Statistical Methods for Combining Information: Stryker Family of Vehicles Reliability Case Study

Rebecca Dickinson
Department of Statistics, Virginia Tech, Blacksburg, VA

Laura Freeman and Bruce Simpson
Institute for Defense Analyses, Alexandria, VA

Alyson Wilson
North Carolina State University, Raleigh, NC

Problem: Reliability is an essential element in assessing the operational suitability of Department of Defense weapon systems. Reliability takes a prominent role in both the design and analysis of operational tests. In the current era of reduced budgets and increased reliability requirements however, it is challenging to verify reliability requirements in a single test.

Approach: This paper describes the benefits of using parametric statistical models to combine information across multiple testing events. Both frequentist and Bayesian inference techniques are employed, and they are compared and contrasted to illustrate different statistical methods for combining information. We apply these methods to data collected during the developmental and operational test phases for the Stryker family of vehicles.

Results: We show that when we combining information across two test phases for the Stryker family of vehicles, reliability estimates are more accurate and precise than those reported previously using traditional methods that use only operational test data in their reliability assessments.

Key Words: Combining Information; Defense Acquisition; Exponential Distribution; Reliability; Weibull Distribution

PROCESS DESCRIPTION

In today's data-driven society, we often find ourselves with many sources of data that, when looked at collectively, may paint a different picture from what emerges when those data are analyzed in isolation. Most statistical analysis methodologies, however, focus on methods for one sample of data. The medical field has pioneered meta-analysis, which is typically associated with studies in which several clinical trials are combined into a single analysis.

In the Department of Defense, data are often collected over several phases of tests. A primary initiative across the Department of Defense (DoD) testing and evaluation community has been to

integrate testing through collaborative test planning and execution among all test agencies, resulting in data that can be shared by all. However, no one has yet capitalized on the knowledge that can be gained when one properly combines information across all of these test venues. Anderson-Cook (2009) highlights these potential gains, stating, "*If we have multiple datasets that are individually insufficient to answer the question of interest, then combining them and incorporating engineering or scientific understanding of the process should allow us to extract more from that collection of data compared to just looking at the pieces alone.*" There are many challenges that have limited the use of all test data, including limitations on data sharing, stark differences in how tests are conducted between test phases, and differences in the types of data collected by test phase. Reliability, however, has been a high priority in the testing of military systems for several years now. This emphasis has resulted in increased attention to capturing reliability data consistently across all stages of testing.

Reliability is the probability that a system will perform its intended function under appropriate operating conditions for a specified period of time. It is an essential component in the assessment of the operational suitability of many major defense systems; our armed forces need weapon systems that are available for combat when needed, reliable enough to accomplish their missions, operable by service personnel, and have a reasonable logistics burden. The current analyses employed by the Director Operational Test and Evaluation (DOT&E) and many of the services' operational test agencies use only operational test data in assessing operational reliability. Operational test (OT) data are collected under a limited set of test conditions—specifically, under conditions that replicate, as much as possible, actual "in the field" use. Using only OT data ensures that the results will be representative of the reliability under "in the field" conditions, but, in doing so we are discarding valuable information from previous testing on system reliability.

The National Academies Press (NAP) has published a series of studies conducted by the National Research Council through the Committee on National Statistics on defense acquisition, testing, and evaluation of defense systems. These reports are in direct response to requests made by the DoD and include the following titles: *Statistics, Testing, and Defense Acquisition* (1999); *Improved Operational Testing and Evaluation: Phase I Report* (2003); *Improved Operational Testing and Evaluation: Phase II Report* (2004); *Testing of Defense Systems in an Evolutionary Acquisition Environment* (2006); and *Industrial Methods for Effective Development and Testing of Defense Systems* (2012). These reports have repeatedly encouraged the use of all relevant information in both the design and evaluation of operational tests. They stress that state-of-the-art statistical methods for combining information should be used when appropriate to make testing and the associated evaluations as cost-efficient as possible (NAP 1998). Additionally, they note that the use of all relevant information can improve test design and estimation (NAP 2004). In combining information, not only are the estimates likely to be more accurate, but their uncertainty will also be estimated more precisely.

A 2004 NAP report focuses specifically on combining test information for the Stryker family of vehicles (FOV). However, their recommendations are general and broadly relevant. The panel highlights that, "*experience with the Stryker/SBCT test and evaluation shows that operational testing alone often does not include enough data to permit definitive conclusions. It is therefore necessary to also use data from developmental testing, training, and field experience of the given system and of related systems.*" This case study provides the data and detailed methodologies necessary to implement part of the 2004 NAP recommendations.

The purpose of this case study is to demonstrate a proof of concept. We explore both frequentist and Bayesian inference techniques to combine information through the use of formal statistical models for the Stryker FOV from two test phases: the Developmental Test (DT) phase and the Operational Test (OT) phase. The remainder of this case study is organized as follows. First, we provide an overview of the defense acquisition process, highlighting testing that occurs within each phase. Then we provide an introduction to the Stryker family of vehicles. In the data collection section, we describe the DT and OT data that was used in the analysis, along with more specific details regarding the differences between DT and OT for this specific testing of the Stryker FOV. In the Analysis and Interpretation section, we start by illustrating the current reliability analysis that is widely employed by DoD and uses only OT data in its assessment of a Stryker vehicle's reliability. This is followed by a discussion of both the frequentist and Bayesian inference techniques that were used to combine DT and OT information through the use of a parametric model. The results of these analyses are then compared and interpreted. We show that one can improve upon current DoD reliability analyses by incorporating information from both DT and OT testing.

The Defense Acquisition System

The primary objective of the DoD acquisition process is to obtain quality weapon systems that meet an operational need, in a cost-effective and timely manner. From initial concept to deployment, the procurement of a major weapon system follows a series of event-based decision points and milestone reviews. This process is formally referred to as the Defense Acquisition System and is the management process by which DoD develops and buys weapons and other systems. At each milestone, a set of specific criteria must be met for entry into a new program phase to be authorized. There are three milestones:

- **Milestone A:** entry into Technology Development
- **Milestone B:** entry into Engineering and Manufacturing Development
- **Milestone C:** entry into Production and Deployment

Testing and evaluation plays an extremely important part throughout this acquisition process. The two broad types of testing used are developmental testing and operational testing. In a developmental test (DT), the primary purpose is to verify that the system meets its specifications. Developmental testing and evaluation is done throughout the system's life cycle, from program

initiation through system sustainment and includes component testing, modeling and simulation, and engineering systems testing of a more complete system. This testing can occur as contractor testing, government testing, or a mixture of both, and it is usually carried out in a more controlled environment. This testing can last years and the design of the system itself may change multiple times during this period. Developmental testing and evaluation is used to support the low rate initial production decision at Milestone C.

Operational testing follows the low rate initial production of the system. In an operational test, production representative systems are tested by end user (DoD civilian) or real user (soldier) test teams in an operationally realistic environment. The primary purpose of the OT is to validate the system; to determine whether the system is operationally effective and suitable for its intended use before full rate production is approved and contracts are awarded. The duration of these tests is typically much shorter than in the DT phase because of the limited purpose of the testing and considerable cost of conducting operational tests. In order to be operationally suitable, a system must be available for combat when needed, reliable enough to accomplish its intended mission in its anticipated environment, operate satisfactorily with service personnel and with other systems, and not impose an undue logistics burden in peacetime or wartime. Three of the primary components used to asses a system's suitability are reliability, availability, and maintainability. The operational effectiveness of a system refers to its capability to perform its mission in the operational environment in the face of an expected threat.

The Stryker Family of Vehicles

The Stryker is a family of wheeled armored combat vehicles built for the U.S. Army. The family of vehicles includes ten separate system configurations, with two main versions of the vehicle under which these ten systems can be organized: the Infantry Carrier Vehicle (ICV) and the Mobile Gun System (MGS). This case study focuses on the ICV, which provides protected transport and supporting fire for its two-man crew and squad of nine infantry soldiers.

