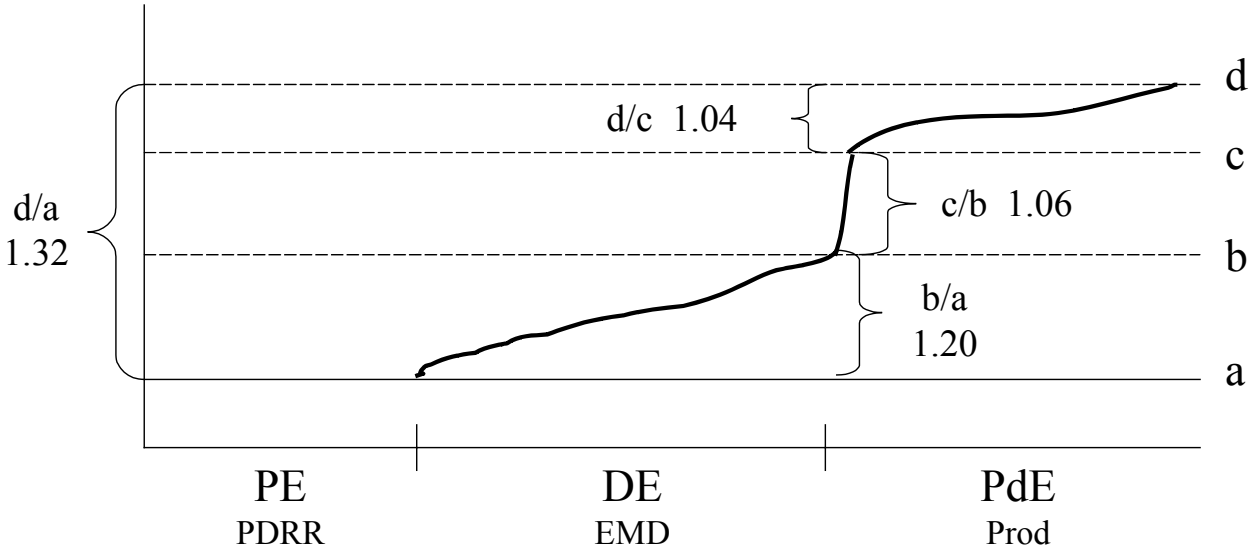
DoD Programs with DE and PdE only:
The RAND 93 data set contained 53 DoD programs with DE and PdE only. SSN 21 was removed for this analysis as it was greatly influenced by political forces and suffered massive restructure at the fall of the Berlin Wall. The scatter plot below shows the PdE Baseline and the PdE CGF for Procurement. SSN 21 had a Procurement CGF of only 0.33, which skewed the overall results.



The Procurement results for DoD programs with DE and PdE only are shown below. These programs had an average Procurement CGF without discontinuities for Acquisition of 1.25, as compared to 1.35[2] for DoD programs with all three phases.

Proc RAND 93 DE and PdE Only - DoD						
Proc	Total $	Number	Raw Mean	Raw Std. Dev	CV	Dollar Weighted Mean
PE						
PE to DE						
DE	277352	53	1.32	0.35	27%	1.20
DE to PdE	332494	53	1.09	0.50	46%	1.06
PdE	351799	53	1.07	0.22	21%	1.04
Total during R&D (compound)						1.20
Total during Proc						1.04
Total Proc			1.49	0.82	55%	1.32
Total Proc without discontinuities (compound)						1.25

The following graph shows the cost growth in each phase.

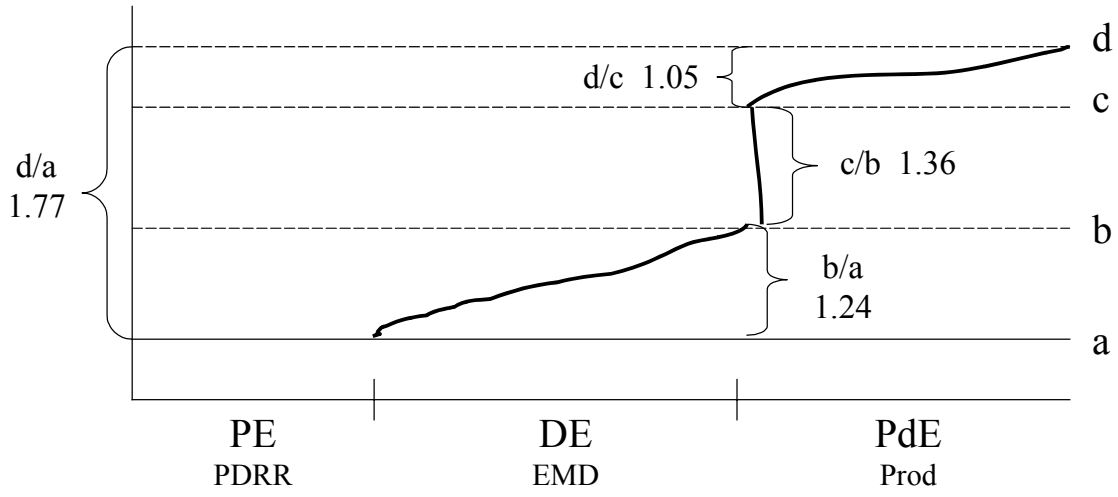


[2] Note that the sample size was small (n=6) for DoD programs with PE, DE, and PdE

NAVAIR Programs with DE and PdE only:
There were 10 NAVAIR programs with DE and PdE only. The table below summarizes the results for this category. NAVAIR programs with only DE and PdE experienced lower growth in Procurement during Acquisition without discontinuities (1.30) than did NAVAIR programs with all three phases (1.64, but only one data point). This subgroup had about the same growth as DoD programs with DE and PdE only (1.25).

Proc RAND 93 DE and PdE Only - NAVAIR						
	Total $	Number	Raw Mean	Raw Std. Dev	CV	Dollar Weighted Mean
PE						
PE to DE						
DE	54206	10	1.35	0.48	36%	1.24
DE to PdE	66985	10	1.21	0.44	36%	1.36
PdE	91275	10	1.12	0.22	20%	1.05
Total during R&D (compound)						1.24
Total during Proc						1.05
Total Proc			1.92	1.47	77%	1.77
Total Proc without discontinuities (compound)						1.30

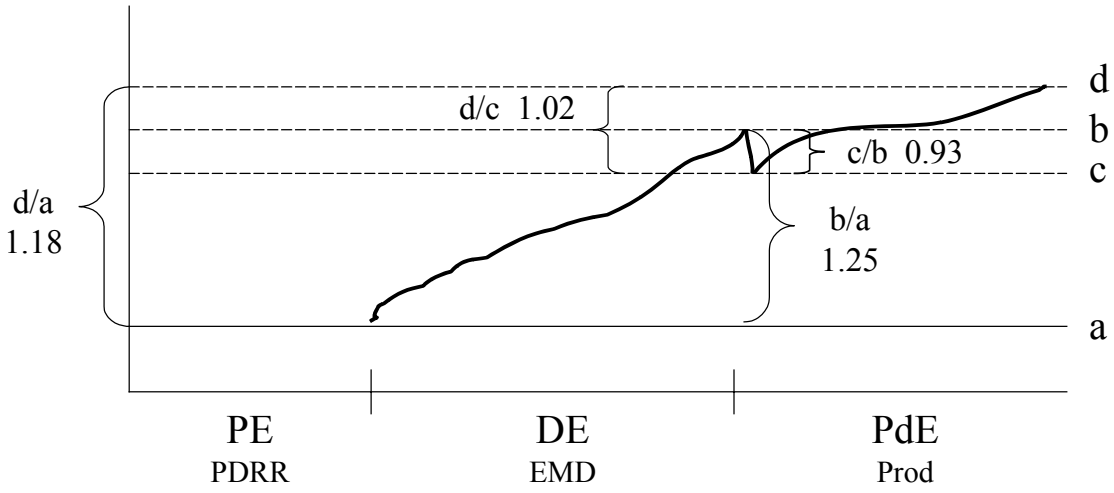
The graph below shows the growth during the phases.



Ship Programs with DE and PdE only:
There were 4 Ship programs with DE and PdE only. SSN 21 was again removed for this analysis. The table below summarizes the results for this group. Ship programs with only DE and PdE experienced higher growth in Procurement during Acquisition without discontinuities (1.28) than did Ship programs with all three phases (0.89, but only one data point). This subgroup had about the same growth as DoD programs with DE and PdE only (1.25).

Proc RAND 93 DE and PdE Only - Ship						
Proc	Total $	Number	Raw Mean	Raw Std. Dev	CV	Dollar Weighted Mean
PE						
PE to DE						
DE	34403	4	1.43	0.44	31%	1.25
DE to PdE	43069	4	0.79	0.41	52%	0.93
PdE	39932	4	1.12	0.20	18%	1.02
Total during R&D (compound)						1.25
Total during Proc						1.02
Total Proc			1.39	0.81	58%	1.18
Total Proc without discontinuities (compound)						1.28

The graph below displays the cost growth during the phases.



DoD Programs with DE only:
The RAND 93 data set contained 94 DoD programs with DE only. The results are displayed below. DoD programs with DE only experienced lower Procurement growth during Acquisition without discontinuities (1.15) than did DoD programs with PE, DE, and PdE (1.35, but small n) and DoD programs with DE and PdE (1.25).

Proc RAND 93 DE only - DoD						
Proc	Total $	Number	Raw Mean	Raw Std. Dev	CV	Dollar Weighted Mean
PE						
PE to DE						
DE	502470	94	1.32	0.88	66%	1.15
DE to PdE						
PdE						
Total during R&D (compound)						1.15
Total during Proc						
Total Proc						
Total Proc without discontinuities (compound)						

NAVAIR Programs with DE only:
The RAND 93 data set contained 12 NAVAIR programs with DE only. The results are displayed below. NAVAIR programs with DE only experienced lower Procurement growth during Acquisition without discontinuities (1.04) than did NAVAIR programs with PE, DE, and PdE (1.64, but n=1) and NAVAIR programs with DE and PdE (1.30).

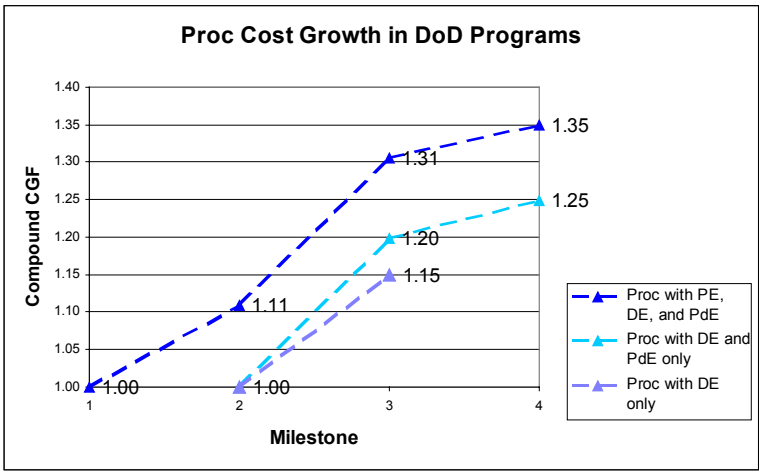
Proc RAND 93 DE only - NAVAIR						
Proc	Total $	Number	Raw Mean	Raw Std. Dev	CV	Dollar Weighted Mean
PE						
PE to DE						
DE	122430	12	1.15	0.31	27%	1.04
DE to PdE						
PdE						
Total during R&D (compound)						1.04
Total during Proc						
Total Proc						
Total Proc without discontinuities (compound)						

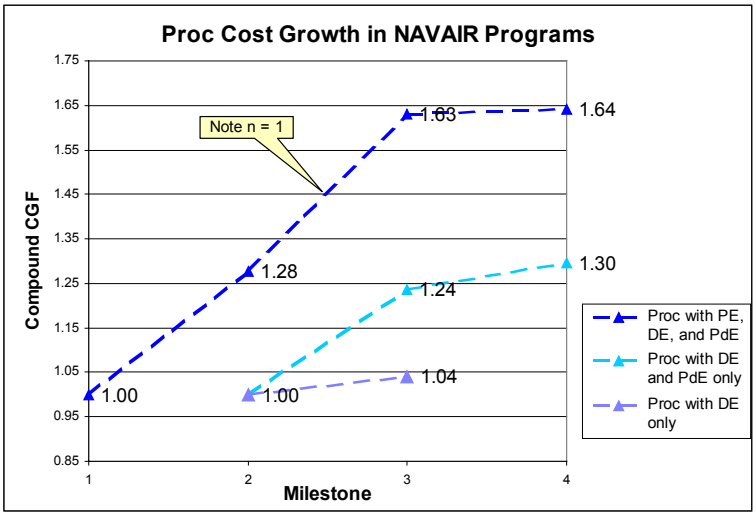
Ship Programs with DE only:
The RAND 93 data set contained 11 Ship programs with DE only. The results are displayed below. Ship programs with DE only experienced higher Procurement growth during Acquisition without discontinuities (1.03) than did Ship programs with PE, DE, and PdE (0.89, but n=1) and NAVAIR programs with DE and PdE (1.28).

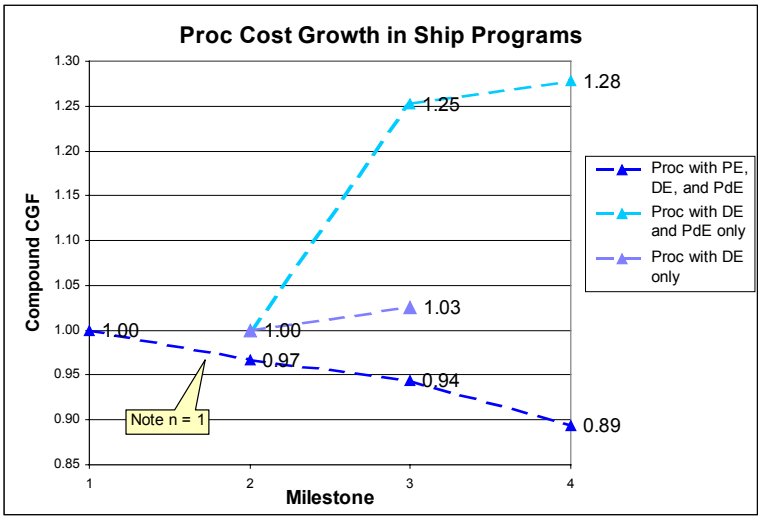
Proc RAND 93 DE Only - Ship						
Proc	Total $	Number	Raw Mean	Raw Std. Dev	CV	Dollar Weighted Mean
PE						
PE to DE						
DE	89554	11	1.04	0.18	17%	1.03
DE to PdE						
PdE						
Total during R&D (compound)						1.03
Total during Proc						
Total Proc						
Total Proc without discontinuities (compound)						

Procurement Graphs:

The following charts give a visual image of Procurement cost growth for the various commodities. For DoD, the more phases a program had, the more cost growth occurred. Programs tended to have about the same DE Procurement growth regardless of whether the program had a PDRR. This pattern is less clear for NAVAIR and Ships, as they had much smaller data sets.

