Technomics
Better Decisions Faster

Booz | Allen | Hamilton

Exploring How Systems Age (and Fail) and the Impact on O&S Cost Estimates

PATRICK MCCARTHY & ALEX KRENCICKI

2018 ICEAA PROFESSIONAL DEVELOPMENT & TRAINING WORKSHOP

ABSTRACT

Significant time is spent estimating the development and production costs of weapon systems. However, estimating how systems age and fail during the O&S phase, as well as trying to identify optimal sustainment and maintenance strategies can be every bit as challenging and have a significant impact on the lifecycle cost. One key aspect of reliability and maintenance studies that is often overlooked is how to properly incorporate right-censored data into the estimation of failure distribution parameters. Right-censored data is a condition commonly encountered in statistics, reliability engineering and medical research in which the true value of a data point is greater than the observed value, but the difference will not be known until a point later in time. This paper will address obstacles associated with predicting fleet aging profiles, including dynamic failure distributions and the integration of right-censored data into the estimate. We will also observe how the impact of right-censored data evolves over the course of the O&S phase. Lastly, we will address the significance of reliability studies on resource requirements and utilize our findings to identify likely areas of cost savings and cost avoidance in the O&S planning/budgeting phase.

TABLE OF CONTENTS

INTRODUCTION

"*Failure is the key to success; each failure teaches us something*" - Morihei Ueshiba, Martial Artist (1883-1969)

The quote above is certainly wrought with wisdom. While we may feel a brief sense of success with each non-failure or victory, we will not truly learn what it takes to improve until we experience failure. In the context of wars fought, games lost and passwords forgotten, this quote usually rings true. However, in the field of cost estimating, we do not always have the luxury of waiting until something fails to educate ourselves on what an anticipated rate of failure (or non-failure) will be – and we certainly do not have time to wait for failure data to estimate how the failure rate impacts the system cost.

The purpose of this paper is to stimulate thought and discussion on how best to estimate system aging and failures at various points of the program lifecycle. We address various ways to estimate how weapon systems, and more specifically their sub-systems, will age and fail. In particular, we address various techniques to use in performing this estimating as the program lifecycle evolves. We briefly review reliability theory and then analyze failure rate estimating approaches at two distinctly different periods of the program lifecycle. In doing so, classic reliability and statistical subject matter, such as censored data and Maximum Likelihood Estimation, are introduced and applied. Lastly, we address how these various techniques of estimating system reliability will influence cost estimates.

PROBLEM STATEMENT & EXAMPLE

If we were to consider the discussion above in terms of learning from weapon system failures, the discussion immediately becomes more complex. For instance, several key questions would need to be asked relative to the weapon system and our cost estimate:

1. Are we analyzing the reliability and aging of the weapon system as a single end item or are we analyzing the reliability of sub-systems and the subsequent impact on reliability and availability of the end item?
2. What point in the program lifecycle is the weapon system?
3. How often will the reliability of the system or sub-system be assessed?
4. What is the sustainment strategy for the weapon system or sub-system?

In this paper, we primarily focus on developing failure estimates dependent upon how questions 1 and 2 above are answered. We begin by addressing question 1 with a general overview of reliability theory.

RELIABILITY THEORY OVERVIEW

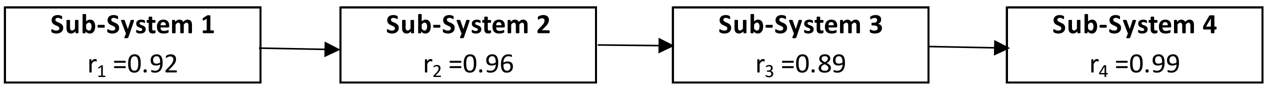
Before jumping right in to how reliability and system aging is evaluated and considered in the context of cost estimates, let us start by quickly reviewing some of the fundamental aspects of reliability theory. Ross provides a very general definition of reliability – "Reliability theory is concerned with determining the probability that a system, possibly consisting of many components, will function". Beyond this definition, it is also important to consider how the functionality of these components or sub-systems will affect the functionality of the primary system. The two primary systems to consider when assessing system reliability are series systems and parallel systems.

Series Systems

Series systems are primary systems that only function or perform satisfactorily if all of its components or sub-systems are functional. Below we consider a series system with *n* independent components or sub-systems where the reliability of the *i*th component is r_i. The primary system reliability, R_n, is defined as:

$$R_n = \prod_{i=1}^{n} r_n$$

Below we consider a primary system made up of four independent sub-systems in series, each with a sub-system probability of r_i:



$$R_n = (0.92) \times (0.96) \times (0.89) \times (0.99) = 0.778$$

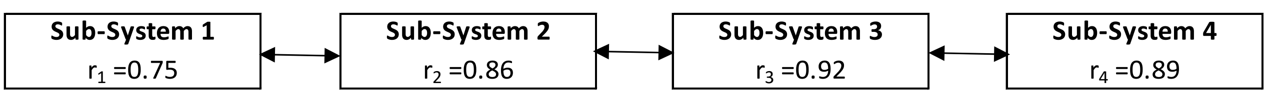
The reliability of the primary system is 77.8% when made up of these four independent sub-systems in series.

Parallel Systems

Parallel systems are primary systems that will function if and only if at least one of its components or sub-systems are functional. Below we consider a parallel system with *n* components or sub-systems where the reliability of the *i*th component is r_i. The primary system reliability, R_n, is defined as:

$$R_n = 1 - \prod_{i=1}^{n} (1 - r_n)$$

Below we consider a primary system made up of four sub-systems in series, each with a sub-system probability of r_i:



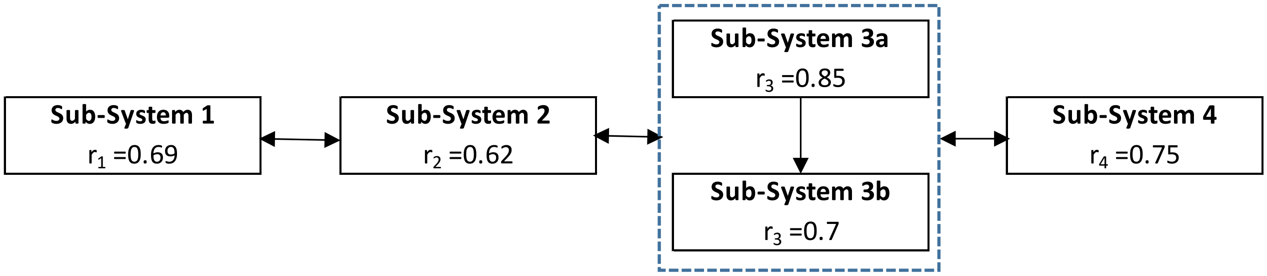
$$R_n = 1 - ((1 - 0.75) \times (1 - 0.86) \times (1 - 0.92) \times (1 - 0.89)) = 0.997$$

The reliability of the primary system is 99.7% when made up of these four sub-systems in parallel.

Hybrid Systems (Series-Parallel)

Primary systems also exist where the components or sub-systems consist of a combination of parallel and series systems. These systems are known as Series-Parallel systems. The reliability of such systems is calculated by using a combination of the series and parallel reliability equations.