Figure 1. The Infantry Carrier Vehicle (ICV) serves as the base vehicle for eight additional system configurations. SOURCE: www.sbct.army.mil.

The ICV serves as the base vehicle for the eight remaining system configurations. These vehicles share a common chassis and are then outfitted with additional components that are specific to the mission and purpose of each vehicle. The additional eight configurations are the Antitank Guided Missile Vehicle (ATGMV), Commander's Vehicle (CV), Engineer Squad Vehicle (ESV), Fire Support Vehicle (FSV), Medical Evacuation Vehicle (MEV), Mortar Carrier Vehicle (MCV), Reconnaissance Vehicle (RV) and the Nuclear, Biological and Chemical Reconnaissance Vehicle (NBC RV). The NBC RV was excluded from this case study because it was on a different acquisition timeline, and therefore does not have data from the same tests as the other variants. Each of the ICV variants is outfitted with specialized equipment to accomplish the particular mission of the vehicle. For example, the Engineer Squad Vehicle (ESV) provides the ability to clear obstacles, including both surface and subsurface mines, generate smoke, and mark lanes for safe passage. To accomplish its mission, this vehicle is outfitted with a surface mine plow, a lightweight mine roller, a counter-mine magnetic signature duplicator system, and a trailer with a lane marking system and either a mine-clearing line charge or a multiple delivery mine system.

The reliability requirement for the Stryker is based on the mean number of miles between failures. In general, a failure can be defined as an event in which an item or part of an item does not perform as intended. The Army Failure Definition Scoring Criteria (FDSC) describes four essential functions that the Stryker must be capable of performing. It must be able to move, shoot, communicate, and survive. Each of these four essential functions has a specific definition of what it means to satisfy the requirement. The FDSC categorizes the severity of failures into three levels: System Aborts, Essential Function Failures, and Non-Essential Function Failures. These failure types are further defined by the FDSC as follows:

- System Abort (SA) – A failure that results in the loss or degradation of an essential function that renders the system unable to enter service or causes immediate removal from service.

- Essential Function Failure (EFF) – A failure that results in loss or degradation of an essential function. By definition all SAs are also EFFs.

- Non-Essential Function Failure (NEFF) – An event that does not result in the degradation of an essential function or can be deferred to the next maintenance period.

A system abort occurs when the system is degraded by the loss of one or more systems such that it cannot complete the assigned mission. An EFF occurs when an essential function is lost, but it is not required to complete the specific mission at hand. The Army reliability requirement for the Stryker is that the vehicle have a mean of 1,000 miles between SAs.

In this case study we focus on SA failures, as this allows for a direct comparison between current DoD analysis and an analysis that combines information across multiple phases. Future analyses might also use data on EFFs and NEFFs, which would undoubtedly provide more information about system reliability.

DATA COLLECTION

The data used in this case study come from the Stryker FOV developmental and operational testing in 2003. There are several differences between DT and OT that should result in practical differences in their reliability estimates. The road conditions, vehicle drivers, and individual mission durations varied between DT and OT. Additionally, the DT testing spanned a much longer time period than OT. Most importantly, the operators in OT were field-representative operators, and the test was limited to two weeks. Because of these differences, we expect differences in the reliability estimates from DT and OT. Robinson and Dietrich (1989) and Erkanli et al. (1998) suggest that systems reliability will increase from phase to phase under the assumption that engineering modifications being made to the system on the discovery of a failure will improve the system. This is not often the case in operational testing; in a typical vehicle program, degradations in reliability between 20 and 50 percent have been observed due to the operational nature of the test.

Table 1 provides a summary of the Stryker FOV reliability data by test phase and vehicle variant. Included in this table are the total number of miles driven, the number of SAs, and the number of right censored observations. Of the 263 observations recorded, 199 of these were SA failures (131 SAs in DT and 68 SAs in OT) and the remaining 64 observations were right censored (12 in DT and 52 in OT).

Table 1. Stryker 2003 Data Summary by Vehicle Variant and Test Phase

Vehicle Variant	Developmental Test			Operational Test		
	Total Miles	System Aborts	Right Censored	Total Miles	System Aborts	Right Censored
ATGMV	30086	17	1	10334	12	9
CV	24160	11	2	8494	1	6
ESV	25095	35	2	3771	13	3
FSV	24385	11	2	2306	1	2
ICV	61623	39	3	29982	35	23
MCV	3702	7	1	4521	4	4
MEV	--	--	--	1967	0	2
RV	23742	11	1	5374	2	3
Total	192793	131	12	66749	68	52

Type I (time-based) right censoring occurs when the testing of the vehicle was terminated before a SA was observed. As can be seen in Table 1 and Figure 2, there was a higher rate of censoring in OT because of the limited test time and larger numbers of test assets. This higher rate of censoring also affects the spread of the data; there is more variability in the DT failure distances than there is in the OT failure distances.

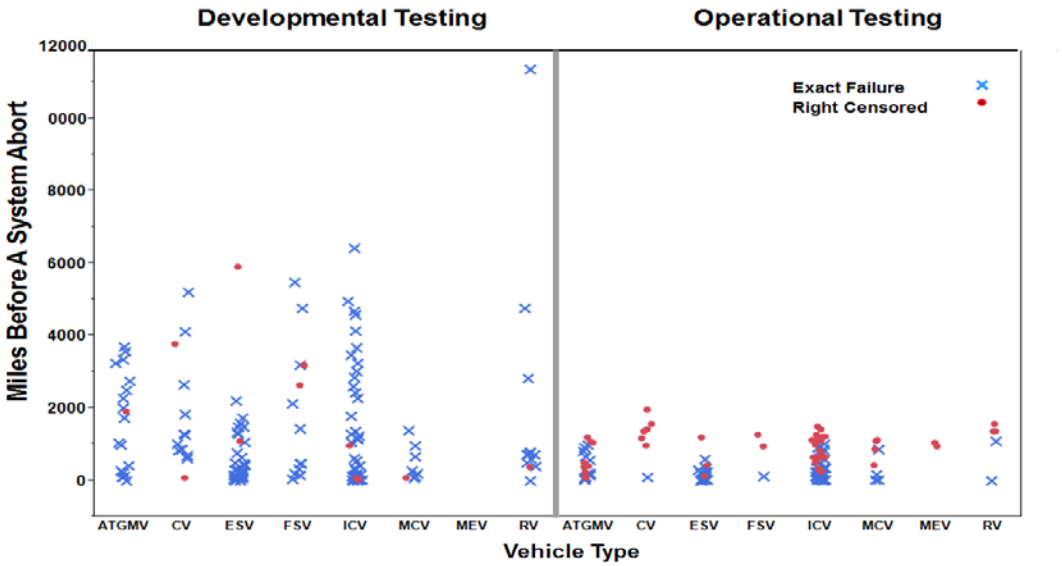

Figure 2. A scatter plot of the data grouped by test phase and vehicle type.

A limitation in that data is that there are nine instances in which the vehicle's SA failure mileage was recorded as a zero. These nine responses were spread over the different vehicle variants and the two test phases, are are the result of finding another SA failure during the repair of a primary SA failure. To account for these special cases in the data, two separate entries are recorded: the first entry records the exact mileage for the initial SA that stopped the vehicle, while the second entry records the discovered SA and uses the response value of zero in the mileage column. This

data limitation and its correction will be discussed in more detail in the Analysis and Interpretation section.

Note in Table 1 and Figure 2 there are no observations recorded for the MEV in DT and there are just two right censored observations recorded for the MEV in OT. With no SAs, we are limited in the results that we can provide for this vehicle under the frequentist paradigm. Using Bayesian inference techniques, however, we can take advantage of the available information for the other vehicles and their relationship to each other to better understand the MEV reliability.