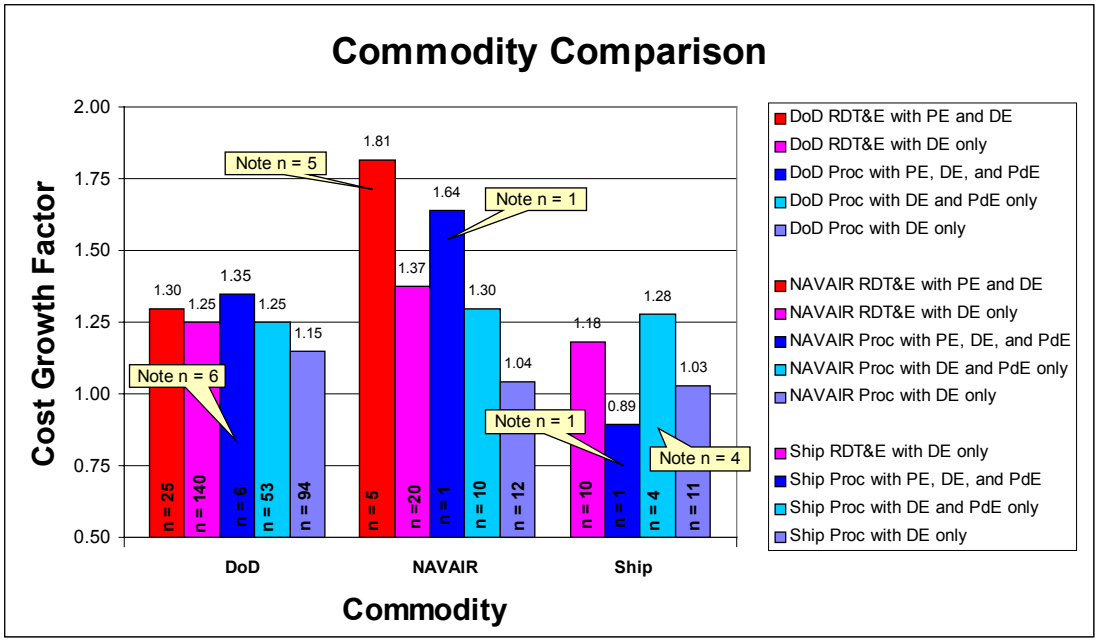

Proc Cost Growth in DoD Programs
Compound CGF
1.40
1.35
1.30
1.25
1.20
1.15
1.10
1.05
1.00
1
2
3
4
Milestone
1.35
1.31
1.25
1.20
1.15
1.11
1.00
1.00
Proc with PE, DE, and PdE
Proc with DE and PdE only
Proc with DE only


Proc Cost Growth in NAVAIR Programs
Compound CGF
1.75
1.65
1.55
1.45
1.35
1.25
1.15
1.05
0.95
0.85
1
2
3
4
Milestone
Note n = 1
1.64
1.63
1.30
1.28
1.24
1.04
1.00
1.00
Proc with PE, DE, and PdE
Proc with DE and PdE only
Proc with DE only


Proc Cost Growth in Ship Programs
Compound CGF
1.30
1.25
1.20
1.15
1.10
1.05
1.00
0.95
0.90
0.85
1
2
3
4
Milestone
1.28
1.25
1.03
1.00
1.00
0.97
0.94
0.89
Note n = 1
Proc with PE, DE, and PdE
Proc with DE and PdE only
Proc with DE only

Commodity Comparison:

The following chart displays the dollar weighted mean CGFs for each of the cohort groups. Note that the overall pattern between cohort groups is about the same across commodities. In RDT&E, programs with both PE and DE had somewhat higher total growth than programs with DE only. For Procurement, programs with all three phases generally had the highest growth, followed by programs with programs with both DE and PdE, then by programs with DE only. Note that many of the cohort groups had very small sample sizes, limiting the ability to draw conclusions in those areas.

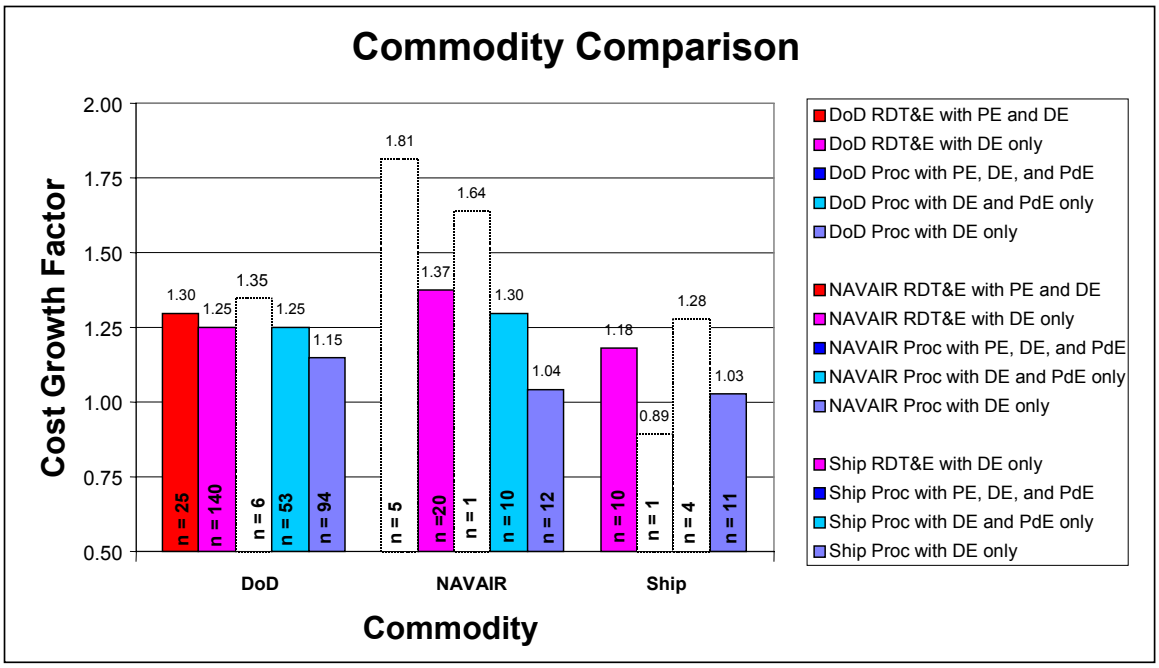


The problem of small sample sizes gives rise to the question of what is a sufficiently large n. In general, n should be large, but how large is a function of tolerance for error. In the analysis conducted above, the CVs appear to range around 30%. The table below shows the 67% confidence intervals for various values of n, assuming a 30% CV3. So, a sample size of approximately 25 would be required to make judgments within about 5%. When using small sample sizes, it is necessary to be aware of the error range.

N	+/-
4	15%
9	10%
16	8%
25	6%
36	5%

The chart below shows the results by commodity for a sufficient n only (n = 10).

[3] The confidence interval is calculated as CV/sqrt(n), based on the confidence interval of a t-test.



Conclusions:

It appears that RDT&E cost growth is discovered early. DoD programs with both PDRR and EMD tend to have RDT&E growth of about 20% in PDRR and 10% in EMD (total of 30%). DoD programs without PDRR seem to have about a 25% RDT&E growth during EMD. Most Procurement growth is discovered fairly early, during RDT&E. DoD Procurement growth was about 10% in PDRR, 20% in EMD, and 5% in Production (total of 35%) for programs with all three phases. DoD programs without PDRR had 20% growth during EMD and 5% during Production. The same basic pattern was present for both NAVAIR and Ship.

Programs with PDRR tended to experience somewhat more growth than those without. Also, it appears that cost growth is discovered about one phase in advance. So, most growth occurs during PDRR for RDT&E and during EMD for Procurement. This is a significant finding, and bears further discussion. The reasons for the effect are, of course, not known, but it is tempting to think that it is because the focus of effort (the "time horizon") is only about one phase ahead.

The following table summarizes NAVAIR and Ship growth as compared to the baseline of DoD. NAVAIR programs had higher RDT&E growth than DoD, but lower Procurement growth. Ship growth was lower than DoD for both RDT&E and Procurement.

	NAVAIR	Ship
RDT&E	+	-
Procurement	-	-

SAR 00 Data:

Principal Analyst – Cari L. Pullen

Purpose:

The SAR 00 data set contains DE SARs only, but is not "cohorted." Averaging programs that have a PDRR with those that do not may distort the result.

Approach:

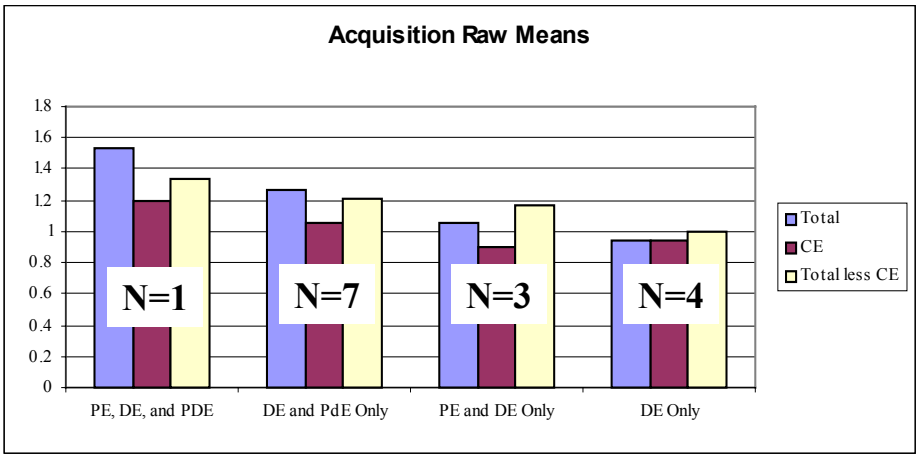
The RAND 93 data set was used to determine which phases the programs in the SAR 00 data set actually had (PDRR, EMD, and Production). Programs were grouped according to their phases as follows:

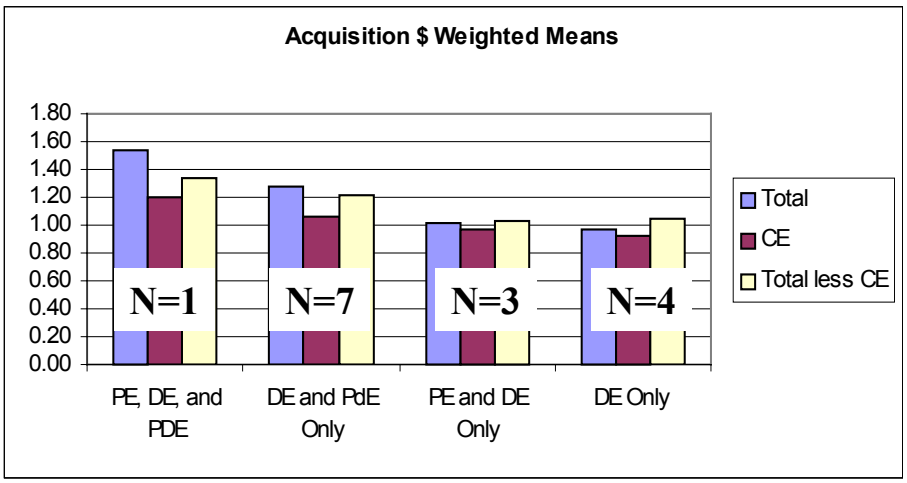
–Programs with PE, DE, and PdE
–Programs with PE and DE only
–Programs with DE and PdE only
–Programs with DE only

CGFs were averaged and compared for each cohorted group. In other words, the data has only DE, but was grouped as to whether there was a PDRR.

Acquisition Results:

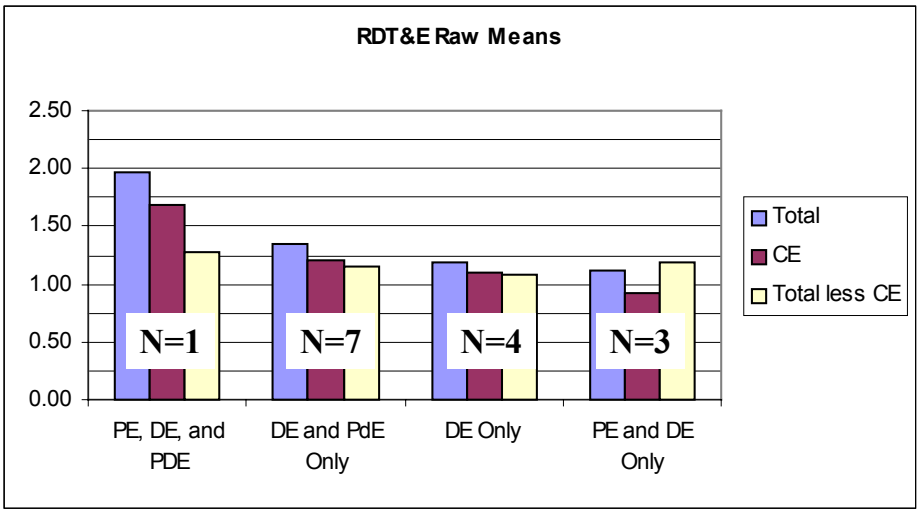
These results differ from the RAND 93 results where DE is lowest when PE is present. The DE CGF is lowest when there is a DE only, still quite low with 2 phases, and highest when all 3 phases are present.





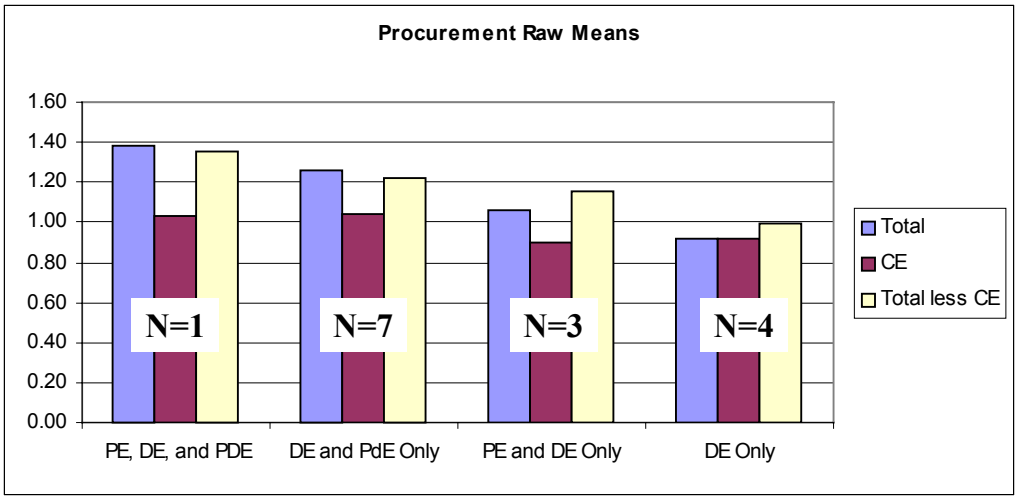
RDT&E Results:

These results are more consistent with RAND 93 results where DE is lowest when PE is present. Here, the DE CGF is highest when all 3 phases are present.



Procurement Results:

These results differ from the RAND 93 results where DE is lowest when PE is present. The DE CGF is lowest when there is a DE only, still quite low with 2 phases, and highest when all 3 phases are present.



Conclusions:

SAR 00 cohorts coincide with RAND 93 cohorts *only* in RDT&E and differ in Procurement and Acquisition, but the differences are small in size. The sample size for the SAR 00 cohorted groups is too small to draw meaningful conclusions, therefore cohort conclusions must come from RAND 93 DoD data.

Commodity Comparisons

Principal Analyst – Jessica R. Summerville

Purpose:

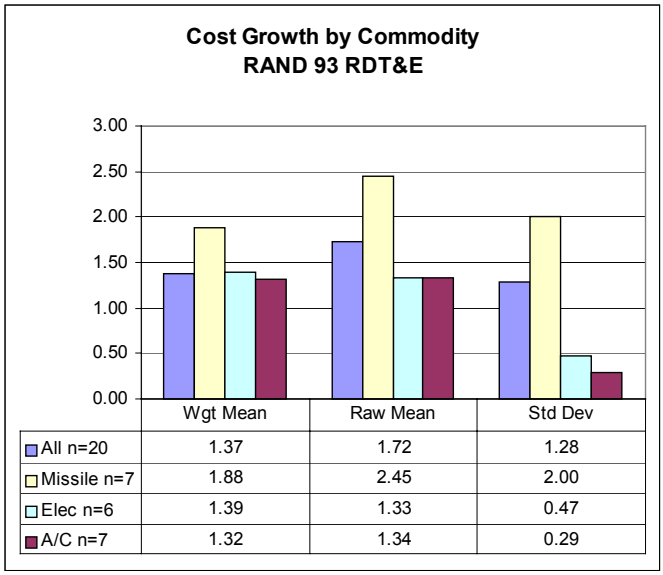
To explore and compare the cost growth in different types of commodities.

