ABSTRACT

Title of Document: LATENT FAILURES AND MIXED DISTRIBUTIONS: USING MIXED DISTRIBUTIONS AND COST MODELING TO OPTIMIZE THE MANAGEMENT OF SYSTEMS WITH WEAK LATENT DEFECT SUBPOPULATIONS

Anduin E. Touw, PhD, 2008

Directed By: Dr. Peter Sandborn, Mechanical Engineering Department

Abstract

Under most reliability model assumptions, all failures in a population are considered to come from the same distribution. Each individual failure time is assumed to provide information about the likely failure times of all other devices in the population. However, from time to time, process variation or an unexpected event will lead to the development of a weak subpopulation while other devices remained durable. In this paper, estimation techniques for this situation are explored.

Ideally, when such situations arise, the weak subpopulation could be identified through determination of root cause and sequestering of impacted devices. But many

times, for practical reasons, the overall population is a mixture of the weak and strong subpopulations; there may be no non-destructive way to identify the weak devices.

If the defect is not inspectable, statistical estimation methods must be used, either with or without root cause information, to quantify the reliability risk to the population and develop appropriate screening. The accuracy of these estimates may be critical to the management of the product, but estimation in these circumstances is difficult. The mixed Weibull distribution is a common form for modeling latent failures. However, estimation of the mixed Weibull parameters is not straightforward. The number of parameters involved, and frequently the sparseness of the data, can lead to estimation biases and instabilities that produce misleading results. Bayesian techniques can stabilize these estimates through the priors, but there is no closed-form conjugate family for the Weibull distribution. This dissertation, using Monte Carlo simulation, examines bias and random error for three estimation techniques: standard maximum likelihood estimation, the Trunsored method, and Bayesian estimation.

To determine how errors in the estimation methods impacts decisions about screening, a cost model was developed by generalizing existing screening cost models through the addition of the impact of schedule slippage cost and capacity. The cost model was used in determining the optimal screen length based on total life-cycle cost. The estimated optimal screen length for each method was compared to the true optimal screen length. Recommendations about when each estimation method is appropriate were developed.

LATENT FAILURES AND MIXED DISTRIBUTIONS: USING MIXED DISTRIBUTIONS AND COST MODELING TO OPTIMIZE THE MANAGEMENT OF SYSTEMS WITH WEAK LATENT DEFECT SUBPOPULATIONS

By

Anduin E. Touw

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:
Associate Professor Peter Sandborn, Chair
Associate Professor Joseph Bernstein
Professor Abhijit Dasgupta
Associate Professor F. Patrick McCluskey
Associate Professor Gilvan Souza

Dedication

This work is dedicated to my family, who has encouraged me to pursue higher learning in science and math. It is also especially dedicated to Brian, Lorien, and Mendel. Thanks for keeping me company on this journey.

Acknowledgements

I cannot express my appreciation for the help of Dr. Tony Lin and Dr. Leonard Rosenheck of Boeing. The inspiration for this topic came from discussions with them and they were excellent sounding boards throughout the project. Probing questions from Walt Thomas and Henning Leidecker of NASA also helped inspire some of the work. Peter Sandborn provided excellent direction, particularly on the cost work, and important motivation for finishing, even in the face of many distractions.

Table of Contents

Chapter 1: Introduction 1
Motivation 1
Background 7
Research Goals 8
Approach Summary and Analysis Strategy 9
Chapter 2: Previous Work 11
Methods of Estimation for Mixed Models 11
Optimizing Test and Screen Lengths 17
Summary of Weaknesses of Current Research 18
Chapter 3: Monte Carlo Methodology 20
Chapter 4: Maximum Likelihood Estimation 22
Methodology 22
Results 22
Chapter Conclusions 32
Chapter 5: Trunsored Estimation 33
Methodology 33
Results 34
Chapter Conclusions 46
Chapter 6: Bayesian Estimation 47
Methodology 47
Results 51
Chapter Conclusions 62
Chapter 7: Cost Model 64
Previous Cost Models 64
Case Study Parameter Values 68
Discussion 71
Chapter Summary 79
Chapter 8: Impact of Model Error on Burn-in Decision Rules 80
Chapter Summary 98
Chapter 9: Conclusions, Contributions, and Future Work 100
Review of Findings and Contributions 100
Future Work 103
Appendix 1: Median Rank Regression 109
Appendix 2: Estimation Errors 117

Appendix 2: Estimation Errors 117

Appendix 3: Detailed Calculations for Bayesian Posterior 120
Appendix 4: Boeing Copyright Assignment 125
References 126

Nomenclature

p	Mixture ratio (usually estimated)
p_0	True mixture ratio
θ	Characteristic life for Weibull distribution
θ_w, θ_s	Characteristic life for weak & strong subpopulations if strong modeled
β	Shape parameter for Weibull distribution
β_w, β_s	Shape parameter for weak & strong subpopulations if strong modeled
λ	Failure rate for Exponential distribution
R(t)	Reliability at time t
F(t)	1-R(t)
L(t)	Likelihood function
LRT	Likelihood Ratio Test (for presence of subpopultions, p<1)
a, b	Parameters for inverted gamma distribution (assumed on θ in Bayesian estimation)
α, β	Parameters for beta distribution (assumed on p in Bayesian estimation)
δ_1, δ_2	Parameters for uniform distribution (assumed on β in Bayesian estimation)
D	Observed data
t_{BI}	Burn-in time (length)
t_{cens}	Censor time
t_{miss}	mission length (or time in field under warranty)
C_{setup}	fixed setup cost

C_B	time dependent cost for burn-in per unit per unit time
C_{field}	cost of field repairs during the warranty period
N	number of parts
C_{slip}	Cost of schedule slippage
C_{chng}	Cost of changing batches in burn-in
C_p	Cost of parts
C_{loss}	Cost of loss in the field
C_{incent}	Cost of loss of incentives due to field failures

Chapter 1: Introduction

Motivation

In addition to the failure rates observed during the three typical periods of product life (infant mortality, useful life, and wear-out), unexpected clusters of failures may occur that deviate from expected, failure rates (see Figure 1). These latent failures occur after the infancy period, during the constant failure rate portion of the bathtub curve.

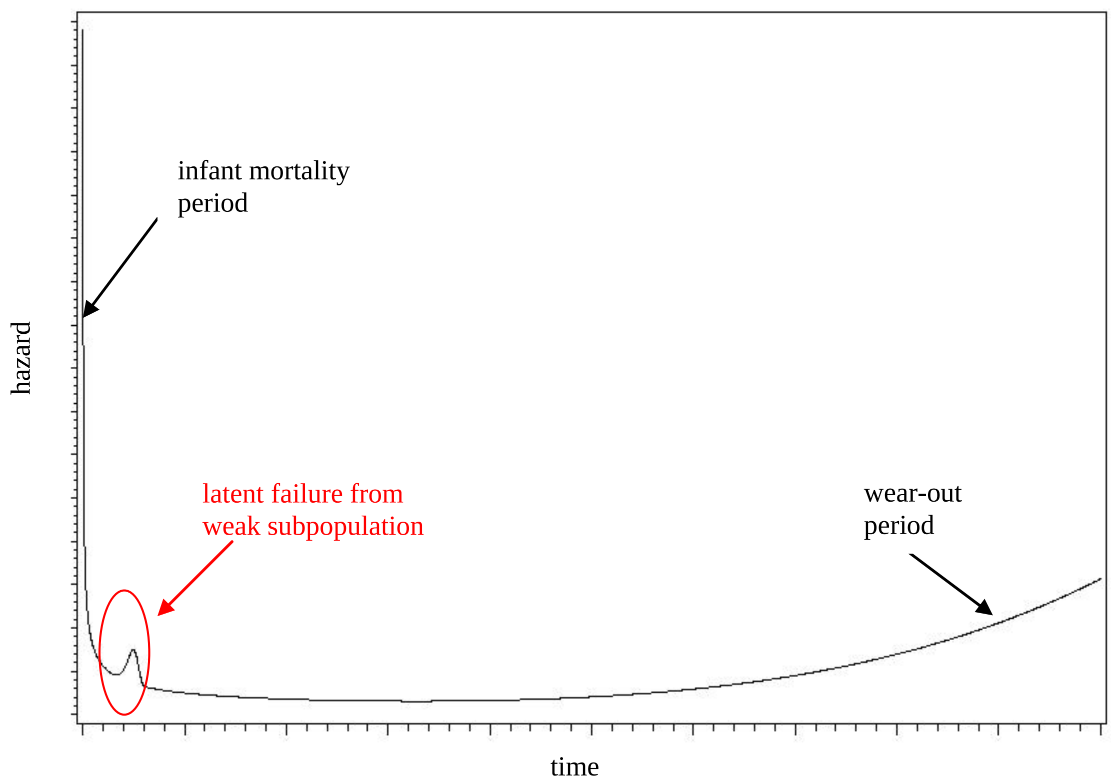


Figure 1. Bathtub curve with additional latent failures

Under most reliability model assumptions, all failures in a population are considered to come from the same distribution. Each individual failure time is assumed to provide information about the likely failure times of all other devices in the population. However, in situations with latent failures, it may be the case that there is a weak subpopulation while some percentage of the devices remained durable, [Paulsen, 2004][Meeker, 1987].

For example, consider some of the thermal cycle test results from the Joint Council on Aging Aircraft/Joint Group on Pollution Prevention (JCAA/JG-PP) Lead-free Solder Project. They tested BGAs under thermal cycles of -20^{o}C to +80^{o}C with dwell times of 30 minutes at hot and 10 minutes at cold. They tested single material solder combinations (SnPb balls with SnPb solder and SnAgCu balls with SnAgCu solder) and mixture situations (SnPb balls with SnAgCu and SnAgCuBi Pb-free solders). They noted that in the mixture situations, particularly the SnAgCuBi case, “had a population that failed very early and a population that had reliability numbers equivalent to the SnPb controls.” [Woodrow, 2006] This conclusion was easy to observe in their experiment because they tested most devices to failure. When all parts are driven to failure, the mixture is evident in Weibull plots by a bend in the Weibull curve, as shown in Figure 2. A single Weibull curve does not fit this data. The weak subpopulation parts and the durable parts follow two different distributions. The durable parts have lifetimes of 8000 to 10,000 cycles while the many of the weak subpopulation parts fail in less than 1000 cycles. Additional investigation revealed that a weak subset of parts was created by "lead poisoning," a separate lead phase in the metallurgy of the joint, particularly on reworked devices. The non-uniform joint resulted in early cracking. [Hillman, 2006]

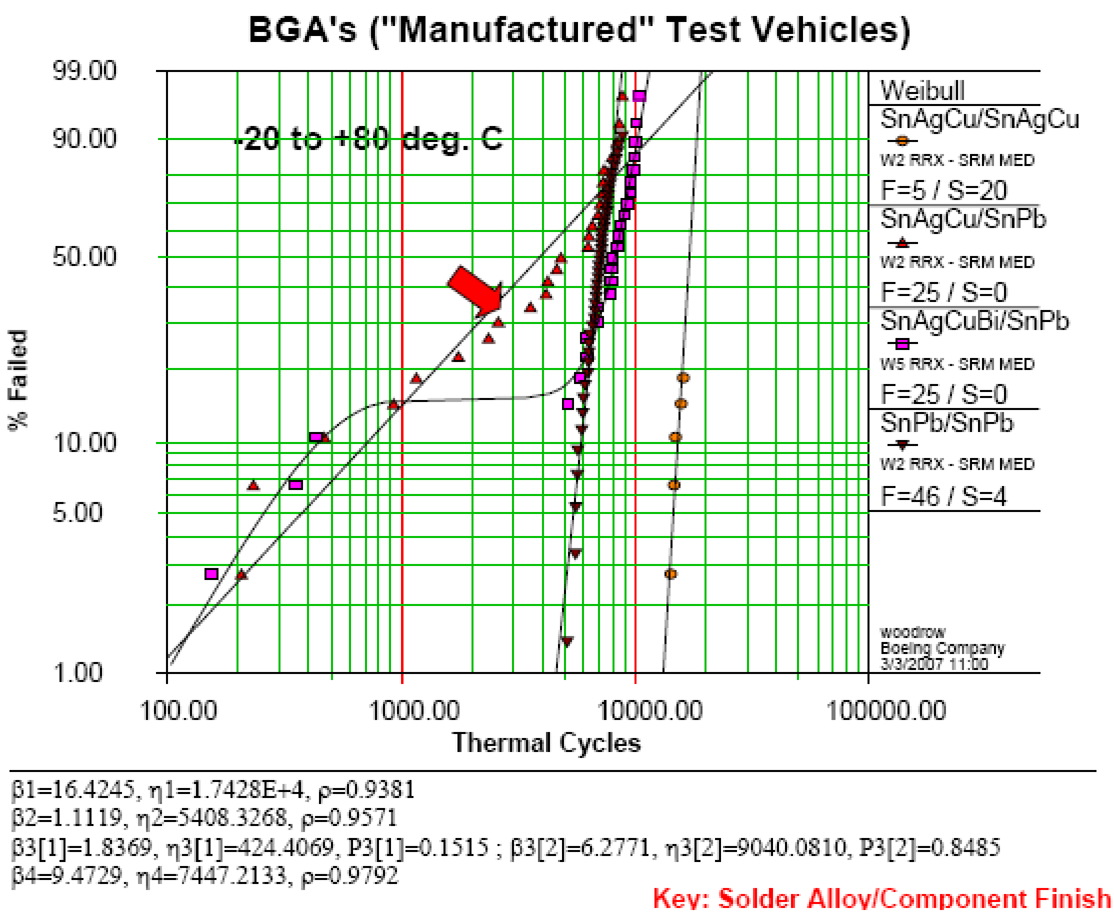


Figure 2. Weibull plot of thermal cycle test data (-20°C to +80°C) from JCAA/JG-PP BGA testing. [Woodrow, 2006]

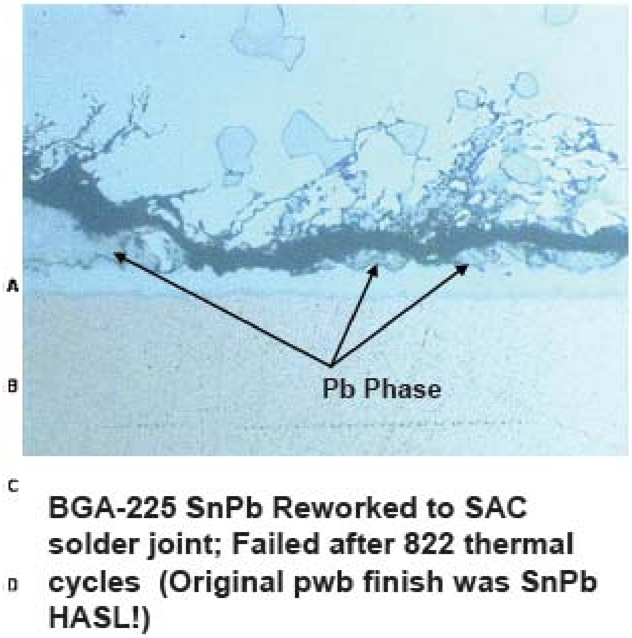


Figure 3. Lead phase and cracks observed in weak subpopulation devices from the JCAA/JG-PP thermal cycling tests of mixed material BGAs. [Hillman, 2006]

In the experiment, there was a lot of information about the Weibull mixture. Because it was an experiment, it was possible to drive all the parts to failure and create Weibull curves for each subpopulation. It was also possible to perform analyses on the failures to determine why the weak subpopulation parts failed early. However, in a production situation, there would be much less information. Perhaps thermal cycling was performed as part of unit assembly and test and some of the weak subpopulation failures occurred. Perhaps their previous testing had been lucky, and there had been very few cases of “lead poisoning.” They did not expect failures in a few hundred cycles. They would not necessarily know if the early failures were suggesting that there was a systemic problem with the entire production run or whether, as was the case in the experimental data, there was only a percentage of devices that were at risk to fail early. Even if they knew of the possibility of early failures, the percentage of parts in the weak subpopulation would vary. For a given production run, it would be unknown how many parts were at risk.

In this example, obviously it would be preferable in a production situation to use one of the single material processes and avoid the root cause all together. However, say that a mixed material situation was unavoidable or the boards had already been built. How many thermal cycles should they perform to optimize field performance? In this case, with the experimental results, they may have some information about the expected values for the Weibull parameters. They also have the failure data from the production lot. How should they estimate the reliability for the production lot? If they didn't have the experimental results and could only depend on the highly censored production lot results, how should they perform the modeling?

The causes of latent failures are many. In this example, it was a random failure mode that only appears in a percentage of the lot. However, they may also be due to lot or batch related production problems, and it may be that traceability has been lost. They may be from a failure mechanism that has always been present in the product at a low rate and was historically masked by other mechanisms. These latent failures only become visible after other process improvements have reduced the constant failure rate enough to unmask them. In other cases, the latent defects are introduced at higher levels of processing and test. This might include problems with connections, exposure of parts to a damaging environment, or electrical overstress during testing.

Ideally, when a mixed distribution is suspected, the weak subpopulation could be identified through determination of root cause, inspecting, and sequestering of impacted devices. But many times, for practical reasons, when the overall population is a mixture of the weak and strong subpopulations; there may be no non-destructive way to identify the weak devices. In the earlier example, there is no non-destructive means to inspect solder balls for the Pb-phase. In another example, a section of a wafer might be contaminated but the problem not discovered until after traceability to row and column is lost for individual die. It is unknown what percentage of the die were contaminated, and if the die are already packaged, inspection may not be possible. Or, early "freak" failures might be observed in qualification or characterization life tests, after infant mortals were already removed. Root cause may be completely unknown. The numbers of and causes for the "freaks" may be ignored for characterizing wear-out for the overall population, but the customer and supplier may want to quantify the number of early failures that can be expected in a production run, [Paulsen, 2004].

In these cases, statistical estimation methods must be used, either with or without root cause information, to quantify the reliability risk to the population, [Meeker, 1987]. The accuracy of these estimates may be critical to the decision making regarding release of the product, but estimation in these circumstances is difficult. There may be very few failures, with the majority of devices simply censored. There may be no way of determining what percentage of the censored devices is in the weak subpopulation, or even verifying that the weak devices represent a subpopulation instead of an early sign that the entire population is weaker than expected. These data limitations can lead to instabilities in some estimation techniques and it has been theorized that some estimation methods may have heavily biased results [Kececioglu, 1998].

Once estimation of the distribution or mixture of distributions is established, decisions need to be made. If it is determined that there is a mixture, screens and additional burn-in may be possible for removing the weak subpopulation. However, the length and cost of those tests must be weighed against the risk of field failures and the potential that screens and additional burn-in may remove useful life from products in the main subpopulation. Appropriate cost models must be developed to determine the best course of action. The model needs to include the cost of screening, the impact of the screening on the product life-cycle, the cost of field failures, and uncertainties in the reliability modeling.

Background

If it can be determined that there is a weak subpopulation, one response might be to develop a test or screen by which the subpopulation can be detected and removed. Ideally the application of such a screen would be based on some sort of easily detected characteristic, such as a particular wafer or processing run. It might be that the devices can be visually screened or put through some other non-destructive physical analysis, such as EDX or a bond pull. However, for many of these situations there is either not enough information about the failure mechanism to develop such a test or simple inspection is not possible. It may be the case that the only feasible means of removing the weak subpopulation is to perform an extended burn-in or Environmental Stress Screen (ESS). It may also be the case that performing such screening is necessary to get additional data for the modeling of the parameters of the mixed populations.

ESS and burn-in, although having similar goals, are not identical. ESS typically uses high stress conditions to cause certain types of failure mechanisms to occur more quickly than under normal operating conditions [Chan, 1994]. Since different types of stresses accelerate different mechanisms, the underlying cause of the weak subpopulation must be understood before an ESS can be effectively applied. Burn-in is typically performed under much lower stresses, much closer to the operating conditions [Jensen and Petersen, 1982]. Although performing these tests will have different logistics and costs, from a modeling perspective they can be treated virtually interchangeably. For example, for ESS, it is necessary to have an acceleration model to convert time under ESS into operational time. Once this conversion has taken place, both ESS and burn-in represent consumption of early life for the purposes of avoiding latent failures.

Research Goals

There are two primary objectives for this dissertation:

1. Determine the best statistical methodology for evaluating latent failures, including the development of a methodology for testing whether a weak subpopulation exists.
2. Develop cost models and decision guidelines to evaluate responses to latent failures, including extended burn-in, product replacement, or use-as-is.

Approach Summary and Analysis Strategy

Different estimation methodologies for mixed Weibull distributions were characterized and compared using a series of Monte Carlo analyses. The amount of bias and random error for the different methods was examined. These results have been used to guide recommendations about when different methods are most appropriate.[1]

A cost model was developed and characterized for a variety of mixed Weibull situations and input costs. The impact of the biases and random errors associated with the different estimation methods on the cost estimates was examined. Recommendations are made about the use of the different methods when cost-based decision making is used.

[1] Throughout this dissertation, references are made to electrical parts and boards. However, these are merely an example of the situation where these techniques might be applied. The results apply equally well to mechanical situations or even to survival problems in medical applications.

In Chapter 2, previous work on mixed model estimation, screen length optimization, and cost models will be summarized. Chapter 3 introduces the Monte Carlo methodology used to examine three estimation methods. Chapters 4 through 6 present the Monte Carlo results for each method. Chapter 7 develops a burn-in cost model. Chapter 8 discusses the impact of modeling error for each method on the decision to perform burn-in using both a minimum reliability and a cost decision rule. Finally, Chapter 9 describes plans for future work.

Chapter 2: Previous Work

Estimation of mixed population parameters, confidence bounds, and the cost and optimal screen length are all areas that have been explored to varying degrees in the existing literature. However, there are still many open questions in all three areas. This section reviews the current research for each area and then discusses what is lacking to make good decisions when latent failures are observed in electronic products.

Methods of Estimation for Mixed Models

The basic form of a mixed distribution model is:

y_i from the distribution $F(y)$
z_i from the distribution $\mathrm{G}(z)$
x_i from the distribution $H(x) = p \times F(x) + (1-p) \times G(x)$

where p is the mixture ratio. For the reliability and electronics world, the two distributions are typically taken to be exponential, Weibull, or log-normal: [Ebeling, 1997], [Chan, 1999]

$$
\begin{aligned}
R(t) &= p \times F(t) + (1-p) \times G(t) \\
&= p \times \exp\left(-\frac{t}{\theta_1}\right)^{\beta_1} + (1-p) \times \exp\left(-\frac{t}{\theta_2}\right)^{\beta_2} \quad \text{(Weibull)}
\end{aligned}
$$

or

$$
R(t) = p \times \left[1 - \Phi\left(\frac{1}{s_1} \ln \frac{t}{t_{med,1}}\right)\right] + (1-p) \times \left[1 - \Phi\left(\frac{1}{s_2} \ln \frac{t}{t_{med,2}}\right)\right] \quad \text{(Lognormal)}
$$

In some cases, in both the theoretical and the reliability fields, these are simplified, by assuming that samples in the strong subpopulation have a reliability of 1, independent of time. This assumption is sometimes called the limited failure population model [Meeker, 1987], [Hirose, 2005]. This is an appropriate assumption if the MTTF for the strong subpopulation is much larger than the maximum time of interest. It reduces the number of parameters to be estimated from five to three.

Work has been done on estimating mixture models in a variety of fields since the 1970s. Much of the early work was in the realm of theoretical statistics and biostatistics and focused on cases where only the proportion of different elements of the mixture is of interest or where small samples from each of the subpopulations are observable independent of each other, as well as in the mixture (i.e., samples x_i, y_i, and z_i). [Hosmer, 1973], [Murray, 1978], [Hall, 1984]

Even with data from each of the underlying distributions, this problem can be difficult, particularly if there is uncertainty about the form of the distributions. [Qin, 1999] proposes a likelihood type method for estimating mixture proportion under a proportional hazard model. However, Qin barely touches on the issue of estimating the distribution parameters.

Although the work in theoretical statistics and biostatistics may have some applicability in electronics reliability, it does not address many of the problems faced in a production situation. First, it is unlikely that data is available from each of the underlying populations; i.e., it is unlikely that the weak subpopulation can be isolated and tested without some of the main population being present. Second, the estimates of the parameters, at least for the weak subpopulation, are very important. Without information about the mean time to failure and spread of failures in the subpopulation, it cannot be determined whether products from the weak subpopulation will meet the reliability requirements, whether they can be screened, or whether the mixed population must be quarantined.

Mixed distributions have also been examined in the reliability engineering and electronics fields, the earliest methods of estimation were based interpretation of plots. Surveys of these methods are presented in [Cran, 1976], [Jiang, 1992b], and [Jiang, 1995]. The technique involves using Weibull-paper and looking at bends in the data. Various methods exist for determining the appropriate two Weibull curves on the graph, but the methods depend on "eye-balling" the data and are subject to the same sorts of errors that occur when single Weibull parameters are estimated entirely from graphing. Recent papers acknowledge that extrapolations from these methods can have large errors [Kececioglu, 1998].

With today's computing power, the most obvious method of estimating the parameters is maximum likelihood estimation (MLE) [Meeker, 1987] [Hirose, 2005]. Due to the complexity of the reliability function for the mixed population, there are no closed form solutions for the maximum likelihood estimates of the parameters.

Instead, numerical methods, such as the Expectation and Maximization (EM) algorithm or grid searches are used [Kececioglu, 1998], [Jiang, 1992b].

There are potential problems with MLE, particularly in cases where only a limited number of failures are observed. Since numerical methods must be used to maximize the likelihood, the speed of convergence may be an issue. In some cases, calculation of the MLE parameters may require more computing power or more sophisticated software than is available to some practitioners. There may also be problems with convergence of the estimation algorithms or with local minima if starting values are not chosen appropriately [Schifani, 1999] [Kececioglu, 1998].

Even more importantly, MLE on samples with few failures can produce biased estimates, even when it is used to estimate only one distribution's parameters. [Abernethy, 1999] [Mardia, 1999]. Very few papers actually quantify the amount of this bias.

[Hirose, 2005] appears to be the first to systematically quantify bias in the MLE parameter estimates for mixed distributions. Simulations were run on a case where the strong subpopulation had a reliability of 1 and was ignored in the modeling. Hirose's results demonstrated that the MLE estimates of the mixture proportion and estimates of the mean time to failure (MTTF) were highly correlated for failure times from a mixture of exponential distributions (see Figure 4). This correlation is a major contributor to the bias in the parameter estimates.

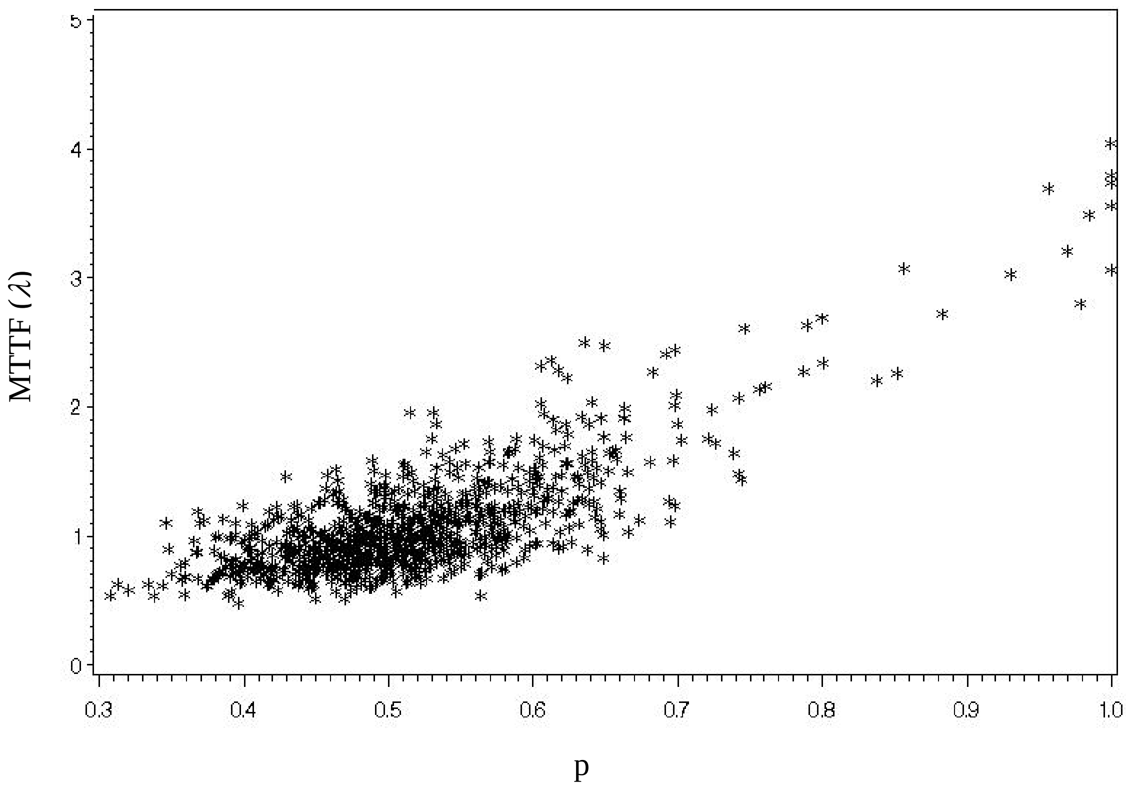


Figure 4. Graphic produced from reproducing the methodology in Hirose [Hirose, 2005]. The plot shows the estimated mixture ratio (*p*) versus the estimated MTTF (λ) when the true mixture ratio was 50% and the two MTTF were 1 and 1x10^8.

In an attempt to reduce the correlation and the bias, Hirose explored changing the assumptions on the mixed model by looking at the impact of removing the constraint that the two subpopulations sum to 100% (Trunsored model), ignoring the possibility of a mixture (censored model), and using a conditional likelihood for failures only (truncated model). Hirose only looked at cases where the weak subpopulation was at least 50% of the overall population, and demonstrated that the estimate of the MTTF could be highly biased and the amount of bias varied with the true mixture proportion. Hirose concluded that if the proportion of the weak subpopulation was close to 1 then it was better to ignore the possibility of the mixture (censored model) and otherwise it was best to use the Trunsored model.

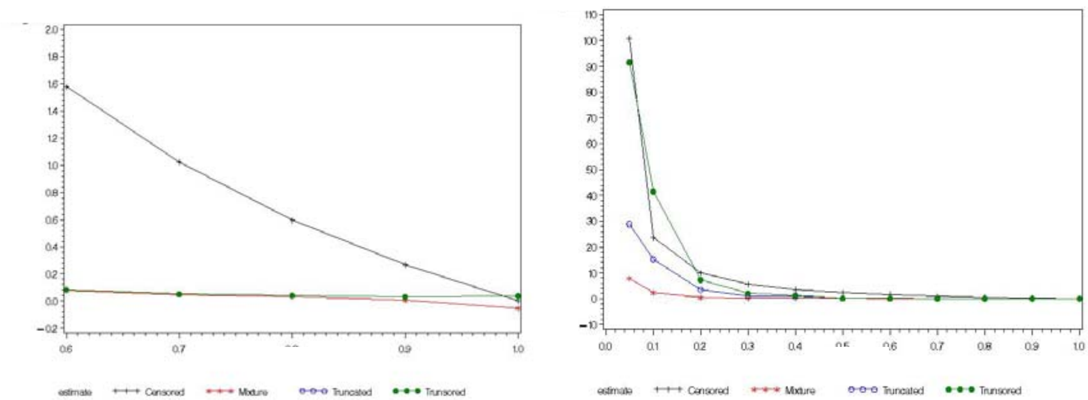


Figure 5. Graphic produced by reproducing the methodology in Hirose over the range of mixture proportions of 0.6 to 1. Bias in the estimates of MTTF (λ) versus the true mixture proportion (*p*) for the four model assumptions considered in [Hirose, 2005].

The figure on the left of Figure 5 reproduces the results from Hirose. The figure on the right expands those results to smaller mixture ratios. Expanding the simulations that Hirose performed shows that for situations where the weak subpopulation is very small, use of the Trunsored model produces a large bias in the estimate of the MTTF. Using the standard mixture model, with the constraint that the two subpopulations sum to 100%, has substantially smaller biases, at least in the case where the strong subpopulation has a reliability of 1. It is expected that varying the other parameters will have similar impacts on the amount of bias. However, Hirose's results to not demonstrate under what circumstances the biases are substantial enough to be a true concern.

Other iterative procedures have also been proposed that claim, although do not prove, that they eliminate the bias and convergence problems that may exist for MLE

[Kececioglu, 1998] [Perlstein, 2001] [Schifani, 1999]. Without quantifying the bias in the MLE or their new methods, it is unclear whether these methods are necessary or even provide any improvement over MLE.

Optimizing Test and Screen Lengths

Optimal lengths for burn-in and ESS have been explored for many years, particularly from the perspective of removing infant mortals and using decreasing failure rate Weibull curves. There are a number of different criteria that are used for optimization. In some cases, mean residual life, post screen, is taken as the objective function. However, if there is a set lifetime of interest, such as a mission length or a warranty period, minimizing the chance of failure during that time period may be the more appropriate measure. In still other cases, there may be a requirement that the probability of survival over a time period is at least some value, such as 0.90 reliability at 100,000 hours. [Leemis, 1990] Although these may be appropriate measures of optimal screen length, in most real world applications, another important consideration is the lifecycle cost impact.

The idea of balancing the lifecycle cost of field failures with the cost of performing tests to remove the latent failures has been explored with varying degrees of success. In some cases, a single distribution is assumed, and the test is designed to remove some percentage of infant mortals from a decreasing failure rate Weibull distribution, [Mi, 1996]. Others consider latent failures and mixed distributions, including mixed Weibull distributions [Perlstein, 1989] [Reddy, 1994] [Pohl, 1995] [Yan, 1997] [Kim, 1998]. However, they also frequently assume that the parameters of the Weibull (and the costs) are known with a high degree of certainty and make other simplifying assumptions.

Summary of Weaknesses of Current Research

As illustrated in the literature review, there are weaknesses in the current research. Although there is mention of maximum likelihood estimation methods producing biased results, there is little guidance on how biased the results might be or under what conditions bias is most likely to occur. Small sample sizes are mentioned as a possible factor in likelihood of bias, but it is unclear how many failures must be observed to avoid the problem. Understanding of, and attention to these problems also varies: some authors do not even mention the possible problems with the MLE, while others propose entirely new estimation schemes to avoid the problems with the technique.

Many simulation studies have been conducted to look at the effectiveness of likelihood ratio methods for confidence bound estimation on single distributions, but there does not appear to be similar work on its effectiveness on mixed distributions. It is unclear whether some of the proposed correction factors to the likelihood ratio method would be appropriate under mixed conditions.

Finally, it is unclear how uncertainties and errors in the parameters of the mixed distribution might impact the cost tradeoff regarding further screening and testing.

The lack of industry consensus on how to address mixed distributions and the possible cost of releasing a product that contains devices from several failure time distributions makes it difficult for decision makers to know how to proceed when latent failures appear in their products.

Chapter 3: Monte Carlo Methodology

In order to determine which estimation methods are most accurate, Monte Carlo simulations were developed to estimate parameters under a variety of known underlying conditions[2]. The cases were selected to be similar to those presented in [Hirose, 2005], but also to expand them to a larger range of situations.