Below we consider a primary system made up of four sub-systems in series, each with a sub-system probability of r_i. Note though that sub-system 3 consists of two sub-systems (3a and 3b) in parallel:



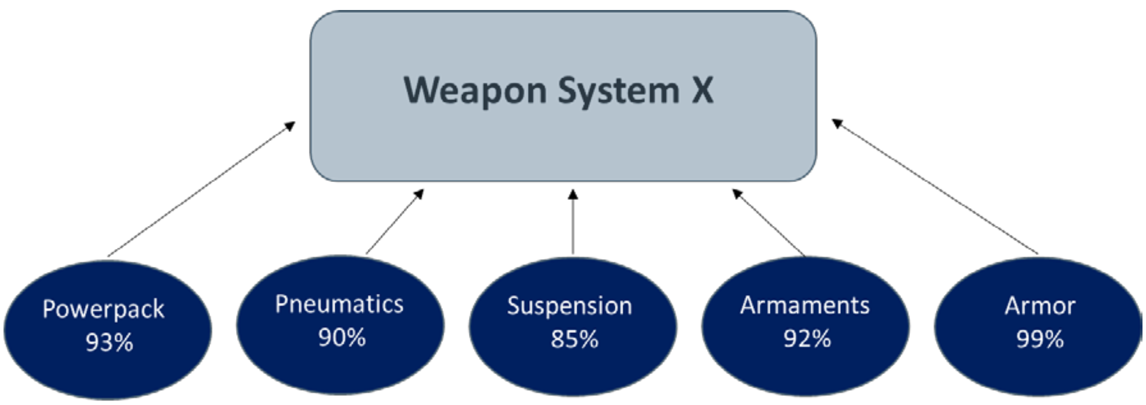
$$R_n = 1 - ((1 - 0.69) \times (1 - 0.62) \times (1 - (0.85 * 0.7)) \times (1 - 0.75)) = 0.988$$

The reliability of the primary system is 98.8% when made up of these four sub-systems in parallel and sub-system 3 in series.

Weapon Systems Application

For the purpose of this paper, it is assumed that the primary system in question will consist of sub-systems and/or components in series. For example, we will assume that our primary system (Weapon System X) consists of five major sub-systems as shown below in Figure 1:

Figure 1



93% * 90% * 85% * 92% * 99% = 64.8%

The overall reliability of Weapon System X is 64.8%

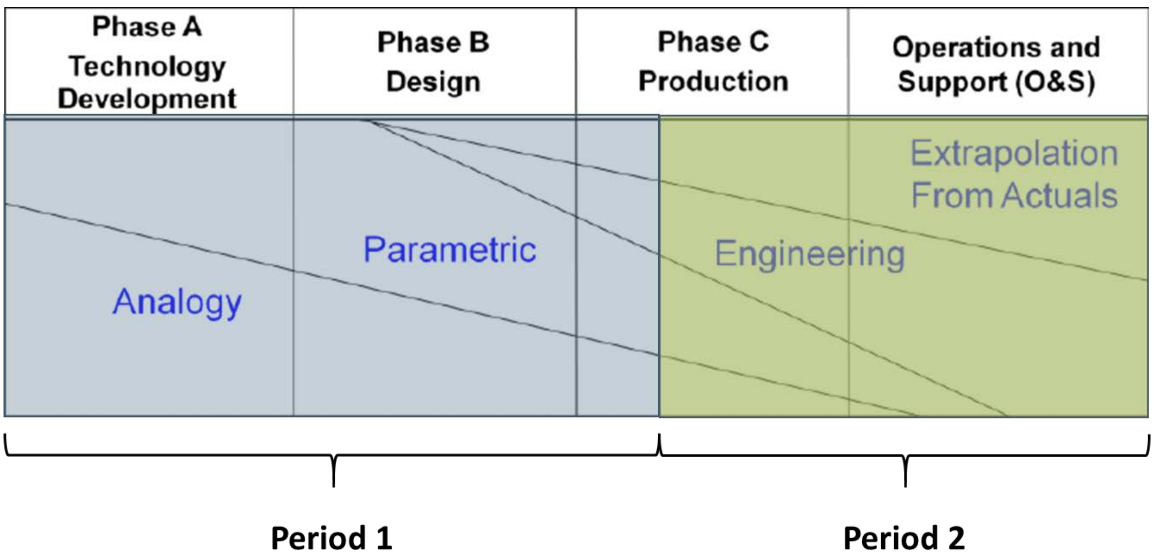
We will assume that as particular sub-systems or components fail, those systems are replaced in the field and the failed sub-system or component is returned for repair either to a field maintenance site or to a facility designated for repairing specific systems or sub-systems. The goal for the remainder of this paper is to develop estimating methodologies for predicting how each of these sub-systems will fail as a function of time over the weapon system's lifecycle.

PROGRAM LIFECYCLE IMPACT

As with any cost estimate, where we are in the program lifecycle strongly dictates the estimating methodology that should be used. Figure 2 illustrates that as time passes, more reliance can be placed on data from the program for which the estimate is being performed. The horizontal nature of the lines indicating estimating type is also worth noting. For example, there are points in the Design, Production and O&S phases where each of the four major cost-estimating techniques could be used.

For the purposes of this paper, we have chosen to draw a vertical line in the sand during the production phase and label the area to left as Period 1 and the area to the right as Period 2. In their research, Li, Wang and Zhou found that significant impacts in the estimating process occur once a system has experienced at least 20 failures. We will expand on this phenomenon later in this paper. For now, we will assume that this event occurs at some point following the commencement of production.

Figure 2



Source: Integrated Defense Acquisition, Technology and Logistics Life Cycle Management Chart, Defense Acquisition University, https://ilc.dau.mil

APPLICATION – ANALOGOUS SYSTEM DATA

In the area identified in Figure 2 as Period 1, it is assumed that there is not enough failure data available for the weapon system (or sub-systems) in question to be utilized for accurate and useful estimates. Instead, the estimate will be based on a combination of analogy and parametric analysis.

Before we can identify analogous systems to base our estimate on, we must consider what attributes a good analogous system should have relative to the one being estimated and where data for these attributes resides. Prior to identifying these attributes, let us define a good objective statement of what we are hoping to accomplish:

Objective: Identify systems or sub-systems that are comparable to Weapon System X and its sub-systems and have experienced failures that will be representative of those expected for Weapon System X and/or it sub-systems.

While general in scope, this objective states that we should be looking for systems that are comparable in physical specifications, but will also be utilized in a similar manner. Below are some specific areas of utilization to consider when selecting an analogous system or systems:

- Mission Objective – How will this system be used? Combat? Transport?
- Usage/OPTEMPO – Can a system that is used 16 hours per day be compared to a system that is used 16 hours per week? 16 hours per month?
- Fleet Size – Can a system with 10,000 units be compared to a system with 100 units? 100,000 units?
- Field Locations – Will the systems be used in the same conditions? Can we assume that a system will age the same in desert conditions versus cold, winter-like conditions?
- Maintenance Strategies – How will this system or subsystem be repaired? Will it always be repaired in the field? Will it be repaired at a maintenance facility? Will there be preventative, condition-based repair performed as well?

Based on these questions and the variables within, it is clear that picking an analogous system, much less the best analogous system, can prove to be a challenge. In the next section, we will begin a two-step process of identifying an analogous system and using that system's reliability profile to develop estimating parameters for our future system.

Identifying an Analogous System

When identifying an analogous system, we typically try to estimate costs for a future system by comparing a known or anticipated attribute of that system to a similar system that is further along in its lifecycle and has data more readily available for analysis. When trying to find an analogous system for aging and failure predictions, it is not as straightforward.