Approach:

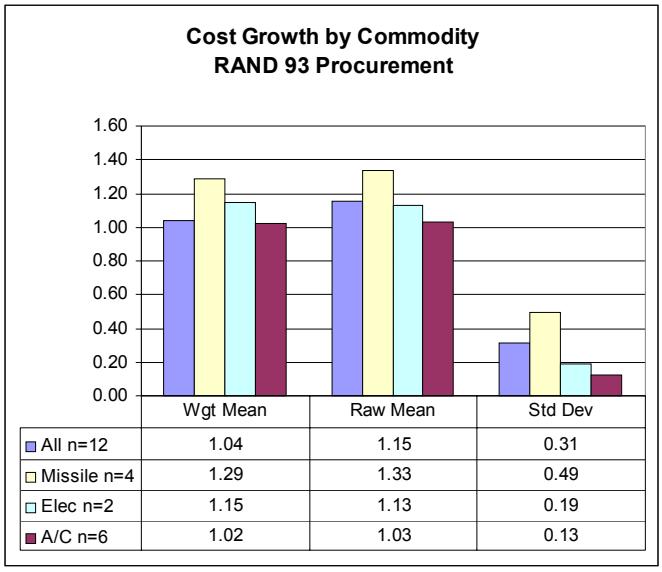
Analysis was conducted for RAND 93, SAR 00, and Contract data. For each data set, the programs were divided into commodity groups. Both RAND 93 and SAR 00 contained missiles, aircraft, and electronics. The Contract data was divided into EW, radar, radio, and targets. Descriptive analysis was performed for each group, and the results were compared. Inferential analysis was not useful do to the small sample sizes.

RAND 93 Results:

Analysis was performed for both RDT&E and Procurement using NAVAIR programs with DE only[4], with the results shown below. Missiles incurred the most growth for both RDT&E and Procurement. In RDT&E, Aircraft and Electronics experienced about the same growth. In Procurement, Electronics incurred growth only somewhat below that of Missiles, but Aircraft had almost no growth.



	Wgt Mean	Raw Mean	Std Dev
All n=20	1.37	1.72	1.28
Missile n=7	1.88	2.45	2.00
Elec n=6	1.39	1.33	0.47
A/C n=7	1.32	1.34	0.29



	Wgt Mean	Raw Mean	Std Dev
All n=12	1.04	1.15	0.31
Missile n=4	1.29	1.33	0.49
Elec n=2	1.15	1.13	0.19
A/C n=6	1.02	1.03	0.13

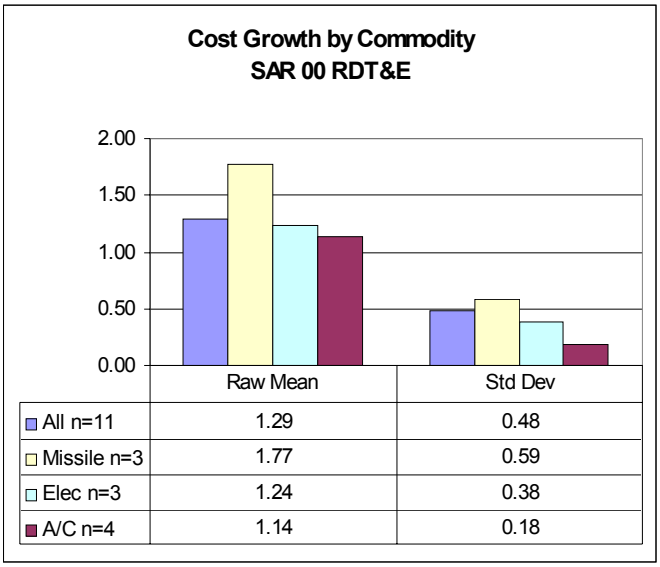
SAR 00 Results:

The SAR 00 data set consists only of DE SARs, but for this analysis, programs that previously had PEs were removed[5]. Weighted averages could not be calculated for RDT&E and Procurement as dollars were not broken out by appropriation. The results

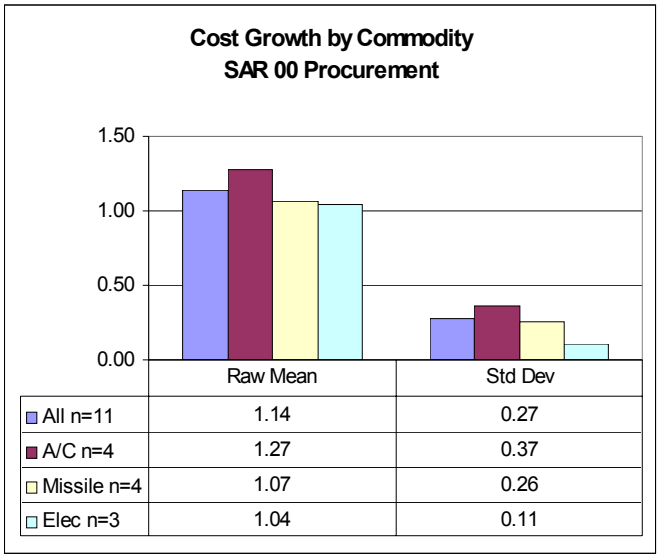
[4] DE only was the only RAND 93 cohort group with enough data to break down by commodity.

[5] Cohort tracking results suggest that programs with PDRR should not be combined with those without a planning phase.

are shown below. In RDT&E, Missiles incurred the most growth, followed by Electronics and then Aircraft. This finding is similar to the RAND 93 results. For Procurement, Aircraft displayed the most growth, while Missiles and Electronics displayed very little growth. This result is opposite of the RAND 93 outcome.



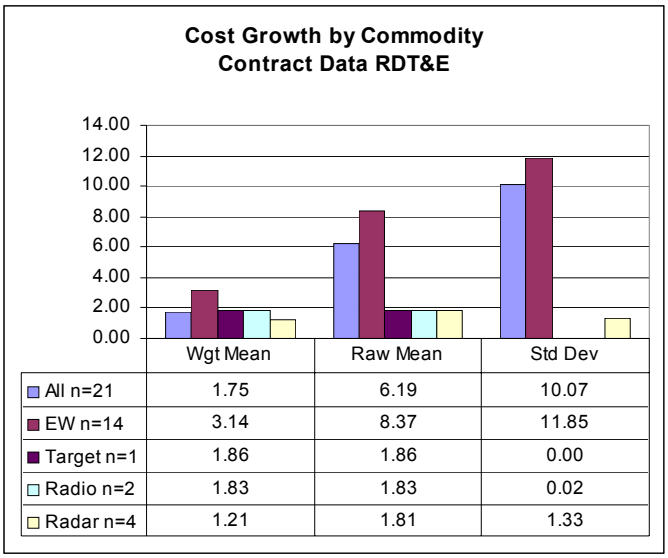
	Raw Mean	Std Dev
All n=11	1.29	0.48
Missile n=3	1.77	0.59
Elec n=3	1.24	0.38
A/C n=4	1.14	0.18



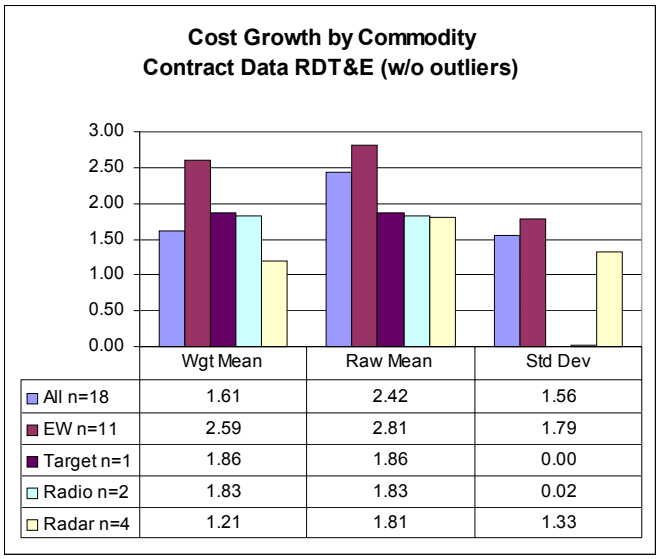
	Raw Mean	Std Dev
All n=11	1.14	0.27
A/C n=4	1.27	0.37
Missile n=4	1.07	0.26
Elec n=3	1.04	0.11

Contract Data Results:

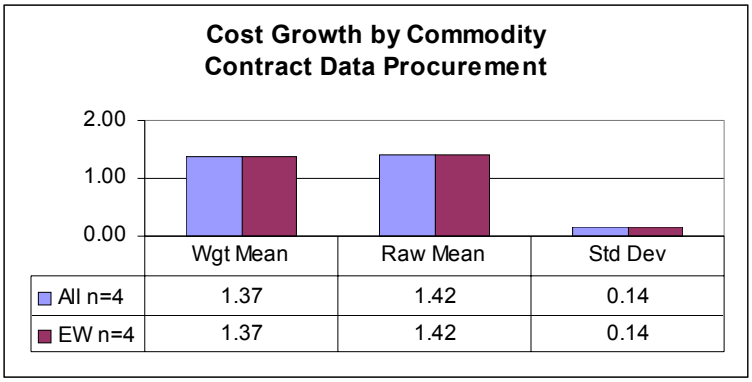
The mix of PDRR and EMD data is unknown for the Contract data set. Analysis is shown both with and without three potential outliers (each having growth of over 20 times the original estimate). The results of the analysis are shown below. In RDT&E, EW programs incurred the most growth and radars the least. For Procurement, EW was the only commodity available.



	Wgt Mean	Raw Mean	Std Dev
All n=21	1.75	6.19	10.07
EW n=14	3.14	8.37	11.85
Target n=1	1.86	1.86	0.00
Radio n=2	1.83	1.83	0.02
Radar n=4	1.21	1.81	1.33

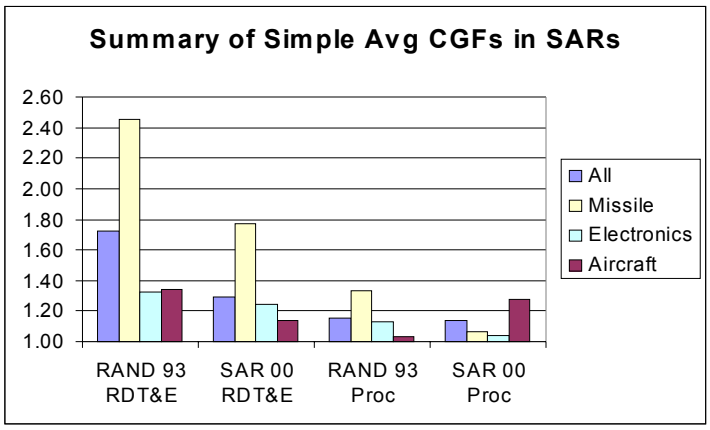


	Wgt Mean	Raw Mean	Std Dev
All n=18	1.61	2.42	1.56
EW n=11	2.59	2.81	1.79
Target n=1	1.86	1.86	0.00
Radio n=2	1.83	1.83	0.02
Radar n=4	1.21	1.81	1.33

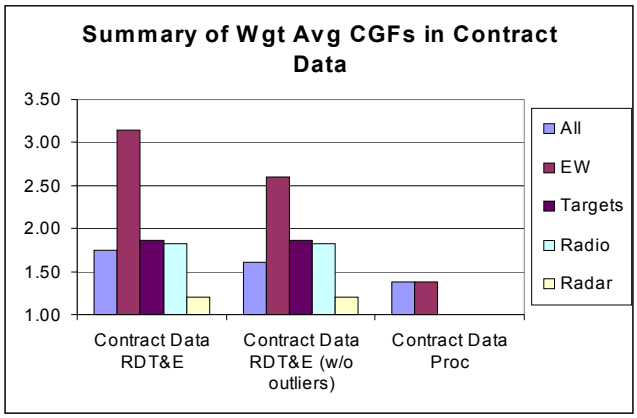


Conclusions:

The chart below summarizes the simple average CGFs from the RAND 93 and SAR 00 commodity analysis[6]. For RDT&E, cost growth in Missiles is much higher than other commodities in both data sets. Procurement results are conflicting between RAND 93 and SAR 00.



The next chart shows the weighted average CGFs in the Contract data. Here, EW programs showed the most growth.



[6] Weighted averages were not available for SAR 00.

Effects of Program Size on Cost Growth

Principal Analyst – Megan E. Dameron

Purpose:

To determine the effect of program size on cost growth for both RDT&E and Procurement. It was hypothesized that small programs experience more cost growth than larger programs.

Data:

The initial analysis was performed using DoD programs with DE only from the RAND 93 dataset. The same analysis was then repeated with the SAR 00 and Contract data. Various shortcomings are associated with each of the three datasets. For example, the RAND 93 data cannot be broken out between Cost Estimating (CE) risk and Schedule/Technical (S/T) risk. The SAR 00 data cannot be dollar weighted for RDT&E and Procurement because total dollars are not broken out by appropriation. The Contract data analysis focuses only on RDT&E as there are only four data points for Procurement.

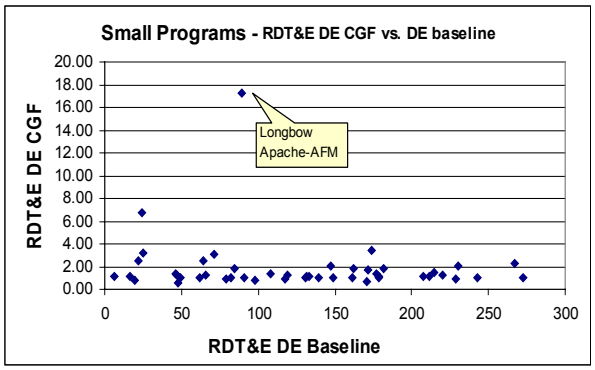
Approach:

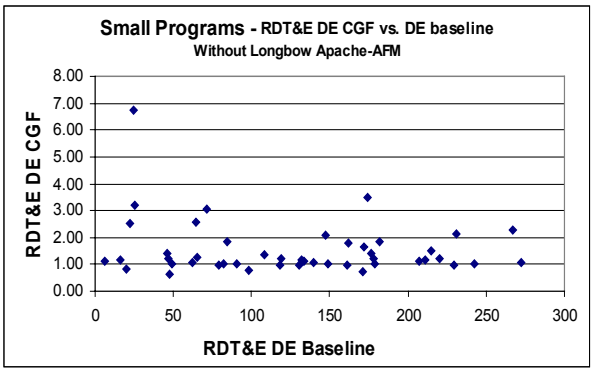
The data was divided into groups based on program size, as measured by the initial baseline estimate. Three groups were made for the RAND 93 dataset, but only two groups for SAR 00 and Contract data. Simple and weighted averages of Cost Growth Factors (CGFs) were calculated and compared. T-tests were performed to test for statistically significant differences between the average CGFs for each of the size groups. For the RAND 93 data, a Kolmorgov-Smirnov test was conducted to determine if the size groupings had identical distributions.

Analysis – RAND 93 Data:

RDT&E:

The RAND 93 dataset contained 140 observations for RDT&E programs with DE only. The data was divided into three size groups containing equal numbers of programs based on the initial DE RDT&E baseline. The Longbow Apache-AFM was removed from the small group as an outlier, as demonstrated in the below scatter plot. It is believed that the Longbow Apache may be a mixture of a SAR on airframe and one on electronics, preventing an "apples-to-apples" comparison.





The following table summarizes the descriptive statistics for RDT&E. Note that the weighted averages are all lower than the simple averages, a result consistent with earlier studies. This supports the idea that smaller programs have higher growth. The average RDT&E CGF appears to be decreasing as the size of the program increases. This effect will be tested for statistical significance.

	Number	Total Dev $	Average Dev $	Raw Mean	Raw Std. Dev	CV	$ Weighted Mean
Small Programs	46	5827	127	1.54	1.03	67%	1.43
Medium Programs	46	20740	451	1.41	0.48	34%	1.42
Large Programs	47	146115	3109	1.34	0.60	45%	1.22

T-tests assuming unequal variances[7] were performed to test for differences in the means between any two of the three size groupings. As displayed in the table below, there is no statistical difference between the means of the three groups.

t-Test Summary: Two-Sample Assuming Unequal Variances				
	t stat	critical value	p-value	Significant Difference?
Small and Medium	0.81	2.00	0.42	No
Medium and Large	0.55	1.99	0.58	No
Small and Large	1.13	1.99	0.26	No

An ANOVA test was also performed with the null hypothesis of equal means across the three size groups. Again, there is no statistical difference between the RDT&E means of the three groups. However, the trend in means does seem to indicate that smaller programs do have larger growth.