The true distribution of the data was always taken from a limited failure population. This is a simplification of the general mixed Weibull where it is assumed that samples in the strong subpopulation have reliability of 1, independent of time. This is a reasonable model for at least some electronics where the product has been designed to have very little probability of failure during normal operating life with margin. The overall reliability had the form:

$$R(t) = p \times \exp\left[-\left(\frac{t}{\theta}\right)^{\beta}\right] + (1 - p)$$

Values of the mixture ratio, *p*, were varied. In order to allow greater sampling under difficult estimation situations, more values of *p* were examined close to 0 and 1. Values of *p* included in the simulation were: {0.05, 0.10, 0.15, 0.20, 0.25, 0.50, 0.75, 0.80, 0.85, 0.90, 0.95, 1.00}. The characteristic life, θ, was taken to be close to 1 for all cases.[3] Both a constant failure rate case (β=1) and increasing failure rate (β=3)

---

[2] All simulations and estimations were performed using the SAS Software System version 9.1 (Copyright (c) 2002-2003 by SAS Institute Inc., Cary, NC, USA)

[3] For the case of increasing failure rate, the characteristics life was adjusted from 1 to maintain a consistent number of observed failures.

case were examined. A censor time of 2 (about twice the characteristic life) was chosen .[4] Beyond that time, the exact failure time of surviving samples was not known; they were only included in the models as surviving past t =2.

For each case, 200 datasets, each containing 100 observations, were generated[5]. The estimation method was applied to each dataset. The bias was calculated as the average difference between the estimated value and the true value; the random error was calculated as the standard deviation of the difference. In addition, the impact of these parameter estimates on the estimates of reliability was also examined. The bias was calculated as the average difference between the estimated value and the true value; the random error was calculated as the standard deviation of the difference.

[4] Consistent with the censor time in [Hirose, 2005].
[5] This was based on the sample size in [Hirose, 2005] and should be a conservative example of production lot sizes. If sample sizes are larger, more failures will be observed which will only decrease the errors in the methods.

Chapter 4: Maximum Likelihood Estimation

This chapter summarizes the results of the Monte Carlo simulations using Maximum Likelihood Estimation (MLE), including the impact of variation is the underlying size of the weak subpopulation and the censor time. The impact of likelihood ratio tests for determining the presence of a weak subpopulation is also explored.

Methodology

The likelihood was maximized using the Newton-Raphson optimization technique as the default. Although some of the literature has proposed other methods of optimization, particularly for the purposes of speed of convergence ([Jiang, 1992b] [Kaylan, 1981]), exploration of the impact of optimization technique was not performed.

Results

Most of the alternative methods were developed as a result of concerns about bias or errors in the MLE method. However, the papers developing them do not establish the magnitude of those errors. Thus the first step in this analysis is to examine the performance of the MLE method under a variety of conditions.

Some bias was found in the MLE estimates of the mixture ratio. The amount of bias depends on the number of observed failures. The more failures observed, the smaller the observed bias. For example, when 30 or more failures in the 100 samples were observed the average bias on p was only about 2%, and 90% of the predictions were within 20% of the true p. However, if the true p is very small, the number of observed failures in 100 samples will be small. In the case of the true mixture ratio being 0.1, on average, only 8 failures were observed. The bias on p was 32%, and only 50% of the estimates were within 20% of the true mixture ratio. These high errors are illustrated in **Figure 7**. Because the estimates of all parameters are correlated, the bias and random errors in the other Weibull parameters, particularly for small mixture ratios and few observed failures are even worse. The errors are summarized in Table 8.

Table 1. MLE: Errors in Parameters and Reliability Estimates for θ=1, β=1, censor time=2.

True p	Bias in p (SD)	Bias in θ (SD)	Bias in β (SD)	Bias in R(10)[6] (SD)
0.05	0.09 (0.26)	28.24 (160.30)	43.45 (428.99)	-0.02 (0.07)
0.1	0.09 (0.26)	10.69 (60.76)	2.74 (63.02)	-0.02 (0.18)
0.25	0.07 (0.21)	1.80 (8.45)	0.08 (0.28)	-0.03 (0.11)
0.5	0.04 (0.15)	0.29 (0.96)	0.04 (0.19)	-0.03 (0.12)
0.75	0.03 (0.11)	0.11 (0.38)	0.02 (0.14)	-0.03 (0.11)
0.9	0 (0.08)	0.01 (0.22)	0.03 (0.13)	0 (0.08)
1	-0.03 (0.04)	-0.06 (0.13)	0.05 (0.11)	0.03 (0.04)

[6] 10 times longer than the censor time. In truth all the at risk subpopulation has failed by this point, so R(20)=1- p .

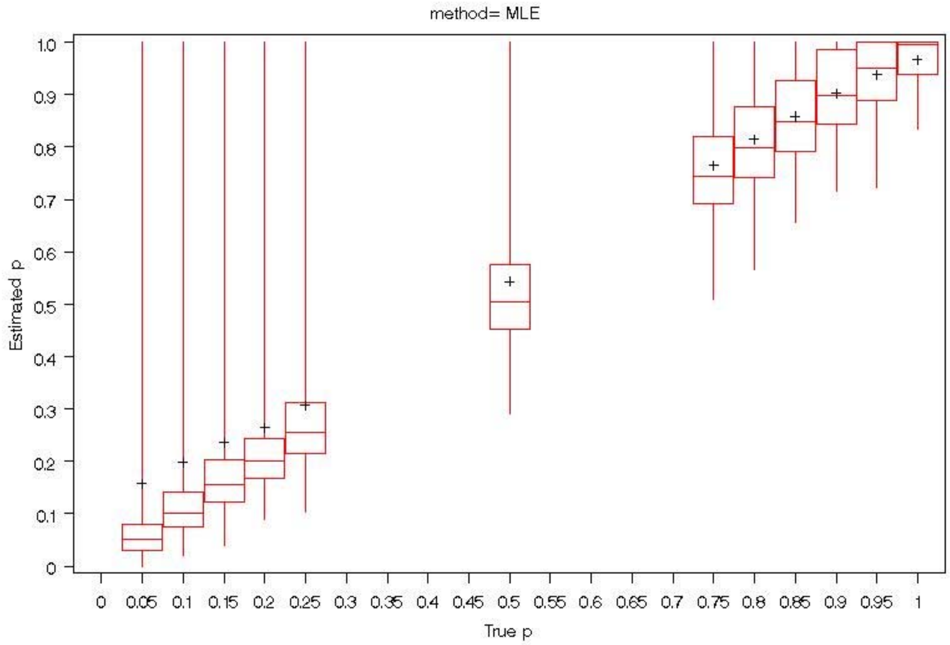


Figure 6. Performance of MLE relative to mixture ratio p. The mean estimates are represented by "+". The boxes represent the mean and interquartile range. The whiskers represent minimum and maximum.

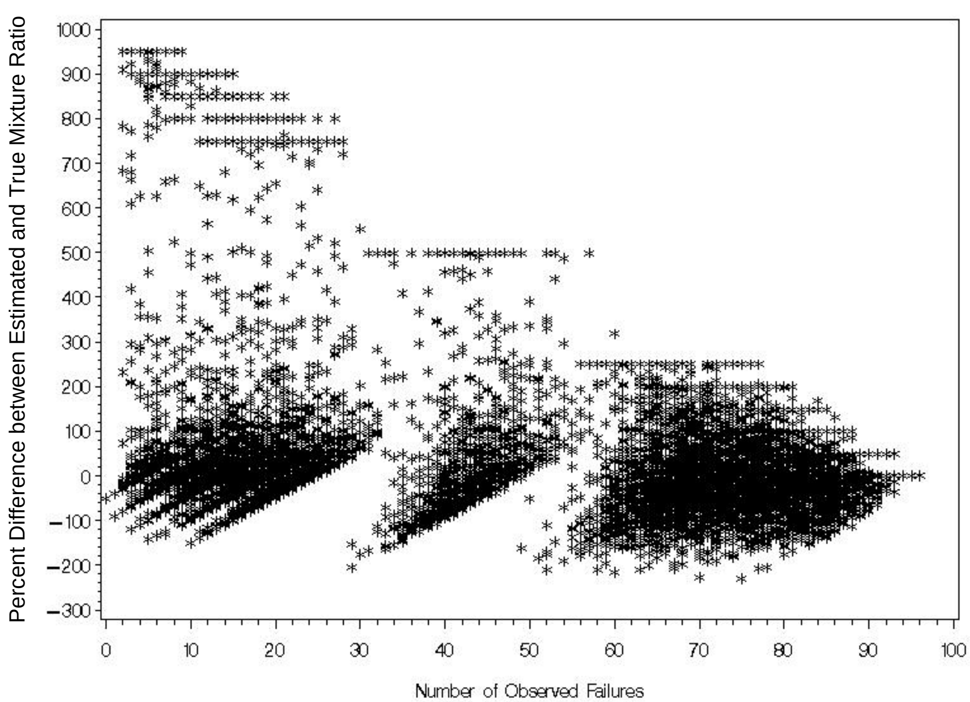
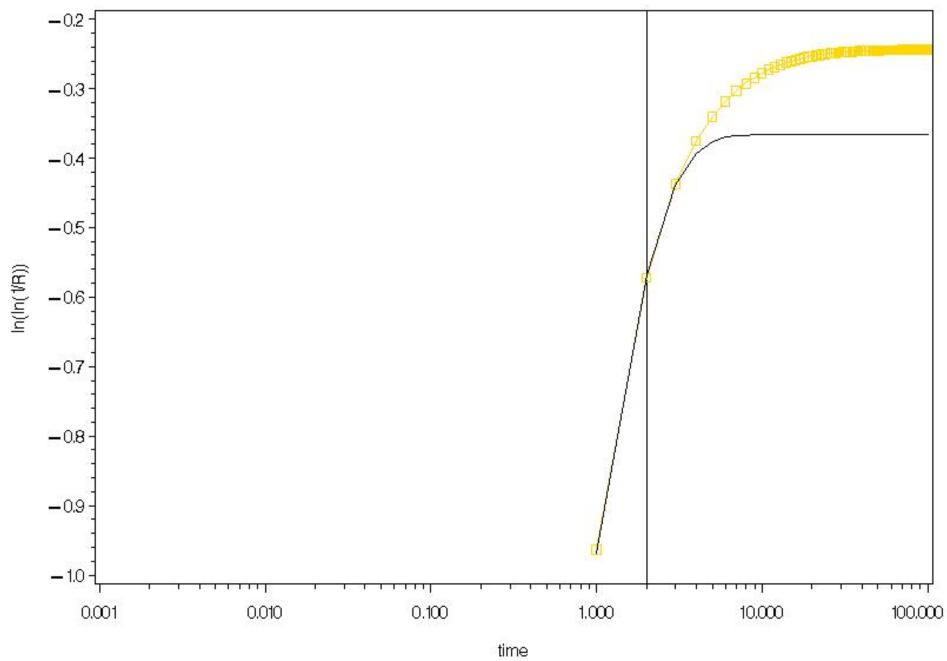


Figure 7. Percent Error in Estimates of Mixture Ratio p when the true p is 0.1. The 100% cases are situations where the mixture ratio was estimated to be 1.

However, in most applications, the error in the parameter estimates is much less important than the errors in the reliability estimate at some time point. The MLE method performs much better with regards to this criterion (see Table 8). Even in the case where the true mixture ratio is 0.1, where there were very high errors in all three Weibull parameters, the bias in the reliability estimate is about a 5% under prediction in the reliability, as illustrated in Figure 8. Although it would be preferable to have no bias, the fact that the method tends to under predict the reliability is typically considered preferable to over-prediction in practice as it is more conservative.



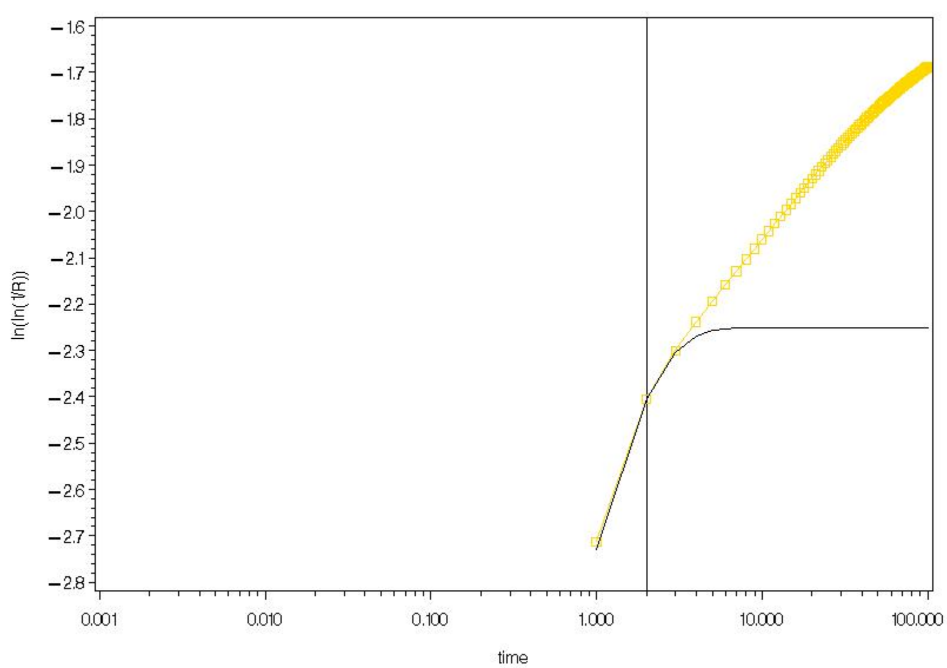


Figure 8. Bias in reliability estimates from the MLE method under assumed mixed Weibull distribution. True conditions were a mixed exponential distribution with mixture ratio of 0.5 (top) and 0.1 (bottom). The censor time was 2. The true reliability is represented by the solid line and the prediction by the dashed line.

Caution should be taken when the true mixture ratio is close to 1. In these cases, the MLE method tended to predict a higher reliability than truth (see Figure 9). This is because the model form allows for a mixture even when all devices are at risk. The bias in these cases can be substantially reduced by performing a Likelihood Ratio Test (LRT) with one degree of freedom to test if the mixture ratio p is significantly different from 1. There are problems with performing a simple likelihood ratio test on MLE situations, as described in [Hirose, 2005] and [Maller, 1996], because the chi-squared distribution is a poor approximation of the true distribution, particularly near boundary conditions. However, it was also felt that this simple test is the one most likely to be incorporated into practice if the MLE is used. Thus it was applied without any correction. For cases where the mixture ratio is close to 1, even the inaccurate, simple chi-squared LRT improved the reliability estimate errors. The reduction in bias when the true mixture ratio is 1 is illustrated in Figure 9.

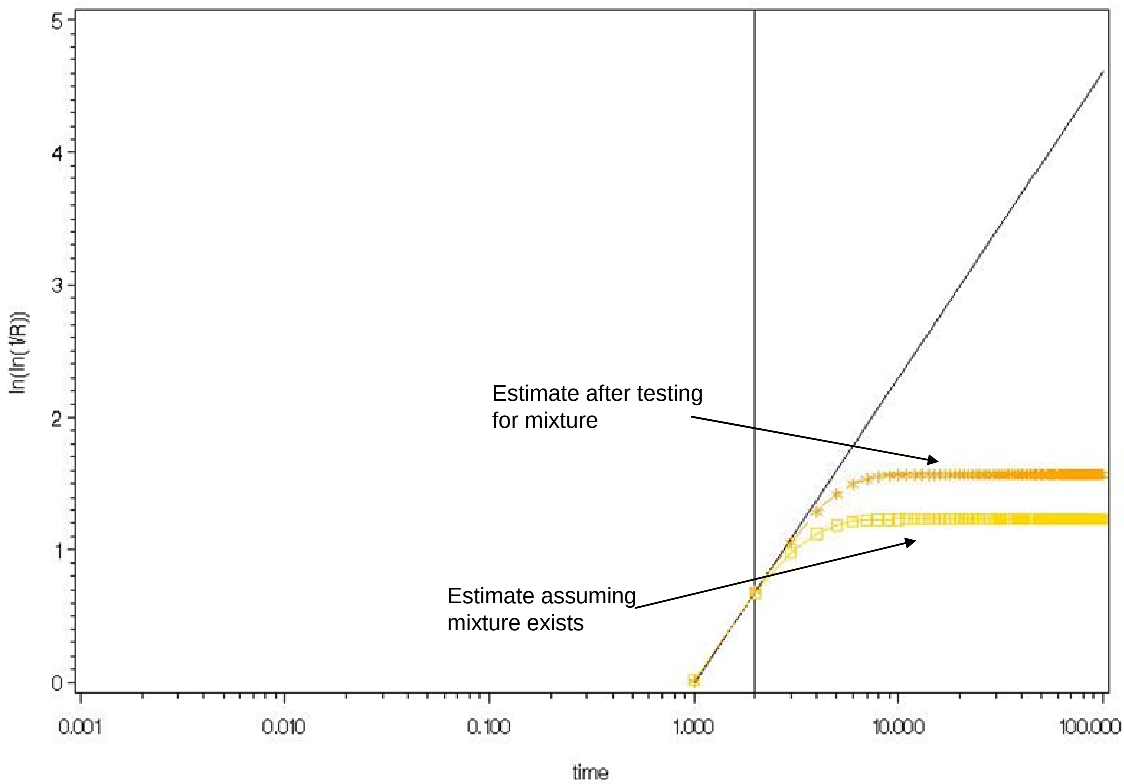


Figure 9. Bias in reliability estimates for the MLE method under assumed mixed Weibull distribution. True conditions were an exponential distribution (with mixture ratio of 1 – no real mixture). The censor time was 2. The true reliability is represented by the solid line, the prediction without testing the significance of *p* by the yellow squares, and the prediction after testing for the significance of *p* by the orange stars.

There is a complication with using the likelihood ratio test on *p* when it is small. When the true mixture ratio is small and few failures are observed, the test will frequently determine that the estimate *p* is not significantly different from 1. In these cases, the model will predict reliabilities much lower than the true reliability at extrapolated times. Applying the LRT when *p* is small increases the bias and random error substantially, as illustrated in Figure 10. The MLE method without testing to see if the mixture ratio *p* is less than 1 predicts slightly lower reliability than truth up until p_0 is about 90%. Above 90%, it predicts slightly higher reliabilities than truth.

When the LRT is applied, the model predicts much too conservative reliabilities unless *p* is very close to 100%. Performing the test preserves the conservatism of the model when *p* is greater than 90%, but at a high cost when *p* is actually small. Further work is needed on adjusting the LRT to be more accurate over a variety of *p* values. It is currently recommended that the test only be applied when the predicted *p* is greater than 0.9 or when additional conservatism in the reliability estimates is needed.

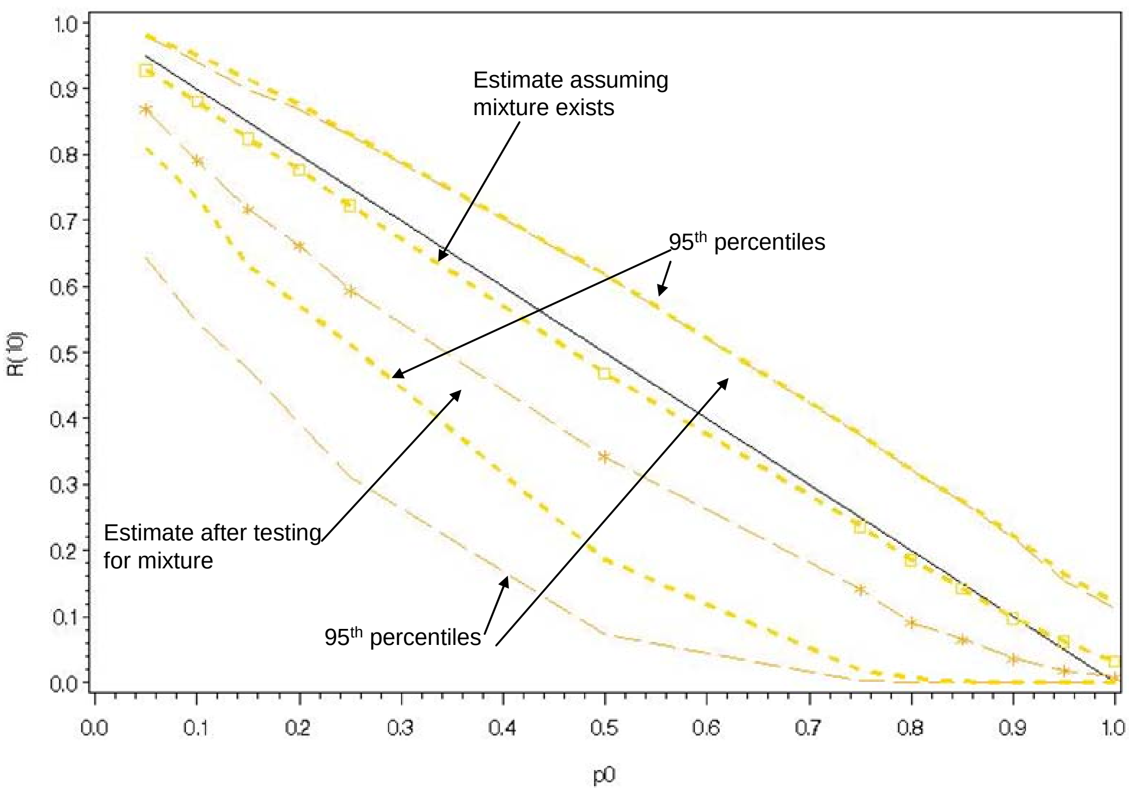


Figure 10. Black is the true reliability. Gold square is without test. Orange stars is with test.

To examine the impact of an increasing failure rate for the weak subpopulation, cases with β=3 were also examined. If β=3 and θ=1, then the reliability of the weak subpopulation at the censor time of 2 would be very close to 0. Since this would result in more observed failures, and likely less error in the estimates, the characteristic life was also adjusted. At β=3 and θ=1.6, the reliability of the weak subpopulation at the censor time of 2 would be 14%, the same as in the case of β=1 and θ=1. Comparisons of reliability predictions at extrapolations to time=4 and time=10 are shown in **Figure 11**. An increasing failure rate produces larger bias and random errors for the MLE model, particularly when p is small and few failures have been observed. At small mixture ratios, the bias can double, but the bias is still in the conservative direction. The random error can be very large, particularly when extrapolating far from the censor time. Performing tests on the significance of p does not improve the errors.

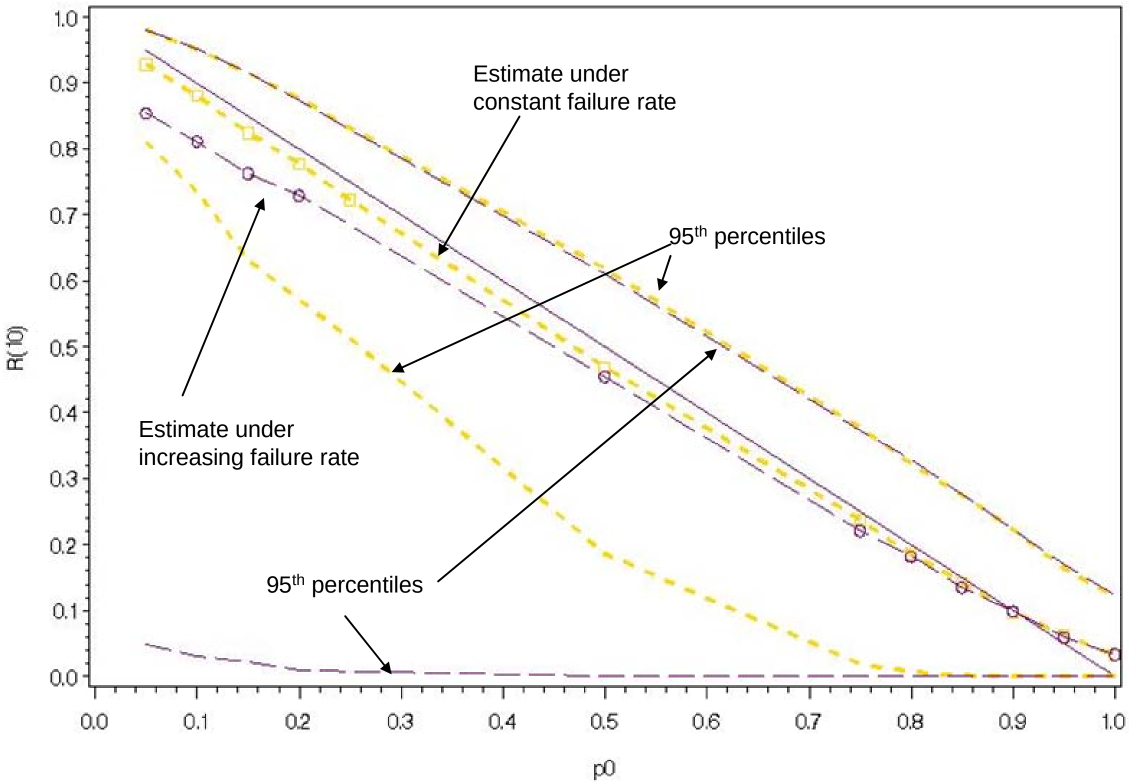


Figure 11. Black is true reliability. Gold is the predictions when the true failure rate is constant. Purple is the prediction when the true failure rate is increasing.

The impact of censor time was also investigated and summarized in Figure 12. As might be expected, the later the censor time, the better the estimate because there will be more information about the distribution of the failures. This is true whether one is predicting a relatively small distance from the censor time or substantially farther from the censor time. In a few cases, when the censor time is particularly short and the extrapolation time is close to the censor time, the bias is in the direction of over-predicting the reliability.

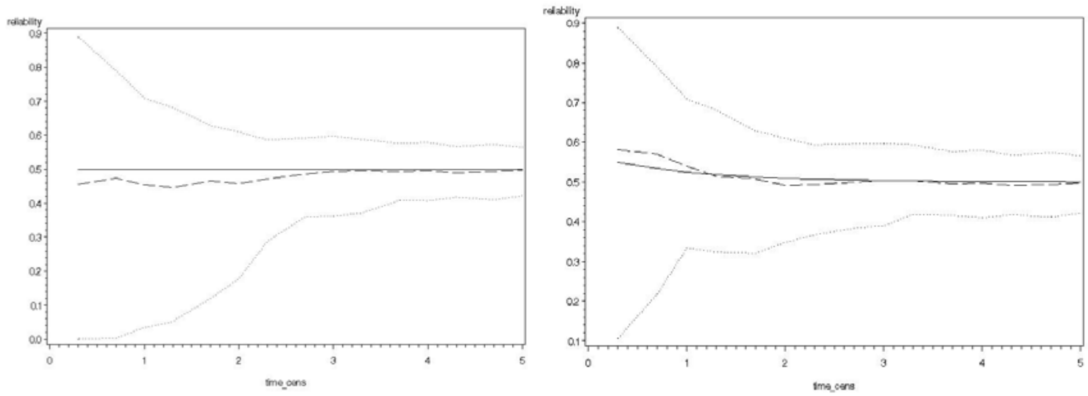


Figure 12. Impact of censor time when $p_0 = 0.5$. Errors in R(10) are on the right; errors in R(censor time+2) are on the left. Solid line is true reliability, dash is average predicted, dots are 5th and 95th percentiles.

Chapter Conclusions

Overall, the MLE ratio does reasonably well in predicting the reliability. However, there is room for improvement, particularly in the errors in the predictions of the mixture ratio and Weibull parameters and in cases where few failures are observed. Now that baseline performance of the MLE method has been established, it will be considered in the next two chapters whether alternate methods proposed in the literature reduce these errors.

Chapter 5: Trunsored Estimation

This chapter summarizes the results of the Monte Carlo simulations using Trunsored Estimation, including the impact of variation in the underlying size of the weak subpopulation and the impact of likelihood ratio tests for determining the presence of a weak subpopulation. These results are compared to those on MLE and the results presented in [Hirose, 2005].

Methodology

The Censored MLE estimation method ignores the possibility of a mixture and only estimates a single Weibull. The Truncated estimation method uses a conditional likelihood on the failures and ignores censored observations:

$$L = \prod_{i=1}^{r} \frac{f\left(t_i;\{\theta,\beta\}\right)}{F\left(T;\{\theta,\beta\}\right)}$$, where r is the number of observed failures

Trunsored method, as developed by [Hirose, 2005], attempts to unify the censored and truncated estimation methods. It removes the constraint that the mixture parameters for the different subpopulations sum must to 100%. As shown in [Hirose, 2005] for the mixed exponential, solving the likelihood equation for the Trunsored model corresponds to the solutions of a truncated model[7], and the solution also corresponds to the censored model if the mixture ratio is set to 1. This means that it unifies the censored and truncated models and allows better use of the likelihood ratio

[7] Parameter estimates are the same; the likelihood values differ.

test. The likelihood was maximized using the Newton-Raphson optimization technique as the default.

Results

The first step in examining the Trunsored method was to attempt to replicate the results on mixed exponential models reported in [Hirose, 2005]. He compared the performance of Trunsored and MLE for a mixture of exponentially distributed life-times. His work also focused on situation where the weak subpopulation made up at least 60% of the overall subpopulation because he was interested in the statistical behaviors near the boundary condition, when the weak subpopulation was close to 100% of the overall population.

[Hirose, 2005] only reports bias and standard errors for characteristic life (failure rate). However, for at least that parameter, the simulations generated for this dissertation produced very similar results to those reported in [Hirose, 2005]. For example, Table 2 provides a comparison of the results for a true mixture ratio of 0.9.

Table 2. Comparison of [Hirose, 2005] and Results from this Dissertation on Mixed Exponential with True Mixture Ratio 0.9

	Bias in θ	STD in θ	5%	95%
[Hirose, 2005]	0.013	0.195	0.729	1.364
Replication for Dissertation	0.033	0.248	0.731	1.536

Once it was determined that the results from this dissertation were in line with those reported by [Hirose, 2005] for comparable cases, additional cases for mixed exponential were examined. As shown in Figure 13, the errors increase dramatically as the mixture ratio becomes small. This is believed to be due to the fact that under low mixture ratios, there are few failures observed and the method can become unstable.

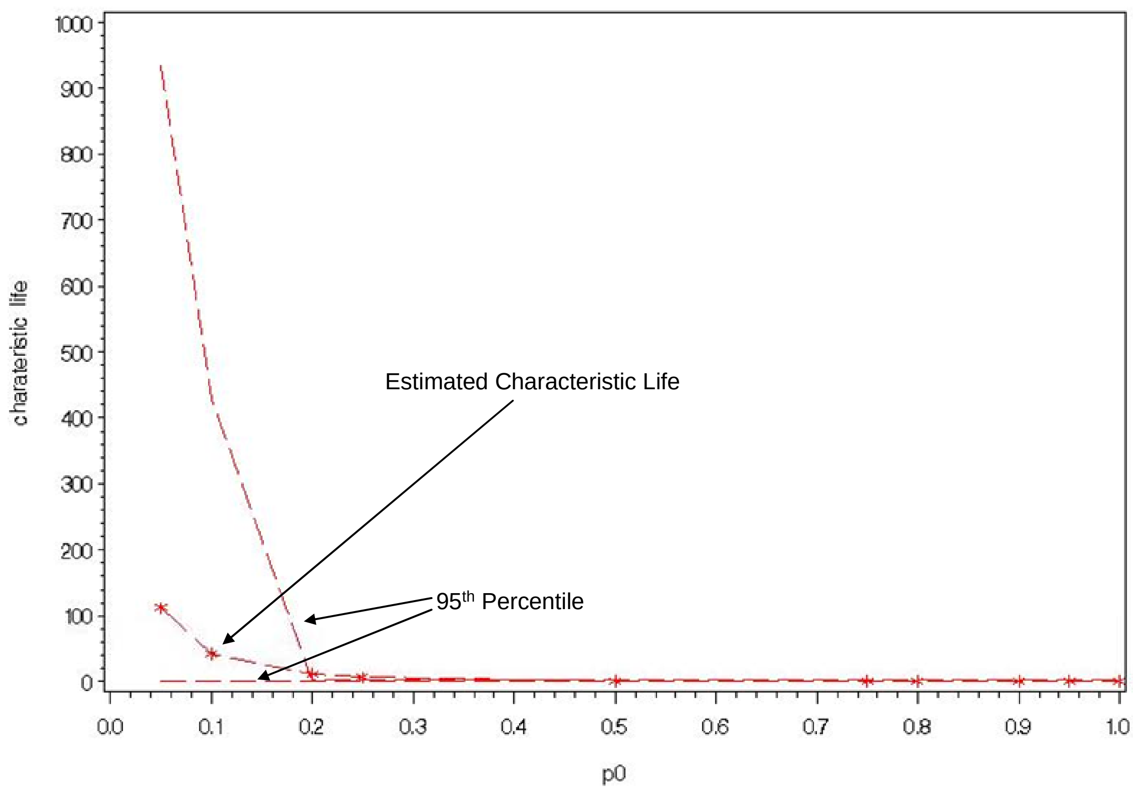


Figure 13. Estimated Characteristic Life for Assumed Mixed Exponential. Results greater than or equal to 0.6 correspond to those reported in [Hirose, 2005]. Stars are biases and dashes without stars are the 5th and 95th percentiles.

Finally, [Hirose, 2005] results were generalized to assuming a mixed Weibull distribution, instead of a mixed Exponential. The addition of the estimation of the shape parameter results in higher error in the parameter estimates, for both the MLE and Trunsored methods but the errors are larger for the Trunsored method. For example, for the case when the true mixture ratio was 0.5, when mixed exponential model is assumed, the bias in the Trunsored of *p* was about 0.1 or about 20%. However, on the same data, when a mixed Weibull model is assumed, the Trunsored of *p* had a bias of 0.5 or about 100%. Similarly, the bias on the estimate of the characteristic life was about 0.02[8] for the mixed exponential but was 16 for the mixed Weibull case.

---

[8] Note that the bias is for $1/\lambda$ in the exponential case to be consistent with our parameterization of the Weibull.