Let us present an approach by way of example. As discussed above, we are not only trying to find systems that have similar physical attributes, but we also want to identify systems that will be utilized in a similar manner. Let us start by assuming we are going to estimate the failures for a sub-system of a wheeled vehicle system with the following utilization attributes:

Weapon System Type: Wheeled Vehicle
Period of Performance: 10 Years
Sub-System Weight: 4,500 lbs.
System Location 1 Fleet Size: 10 Units
System Location 1 OPTEMPO: 50 Hrs/Month
System Location 2 Fleet Size: 12 Units
System Location 2 OPTEMPO: 50 Hrs/Month
System Location 3 Fleet Size: 15 Units
System Location 3 OPTEMPO: 50 Hrs/Month
Total Fleet Size: 37 Units

We have identified 10 potential analogous sub-systems to base our estimate on. Each of the ten sub-systems were used in wheeled vehicles that had OPTEMPOS of approximately 50 hours per month and were used in similar conditions as the sub-system we are estimating. For analogous estimates, data such as this will usually be found in databases maintained by each of military departments within DoD. For example, the Army houses data similar to this in the Operating and Support Management Information System (OSMIS). Table 1 below details the observed failures, mean failure time, standard deviation of failure time and the sub-system weight:

Table 1

	Failures Observed	Mean Failure Time (Hrs)	Std. Deviation	Sub-System Weight (lbs.)
Sub-System 1	1,500	854.8	334.5	4,275.8
Sub-System 2	1,441	846.4	314.7	4,077.0
Sub-System 3	1,435	854.7	392.7	4,324.8
Sub-System 4	1,386	898.8	348.1	5,899.1
Sub-System 5	1,367	864.1	358.4	4,923.5
Sub-System 6	1,342	865.7	319.2	5,259.3
Sub-System 7	1,322	868.9	405.7	4,678.3
Sub-System 8	1,304	858.1	356.9	4,488.7
Sub-System 9	1,280	878.7	445.0	4,834.8
Sub-System 10	1,271	907.3	339.4	6,118.1

The mean failure time referenced in Table 1 for *n* sub-system failures is defined as the mean time between sub-system integration with the primary system and the time of sub-system failure. While it would be easy to select Sub-System 8 as the best analogous system for our estimate due to the similar system weight, we first test whether or not sub-system weight is a good indicator of mean failure time by developing a scatter plot and ANOVA table of the proposed CER (Figures 3 and 4):

Figure 3

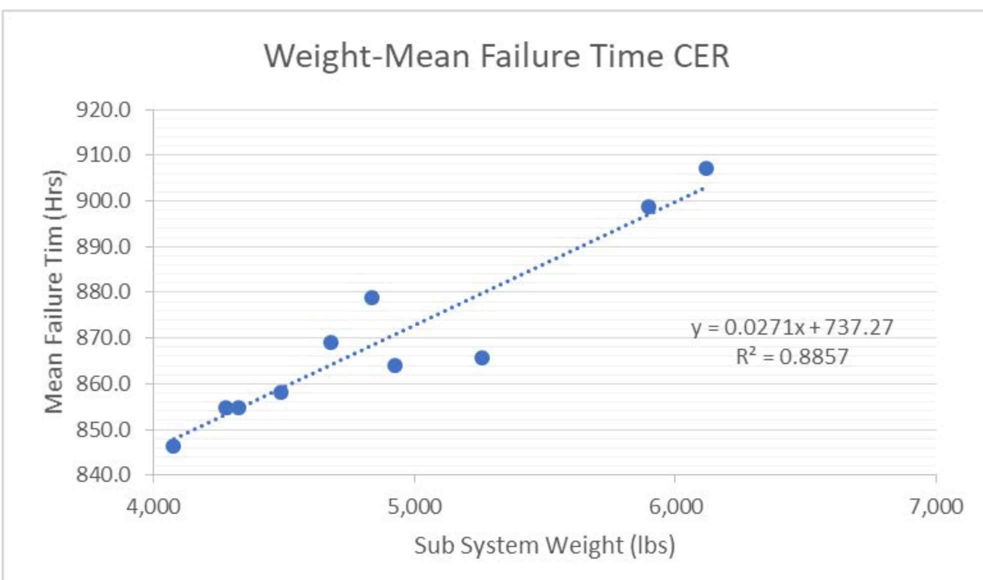


Figure 4

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.941111908
R Square	0.885691623
Adjusted R Square	-1.25
Standard Error	7.080845993
Observations	1

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	10	3107.884	310.7884	61.98612	#NUM!
Residual	8	401.107	50.13838		
Total	18	3508.991			

	Coefficients	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept							7.080845993	7.080845993
X Variable 1							0.885691623	0.885691623
X Variable 2							-1.6796E-185	1.6796E-185
X Variable 3							-1.6796E-185	1.6796E-185
X Variable 4							7.3136E-186	7.3136E-186
X Variable 5							-9.5322E-186	2.416E-185
X Variable 6							-9.5299E-186	2.4097E-185
X Variable 7							-9.5299E-186	2.4097E-185
X Variable 8							6.0344E-270	-1.5214E-269
X Variable 9	737.2730029	16.97547	43.43166	8.71E-11	698.1274917	776.418514	698.1274917	776.418514
X Variable 10	0.027103873	0.003443	7.873127	4.9E-05	0.019165268	0.035042478	0.019165268	0.035042478

Between the scatter plot and some of the key statistics from Microsoft Excel regression output table highlighted in Figure 4, it is clear that a statistically significant relationship exists between sub-system weight and the mean failure time of the sub-system. It is worth noting that the F Significance value returned was “#NUM!”. Excel will often return this error if the value is smaller or bigger than what excel can represent. Given the low p-values and strong R^2 value, we can assume that the value was too small for Excel to represent. Based on this, we can feel confident in selecting Sub-system 8 as a good analogous system to base our estimate on since it is so close in weight to the new system. However, just knowing the mean failure time will not help us in predicting *when* we can anticipate the failures. In order to deduce this information, we must consider the individual failures of Sub-System 8.

Aging Parameter Identification and System Model

Instead of listing out all 1,304 failures from Sub-System 8, we have fit the failure data to 17 different continuous distributions using statistical software. The parameters calculated for each distribution are in the far-right column of Table 2. We will discuss the mathematical technique(s) for calculating these parameters later in this paper. Table 2 also details the quality of the fit for this dataset to each distribution. Three different goodness of fit tests were performed on each distribution using statistical software. For each test, a test statistic was calculated and the distributions were ranked using that test. We then took an average ranking over the three tests to order our distributions in terms of goodness of fit. The Weibull distribution ranked first using all three tests and is the distribution that should be selected for modeling how failures will occur.