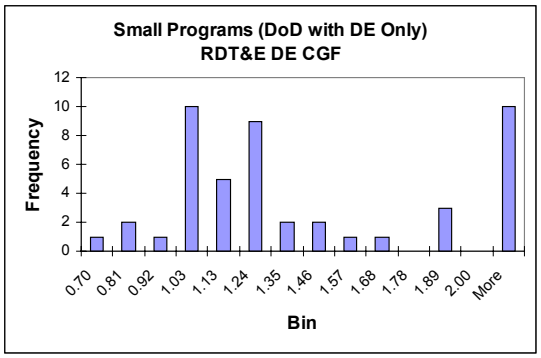
ANOVA

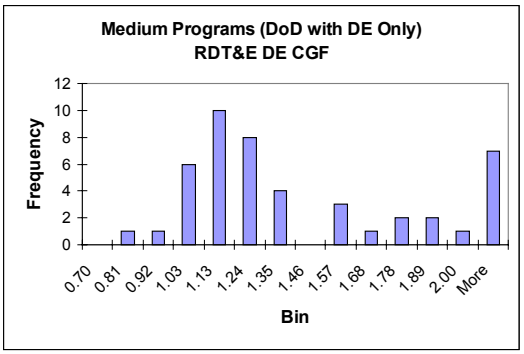
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.95	2	0.48	0.87	0.42	3.06
Within Groups	74.61	136	0.55			
Total	75.56	138				

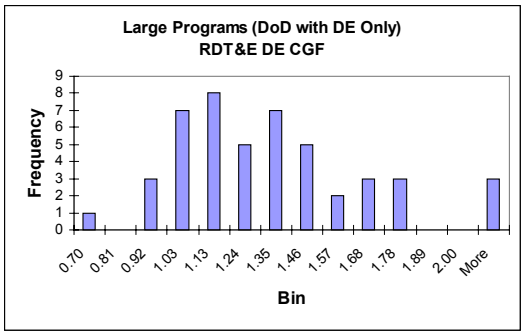
Next, analysis was conducted to determine if the three size groupings are identically distributed. If they are not identically distributed, then there must be some underlying difference between small, medium and large programs. Histograms of the CGFs for each of the three size groups were plotted, and a Kolmogorov-Smirnov (K-S) test for identical distributions was performed.

[7] T-tests assuming equal variances were also performed with the same results

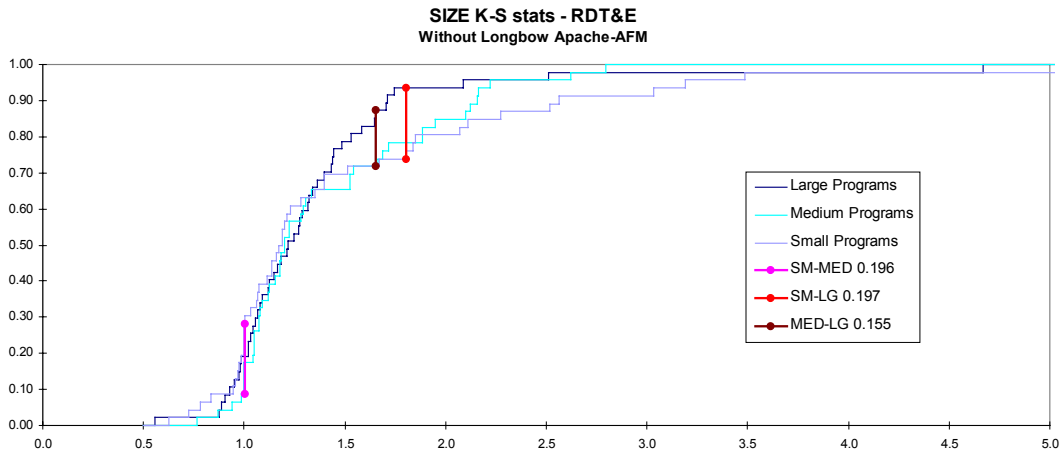
As demonstrated in the histograms below, the distributions look similar for all three size groups, but with more high-risk extrema in the smaller groups.







The K-S test statistic is the maximum distance between empirical Cumulative Density Functions (CDFs). The graph below shows the CDF for each size group, as well as the maximum distance between the curves.



The K-S test statistics are summarized below. As shown, the small, medium, and large program groups are all identically distributed. As visible in the histograms above, there is a pattern towards more high-risk extrema in smaller size groups than in larger size

groups, although not statistically significant. It may be that high-risk programs are terminated early if they are large (expensive), but are tolerated if they are small (relatively inexpensive).

	K-S Critical Value	K-S Test Statistic	Identical Distribution?
Small to Medium	0.283	0.196	Yes
Small to Large	0.282	0.197	Yes
Medium to Large	0.282	0.155	Yes

Procurement:
The same analysis performed for RDT&E was conducted for the Procurement data. There were 94 Procurement observations for DoD programs with DE only. This data was again divided into three groups of equal numbers based on the initial DE Procurement baseline.

The descriptive statistics for the Procurement size groups are summarized below. Again, the weighted averages are all smaller than the simple averages, suggesting that program size may affect cost growth. As found in RDT&E, the average Procurement CGF appears to be decreasing as the size of the program increases, i.e., smaller programs seem to have larger cost growth.

	Number	Total Dev $	Average Size	Raw Mean	Raw Std. Dev	CV	$ Weighted Mean
Small Programs	31	20206	652	1.61	1.36	84%	1.54
Medium Programs	31	69334	2237	1.24	0.56	45%	1.18
Large Programs	32	412930	12904	1.13	0.31	28%	1.13

The relationship described above was then tested for statistical significance using a t-test and an ANOVA test. T-tests assuming unequal variances[8] were performed to test for differences in the means between any two of the three groups. A summary is provided below. As for RDT&E, there was no statistical difference between the means of the three Procurement size groups.

t-Test Summary: Two-Sample Assuming Unequal Variances				
	t stat	critical value	p-value	Significant Difference?
Small and Medium	1.43	2.02	0.16	No
Medium and Large	0.95	2.01	0.34	No
Small and Large	1.94	2.03	0.06	No

[8] T-tests assuming equal variances were also performed with the same results.

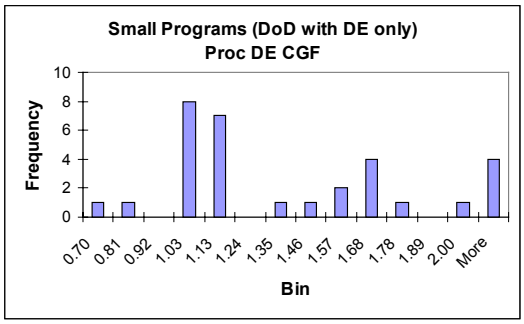
An ANOVA test was also performed with the null hypothesis of equal means across the three size groups. Again, there was no statistical difference between the means of the small, medium, and large Procurement groups. However, the trend in means does seem to indicate that smaller programs have larger cost growth.

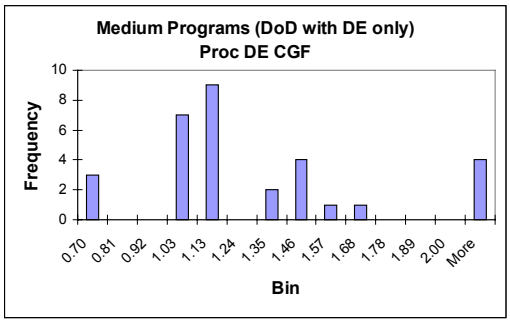
ANOVA

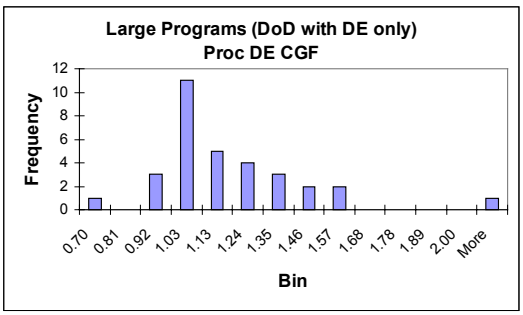
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	4.06	2	2.03	2.72	0.07	3.10
Within Groups	67.90	91	0.75			
Total	71.96	93				

Next, analysis was conducted to determine if the three Procurement size groupings are identically distributed. If they are not identically distributed, then there must be some underlying difference between small, medium and large programs. Histograms of the CGFs for each of the three size groups were plotted, and a Kolmogorov-Smirnov (K-S) test for identical distributions was performed.

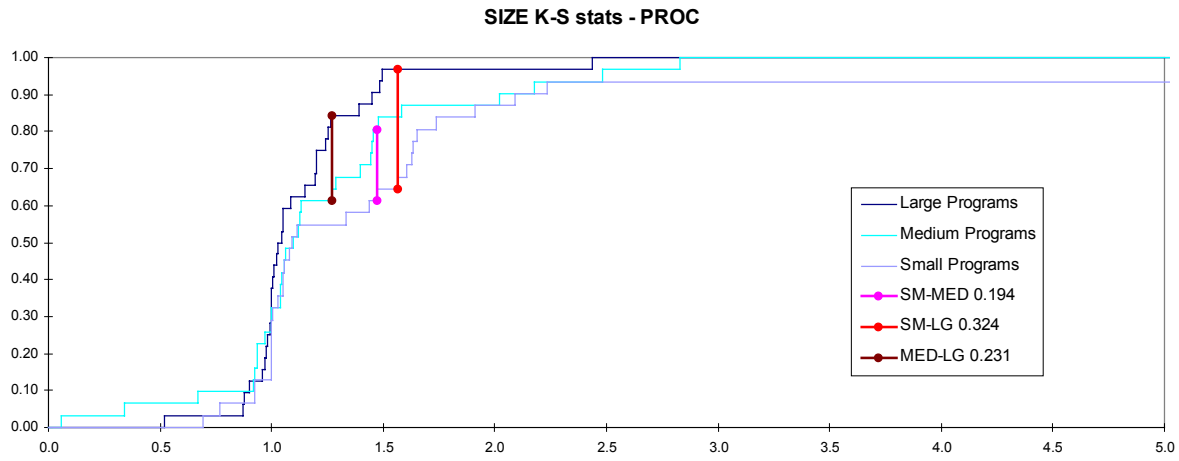
As shown below, the distributions are sparse, but they appear similar for all three size groupings, with more high risk extrema in the smaller groups.







A K-S test for identical distributions was conducted on the three Procurement size groups. The graph below shows the empirical CDFs and the maximum distances between curves used for the test.



The Procurement K-S test statistics are summarized below. As found for RDT&E, the small, medium, and large Procurement program groups are all identically distributed. As visible in the histograms above, there is again a pattern towards more high-risk extrema in smaller size groups than in larger size groups, although not statistically significant. It may be that high-risk programs are terminated early if they are large (expensive), but are tolerated if they are small (relatively inexpensive).

	K-S Critical Value	K-S Test Statistic	Identical Distribution?
Small to Medium	0.323	0.194	Yes
Small to Large	0.343	0.324	Yes
Medium to Large	0.343	0.231	Yes

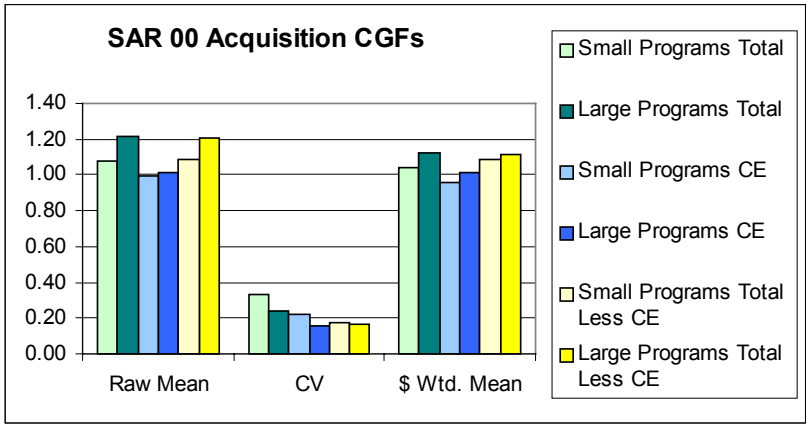
Analysis - SAR 00 Data:

The same analysis was performed using the SAR 00 dataset, which consists of DE estimates for 16 NAVAIR programs. The data was divided into two groups of equal numbers based on size, as measured by the initial DE baseline estimate. The SAR 00 dataset has CE and S/T risk broken out separately, allowing for new analysis. However, dollar weighted averages are not always available as RDT&E and Procurement dollars are not specified. In addition, the dataset is small, so it is difficult to draw conclusions.

Acquisition:
Descriptive statistics for Total Acquisition are displayed below. The dollar-weighted averages are again lower than the simple averages, suggesting that smaller programs have larger growth. However, in this data, large programs seem to have slightly higher growth, opposite of the RAND 93 data results. This relationship will be tested for statistical significance.

Acquisition CGFs					
	Number	Raw Mean	Raw Std. Dev	CV	$ Wtd. Mean
Small Programs Total	8	1.08	0.35	33%	1.04
Large Programs Total	8	1.22	0.29	24%	1.13
Small Programs CE	8	0.99	0.22	22%	0.96
Large Programs CE	8	1.02	0.16	15%	1.01
Small Programs Total Less CE	8	1.08	0.19	17%	1.09
Large Programs Total Less CE	8	1.20	0.20	16%	1.11

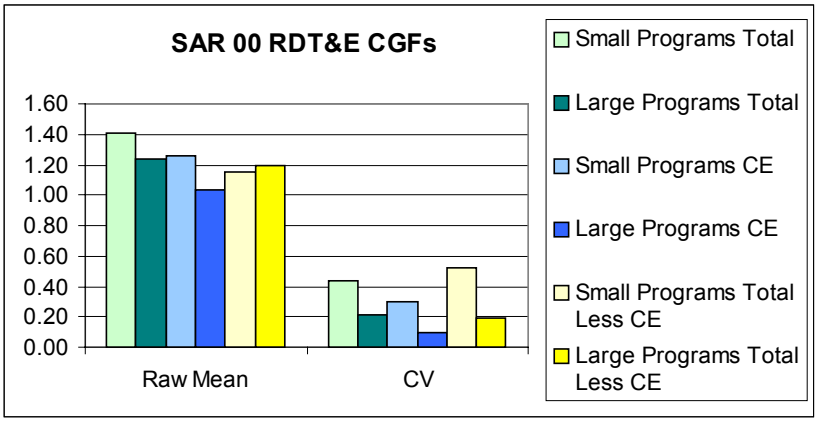
A graphical representation of the Acquisition CGFs by size is provided below.



RDT&E:
The table below summarizes the SAR 00 RDT&E results. Note that dollar-weighted means could not be calculated since only Total Acquisition dollars were available. As found in the RAND 93 data, it appears that small programs tend to have larger growth. This effect will be tested for statistical significance.

RDT&E CGFs				
	Number	Raw Mean	Raw Std. Dev	CV
Small Programs Total	8	1.41	0.61	43%
Large Programs Total	8	1.23	0.26	21%
Small Programs CE	8	1.26	0.38	30%
Large Programs CE	8	1.04	0.10	10%
Small Programs Total Less CE	8	1.15	0.60	52%
Large Programs Total Less CE	8	1.19	0.23	19%

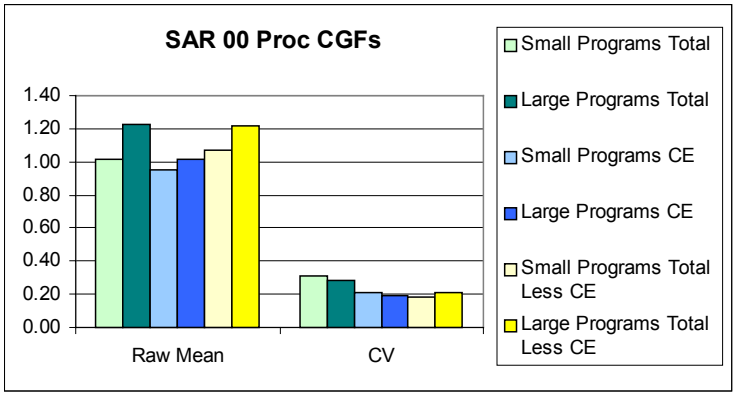
A graphical representation of the descriptive statistics for RDT&E is provided below.