ANALYSIS AND INTERPRETATION

The Current DoD Analysis

A standard reliability analysis employed by the DoD test community considers each test phase (and each vehicle type in this case) independently and uses the exponential distribution to model the miles (or time) between SAs (Director, Defense Test and Evaluation, 1982). In an analysis using the exponential distribution, the reliability of the Stryker is expressed in terms of the mean miles between an SA (MMBSA), and can be estimated as

$$\widehat{\mathrm{MMBSA}} = \frac{\text{Total Miles Driven}}{\#\text{ of System Aborts}}$$

An exact two-sided confidence interval for the MMBSA can be calculated directly by

$$\left[\underline{\mathrm{MMBSA}},\ \overline{\mathrm{MMBSA}}\right] = \left(\frac{2\mathrm{T}}{\chi^2_{\frac{\alpha}{2},2r+2}}, \frac{2\mathrm{T}}{\chi^2_{1-\frac{\alpha}{2},2r}}\right),$$

where T is the total number of miles driven and r is the number of SAs. If there are no SAs recorded (0 failures), then a one-sided confidence bound for MMBSA can be calculated. A conservative 100(1-α)% one-sided lower confidence bound for MMBSA is

$$\underline{\mathrm{MMBSA}} = \frac{2\mathrm{T}}{\mathrm{X}^2_{\alpha,2\mathrm{r}+2}}.$$

Table 2 illustrates the results of this analysis using OT data only. It is very similar to the analysis that DOT&E provided to Congress, except that we employ 95 percent two-sided confidence intervals, DOT&E used one-sided 80 percent lower confidence bounds. We will use these results as reference when comparing the new methods that consider incorporating information across the DT and OT test phases.

Table 2. Stryker Reliability by Vehicle Variant in Operational Testing

Vehicle	Total	System	MMBSA	MMBSA	MMBSA

Variant	Miles	Aborts		95% LCL	95% UCL
ATGMV	10334	12	861	493	1667
CV	8494	**1**	8494	1525	335495
ESV	3771	13	290	170	545
FSV	2306	**1**	2306	414	91082
ICV	29982	35	857	616	1230
MC	4521	4	1130	441	4148
MEV	1967	**0**	--	657	--
RV	5374	**2**	2687	743	22187
Total	66749	68	982	774	1264

Notice that for four of the eight vehicles under consideration (CV, FSV, MEV, and RV), there are no more than two SAs in OT. We expect the benefits of combining information will be the greatest in these cases where only limited OT data are available. The CV additionally stands out as potentially having an optimistically high MMBSA, considering that it was based on six censored observations and no individual vehicle traveled more than 2,000 miles in OT. Additionally, if we use the simple exponential estimate of the MMBSA in DT, we find that the MMBSA was less than 2,200 miles. It is highly unlikely that we would see such a large improvement in the reliability between late DT and OT because no major changes were made to the system configuration.

Statistical Models for Combining DT and OT Data

Instead of considering the DT and OT phases and the vehicle variants independently of each other, we aim to improve on this current reliability analysis by using parametric statistical models to formally combine the data and make inference. We begin with the assumption that the observations, vehicle failure miles, t, follow a known distribution, $f(t|\boldsymbol{\theta})$.

In Figure 3, we compare the distribution of the data across all vehicle variants and test phases to both the exponential distribution and the Weibull distribution - two common distributions for modeling lifetime data. The Weibull distribution appears to be a better fit to the data, as the data points are closer to a straight line in the probability plot. This comes as no surprise, because the Weibull is a two-parameter distribution and therefore offers more flexibility than the exponential distribution. In the analyses to follow, we use the Weibull distribution to model the vehicle failure miles. However, if we set the Weibull shape parameter, $\beta = 1$, all of the analyses reduce to the exponential case and are therefore comparable to the current analysis results presented in Table 2.

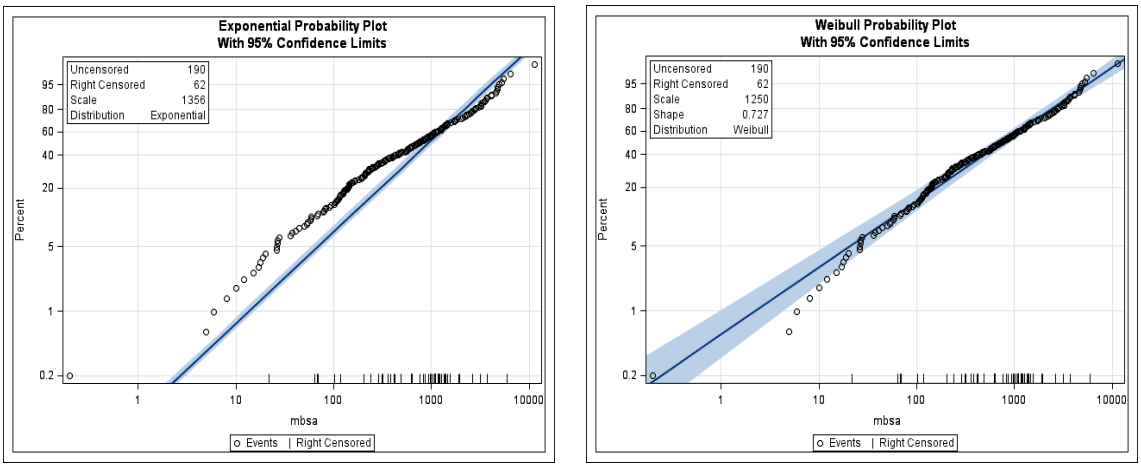

Figure 3: Exponential (left) and Weibull (right) probability plots.

A Frequentist Model for Combining DT and OT Data

A Weibull regression model was used to formally combine the information from the DT and OT phases and the information from the individual vehicle variants. To combine DT and OT data for the Stryker FOV, we treated test phase as an explanatory variable to be included in the model. All of the vehicle variants were also included as explanatory variables in the model so that individual reliability estimates for each of the vehicles within each test phase could be estimated. The MEV data were removed from this analysis because there were just two right censored observations.

Since the Weibull distribution has two parameters, it is possible for both the shape parameter β and the scale parameter η to depend on the explanatory variables (test phase and vehicle variant). It often assumed that the shape parameter β is constant and does not depend on the explanatory variables. From an engineering perspective, this assumption is valid as long as the failure mechanism is not expected to change with the different levels of the explanatory variables. For completeness, however, we considered two Weibull regression models:

Model 1: Both η and β are functions of test phase and vehicle variant.

$$\mu_i = \log(\eta_i) = \gamma_{\eta 0} + \gamma_{\eta 1}\text{Phase}_i + \gamma_{\eta 2}\,\text{ATGMV}_i + \gamma_{\eta 3}\,\text{CV}_i + \gamma_{\eta 4}\text{ESV}_\text{i} + \gamma_{\eta 5}\text{FSV}_\text{i} + \gamma_{\eta 6}\text{ICV}_\text{i} + \gamma_{\eta 7}\text{MCV}_\text{i}$$

$$\beta_i = \gamma_{\beta 0} + \gamma_{\beta 1}\text{Phase}_i + \gamma_{\beta 2}\,\text{ATGMV}_i + \gamma_{\beta 3}\,\text{CV}_i + \gamma_{\beta 4}\text{ESV}_\text{i} + \gamma_{\beta 5}\text{FSV}_\text{i} + \gamma_{\beta 6}\text{ICV}_\text{i} + \gamma_{\beta 7}\text{MCV}_\text{i}$$

Model 2: Only η is a function of test phase and vehicle variant; β remains constant.