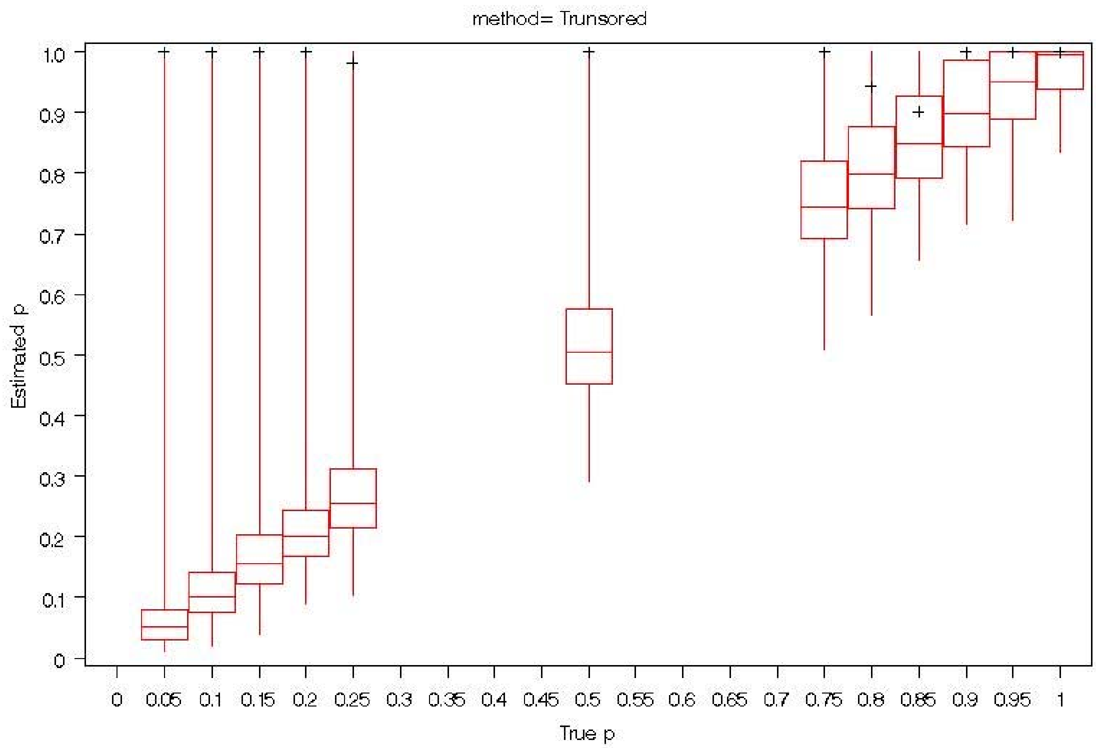


Figure 14. Performance of Trunsored relative to mixture ratio p for mixed Weibull. The mean estimates are represented by "+". The boxes represent the mean and interquartile range. The whiskers represent minimum and maximum.

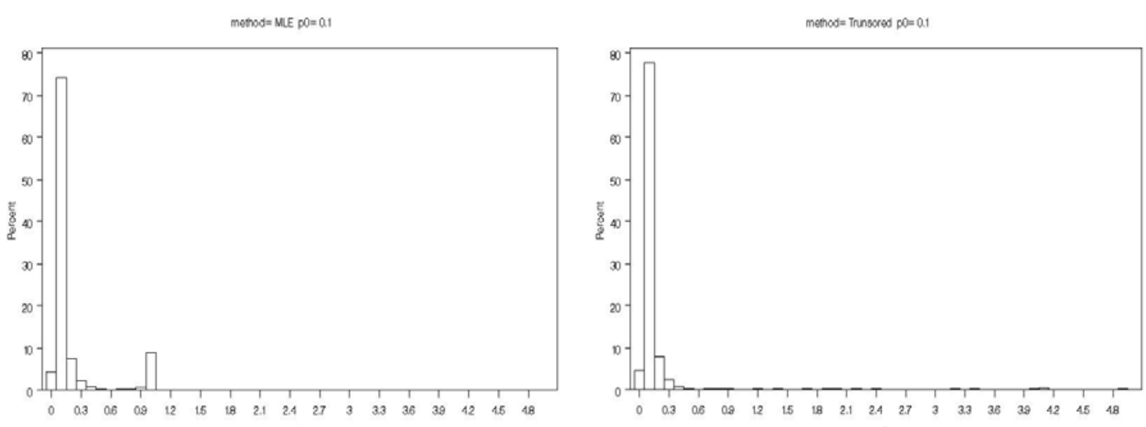


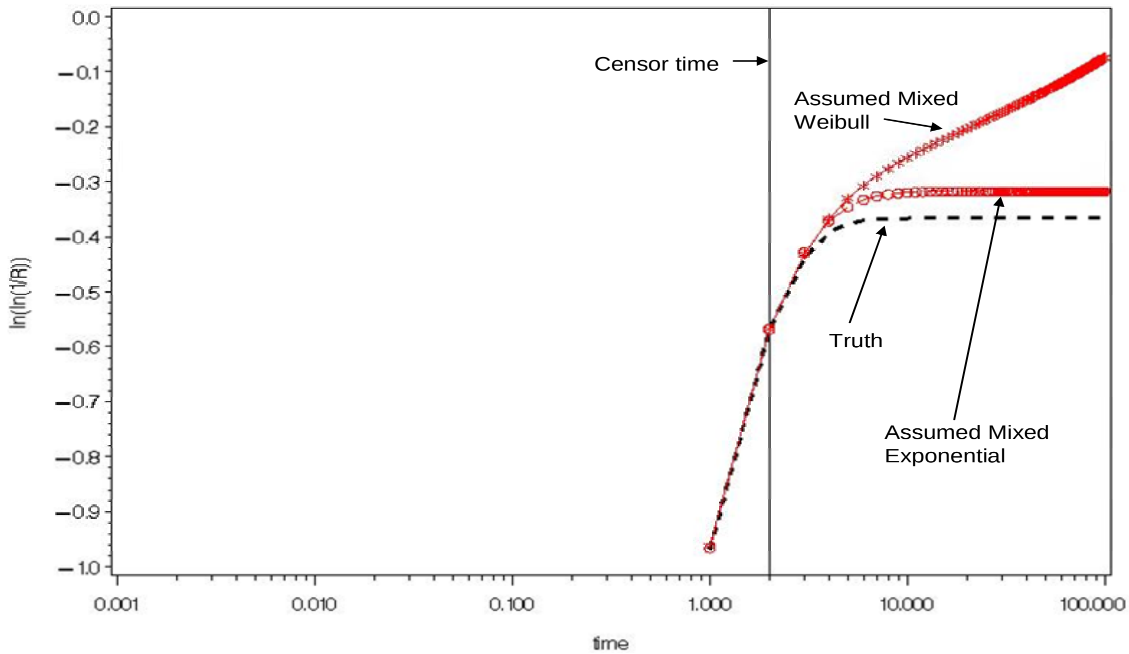
Figure 15. Distribution of estimates for *p* for the MLE and Trunsored methods.

The biases, particularly in the mixture ratio, are worst when the true mixture ratio is small, as illustrated in Figure 14. This is due to the large skew in the distribution of estimated mixture ratios for these cases (see Figure 15). When the true mixture ratio is small, fewer failures are observed. The overall likelihood is fairly flat. Whereas the MLE bounds *p*, the Trunsored estimates are unconstrained. Thus, when few failures are observed, there is a higher chance that an "extreme" set of parameters, with a lot of bias, will be selected as the optimal estimates. For example, consider a case where four failures are observed at {0.14, 0.65, 1.41, 1.97} and the other 96 pieces are censored (at t=2). With the constraint that p<1, the MLE method estimates the parameters to be {p=1; θ=55.44; β=0.96} and the log-likelihood is -19.5677. The Trunsored method, without the constraint on p, selects the parameter estimates {p=27.34, θ=1847.51; β=0.96}. The log-likelihood is only slightly improved to -19.5629. At the censor time, t=2, the two sets give almost identical reliabilities, 0.959 for MLE and 0.961 for Trunsored. The start to diverge at t=10 with 0.824 for MLE and 0.818 for Trunsored. By t=50, they are quite different with 0.404 for MLE and 0.159 for Trunsored. The lack of constraints on the Trunsored method, combined with relatively flat likelihoods, means that it will occasionally select parameters with extremely large errors. The flatness of the likelihood is related to the number of observed failures. Thus if the number of parts tested is large, even if the true mixture ratio is small, more failures will be observed and the chances of these occasional bad cases for the Trunsored methods will go down. For example, when the sample size is 100, the average estimate of p, when $p_0 = 0.1$, is 2.19. However, when the sample size is increased to 500, creating 5 times as many observed failures, the average estimate falls to only 0.18. The median estimate only changes by 0.02, so the increased sample size is only reducing the tail of the distribution. Similarly, the average estimated characteristic life falls from 94.5 to 13.7 while the median only changes by 0.1. The unconstrained nature of the Trunsored distribution makes is susceptible to highly misleading estimates when few failures are observed and the likelihood is relatively flat.

In general, when the reliability estimates were extrapolated far from the censor time, the error in the reliability became increasingly worse, as illustrated in Figure 16. This is because at higher time points, the value of the mixture ratio p becomes increasingly important. With three parameters in the Weibull model and no constraints on p for the Trunsored method, there was a much greater chance of the best fit will involve a p greater than 1. Close to the censor time, the Trunsored method for Weibull distributions provided a good fit, but extrapolated, and the reliability estimate was likely to be biased very low.

Beyond the specific bias and random errors of the parameter estimates under the Trunsored method, there are also theoretical problems when the Trunsored model estimates $p>1$. Although calculating the reliability for the mixed population using these estimated parameters produces good reliability estimates when t is close to the censor time, it has problems when extrapolated far from the censor time. In fact, if the estimates mixture ratio is greater than 1, then there will be a time at which the estimated reliability will be negative. This happens for the mixed exponential case for the Trunsored method as well, but because the shape parameter is de facto constrained for the mixed exponential case, there are fewer times when extremely large mixture ratios are estimated. Thus, the problem is exacerbated for the mixed Weibull.

Another question was whether any of the mixture estimates (Trunsored or MLE) were an improvement over simply applying a single Weibull model. The single Weibull model has the same number of parameters as the mixed exponential, so perhaps it was possible that it would provide an adequate estimate. However, as shown in Figure 17, all of the mixture estimates provide more accurate extrapolations of reliability than does a single Weibull model.



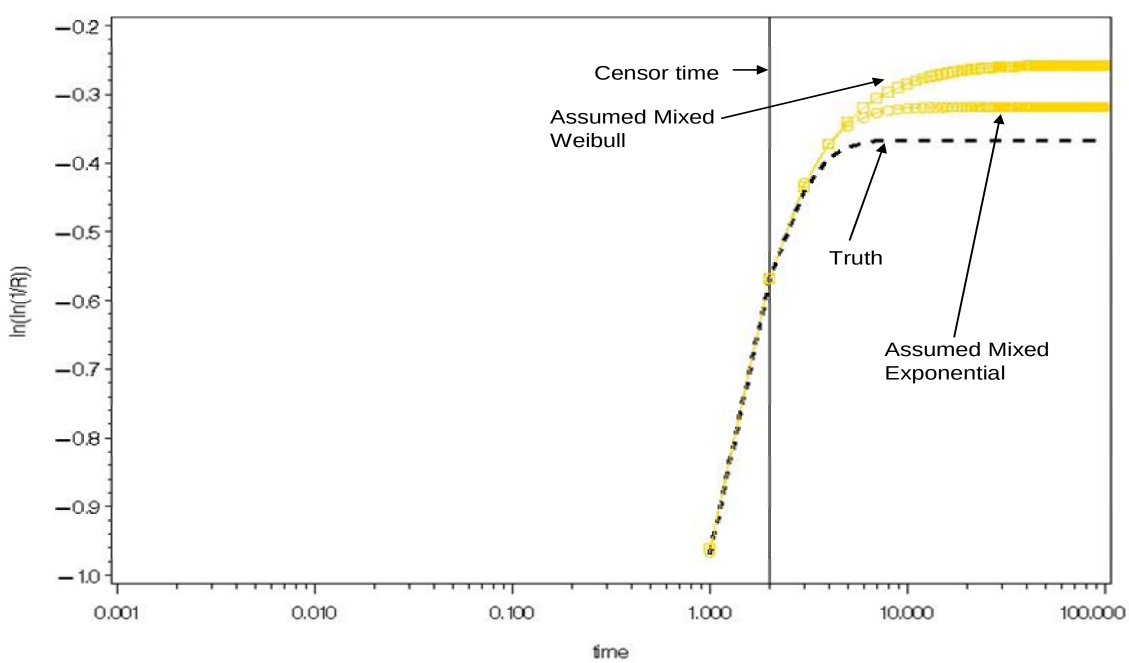


Figure 16. Comparison of bias in reliability estimates for the under assumed mixed exponential model (circles) and assumed mixed Weibull model (stars). True conditions were a mixed exponential distribution. The censor time was 2 and mixture ratio 50%. The results for the Trunsored method is on the top and the MLE method on the bottom.

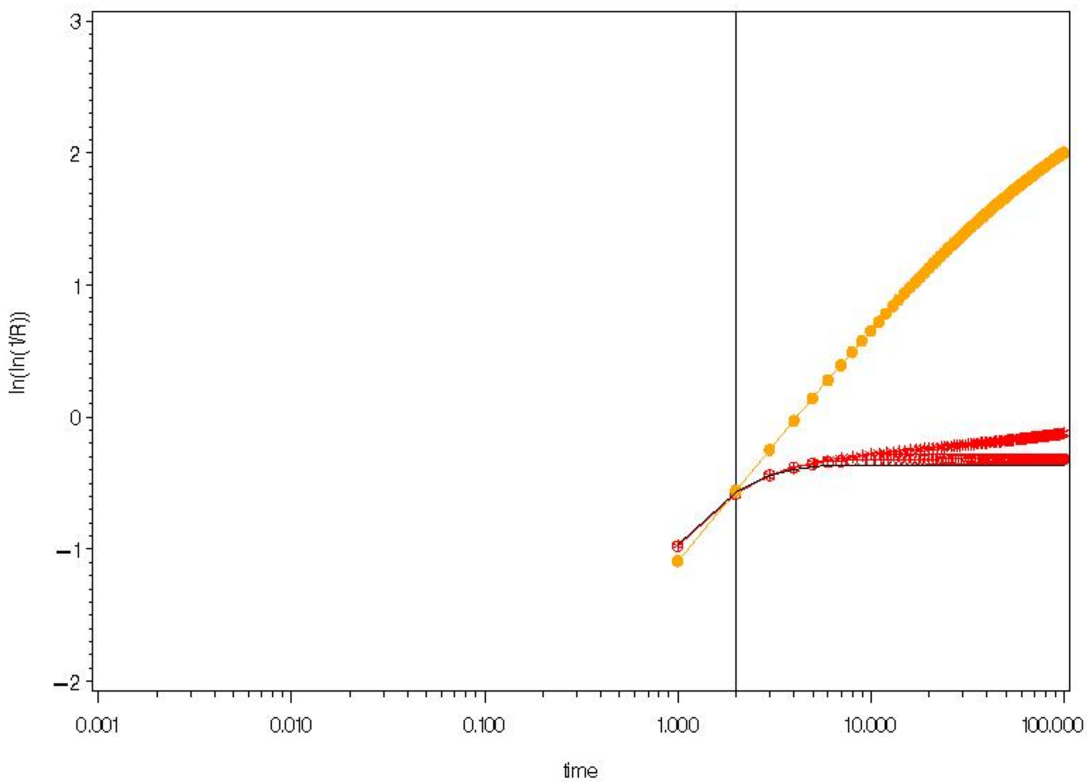


Figure 17. Comparison of Trunsored results (mixed Weibull and mixed exponential, from Figure 16) to a single Weibull. Single Weibull is shown by solid dots. True conditions were a mixed exponential distribution. The censor time was 2 and mixture ratio 50%.

It is possible that some of the error observed is because the mixture ratio is not significantly different from zero. The Trunsored method was primarily proposed to improve statistical characteristics for hypothesis testing and confidence intervals, not for an inherent improvement in point estimate errors. There are problems with performing a simple likelihood ratio test on MLE situations, as described in [Hirose, 2005] and [Maller, 1996], because the chi-squared distribution is a poor approximation of the true distribution. [Hirose, 2005] developed the Trunsored method to be more compatible with the use of the likelihood ratio test for confidence bounds and hypothesis testing. He shows that the percentiles from the log-likelihood ratio from the Trunsored method are closer to the chi-square distribution than are those from the MLE method. Because the use of the likelihood ratio test theoretically should be more appropriate for the Trunsored method than for the MLE method, it was expected that the when the test (of whether the mixture ratio *p* was significantly different from 0) was applied to both methods, it should result in the Trunsored method errors being smaller than those from the MLE method.

However, in almost all cases, the Trunsored method with LRT had more bias and random error than either the Trunsored method prior to the test or the MLE method, as illustrated in Figure 18. For both the MLE and Trunsored methods, the addition of the LRT increased bias and error under most circumstances. The test indicated the mixture ratio was insignificant a very high percentage of the time and predicted substantially lower reliabilities than truth. These errors occurred with both the MLE and Trunsored estimation methods, although the biases for the Trunsored methods were a little higher than those from the MLE method. A summary of the errors under the LRT is provided in Table 9.

The only test that improved accuracy is the MLE method when p_0 is close to one. This was unexpected since it is close to a boundary condition and the LRT is not supposed to be accurate for the MLE method under these circumstances. However, the error in estimates under these circumstances is greatly reduced. The reduction in bias when the true mixture ratio is 1 is illustrated in Figure 18. Based on these results, it appears that the LRT should only be used if increased conservatism is needed or perhaps when the original estimate of p is already very close to 1.

Based on the results of these simulations, it is believed that the power of the likelihood ratio test for detecting a true mixture is low. In the cases examined, the majority of which had a mixture, the test was likely to revert to the null hypothesis, that there was no mixture. In cases where there truly was no mixture, performing the test reduced the errors. In the field, the applicability of the test depends on the prior knowledge of the existence of a mixture and how much conservatism is desired. If the user in the field believes it likely that there is no mixture then performing the test is a good idea. If there is a desire to err on the side of conservatism, that the risk of over predicting reliability is too costly, then performing the test is a good idea. However, if physical evidence and knowledge of the root cause strongly suggests that a mixture exists, performing the test may introduce more conservatism on the reliability estimate than is necessary.

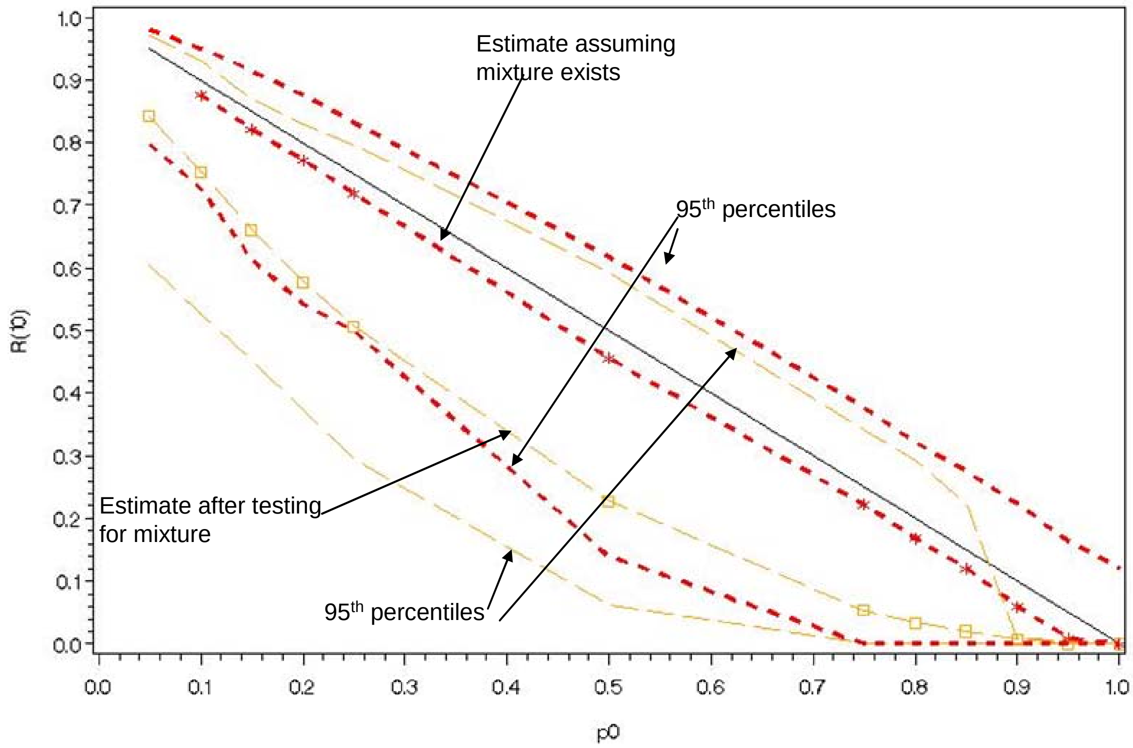


Figure 18. Impact of Testing on Trunsored Method. Red stars is without LRT on *p*; orange squares is with LRT on *p*. Dash line is the average estimate of the reliability at time=10. The light dotted lines represent the 5th and 95th percentiles on the estimate.

Since the model assumes a non-constant failure rate and to examine the impact of an increasing failure rate for the weak subpopulation, cases with β=3 were also examined. If β=3 and θ=1, then the reliability of the weak subpopulation at the censor time of 2 would be very close to 0. Since this would result in more observed failures, and likely less error in the estimates, the characteristic life was also adjusted. At β=3 and θ=1.6, the reliability of the weak subpopulation at the censor time of 2 would be 14%, the same as in the case of β=1 and θ=1. An increasing failure rate produces larger bias and random errors for the reliability estimate for both methods.

Once again, the impact on the Trunsored method is more dramatic than for the MLE method. The results for a few conditions are provided in **Appendix 2: Estimation Errors**. The errors for the characteristic life are actually smaller for the increasing failure rate case, but the errors in the shape parameter have increased substantially. This leads to the overall increase in error for the reliability estimate.

Chapter Conclusions

The Trunsored method performed better under mixed Exponential assumptions than under mixed Weibull assumptions. Under mixed Weibull, it occasionally selects parameter values which produce nonsensical reliability values, even relatively close to the censor time. This typically happens when few failures are observed and the likelihood is relatively flat. The lack of constraints on the parameter values under the Trunsored method means that extreme parameter values may be selected when only minimal improvement is made to the likelihood. These results happen less frequently if the population is larger and thus more failures are observed. This situation is not improved by testing for the presence of the weak subpopulation nor if there is an increasing failure rate.

Chapter 6: Bayesian Estimation

This chapter develops the methodology for performing Bayesian estimation in mixed Weibull situations and summarizes the results of the Monte Carlo simulations using the Bayesian technique. These results are compared to those from MLE.

Methodology

It is assumed that a product comes from either a weak or strong subpopulation with a mixture ratio *p*.[9]

$$R(t) = p \times \exp\left[-\left(\frac{t}{\theta_W}\right)^{\beta_W}\right] + (1-p) \times \exp\left[-\left(\frac{t}{\theta_S}\right)^{\beta_S}\right]$$

where $0 < p < 1$

and $\theta_W, \theta_S, \beta_W, \beta_S > 0$

As with all Bayesian analyses, assumptions must be made about the prior knowledge of the parameters. It is assumed that the mixture and characteristic life parameters are all independent of each other. It is also assumed here that the shape parameters are independent from the other variables. This is perhaps a more difficult assumption, but is common in the Bayesian reliability literature [Martz, 1982] [Canavos, 1973].

One commonly used set of priors on the Weibull is the inverted gamma distribution on Θ and the uniform distribution on B [Martz, 1982] [Canavos, 1973][10]. This is

---

[9] Note that this is a slightly different parameterization of the Weibull than is typically used: the shape parameter β is only applied to the random variable *t*. This is to allow easier separation of the shape and characteristic life parameters in the integrations needed to calculate the posterior distribution. It is recommended in Martz and Waller [1982].

similar to the assumption made in [Perlstein, 2001], who assumed a beta prior on the mixture parameter *p* and gamma priors on each of the failure rates. Since for the Weibull distribution, the standard parameterization uses characteristic life, the parameter of interest is similar to the inverse of the failure rate parameter in the exponential case. If Λ is a random variable having a Gamma distribution with parameters $\left(a, b^{-1}\right)$, then $\Theta = \frac{1}{\Lambda}$ has an inverted gamma distribution with pdf

$$g(\theta) = \frac{b^a}{\Gamma(a)} \left(\frac{1}{\theta}\right)^{a+1} \exp\left(-\frac{b}{\theta}\right) \quad a, b > 0$$

Thus, for consistency with the previous literature, it will be assumed that the characteristic life parameters Θ_W, Θ_S follow inverted gamma priors, the shape parameter B follows a uniform distribution, and the mixture parameter *p* follows a beta prior.

Using a generalization of the likelihood parameterization proposed in [Perlstein, 2001],

[10] Capital Greek letters are used here for the random variables of the parameters, as is customary for Bayesian analysis.

$$L\left(p,\theta_W,\theta_S,\beta_W,\beta_S\right)=\sum_{\ell=0}^{n}\sum_{m=1}^{\binom{n}{\ell}}\sum_{v=0}^{u}\left\{\binom{u}{v}p^{\ell+v}\left(1-p\right)^{n-\ell+u-v}\right.$$

$$\times\left(\frac{\beta_W}{\theta_W}\right)^{\ell}\prod\left(t_m^{(\ell)}\right)^{\beta_W-1}\exp\left(-\frac{\sum\left(t_m^{(\ell)}\right)^{\beta_W}+vt_*^{\beta_W}}{\theta_W}\right)$$

$$\left.\times\left(\frac{\beta_S}{\theta_S}\right)^{n-\ell}\prod\left(t_m^{(n-\ell)}\right)^{\beta_S-1}\exp\left(-\frac{\sum\left(t_m^{(n-\ell)}\right)^{\beta_S}+\left(u-v\right)t_*^{\beta_S}}{\theta_S}\right)\right\}$$

where $\sum \mathrm{t}_{\mathrm{m}}^{\ell}=$ the sum of the elements of the m[th] permutation of vectors of size ℓ from $\underline{t}=\{t_1,...,t_n\}$.

The joint posterior distribution has the form:[11]

$$g\left(p,\beta_W,\beta_S,\theta_W,\theta_S\mid D\right)=\frac{g\left(p,\beta_W,\beta_S,\theta_W,\theta_S\right)\times L\left(p,\beta_W,\beta_S,\theta_W,\theta_S\mid D\right)}{C\left(n,u\right)}$$

For simplicity, a constant for this denominator is defined to be:

$$C\left(n,u\right)=\sum_{\ell=0}^{n}\sum_{m=1}^{\binom{n}{\ell}}\sum_{v=0}^{u}\left\{\binom{u}{v}\frac{\Gamma\left(\alpha^*\right)\Gamma\left(\beta^*\right)}{\Gamma\left(\alpha^*+\beta^*\right)}\times\Gamma\left(a_W^*\right)\times\Gamma\left(a_S^*\right)\right.$$

$$\left.\int_{\delta_{W1}}^{\delta_{W2}}\int_{\delta_{S1}}^{\delta_{S2}}\left(b_S^*\right)^{a_S^*}\beta_S^{\,n-\ell}\prod\left(t_m^{(n-\ell)}\right)^{\beta_S-1}\times\left(b_W^*\right)^{a_W^*}\beta_W^{\,\ell}\prod\left(t_m^{(\ell)}\right)^{\beta_W-1}\mathrm{d}\beta_S\,\mathrm{d}\beta_W\right\}$$

where $\left(\delta_{W1},\delta_{W2}\right)$ and $\left(\delta_{S1},\delta_{S2}\right)$ are the limits of the prior uniform distribution assumed for β_W and β_S respectively.

[11] D represents the observed Data.

Using the joint posterior distribution, the posterior expected values of the parameters and the reliability can be calculated.

$$
\begin{aligned}
E(p) = \frac{1}{C(n,u)} \times \\
\sum_{\ell=0}^{n} \sum_{m=1}^{\binom{n}{\ell}} \sum_{v=0}^{u} \Bigg\{ \binom{u}{v} \frac{\Gamma(\alpha^{*}+1)\Gamma(\beta^{*})}{\Gamma(\alpha^{*}+1+\beta^{*})} \times \Gamma(a_W^{*}) \times \Gamma(a_S^{*}) \\
\int_{\delta_{W1}}^{\delta_{W2}} \int_{\delta_{S1}}^{\delta_{S2}} \left(b_S^{*}\right)^{a_S^{*}} \beta_S{}^{n-\ell} \prod \left(t_m^{(n-\ell)}\right)^{\beta_S-1} \times \left(b_W^{*}\right) \beta_W{}^{\ell} \prod \left(t_m^{(\ell)}\right)^{\beta_W-1} \mathrm{d}\beta_S \, \mathrm{d}\beta_W \Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
E(\theta_W) = \frac{1}{C(n,u)} \times \\
\sum_{\ell=0}^{n} \sum_{m=1}^{\binom{n}{\ell}} \sum_{v=0}^{u} \Bigg\{ \binom{u}{v} \frac{\Gamma(\alpha^{*})\Gamma(\beta^{*})}{\Gamma(\alpha^{*}+\beta^{*})} \times \Gamma(a_W^{*}-1) \times \Gamma(a_S^{*}) \\
\int_{\delta_{W1}}^{\delta_{W2}} \int_{\delta_{S1}}^{\delta_{S2}} \left(b_S^{*}\right)^{a_S^{*}} \beta_S{}^{n-\ell} \prod \left(t_m^{(n-\ell)}\right)^{\beta_S-1} \times \left(b_W^{*}\right)^{a_W^{*}-1} \beta_W{}^{\ell} \prod \left(t_m^{(\ell)}\right)^{\beta_W-1} \mathrm{d}\beta_S \, \mathrm{d}\beta_W \Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
E(\theta_S) = \frac{1}{C(n,u)} \times \\
\sum_{\ell=0}^{n} \sum_{m=1}^{\binom{n}{\ell}} \sum_{v=0}^{u} \Bigg\{ \binom{u}{v} \frac{\Gamma(\alpha^{*})\Gamma(\beta^{*})}{\Gamma(\alpha^{*}+\beta^{*})} \times \Gamma(a_W^{*}) \times \Gamma(a_S^{*}-1) \\
\int_{\delta_{W1}}^{\delta_{W2}} \int_{\delta_{S1}}^{\delta_{S2}} \left(b_S^{*}\right)^{a_S^{*}-1} \beta_S{}^{n-\ell} \prod \left(t_m^{(n-\ell)}\right)^{\beta_S-1} \times \left(b_W^{*}\right)^{a_W^{*}} \beta_W{}^{\ell} \prod \left(t_m^{(\ell)}\right)^{\beta_W-1} \mathrm{d}\beta_S \, \mathrm{d}\beta_W \Bigg\}
\end{aligned}
$$

$$E\left(\beta_W\right)=\frac{1}{C\left(n,u\right)}\times$$

$$\sum_{\ell=0}^{n}\sum_{m=1}^{\binom{n}{\ell}}\sum_{v=0}^{u}\left\{\binom{u}{v}\frac{\Gamma\left(\alpha^*\right)\Gamma\left(\beta^*\right)}{\Gamma\left(\alpha^*+\beta^*\right)}\times\Gamma\left(a_W^*\right)\times\Gamma\left(a_S^*\right)\right.$$

$$\left.\int_{\delta_{W1}}^{\delta_{W2}}\int_{\delta_{S1}}^{\delta_{S2}}\left(b_S^*\right)^{a_S^*}\beta_S{}^{n-\ell}\prod\left(t_m^{(n-\ell)}\right)^{\beta_S-1}\times\left(b_W^*\right)^{a_W^*}\beta_W{}^{\ell+1}\prod\left(t_m^{(\ell)}\right)^{\beta_W-1}\mathrm{d}\beta_S\,\mathrm{d}\beta_W\right\}$$

$$E\left(\beta_S\right)=\frac{1}{C\left(n,u\right)}\times$$

$$\sum_{\ell=0}^{n}\sum_{m=1}^{\binom{n}{\ell}}\sum_{v=0}^{u}\left\{\binom{u}{v}\frac{\Gamma\left(\alpha^*\right)\Gamma\left(\beta^*\right)}{\Gamma\left(\alpha^*+\beta^*\right)}\times\Gamma\left(a_W^*\right)\times\Gamma\left(a_S^*\right)\right.$$

$$\left.\int_{\delta_{W1}}^{\delta_{W2}}\int_{\delta_{S1}}^{\delta_{S2}}\left(b_S^*\right)^{a_S^*}\beta_S{}^{n-\ell+1}\prod\left(t_m^{(n-\ell)}\right)^{\beta_S-1}\times\left(b_W^*\right)^{a_W^*}\beta_W{}^{\ell}\prod\left(t_m^{(\ell)}\right)^{\beta_W-1}\mathrm{d}\beta_S\,\mathrm{d}\beta_W\right\}$$

The remaining integrals with respect to β_S and β_W do not have closed form solutions and must be numerically integrated. However, limiting the numeric integration to these two integrals, as opposed to numerically integrating over all five parameters, greatly simplifies the problem and decreases the computational time.

Results

In the baseline case, *p* was taken to have prior distribution Beta(0.5, 0.5) with likely values either close to 0 or 1, with less likelihood of middle values; θ was taken to have the prior Γ(4, 3) an expected value of 1; β was given a uniform distribution prior between 0 and 6. These prior distributions are illustrated in Figure 19.

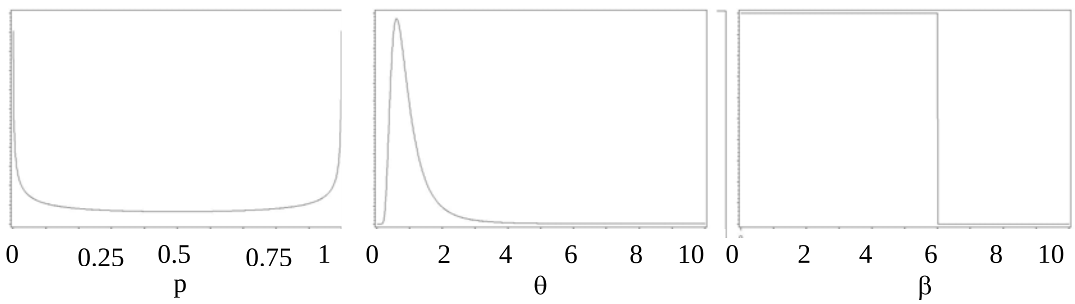


Figure 19. Prior distributions on *p* (left), θ (middle) and β (right) for baseline case.

As illustrated in Figure 20 and Figure 21, the Bayesian method performed very well. It had less bias and random error in its estimates of the mixture ratio than the MLE method did in almost all cases, particularly for small values of *p*. For example, for *p*=0.05, the bias and random error in a were reduced by an order of magnitude; this translated to a reduction in the reliability estimate at time 10 from -0.02 to 0 in bias and 0.07 to 0.03 in random error. However, given that the prior gave high weight to very small and very large values of *p*, the strong performance of the Bayesian method under these conditions is not unexpected. The performance of the Bayesian method at *p*=0.5, with a lower likelihood under the prior, is a harder test. Under this condition, the Bayesian method performed comparably to the MLE method. The bias in its estimate of *p* was a little higher (0.06 versus 0.04 for mixture) but its random error was smaller (0.09 versus 0.15). The difference in the estimate of the reliability at time 10 was also very small, (-0.05 versus -0.03 bias and 0.09 versus 0.12 random error). These results suggest that the Bayesian method would be a better choice than the MLE method, since when the prior is close to correct the Bayesian method produces more accurate estimates of the mixture ratio and even when the prior gives low likelihood the Bayesian method performs about as well.