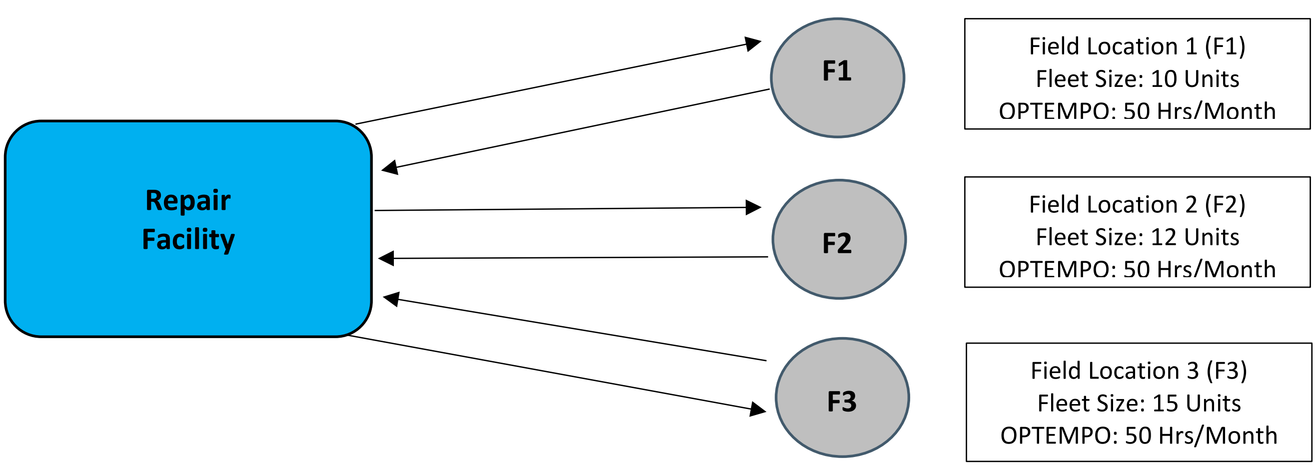
Table 2

	Kolmogorov Smirnov		Anderson Darling		Chi-Squared			
Distribution	**Statistic**	**Rank**	**Statistic**	**Rank**	**Statistic**	**Rank**	**Average Rank**	**Parameters**
Weibull	**0.01864**	**1**	**0.47908**	**1**	**5.1576**	**1**	**1.0**	**α=2.6061 β=964.53**
Beta	0.01899	2	3.2993	3	9.974	2	2.3	α_1=2.4187 α_2=3.3906 a=57.231 b=1960.8
Normal	0.03638	3	2.8918	2	14.398	3	2.7	σ=356.9 μ=858.12
Gamma	0.04901	4	5.9959	4	44.948	5	4.3	α=5.7811 β=148.44
Logistic	0.05384	5	7.2435	5	36.416	4	4.7	σ=196.77 μ=858.12
Triangular	0.0609	7	12.919	6	80.07	6	6.3	m=542.04 a=57.231 b=1960.8
Rayleigh	0.06536	8	17.77	9	98.855	7	8.0	σ=684.68
Gumbel Max	0.06632	9	13.266	7	106.33	9	8.3	σ=278.27 μ=697.5
Lognormal	0.06703	10	13.755	8	99.225	8	8.7	σ=0.48781 μ=6.6517
Uniform	0.058	6	255.21	14	N/A		10.0	a=239.96 b=1476.3
Cauchy	0.09732	11	22.639	10	166.85	11	10.7	σ=230.7 μ=829.48
Gumbel Min	0.10273	12	39.048	11	133.07	10	11.0	σ=278.27 μ=1018.7
Erlang	0.16685	13	87.864	12	214.22	12	12.3	m=5 β=148.44
Exponential	0.27922	14	196	13	1043.1	13	13.3	λ=0.00117
Pareto	0.4258	16	385.98	15	2965.3	14	15.0	α=0.41991 β=71.538
Chi-Squared	0.42065	15	7875.1	16	4461.6	15	15.3	ν=858
Student's t	0.9999	17	16252	17	1.39E+08	16	16.7	ν=2

After identifying the best-fit distribution and parameters for our estimate, we are ready to translate these into our predicted failures. Since we will be interested in when these failures will occur as well, we will summarize the predicted failures by year for the 10 year period of performance. In Figure 5, we introduce a basic system consisting of three field sites and one repair site. Some key assumptions and limitations of the model are as follows:

- System run time = 10 years
- Number of trials = 30
- Failure times are measured in hours
- System failures at the field sites are immediately replaced (i.e. no swap-out or integration time considered)
- Repair time and transportation time is not considered
- All sites begin with sub-systems at 0 hours
- Failures occur at each site following a Weibull distribution with shape (α) = 2.6061 and scale (β) = 964.53

Figure 5



After running 30 trials, the mean and standard deviation of the total sub-system failures for each of the ten years was calculated (Table 3).

Table 3

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10
Mean	9.4	26.8	26.3	25.7	25.1	26.7	25.2	27.3	25.3	25.9
Std. Dev.	2.6	4.0	4.3	3.5	3.5	3.5	3.1	3.9	3.5	3.6

While this prediction of sub-system failures will certainly assist in estimating operations and sustainment (O&S) costs for its parent system while we are in early phases of the program lifecycle, as data becomes readily available for the parent system and its sub-systems, we should expect estimates that are extrapolated from failure data for that particular sub-system will be far more accurate. In the next section, we will address when we should consider failure data for the new system and how it should be used to predict system (or sub-subsystem) aging and failures.

APPLICATION – EXTRAPOLATION FROM ACTUALS

As we mentioned earlier when discussing Program Lifecycle Impact, Li, Wang and Zhou found that significant impacts in the estimating process occur once a system has experienced at least 20 failures. Their research focused on statistical lifetime modelling for power transformers. By performing a sensitivity analysis using the relative root mean square error (RRMSE) as output, it was found that "correlations suggest that a minimum failure number of 20 may serve as an appropriate criterion for evaluating whether a set of lifetime data qualifies for Weibull lifetime modelling at a given accuracy level". The RRMSE was defined as:

$$\text{RRMSE} = \frac{\sqrt{SD(\theta)^2 + Bias\ (\theta)^2}}{\vartheta}$$

$$\theta = \text{ estimated result of the concerned parameter;}$$
$$\vartheta = \text{ true value of the concerned parameter}$$

We will use these findings to define the starting point for Period 2 from Figure 2 as the point in a system's or sub-system's lifecycle when 20 failures have occurred. We will now introduce an alternative method for estimating the number of failures during this period of the program lifecycle.

Introduction to Lifetime and Right-Censored Data

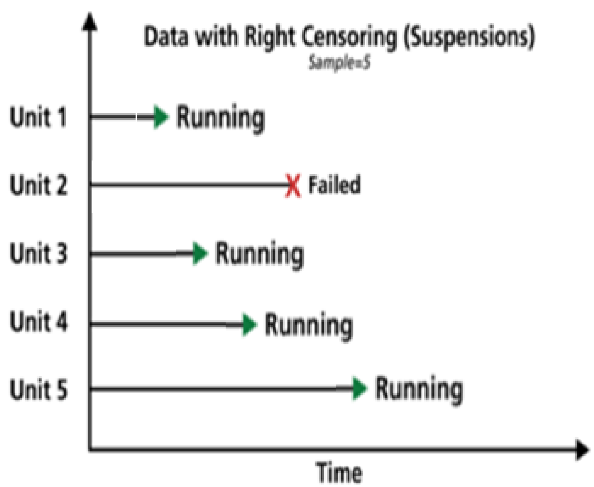
We will continue by analyzing a single sub-system for a weapon system made up of sub-systems in series. Once a sub-system has been fully integrated with its end item, the sub-system's O&S life begins and will continue until it experiences a distinct failure at some unknown time during the end item's estimated useful life**.** At any instance during an end item's O&S life, sub-systems can exist in two states, failed or operational. During these states, sub-system age is usually measured in units that are associated with failure, such as the number of miles driven on an engine or the number of rounds fired out of a barrel. When data is collected on fielded sub-systems, two questions need to be answered:

1) Is the sub-system operational, yes or no?
2) If yes, what was its age at the time of the data collection, if no, what was its age at time of failure?

For sub-systems whose times of failure are known, the data is classified as failure data. However, when data is collected on the age of fielded sub-systems that are still operational, the data is classified as right-censored data. Understanding the difference between failure and right-censored data is paramount to creating the most accurate aging profile and predictive failure model for the sub-system.

To better understand right-censored data, imagine a scenario where a lightbulb company wants to know the average time to failure for its incandescent bulb. The company can conduct an experiment where they measure the lifetime of a large sample of incandescent bulbs. During the experiment, data is collected for each bulb to answer the two critical questions mentioned above: Whether or not it had failed at the time of data collection and the age of the bulb at that time or at the time of failure. We will run through a series of examples to illustrate how the sample bulbs age over time. Shortly after the beginning of the experiment, nearly all the bulbs are working and the majority of units are still operational (right-censored) as shown in Figure 6:

Figure 6



Unit ID	Operational	Age (Hrs)
1	Yes	1
2	No	3
3	Yes	2
4	Yes	3
5	Yes	5

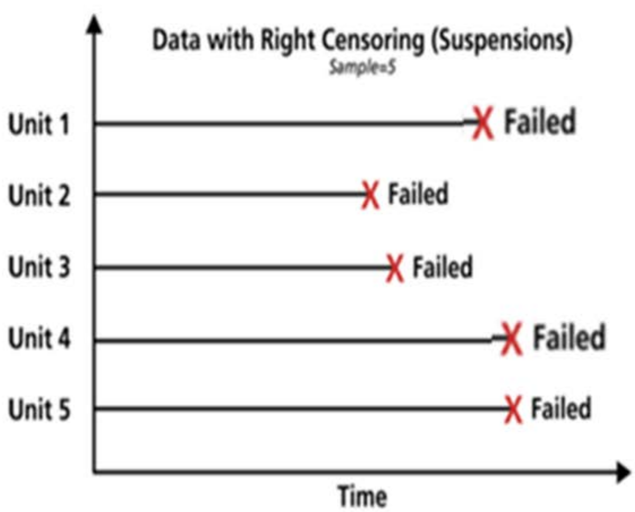
As time progresses, bulbs will fail one-by-one (Figure 7) until all the bulbs have failed (Figure 8). Figure 7 illustrates how the proportion of failed and operational sub-system units become similar to one-another as the sub-system as a whole progressively ages.