$$\mu_i = \log(\eta_i) = \gamma_0 + \gamma_1\text{Phase}_i + \gamma_2\,\text{ATGMV}_i + \gamma_3\,\text{CV}_i + \gamma_4\text{ESV}_\text{i} + \gamma_5\text{FSV}_\text{i} + \gamma_6\text{ICV}_\text{i} + \gamma_7\text{MCV}_\text{i} \qquad (1)$$

We found neither test phase nor vehicle variant to have a significant effect on the shape parameter β. Thus a common value for β is appropriate, and the expression for the scale parameter η given in Equation 1 was used. The indicator variables in this expression were coded as

$$\text{Phase}_i = \begin{cases} 1 & if\ t_i\ is\ a\ DT\ observation \\ 0 & otherwise \end{cases}$$

$$\text{ATGMV}_i = \begin{cases} 1 & if\ t_i\ is\ an\ ATGMV\ observation \\ 0 & otherwise \end{cases}$$

The Reconnaissance Vehicle serves as the vehicle reference group. Maximum likelihood estimation is used to estimate the Equation 1 regression coefficients and the constant shape parameter β. The total likelihood to be maximized is (from Meeker and Escobar 1998)

$$\begin{aligned} L(\gamma_0, \gamma_1, \ldots \gamma_7, \beta | \boldsymbol{t}) &= C \prod_{i=1}^{n} [f(t_i)]^{\delta_i} [1 - F(t_i)]^{1-\delta_i} \\ &= C \prod_{i=1}^{n} \left[\frac{\beta}{\eta_i} \left(\frac{t_i}{\eta_i} \right)^{\beta-1} \exp\left(- \left(\frac{t_i}{\eta_i} \right)^{\beta} \right) \right]^{\delta_i} \left[\exp\left(- \left(\frac{t_i}{\eta_i} \right)^{\beta} \right) \right]^{1-\delta_i}, \end{aligned}$$

where η_i is replaced by the expression given in Equation 1, C is a constant dependent on the sampling scheme but not dependent on the unknown parameters and for simplicity can be set to $C = 1$, $f(t_i)$ is the Weibull pdf, $F(t_i)$ is Weibull cdf, and the indicator δ_i is defined by

$$\delta_i = \begin{cases} 1 & if\ t_i\ is\ an\ exact\ observation \\ 0 & if\ t_i\ is\ a\ right\ censored\ observation \end{cases}.$$

For this regression model, while the regression coefficients $\gamma_0, \gamma_1, \ldots \gamma_6$, and γ_7 are important, the quantities we are most interested in interpreting are the mean miles between a system abort (MMBSA) for each of the Stryker vehicle variants. For the Weibull distribution, the maximum likelihood estimate for MMBSA is

$$\widehat{MMBSA} = \hat{\eta} \left[\Gamma\left(1 + \frac{1}{\hat{\beta}} \right) \right].$$

The MMBSA estimates for each vehicle variant in both the DT and OT phases can be calculated by replacing $\hat{\eta}$ with its estimated expression given in Equation 1. Because these MMBSA estimates are functions of both β and the regression coefficients $\gamma_0, \gamma_1, \ldots \gamma_6, \gamma_7$, the multivariate delta method can be used to calculate their standard errors, making inference on the MMBSA for each of the vehicle variants possible. Using these standard errors, a 100(1- α)% Wald confidence interval for each vehicle's MMBSA reliability estimates is

$$\left[\underline{g(\widehat{\boldsymbol{\theta}})}, \overline{g(\widehat{\boldsymbol{\theta}})}\right] = g(\widehat{\boldsymbol{\theta}}) \pm z_{1-\frac{\alpha}{2}}\sqrt{var\left(g(\widehat{\boldsymbol{\theta}})\right)}$$

Where, $\widehat{\boldsymbol{\theta}}$ is the vector of the parameter estimates, $g(\widehat{\boldsymbol{\theta}})$ represents a scalar function of the parameters, and $var\left(g(\widehat{\boldsymbol{\theta}})\right)$ is the m^{th} diagonal element of the variance-covariance matrix of $g(\boldsymbol{\theta})$. For example, to estimate the OT MMBSA for the Stryker ICV, the scalar function $g(\widehat{\boldsymbol{\theta}})$ would be

$$g(\widehat{\boldsymbol{\theta}}) = g(\hat{\gamma}_0, \hat{\gamma}_6, \ldots, \hat{\gamma}_7, \hat{\beta}) = \exp(\hat{\gamma}_0 + \hat{\gamma}_6)\,\Gamma(1 + \tfrac{1}{\hat{\beta}}).$$

A Bayesian Model for Combining DT and OT Data

One of the main attractions of using a Bayesian approach is that it provides a formal procedure for combining multiple sources and types of information. Even when multiple types of information are being used (e.g. component-level data, system-level data, and subject-matter expertise), system reliability estimates can still be obtained through a single analysis. Much of the Bayesian reliability literature focuses on combining information for complex systems where it is not always possible to conduct many full system tests due to cost, practicality, and permissibility. Anderson-Cook et al. (2007) show this to be especially true in the assessment of the reliability of the stockpile of nuclear weapons. Hamada et al. (2004), Graves et al. (2010), and Wilson et al. (2011) all consider methods that allow for the combination of component-level data or data from other variants of the system that can be incorporated into the system-level analysis. Bayesian hierarchical modeling provides a rigorous way to combine information from multiple sources and different types of information. Johnson et al. (2003) and Reese et al. (2011) show how one can use Bayesian hierarchical models to integrate component, subsystem and system data, along with prior expert opinion, to assess the reliability of a complex system. Anderson-Cook (2009) looks at both the opportunities and issues in data combination while Wilson et al. (2006) addresses some of the advances in data combination; both provide helpful examples.

We employ a Bayesian hierarchical model framework to formally combine DT and OT data for the Stryker FOV. The following model specification was used to model the failure miles, t

$$t_{DT} \sim Weibull(n_j, \beta) \qquad t_{OT} \sim Weibull(\delta n_j, \beta) \qquad j = 1,2,\ldots,8 \qquad (2)$$

A multiplicative model structure was chosen specifically because it is analogous to the Weibull regression model defined in Equation 1. It is comparable because of the parameterization that is used in the Weibull regression model expression for the scale parameter, $\mu = \log(\eta)$, where the difference between the OT and DT phases for a vehicle variant is

$$\mu_{OT,variant} - \mu_{DT,variant} = \log\left(\frac{\eta_{OT,variant}}{\eta_{DT,variant}}\right).$$

Expressing this difference then, in terms of the scale parameter η, gives

$$\frac{\eta_{OT,variant}}{\eta_{DT,variant}} = \exp(\mu_{OT,variant} - \mu_{DT,variant}) = \exp(-\gamma_1).$$

In the Bayesian multiplicative model, the shift in the scale parameter η from the DT to OT phase is represented by δ. Like the Weibull regression model, the model given in Equation 2 assumes a common shape parameter β. By indexing, η, we are also allowing for the reliability estimates to be different for the eight vehicles but still related since we are assuming the $\eta_j's$ come from a common distribution. An immediate advantage to using this type of model is that a reliability estimate for the MEV can now be obtained. This estimate is driven by the information that we have for the seven other vehicles.