Procurement:
The table below summarizes the SAR 00 Procurement results. Note that dollar-weighted means could not be calculated since only Total Acquisition dollars were available. It appears that large programs have higher growth, opposite of the results from the RAND 93 data. This effect will be tested for statistical significance.

Proc CGFs				
	Number	Raw Mean	Raw Std. Dev	CV
Small Programs Total	8	1.02	0.32	31%
Large Programs Total	8	1.23	0.35	28%
Small Programs CE	8	0.95	0.20	21%
Large Programs CE	8	1.01	0.20	20%
Small Programs Total Less CE	8	1.07	0.19	18%
Large Programs Total Less CE	8	1.22	0.26	21%

A graphical representation of the SAR 00 Procurement CGFs is provided below.



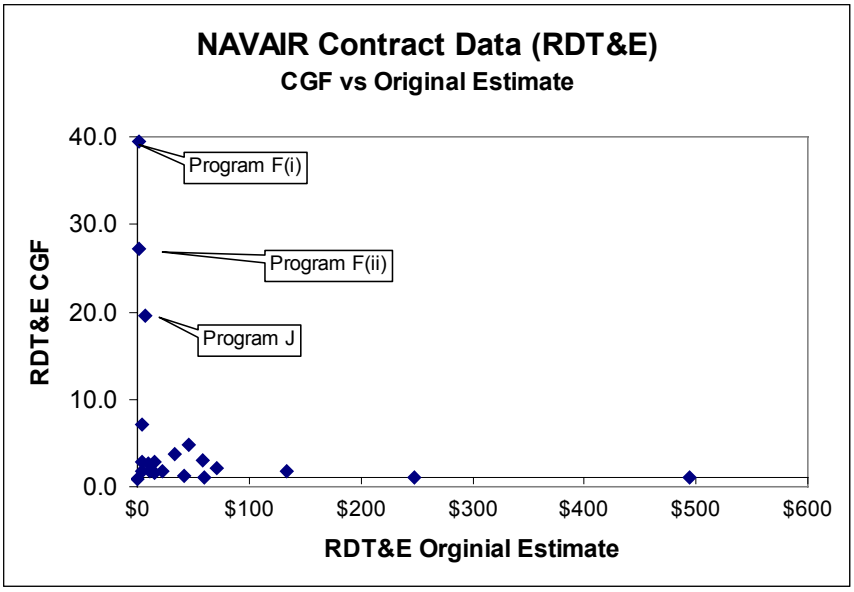
SAR 00 t-tests:
T-tests assuming unequal variances[9] were conducted to determine if there is a size difference in CGF means for Total, CE, and Total Less CE (S/T) in Acquisition, RDT&E, and Procurement. The results are displayed in the table below. There was no statistically significant difference found for any of the relationships.

	Critical Value	t Stat	p-value	Significant?
Small to Large Total Acquisition CGFs	2.14	-0.88	0.40	No
Small to Large Total RDT&E CGFs	2.23	0.76	0.47	No
Small to Large Total Proc CGFs	2.14	-1.25	0.23	No
Small to Large CE Acquisition CGFs	2.16	-0.26	0.80	No
Small to Large CE RDT&E CGFs	2.31	1.59	0.15	No
Small to Large CE Proc CGFs	2.14	-0.60	0.56	No
Small to Large S/T Acquisition CGFs	2.14	-1.22	0.24	No
Small to Large S/T RDT&E CGFs	2.26	-0.20	0.85	No
Small to Large S/T Proc CGFs	2.16	-1.32	0.21	No

Analysis – Contract Data:
The same analysis used for the RAND 93 and SAR 00 data was conducted for the Contract data. The data was divided into two groups of equal numbers based on size, as measured by the original estimate. Analysis was performed only for RDT&E as there were only four data points for Procurement. As with the SAR 00 data, the small number of observations in the Contract dataset makes it is difficult to draw conclusions.

A scatterplot of the Contract data is shown below. There are three programs with CGFs of approximately 20.0 and higher that are possible outliers. However, these programs cannot be justifiably removed without further information. Therefore, analysis is conducted both with and without these three points.

[9]T-tests assuming equal variances were also performed with the same results.



The table below summarizes the descriptive statistics for the Contract data. Note that the weighted means are again all smaller than the simple means, suggesting that larger programs have smaller growth. As in the RAND 93 data, there appears to be a trend of larger programs having smaller cost growth. This relationship will be tested for statistical significance.

Contract Data RDT&E CGF							
	Number	Total $	Avg. $ / Program	Raw Mean	Raw Std. Dev	CV	$ Wtd. Mean
Small Programs	11	71.46	6.50	9.82	13.08	133%	4.87
Small Programs less possible outliers	8	63.06	7.88	2.71	1.90	70%	2.53
Large Programs	10	1,208.93	120.89	2.19	1.29	59%	1.56

T-tests assuming unequal variances[10] were conducted to determine if there was a statistically significant difference between the mean CGFs for the two size groups. As displayed in the table below, the difference in means was not statistically significant. Inclusion of the possible outliers did not affect the outcome of the t-test.

	t-stat	Critical Value	p-value	Significant?
Small to Large	1.92	2.23	0.08	No
Small less possible ouliers to Large	0.67	2.18	0.52	No

[10] T-tests assuming equal variances were also performed with the same results.

Conclusions:

Descriptive statistics performed support the hypothesis of a decrease in cost growth with an increase in program size. For all three datasets, the weighted averages were lower than the simple averages, suggesting that larger programs with smaller growth pull the average down. For RAND 93 and Contract data, size-grouped average CGFs were lower for large programs. This pattern did not hold in the SAR 00 data, but n was small and the difference in means was not statistically significant.

Although a size related trend was apparent, there was no statistically significant difference in mean CGFs among the size groups. This finding is consistent with a 1996 Allison/Coleman study[11], which concluded that "smaller programs tend to incur cost growth greater than larger programs." In this earlier study, descriptive graphs suggested that larger programs tend to have lower cost growth, though a regression showed it was not a statistically significant result.

For both RDT&E and Procurement, the three size groupings were shown to be identically distributed.[12] There is appears to be a trend of more high end extrema in the smaller size classes, although the difference is not statistically significant. One explanation for this phenomenon is that high risk programs may be terminated early if they are large (expensive), but tolerated if they are small (relatively inexpensive). This result is not consistent with the Allison/Coleman study, in which a K-S test showed a statistically significant difference between the CGF distributions for small and large program groups. However, this earlier analysis was done at the Total Acquisition level, instead of for RDT&E and Procurement separately. In addition, the Allison/Coleman study did not use cohorted data.

11 *Weapon System Cost Growth As a Function of Maturity*, DoDCAS 1996, K. J. Allison, R. L. Coleman

12 This was tested only in the RAND 93 data set due to the small sample size in SAR 00 and Contract.

Effects of Era on Cost Growth

Principal Analyst – Cari L. Pullen

Purpose:

Analysis was performed to determine if era affects cost growth in RDT&E and Procurement. It is hypothesized that more recent programs have less cost growth as compared to older programs.

Data:

The initial analysis was performed using DoD programs with DE only from the RAND 93 dataset. The same analysis was then repeated with the SAR 00 and Contract data, which include NAVAIR programs only. Various shortcomings are associated with each of the three datasets. For example, the RAND 93 data cannot be broken out between Cost Estimating (CE) risk and Schedule/Technical (S/T) risk. The SAR 00 data cannot be dollar weighted for RDT&E and Procurement because total dollars are not broken out by appropriation. The Contract data analysis focuses only on RDT&E, as there are only four data points for Procurement.

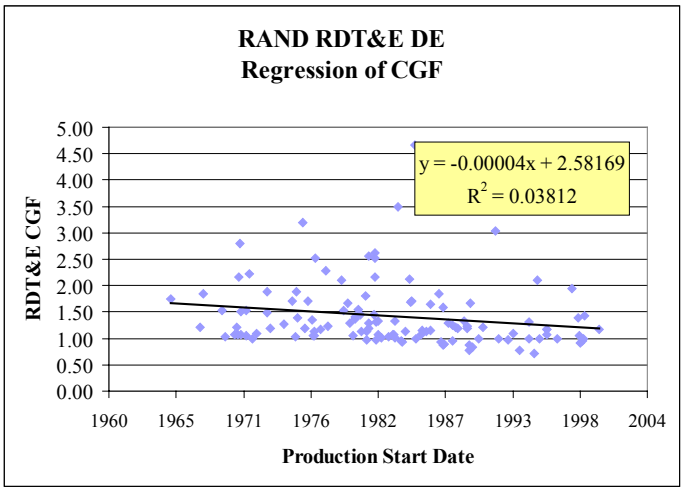
Approach:

Descriptive analysis included regression and moving averages over time for each data set. Inferential analysis noted the differences in each data set, then grouped programs by era and compared the CGFs. Furthermore, a t-test was performed to determine if the mean CGF was significantly different for each era. Results were then compared to previous studies on this subject.

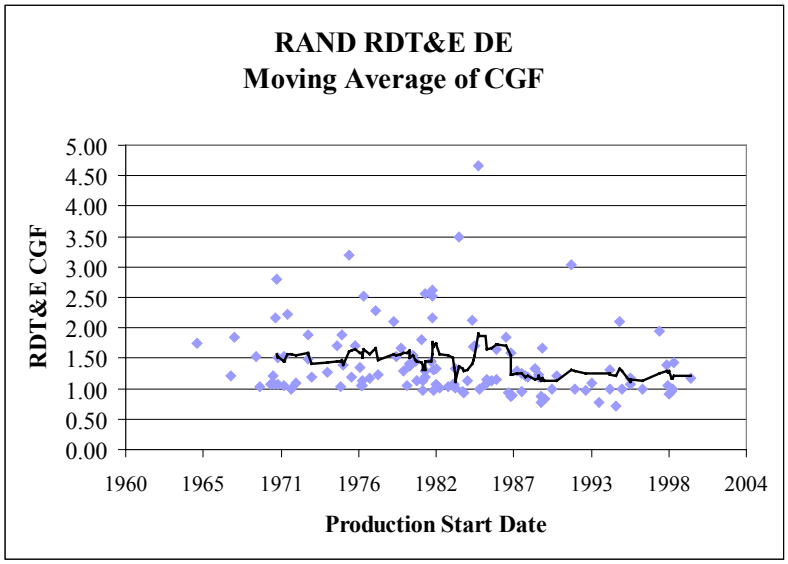
Analysis—RAND 93:

RDT&E:

A scatter plot of Production Start Date and RDT&E CGF is shown below. There does appear to be a slight downward trend of lower cost growth in later years. While the correlation is significant (F statistic of 4.72), the R^2 value is very small, indicating that era does not play a large role in the decrease in the CGF.



A moving average was computed using periods of 10. There is a slight decline in the average, but there are no significant drops indicating a point in time responsible for this decrease.



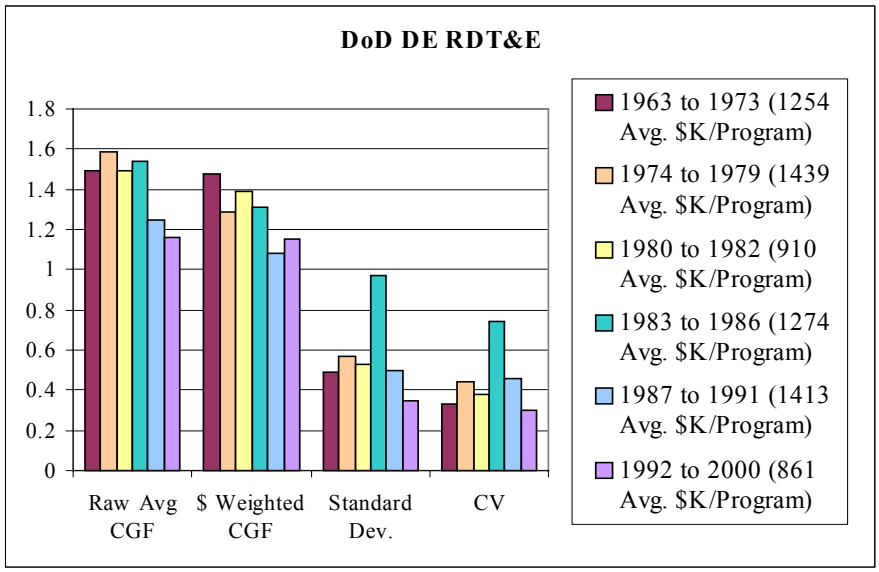
Next, the data was divided into two groups, up to and including 1986 and after 1986 (as measured by the Year of Production Start). 1986 was chosen as the dividing year as it is the end of the Reagan plus-up in defense spending. Results are as follows:

	Count	Total Dev. $	Avg. Dev. $	Raw Avg. CGF	$ Weighted CGF	St. Dev	CV
Production Start Date <= 1986	80	97177	1215	1.53	1.36	0.65	0.48
Production Start Date > 1986	41	46895	1144	1.20	1.11	0.43	0.39

The average RDT&E CGF appears to be decreasing through time. To take a closer look, programs were grouped as indicated below:

- 20 programs with a production start date between 1963 and 1973 and with an average cost per program of $1254K.
- 20 programs with a production start date between 1974 and 1979 and with an average cost per program of $1439K.
- 21 programs with a production start date between 1980 and 1982 and with an average cost per program of $910K.
- 19 programs with a production start date between 1983 and 1986 and with an average cost per program of $1274K.
- 21 programs with a production start date between 1987 and 1991 and with an average cost per program of $1413K.
- 20 programs with a production start date between 1992 and 2000 and with an average cost per program of $861K.

Graphical results are as follows:



A t-Test assuming unequal variance[13] was performed to test for differences in the means between the two era groups (<= 1986 and > 1986).

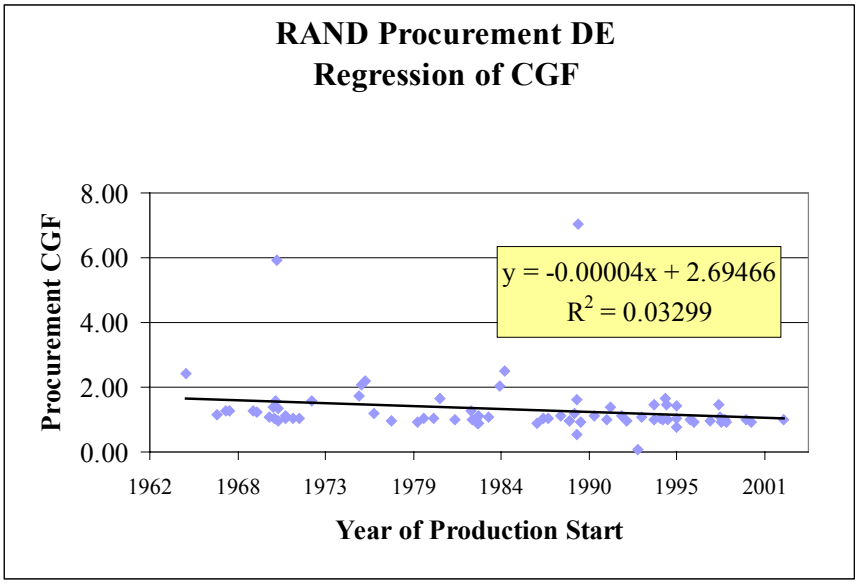
t-Test: Two-Sample Assuming Unequal Variances				
	t stat	critical value	p-value	Significant Difference?
DoD DE RDT&E	3.25	1.98	0.002	Yes

The difference between the means of the "old" and "new" groups is statistically significant.