Figure 7

Unit ID	Operational	Age (Hrs)
1	Yes	5
2	No	3
3	No	3.5
4	Yes	7
5	No	6

Figure 8 illustrates how the proportion of failure data continues to grow as the sub-system ages and eventually reaches 100%, as they all reach the end of their useful lives.

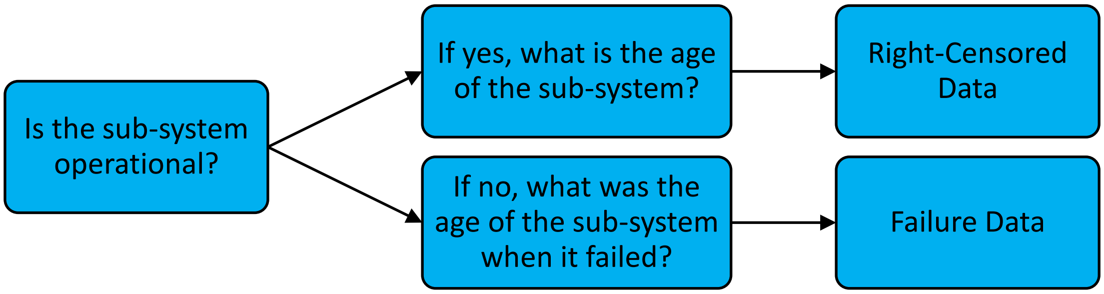
Figure 8



Unit ID	Operational	Age (Hrs)
1	No	6
2	No	3
3	No	3.5
4	No	8
5	No	6

The bulbs that have failed and the bulbs that have yet to fail can be classified as failure data and right censored data respectively (Figure 9). This experiment also demonstrates how the proportion of data that is right-censored will continue to decrease as more units fail and become failure data as the sample ages and experiences more failures.

Figure 9



Parameter Identification

As was done with the analogous model for Period 1, the goal for Period 2 is to create an aging model for a particular sub-system that will help stakeholders plan for future demands and estimate the associated costs. In order to achieve this goal, the probability of failure occurring at any instance during a sub-system's life needs to be known. This is accomplished by using a probability density function (PDF) and its parent function, a cumulative density function (CDF). For every statistical distribution used to model failures, the probability (p) of observing a failure at time "t" for any unit during the system's useful life, is a function, F(t), of the given distribution's parameter "θ". An example of a PDF would be the normal distribution with its parameters being its mean (μ) variance (σ^2) and time of failure (x):

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}}$$

The correct identification and calculation of these parameters for the lifetime data being analyzed is crucial to the most accurate modeling of the sub-system's aging profile. Unlike the analogous model, where we were able to fit a complete population of failures to a distribution and derive parameters, deriving parameters for a system where only a portion of the population has failed can be a little more complex.

Maximum Likelihood Estimation

Traditionally, the most accurate way to correctly model the parameters for the PDF/CDF that governs the aging profile for the sub-system was to perform maximum likelihood estimation (MLE) on observed failure data. MLE calculates the most accurate set of parameters for a PDF by maximizing the probability of observing each failure in the data according to the likelihood equation. The likelihood equation (figure 11) quantifies the probability of observing the failures actually occurring at their given hours according to the estimated parameters θ, of the distribution:

$$L(\theta) = C \prod_{i=1}^{n} L_i(\theta; \mathrm{data}_i)$$

To find the parameter values that maximizes this probability, the log of the likelihood equation is taken, to get the log likelihood equation:

$$\mathcal{L}(\theta) = \log[L(\theta)] = \sum_{i=1}^{n} \mathcal{L}_i(\theta)$$

The derivative of the log likelihood equation is taken and the result is set to zero to solve for the parameters, where the solutions to the equation are the parameters:

$$\frac{\delta \mathcal{L}(\theta)}{\delta(\theta)} \mathcal{L}(\theta) = 0$$

In practice, the true value of these parameters cannot be known because only a sample of the total fleet of sub-systems is recorded in the failure data. When MLE is used with samples of the weapon system's units and not the entire population, the calculated values of the parameters from the MLE are known as maximum likelihood estimates (MLEs). To provide an example of an MLE, recall the example of the incandescent lightbulb study. Once every bulb in the sample has failed, the company can perform MLE to identify a PDF that will be used to model the failure of the entire population of the incandescent bulbs.

Unfortunately, many DoD weapon sub-systems have been engineered to survive for many years, if not decades, unlike everyday items like incandescent lightbulbs. Therefore, an adaption of MLE must be used, as it is impractical to wait for a statistically significant number of sub-systems to fail until an O&S aging profile is created. Imagine if the lightbulb company conducted the same aging experiment with their LED bulbs (whose lifetimes can be up to ten times longer) as they did with their incandescent bulbs. Only a small percentage of their sample would have experienced a failure before the company would need to create estimates of the aging profile. Their MLE aging profile would not be an accurate representation of the entire population of bulbs because the overwhelming majority of the data collected, the right-censored data of units still operational, was not considered in the analysis.

In order to produce an aging profile earlier in a sub-system's O&S lifetime, the entire spectrum of lifetime data needs to be analyzed, both failure data and particularly, right censored data. The adaption of traditional MLE to include right-censored data is simply known as maximum likelihood estimation with right-censored data (MLE R-C). Unlike traditional MLE, MLE R-C takes advantage of the data that has yet to fail to build a more accurate aging profile and predictive failure model. MLE R-C works exactly the same as traditional MLE (take the derivate of the log of the likelihood equation, set it to zero and solve) except for one key step. Rather than maximizing the likelihood of observing all the failures, MLE R-C likelihood equation maximizes both the likelihood of observing all the failures observed *and* the likelihoods of all the right censored data points not failing up until their current age. To denote the failure or operational status of a unit, sub-systems are coded δ = 1 for failures and δ = 0 for operational units:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \{f(t_i;\boldsymbol{\theta})\}^{\delta_i} \{1 - F(t_i;\boldsymbol{\theta})\}^{1-\delta_i},$$

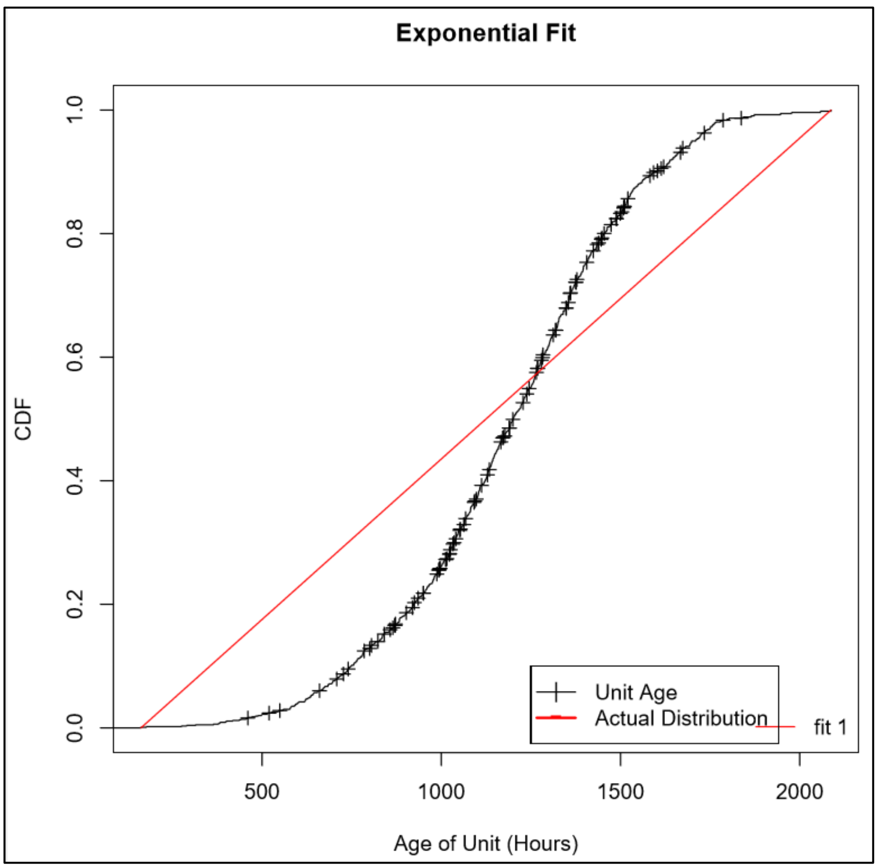
Dynamic Failure Distributions and Best Fit Identification

In order to create an aging profile and predictive failure model for a particular sub-system over the course of its life, it is important to reassess the failure distribution or, at the very least, the optimal parameter values for a distribution throughout the O&S phase. For example, the distribution and subsequent parameters identified as best fit in the first year of O&S may be different than the best fit distribution and parameters in the second year and the third year and so forth. Therefore, to ensure the aging model is the most accurate representation of the sub-system, a wide variety of probability distributions must be fit via MLE or MLE R-C to the lifetime data so each model can be evaluated for accuracy. Common distributions evaluated for lifetime data include the normal, exponential, log-normal and Weibull.