We used a hierarchical prior for the parameter η_j specified by the gamma distribution and expressed as

$$\pi(\eta_i|\alpha_\eta, b_\eta) = \frac{b_\eta^{a_\eta}}{\Gamma(a_\eta)}\eta_j^{a_\eta - 1}e^{-(b_\eta \eta_j)}.$$

Completing this hierarchical specification for η_j, we assumed that the hyperparameters α_η, b_η have independent prior gamma distributions. A diffuse prior distribution reflects little prior knowledge and allows the results to be driven by the data. Specifically, we used the priors:

$$\pi(\alpha_\eta) \sim gamma(.001, .001)$$

$$\pi(b_\eta) \sim gamma(.001, .001)$$

The prior variance for these parameters is 1,000. There is little additional information to inform this choice. Assuming that η_j and β are independent, we used a diffuse gamma prior for β, expressed as

$$\pi(\beta|\alpha_\beta = .001\,, b_\beta = .001) = \frac{b_\beta^{a_\beta}}{\Gamma(a_\beta)}\beta^{a_\beta - 1}e^{-(b_\beta \beta)}$$

A diffuse gamma prior was also used for the multiplicative shift parameter δ. This choice in prior allows for the possibility that the vehicle MMBSA estimates could improve when testing moves from the DT phase to OT phase,

$$\pi(\delta) = gamma(.001, .001).$$

O'Hagan (2004), Bedford et al. (2006) and Wilson et al. (2007) discuss how to elicit information from subject-matter experts.

Inferences for this multiplicative model are made by using the joint posterior distribution, which is proportional to the product of likelihood and priors, and is

$$f(\eta_1, \ldots, \eta_8, \beta, a_\eta, b_\eta, \delta | \boldsymbol{t})$$

$$\propto \left[\prod_{i=1}^{n} [f(t_i)]^{\delta_i} [1 - F(t_i)]^{1-\delta_i}\right]\left[\prod_{j=1}^{8} \pi(\eta_j) * \pi(\beta) * \pi(a_\eta) * \pi(b_\eta) * \pi(\delta)\right],$$

where f(t_i) is the Weibull pdf, $F(t_i)$ is the Weibull cdf, and the indicator δ_i is defined by

$$\delta_i = \begin{cases} 1 & if\ t_i\ is\ an\ exact\ observation \\ 0 & if\ t_i\ is\ a\ right\ censored\ observation \end{cases}.$$

To obtain draws from this joint posterior distribution of the model parameters, we implemented a Metropolis-in-Gibbs algorithm using the programming language R. Gaussian proposal densities were used for the parameters β, η_i, δ and for the hyper-parameters in the hierarchical specification of η_j. Initial runs were used to determine appropriate standard deviations for the Gaussian proposal densities. The Raftery-Lewis diagnostic (Raftery and Lewis 1996) was used to help in determining the number of iterations needed to adequately estimate the 2.5% and 97.5% quantiles. To determine adequate mixing and a sufficient burn-in time, we considered trace plots and auto-correlation. In the end, we based our results on 1,000,000 draws from the posterior distribution. This was done to ensure that we were properly exploring the space because there was significant auto-correlation between the hyper-parameters a_η and b_η.

Imputing Missing Data

As was mentioned earlier, nine of the SA failure miles were recorded as a zero. These responses become an issue in the analysis of the reliability data when using parametric models. Many of the commonly used parametric distributions in reliability analysis, such as the Weibull distribution, require that the random variable be positive ($t > 0$). Additionally, we know that the failure occurred in the range between the last failure and the current failure under investigation. Therefore, we treated these responses as missing values and imputed plausible values for $t_{missing}$ to complete the dataset. In this case study, when imputing new data points to replace the zeros, we must also correct the initial SA stopping mileage so as not to incorrectly duplicate these miles in the total likelihood. Figure 4 illustrates this concept and the appropriate mileage correction for the response value to be used in the total likelihood.

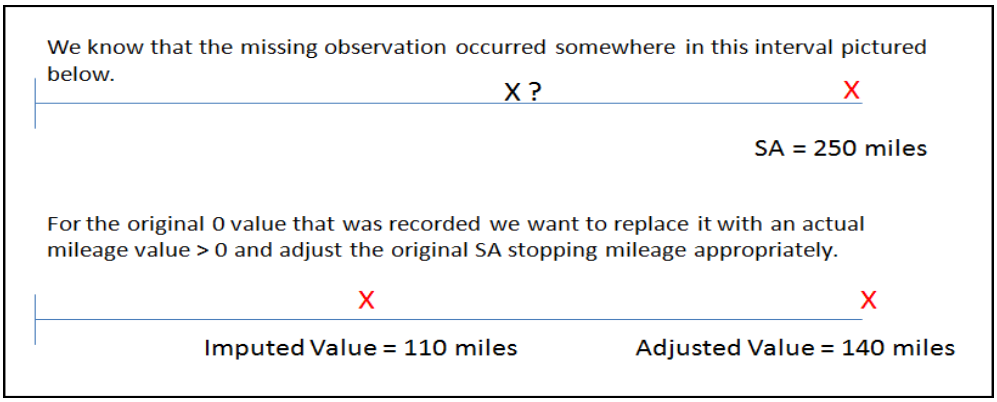

Figure 4. Correcting for missing data

A Frequentist Solution for Missing Data

Before fitting the Weibull regression model in Equation 1, we used the method of multiple imputation to generate three complete datasets. Multiple imputation is a three-step process:

1. **Impute:** propose plausible values (m sets) for missing observations
2. **Analyze:** analyze each of the m complete datasets
3. **Pool:** integrate the m analysis results into a final result

Little and Rubin (1987) suggest that 3-5 imputed datasets will usually be sufficient. The three datasets that we used were completed by replacing the $t_{missing}$ response values with the 20^{th}, 40^{th}, and 50^{th} percentile values of the observed SA stopping mileage respectively. For example, if the observed SA stopping mileage was 1,028 miles, the 20^{th}, 40^{th}, and 50^{th} percentile values used to replace the recorded zero observations in the three datasets are 205.6, 411.2, and 514 miles; the adjusted values are 822.4, 616.8, and 514 miles. Once the m complete datasets are created and analyzed individually, we can calculate the combined coefficient estimates and a covariance matrix for these estimates by averaging the values across the imputations. Details can be found in the Appendix.

It is often recommended that a more rigorous approach be used to impute the data, such as simulating data from an appropriate probability distribution. However, since we know the range of values that were possible (i.e. the value must be less than or equal to the next miles between system abort) a simpler methodology suffices. This methodology eliminates challenges with generating values outside of the range of appropriate values. Additionally, we should note that for this data, the parameter estimates were robust to the imputed values.

A Bayesian Solution for Missing Data

These missing response values can also be accounted for in the Bayesian analysis. This was done by making use of the following relationship,

$$t_{missing}|\beta, \eta_{\,phase,variant} \sim Weibull(\beta, \eta_{\,phase,variant})$$

where the Weibull distribution is truncated at the original SA mileage. β and $\eta_{phase,variant}$ are the current draws for these parameters in kth iteration of the Metropolis algorithm. More specifically, the value of $\eta_{phase,variant}$ is the current value of η for the missing observation's associated test phase and vehicle variant. New values are sampled from the updated distribution in each iteration of the Metropolis algorithm and, as illustrated in Figure 4 its counterpart, the original SA mileage, is adjusted accordingly.

Because it is possible to sample a new value from this $Weibull(\beta, \eta_{\,phase,variant})$ distribution that is larger than the original SA mileage, an adjustment was made so that only appropriate values would be sampled in each iteration of the Metropolis algorithm. For this adjustment, a grid of the possible values (i.e., $\leq$ original SA mileage) was simulated from the $Weibull(\beta, \eta_{\,phase,variant})$ distribution; see Figure 5. From this truncated distribution, a value for $t_{missing}$ was sampled using the $Weibull(\beta, \eta_{\,phase,variant})$ distribution probabilities as weights.