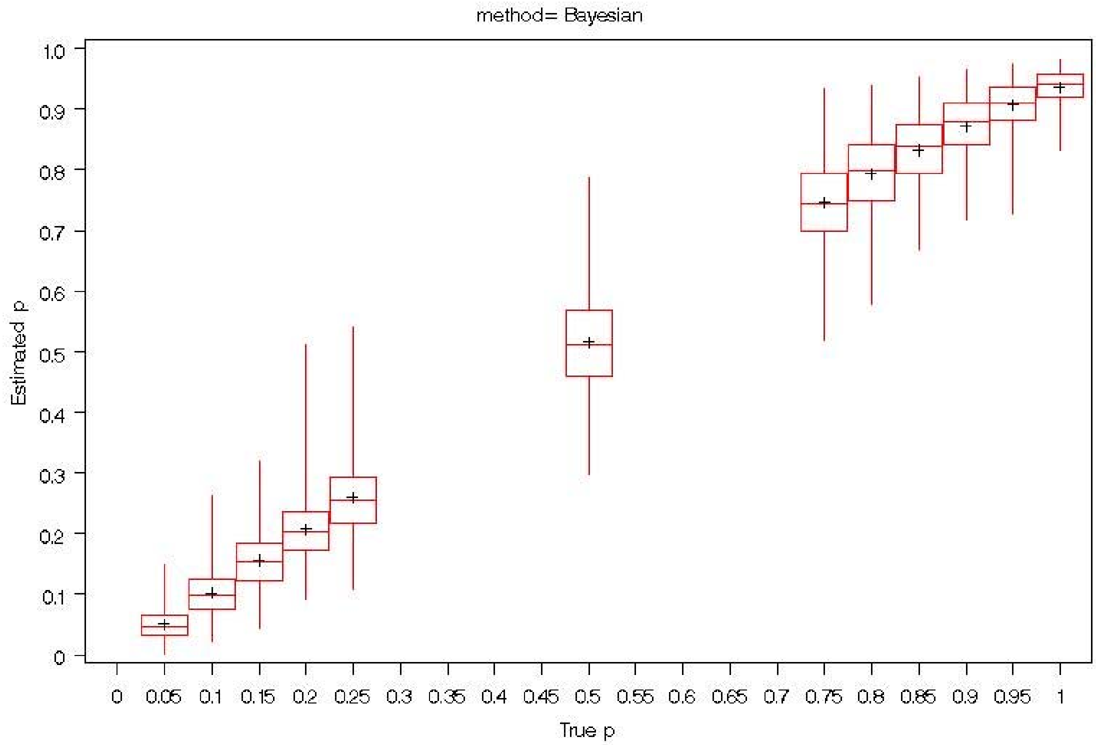


Figure 20. Performance of Bayesian method relative to mixture ratio *p*. [Baseline case: p~Beta(0.5, 0.5); θ~Γ(4, 3); β~U(0, 6)]

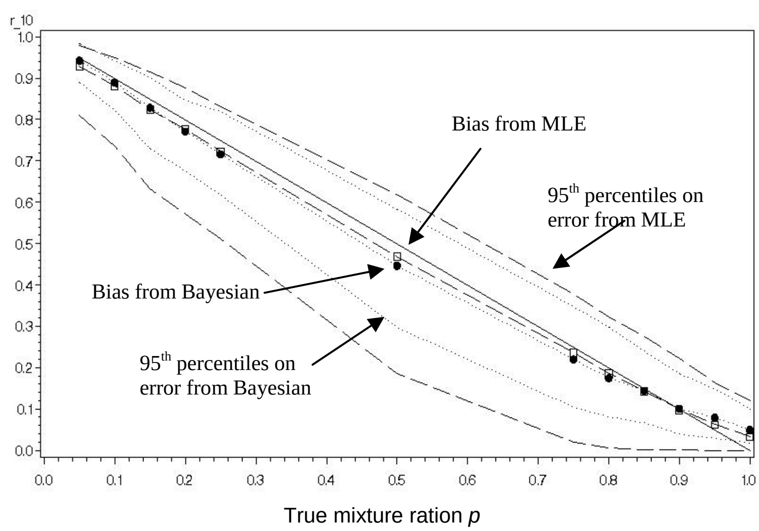


Figure 21. Comparison of Error in R(10), 5 times the censor time, for MLE and Bayesian estimates [Baseline case: p~Beta(0.5, 0.5); θ~Γ(4, 3); β~U(0, 6)]

Next, the priors were left as in the baseline case, but an increasing failure rate was considered. In this case, the shape parameter γ was increased to 3. The characteristic life θ was also increased, to 1.6, in order to maintain similar numbers of observed failures at the censor time. The results for the reliability estimate at t=10 are shown in Figure 22. The method seemed very insensitive to the shape of the Weibull, with errors being virtually identical for the constant failure rate and increasing failure rate cases. This result was expected since a uniform prior was placed on the shape parameter.

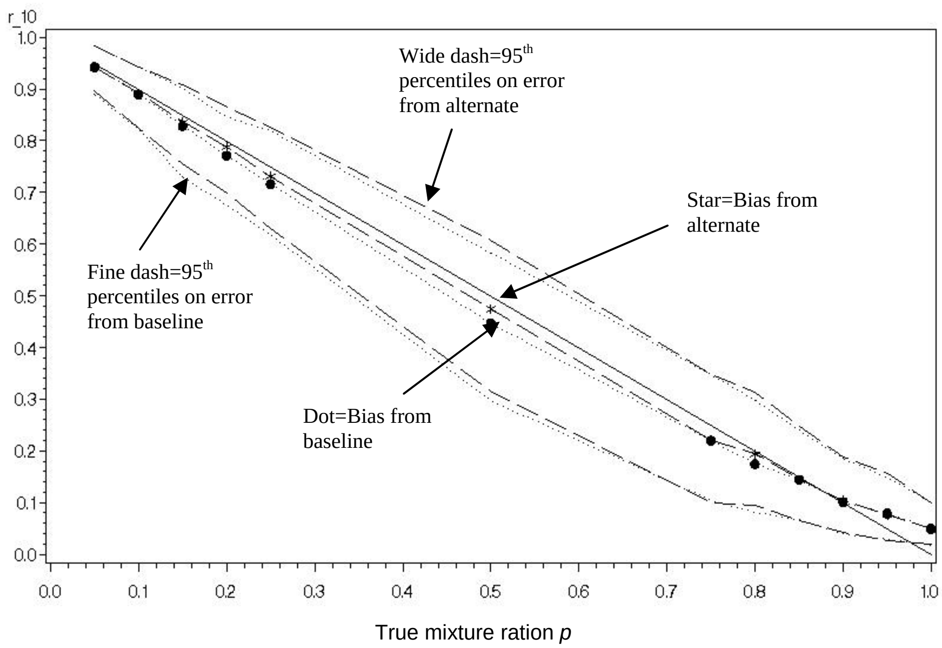


Figure 22. Comparison of Error in R(10) between the baseline case where true characteristic life θ=1 and true shape parameter β=1 and the alternate case where true θ=1.6; true β=3. [Baseline prior case: p~Beta(0.5, 0.5); $\theta \sim \Gamma(4, 3)$; $\beta \sim U(0, 6)$]

Of greater concern was whether the Bayesian method would be robust under inaccurate priors on the mixture ratio p and the characteristic life θ. First a prior distribution that gave higher weight to low values of the mixture ratio was selected (p~Beta(0.5, 2)). As expected, the performance of the Bayesian method improved for low and even medium values of *p*, as illustrated in Figure 23. The performance only got worse at *p*>0.8. At *p*=1, the p~Beta(0.5, 2) case had bias of 0.1 and random error of 0.03 for R(10) as compared to 0.05 and 0.03 for the baseline p~Beta(0.5, 0.5) case. Similarly, as shown in Figure 24, when a prior was selected to give higher likelihood to middle values (p~Beta(2, 2)), bias in the estimates of *p* and the reliability at 10 got 2 to 3 times worse under this prior for true mixture ratios close to 0 or close to 1.

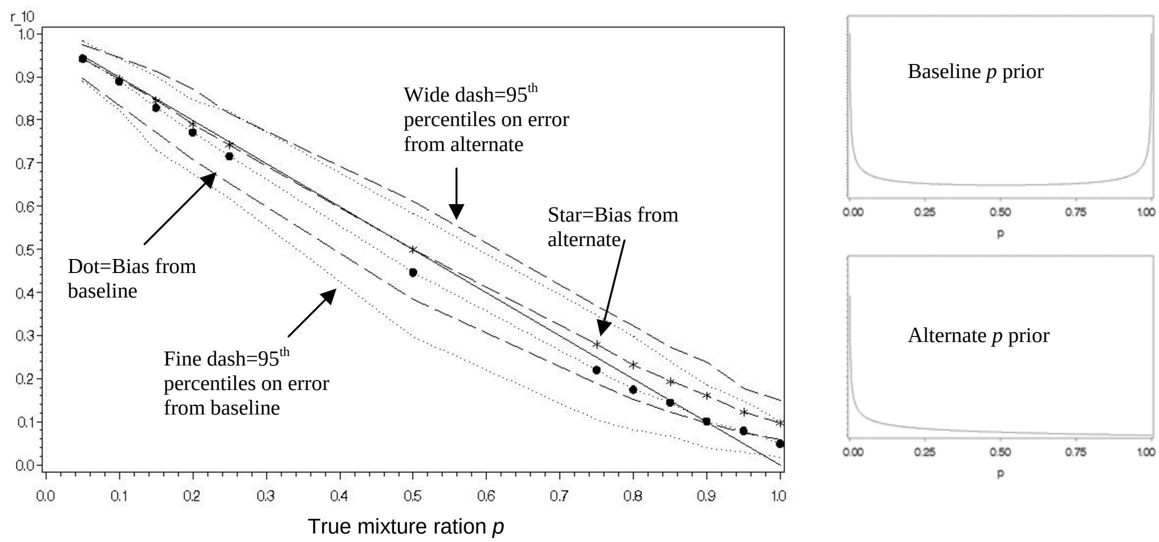


Figure 23. Comparison of Errors in R(10) between baseline Bayesian case (*p*~Beta(0.5, 0.5), θ~Γ(4, 3), and β~U(0, 6)) and alternate case where *p* has high likelihood of being small (*p*~Beta(0.5, 2), θ~Γ (4, 3), β~U(0,6))

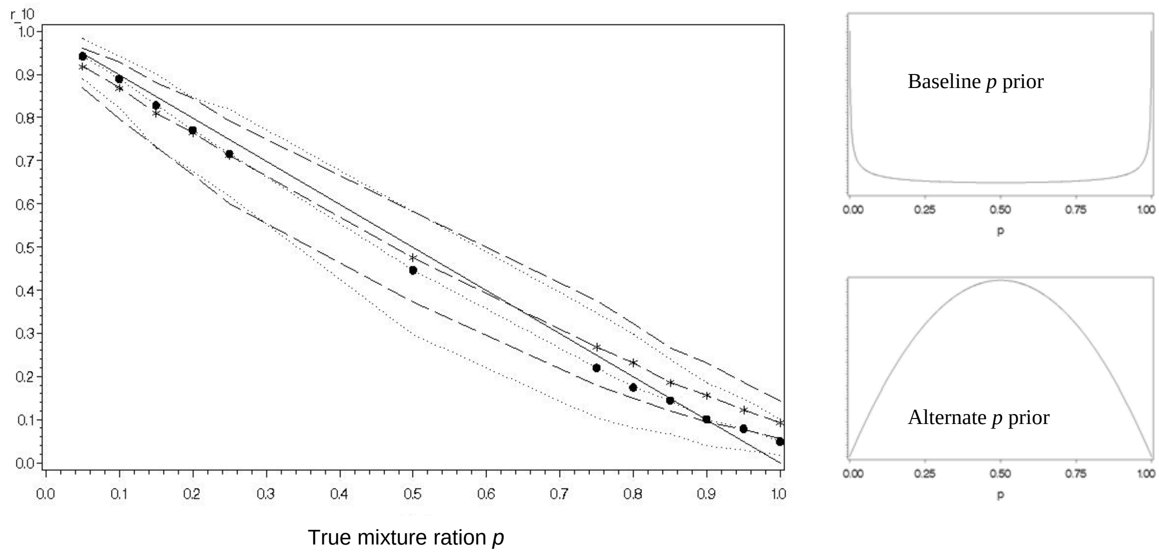


Figure 24. Comparison of Errors in R(10) between baseline Bayesian case (*p*~Beta(0.5, 0.5), θ~Γ(4, 3), and β~U(0, 6)) and alternate case where *p* has high likelihood of being a middle value (*p*~Beta(2, 2), θ~Γ (4, 3), β~U(0,6))

Even though the Bayesian method is sensitive to the choice of prior on *p*, it is still appears to be more robust than the MLE method. None of these increases in error from choices of prior on *p* are comparable to the order of magnitude reduction in the errors in *p* between the MLE and any of the Bayesian cases for small mixture ratios.

The robustness of the Bayesian method with regards to the prior on the characteristic life θ was also examined. In the first case, shown in Figure 25, the prior was adjusted to give an expected value of 0.5 (instead of 1). In this case the probability of θ being greater than or equal to 1 (the true value) was about 0.09. This case produced almost no change in the errors from the baseline case where the expected value of θ was equal to 1. In the second case, shown in Figure 26, the prior was adjusted to give an expected value of 4.3, with the chance of θ being less than or equal to 1 being about 0.09 for symmetry. In this case, the Bayesian estimate produced much worst estimates of the reliability than did the baseline case or the MLE method. Reducing the expected value to 2 did not substantially correct the problem, as illustrated in Figure 27, nor did more than doubling the chance of θ being less than or equal to 1, as illustrated in Figure 28.

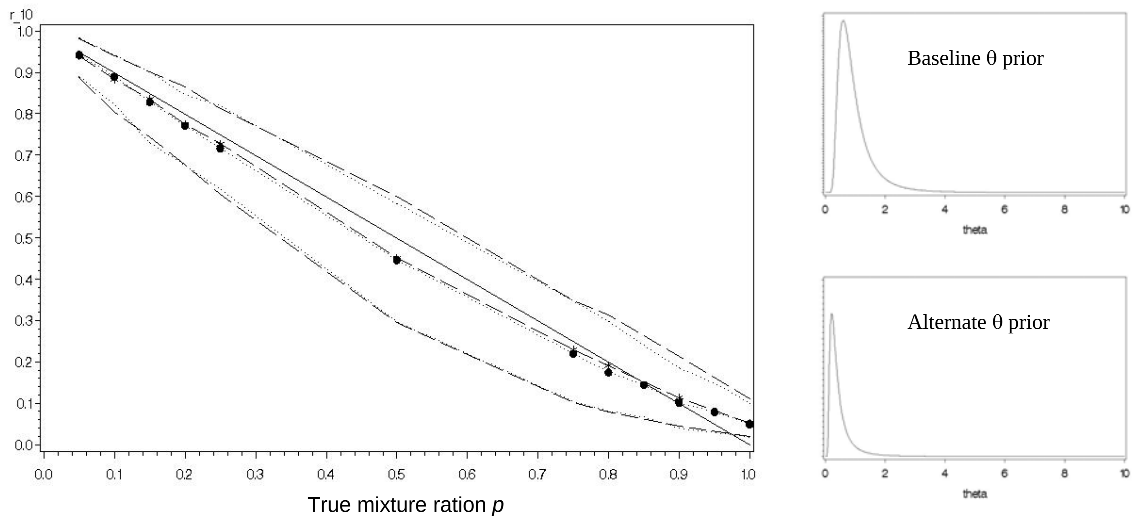


Figure 25. Comparison of Errors in R(10) between baseline Bayesian case (*p*~Beta(0.5, 0.5), θ~Γ(4, 3), and β~U(0, 6)) and alternate case where θ is likely to be small (*p*~Beta(0.5, 0.5), θ~Γ (2.5, 0.75), β~U(0,6)). For the alternate case, the expected value of θ is 0.5 and the chance of θ being greater or equal to 1 is about 0.09. The dots represent the baseline case and the stars the alternate case. There is almost no difference in the results.

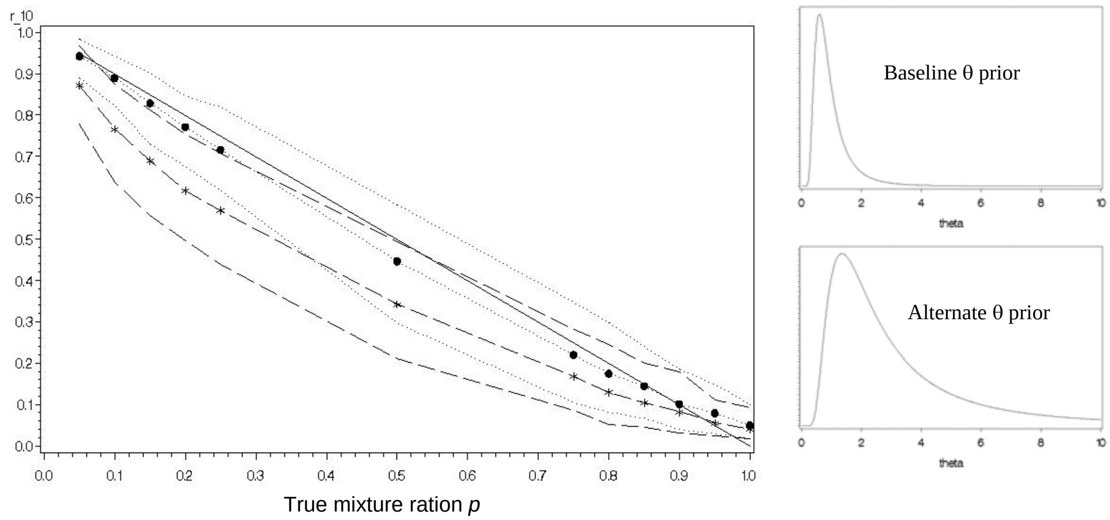


Figure 26. Comparison of Errors in R(10) between baseline Bayesian case (*p*~Beta(0.5, 0.5), θ~Γ(4, 3), and β~U(0, 6)) and alternate case where θ is likely to be large (*p*~Beta(0.5, 0.5), θ~Γ (1.915, 3.935), β~U(0,6)). For the alternate case, the expected value of θ is 4.3 and the chance of θ being greater or equal to 1 is about 0.09. The dots represent the baseline case and the stars the alternate case. The alternate case performs much worse unless p_0 is very high.

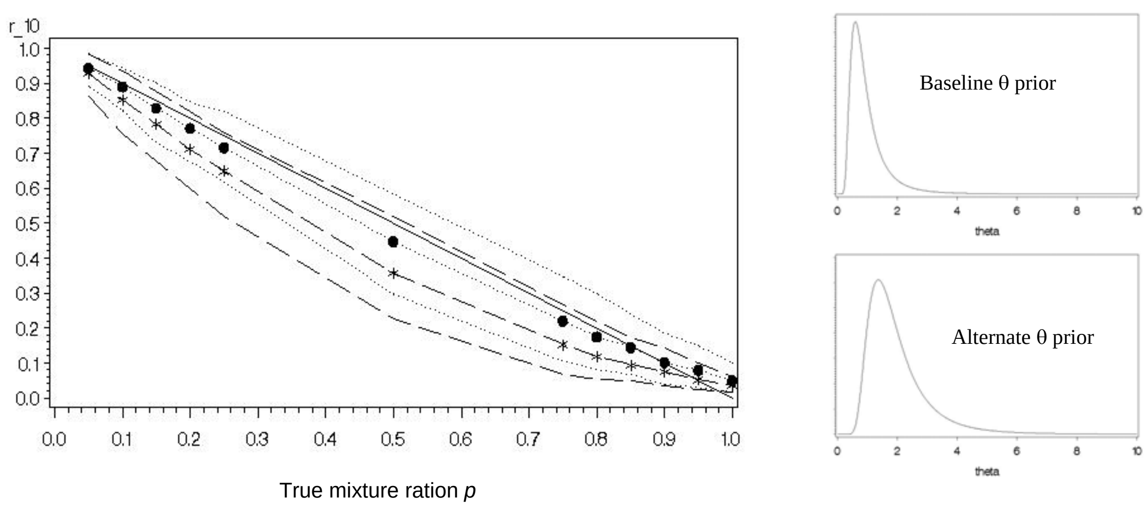


Figure 27. Comparison of Errors in R(10) between baseline Bayesian case (*p*~Beta(0.5, 0.5), θ~Γ(4, 3), and β~U(0, 6)) and alternate case where θ is likely to be large (*p*~Beta(0.5, 0.5), θ~Γ (5.35, 8.68), β~U(0,6)). For the alternate case, the expected value of θ is 2 and the chance of θ being greater or equal to 1 is about 0.09. The dots represent the baseline case and the stars the alternate case. The alternate case performs much worse unless p_0 is very high.

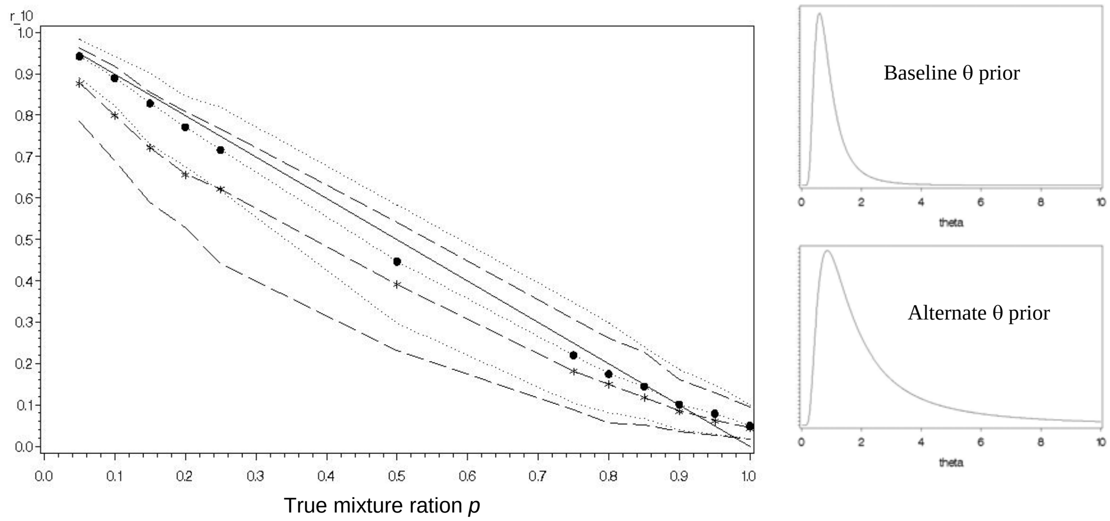


Figure 28. Comparison of Errors in R(10) between baseline Bayesian case (*p*~Beta(0.5, 0.5), θ~Γ(4, 3), and β~U(0, 6)) and alternate case where θ is likely to be large (*p*~Beta(0.5, 0.5), θ~Γ (1.5, 2.15), β~U(0,6)). For the alternate case, the expected value of θ is 4.3 and the chance of θ being greater or equal to 1 is about 0.2. The dots represent the baseline case and the stars the alternate case. Even with a much higher probability of small θ, the alternate case performs much worse unless p_0 is very high.

These for the variants in the prior on θ suggest that the Bayesian model is more sensitive to over estimating θ with the prior than to under estimating θ with the prior. This lack of symmetry seems at first counter-intuitive. The cause is related to the correlation between estimates of the mixture ratio *p* and estimates of the characteristic life θ. Estimates of *p* and θ tend to be correlated with higher estimates of θ being associated with higher estimates of *p*. The likelihood function for the mixed Weibull cases considered in this paper tend to be relatively flat, particularly with low true mixture ratios because there is only limited observed failures. These flat likelihoods are heavily influenced by the prior distribution, as shown in Figure 29. If the prior distribution on θ has a larger weight on high values, then it is likely that a high estimate of *p* will also be selected. This result is shown by the smaller error in the reliability estimates when the true mixture ratio is close to 1 and the higher θ prior is selected in Figure 26 and Figure 27. However, when the true mixture ratio and characteristic life are small, the higher priors dominate the flat likelihood and create more error in the posterior.

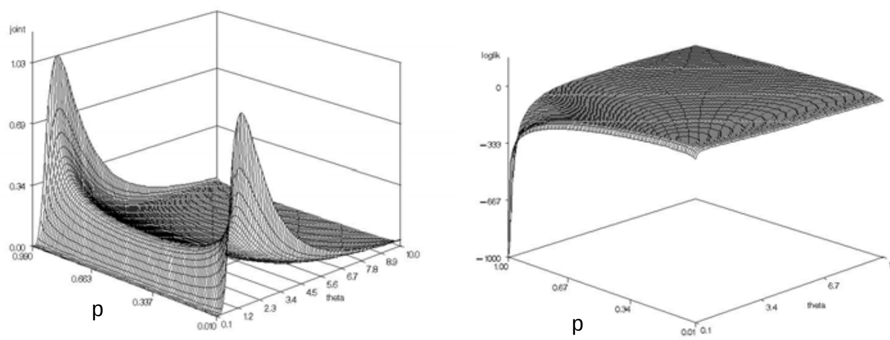


Figure 29. Left is the joint prior for *p* and θ. Right is log=likelihood (β held at 1) for a case where true *p* is 0.1 and 12 failures were observed. The 25th percentile for the log-likelihood was -56 and the 75th percentile was -45. The flatness of the likelihood is dominated by the prior in the posterior. The likelihood function is only able to put low weight on situations with θs and high p_0s.

On the other hand, if the prior on θ has a larger weight on small values, the estimates of *p* will also tend to be small. But if the true mixture ratio *p* is large, more failures will have been observed than suggested by the prior. If more failures are observed, the likelihood function is less flat and the priors have less influence on the estimates from the posterior distribution, as shown in Figure 30.

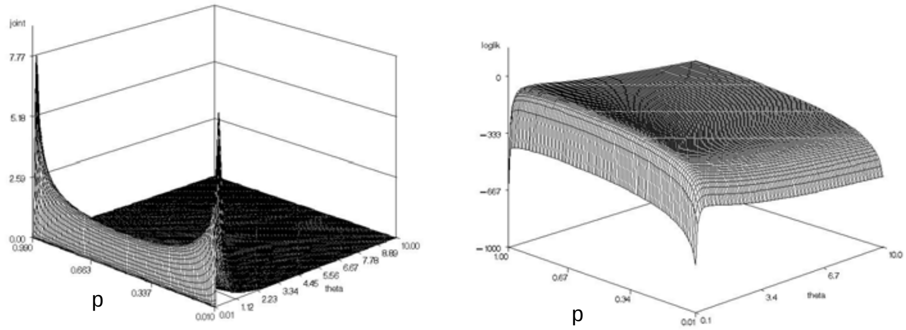


Figure 30. Left is the joint prior for *p* and θ. Right is log=likelihood (β held at 1) for a case where true *p* is 0.9 and 84 failures were observed. The 25th percentile for the log-likelihood was -267 and the 75th percentile was -167. The larger number of failures means that the log-likelihood is less flat and that cases with very small θs have very low likelihoods, countering the inaccurate prior distribution.

This theory was further substantiated by looking at a case where the prior on *p* gave preference to low values and the θ prior continued to over-predict θ. As shown in Figure 31, the errors in reliability were substantially smaller as long as the true *p* was less than about 0.8.

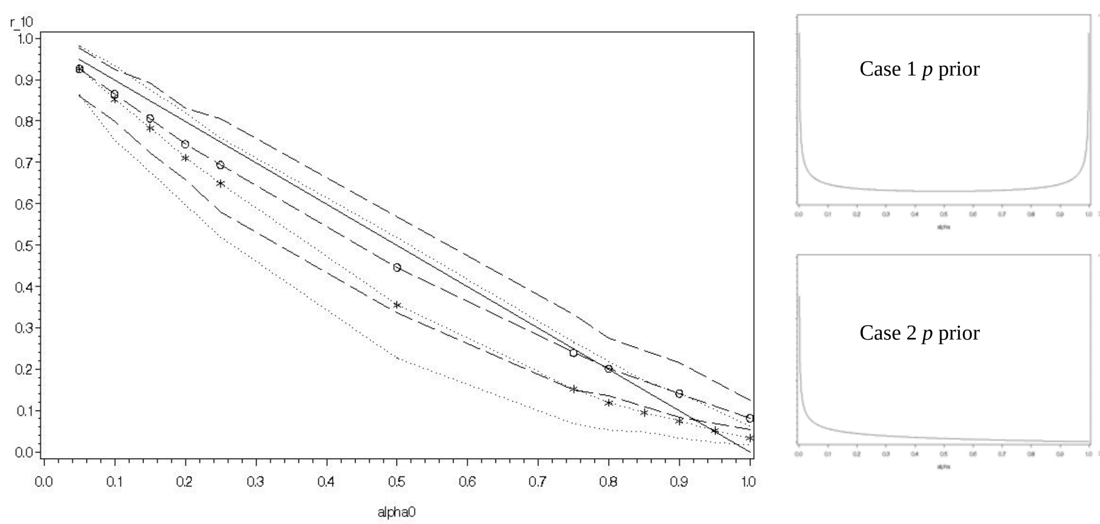


Figure 31. Comparison of Errors in R(10) between Case 1 that allows high mixture ratios (*p*~Beta(0.5, 0.5), θ~Γ (1.915, 3.935), β~U(0,6)) and Case 2 that does not (*p*~Beta(0.5, 2), θ~Γ (1.915, 3.935), β~U(0,6)). Although the prior on θ still allows high values, the change in prior on *p* limits their likelihood on the posterior.

Chapter Conclusions

As expected, the Bayesian method is sensitive to the choice of prior. However, it generally performs well, when compared to MLE, unless poor choices of prior are made on both the mixture ratio p and the characteristic life θ. It is particularly sensitive to situations where the prior distribution is selected to give large probability to high values of the characteristic life and mixture ratio simultaneously. It seems that it is better to select a prior with an expected value less than or equal to the true θ than to possibly overestimate θ.

When a low mixture ratio is suspected, the Bayesian method can greatly reduce the errors in the estimates over the MLE method so long as a good prior is selected for θ. It is suspected this is because at low mixture ratios, particularly few failures will be observed. The MLE then has little data for its parameter estimates, while the Bayesian method gains information from its priors. If the prior distribution does not out-weigh the data into producing results with higher than truth mixture ratios and characteristics lives, the Bayesian method produces good results.

If a poor prior distribution is chosen with the assumption that the mixture ratio is very low (i.e. gives too much probability to mixture ratio values lower than truth), the Bayesian method will perform worse than the MLE method. However, the increase in error under these circumstances is smaller than the decrease in error in circumstances where the mixture ratio really is very small. Based on these results, it is recommended that Bayesian methods be considered when a low mixture ratio is suspected.

Chapter 7: Cost Model

Decisions about performing additional burn-in to find weak parts can be made based on a fixed requirement (such as the overall reliability is required to be 0.99) or may be based on the cost effectiveness of the decision. In order to determine if the burn-in is cost effective, a cost model must be developed. With the cost model, one can find the optimum balance between investing in burn-in and avoiding field failure. In this chapter, previously developed cost models are reviewed, a new model is formulated, and the impact of the new aspects of the model are examined.

Previous Cost Models

Most burn-in cost models in the literature are similar to the one proposed by Perlstein [Perlstein, 2001]. This model is relatively straightforward with a fixed portion, a time dependent portion, and different failure costs corresponding to failures during burn-in and in the field.

$$C_{total} = C_{setup} + C_B \cdot N \cdot t_{BI} + C_{opp} \cdot N \cdot F(t_{BI} \mid \theta) + C_{field} \cdot N \cdot \left[F(t_B + t_{miss} \mid \theta) - F(t_B \mid \theta)\right]$$

where

C_{setup} = fixed setup cost

C_B = time dependent cost for burn-in per unit per unit time

C_{opp} = lost opportunity cost of products which fail during burn-in

C_{field} = cost of field repairs during the warranty period

N = number of parts

t_{BI} = burn-in time (for a given part or batch)

t_{miss} = mission length (or time in field under warranty)

$F(t \mid \theta)$ = CDF of parts given a set of mixed Weibull parameters θ

For the problem addressed in this dissertation, the same general scheme used by Perlstein, of fixed, variable, and two different failure costs, is used as a basis but with some important modifications.

$$C_{total} = C_{setup} + C_B + C_{FB} + C_{FF}$$

where

C_{setup} = fixed set-up cost

C_B = burn-in cost (dependent on number of pieces and length of time)

C_{FB} = cost of failures during burn-in

C_{FF} = cost of failures in the field

As in [Perlstein, 2001], the burn-in costs depend on the number of devices to be burned in and the length of the burn-in[12]. However, the cost may increase if the length of the burn-in exceeds the allotted program schedule or if the capacity of the burn-in fixture is exceeded. The notion of capacity constraints was introduced in [Chi, 1989]. However, Chi's paper did not explicitly relate capacity to the cost impacts, most obviously the cost of schedule slips must be included if the capacity constraint is violated. [Alani, 1996] also considers capacity and schedule constraints, but did not allow for long burn-in times to consume useful life in the strong subpopulation, as was done in [Perlstein, 2001]. The model proposed here does both.

[12] Although this model is written in terms of performing device burn-in, the same model could be adjusted to apply to board or unit burn-in.

Without including the capacity and schedule slip components, the cost of performing the burn-in will be incorrect and performing the burn-in may appear much less expensive than it really is. This would lead to incorrect decisions about the optimal and most cost effective burn-in time.

New Model Formulation

If the number of pieces is less than the capacity of the burn-in fixture and the burn-in time is less than the time allowed in the schedule then burn-in cost is the same as that proposed by [Perlstein, 2001]

$$C_{B1} = N \cdot t_{BI} \cdot C_t$$

where $C_t =$ burn-in test cost per part per unit time

If the number of pieces is less than capacity but the burn-in time is greater than the time allowed in the schedule then there is a cost for the schedule slip. The schedule slippage cost is assumed to be a function of the amount of time over the schedule allowance.

$$C_{B2} = N \cdot t_{BI} \cdot C_t + (t_{BI} - t_{sch}) \cdot C_{slip} \text{, if } t_{BI} > t_{sch}$$

If the number of pieces is over capacity, then the burn-in must be performed in multiple batches. There are costs associated with performing the switch to a new batch in addition to the burn-in cost per part per unit time.

$$C_{BI3} = N \cdot t_{BI} \cdot C_t + C_{chg} \cdot (K - 1)$$

where K = number of batches = $\left\lceil N / Capacity \right\rceil$

Note that the amount of time needed to complete the burn-in for all parts is $t_{BI} \cdot K$.

This assumes that the set up for the first batch is included in the overall set up costs C_{setup}.

If the number of pieces is over capacity and the total time for performing the burn-in on all batches is more than the time allocated in the schedule, then there are penalties for both the batch switch over and the schedule slippage.

$$C_{B4} = N \cdot t_{BI} \cdot C_t + C_{chg} \cdot (K-1) + (t_{BI} \cdot K - t_{sch}) \cdot C_{slip}$$

The cost of failure during burn-in is the cost of the part (C_p) plus the costs associated with the man power to handle the replacement. Of course, the parts from the weak subpopulation are at higher risk for failing during the burn-in.

Field failures have higher cost than failures during burn-in as they may impact availability, warranties, customer payments, and public perception of the product.[13] The cost model described in this dissertation is based on one commonly used in the satellite market where field repairs are not feasible. One failure might impact so-called incentive payments. In these cases, the first failure is noted by the customer

---

[13] The cost of money may also impact the difference in field failure versus burn-in failure costs. Postponing the cost associated with the failure might be preferable in some cases. However, the cost of money is ignored in this model, since it is expected that the increase in cost between a failure occurring in the field and it occurring on test far outweighs the cost of money. However, the cost of money could impact the conclusions under some circumstances.

and they reduce payment to the supplier, but the product is still likely to be functional due to redundancy. However, if two failures occur then there is a program loss and the penalty cost is much higher.[14]

Case Study Parameter Values

This dissertation looks at several cases, which can roughly be divided into a satellite example and a high end consumer example. Neither involves field repairs, but both involve a cost for field failures.