In order to assess which distribution is the best fit for the data at a particular point in time of the O&S phase, each distribution that the data has been fit to must be put through goodness of fit tests, which mathematically quantify the PDF that most accurately represents the observed lifetime data. Common goodness of fit tests include the Anderson-Darling test, the Shapiro-Wilks test and the Lilliefors tests. Additionally, plotting the observed data against the theoretical values of the estimated aging profile will illustrate how well of a fit (or poor of a fit) the PDF is for the data depending on how closely the data points align with one-another.

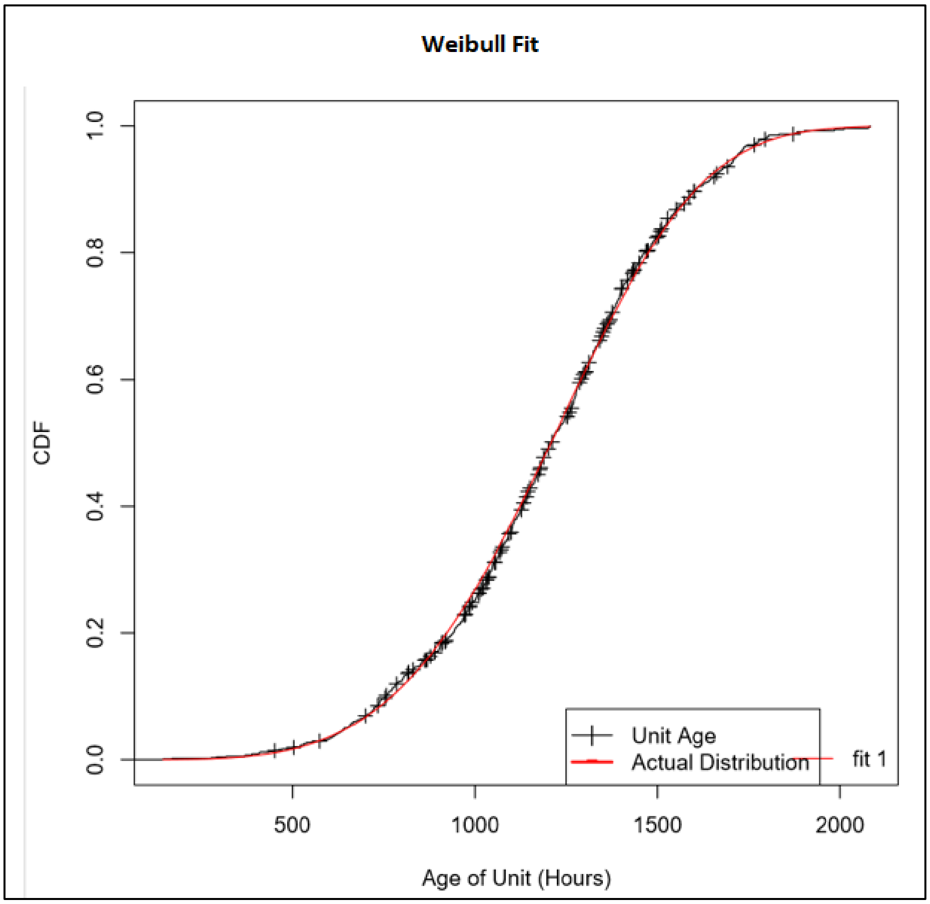
Weapon System X's sub-system distribution of lifetime data had statistically significant evidence at a 5% significance level that it did not follow a normal distribution according to both the Shapiro-Wilk's and Anderson-Darlings tests. Additionally, we plotted the observed lifetime data against theoretical Weilbull and Exponential distributions to evaluate the fits visually. As seen in the Figure 10 below, the data does not fit the Exponential distribution.

Figure 10



However, the lifetime data does follow the theoretical Weibull data (as seen in Figure 11), thus confirming that the sub-system failures can be accurately modeled by a Weibull PDF.

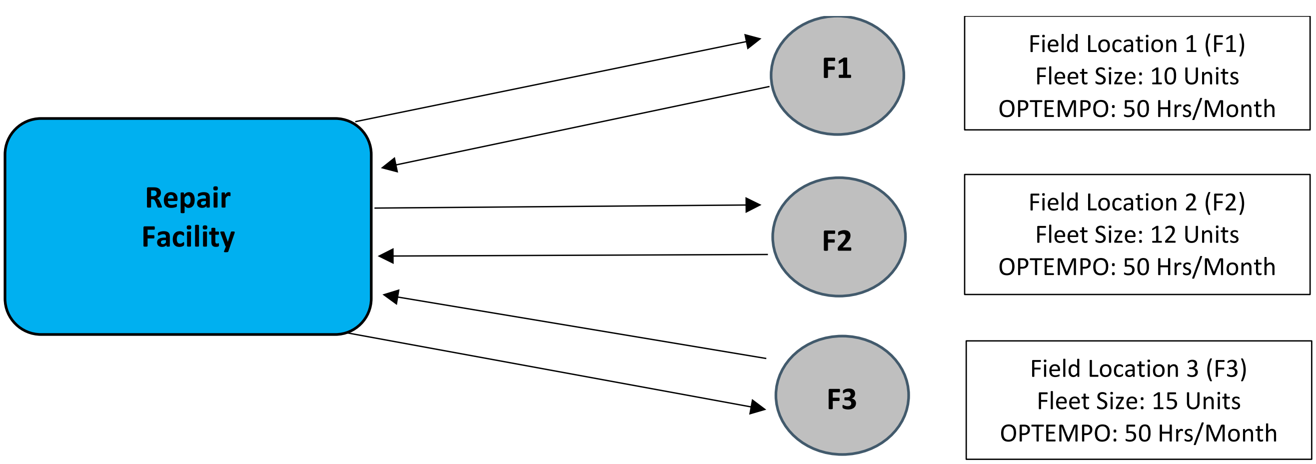
Figure 11



SIMULATION OF SYSTEM

As was the case with the analogous system model, we will now analyze data from a sub-system that has reached Period 2 as defined in Figure 2. For simplicity's sake, we will again assume we are looking at Weapon System X, but a different subsystem (Figure 12). Given this information, each our field site fleet sizes and OPTEMPOS will be the same as the analogous example.

Figure 12



The failure data for this sub-system was evaluated at different stages of the sub-system's period of performance (10 years). The data was broken into nine groups such that the first group had 10% failed and 90% operational units and last group had 90% failed and 10% operational. The data was evaluated in 10% increments between these two extremes. At each of these increments, a Weibull probability density function was fit to the data using MLE R-C. The subsequent parameters were then used to run 30 simulations that tracked total failures over a ten-year period (Table 4). All other assumptions from the analogous model were applicable as well.