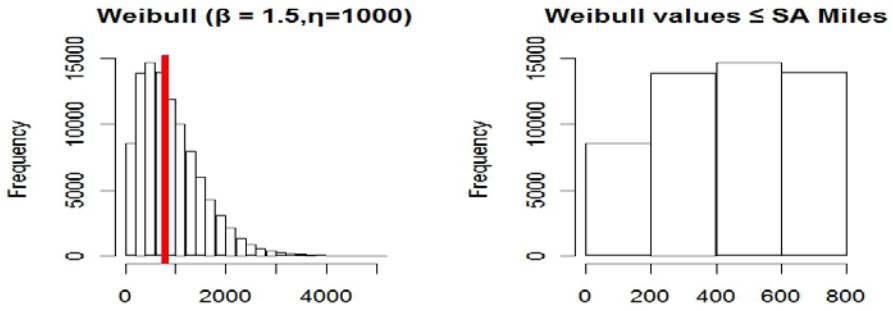

Figure 5. Left: A histogram of simulated data following a Weibull($\boldsymbol{\beta = 1.5, \eta = 1000}$) distribution. Right: A histogram of all possible values ≤ SA 800 miles. From this truncated distribution, a value is selected using the Weibull($\boldsymbol{\beta = 1.5, \eta = 1000}$) probabilities as weights.

Results and Comparison of Methods

In this section we discuss and compare the results of both the frequentist Weibull regression analysis and Bayesian analysis. The parameter estimates for the Weibull regression model described in Equation 1 are shown in Table 3. These estimates were obtained using the three

imputed data sets. The reliability estimates for both DT and OT MMBSA are included in Table 4, along with Wald 95% confidence intervals.

Table 3. Parameter estimates for the Weibull regression analysis using the three imputed data sets.

Model Term	Estimate	Standard Error
Intercept: γ_0	7.479	0.391
DT Phase: γ_1	0.267	0.207
OT Phase	--	--
ATGMV: γ_2	-0.452	0.434
CV: γ_3	0.327	0.520
ESV: γ_4	-1.460	0.407
FSV: γ_5	-0.047	0.520
ICV: γ_6	-0.585	0.393
MCV: γ_7	-0.976	0.536
RV	--	--
SEV Scale σ	1.298	0.076
Weibull Shape β	0.771	0.045

Table 4. MMBSA estimates and Wald 95% confidence intervals based on Weibull regression analysis.

Vehicle Variant	Developmental Test		Operational Test	
	MMBSA Estimate	95% Confidence Interval	MMBSA Estimate	95% Confidence Interval
ATGMV	1714.69	(863.83, 2565.56)	1312.90	(584.01, 2041.79)
CV	3736.82	(920.35, 6553.29)	2861.18	(556.38, 5165.98)
ESV	625.77	(391.01, 860.54)	479.14	(240.93, 717.35)
FSV	2570.85	(669.98, 4471.71)	1968.43	(342.04, 3594.81)
ICV	1501.16	(982.08, 2020.23)	1149.39	(703.73, 1595.06)
MCV	1015.35	(200.18, 1830.53)	777.43	(159.43, 1395.43)
RV	2694.56	(765.65, 4623.47)	2063.15	(473.47, 3652.83)

The parameter estimates for the Bayesian model given in Equation 2 are shown in Table 5. In this table, we see that the estimate for the shift parameter $\delta = 0.83$, but that the 95% credible interval covers the range (0.53, 1.26), indicating that it is possible for OT reliability estimates to be better than DT reliability estimates. For this reason, the histogram seen in Figure 6 is included to show that the probability of $\delta \leq 1$ (i.e. DT estimates are higher than OT estimates) is 83%.

The results of this Bayesian analysis can be easily expressed in terms of MMBSA and are included in Table 6. Notice that this table now includes an estimate for the MEV.

Table 5. Parameter estimates from Bayesian analysis.

Model Term	Estimate	95% Credible Interval
η_{ATGMV}	1542	(933.5, 2456.4)
η_{CV}	2662	(1374.3, 5013)
η_{ESV}	604	(388.3, 919.5)
η_{FSV}	2103	(1091.1, 3904.5)
η_{ICV}	1309	(904.6, 1844.4)
η_{MCV}	1181	(511, 2330.6)
η_{RV}	2127	(1113.5, 3925.7)
η_{MEV}	2543	(763.7, 7188.6)
β	0.74	(0.65, 0.83)
δ	0.83	(0.53, 1.26)

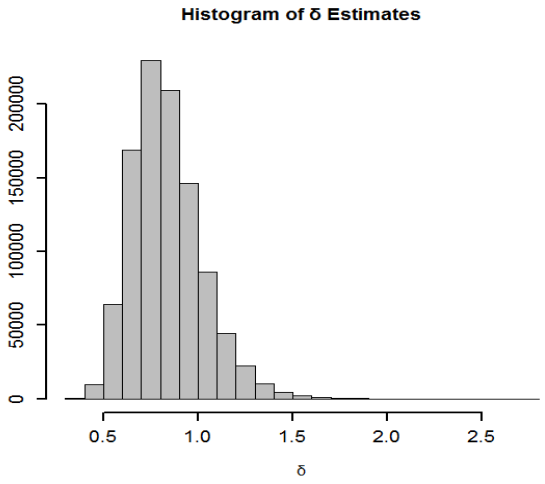

Figure 6. A histogram of the estimates for δ, the multiplicative shift parameter. Note that the $\Pr(\delta \leq 1) = .828$.

Table 6 MMBSA estimates and 95% credible intervals from Bayesian analysis.

Vehicle Variant	Developmental Test		Operational Test	
	MMBSA Estimate	95% Credible Interval	MMBSA Estimate	95% Credible Interval
ATGMV	1872.00	(1136.91, 3008.76)	1541.45	(870.14, 2649.30)
CV	3230.95	(1677.35, 6135.88)	2663.58	(1290.69, 5345.95)
ESV	732.82	(475.45, 1126.41)	607.50	(343.96, 1043.18)
FSV	2551.70	(1330.47, 4769.15)	2114.31	(993.59, 4247.07)
ICV	1587.88	(1107.22, 2254.91)	1302.07	(863.50, 1959.57)
MCV	1434.13	(620.96, 2856.35)	1175.10	(503.99, 2412.21)
RV	2580.49	(1360.34, 4792.85)	2130.08	(1037.61, 4184.73)
MEV	3084.37	(929.24, 8735.19)	2529.22	(751.14, 7245.12)

Figure 7 compares the frequentist Weibull regression analysis to the Bayesian analysis. Since we specified the model form of the Weibull scale parameter in the same manner, we see comparable results between the two analyses. One benefit of the Bayesian analysis is that we can obtain a point estimate for the reliability of the MEV, which is not possible in a frequentist analysis.

Additionally, the Bayesian credible intervals are easier to calculate then the frequentist confidence intervals on the MMBSA. However, the frequentist analysis is available in many standard statistical packages.

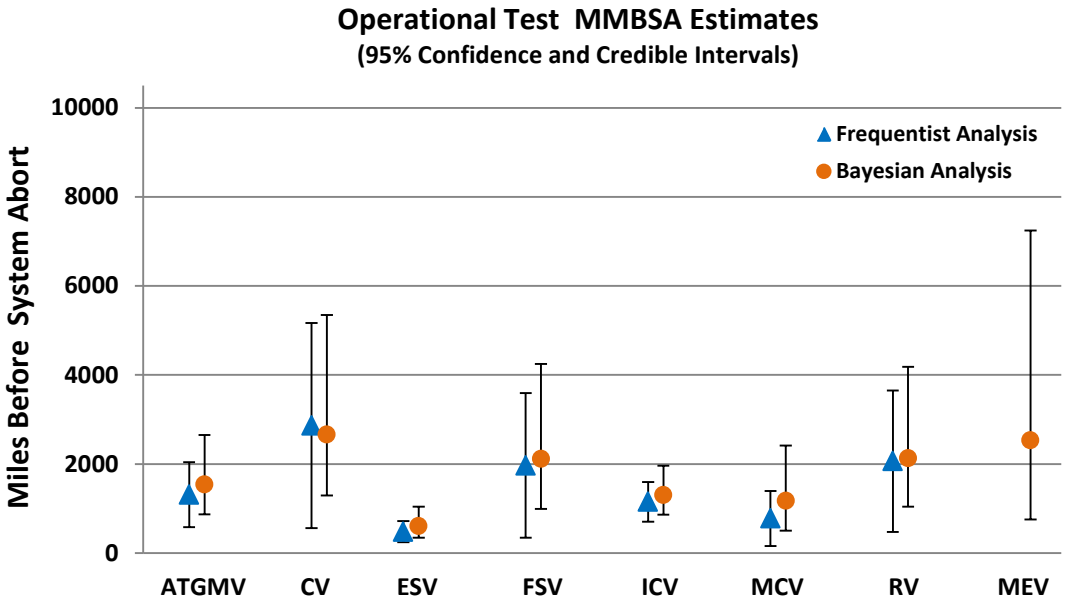

Figure 7 A comparison of the OT MMBSA vehicle variant estimates for the frequentist analysis and Bayesian analysis using the Weibull distribution.