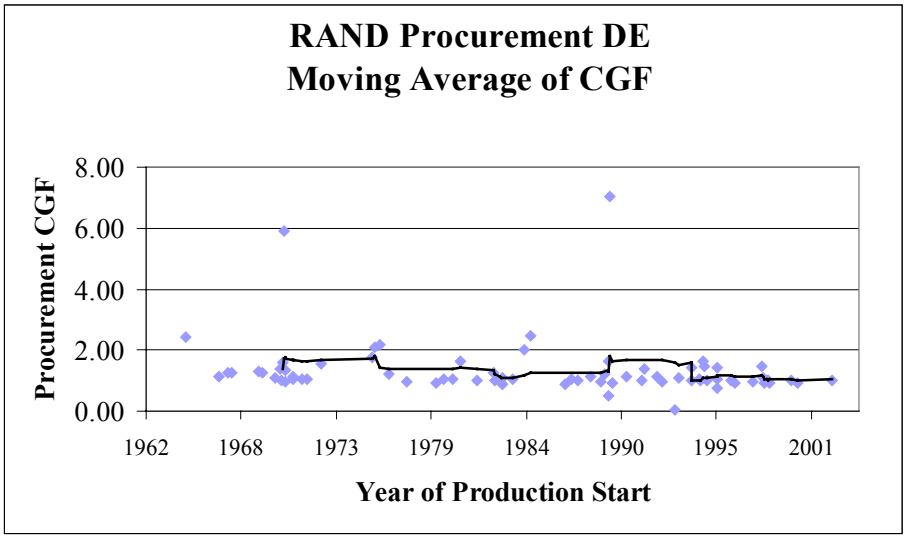
Procurement:

A scatter plot of Production Start Date and Procurement CGF is shown below. There does appear to be a slight downward trend of lower cost growth in later years. However, the correlation is not significant (F statistic of 2.52) and the R^2 value is very small, indicating that era does not play a large role in the decrease in the CGF.

[13] A t-Test assuming equal variance was also performed with the same results



A moving average was computed using periods of 10. There is a slight decline in the average, but there are no significant drops indicating a point in time responsible for this decrease.

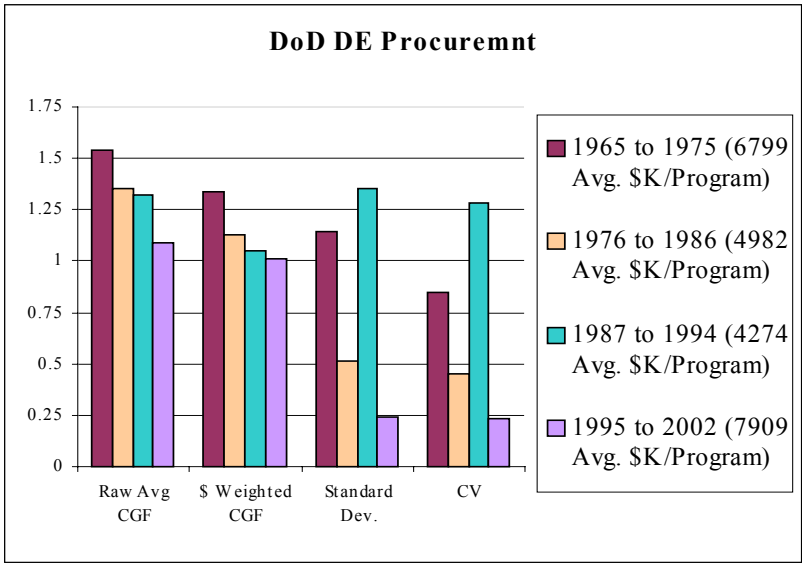


Next, the data was divided into two groups, up to and including 1986 and after 1986 (as measured by the Year of Production Start). Results are as follows:

	Count	Total Proc. $	Avg. Proc. $	Raw Avg. CGF	$ Weighted CGF	St. Dev	CV
Production Start Date <=1986	37	217044	5866	1.44	1.25	0.87	69%
Production Start Date > 1986	39	232114	5952	1.21	1.03	1.00	97%

The average Procurement CGF appears to be decreasing through time. To take a closer look, programs were grouped as indicated below:

- 18 programs with a production start date between 1965 and 1975 and with an average cost per program of $6799K.
- 19 programs with a production start date between 1976 and 1986 and with an average cost per program of $4982K.
- 21 programs with a production start date between 1987 and 1994 and with an average cost per program of $4274K.
- 18 programs with a production start date between 1995 and 2002 and with an average cost per program of $7909K.



A t-test assuming unequal variance[14] was performed to test for differences between the means of the two groups (<= 1986 and > 1986).

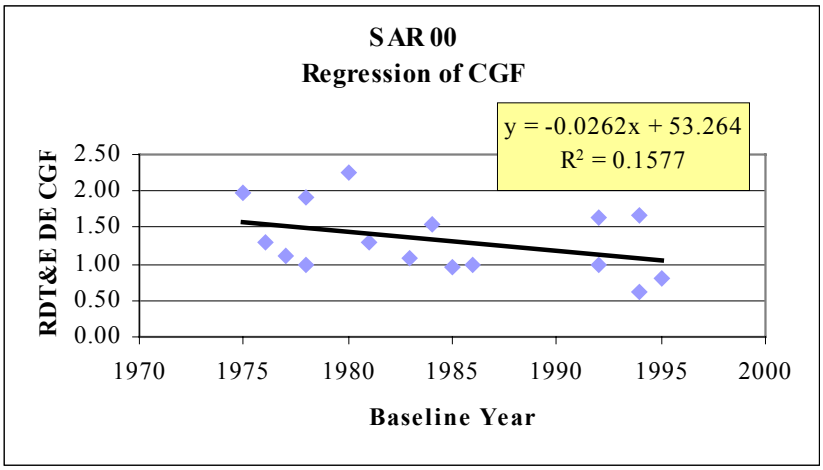
t-Test: Two-Sample Assuming Unequal Variances				
	t stat	critical value	p-value	Significant Difference?
DoD DE Procurement	1.08	1.99	0.29	No

[14] A t-Test assuming equal variance was also performed with the same results.

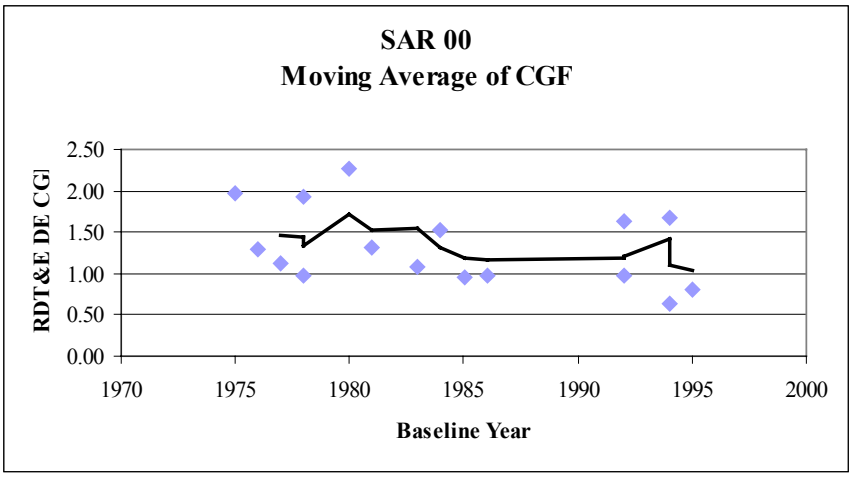
While the CGF decreases over time, the means of the two segments are not significantly different.

Analysis—SAR 00:

RDT&E:
A scatter plot of the Baseline Year and RDT&E Acquisition CGF is shown below. There does appear to be a downward trend of lower cost growth in later years. However, the correlation is not significant[15] (F statistic of 2.62), and the R^2 value is small, indicating that era does not play a large role in the decrease in the CGF.



A moving average was computed using periods of 3. There is a slight decline in the average, but there are no significant drops indicating a point in time responsible for this decrease.



[15] In the RAND 93 data, there was a significant correlation between year and RDT&E CGF, but the R^2 was very small.

Next, the data was divided into two groups, up to and including 1986 and after 1986 (as measured by the Baseline Year). Results are as follows:

	Count	Raw Avg	St. Dev	CV	Avg. $K per Program
Baseline Year <= 1986	11	1.40	0.46	33%	$7,960.86
Baseline Year > 1986	5	1.15	0.48	42%	$15,372.06

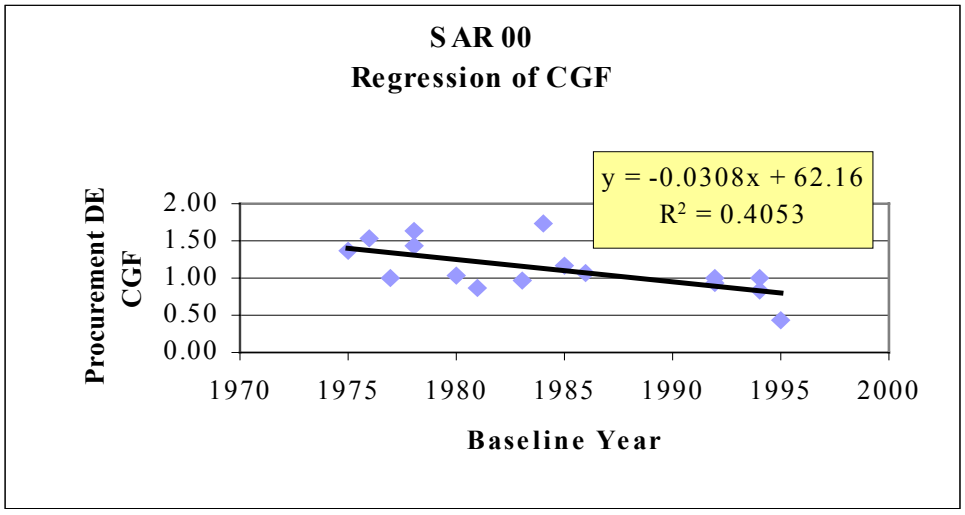
A t-test assuming unequal variance was performed to test for differences in the means between the two groups (<= 1986 and > 1986).

t-Test Summary	t-stat	Critical Value	P value	Significance?
<>Variance RDT&E	1.02	2.14	0.3243	No
= Variance RDT&E	1.01	2.31	0.3441	No

The difference between the means of the "old" and "new" groups is not statistically significant[16].

Procurement:

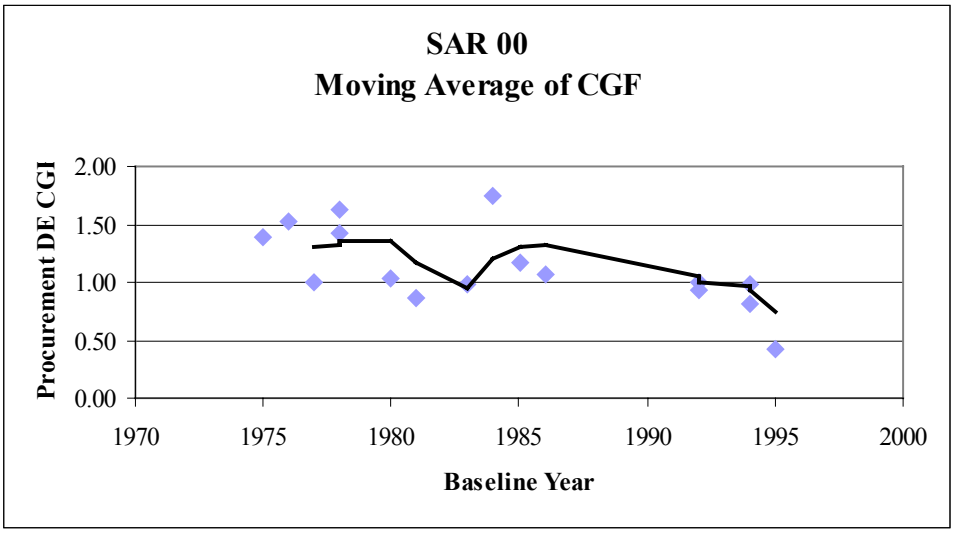
A scatter plot of the Baseline Year and Procurement Acquisition CGF is shown below. There does appear to be a downward trend of lower cost growth in later years. The correlation is significant[17] (F statistic of 9.54), and the R^2 value is fairly large, indicating that era does play a role in the decrease in the CGF.



.

[16] In RAND 93, the difference in RDT&E CGF era-grouped means ***was*** statistically significant.
[17] In the RAND 93 data, the correlation between year and Procurement CGF was ***not*** statistically significant, and the R^2 was very small.

A moving average was computed using periods of 3. There is a slight decline in the average, but there are no significant drops indicating a point in time responsible for this decrease.



Next, the data was divided into two groups, up to and including 1986 and after 1986 (as measured by the Baseline Year). Results are as follows:

	Count	Raw Avg	St. Dev	CV	Avg. $K Per Program
Baseline Year <= 1986	11	1.26	0.30	24%	$7,960.86
Baseline Year > 1986	5	0.83	0.24	28%	$15,372.06

The average Procurement CGF appears to be decreasing over time. A t-test assuming unequal variance was performed to test for differences in the means between the two groups (<= 1986 and > 1986).

t-Test Summary	t-stat	Critical Value	P value	Significance?
<> Variance Procurement	2.77	2.14	0.0152	Yes
= Variance Procurement	3.04	2.23	0.0125	Yes

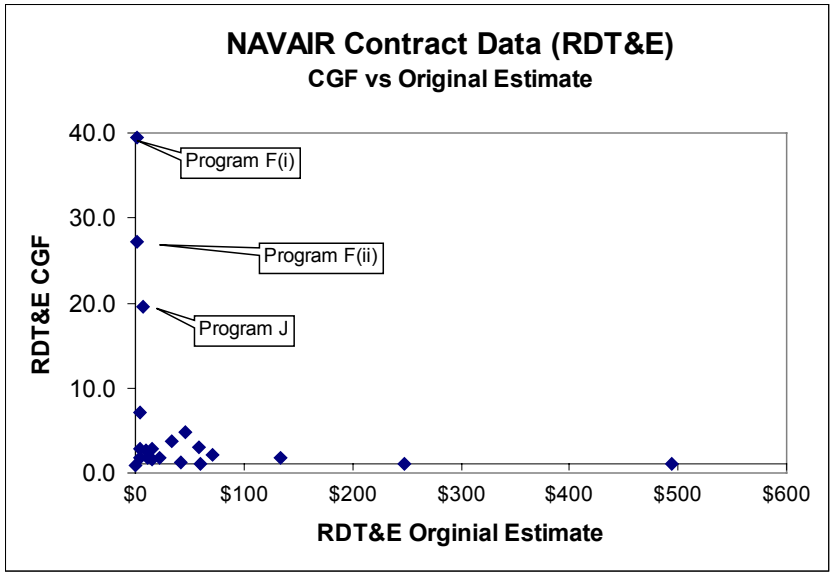
The difference between the means of the "old" and "new" groups is statistically significant[18].

[18] In RAND 93, the difference in Procurement CGF era-grouped means was ***not*** statistically significant.

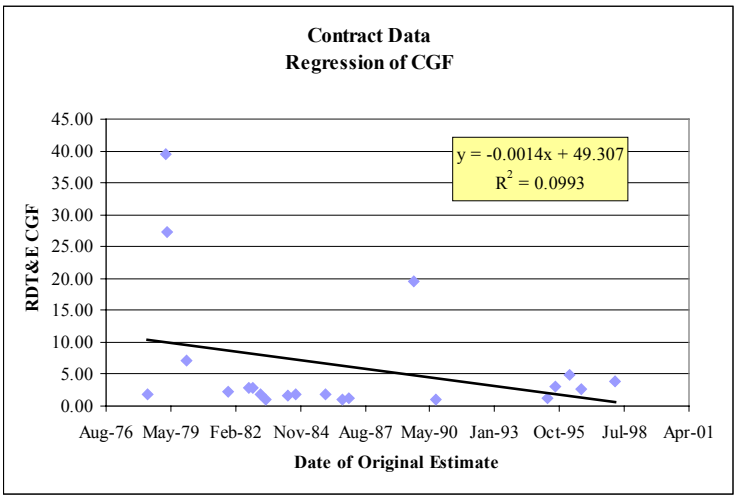
Analysis—Contract Data:

RDT&E:

The Contract data analysis was done the full set of data points, and with a sub-set. Three of the programs may be outliers, but cannot be removed without further information on the programs:

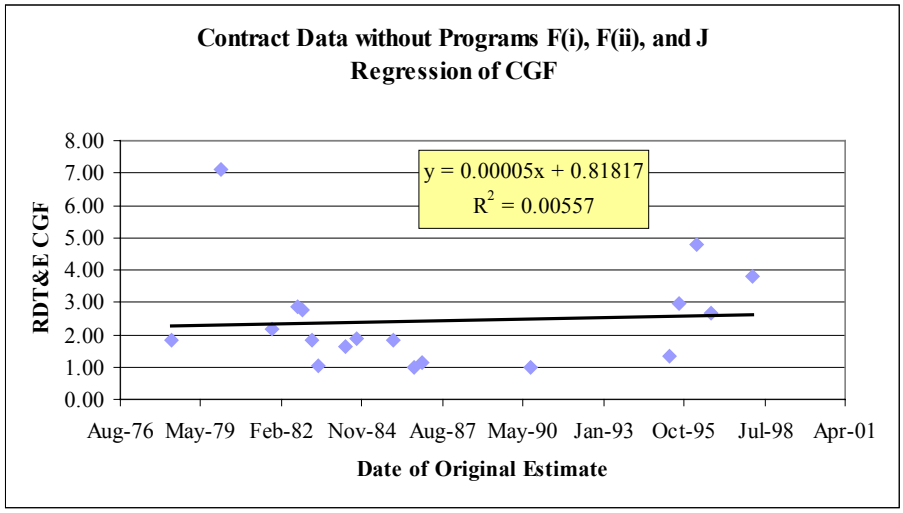


A scatter plot of the Original Estimate Date and RDT&E CGF for the full Contract dataset is shown below. There does appear to be a downward trend of lower cost growth in later years. However, the correlation is not significant[19] (F statistic of 2.09), and the R^2 value is very small, indicating that era does not play a role in the decrease in the CGF.