Table 4

Total Failures By Year for All Sites with MLE R-C (30 Trials)																		
Censoring Rate	**90%**		**80%**		**70%**		**60%**		**50%**		**40%**		**30%**		**20%**		**10%**	
Shape	**4.5**		**4.08**		**4.02**		**4.19**		**4.18**		**4.21**		**4.22**		**4.28**		**4.2**	
Scale	**2140**		**1884**		**1711**		**1585**		**1503**		**1438**		**1387**		**1344**		**1306**	
Mean ($\bar{x}$), Std. Dev. (s)	**$\bar{x}$**	**s**	**$\bar{x}$**	**s**	**$\bar{x}$**	**s**	**$\bar{x}$**	**s**	**$\bar{x}$**	**s**	**$\bar{x}$**	**s**	**$\bar{x}$**	**s**	**$\bar{x}$**	**s**	**$\bar{x}$**	**s**
Year 1	0.1	0.3	0.6	0.8	0.4	0.6	0.5	0.8	0.9	0.7	0.9	0.8	1.2	1.0	0.9	0.8	1.6	1.3
Year 2	2.4	1.8	5.2	2.4	7.8	2.4	8.9	3.0	12.0	2.5	13.7	3.2	14.3	2.8	16.0	2.2	18.4	2.9
Year 3	11.9	2.8	15.4	3.1	18.1	3.6	21.1	3.3	20.7	2.5	21.8	3.4	21.9	3.0	22.2	2.3	20.6	2.6
Year 4	16.4	3.5	15.9	3.4	15.0	3.1	12.9	3.4	13.3	2.9	13.3	3.1	14.7	2.9	14.8	2.5	17.5	3.0
Year 5	9.7	2.9	10.9	3.2	11.6	2.4	15.0	3.4	17.6	2.3	18.3	3.5	18.6	2.9	21.0	2.9	20.5	2.7
Year 6	8.9	2.7	13.0	2.7	16.3	2.4	17.0	3.1	15.8	3.0	16.7	3.1	16.2	3.2	16.1	3.1	17.0	2.7
Year 7	13.2	2.9	13.5	2.5	13.6	3.1	14.0	2.9	15.8	2.9	17.0	3.1	18.4	2.8	19.3	2.9	19.8	3.3
Year 8	12.0	3.1	12.7	2.4	13.2	2.3	15.1	3.6	17.0	2.5	17.2	3.6	17.5	3.0	17.7	2.9	18.1	3.3
Year 9	10.8	2.6	13.5	3.3	15.6	3.1	16.8	3.8	15.4	3.1	16.4	3.9	17.5	3.3	18.3	2.6	19.0	2.9
Year 10	11.6	3.0	12.0	2.4	14.4	2.9	14.5	2.4	16.2	2.8	17.3	3.1	18.5	3.5	18.4	3.3	18.6	3.0

Notice that early in the life of the sub-system (high censoring rates), that the difference in estimated values for the parameters are quite different, however, as the amount of failures observed grows closer to the number of operational units, the parameter estimates become closer and closer until they differences are nearly statistically insignificant. In order to illustrate the differences that can arise when right-censored data is not used in estimating the failure distribution, Weibull probability distribution functions were also fit at each data grouping, but only on the failure data. The subsequent parameters were then used to run 30 simulations that tracked total failures over a ten-year period (Table 5). All other assumptions from the analogous model were applicable as well.

Table 5

Total Failures By Year for All Sites with MLE (30 Trials)																		
Censoring Rate	**90%**		**80%**		**70%**		**60%**		**50%**		**40%**		**30%**		**20%**		**10%**	
Shape	**4.8**		**4.04**		**4.05**		**4.34**		**4.21**		**4.24**		**4.16**		**4.21**		**4.19**	
Scale	**1292**		**1268**		**1264**		**1264**		**1270**		**1273**		**1276**		**1281**		**1272**	
Mean ($\bar{x}$), Std. Dev. (s)	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
Year 1	1.2	0.7	1.6	1.2	1.8	1.4	1.5	1.2	1.4	1.1	2.0	1.3	1.1	1.1	1.0	1.0	1.7	1.3
Year 2	17.5	3.0	18.9	2.8	18.4	3.0	19.1	3.8	18.1	3.5	17.5	3.7	20.1	2.3	18.1	3.2	18.2	3.0
Year 3	21.4	3.1	20.9	3.0	20.9	3.2	20.8	3.6	21.2	3.1	21.5	3.6	19.7	2.4	21.8	2.7	21.1	3.2
Year 4	17.2	2.8	18.5	3.8	19.4	2.8	18.5	3.6	17.4	3.0	18.1	3.3	18.2	3.8	17.6	2.9	17.9	3.1
Year 5	20.7	2.3	20.6	3.7	19.2	2.7	19.9	3.2	20.1	3.7	20.0	3.3	20.5	3.8	20.7	3.6	20.3	2.5
Year 6	17.9	2.3	18.4	3.0	19.2	3.2	18.6	3.0	18.9	3.2	19.1	3.1	17.8	3.3	17.2	3.0	18.4	3.2
Year 7	19.2	3.0	19.6	2.9	19.7	2.9	20.4	3.2	19.6	3.2	19.5	2.7	21.5	4.0	19.3	3.9	19.4	3.2
Year 8	18.5	2.9	19.2	2.7	19.7	3.4	19.2	3.2	18.9	2.9	18.7	3.2	18.0	3.2	18.6	3.3	19.4	3.4
Year 9	18.8	3.0	19.7	2.6	18.8	3.3	19.4	2.8	19.4	3.1	19.2	2.8	19.8	2.8	19.0	2.7	18.9	2.7
Year 10	18.7	2.9	18.2	2.4	19.9	3.5	19.1	3.1	19.4	3.5	19.1	3.5	18.9	3.3	19.4	2.7	19.1	3.0

To quantify the difference in failure distributions that can be created when using traditional MLE (Table 5) versus MLE R-C (Table 4), we ran two-tailed difference of mean t-tests with unequal variances to see if these aging profiles' mean predicted number of failures over a ten year period of performance are statistically significantly different.

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Figure 13

Data is 60% Censored

t-Test: Two-Sample Assuming Unequal Variances

	MLE-RC	*MLE*
Mean	13.56333	17.63333333
Variance	30.82974	32.80938272
Observations	10	10
Hypothesized Mean Difference	0	
df	18	
t Stat	-1.61336	
P(T<=t) one-tail	0.06203	
t Critical one-tail	1.734064	
P(T<=t) two-tail	0.124059	
t Critical two-tail	2.100922	

Figure 14

Data is 70% Censored

t-Test: Two-Sample Assuming Unequal Variances

	MLE-RC	*MLE*
Mean	12.58667	17.69
Variance	26.2319	31.75359
Observations	10	10
Hypothesized Mean Difference	0	
df	18	
t Stat	-2.11931	
P(T<=t) one-tail	0.024114	
t Critical one-tail	1.734064	
P(T<=t) two-tail	0.048228	
t Critical two-tail	2.100922	

As seen in Figures 13 and 14, the mean predicted number of failures is no longer statistically significantly different between the MLE-RC and MLE failure models sometime in the weapon system x's life where its sub-system has between

60% and 70% right-censored data. This to say that there is no data to suggest the mean number of failures predicted by the two models are statistically different once the lifetime data becomes roughly 60%-70% censored (e.g. one failure for every two operational units). This does not imply that the predicted number of failures becomes identical to one-another, but given the natural variability in the weapon system's failures, the differences between MLE and MLE R-C are not significant. The impact of identifying when this occurs can help the analyst decide when the necessity of collecting and integrating right-censored data into the estimate lessens. This is not guaranteed to happen at this point for every weapon system, but it is an analysis that should be performed.