Some parameters were assumed to be the same for both sets. Others, such as mission life and field failure cost were different. The common parameters are,

$C_{setup} = \$200$ [15]

$C_t = \$0.01$ per part, per hour [1]

$C_{chg} = \$15$ per batch[16]

$C_p = \$5$ per part[17]

$n = 100$ parts

$\theta_s = 10000000$ hours

$\beta_s = 2$

[14] There does not seem to be an analytical solution to minimizing the cost model with respect to t_{BI}. Numeric methods are required.
[15] [Reddy, 1994], [Pohl, 1995], [Yan, 1997]
[16] [Perlstein, 2001]
[17] [Yan, 1997]

In satellite contract pricing, typically there is an up-front cost of purchasing the satellite and a series of performance incentives that are only paid after certain milestones. These incentives are typically 10% to 20% of the cost of the satellite. [de Selding, 2005] It is only the incentive payments[18] that are impacted by a field failure. Satellites are typically \$200 million to \$1 billion. [Bearden, 2001] So that would put incentive payments between \$20 million and \$200 million. \$50 million for the total incentive package was selected as representative of a satellite case. Incentive payments are tied to several different units, and loss of performance on any one of these units results in a loss of payment, with a total loss of incentive only coming from a total satellite loss (C_{loss}). It was assumed that the payments were tied to the performance of 10 critical unit types (each with primary and redundant units) and loss of the primary unit for any of them would result in a 10% loss of incentive payments (C_{incent}).[19] Since the weak parts were assumed to only be in one of these critical unit types, a part field failure was assumed to cost \$5 million in incentive payments. So, for the satellite case,

$$C_{loss} = \$50\,\text{mil}$$

$$C_{incent} = \$5\,\text{mil}$$

$$t_{mission} = 100000 \text{ hours (approximately 11 years)}$$

[18] Also the potential for future business and insurance costs on future programs, but these are ignored for this study.

[19] In reality, this amount would also depend on when in the mission the field failure occurred, but this complexity was dropped for the purposes of the model developed here. Note that this also means that in some cases, if enough critical, primary units fail, incentive payments could fall to \$0, even if the satellite is fully functional. The exact impact of individual unit failures varies greatly between contracts.

For the consumer case,

$$C_{loss} = \$5000$$

$$C_{incent} = \$0$$

$$t_{mission} = 50000 \text{ hours (between 5 and 6 years)}$$

The cost of schedule slip, C_{slip}, may include only the cost to having a few engineers without work for a few hours or it may be a much bigger cost. If the parts go into a unit on the critical path of the program, a slip in schedule for the parts may result in a "marching army" being without work awaiting delivery of the parts to the unit an the unit to the system. It may also be the case that a slip in the schedule of this program could delay other programs – and cause them to have slip costs, impacting the manufacturer for both programs. For this reason, two values of C_{slip} were considered for both models.

$$C_{slip} = \{\$0, \$100, \$1000\} \text{ per hour}$$

In both cases other parameters were allowed to vary to determine their impact on the cost model. In particular,

$$t_{sch} = \{10, 50, 100\} \text{ hours}$$

$$capacity = \{10, 20, 50, 100\} \text{ parts per batch}$$

The parameters for the weak subpopulation were also varied:

$$p = \{0.01, 0.1, 0.5, 0.9, 0.99\}$$

$$\theta_w = \{0.5, 1, 5\}$$

$$\beta_w = \{0.5, 1, 2\}$$

Discussion

The model allows for a balance between the cost of performing longer burn-ins and the cost of field failures. The cost of field failures is, of course, driven by the mixture ratio and the parameters of the weak subpopulation. Thus different Weibull parameters lead to different optimal burn-in times, as shown in Figure 32. Since the characteristic life of the strong subpopulation is so much longer than the mission life, the only measureable disadvantage of a longer burn-in is its impact on schedule. Changes in the value of the mixture ratio p do not have a big impact on the optimal burn-in time but impacts the total cost, as illustrated in Figure 33.

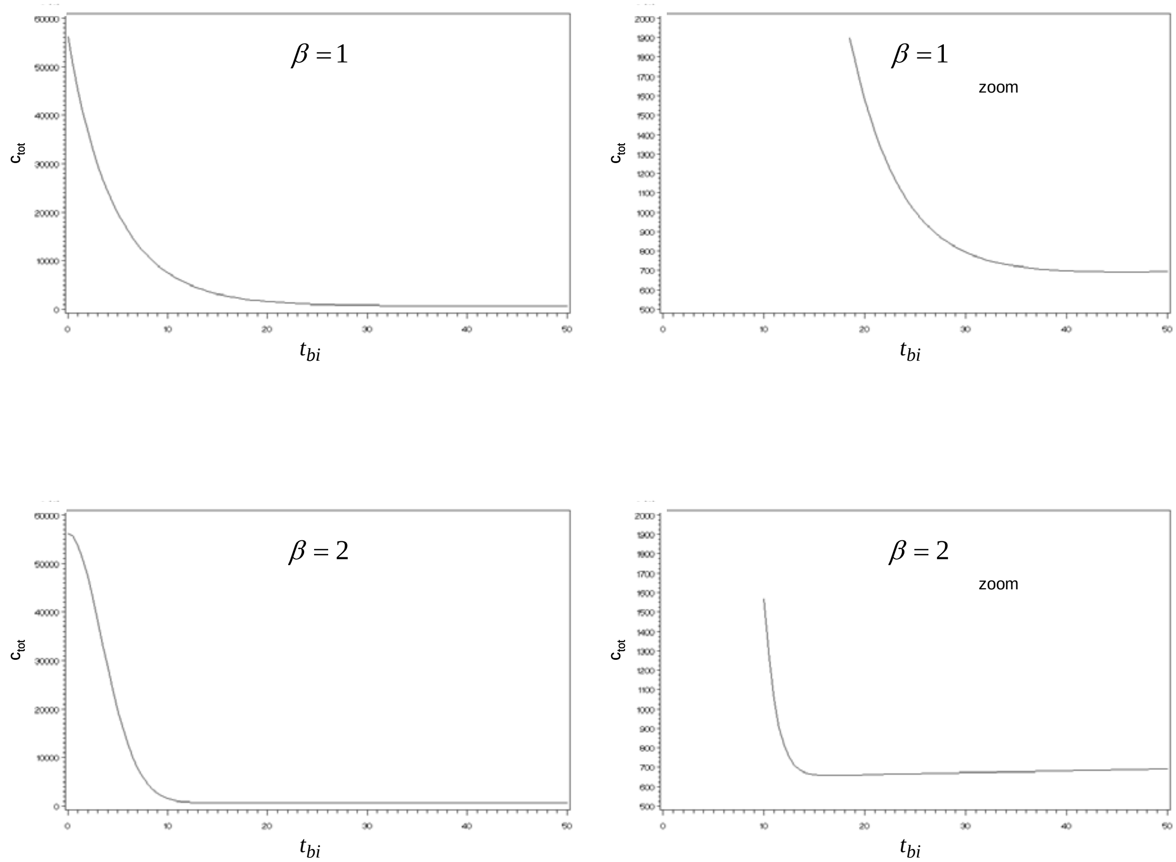


Figure 32. Impact of weak subpopulation's shape parameter on optimal burn-in time. If $\beta_w = 1$, then the optimal burn in time is 46 hours. However, if $\beta_w = 2$, then the weak parts can be weeded out faster and the optimal burn-in time is only about 16.5 hours. ($p_0 = 0.01$**,** $t_{mission=100,000}$**,** $\theta_w = 5$**,** $capacity = 10$**,** $c_{slip} = 0$**,** $c_{incent} = 5{,}000{,}000$**,** $c_{loss} = 50{,}000{,}000$ **)**

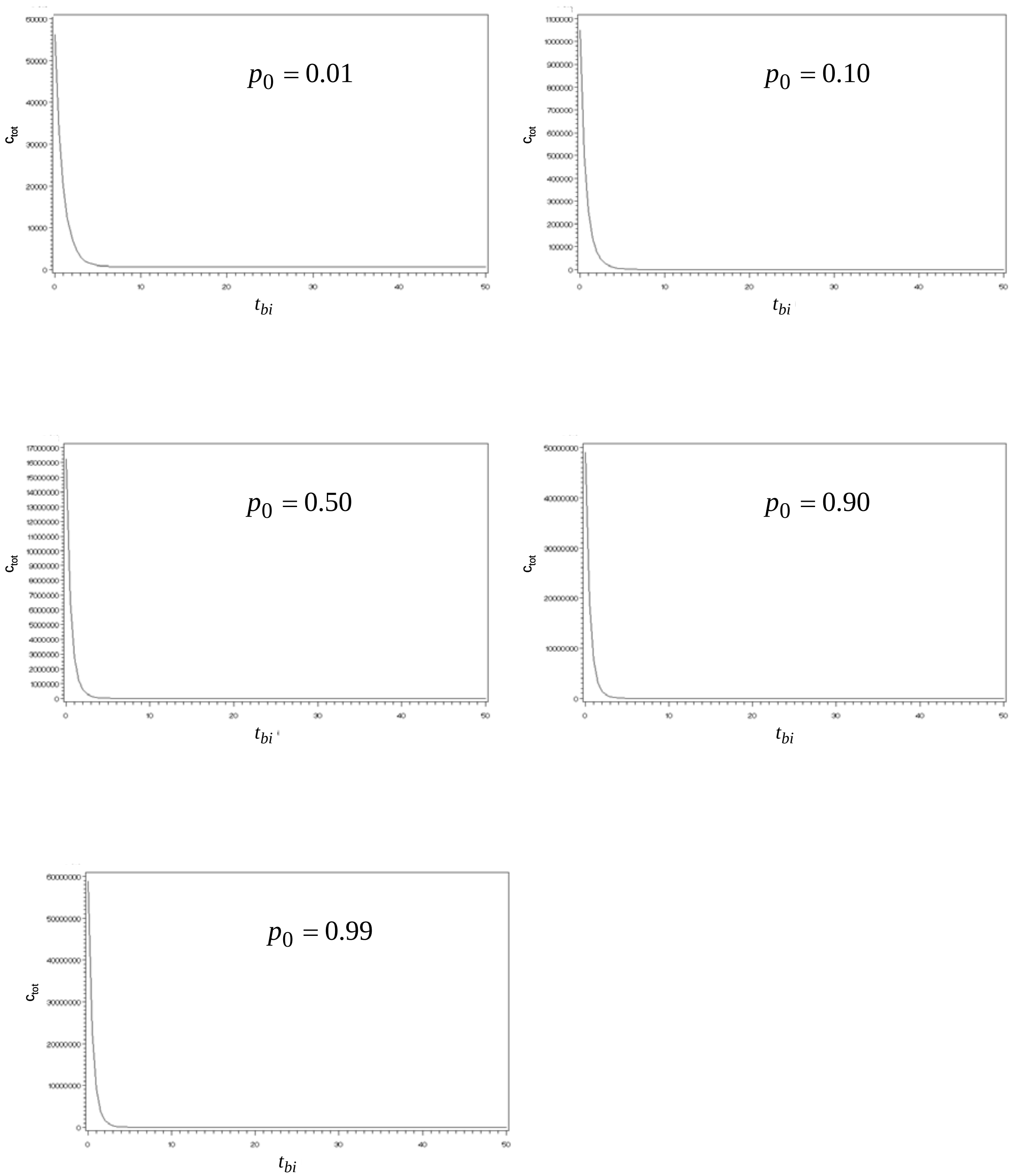


Figure 33. Impact of varying mixture ratio. The five plots represent values of p_0 **{0.01, 0.1, 0.5, 0.9, and 0.99}. (**$t_{mission=100,000}$, $\theta_w = 1$, $\beta_w = 1$, $capacity = 10$, $c_{slip} = 0$, $c_{incent} = 5{,}000{,}000$, $c_{loss} = 50{,}000{,}000$ **)**

Similarly, other parameters tend to impact the total cost more than the optimal burn-in time. For example, increasing the parts cost (c_p) from \$5 to \$50 raises the total cost by about \$90 but doesn't change the optimal burn-in time.[20] There are similar results when the batch change cost (c_{change}) is adjusted. Changing the per part, per hour test cost (c_t) can impact the optimal burn-in time. If slippage cost is low, then the optimal burn-in time is driven by the test cost. However, if slippage cost is high, the optimal burn-in time is insensitive to the test cost and only the total cost is impacted. For example, if the slippage cost is \$0, adjusting the per part, per hour test cost from \$0.01 to \$0.10 changes the optimal burn-in time from 102 hours to 64.5 hours and almost doubled the total cost from \$719 to \$1403.[21] However, if the slippage time is \$100 per hour, both hourly test costs result in an optimal test time of 17 hours. The total cost changes from \$24,854 to \$25,007.

The differences between the satellite case and the commercial case, in terms of mission length and cost of field failure, have a large impact on the optimal burn-in time as well. For example, for the same case shown in Figure 34 with $\beta_w = 1$, while the satellite case showed a steady decrease in cost over burn-in time, for the commercial case, there is a steady increase in cost with burn-in time. For all cases with mixture ratio of 1%, there was no benefit of performing burn-in for the

[20] This was true for at least 3 cases of other parameters for satellites:
Case 1={ $p = 0.1, \theta_w = 1, \beta_w = 0.5, c_{slip} = 0, t_{sch} = 10, capacity = 10$ };
Case 2={ $p = 0.5, \theta_w = 1, \beta_w = 1, c_{slip} = 0, t_{sch} = 10, capacity = 10$ };
Case 3={ $p = 0.1, \theta_w = 5, \beta_w = 1, c_{slip} = 100, t_{sch} = 10, capacity = 10$ }
[21] Satellite case and $p = 0.1, \theta_w = 1, \beta_w = 0.5, t_{sch} = 10, capacity = 10$

commercial case. Burn-in only became cost effective if the mixture ratio was more than 10%, shown in Figure 35. This illustrates that the balance between doing burn-in or not depends a lot on the expected loss from field failures. If only a very small percentage of parts are at risk and the cost of a field failure is small then there will be little value to burn in. For the commercial case, when only 1% are at risk, the expected field failure loss is only $0.01^2 \cdot \$5000 = \$0.50$. It is unlikely that such a small expected loss would ever justify spending money on burn-in. It is only when the cost of a field failure is higher or a higher percentage of parts are at risk for failure that a burn-in is warranted.

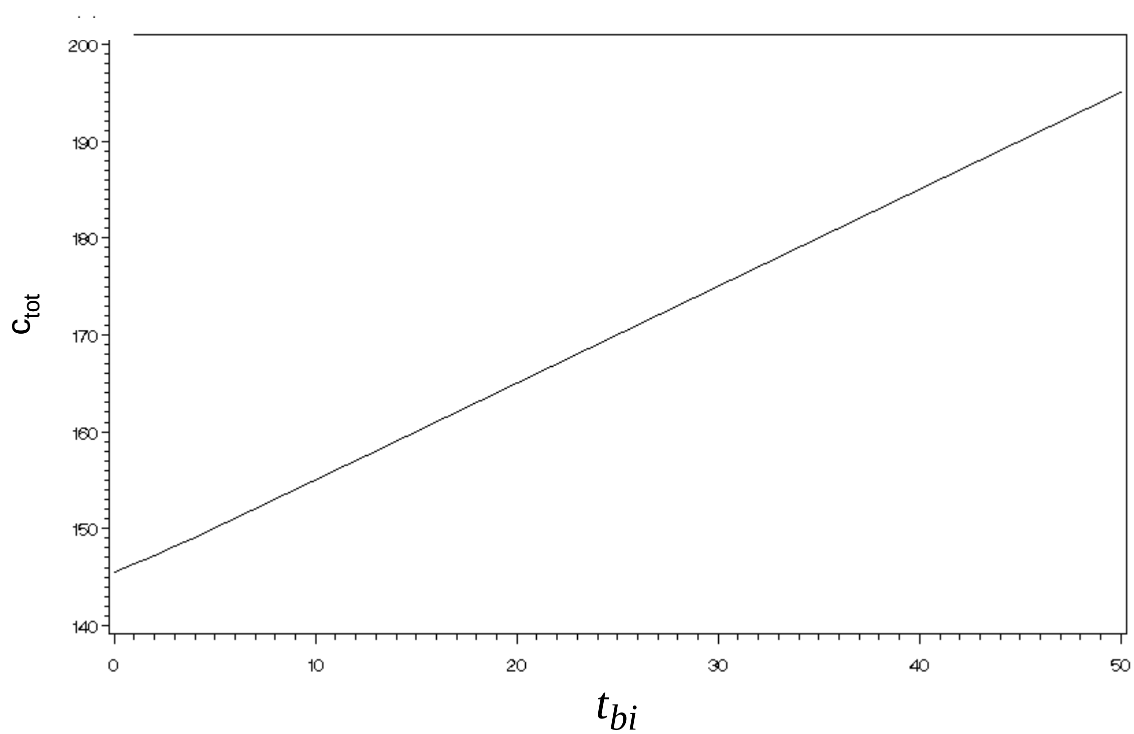


Figure 34 Commercial case with small percentage of weak parts. Note that when the percentage of weak parts are small, the commercial case almost always shows little value to burn-in.
($p_0 = 0.01$, $t_{mission=50,000}$, $\theta_w = 5$, $\beta_w = 1$, $capacity = 10$, $c_{slip} = 0$, $c_{incent} = 0$, $c_{loss} = 5,000$ **)**

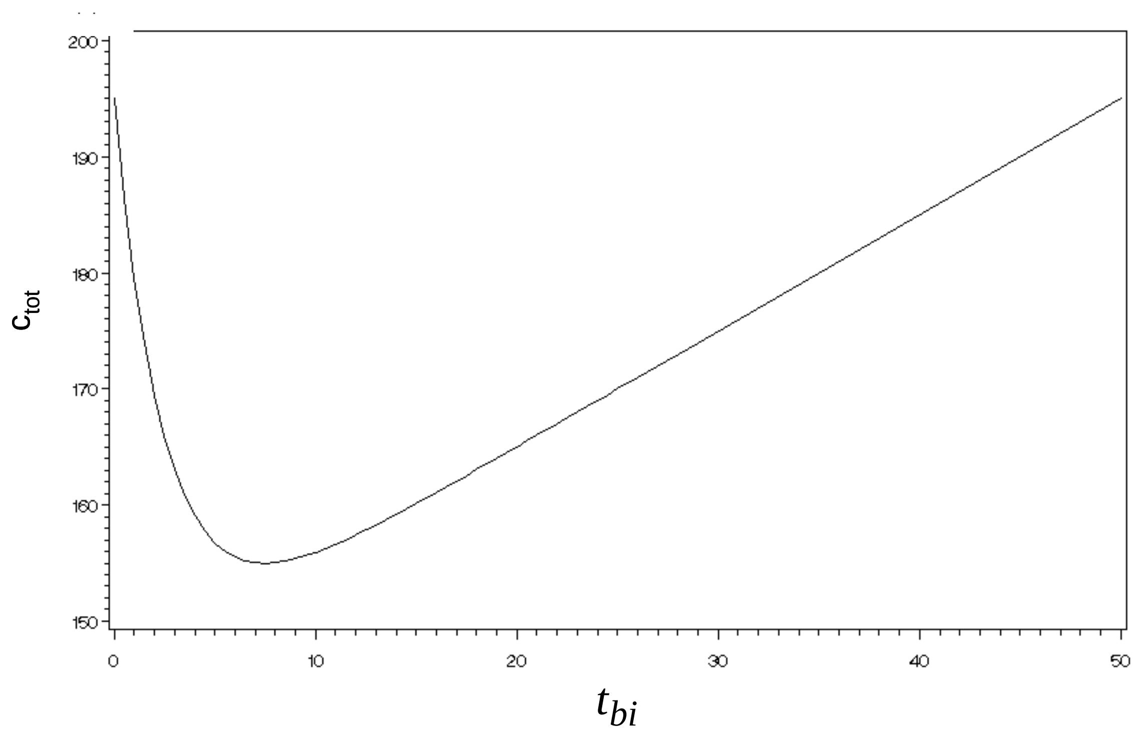
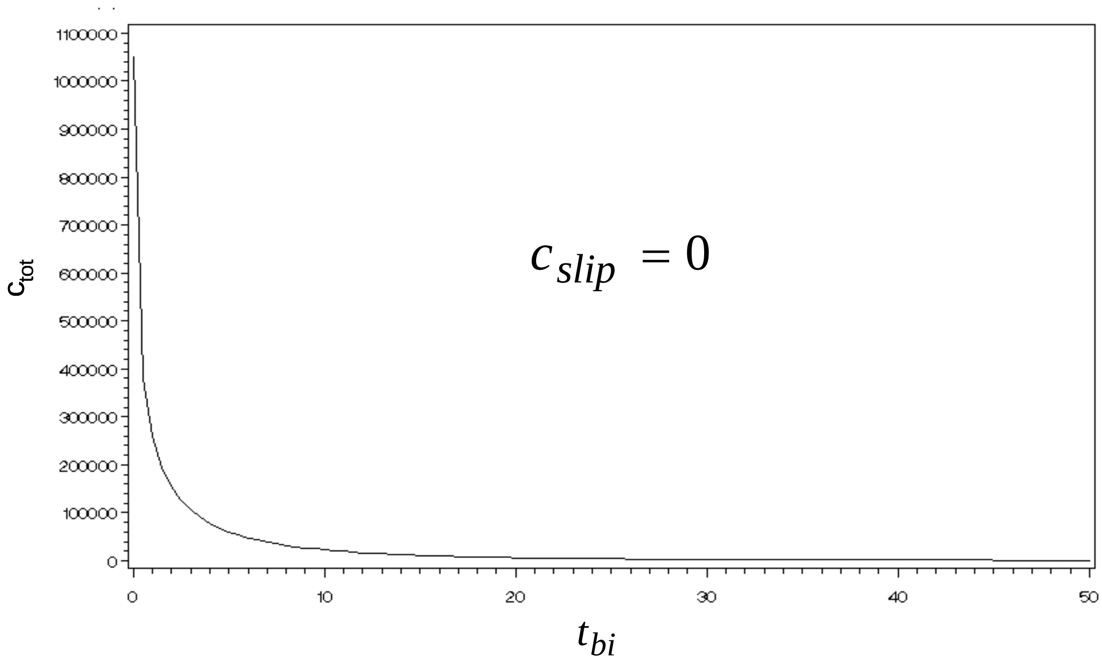


Figure 35. With larger percentages of parts at risk, short burn-ins may be effective. ($p_0 = 0.10$ **,** $t_{mission=50,000}$ **,** $\theta_w = 5$ **,** $\beta_w = 1$ **,** $capacity = 10$ **,** $c_{slip} = 0$ **,** $c_{incent} = 0$ **,** $c_{loss} = 5{,}000$ **)**

The inclusion of the cost of slipping schedule impacts the decision on optimal burn-in time. For example, consider the satellite case where the mixture ratio is 10%, the Weibull parameters for the weak subpopulation are $\theta_w = 1$ and $\beta_w = 0.5$, and the schedule only allows for a slip of 10 hours, shown in

Figure 36. If the issue of slipping schedule ($C_{slip} = 0$) is ignored, similar to the models proposed by [Perlstein, 2001] then the optimal burn-in time is past 50 hours. The total cost at 50 hours is about $936. However, using the new model, if each hour of slippage cost $1000, then the optimal burn-in time is 17 hours and the total cost is about $15,719. The use of the new model, which realistically takes into account that there are costs to slipping schedule, significantly shortens the recommended burn-in time and reflects a much larger cost. Previous models would have made an incorrect recommendation.

This impact only increases when a capacity constraint is also implemented. Consider that the capacity is only 10 parts, so to burn-in all parts will take 10 batches. If there is no cost of schedule slippage, the optimal burn-in time is still past 50 hours (even though there is now a cost of swapping batches). However, now instead of a slow increase in cost past 17 hours, there is a dramatic increase in costs after 6 hours. This is illustrated in **Figure 37**.



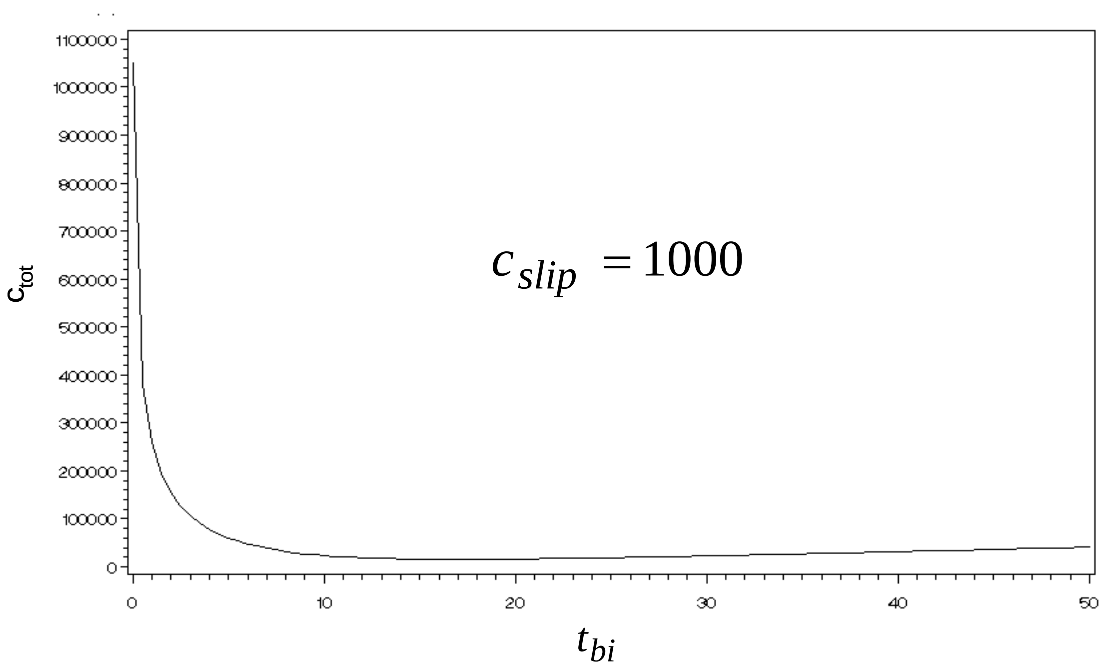


Figure 36. Impact of schedule slippage cost. When there is no cost to schedule slippage, then the optimal burn-in is past 50 hours; more burn-in always reduces costs. However, if the cost of schedule slippage is included, then past 17 hours the burn-in costs more money than it saves. ($p_0 = 0.10$, $t_{mission=100,000}$, $\theta_w = 1$, $\beta_w = 0.5$, $capacity = 100$, $t_{sch} = 10$, $c_{incent} = 5{,}000{,}000$, $c_{loss} = 50{,}000{,}000$)

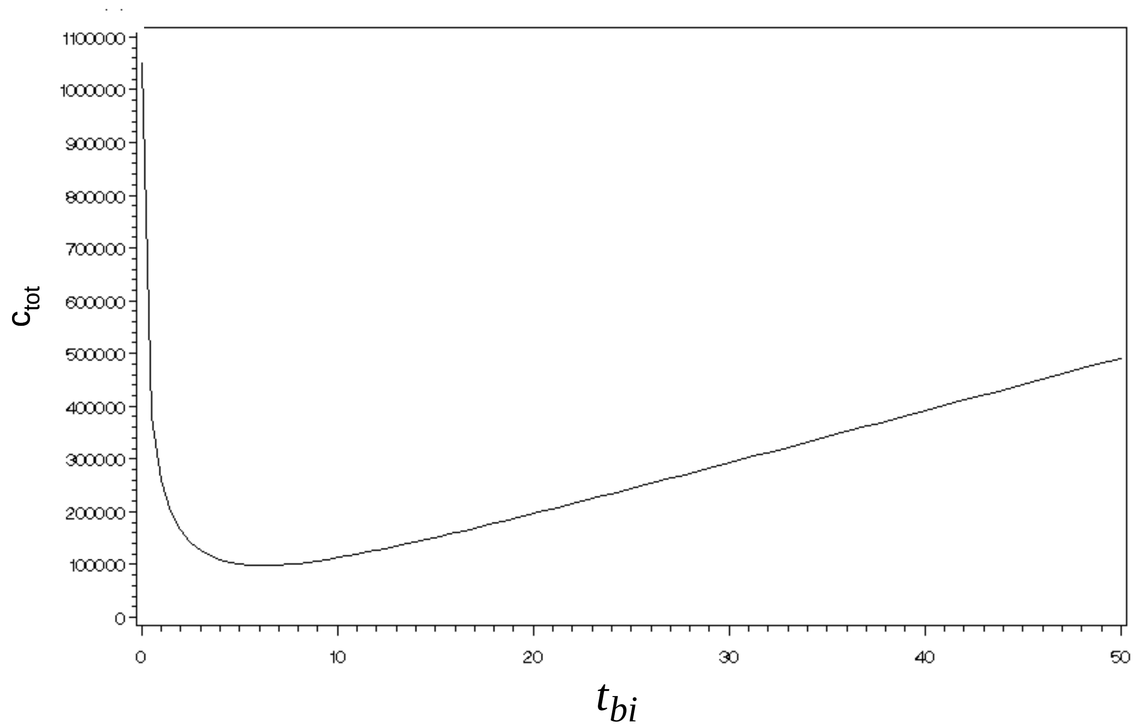


Figure 37. If capacity is also limited, the schedule slippage costs are multiplied by the number of batches, so the increase in costs of going past the schedule is much more. ($p_0 = 0.10$ **,** $t_{mission}$=100,000 **,** $\theta_w = 1$ **,** $\beta_w = 0.5$ **,** $capacity = 10$ **,** $t_{sch} = 10$ **,** $c_{slip} = 1000$ $c_{incent} = 5{,}000{,}000$ **,** $c_{loss} = 50{,}000{,}000$ **)**

Chapter Summary

One method of determining if the burn-in is cost effective is to find the burn-in length that optimizes the life-cycle costs. With the cost model, one can find the optimum balance between investing in burn-in and avoiding field failure. In this chapter, a new cost model was developed that added the impacts of schedule slippage and capacity to existing cost models for burn-in testing. It was shown that these improvements can have significant impact on determination of the optimal burn-in length.

Chapter 8: Impact of Model Error on Burn-in Decision Rules

As discussed in Chapter 2, there are a number of different decision criteria that are used for determining whether extended burn-in should be performed and, if so, the length of the optimal burn-in. Two types of decision rules are examined here.

The first is common in government contracts. It requires that a certain probability of survival at end of life is achieved. To that end, if a weak subpopulation or infant mortality problem is suspected, the burn-in time must produce a conditional reliability that meets the requirement.

The second decision rule takes advantage of the cost model developed in Chapter 7 and is more likely applicable to more commercial cases. It determines the optimal burn-in length based on minimizing overall lifecycle cost.

Optimal Burn-in Time Given an End of Life Reliability Requirement

Chapters 4 through 6 explored the impact of the estimation method errors on the reliability estimate. However, when a specific reliability goal must be met, it is a conditional reliability that is of interest. Recall that, for a standard reliability estimate, the Bayesian technique performed better than the others when the true mixture ratio was small. The MLE and Trunsored techniques performed similarly unless the mixture ratio was very large. When the mixture ratio was large, the Trunsored method would occasionally predict unreasonably large values for *p* and thus had much higher variability than the other methods. The MLE method performed the best at high mixture ratios. These results are once again illustrated in Figure 38.

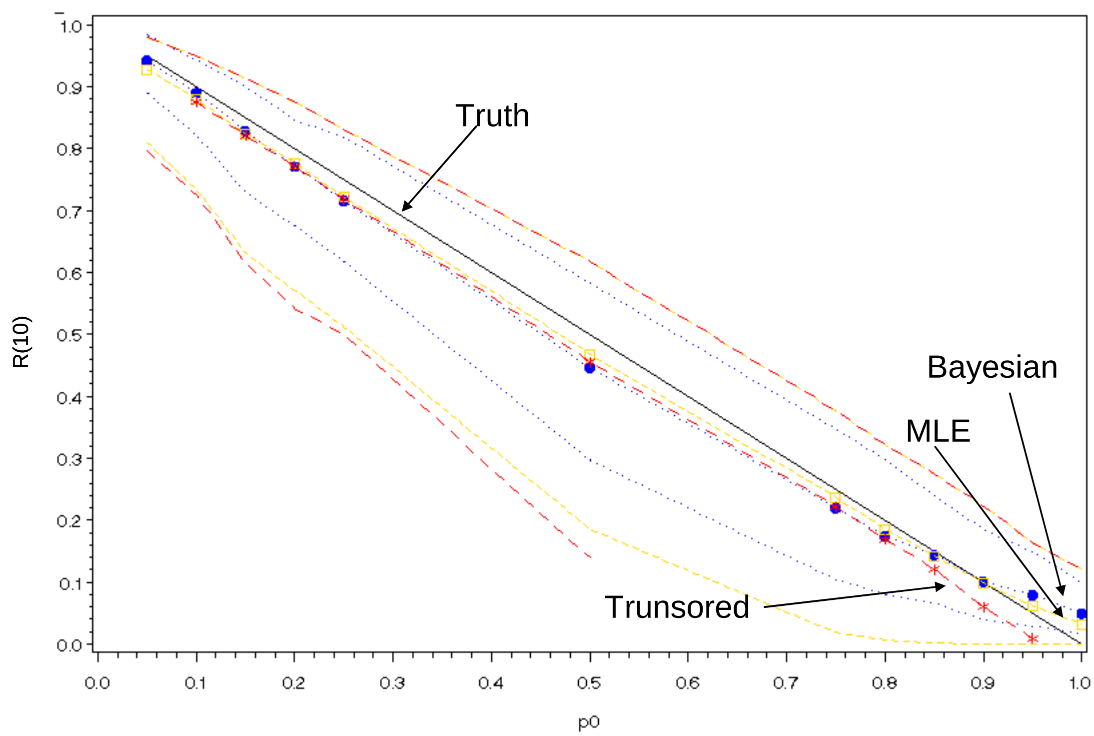


Figure 38. True and estimated reliability at time=10 ($\beta_w = 1$, $\theta_w = 1$) for all three estimation methods. Blue is Bayesian; gold is MLE; red is Trunsored.

These results continue to be true, only to a larger extent, when the conditional reliability is examined. This is illustrated in **Figure 39**. The Baysian estimate of the conditional reliability is much better than the others when the mixture ratio *p* is small. The MLE estimate tends to be better at larger mixture ratios. The Trunsored method has very wide variability, skewing its results.