Although the Weibull distribution provides a better fit to the data than the exponential distribution, it may still be reasonable to use the exponential distribution, especially for inference on the means. Figure 8 compares the results of the current DoD analysis, which includes only data from the OT, to the frequentist and Bayesian analysis methods described in this paper using the exponential distribution ($\beta = 1$). Notice that using a statistical model to account for changes in test phase and vehicle variant has a large practical impact on the reliability results. Consider, for example, the CV variant, in operational testing these vehicles traveled a total of 8,494 miles with one failure and six censored observations. Recall that no individual CV traveled more than approximately 2,000 miles during the operational test. Furthermore, the reliability estimate of the CV in developmental testing was 2,197 mean miles between failures, under the current exponential approach. Clearly, the estimate of the CV reliability using a statistical model for the exponential mean provides a more realistic estimate of the reliability.

The model-based analyses also improve the precision of the interval estimates (confidence intervals, credible intervals) of system reliability for the vehicles with a small number of failures (CV, FSV, MEV and RV) by leveraging the failure information from the other variants and developmental testing. Again, notice that the Bayesian and frequentist model-based approaches give similar results.

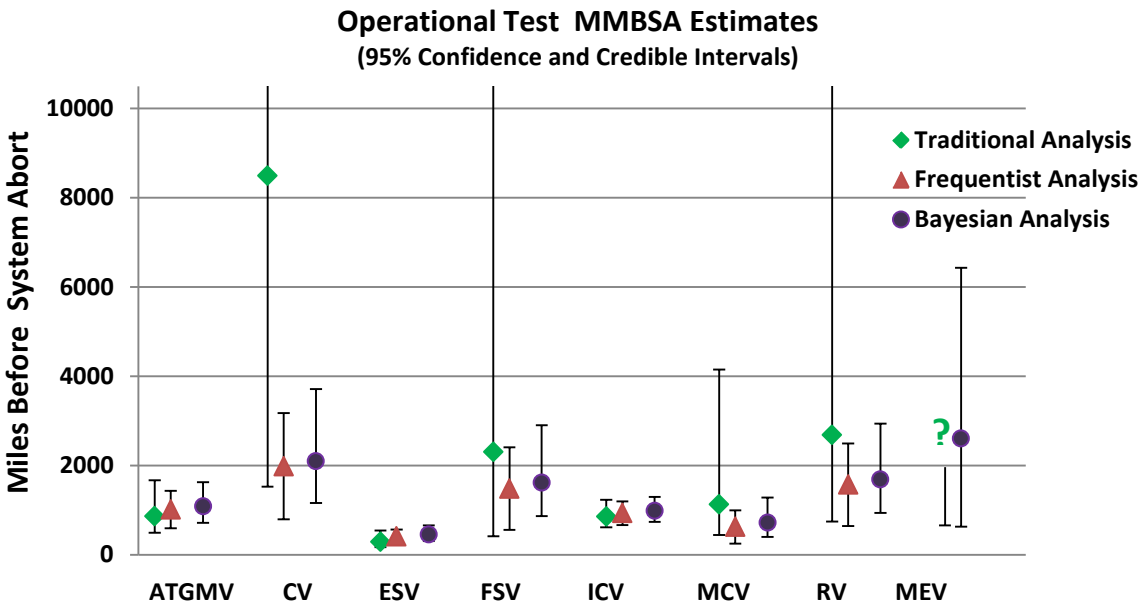

Figure 8. A comparison of the Operational Test MMBSA Vehicle Variant Estimates for the Current Analysis, Frequentist Analysis and Bayesian Analysis using the exponential distribution.

SOFTWARE RECOMMENDATIONS

The frequentist failure time regression analysis when using a complete dataset can be implemented in the current version of the JMP and Minitab. Additionally, one might consider using include the SAS software and the programming language R, both of which have built-in functions that can be used to fit failure time regression models under various distributional assumptions and censoring schemes. In the SAS software this can be done by using the PROC LIFEREG statement; in the programming language R it can be done through the command survreg() found in the library survival. Regardless of the software decision, one will still need to manually calculate the point and standard error estimates for MMBSA.

The Metropolis-in-Gibbs algorithm used in the Bayesian analysis was implemented using the programming language R. Other software programs that one might consider using to carry out the Bayesian analysis include the popular OpenBUGS and SAS (PROC MCMC) software. Less coding is required when using OpenBUGS and the PROC MCMC command in the SAS software, but the user must still be able to properly specify the correct likelihood function for the model with the appropriately coded censoring scheme (if applicable) and choose priors for the model parameters. Ultimately, we decided to use R for this case study. This choice was due both to the control one has when writing their own code and the need to impute values for the missing data and adjust these sampled values and their counterparts accordingly within the algorithm. The details of the algorithm we used can be found in the appendix and code can be made available upon request.

CONCLUSIONS AND FUTURE WORK

This case study on the Stryker FOV presents a paradigm shift in how the DoD test and evaluation community could analyzes reliability data. The case study illustrates the advantages of using data from multiple phases of testing and leveraging data from systems with common infrastructure. The results are (1) better estimates of system reliability (specifically for the CV variant) and (2) more precise inferences (especially in the cases where only a small number of failures occurred in OT). However, there are caveats to these benefits. First, the analyses presented in this paper require a strong statistical understanding of many statistical techniques, including reliability analysis, likelihoods, and missing data imputation. While some of the complications (i.e. imputation) could be removed by better data collection methods, the remainder of the analysis is considerably more complex than the currently employed methods. Furthermore, when combining information, there is no omnibus solution, as is currently employed in the exponential analysis for each test phase and variant. Rather, models need to be carefully considered and evaluated to ensure that they accurately reflect the data and the underlying physical processes.

This case study provides a proof of concept for using both Weibull regression and Bayesian analysis methods in reliability analysis when data come from different testing phases. We elected to use diffuse priors in this analysis to illustrate the comparability of frequentist and Bayesian approaches when similar model choices are made. Further improvements in reliability estimates might be achieved by leveraging information from essential function failures (EFFs) and non-essential function failures (NEFFs), and using appropriately selected informative priors. Finally, the Stryker FOV represents only one type of system that the DoD must test and evaluate. Case studies showing the value of combining information on other types of systems are needed to advance the effort across the broader DoD test and evaluation community.

REFERENCES

Anderson-Cook, C. M. (2009). Opportunities and issues in multiple data type meta-analyses, *Quality Engineering*, 21:3, 241-253.

Anderson-Cook, C. M., Graves, T., Hamada, M., Hengartner, N., Johnson, V., Reese, C.S., Wilson, A.G. (2007). Bayesian stockpile reliability methodology for complex systems" *Journal of the Military Operations Research Society* 12 25-37.