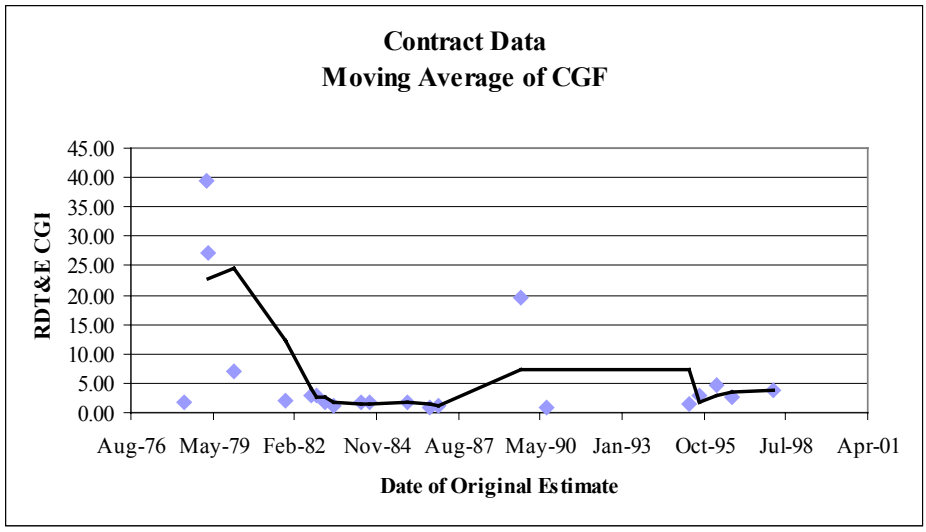


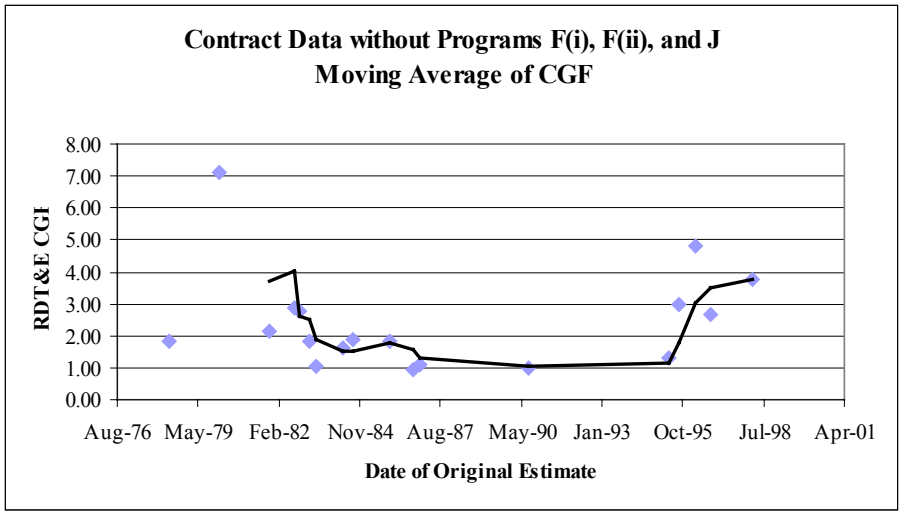
[19] In the RAND 93 data, there was a significant correlation between year and RDT&E CGF, but the R^2 was very small.

The following scatter plot shows the Original Estimate Date and RDT&E CGF for the Contract data with the possible outliers removed. In this case, there is a slight upward trend in CGF over time, but the correlation is not significant (F statistic of .09) and the R^2 value is very small, indicating that that era does not play a role in this increase in the CGF.



Moving averages were computed using periods of 3 for both the full dataset and with possible outliers removed. There is a slight decline in the average, but there are no significant drops indicating a point in time responsible for this decrease.





Next, the data was divided into two groups, up to and including 1986 and after 1986 (as measured by the Date of Original Estimate). Results are as follows:

RDT&E		Count	Raw Avg	$ Weignted Avg.	St. Dev	CV	Avg. $M per Program
All Programs	Date of Original Estimate <= 1986	14	6.70	1.50	11.66	174%	$59.79
	Date of Original Estimate > 1986	7	5.15	2.21	6.46	125%	$63.33
No Outliers	Date of Original Estimate <= 1986	12	2.25	1.43	1.65	73%	$69.61
	Date of Original Estimate > 1986	6	2.76	1.95	1.45	52%	$72.79

A t-test assuming unequal variance was performed to test for differences in the means between the two groups (<= 1986 and > 1986).

t-Test Summary	t-stat	Critical Value	P value	Significance?
<>Variance with outliers	0.39	2.09	0.700	No
<> Variance without outliers	-0.67	2.20	0.516	No
= variance with outliers	0.32	2.09	0.749	No
= variance without outliers	-0.64	2.12	0.531	No

The differences between the means of the "old" and "new" groups are not statistically significant[20].

[20] In RAND 93, the difference in RDT&E CGF era-grouped means ***was*** statistically significant.

Conclusions:

Differences by Era

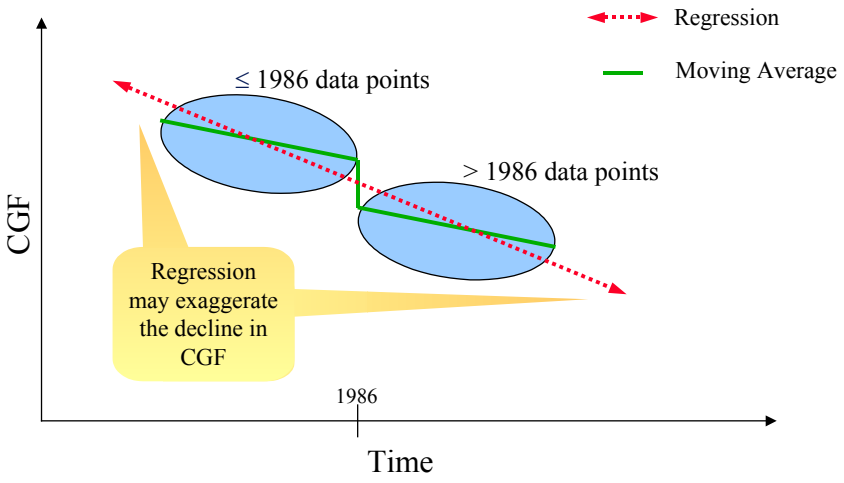
	Previous Studies		Current Studies		
	RAND	RLC/KJA	RAND 93	SAR 00	Contract Data
RDT&E	--	--	Y, Y	N, N	N, N
Procurement	--	--	N, N	Y, Y	--
Total	N	N, N	--	--	--

Based on Histograms

Legend: Y, N

Y = Statistically Significant
N = Not Statistically Significant
Where the first response (Y) corresponds to the t-test and the second response (N) corresponds to the regression.

Regression analysis and moving averages indicate a decrease over time in RDT&E and Procurement cost growth. They depict a decline, with one exception: Contract Data without outliers. However, the R^2 value however was quite low for each data set, with one exception: SAR 00 Procurement. However, caution must be taken when using regression analysis, as it may mislead.



T-tests show inconsistent results:

- RAND 93: The means of programs through 1986 and those after 1986 *did* show a statistical difference for RDT&E, but *not* for Procurement
- SAR 00: The means of programs through 1986 and those after 1986 *did* show a statistical difference for Procurement, but *not* RDT&E.
- Contract Data: The means of programs through 1986 and those after 1986 *did not* show a statistical difference for RDT&E *(for the entire data set and the subset)*

Results from previous studies support the current findings presented in this paper:

- Rand[21] Conclusion: "Weighted average total program cost growth has been fairly constant over time, averaging around 20 percent"
    - Based on descriptive histogram of CGFs from 1965 to 1989
- Allison/Coleman[22] Conclusion: "There has not been a significant change in total cost growth between the two time periods [prior-to-1989 and 1991]"
    - Basis:
        1. K-S test showed 2 databases (prior-to-1989 and 1991) had identical distributions.
        2. Regression analysis showed there was no significant relationship between cost growth and system base year.

[21] Analyses reported here were done using DoD data at the Acquisition level.

[22] Cohort tracking was not used in either of these studies.

Phase-to-Phase Correlation of Cost Growth

Principal Analyst – Megan E. Dameron

Purpose:

To determine the presence of phase-to-phase correlation of cost growth. The following relationships were analyzed:

- Growth of RDT&E and Procurement appropriations *during* a phase (e.g., correlation between growth of the RDT&E estimate during EMD and growth of the Procurement estimate during EMD)
- Growth of all appropriations *between* phases (e.g., correlation between growth of the RDT&E estimate during PDRR and during EMD).

Approach:

Analysis was conducted using all DoD programs in the RAND 93 data set. This data set is based on Selected Acquisition Reports (SARs). Estimates are divided into three categories: Planning Estimates (PE), corresponding to PDRR; Development Estimates (DE), corresponding to EMD; and Production Estimates (PdE), corresponding to Production. Programs were divided into three "cohorted" groups – those with PE, DE, and PdE; those with PE and DE only; and those with DE and PdE only. The following relationships were tested for correlation using a Kendall Test at the $\alpha = 0.05$ level of significance:

	Correlation	**n**
RDT&E		
Programs with PE and DE	PE to DE	25
Procurement		
Programs with PE, DE, and PdE	PE to DE	6
	DE to PdE	6
	PE to PdE	6
Programs with DE and PdE only	DE to PdE	54
RDT&E and Procurement		
Programs with PE, DE, and PdE	RDT&E PE and Proc PE	6
	RDT&E DE and Proc DE	6
	RDT&E PE and Proc PdE	6
	RDT&E DE and Proc PdE	6
Programs with DE and PdE only	RDT&E DE and Proc DE	54
	RDT&E DE and Proc PdE	54

Regression analysis was also used to determine R^2, correlation (r), and coefficient of variation (CV).

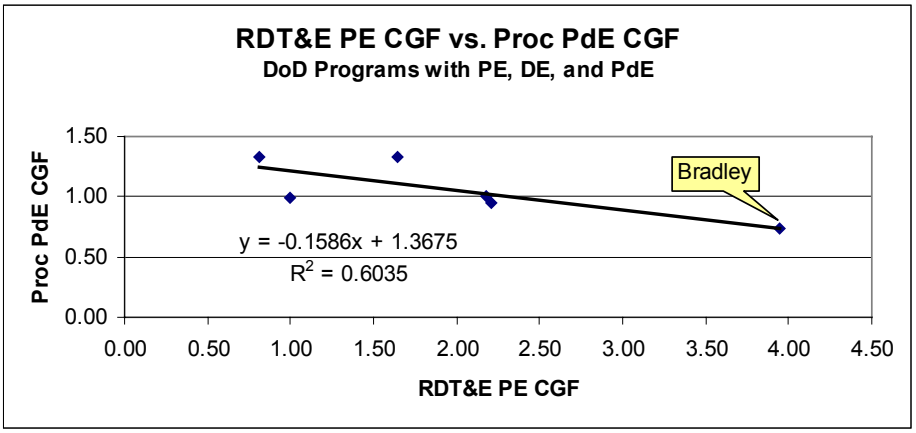
Results:

The results of the analysis are summarized in the table below. The following relationships showed a statistically significant correlation:

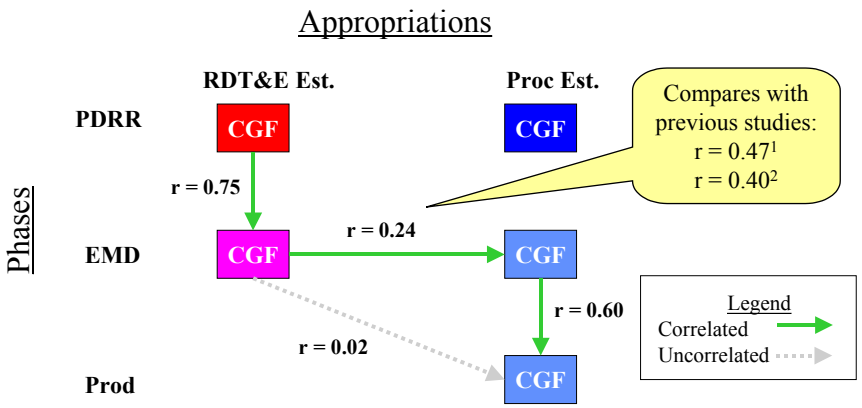
- RDT&E PE and RDT&E DE of programs with PE and DE only;
- Procurement DE and Procurement PdE of programs with DE and PdE only;
- RDT&E DE and Procurement DE of programs with DE and PdE only.

	Correlation	n	Kendall Critical Value	Kendall Test Statistic	Significant?	r	CV
RDT&E							
Programs with PE and DE	PE to DE	25	84	201	Yes	0.75	34%
Procurement							
Programs with PE, DE, and PdE	PE to DE	6	11	7	No	0.38	30%
	DE to PdE	6	11	3	No	-0.38	23%
	PE to PdE	6	11	-1	No	-0.40	22%
Programs with DE and PdE only	DE to PdE	54	263	561	Yes	0.60	19%
RDT&E and Procurement							
Programs with PE, DE, and PdE	RDT&E PE and Proc PE	6	11	3	No	0.44	44%
	RDT&E DE and Proc DE	6	11	7	No	0.82	19%
	RDT&E PE and Proc PdE	6	11	-9	No	-0.78	15%
	RDT&E DE and Proc PdE	6	11	-5	No	0.71	17%
Programs with DE and PdE only	RDT&E DE and Proc DE	54	263	267	Yes	0.24	27%
	RDT&E DE and Proc PdE	54	263	-55	No	0.02	23%

The three scenarios demonstrating negative correlation were small samples, skewed by Bradley. The Bradley program experienced high growth in all areas except Procurement PdE, where it experienced negative growth (CGF of 0.73). For example, consider the scatter plot for the RDT&E PE and Procurement PdE for programs with PE, DE, and PdE. The regression line is pulled down by Bradley. These negative correlations are not statistically significant, and the effect should be ignored.



The following diagram depicts the presence of correlation. There were many areas with too few data points to draw conclusions, so only those with sufficient data are indicated. RDT&E PDRR is correlated to RDT&E EMD, which is correlated to Procurement EMD, which is then correlated to Procurement PdE.



Conclusions:

Significant correlations were found between:

- RDT&E growth during PDRR and during EMD
- Procurement growth during EMD and during Production
- RDT&E growth and Procurement growth during EMD.

So, correlation does exist between appropriations and between phases. The finding of correlation between appropriations is consistent with both a 1996 Allison/Coleman study[23] and a 1997 Coleman/Gupta/Summerville/Hartigan paper[24]. Both of these previous studies found a significant correlation between RDT&E and Procurement, although somewhat higher than the correlation found in the work presented here.