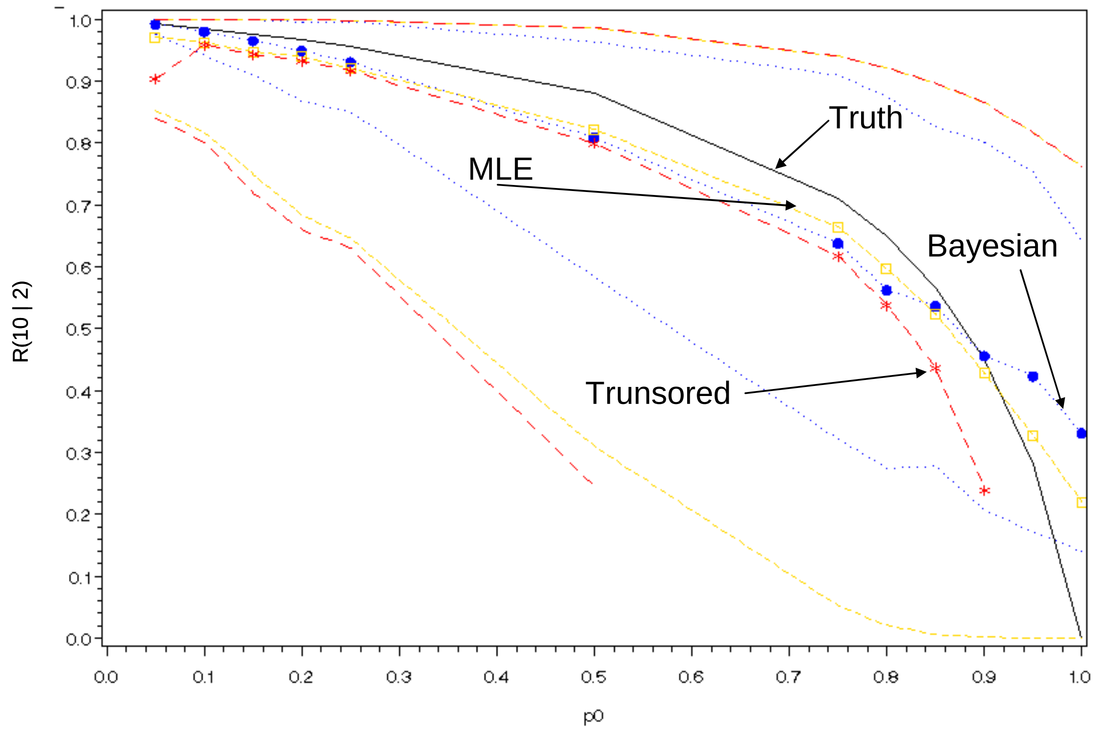


Figure 39. Reliability at time=10 conditioned on time=2 ($\beta_W = 1$, $\theta_W = 1$). Blue is Bayesian; gold is MLE; red is Trunsored.

To examine the impact of these errors on conditional reliability on the determination of the appropriate burn-in time, it was assumed that a conditional reliability of 0.99 was required at end of life. The results if a mission was relatively short, only 10 compared to an original censor time of 2 and a characteristic life of 1 were examined.[22] The simulation cases from the previous chapters were used. For each set of estimated parameters, it was determined what length of burn-in (t_{bi}) was required such that $R(10 + t_{bi} \mid t_{bi}) > 0.99$. This burn-in time was compared to the true required burn-in time, based on the true parameter values. The results are summarized in

[22] The results are almost identical for longer life times because the reliability at t=10 larger times are so close with these parameter values. Using the lifetimes proposed for the satellite and commercial cases described in Chapter 7 tent to only change the times by less than 0.1.

Table 3. Since the reliability at end of life for the Trunsored model was frequently negative, any negative reliability was set to 0. Once this condition was applied, the results of the Trunsored and MLE methods were identical. The end of life reliability depends almost exclusively on the value of the mixture ratio *p* for our case studies (since the characteristic life is relatively small). The cases where the Trunsored model was selecting a value of *p* greater than 1 result in the negative reliability conditions. By forcing the reliability to be 0, in essence, the same limit as on the MLE case is being imposed. Thus the two methods produce the same burn-in results if negative reliability is disallowed. They would not necessarily produce the same results if the characteristic life was larger than the censor time and closer to the end of life time.

Table 3. Optimal burn-in time given an end of life reliability requirement

	Mean	**STD**	**Median**	**Q1**	**Q3**	**5th percentile**	**95th percentile**	**Percentage of cases where $R(10+t_{bi} \mid t_{bi}) > 0.99$ could not be achieved**
p=0.1	**True burn-in time to achieve $R(10+t_{bi} \mid t_{bi}) > 0.99$: 2.5**							
Bayes	2.8	1.5	2.5	1.7	3.9	1.2	5.9	0
MLE / Trunsored	3.9	8.7	1.7	1.3	2.9	1.1	13.6	10%
p=0.5	**True burn-in time to achieve $R(10+t_{bi} \mid t_{bi}) > 0.99$: 4.7**							
Bayes	6.4	3.3	5.5	3.9	7.6	2.8	13.6	0
MLE / Trunsored	6.4	7.4	4.3	3.1	6.5	2.3	18.8	5%
p=0.9	**True burn-in time to achieve $R(10+t_{bi} \mid t_{bi}) > 0.99$: 6.9**							
Bayes	7.0	2.2	6.7	5.5	8.0	4.0	10.5	0
MLE / Trunsored	6.4	2.8	5.8	4.5	7.6	3.4	11.9	22%
p=1.0	**True burn-in time to achieve $R(10+t_{bi} \mid t_{bi}) > 0.99$: Not possible**							
Bayes	6.6	1.6	6.3	5.5	7.4	4.6	9.2	0
MLE / Trunsored	6.2	2.2	5.7	4.6	7.3	3.5	10.6	48%

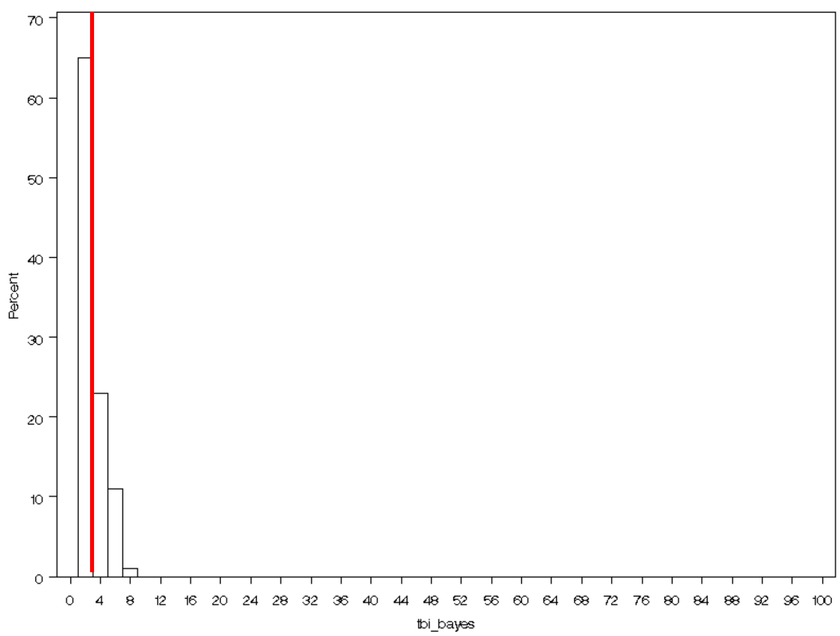


Figure 40. Bayesian estimates of optimal burn-in time for p0=0.1.

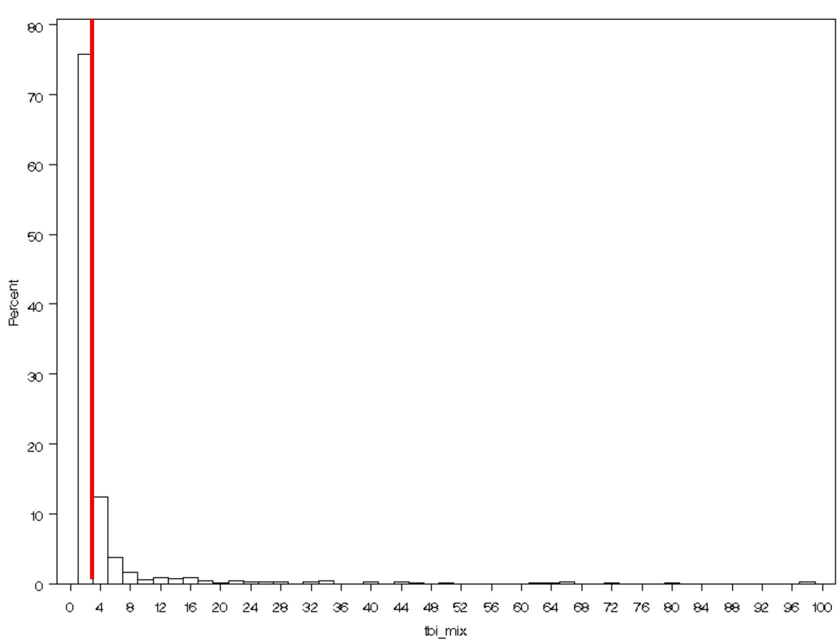


Figure 41. MLE and Trunsored estimates of optimal burn-in time for p0=0.1.

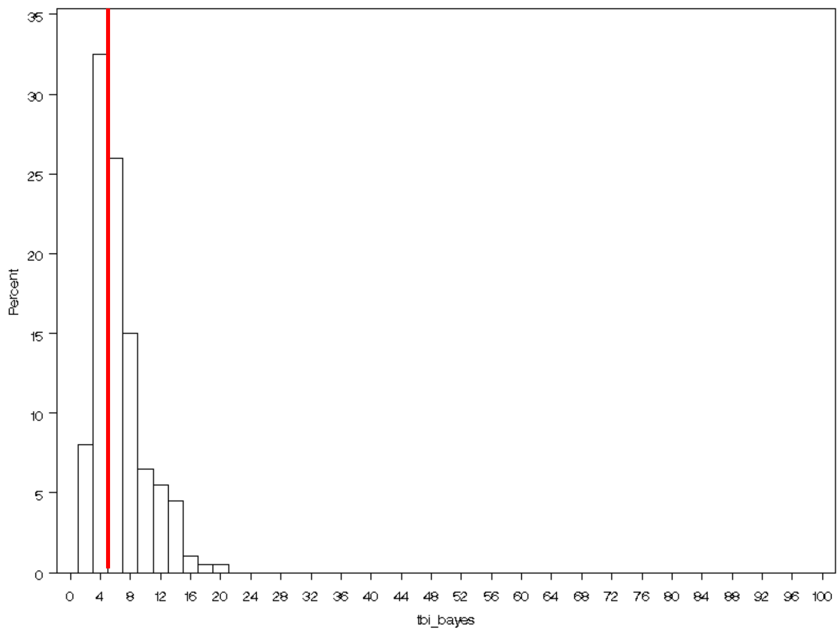


Figure 42. Bayesian estimates of optimal burn-in time for p0=0.5.

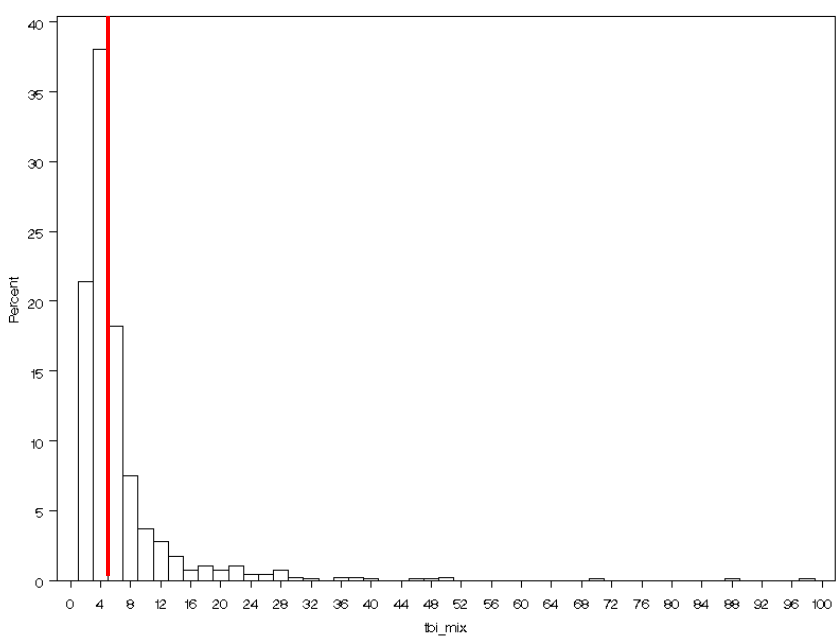


Figure 43. MLE and Trunsored estimates of optimal burn-in time for p0=0.5.

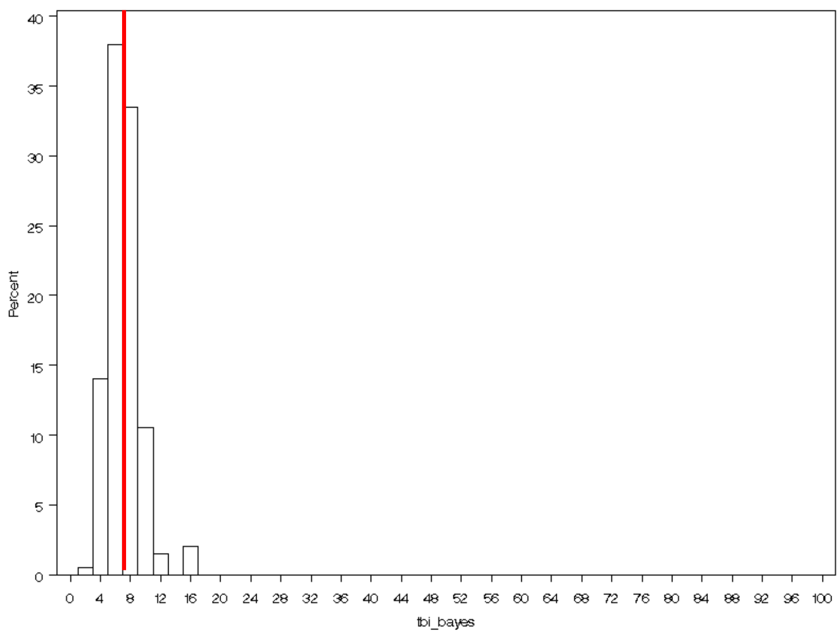


Figure 44. Bayesian estimates of optimal burn-in time for p0=0.9.

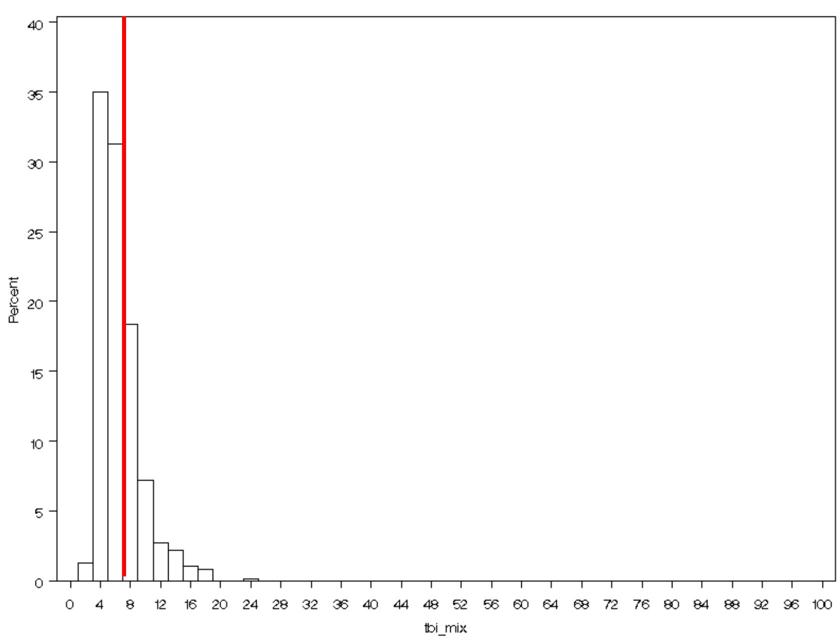


Figure 45. MLE and Trunsored estimates of optimal burn-in time for p0=0.9.

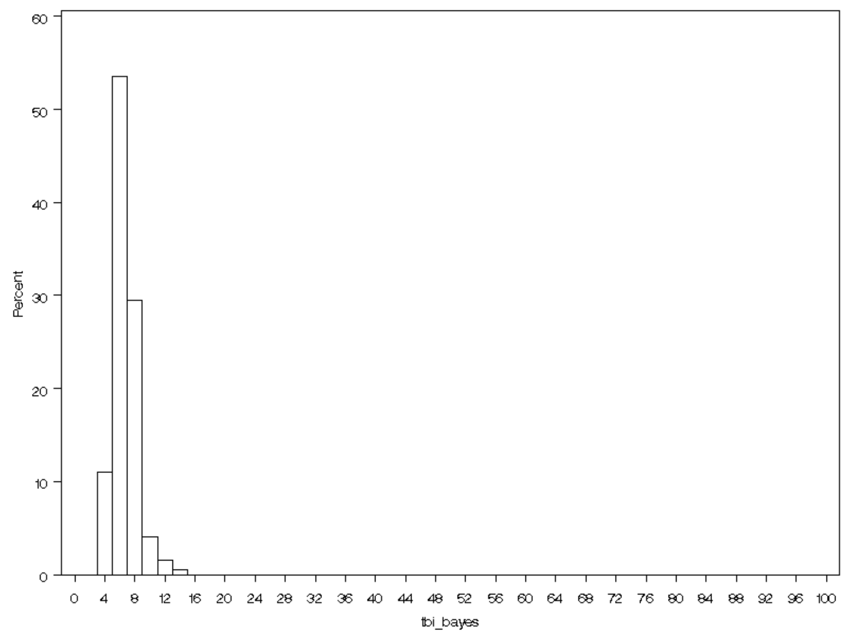


Figure 46. Bayesian estimates of optimal burn-in time for p0=1.0.

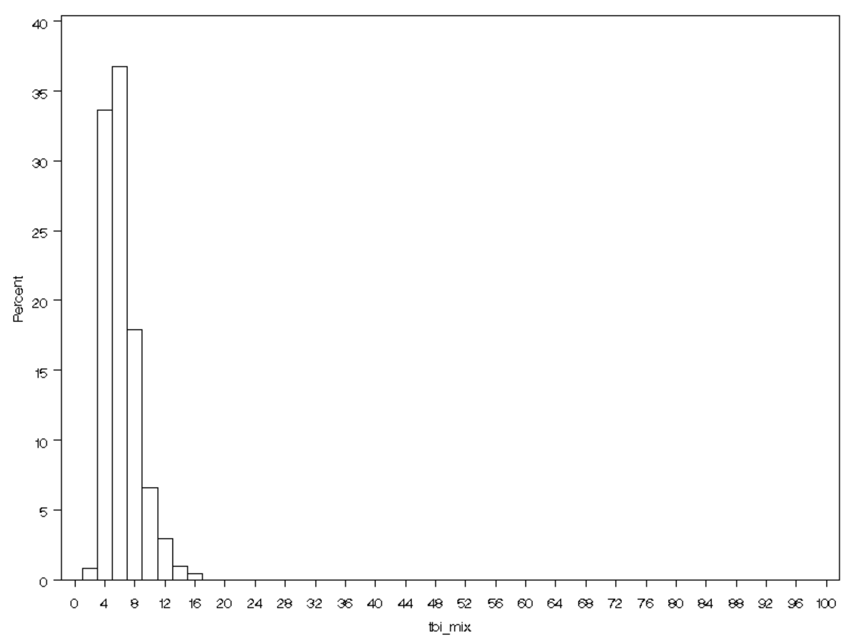


Figure 47. MLE and Trunsored estimates of optimal burn-in time for p0=1.0.

On average, the methods recommend a longer burn-in than is necessary. However, this is in part due to the skewness of the distributions of estimated values. When the median optimal burn-in times are examined, the methods tend to suggest burn-ins that are slightly too short. In general, the Bayesian results provide the most accurate estimate of the burn-in time required to meet the decision rule. Even at the mixture ratio of 0.5, which is unlikely under the baseline Bayesian prior examined, the errors in the Bayesian method are not much higher than the others. In other cases, where the prior is more accurate, the Bayesian method performed better than the others.

However, the Bayesian method does not ever predict that the mixture ratio is 1 and the burn-in will be ineffective. In all cases, it provides estimates such that the 0.99 reliability can be achieved. In the case where p_0=1.0 and β=1, this is incorrect. All parts are at risk and there is a constant failure rate. In this situation, the burn-in is useless at screening for the anomaly. However, the Bayesian method still predicts that burn-in will screen weak parts and improve the end of life reliability.

Optimal Burn-in Time Given a Life-cycle Cost Decision Rule

A cost-based decision rule was examined for both the satellite and consumer assumptions presented in Chapter 7. For each, four cases were examined, varying the mixture ratio, capacity, and schedule slip costs. These cases are summarized in Table 4.

Table 4. Cost-based decision rule case studies

	p_0	*Capacity*	C_{slip}	t_{sch}
Case 1	0.1	10	0	10
Case 2	0.1	10	1000	10
Case 3	0.1	100	1000	10
Case 4	0.05	10	0	10

First the estimated costs under each method were examined for each case study. These are shown in Figure 48 for the satellite cases and Figure 50 for the commercial cases. 95th percentiles on the Trunsored case were very wide, with maximum costs more than an order of magnitude larger than those from the other methods. This was expected since the distribution of Trunsored estimates had such a long tail. However, the medians for all methods were close to the true costs, as illustrated in the enlarged figures in Figure 49 and Figure 51.

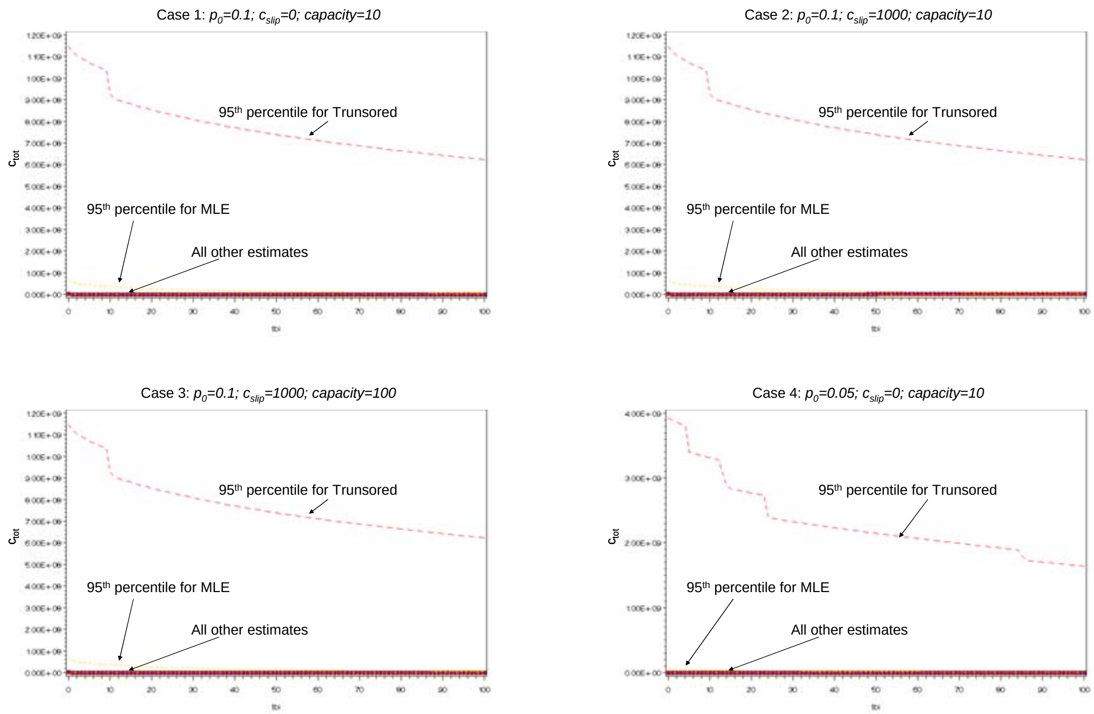


Figure 48. Satellite case full scale with 5th and 95th percentiles shown. Blue is Bayesian; gold is MLE; red is Trunsored.

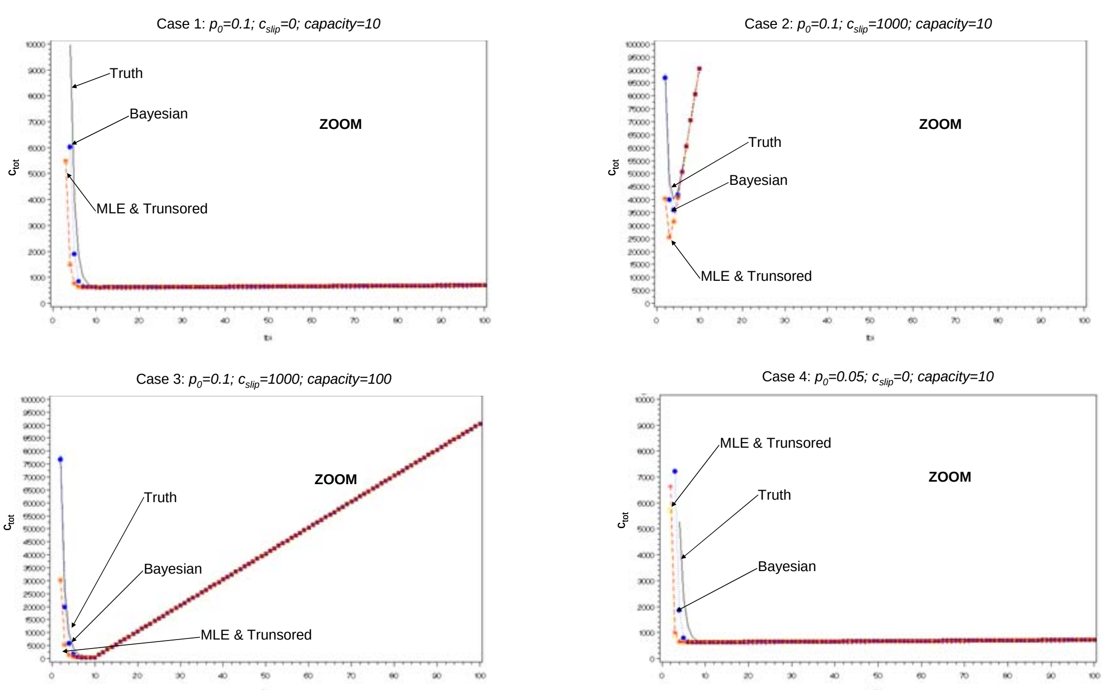


Figure 49. Satellite Case: Zoom with only medians shown. Blue is Bayesian; gold is MLE; red is Trunsored.

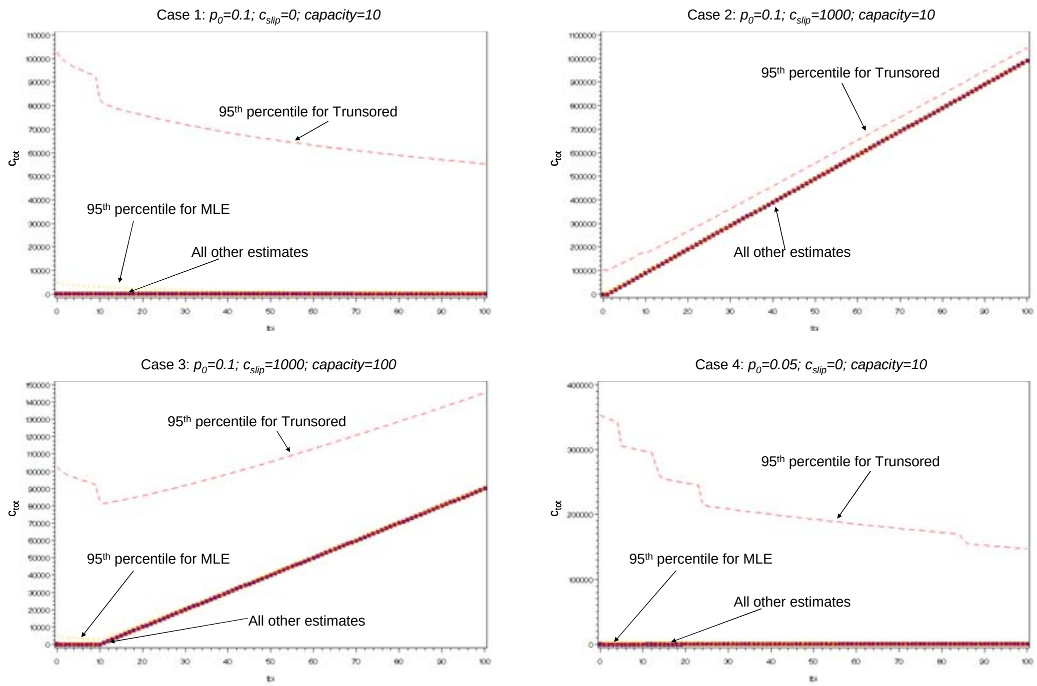


Figure 50. Commercial case full scale with 5th and 95th percentiles shown. Blue is Bayesian; gold is MLE; red is Trunsored.

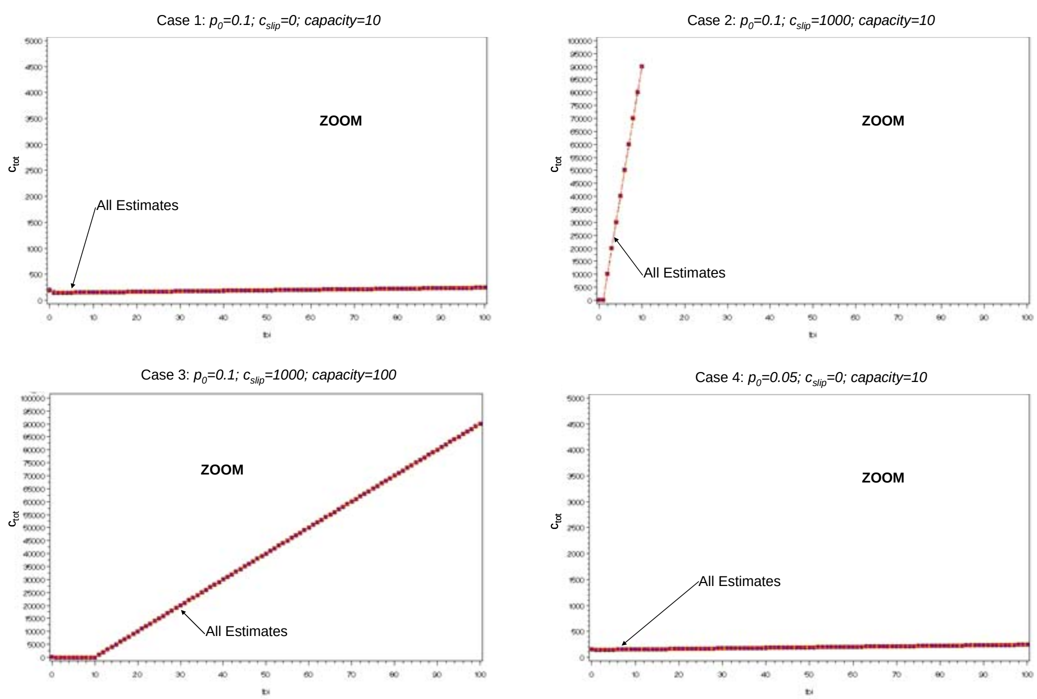


Figure 51. Commercial Case: Zoom with only medians shown. Blue is Bayesian; gold is MLE; red is Trunsored.

Perhaps of greater importance than the estimated cost at a particular burn-in time is the ability of the estimation method to predict the correct optimal burn-in time. As with the evaluation of minimum reliability based decision rule, the simulation cases from the previous chapters were used. For each set of estimated parameters, it was determined what length of burn-in (t_{bi}) minimized the estimated total life-cycle costs. This burn-in time was compared to the true required burn-in time, based on the true parameter values. The results are summarized in Table 5 for the satellite case and Table 6 for the commercial case. In some cases, the optimal burn-in time was estimated to be quite large. To reduce computation time, the maximum burn-in time examined was 100 hours. If the estimated parameters predicted an optimal burn-in time greater than 100 hours, the optimal burn-in time was set to 100 hours. This means that the summary statistics presented in Table 5 and Table 6 are biased a bit low. The percentage of cases where this truncation was applied is given in the tables. This bias is greatest in the lines with the greatest percentage of truncated cases and will be most severe in the estimated mean and standard deviation. So, for example, while the mean optimal burn-in time for the Trunsored method in Case 2 is reported as 14.4, already much higher than the true optimal of 4.2, the true mean for that method would be higher because 10.1% of cases should have been averaged in at values greater than 100.

Examples of the distributions of the optimal burn-in times are provided in Figure 52 through Figure 57. Notice that in some of the distributions, such as those shown in Figure 53 and Figure 56, there is a bimodality in the estimated optimal burn-in times. The peak around 10 is associated with cases where the optimal burn-in time is thought to be the maximum time before the schedule is exceeded (since $t_{sch} = 10$). In the satellite case in Figure 53, the true optimal is at about 10. However, there is still a peak around 4 or 5. These cases are associated with an estimated shape parameter $\beta > 1$ and an estimated mixture ratio close to correct. When the weak subpopulation is thought to have an increasing failure rate, the burn-in time can be shorter, corresponding to the smooth distribution of values less than 10 in the figure.

The skewness of the MLE and Trunsored method pulls the average optimal burn-in times much higher than truth, particularly for the satellite cases. Thus, on average the burn-ins would be longer than necessary, spending more money on the burn-in than required. However, the median optimal burn-ins are, for almost all cases, lower than the true optimal time. Thus in most situations, the methods would suggest burn-ins that were too short. Most of the time, these mistakenly too short bur-ins are not terribly costly, only about 10% to 20% more than the optimal burn-in's cost for satellite situations and only about 1% to 5% for the commercial cases. The median estimated values are very similar for all three methods. The more costly situations are where a longer burn-in is recommended than the true optimal. These occur much more often, and more severely, in the MLE and Trunsored methods as their predicted optimal burn-in times are very skewed. In cases where schedules slips are costly, such as cases 2 and 3 for the satellite systems, the average predicted length for the Trunsored and MLE methods is about 10 hours longer than optimal. These result in costs 2 to 20 times more than the optimal costs.