COST ESTIMATE IMPACT

Up to this point, we have primarily addressed how many of a particular item will fail and when. In order to capture the impact of these failures from a cost perspective, the results would then need to be multiplied by an estimated cost per failure and/or the subsequent repair cost. For our MLE examples above, we will assume the cost of repair for each sub-system failure is $150,000. As was the case with the analogous estimate, OSMIS can be used to estimate the average cost of repair. In Table 6, we present the difference in the estimated mean failure totals, by year and censoring rate, between the MLE and MLE R-C methods as well as the cost impact:

Table 6

Difference in Estimated Failures (μ_{MLE} - $\mu_{MLE\ R\text{-}C}$)									
Censoring Rate	**90%**	**80%**	**70%**	**60%**	**50%**	**40%**	**30%**	**20%**	**10%**
Year 1	1.0	1.0	1.4	1.0	0.6	1.1	-0.1	0.1	0.1
Year 2	15.2	13.6	10.6	10.2	6.0	3.8	5.8	2.1	-0.2
Year 3	9.5	5.5	2.8	-0.3	0.4	-0.3	-2.2	-0.4	0.5
Year 4	0.8	2.6	4.4	5.6	4.1	4.8	3.5	2.8	0.4
Year 5	11.0	9.7	7.6	4.9	2.6	1.7	1.9	-0.3	-0.2
Year 6	9.0	5.4	3.0	1.6	3.1	2.4	1.6	1.1	1.4
Year 7	6.0	6.1	6.0	6.4	3.8	2.5	3.1	0.0	-0.4
Year 8	6.5	6.4	6.5	4.0	1.8	1.5	0.5	0.9	1.3
Year 9	8.0	6.2	3.2	2.6	4.0	2.8	2.3	0.7	-0.1
Year 10	7.1	6.2	5.5	4.6	3.3	1.8	0.4	1.0	0.5
Net Difference	74.1	62.8	51.0	40.7	29.7	22.1	16.8	8.1	3.3
Cost Differential	$11,115,000	$9,420,000	$7,655,000	$6,105,000	$4,460,000	$3,310,000	$2,520,000	$1,210,000	$490,000

In other words, if when 10% of the fleet had failed and we only considered the failure data in our estimate, we would have overestimated by $11.1M for this one sub-system alone. As more failures occur and additional failure data is collected, the censoring rate decreases as does the risk of over-estimating costs from using traditional MLE as opposed to MLE R-C.

MODEL LIMITATIONS AND OTHER CONSIDERATIONS

As with most models and studies, this exploration into the consideration of right-censored data and how to apply it in cost estimates is not without its limitations as well as peripheral topics to consider.

Model Limitations

As we noted above, the first limitation of this model is that it only predicts the failures of the system or sub-system. If we were to build out the scope and capabilities of the simulation, several other areas of cost impact could be tracked. If we were to also consider other variables in the system such as repair time, transportation time and swap-out time, we would be able to make informed decisions regarding the fleet size of the sub-system. For example, if our simulation resulted in excessive idle and/or wait time in the field for the primary weapon system, we may determine that the fleet size is too small. Another solution may be to add additional resources at the repair site in order to increase throughput. Conversely, our simulation may also result in our model predicting that our system will be overly saturated with spare sub-systems. This could help inform the level of production and/or maintenance that needs to take place throughout the period of performance. The last major limitation of this model is timeliness. In particular, a large amount of time may pass before enough failures have occurred for us to utilize the MLE R-C technique.

A real world limitation of implementing a model such as this is data collection. The models used in this paper assumed only three field locations. In reality, weapon systems are often fielded throughout the world at a wide variety of locations and used for many different mission objectives. The challenge of collecting, storing and normalizing quality data, especially right-censored data, from many different locations is often a very big obstacle. For example, the chances that any two field locations will have the exact same failure distributions and parameters is miniscule. In reality, the methodologies implemented throughout this paper must be performed at all locations where the system or sub-systems will be utilized and failures will occur.

Other Considerations

This paper was written in order to initiate thought and discussion on how best to estimate system aging and failures at various points of the program lifecycle. In doing so, we have touched on a myriad of topics ranging from advanced statistical analysis to discrete-event-simulation. Several additional topics could and should be explored as part of this research.

In addition to tracking failure data and right censored data, another key piece of information that could be tracked is the reason for failure. Most systems or sub-systems typically have a finite list of failure modes or reasons for failure. Tracking the reasons for failure and frequency with which they occur in addition to the hours at the time of failure would allow us to add another facet to the simulation. Returning to our system from Figure 12, let's assume that after a certain amount of time we have observed the following number of failures at Field Site 1 that can be classified into one of five failure modes for this sub-system:

Table 7

	Number of Failures	P(Failure Mode)	Material Repair Cost	Labor Repair Cost
Failure Mode 1	16	13.4%	$ 178,000	$ 15,000
Failure Mode 2	41	34.5%	$ 162,000	$ 12,500
Failure Mode 3	14	11.8%	$ 123,000	$ 20,000
Failure Mode 4	15	12.6%	$ 200,000	$ 40,000
Failure Mode 5	33	27.7%	$ 95,000	$ 15,700

We could then take our simulation one step further by saying that each time a failure occurs, it does so because of one of the five failure modes following the probability mass function in Table 7. Having this additional information as well as the mean repair material and labor cost would not only help in providing additional accuracy to our estimate, it could also help in making several important programmatic decisions. For instance, in looking at Failure Mode 4, it is by far the most expensive repair for both material and labor. Knowing this information and its impact on driving cost up, the program manager and engineers may work on Engineering Change Proposals (ECPs) to help minimize the occurrence of this failure mode or perform trade studies to find less expensive repair strategies while not sacrificing performance.

Another key concept to consider is condition based maintenance. For instance, there may be a pre-defined maintenance package for a system that fails that is dictated by the number of hours or miles it has been in service. However, this package may or may not include the parts and repair specifications to address the reason that the system has failed. For example, we might have a system that is consistently failing between 100-150 miles due to transfer case issues. However, maintenance of the transfer case is not scheduled unless the system has greater than 500 miles. As these systems return, they will still receive the preventive maintenance for having between 100-150 miles and it will receive the corrective transfer case maintenance. If these occurrences continue, we might make one of the following decisions:

1. Perform additional engineering research into why the transfer cases are failing earlier than expected and develop an ECP
2. Include transfer case maintenance in the 100-150 mile preventative maintenance package.

By tracking the reason for failure and the Bill of Materials (BOM) for each maintenance package, optimal maintenance packages can be defined for systems that fail at various points of their useful life.

CONCLUSIONS

As discussed throughout this paper, where a weapon system is during its program lifecycle will strongly dictate the best methodology for estimating how that weapon system and its sub-systems will age and fail. Before performing such estimates, it is imperative that he estimator have a complete and thorough understanding of the weapon system's maintenance strategy. During earlier phases of the program lifecycle, it is just as important to find analogous systems that have similar utilization profiles as it is it is to find systems with comparable physical specifications. As a system enters into the O&S phase, the accuracy of the aging model can be improved by considering the right-censored data. The rate of improvement is defined by the censoring rate of the data.

REFERENCES

1. Ross, Sheldon M. (1997). *Introduction to Probability Models (6th ed.)* (pp. 475). San Diego, CA: Academic Press.
2. Li, C., Wang, Z., Zhou, D. (2013) Data Requisites for Transformer Statistical Lifetime Modelling—Part I: Aging-Related Failures. *IEEE Transactions on Power Delivery, Volume 28 (No. 3),* Pages 1750-1757.
3. Lawless, Jerald F. (2002). Statistical Models and Methods for Lifetime Data (2nd ed.). Hoboken, NJ: Wiley.

BIOGRAPHY

Pat McCarthy, a Project Manager with Technomics, has over 15 years of experience in Federal and private industry performing cost analysis and Industrial Engineering studies. He has previously worked as a Cost/Price Analyst and Team Lead with Booz Allen Hamilton and the Army Contracting Command and as a Senior Industrial Engineer and Team Leader with General Dynamics Land Systems. Pat is CCEA certified and holds B.S. and M.S. degrees in Industrial Engineering from Purdue University.

Alex Krencicki performs cost estimation and operations research for Booz Allen Hamilton as a Senior Consultant. Previously, he has gained professional experience in data analysis working in the health insurance industry as well as in automotive consulting. Alex holds a B.S. in Statistics and a minor in Financial Mathematics from the University of Michigan and is currently working on his PCEA certification.