23 *Weapon System Cost Growth As a Function of Maturity*, DoDCAS 1996, K. J. Allison, R. L. Coleman

24 *Cost Risk Estimates Incorporating Functional Correlation, Acquisition Phase Relationships, and Realized Risk*, SCEA National Conference 1997, R. L. Coleman, S. S. Gupta, J. R. Summerville, G. E. Hartigan

Conclusions

In order to gain insight into historical cost growth in NAVAIR programs, it was necessary to divide this analysis into smaller components. Detailed statistical analysis of each component resulted in a conclusion describing the magnitude and nature of the impact on cost growth. The results are useful for determining areas that need to be addressed in the assessment of cost risk on all programs within NAVAIR. The components and conclusions are summarized as follows:

Does program size affect cost growth? Distributions for programs in different size groups tested as being identical. However, there is a clear trend that shows there is a higher tolerance for cost growth in smaller programs. Therefore, it is important to keep in mind that high risk programs may be terminated earlier if large, but tolerated if small.

Does the era of the program effect cost growth? There is convincing evidence that cost growth has decreased over time. Specifically, programs before 1986 incurred larger growth than those after 1986. It is suggested that using data from only programs in the later group will provide for a more accurate prediction of cost risk.

Does correlation exist? Analysis shows that correlation exists between phases and between appropriations in the EMD phase. Appropriation-to-appropriation correlation suggests that if R&D cost grows and Procurement does not, then it is likely that something has been missed. The two should generally rise and fall together. Phase-to-phase correlation suggests that if R&D goes up in PE, expect it to happen again in DE. This holds true for Procurement from DE to PdE as well.

Does the structure of a program's life cycle affect cost growth? The data this analysis was split into cohorted groups to determine if the inclusion of a PDRR phase as well as the maturity of a program impacts its cost growth. If a program has a planning phase (PDRR) then much of the R&D growth will occur in that phase. As we know from the correlation results, R&D will continue to grow in EMD, but to a lesser extent. If a program does not have a planning phase, then a large amount of growth in EMD should be expected (though typically not as high as it becomes with programs having both a PDRR and EMD phase). Most Procurement growth occurs during EMD to the Procurement estimate and a smaller amount of growth occurs after the start of Production. These trends should be kept in mind while conducting risk analysis on a program throughout its life cycle.

Is cost growth affected by commodity type? NAVAIR data is not plentiful enough to validate any conclusions with inferential statistics, but the descriptive results from this study suggest that Missiles incur the largest amount of growth among SAR NAVAIR programs and EW programs incur the largest amount of growth among the smaller programs (with only contract data) within NAVAIR. Therefore commodity should be considered in risk modeling, particularly for Missile and EW programs.

While incorporating all these results to achieve an appropriate data set for assessing cost risk, it is critical to keep in mind that the number of data points must be kept sufficiently large. It has been observed that NAVAIR data follow similar trends to all DoD data, which are plentiful. Therefore, conclusions drawn from the DoD data set could be used to assist in validating or refining results from strictly NAVAIR data sets with a small number of programs. Modifications using adjustment factors calculated by comparing a NAVAIR dataset that does have somewhat sufficient n (i.e. RDT&E growth in DE only programs) with the equivalent DoD set of data are one possibility. This is an area that will require additional thought and future exploratory analysis in order to develop a cost risk model for NAVAIR.

Appendix: NAVAIR Cost Growth Study "Thumb Rules"

Purpose:

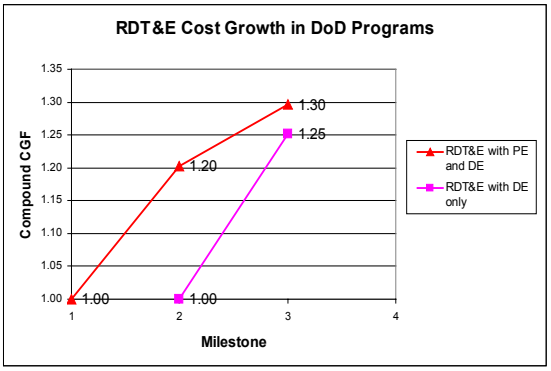
This paper is designed to provide a concise overview of results from an intensive cost growth study performed for NAVAIR in 2000. Details of the analysis can be found in the final documentation, *NAVAIR Cost Growth Study: A Cohorted Study of The Effects of Era, Size, Acquisition Phase, Phase Correlation and Cost Drivers* by M.E. Dameron, C.L. Pullen, J.R. Summerville, R.L. Coleman, and D.M. Snead. The intent is for this study to provide a basis to develop a methodology to assess cost risk for NAVAIR. Since this product will take time to develop, the results thus far are outlined below to provide NAVAIR analysts insight into expectations for cost growth, as well as a general idea of the magnitude and type of results that the model will likely produce. The following topics will be addressed:

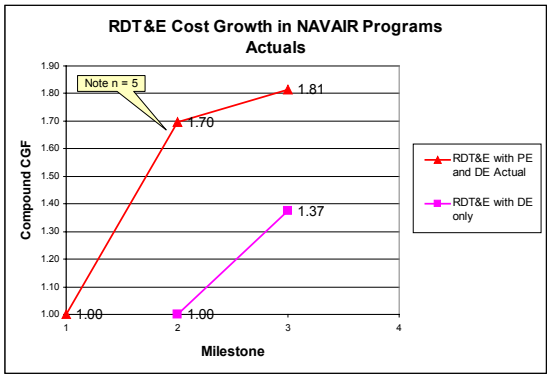
- RDT&E Cost Growth Factors (CGFs)
- Procurement CGFs
- Impact of program size on cost growth
- Impact of program era on cost growth
- Comparison of cost growth across different commodities
- Correlation of cost growth between phases and appropriations

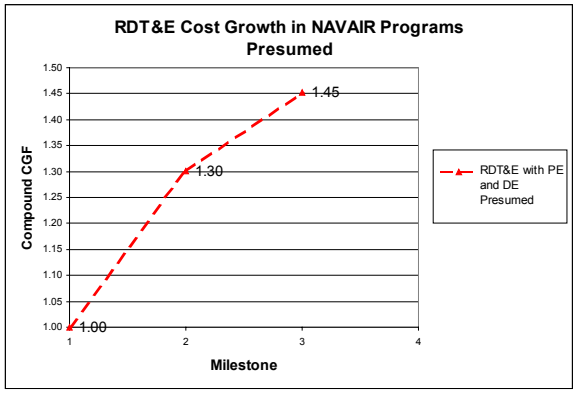
RDT&E Cost Growth Factors:

The following tables and graphs show the historical Cost Growth Factors (CGFs) derived from RDT&E estimates reported in SARs. These factors are adjusted to ignore effects of inflation and quantity changes, but otherwise incorporate total cost growth. As evident below, programs with a Program Definition and Risk Reduction (PDRR) phase experience slightly more growth than programs that enter directly in an Engineering and Manufacturing Development (EMD) phase without performing PDRR. This is not to say that simply excluding PDRR will lead to less growth. Programs that have this phase are inherently harder, and thus, have more potential to incur growth.

Also note that the bulk of growth is discovered during PDRR if it is performed. The figures below suggest that this pattern is the same for DoD and NAVAIR. Note that the sample size for NAVAIR programs with both PDRR and EMD is too small (n = 5) to have confidence in the actual number (CGF of 1.81). However, since the other data sets have large sample sizes, and the patterns of growth are similar, it would be reasonable to use that information to analytically determine a believable result. I.e, instead of using 1.81, the 1.37 NAVAIR CGF for "programs with EMD only" could be adjusted by scaling in proportion to DoD results, giving a CGF of 1.45.







Programs with PDRR and EMD

		PE (PDRR)		DE (EMD)		PdE (Prod)		Total
	N	CGF	CV	CGF	CV	CGF	CV	CGF
DoD	25	1.20	62%	1.08	49%			1.30
NAVAIR	5	1.70	36%	1.07	57%			1.81

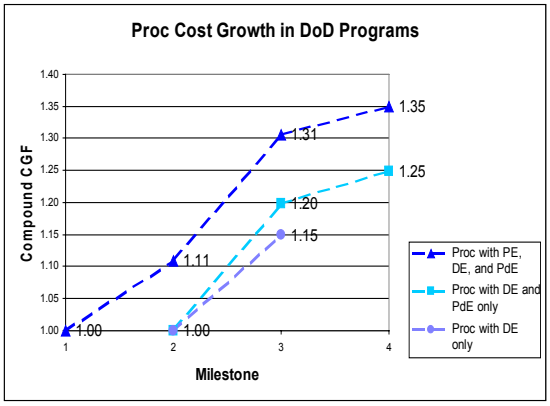
Programs with EMD only

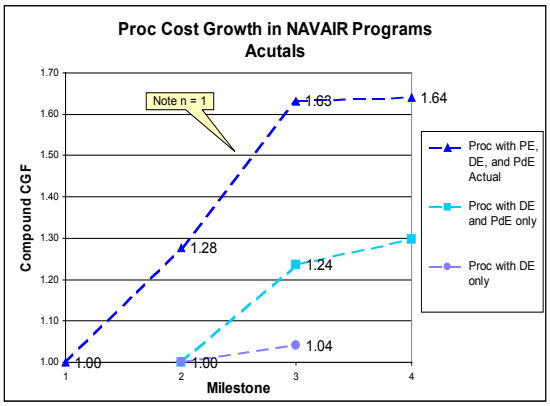
		PE (PDRR)		DE (EMD)		PdE (Prod)		Total
	N	CGF	CV	CGF	CV	CGF	CV	CGF
DoD	139			1.25	51.8%			1.25
NAVAIR	20			1.37	74%			1.37

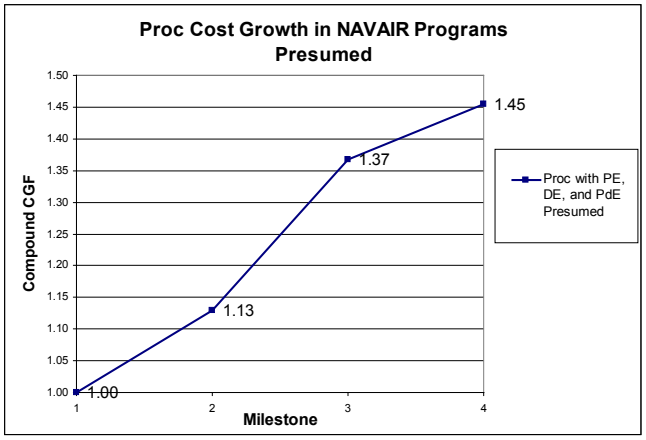
Procurement Cost Growth Factors:

The following tables and graphs show the Cost Growth Factors (CGFs) derived from Procurement estimates reported in SARs. Again, these factors incorporate total cost growth less the effects of inflation and changes in quantity. As evident below, programs that perform a PDRR phase experience more growth than programs that do not perform this phase. This is not to say that excluding a PDRR phase will lead to less growth.

Programs that perform this phase are inherently harder, and thus, have more growth. Also note that most of the growth is discovered fairly early, during EMD. In addition, the pattern of growth is the same for DoD and NAVAIR. The sample size for NAVAIR programs with all three phases is too small (n = 1) to believe the actual numbers, but the pattern holds. Again, this CGF could be estimated by adjusting the NAVAIR CGF for "programs with EMD and Prod" by scaling in proportion to DoD results, yielding a CGF of 1.45.







Programs with PDRR, EMD, and Prod

		PE (PDRR)		DE (EMD)		PdE (Prod)		Total
	N	CGF	CV	CGF	CV	CGF	CV	CGF
DoD	6	1.11	44%	1.18	29%	1.03	22%	1.35
NAVAIR	1	1.28		1.28		1.01		1.64

Programs with EMD and Prod

		PE (PDRR)		DE (EMD)		PdE (Prod)		Total
	N	CGF	CV	CGF	CV	CGF	CV	CGF
DoD	53			1.20	27%	1.04	21%	1.25
NAVAIR	10			1.24	36%	1.05	20%	1.30

Programs with EMD only

		PE (PDRR)		DE (EMD)		PdE (Prod)		Total
	N	CGF	CV	CGF	CV	CGF	CV	CGF
DoD	94			1.15	66%			1.15
NAVAIR	12			1.04	27%			1.04

Program Size:

There appears to be a trend that smaller programs experience larger cost growth; however, the difference is not statistically significant. Nonetheless, the CGFs specific to the size groups shown below should still be used as the maximum likelihood estimators for each particular group.

RAND 93 DoD Programs with DE (EMD) only					
RDT&E	N	Range DE (EMD) baseline (FY96$M)	Avg DE (EMD) baseline (FY96$M)	DE (EMD) CGF	DE (EMD) CV
DoD Small	46	6 - 272	127	1.43	67%
DoD Medium	46	276 - 800	451	1.42	34%
DoD Large	47	816 - 13622	3109	1.22	45%

RAND 93 DoD Programs with DE (EMD) only					
Procurement	N	Range DE (EMD) baseline (FY96$M)	Avg DE (EMD) baseline (FY96$M)	DE (EMD) CGF	DE (EMD) CV
DoD Small	31	151 – 1136	652	1.54	84%
DoD Medium	31	1159 - 3515	2237	1.18	45%
DoD Large	32	3586 - 61632	12904	1.13	28%

Program Era:

A shift in cost growth was found at the end of the Reagan ramp-up in 1986, with higher cost growth occurring before 1986. Multiple data sets showed a tendency towards this shift, supporting the existence of two generations of growth. Therefore, mixing in data from the older generation could potentially skew results upward. If possible, cost growth studies should use post-1986 data only.

RAND 93 DoD Programs with DE (EMD) only								
	RDT&E				**Procurement**			
	N	Raw CGF	$ Weighted CGF	CV	N	Raw CGF	$ Weighted CGF	CV
Production Start Date <= 1986	80	1.53	1.36	48%	37	1.44	1.25	69%
Production Start Date > 1986	41	1.20	1.11	39%	39	1.21	1.03	97%

Commodities:

Missiles incurred the largest RDT&E cost growth in both the SAR 00 and RAND 93 datasets. For Procurement, Missiles again showed the most growth in the RAND 93 data, but Aircraft had the highest growth in the SAR 00 dataset. In the Contractor data, EW systems had the largest cost growth in both RDT&E and Procurement. It should be noted that when datasets are broken down by commodity, the sample sizes become very small (e.g., n < 10 for most of the groups in this study). This makes it impossible to draw any statistical conclusions. Therefore, these results should be viewed cautiously, as they are simply preliminary suggestions based on raw calculations from very small sets of data.

RAND 93 DoD Programs with DE (EMD) only								
	RDT&E				**Procurement**			
	N	Raw CGF	$ Weighted CGF	CV	N	Raw CGF	$ Weighted CGF	CV
All	20	1.72	1.37	74%	12	1.15	1.04	27%
Aircraft	7	1.34	1.32	21%	6	1.03	1.02	12%
Missile	7	2.45	1.88	82%	4	1.33	1.29	37%
Electronics	6	1.33	1.39	36%	2	1.13	1.15	16%

SAR 00 NAVAIR data						
	RDT&E			**Procurement**		
	N	Raw CGF	CV	N	Raw CGF	CV
All	11	1.29	37%	11	1.14	24%
Aircraft	4	1.14	16%	4	1.27	29%
Missile	3	1.77	33%	4	1.07	24%
Electronics	3	1.24	30%	3	1.04	10%

Contract Data NAVAIR programs				
RDT&E				
	N	Raw CGF	$ Weighted CGF	CV
All	21	2.25	1.98	67%
EW	11	2.81	3.24	64%
Radar	5	1.63	1.39	74%
Radio	2	1.83	1.83	1%
Target	1	1.86	1.86	-
Avionics	2	1.40	1.41	7%

Correlation:

CGFs from the RAND93 SAR data were tested for correlation between phases and between appropriations. Significant phase-to-phase correlation was found between RDT&E growth in PDRR and in EMD. In addition, there was a significant correlation present between Procurement growth during EMD and during Production. Correlation also extends across appropriations in EMD. A summary of these correlations is shown in the figure below.

This suggests that program managers should expect programs with above average RDT&E growth in PDRR to also have above average growth in EMD. The same is true for Procurement growth in EMD and Production. Likewise, large growth in the RDT&E estimate during EMD is usually associated with large growth in the Procurement estimate during EMD.