Table 5: Optimal burn-in time using cost-based decision rule and satellite case

	Mean	STD	Median	Q1	Q3	5th percentile	95th percentile	Percentage of cases where cost still decreasing with tbi>100
Case 1: True Optimal Burn-in Time=13.0								
Bayes	15.1	15.3	9.9	6.0	17.8	2.8	45.5	0.5%
MLE	23.6	32.8	8.2	4.4	22.1	2.2	>100	13.4%
Trunsored	23.8	33.0	8.2	4.4	22.2	2.2	>100	13.9%
Case 2: True Optimal Burn-in Time=4.0								
Bayes	4.2	1.7	3.9	2.9	5.1	1.8	7.5	0%
MLE	13.7	28.1	3.1	2.2	5.6	1.4	>100	8.2%
Trunsored	14.4	29.5	3.1	2.2	5.6	1.4	>100	10.1%
Case 3: True Optimal Burn-in Time=10.0								
Bayes	8.6	3.4	9.9	6.0	10.2	2.8	14.1	0%
MLE	18.4	29.2	8.2	4.4	10.2	2.2	>100	9.9%
Trunsored	18.7	29.8	8.2	4.4	10.2	2.2	>100	10.9%
Case 4: True Optimal Burn-in Time=12.4								
Bayes	12.2	15.5	7.9	4.3	13.4	2.0	39.9	1.8%
MLE	27.6	38.3	6.2	2.8	33.1	0	>100	19.9%
Trunsored	28.3	38.9	6.2	2.8	35.6	0	>100	20.9%

Table 6: Optimal burn-in time using cost-based decision rule and commercial case

	Mean	STD	Median	Q1	Q3	5th percentile	95th percentile	Percentage of cases where cost still decreasing with tbi>100
Case 1: True Optimal Burn-in Time=2.3								
Bayes	2.6	1.1	2.6	1.8	3.2	1.2	4.9	0%
MLE	12.0	27.7	2.0	1.5	3.4	0.7	>100	7.8%
Trunsored	12.9	29.6	2.0	1.5	3.4	0.7	>100	9.9%
Case 2: True Optimal Burn-in Time=0.3								
Bayes	1.2	0.2	1.2	1.2	1.2	1.2	1.2	0%
MLE	1.1	0.2	1.2	1.2	1.2	0.7	1.2	0%
Trunsored	2.0	9.3	1.2	1.2	1.2	0.7	1.2	0.9%
Case 3: True Optimal Burn-in Time=2.3								
Bayes	2.6	1.1	2.5	1.9	3.2	1.2	4.9	0%
MLE	3.3	3.0	2.0	1.5	3.4	0.7	10.2	0%
Trunsored	6.3	17.4	2.0	1.5	3.4	0.7	10.2	3%
Case 4: True Optimal Burn-in Time=1.8								
Bayes	1.7	0.9	1.8	1.2	2.2	0	3.0	0%
MLE	17.5	35.4	1.4	0.8	3.1	0	>100	14.3%
Trunsored	18.5	36.8	1.4	0.8	3.2	0	>100	16.5%

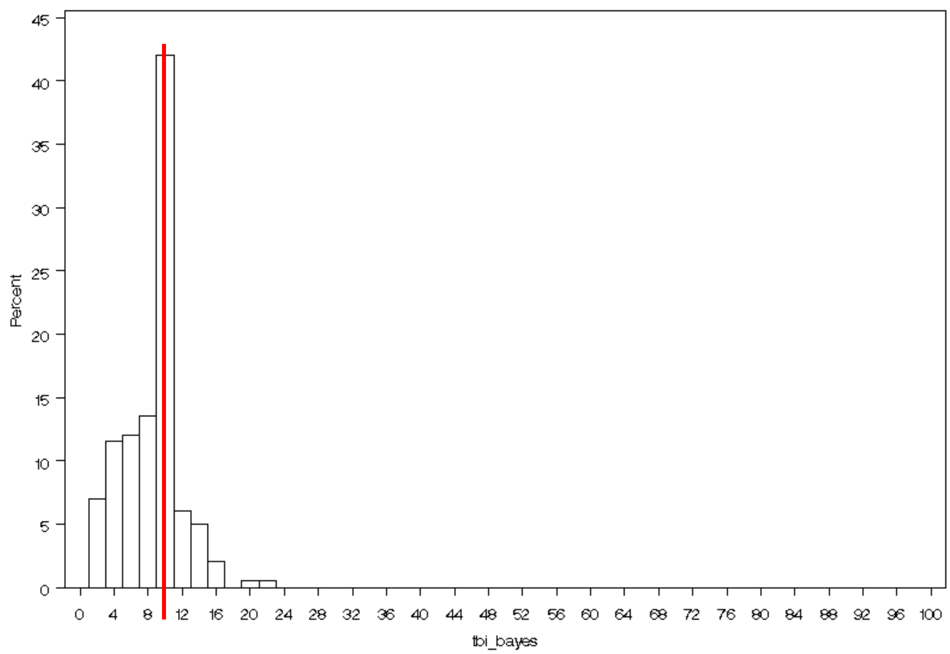


Figure 52 Distribution of optimal burn-in times for Case 3 using Bayesian estimation and satellite costs

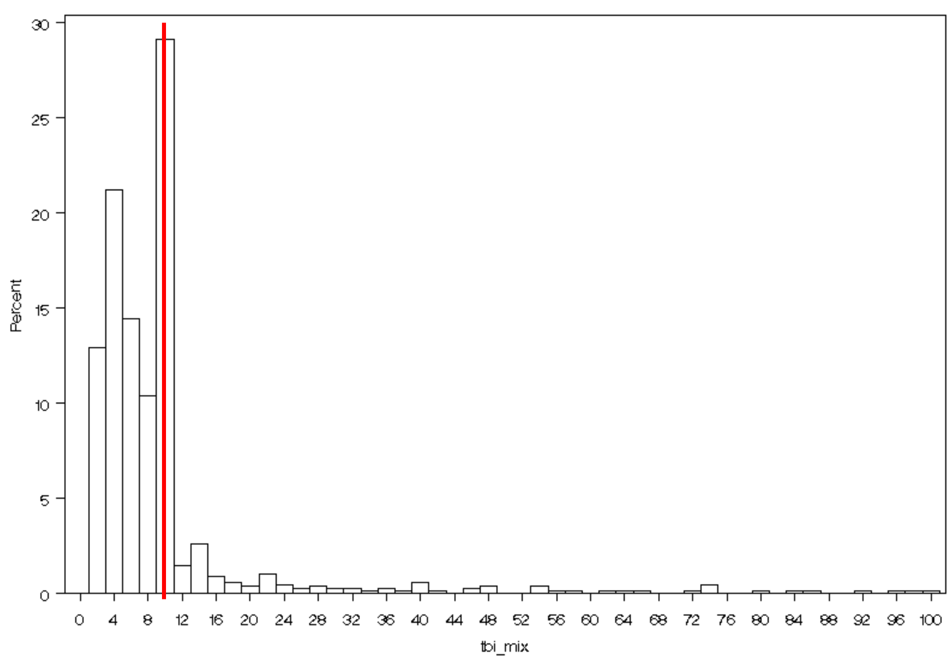


Figure 53. Distribution of optimal burn-in times for Case 3 using MLE and satellite costs

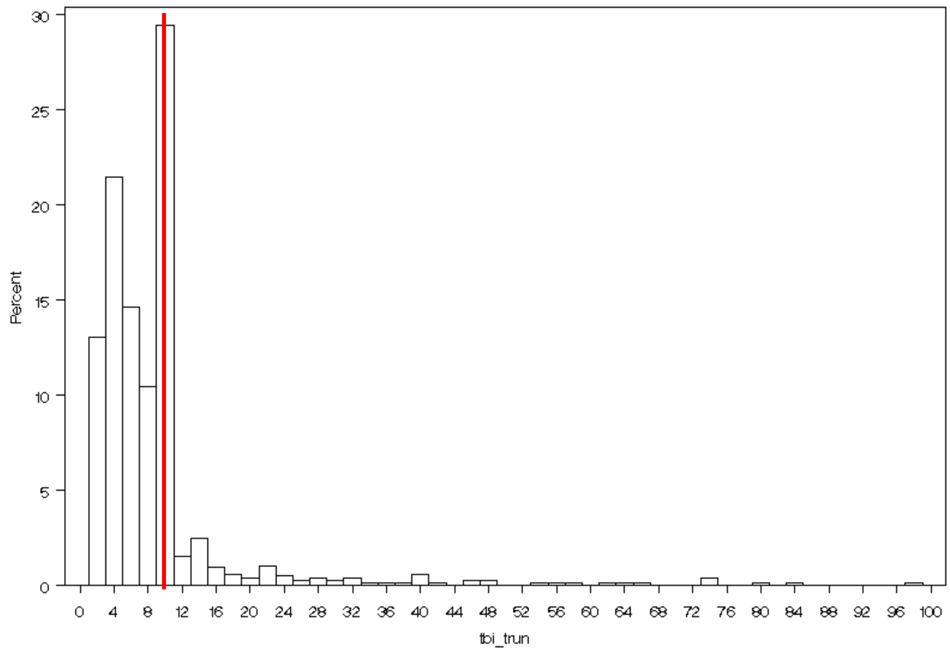


Figure 54. Distribution of optimal burn-in times for Case 3 using Trunsored estimation and satellite costs

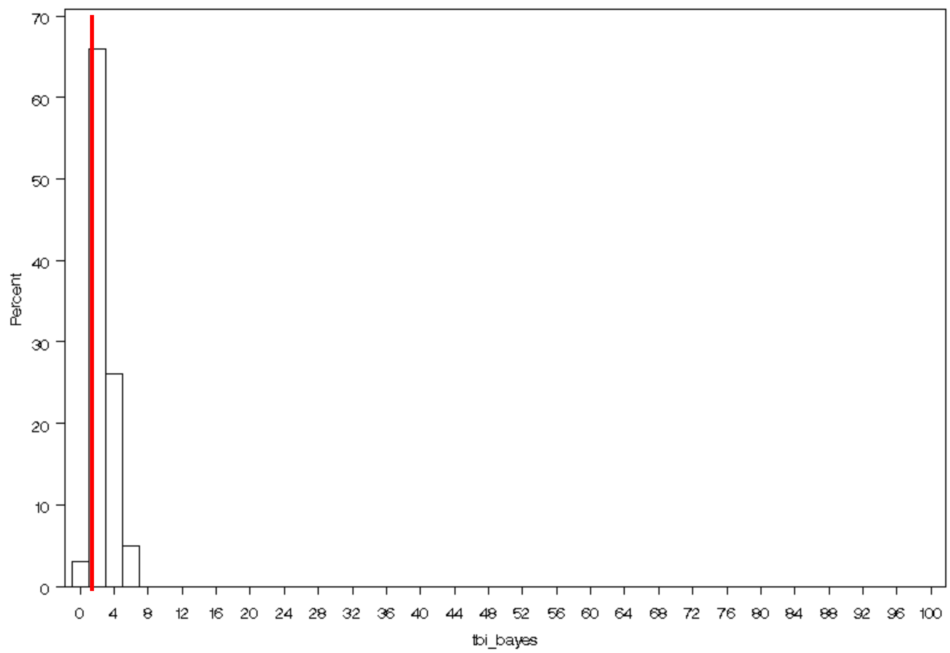


Figure 55 Distribution of optimal burn-in times for Case 3 using Bayesian estimation and commercial costs

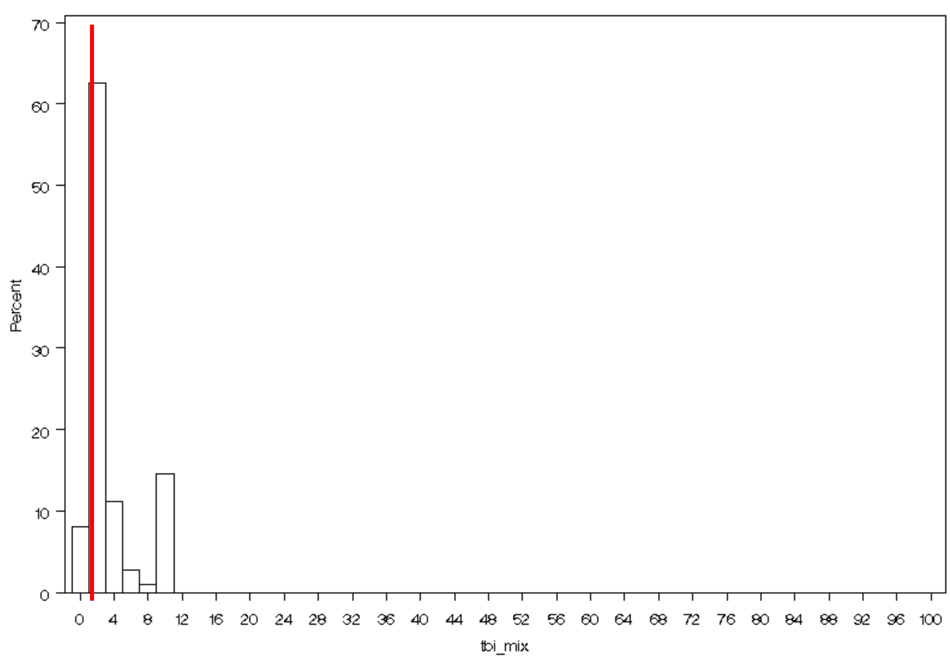


Figure 56. Distribution of optimal burn-in times for Case 3 using MLE and commercial costs

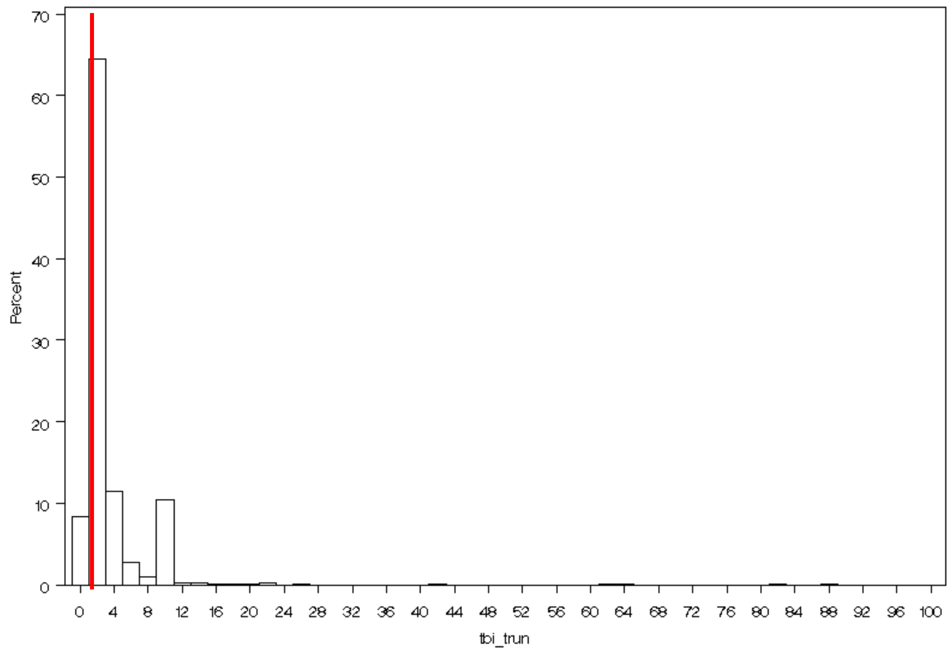


Figure 57. Distribution of optimal burn-in times for Case 3 using Trunsored estimation and commercial costs

Chapter Summary

As with the estimates of reliability examined in Chapters 4 through 6, the Bayesian method appears to perform the best in terms of predicting optimal burn-in time. However, for these decision rule cases, only small mixture ratios and only a reasonably accurate prior were examined. It is expected that, as with reliability estimation, when the mixture ratio is larger (and more failures observed) or when the prior is less accurate, the Bayesian method will not perform as well.

With both the end of life reliability requirement and cost-based decision rules, all methods performed reasonably well most of the time. The medians for the distributions of optimal burn-in times were quite accurate for all three estimation methods and all cases examined. However, the weaknesses in the MLE and Trunsored methods occur because of their skewed distributions. Occasionally, they will predict much longer burn-in times than are truly required. These can lead to substantial unnecessary costs. Unfortunately, there does not seem to be a pattern by which these cases can be identified based on the data observed.

Chapter 9: Conclusions, Contributions, and Future Work

Review of Findings and Contributions

The purpose of this dissertation was to developed recommended methodology for when a limited number of latent failures have been observed and a weak subpopulation is suspected. The circumstances under which the work in this dissertation is most likely to be of use is summarized in Figure 58. There must be early failures observed before all the product has been delivered. This would most likely be due to a limited burn-in or ESS test where early, unexpected failures have been observed. Of course, if non-destructive inspection methods are available, it would likely be the most effective technique. If the failure mechanism is well understood and can be quickly accelerated without consuming too much of the useful life of strong devices, the techniques may or may not be needed to confirm that all (or a sufficient percentage) of the parts have been removed from the population. If the failure mechanism is unknown or if a balance between screening and consuming useful life must be struck, the empirical methods such as those developed in this dissertation must be applied.

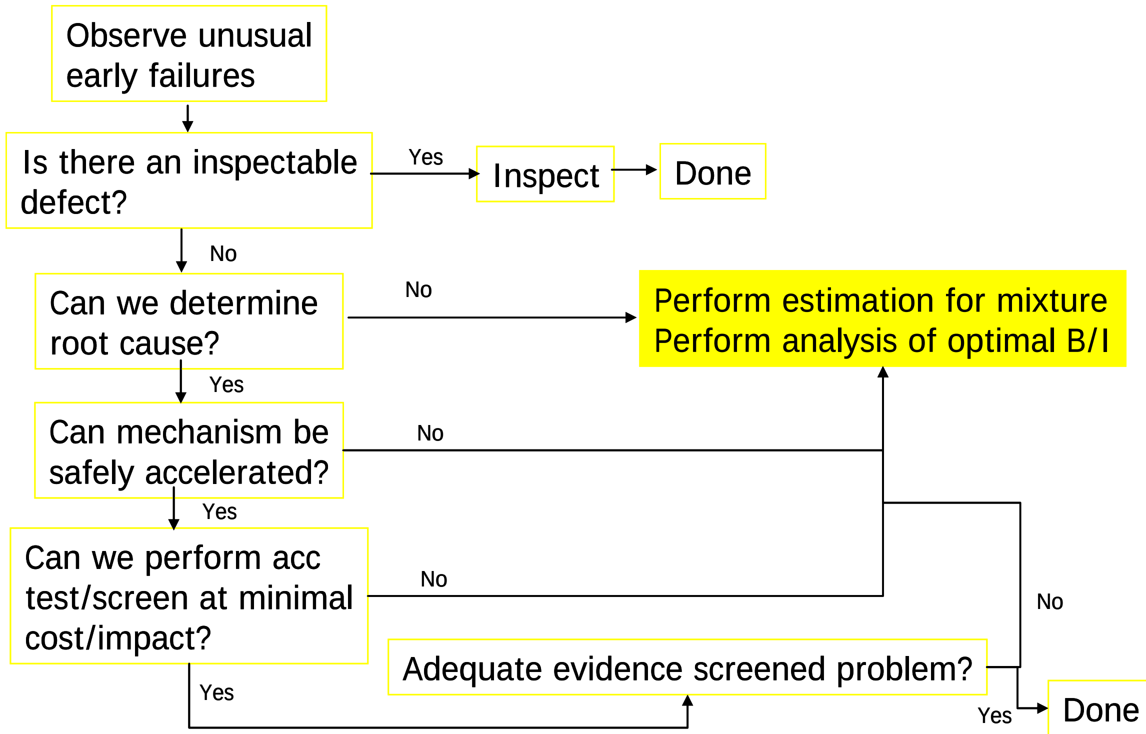


Figure 58. Summary of circumstances when empirical mixture estimation methods are needed.

This dissertation is the first work to extensively examine the estimation properties of Bayesian, MLE, and Trunsored estimation for mixed Weibull situations using Monte Carlo analysis. These results are summarized in Figure 59. It demonstrated problems with the Trunsored method that were not identified under the mixed exponential situations used to develop it [Hirose, 2005]. This dissertation developed a simplified technique to reduce the number of numeric integrations needed for using Bayesian estimation for mixed Weibulls, and showed the Bayesian technique to be robust, particularly when few failures are observed.

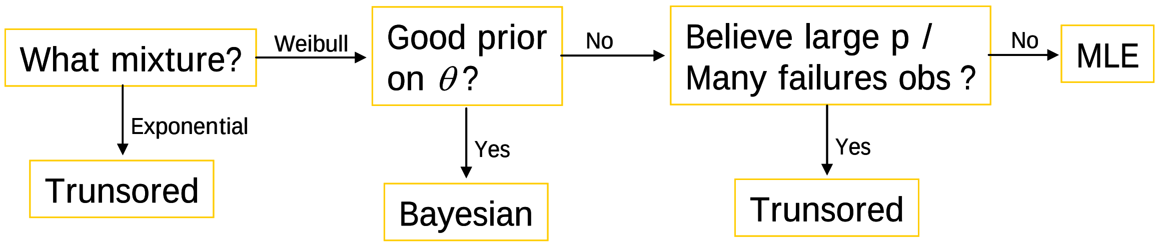


Figure 59. Summary of when each estimation method is most likely to be appropriate.

A new cost model for life cycles with possible burn-in was also developed. The new model adds the impacts of schedule slippage and capacity that were not include in previous cost models. It was shown that these cost elements can change decisions about the optimal burn-in length and whether a burn-in is recommended. The use of the new model, which realistically takes into account that there are costs to slipping schedule, significantly shortens the recommended burn-in time and reflects a much larger cost. Previous models would have made an incorrect recommendation.

Finally, errors in the estimation methods were examined with respect to decision making about optimal burn-in lengths. Two decision rules, one where a minimum end of life reliability was required and one based on minimizing life cycle costs, were examined. In both cases, it was discovered that the MLE and Trunsored methods had skewed distributions, occasionally recommending unrealistically long burn-ins. However, the median and first and third quartiles for all methods produced reasonable estimates of the optimal burn-in time.

These results can help practitioners make better decisions about how to estimate parameters and design burn-ins when mixed Weibull distributions are suspected.

Future Work

There are several areas of possible future work. These are natural extensions of the work developed here, but would allow it to be more broadly applicable in practical settings.

Additional Case Studies

Although a variety of true parameters were examined as case studies in this dissertation and many results can be generalized, there are other case studies that would help increase the understanding of how the estimation methods behave. In particular, most case studies presented here focus on situations where the characteristic life was slightly shorter than the censor time (in most cases $\theta = 1$ and $t_{censor} = 2$). It would be useful to confirm that the estimation behaviors do not change when the characteristic life is longer than the censor time.

A few cases were examined where the population size was larger, and thus more failures were observed even when the mixture ratio p was small (see **Chapter 5: Trunsored Estimation**). However, these were only generated for a cases with small mixture ratio. More extensive characterization of the impact of population size, and number of observed failures, would be valuable.

Finally, this dissertation focused on limited failure population cases, where the strong subpopulation was assumed to have a reliability of 1 for all times of interest.

Although this is a reasonable assumption in many applications, where the parts have already been qualified to have a low failure rate during normal operating life, there may be cases where some probability of failure should be considered for the strong subpopulation. This is particularly important to consider in the cost model, where "eating into" useful life can have a large impact on total life-cycle costs.

Confidence Bounds

Although the distribution of the estimates and quantiles were examined, methods for estimating confidence bounds were not developed. Most textbooks on reliability analysis present only the asymptotic-normal theory for calculation of confidence bounds on Weibull parameters, these bounds are only accurate for large sample situations. They can be too small for situations where few failures were observed and their asymmetric properties can create problems when only a one-sided bound is desired [Meeker, 1987]. Methods using the Wald Statistic and Fisher Information have also been examined, but again may have problems in small sample situations [Jeng, 2001].

Another method is the use of the likelihood ratio statistic, which under asymptotic conditions is approximately distributed as a chi-squared random variable with one degree of freedom. Several studies have compared the asymptotic-normal method to the likelihood ratio method for single Weibull distribution confidence bounds and found the likelihood ratio method to have more accurate and better-behaved confidence intervals. A number of modifications to and correction factors for the likelihood ratio method have also been examined [Vander Wiel, 1990] [Doganaksoy, 1991] [Doganaksoy, 1993].

More limited work exists on confidence bounds for mixed distributions. Some work has been done in the theoretical statistics community, but as with the estimation of the parameters, these methods may not be appropriate under real-life conditions [Qin, 1999]. Confidence bounds for the situation where the strong subpopulation can be ignored (because its mean time to failure is well beyond mission life) were examined in [Meeker, 1987], but only for considering the time of first failure. In all other cases found, it is assumed that the likelihood ratio statistic is the appropriate method for confidence bound calculations, as in [Meeker, 1987] and [Hirose, 2005]. Comparisons of the likelihood ratio statistic to other methods of confidence bound estimation do not appear to have been explored.

It had been planned to look at methods such as likelihood ratio bounds for the MLE and Trunsored methods. However, the only methods I was able to develop for calculating the bounds were very computationally intense, requiring grid searches for every case. This method for finding the confidence bounds was not feasible for the extensive simulations performed here under the time constraints of this dissertation. In order to examine confidence bound estimation, new more efficient code will need to be developed.

Role of Accelerated Testing and Changing Life Conditions

Although, as discussed in the introduction, if only the screen conditions are accelerated, performing an accelerated test is equivalent to performing a longer test and is therefore addressed in this dissertation. However, there are times when both the screen and use applications are adjusted, for example if a part is used in two applications with different temperatures and burn-in is performed at 1.5 times the application temperature. Say it is desirable to establish a burn-in length for all applications. It is necessary to know whether the application with the smaller or larger acceleration should be used as “worst case” (i.e., the application which requires the longer burn-in to establish a given reliability or cost).

Because the optimal burn-in time is related to conditional reliability, a ratio is of interest. Obviously, at higher acceleration, the reliability at a given time is lower. Under a standard Weibull, at higher acceleration, the conditional reliability is also lower, as shown in Figure 60. However, under a mixed Weibull, the ratio is no longer monotonic, as shown in Figure 61 and Figure 62. Depending on the time point of interest, the shape of the conditional reliability curve relative to acceleration factor may change. Sometimes higher acceleration factors may produce lower conditional reliability and sometimes lower. This means that establishing a burn-in screen for some “worst case” acceleration condition may be difficult because it may not be obvious what the worst case condition is.

The behavior under acceleration, when both the application and test conditions are adjusted, seems counterintuitive. The behavior is related to derivatives of logarithms of reliability functions, not a space most practitioners are used to thinking of. Because the behavior is counterintuitive, more exploration of conditional reliabilities of mixed Weibull distributions under acceleration is needed.

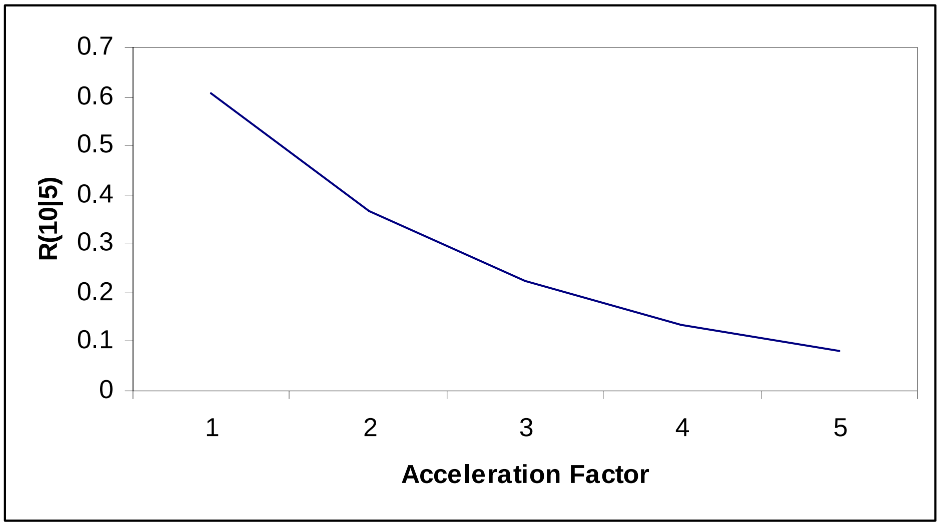


Figure 60. The impact of acceleration on conditional reliability for a standard Weibull (mixture ratio=1). The screen (of length 5) and time point of interest (10) were accelerated using the same factor. Higher acceleration produces lower conditional reliability.

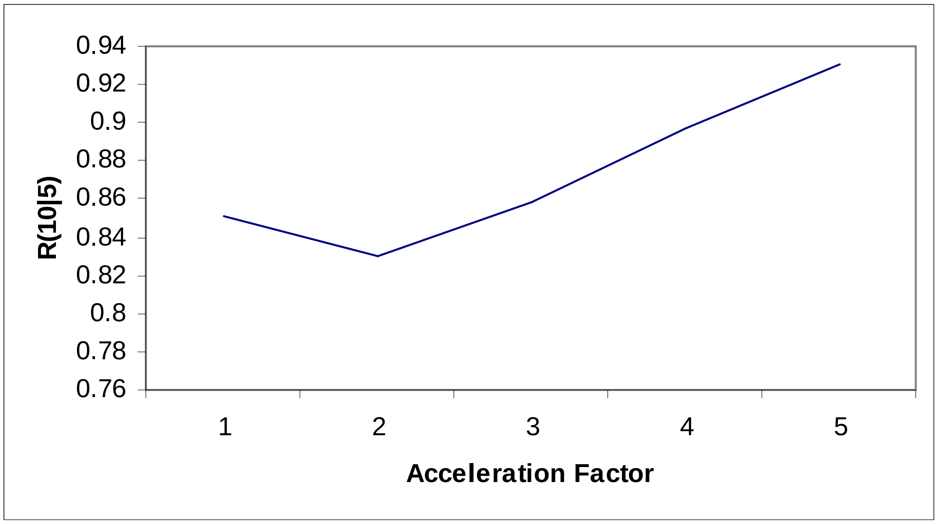


Figure 61. The impact of acceleration on conditional reliability for a mixed Weibull (mixture ratio=0.5). The screen (of length 5) and time point of interest (10) were accelerated using the same factor. Changing the acceleration factor no longer produces a monotonic behavior.

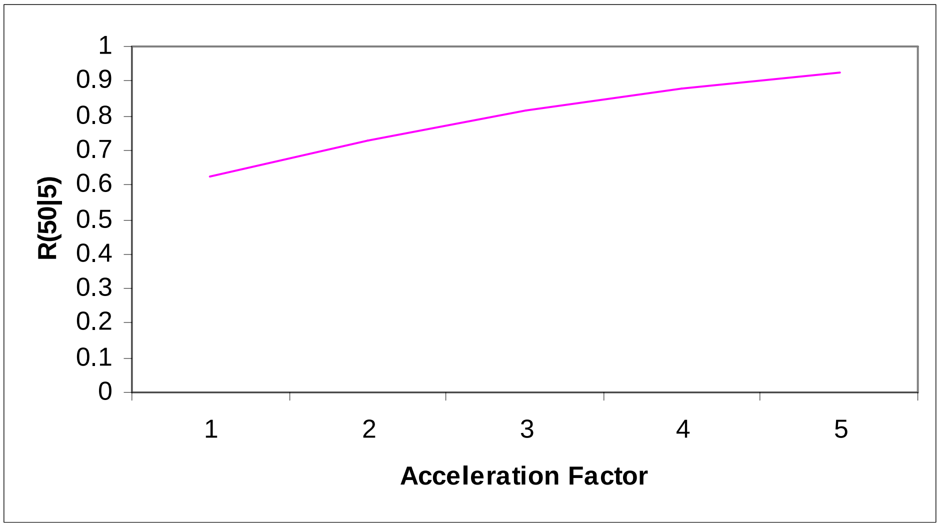


Figure 62. The impact of acceleration on conditional reliability for a mixed Weibull (mixture ratio=0.5). The screen (of length 5) and time point of interest (50) were accelerated using the same factor.

<u>Quick rules for developing burn-in lengths</u>

The work in this dissertation helps establish the appropriate use of different estimation methods and development of optimal screen lengths. However, in many cases, executives will want a quick "rule of thumb" regarding the expected recommended screen length. It is expected that a such a rule could be developed, based on looking at the amount of scatter in the failure times observed to that point. However, such a rule would likely depend on uncertainty in the estimates and confidence bounds, which have not yet been explored.

Appendix 1: Median Rank Regression

Another class of estimations, based on median rank regression, was also examined as part of this work. However, the problems encountered with the method were so extensive that they were not pursued fully. This appendix summarizes some of the problems discovered.

Median Rank Regression (MRR) is commonly used in engineering studies for estimating Weibull parameters. Partially, this is due to the relationship between MRR and the plotting of data on Weibull paper; the line and coefficients typically derived from Weibull paper correspond to those derived from MRR estimation. Furthermore, some studies have shown MRR to be better, in terms of bias and error, than MLE, particularly when sample sizes are small. [Abernethy, 1999] [Mardia, 1999] Others recommend MRR because of the convergence problems and other computational problems with MLE. [Kececioglu, 1998]. Several industry-focused reliability courses and software packages state that MRR is "recommended as the engineering standard" for Weibull analyses of small samples. [Abernathy, 2000] [Amari, 2008]

MRR in a non-mixture situation begins with the calculation of the adjusted rank:

$$O(t_i) = \frac{rr_i \times O(t_{i-1}) + (n+1)}{rr_i + 1}$$

where $rr_i =$ reverse rank of the i^{th} observation

$n =$ number of devices on test

The adjusted rank is then used to calculate the median rank:

$$MR(t_i) = \frac{O(t_i) - 0.3}{n + 0.4}$$

Failure times and median ranks are transformed as:

$$Y_i = \ln(t_i)$$

$$X_i = \ln\left(\ln\left(\frac{1}{1 - MR(t_i)}\right)\right)$$

Weibull parameters are found by performing a basic linear regression of *Y* on *X*. The shape parameter is the slope of the regression. The scale parameter is found by:

$$\theta = \exp\left(\frac{\text{int}}{\beta}\right)$$

A variants of MRR for mixed Weibull situation, based on the work in [Kececioglu, 1998] was considered.

[Kececioglu, 1998] uses all the data and allows for estimation of parameters from both the weak and strong subpopulations. The concept involves assigning a portion of each failure and censored point to each of the subpopulations, based on a Bayesian estimate of the likelihood of the outcome occurring in the subpopulation. First, parameters are postulated for the two subpopulations, either through prior knowledge or from graphical estimation.

Then each failure time is split into the two subpopulations. The proportion of the failure time in the weak subpopulation is:

$$P_1(t_i) = \frac{pf_1(t_i)}{pf_1(t_i) + (1-p)\, f_2(t_i)}$$

The proportion in the strong subpopulation of the failure times in the strong subpopulation are chosen similarly.

The two subpopulations are then treated completely separately and MRR applied to each. Because the number of failures is no longer integral, the calculation of the adjusted rank needs to be changed. [Kececioglu, 1998] presents the calculations for situations with no censored observations. He calculations the adjusted rank (or mean order number) as:[23]

$$O_j(t_i) = \sum_{k=1}^{i} P_j(t_k)$$

where j = 1 or 2 (for each of the subpopulations).

The median ranks are calculated similarly to the single population case:

[23] I'd like to add some theoretical discussion about how this compares to the typical calculation of adjusted rank. However, I've been really struggling with this theory. I'm going to leave it alone for now and see if it is worth going back to.

$$MR_j(t_i) = \frac{O_j(t_i)}{O_j(t_N) + 0.4}$$

This is very similar to the general calculation of the median rank, but is lacking the -0.3 in the numerator. This is likely because in many cases, when only a portion of the failure is assumed to fall in each population, $O_j(t_i)$ is less than 0.3, creating a negative median rank. Kececioglu does not discuss this change to the median rank formulation.

Then the regression proceeds normally. Once new Weibull parameter estimates have been established, the proportion of the failures in each subpopulation can be adjusted. This in turn leads to new parameter estimates. The new mixture ratio is estimated as:

$$\hat{p} = \frac{MON_1(t_N)}{N} = \frac{1}{N}\sum_{i=1}^{N} P_1(t_i)$$

[Kececioglu, 1998] uses the sum of the correlation coefficients for each subpopulation as the objective function to be maximized. [Kececioglu, 1998] proposes that simply iterating through this system will eventually converge to the values that maximize the sum of the correlation coefficients, but he does not prove that the iterations converge, much less to the maximal value.