Bedford, T., Quigley, J., Walls, L. (2006). Expert elicitation for reliable system design. *Statistical Science* 21(4): 428-450.

Colosimo, E. A., Ho, L. L. (1999). Practical approach to interval estimation for the Weibull mean lifetime, *Quality Engineering*, 12:2, 161-167

Director Defense Test and Evaluation (1982). Test and Evaluation of System Reliability Availability Maintainability – A Primer; Third Edition.

Erkanli, A., Mazzuchi, T. A., Soyer, R. (1998). Bayesian computations for a class of reliability growth models. *Technometrics*, 40(1): 14-23.

Graves, T., Anderson-Cook, C. M., and Hamada, M. (2010). Reliability models for almost-series and almost-parallel systems. *Technometrics.* 52(2): 160-171.

Hamada, M., Martz, H. F., Reese C. S., Graves, T., Johnson, V., Wilson, A. G. (2004). A fully Bayesian approach for combining multilevel failure information in fault tree quantification and optimal follow-on resource allocation. *Reliability Engineering and System Safety*, 86(3): 297-305.

Johnson, V. E., Graves, T. L., Hamada, M. S., Reese, C. S. (2003). A hierarchical model for estimating the reliability of complex systems. *Bayesian Statistics*, 7, Oxford University Press.

Little, R. J. A., Rubin D. B. (1987). Statistical Analysis with Missing Data. New York: John Wiley & Sons.

Meeker, W. Q., Escobar, L. (1998). *Statistical Methods for Reliability Data*. New York: John Wiley & Sons.

National Research Council (1998). *Statistics, Testing, and Defense Acquisition: New Approaches and Methodological Improvements*, Washington, DC: The National Academies Press.

National Research Council (2004). *Improved Operational Testing and Evaluation Methods of Combining Test Information for the Stryker Family of Vehicles and Related Army Systems*, Washington, DC: The National Academies Press.

National Research Council (2012). *Industrial Methods for the Effective Development and Testing of Defense Systems*, Washington, DC: The National Academies Press.

O'Hagan, A. and Forster, J. (2004). *The Advanced Theory of Statistics, Vol. 2b: Bayesian Statistics*. New York: Oxford University Press.

Raftery, A. E. and Lewis, S.M. (1996). Implementing MCMC in Markov Chain Monte Carlo in Practice, W.R. Gilks, D.J. Spiefelhalter, and S. Richardson, eds., pp. 115-130. London: Chapman and Hall.

Reese, C. S., Wilson, A.G., Guo, J., Hamada, M.S., Johnson V.E. (2011). A Bayesian Model for Integrating Multiple Sources of Lifetime Information in System-Reliability Assessments. *Journal of Quality Technology*, 43(2): 127-141.

Robinson, D., Dietrich, D. (1989). A nonparametric-Bayes reliability-growth model. *IEEE Transactions on Reliability*, 38(5): 591-598.

Schwartz, M. (2013). Defense Acquisitions: How DoD Acquires Weapon Systems and Recent Efforts to Reform the Process. Library of Congress, Washington DC, Congressional Research Service.

SAS Institute Inc. 2011. SAS/STAT® 9.3 User's Guide. Cary, NC: SAS Institute Inc.

Wilson, A. G., Graves, T. L., Hamada, M. S., Reese, S. C. (2006). Advances in Data Combination, Analysis and Collection for System Reliability Assessment, *Institute of Mathematical Statistics*, 21:4, 514-531.

Wilson, A., Anderson-Cook, C. M., Huzurbazar (2011). A case study for quantifying system reliability and uncertainty. *Reliability Engineering and System Safety*, 96(9): 1076-1084.

Wilson, A.G., McNamara, L., Wilson, G. (2007). Information integration for complex systems. *Reliability Engineering and Systems Safety*. 91(1):121-130.

APPENDIX

More on Multiple Imputation

After imputing *m* complete datasets and analyzing each dataset individually, we can calculate combined coefficient estimates and a covariance matrix for these coefficient estimates. Suppose that $\hat{Q}_i$ and $\hat{W}_i$ are the point and covariance matrix estimates for a p-dimensional parameter Q for the i^{th} imputed dataset, $i = 1,2,..m$. Then the combined coefficient estimates for Q from the multiple imputation is (SAS software documentation for PROC MIANALYZE):

$$\bar{Q} = \frac{1}{m}\sum_{i=1}^{m} \hat{Q}_i$$

The covariance matrix that is associated with $\bar{Q}$ is calculated as follows:

$$T_0 = \overline{\boldsymbol{W}} + \left(1 + \frac{1}{m}\right)\boldsymbol{B}$$

Where $\overline{\boldsymbol{W}}$ is the within-imputation covariance matrix, the average of the m complete-data estimates:

$$\overline{\boldsymbol{W}} = \frac{1}{m}\sum_{i=1}^{m} \widehat{\boldsymbol{W}}_{\boldsymbol{i}}$$

And $\boldsymbol{B}$ is the between-imputation covariance matrix:

$$\boldsymbol{B} = \frac{1}{m-1}\sum_{i=1}^{m} (\hat{Q}_i - \bar{Q})(\hat{Q}_i - \bar{Q})'$$

The MCMC Algorithm

Posterior **

$$f(\eta_1, \dots, \eta_8, \beta, a_\eta, b_\eta, \delta | \boldsymbol{t})$$

$$\propto \prod [f(t_i)]^{\delta_i} [1 - F(t_i)]^{1-\delta_i} * \prod_{j=1}^{8} \pi(\eta_j) * \pi(\beta) * \pi(a_\eta) * \pi(b_\eta) * \pi(\delta)$$

Algorithm

Step 0: Initialize starting values: $\eta_1^{(0)}, \eta_2^{(0)}, .., \eta_8^{(0)}, \beta^{(0)}, a_\eta^{(0)}, b_\eta^{(0)}, \delta^{(0)}$
Initialize starting values for $t_{missing}$ and adjust the related failure times:

$$t_{missing} | \beta^{(0)}, \eta^{(0)}_{phase,variant} \sim Weibull(\beta^{(0)}, \eta^{(0)}_{phase,variant})$$

Step 1: Propose a new value η_1^* from a symmetric proposal distribution.
Calculate the Acceptance Probability:

$$R = \frac{f\left(\eta_1^*, \dots, \eta_8^{(k)}, \beta^{(k)}, a_\eta^{(k)}, b_\eta^{(k)}, \delta^{(k)} \middle| \boldsymbol{t}\right)}{f\left(\eta_1^{(k)}, \dots, \eta_8^{(k)}, \beta^{(k)}, a_\eta^{(k)}, b_\eta^{(k)}, \delta^{(k)} \middle| \boldsymbol{t}\right)}$$

Accept or Reject η_1^*:

- If $R \geq 1$, accept the draw η_1^*. Set $\eta_1^{(k+1)} = \eta_1^*$
- If R < 1, accept the draw η_1^*. Set $\eta_1^{(k+1)} = \eta_1^*$ with probability R.
- We do not accept the draw with probability 1-r. Then $\eta_1^{(k+1)} = \eta_1^{(k)}$

Now using the value of $\eta_1^{(k+1)}$ in R, repeat step 1 for each of the remaining parameters $\eta_2, \dots, \eta_8, \beta, a_\eta, b_\eta, \delta$.

Step 2: Impute new values for $t_{missing}$ and adjust the related failure times accordingly:

$$t_{missing} | \beta^{(k)}, \eta^{(k)}_{phase,variant} \sim Weibull(\beta^{(k)}, \eta^{(k)}_{phase,variant})$$

Step 3: Repeat Steps 1 and 2 a total of N times

**It is often more convenient to use the Log-Posterior. The ratio, R, becomes the difference between two Log-Posteriors using the proposed value and the current parameter value respectively.