Furthermore, the solution to his example at the end of Section 2 neither maximizes the sum of the correlation coefficients nor appears to be the values to which the iterations converge. The example has failure times for 10 observations (3.0, 28.5, 71.6, 91.1, 129.1, 157.8, 188.9, 226.1, 278.0, 367.2). [Kececioglu, 1998] suggests starting values of: $\hat{p} = 0.3, \hat{\beta}_1 = 0.6, \hat{\theta}_1 = 39.4, \hat{\beta}_2 = 2.3, \hat{\theta}_2 = 233.5$. He then reports optimum values for the estimation method as: $\hat{p} = 0.3, \hat{\beta}_1 = 0.5, \hat{\theta}_1 = 50, \hat{\beta}_2 = 1.9, \hat{\theta}_2 = 193$. These produce an total correlation coefficient of $0.972 + 0.981 = 1.953$. Using his same starting values and the iterative methodology described in the paper, the very first iteration produces estimates of $\hat{p} = 0.30, \hat{\beta}_1 = 0.38, \hat{\theta}_1 = 44.25, \hat{\beta}_2 = 1.91, \hat{\theta}_2 = 209.73$, which have a total correlation coefficient of 1.996, higher than the optimal value reported in the paper. Furthermore, subsequent iteration produce lower total correlation coefficients, and the exact value reported in the paper could not be reproduced. When the paper's basic equations are run through a standard non-linear programming routine, the coefficients obtained are $\hat{p} = 0.998, \hat{\beta}_1 = 0.4, \hat{\theta}_1 = 1.31, \hat{\beta}_2 = 0.74, \hat{\theta}_2 = 83.54$ with a total correlation coefficient of 1.999.

Several modifications to [Kececioglu, 1998] are needed to allow it to be used in the general cases. For censored observations, $P_j\left(t_k\right)$ is not known. Instead, there is only the probability of the device surviving to the censor time. To allow those to contribute to the adjusted rank, the reliability at the censor time for any censored devices prior to t_k is included.

$$O_j\left(t_i\right)=\sum_{k=1}^{i} P_j\left(t_k\right)+\sum I_i \times R_j\left(t_m\right)$$

where I_i = indicator variable that censor time t_m is less than the failure time t_i

Since in our simulation, the censoring only occurs after the failure times, the modification is only needed for $O_j\left(t_N\right)$.

$$O_j\left(t_N\right)=\sum_{k=1}^{i} P_j\left(t_k\right)+u \times \frac{p_j \times R_j\left(t_{cens}\right)}{p_1 \times R_1\left(t_{cens}\right)+p_2 \times R_2\left(t_{cens}\right)}$$

where u is the number of devices censored at t_{cens}.

In addition, to prevent having to perform an iterative scheme by hand, the optimization was performed using a standard non-linear programming routine. In this routine, $\hat{p}$ was not calculated as in Kececioglu. Instead it was simply entered as another parameter need to be estimated in the maximization.

Using either this modified $\hat{p}$ or the original proposed by Kececioglu, the errors from MRR were large. Kececioglu MRR had the highest random error of any of the methods in its estimates of α. It also had large bias on α unless the true mixture ratio was between about 0.5 and 0.75, and the bias increased when many failures were observed, whether from a higher α, higher shape parameter or a delayed censor time. The errors in the mixture ratio estimate lead to large errors in the reliability estimates as well, producing biases and random errors each over 5 times larger than the MLE method under almost every condition examined, as shown in Figure 63. Given these large errors, attempts at performing the MRR method were abandoned.

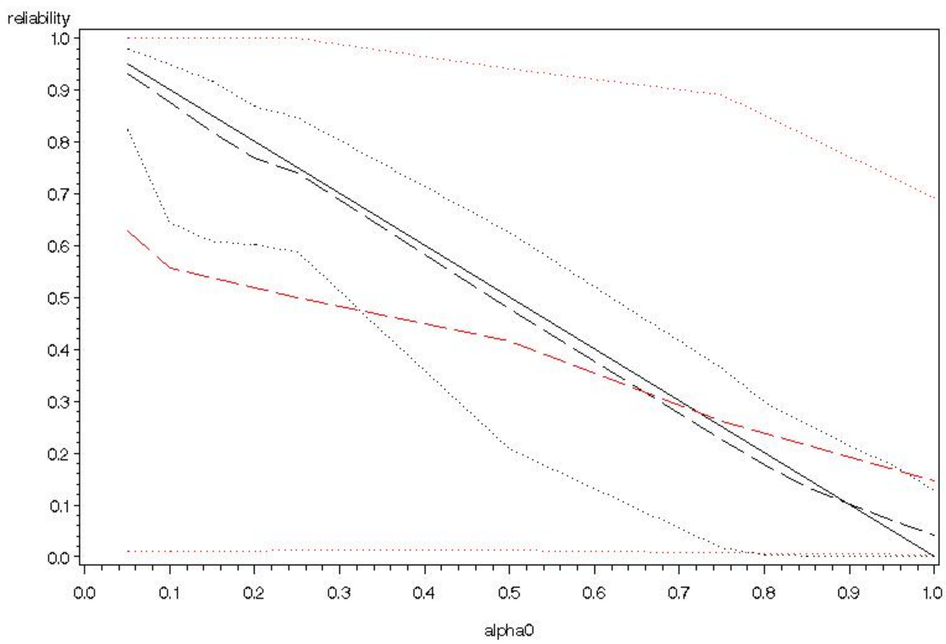


Figure 63. Mixture is Black; Kececioglu MRR is red. Time=10.

Table 7. Kececioglu MRR: Errors in Parameters and Reliability Estimates for θ=1, β=1, censor time=2.

True α	Bias in a (SD)	Bias in θ (SD)	Bias in β (SD)	Bias in R(20)[24] (SD)
0.05	0.33 (0.46)	0.11 (0.18)	-0.13 (0.21)	-0.32 (0.45)
0.1	0.42 (0.47)	0.18 (0.21)	-0.18 (0.21)	-0.42 (0.47)
0.25	0.32 (0.41)	0.23 (0.20)	-0.21 (0.18)	-0.31 (0.41)
0.5	0.11 (0.36)	0.28 (0.20)	-0.19 (0.13)	-0.11 (0.36)
0.75	0.29 (0.30)	0.29 (0.14)	-0.22 (0.10)	-0.01 (0.30)
0.9	-0.04 (0.23)	0.29 (0.10)	-0.21 (0.08)	0.04 (0.23)
1	-0.10 (0.19)	0.26 (0.09)	-0.19 (0.06)	0.10 (0.19)

[24] 10 times longer than the censor time. In truth all the at risk subpopulation has failed by this point, so R(20)=1-α.

Appendix 2: Estimation Errors

Table 8. Errors in Parameters and Reliability Estimates for θ=1, β=1, censor time=2[25].

True p	Estimation Method	Bias in p (SD)	Bias in θ (SD)	Bias in β (SD)	Bias in R(10)[26] (SD)
0.05	MLE	0.12 (0.30)	42.68 (190.63)	0.72 (1.82)	-0.03 (0.09)
	Trunsored	2.61 (18.48)	110.70 (354.19)	0.72 (1.82)	-0.08 (1.11)
	Bayesian	0.01 (0.02)	-0.10 (0.45)	0.25 (0.26)	-0.01 (0.02)
0.1	MLE	0.10 (0.27)	14.73 (77.77)	0.38 (1.02)	-0.02 (0.08)
	Trunsored	0.80 (0.29)	247.88 (1351.60)	-0.04 (1.00)	-0.15 (0.13)
	Bayesian	0.00 (0.04)	-0.04 (0.45)	0.24 (0.26)	0.00 (0.04)
0.25	MLE	0.06 (0.18)	1.05 (4.51)	0.08 (0.28)	-0.03 (0.10)
	Trunsored	0.73 (6.69)	42.84 (310.22)	0.08 (0.28)	-0.03 (0.11)
	Bayesian	0.01 (0.06)	0.04 (0.37)	0.14 (0.22)	-0.01 (0.06)
0.5	MLE	0.04 (0.15)	0.29 (1.00)	0.03 (0.19)	-0.03 (0.12)
	Trunsored	0.55 (4.49)	25.64 (311.80)	0.03 (0.19)	-0.04 (0.17)
	Bayesian	0.02 (0.08)	0.05 (0.30)	0.07 (0.17)	-0.02 (0.08)
0.75	MLE	0.01 (0.11)	0.06 (0.33)	0.03 (0.15)	-0.01 (0.10)
	Trunsored	0.45 (10.23)	9.40 (183.23)	0.03 (0.16)	-0.03 (0.16)
	Bayesian	0.00 (0.07)	-0.01 (0.19)	0.06 (0.14)	0.00 (0.07)
0.9	MLE	0.00 (0.07)	0.01 (0.21)	0.03 (0.12)	0.00 (0.08)
	Trunsored	0.21 (4.38)	3.31 (89.99)	0.02 (0.14)	-0.08 (0.17)
	Bayesian	-0.03 (0.05)	-0.06 (0.14)	0.05 (0.11)	0.03 (0.05)
1	MLE	-0.03 (0.04)	-0.06 (0.13)	0.05 (0.11)	0.03 (0.04)
	Trunsored	0.11 (1.98)	1.39 (39.30)	0.01 (0.13)	-0.04 (0.16)
	Bayesian	-0.06 (0.03)	-0.13 (0.10)	0.09 (0.10)	0.06 (0.03)

---

[25] Cases with <3 failures deleted

[26] 5 times longer than the censor time.

Table 9. After Likelihood Ratio Test on *p*: Errors in Parameters and Reliability Estimates for θ=1, β=1, censor time=2. (Compare to Table 8)[27]

True p	Estimation Method	Bias in p (SD)	Bias in θ (SD)	Bias in β (SD)	Bias in R(10)[28] (SD)
0.05	MLE	0.65 (0.44)	608.13 (2460.29)	0.41 (1.75)	-0.09 (0.12)
	Trunsored	0.86 (0.28)	902.77 (3275.47)	0.11 (1.64)	-0.12 (0.12)
0.1	MLE	0.59 (0.43)	205.43 (1349.25)	0.17 (1.07)	-0.11 (0.14)
	Trunsored	0.80 (0.29)	255.94 (1371.63)	-0.04 (0.99)	-0.15 (0.13)
0.25	MLE	0.42 (0.38)	9.31 (12.50)	-0.04 (0.33)	-0.16 (0.18)
	Trunsored	0.63 (0.28)	13.90 (12.20)	-0.17 (0.29)	-0.24 (0.15)
0.5	MLE	0.23 (0.27)	1.58 (1.94)	-0.04 (0.23)	-0.16 (0.21)
	Trunsored	0.39 (0.22)	2.62 (1.79)	-0.14 (0.22)	-0.27 (0.18)
0.75	MLE	0.12 (0.16)	0.43 (0.58)	-0.03 (0.18)	-0.11 (0.15)
	Trunsored	0.21 (0.11)	0.74 (0.46)	-0.10 (0.15)	-0.20 (0.10)
0.9	MLE	0.06 (0.08)	0.17 (0.24)	-0.02 (0.13)	-0.06 (0.08)
	Trunsored	0.23 (0.10)	3.07 (89.53)	-0.05 (0.10)	-0.09 (0.06)
1	MLE	0.00 (0.03)	-0.01 (0.12)	0.02 (0.10)	-0.01 (0.03)
	Trunsored	0.07 (1.97)	1.26 (39.30)	0.01 (0.09)	0.00 (0.08)

[27] Testing for presence of mixture not performed for Bayesian method
[28] 5 times longer than the censor time.

Table 10. Errors in Parameters and Reliability Estimates for θ=1.6, β=3, censor time=2.

True p	Estimation Method	Bias in p (SD)	Bias in θ (SD)	Bias in β (SD)	Bias in R(10)[29] (SD)
0.05	MLE	0.13 (0.31)	1.17 (4.22)	2.04 (5.27)	-0.09 (0.24)
	Trunsored	12.45 (49.97)	20.04 (115.87)	2.04 (5.27)	-0.89 (6.97)
	Bayesian	0.01 (0.03)	0.68 (0.87)	-0.22 (0.69)	-0.01 (0.03)
0.1	MLE	0.08 (0.26)	0.47 (1.78)	4.17 (98.24)	-0.08 (0.23)
	Trunsored	10.68 (53.87)	6.34 (40.65)	4.16 (98.24)	-0.45 (2.38)
	Bayesian	0.01 (0.04)	1.30 (0.94)	-0.19 (0.60)	-0.01 (0.04)
0.25	MLE	0.07 (0.21)	0.22 (0.73)	0.22 (0.82)	-0.07 (0.21)
	Trunsored	12.55 (73.89)	4.15 (37.32)	0.21 (0.83)	-0.43 (2.11)
	Bayesian	0.02 (0.06)	2.27 (1.33)	-0.07 (0.48)	-0.02 (0.06)
0.5	MLE	0.04 (0.15)	0.07 (0.29)	0.12 (0.58)	-0.04 (0.14)
	Trunsored	2.05 (27.03)	0.49 (4.87)	0.12 (0.59)	-0.16 (1.03)
	Bayesian	0.03 (0.09)	2.35 (0.95)	-0.12 (0.39)	-0.03 (0.09)
0.75	MLE	0.03 (0.11)	0.03 (0.17)	0.08 (0.44)	-0.03 (0.11)
	Trunsored	0.55 (8.48)	0.20 (2.35)	0.07 (0.46)	-0.12 (0.96)
	Bayesian	0.03 (0.08)	2.37 (0.79)	-0.10 (0.35)	-0.03 (0.08)
0.9	MLE	0.00 (0.08)	-0.01 (0.11)	0.10 (0.38)	0.00 (0.08)
	Trunsored	1.04 (17.10)	0.17 (2.03)	0.06 (0.42)	-0.14 (1.33)
	Bayesian	0.00 (0.05)	2.29 (0.75)	-0.04 (0.32)	0.00 (0.05)
1	MLE	-0.04 (0.05)	-0.04 (0.08)	0.15 (0.33)	0.04 (0.05)
	Trunsored	0.22 (3.87)	0.08 (0.90)	0.04 (0.40)	-0.10 (1.12)
	Bayesian	-0.05 (0.03)	2.24 (0.66)	0.12 (0.25)	0.05 (0.03)

[29] 5 times longer than the censor time.

Appendix 3: Detailed Calculations for Bayesian Posterior

[Perlstein, 2001] demonstrates that it is easier to manipulate the expanded form of the likelihood. First, some notation must be established.

Let $\sum \mathrm{t}_{\mathrm{m}}^{\ell}$ = the sum of the elements of the m$^{\text{th}}$ permutation of vectors of size ℓ from $\underline{t} = \{t_1, ..., t_n\}$. So for example, if $n = 4$ there are $\binom{4}{2} = 6$ permutations of vectors of size 2. So for $\ell = 2$, m takes on values from 1 to 6:

$$\sum_{\ell=1}^{4} t_1^{\ell} = t_1 + t_2 \quad \sum_{\ell=1}^{4} t_2^{\ell} = t_1 + t_3 \quad \sum_{\ell=1}^{4} t_3^{\ell} = t_1 + t_4$$
$$\sum_{\ell=1}^{4} t_4^{\ell} = t_2 + t_3 \quad \sum_{\ell=1}^{4} t_5^{\ell} = t_2 + t_4 \quad \sum_{\ell=1}^{4} t_6^{\ell} = t_3 + t_4$$

Similarly, $\prod \mathrm{t}_{\mathrm{m}}^{\ell}$ = the product of the elements of the m$^{\text{th}}$ permutation of vectors of size ℓ.

Using this notation, the likelihood equation can be rewritten as:

$$\begin{aligned} L(p, \theta_W, \theta_S, \beta_W, \beta_S) = \sum_{\ell=0}^{n} \sum_{m=1}^{\binom{n}{\ell}} \sum_{v=0}^{u} \Bigg\{ & \binom{u}{v} p^{\ell+v} (1-p)^{n-\ell+u-v} \\ & \times \left(\frac{\beta_W}{\theta_W}\right)^{\ell} \prod \left(t_m^{(\ell)}\right)^{\beta_W - 1} \exp\left(-\frac{\sum \left(t_m^{(\ell)}\right)^{\beta_W} + v t_*^{\beta_W}}{\theta_W}\right) \\ & \times \left(\frac{\beta_S}{\theta_S}\right)^{n-\ell} \prod \left(t_m^{(n-\ell)}\right)^{\beta_S - 1} \exp\left(-\frac{\sum \left(t_m^{(n-\ell)}\right)^{\beta_S} + (u-v) t_*^{\beta_S}}{\theta_S}\right) \end{aligned}$$

The posterior distribution is calculated using Bayes Theorem:

$$g\left(p,\theta_W,\theta_S,\beta_W,\beta_S \middle| Data\right)=\frac{g\left(p,\theta_W,\theta_S,\beta_W,\beta_S\right)\times L\left(p,\theta_W,\theta_S,\beta_W,\beta_S \middle| D\right)}{\iiiint\!\!\int g\left(p,\theta_W,\theta_S,\beta_W,\beta_S\right)\times L\left(p,\theta_W,\theta_S,\beta_W,\beta_S \middle| D\right)\mathrm{d}p\mathrm{d}\theta_W\mathrm{d}\theta_S\mathrm{d}\beta_W\mathrm{d}\beta_S}$$

$$=\frac{p^{\alpha-1}(1-p)^{\beta-1}\times\left(\frac{1}{\theta_W}\right)^{a_W+1}\exp\left(-\frac{b_W}{\theta_W}\right)\times\left(\frac{1}{\theta_S}\right)^{a_S+1}\exp\left(-\frac{b_S}{\theta_S}\right)\times L\left(p,\theta_W,\theta_S,\beta_W,\beta_S \middle| D\right)}{\iiiint\!\!\int p^{\alpha-1}(1-p)^{\beta-1}\times\left(\frac{1}{\theta_W}\right)^{a_W+1}\exp\left(-\frac{b_W}{\theta_W}\right)\times\left(\frac{1}{\theta_S}\right)^{a_S+1}\exp\left(-\frac{b_S}{\theta_S}\right)\times L\left(p,\theta_W,\theta_S,\beta_W,\beta_S \middle| D\right)\mathrm{d}p\mathrm{d}\theta_W\mathrm{d}\theta_S\mathrm{d}\beta_W\mathrm{d}\beta_S}$$

Let us start by considering the integral in the denominator.

$$\iiiint\!\!\int p^{\alpha-1}(1-p)^{\beta-1}\times\left(\frac{1}{\theta_W}\right)^{a_W+1}\exp\left(-\frac{b_W}{\theta_W}\right)\times\left(\frac{1}{\theta_S}\right)^{a_S+1}\exp\left(-\frac{b_S}{\theta_S}\right)\times L\left(p,\theta_W,\theta_S,\beta_W,\beta_S \middle| D\right)\mathrm{d}\beta_S\,\mathrm{d}\beta_W\,\mathrm{d}\theta_W\,\mathrm{d}\theta_S\,\mathrm{d}p$$

$$=\sum_{\ell=0}^{n}\sum_{m=1}^{\binom{n}{\ell}}\sum_{v=0}^{u}\iiiint \left\{\binom{u}{v}p^{a+\ell+v-1}(1-p)^{\beta+n-\ell+u-v-1}\right.$$

$$\times\left(\frac{1}{\theta_W}\right)^{a_W+\ell+1}\beta_W{}^{\ell}\prod\left(t_m^{(\ell)}\right)^{\beta_W-1}\exp\left(-\frac{b_W+\sum\left(t_m^{(\ell)}\right)^{\beta_W}+vt_*^{\beta_W}}{\theta_W}\right)$$

$$\left.\times\left(\frac{1}{\theta_S}\right)^{a_S+n-\ell+1}\beta_S{}^{n-\ell}\prod\left(t_m^{(n-\ell)}\right)^{\beta_S-1}\exp\left(-\frac{b_S+\sum\left(t_m^{(n-\ell)}\right)^{\beta_S}+(u-v)t_*^{\beta_S}}{\theta_S}\right)\right\}\ \mathrm{d}p\mathrm{d}\theta_W\mathrm{d}\theta_S\mathrm{d}\beta_S\mathrm{d}\beta_W$$

$$=\sum_{\ell=0}^{n}\sum_{m=1}^{\binom{n}{\ell}}\sum_{v=0}^{u}\left\{\binom{u}{v}\int_{\delta_{W1}}^{\delta_{W2}}\int_{\delta_{S1}}^{\delta_{S2}}\beta_S{}^{n-\ell}\prod\left(t_m^{(n-\ell)}\right)^{\beta_S-1}\times\beta_W{}^{\ell}\prod\left(t_m^{(\ell)}\right)^{\beta_W-1}\right.$$

$$\times\int_0^{\infty}\left(\frac{1}{\theta_W}\right)^{a_W+\ell+1}\exp\left(-\frac{b_W+\sum\left(t_m^{(\ell)}\right)^{\beta_W}+vt_*^{\beta_W}}{\theta_W}\right)\times\int_0^{\infty}\left(\frac{1}{\theta_S}\right)^{a_S+n-\ell+1}\exp\left(-\frac{b_S+\sum\left(t_m^{(n-\ell)}\right)^{\beta_S}+(u-v)t_*^{\beta_S}}{\theta_S}\right)$$

$$\left.\times\int_0^1 p^{a+\ell+v-1}(1-p)^{\beta+n-\ell+u-v-1}\right\}\ \mathrm{d}p\mathrm{d}\theta_W\,\mathrm{d}\theta_S\,\mathrm{d}\beta_S\,\mathrm{d}\beta_W$$

$$=\sum_{\ell=0}^{n}\sum_{m=1}^{\binom{n}{\ell}}\sum_{v=0}^{u}\left\{\binom{u}{v}\frac{\Gamma\left(\alpha^*\right)\Gamma\left(\beta^*\right)}{\Gamma\left(\alpha^*+\beta^*\right)}\times\Gamma\left(a_W^*\right)\times\Gamma\left(a_S^*\right)\right.$$

$$\left.\int_{\delta_{W1}}^{\delta_{W2}}\int_{\delta_{S1}}^{\delta_{S2}}\left(b_S^*\right)^{a_S^*}\ \beta_S{}^{n-\ell}\prod\left(t_m^{(n-\ell)}\right)^{\beta_S-1}\times\left(b_W^*\right)\beta_W{}^{\ell}\prod\left(t_m^{(\ell)}\right)^{\beta_W-1}\mathrm{d}\beta_S\,\mathrm{d}\beta_W\right\}$$

where

$$\alpha^* = \alpha + \ell + v \qquad \beta^* = \beta + n - \ell + u - v$$

$$a_S^* = a_S + n - \ell \qquad b_S^* = b_S + \sum \left(t_m^{(n-\ell)} \right)^{\beta_S} + (u - v) t_*^{\beta_S}$$

$$a_W^* = a_W + \ell \qquad b_W^* = b_W + \sum \left(t_m^{(\ell)} \right)^{\beta_W} + v t_*^{\beta_W}$$

The remaining integrals with respect to β_S and β_W do not have closed form solutions and must be numerically integrated. However, limiting the numeric integration to these two integrals, as opposed to numerically integrating over all five parameters, greatly simplifies the problem and decreases the computational time.

For simplicity, a constant for this denominator is defined:

$$C(n,u) = \sum_{\ell=0}^{n} \sum_{m=1}^{\binom{n}{\ell}} \sum_{v=0}^{u} \Bigg\{ \binom{u}{v} \frac{\Gamma(\alpha^*)\Gamma(\beta^*)}{\Gamma(\alpha^* + \beta^*)} \times \Gamma(a_W^*) \times \Gamma(a_S^*)$$

$$\int_{\delta_{W1}}^{\delta_{W2}} \int_{\delta_{S1}}^{\delta_{S2}} \left(b_S^*\right)^{a_S^*} \beta_S^{\,n-\ell} \prod \left(t_m^{(n-\ell)} \right)^{\beta_S - 1} \times \left(b_W^*\right)^{a_W^*} \beta_W^{\,\ell} \prod \left(t_m^{(\ell)} \right)^{\beta_W - 1} \mathrm{d}\beta_S \, \mathrm{d}\beta_W \Bigg\}$$

Using the joint posterior distribution, the posterior expected values of the parameters and the reliability can be calculated.

$$E\left(p\right)=\frac{1}{C\left(n,u\right)}\times$$
$$\sum_{\ell=0}^{n}\sum_{m=1}^{\binom{n}{\ell}}\sum_{v=0}^{u}\left\{\binom{u}{v}\frac{\Gamma\left(\alpha^{*}+1\right)\Gamma\left(\beta^{*}\right)}{\Gamma\left(\alpha^{*}+1+\beta^{*}\right)}\times\Gamma\left(a_{W}^{*}\right)\times\Gamma\left(a_{S}^{*}\right)\right.$$
$$\left.\int_{\delta_{W1}}^{\delta_{W2}}\int_{\delta_{S1}}^{\delta_{S2}}\left(b_{S}^{*}\right)^{a_{S}^{*}}\beta_{S}{}^{n-\ell}\prod\left(t_{m}^{(n-\ell)}\right)^{\beta_{S}-1}\times\left(b_{W}^{*}\right)\beta_{W}{}^{\ell}\prod\left(t_{m}^{(\ell)}\right)^{\beta_{W}-1}\mathrm{d}\beta_{S}\mathrm{d}\beta_{W}\right\}$$

$$E\left(\theta_{W}\right)=\frac{1}{C\left(n,u\right)}\times$$
$$\sum_{\ell=0}^{n}\sum_{m=1}^{\binom{n}{\ell}}\sum_{v=0}^{u}\left\{\binom{u}{v}\frac{\Gamma\left(\alpha^{*}\right)\Gamma\left(\beta^{*}\right)}{\Gamma\left(\alpha^{*}+\beta^{*}\right)}\times\Gamma\left(a_{W}^{*}-1\right)\times\Gamma\left(a_{S}^{*}\right)\right.$$
$$\left.\int_{\delta_{W1}}^{\delta_{W2}}\int_{\delta_{S1}}^{\delta_{S2}}\left(b_{S}^{*}\right)^{a_{S}^{*}}\beta_{S}{}^{n-\ell}\prod\left(t_{m}^{(n-\ell)}\right)^{\beta_{S}-1}\times\left(b_{W}^{*}\right)^{a_{W}^{*}-1}\beta_{W}{}^{\ell}\prod\left(t_{m}^{(\ell)}\right)^{\beta_{W}-1}\mathrm{d}\beta_{S}\mathrm{d}\beta_{W}\right\}$$

$$E\left(\theta_{S}\right)=\frac{1}{C\left(n,u\right)}\times$$
$$\sum_{\ell=0}^{n}\sum_{m=1}^{\binom{n}{\ell}}\sum_{v=0}^{u}\left\{\binom{u}{v}\frac{\Gamma\left(\alpha^{*}\right)\Gamma\left(\beta^{*}\right)}{\Gamma\left(\alpha^{*}+\beta^{*}\right)}\times\Gamma\left(a_{W}^{*}\right)\times\Gamma\left(a_{S}^{*}-1\right)\right.$$
$$\left.\int_{\delta_{W1}}^{\delta_{W2}}\int_{\delta_{S1}}^{\delta_{S2}}\left(b_{S}^{*}\right)^{a_{S}^{*}-1}\beta_{S}{}^{n-\ell}\prod\left(t_{m}^{(n-\ell)}\right)^{\beta_{S}-1}\times\left(b_{W}^{*}\right)^{a_{W}^{*}}\beta_{W}{}^{\ell}\prod\left(t_{m}^{(\ell)}\right)^{\beta_{W}-1}\mathrm{d}\beta_{S}\mathrm{d}\beta_{W}\right\}$$

$$E\left(\beta_{W}\right)=\frac{1}{C\left(n,u\right)}\times$$
$$\sum_{\ell=0}^{n}\sum_{m=1}^{\binom{n}{\ell}}\sum_{v=0}^{u}\left\{\binom{u}{v}\frac{\Gamma\left(\alpha^{*}\right)\Gamma\left(\beta^{*}\right)}{\Gamma\left(\alpha^{*}+\beta^{*}\right)}\times\Gamma\left(a_{W}^{*}\right)\times\Gamma\left(a_{S}^{*}\right)\right.$$
$$\left.\int_{\delta_{W1}}^{\delta_{W2}}\int_{\delta_{S1}}^{\delta_{S2}}\left(b_{S}^{*}\right)^{a_{S}^{*}}\beta_{S}{}^{n-\ell}\prod\left(t_{m}^{(n-\ell)}\right)^{\beta_{S}-1}\times\left(b_{W}^{*}\right)^{a_{W}^{*}}\beta_{W}{}^{\ell+1}\prod\left(t_{m}^{(\ell)}\right)^{\beta_{W}-1}\mathrm{d}\beta_{S}\mathrm{d}\beta_{W}\right\}$$

$$E\left(\beta_S\right)=\frac{1}{C\left(n,u\right)}\times$$

$$\sum_{\ell=0}^{n}\sum_{m=1}^{\binom{n}{\ell}}\sum_{v=0}^{u}\left\{\binom{u}{v}\frac{\Gamma\left(\alpha^*\right)\Gamma\left(\beta^*\right)}{\Gamma\left(\alpha^*+\beta^*\right)}\times\Gamma\left(a_W^*\right)\times\Gamma\left(a_S^*\right)\right.$$

$$\left.\int_{\delta_{W1}}^{\delta_{W2}}\int_{\delta_{S1}}^{\delta_{S2}}\left(b_S^*\right)^{a_S^*}\ \beta_S^{\ n-\ell+1}\prod\left(t_m^{(n-\ell)}\right)^{\beta_S-1}\times\left(b_W^*\right)^{a_W^*}\ \beta_W^{\ \ell}\prod\left(t_m^{(\ell)}\right)^{\beta_W-1}\mathrm{d}\beta_S\,\mathrm{d}\beta_W\right\}$$

Appendix 4: Boeing Copyright Assignment

The Boeing Company
2201 Seal Beach Boulevard
Seal Beach, CA 90740

BOEING

COPYRIGHT ASSIGNMENT

Date of Presentation:	November 17, 2008
Copyright Form To:	ProQuest-CSA LLC, through its UMI® Dissertation Publishing business (ProQuest/UMI) ("Publisher")
Publication Title:	Latent Failures and Mixed Distributions: Using Mixed Distributions and Cost Modeling to Optimize the Management of Systems with Weak Latent Defect Subpopulations - Dissertation ("Work")
Author:	Anduin E. Touw (The Boeing Company)

The author of the above-referenced Work is an employee of The Boeing Company ("Boeing") preparing the Work within the scope of his or her employment. The Boeing employee in question is not authorized to assign or license Boeing's rights therein, or otherwise bind Boeing. However, subject to the limitations set forth below, Boeing is willing to assign its copyright in the Work to Publisher (the "Assignee").

Notwithstanding any assignment or transfer to the Assignee, or any other terms of this agreement, the rights granted by Boeing to Assignee are limited as follows: (i) any rights granted by Boeing to the Assignee are limited to the work-made-for-hire rights Boeing enjoys in the Work; (ii) Boeing makes no representation or warranty of any kind to the Assignee or any other person or business entity regarding the Work, the information contained therein, or any related copyright; and (iii) Boeing retains a non-exclusive, perpetual, worldwide, royalty-free right, without restriction or limitation, to use, reproduce, publicly distribute, display, and perform and make derivative works from the Work, and to license or permit others to do so.

Carol Smith

Carol Smith
Contract Administrator
Intellectual Property Management
Boeing Management Company

References

R.B. Abernethy. *The New Weibull Handbook*. Gulf Publishing Company, Houston, TX. 1999.

A. Alani, C. Dislis, and I. Jalowiecki, “Burn-in Economics Model for Multi-chip Modules,” *Electronics Letters*, Vol. 32, No. 25, 1996. pp. 2349-2351.

S. Amari, “Weibull Parameter Estimation Methods,”

www.relex.com/resources/art/art_weibull4.asp

O.E. Barndorff-Nielsen and D.R. Cox, “Bartlett Adjustments to the Likelihood Ratio Statistic and the Distribution of the Maximum Likelihood Estimator,” *Journal of the Royal Statistical Society Series B*, Vol. 46, No. 3, 1984. pp. 483-495.

D. Bearden, “Small Satellite Costs,” *Crosslink*, The Aerospace Corporation, Winter 2001.

G. Canavos and C. Tsokos. “Bayesian Estimation of Life Parameters in the Weibull Distribution,” *Operations Research*, Vol. 21 1973, pp. 755-763.

H.A. Chan, “A Formulation to Optimize Stress Testing,” *IEEE 44th Electronic Components and Technology Conference*, 1994. pp. 1020-1027.

V. Chan and W.Q. Meeker, “A Failure-Time Model for Infant Mortality and Wearout Failure Modes,” *IEEE Transactions on Reliability*, Vol. 48, No. 4, 1999. pp 377-387.

S.W. Cheng, J.C. Fu, “Estimation of mixed Weibull parameters in life testing”, *IEEE Trans. Reliability*, Vol R-31, 1982. pp 377-381.

D. Chi and W. Kuo, "Burn-in Optimization under Reliability & Capacity Restrictions," *IEEE Transactions on Reliability*, Vol. 38, No. 2, June 1989. pp. 193-198.

G. W. Cran, "Graphical Estimation Methods for Weibull Distributions," *Microelectronics and Reliability*, Vol. 15, 1976. pp. 47-52.

N. Doganaksoy, "Interval Estimation from Censored and Masked System-Failure Data," *IEEE Transactions on Reliability*, Vol. 40, No. 3, 1991. pp. 280-286.

N. Doganaksoy and J. Schmee, "Comparisons of Approximate Confidence Intervals for Distributions Used in Life-Data Analysis," *Technometrics*, Vol. 35, No. 2, 1993. pp. 175-184.

C.E. Ebeling, *An Introduction to Reliability and Maintainability Engineering*. McGraw Hill, 1997.

L. W. Falls, "Estimation of parameters in compound Weibull distributions", *Technometrics*, Vol 12, 1970. pp 399-407.

L. Goglio and M. Rossetto, "Comparison of Fatigue Data Using the Maximum Likelihood Method," *Engineering Facture Mechanics*, Vol. 71, 2004. pp. 725-736.

P. Hall and D.M. Titterington, "Efficient Nonparametric Estimation of Mixture Proportions," *Journal of the Royal Statistical Society Series B*, Vol. 46, No. 3, 1984. pp. 465-473.

V. Hasselblad, "Estimation of finite mixture of distributions from the exponential family", *Journal of the American Statistical Association*, Vol. 64, 1969, pp 1459-1471.

D. Hillman and R. Wilcoxon, "JCAA/JG-PP No-Lead Solder Project: -55^{o}C to +125^{o}C Thermal Cycle Testing Final Report." Contract GST 0504BM3419. May 28, 2006 REV B. As shown on: http://www.teerm.nasa.gov/LFS%20Reliability/Rockwell%20Collins%20Inc%20JCAA%20JGPP%20Final%20Report%20neg55C%20to%20pos125C%20testing%20Rev%20B.pdf

H. Hirose, "The Trunsored Model and Its Applications to Lifetime Analysis: Unified Censored and Truncated Models," *IEEE Transactions on Reliability*, Vol. 54, No. 1, 2005. pp. 11-21

M. Hollander, I.W. McKeague, and J. Yang, "Likelihood Ratio-Based Confidence Bands for Survival Functions," *Journal of the American Statistical Association*, Vol. 92, No. 437, 1997. pp. 215-226.

D.W. Hosmer, Jr., "A Comparison of Iterative Maximum Likelihood Estimates of the Parameters of a Mixture of Two Normal Distributions Under Three Different Types of Samples," *Biometrics*, Vol. 29, No. 4, 1973. pp. 761-770.

S.-L. Jeng and W. Meeker, "Parametric Simultaneous Confidence Bands for Cumulative Distributions from Censored Data," *Technometrics*, Vol. 43, No. 4, 2001. pp. 450-461.

F. Jensen and N.E. Petersen. *Burn-In: An Engineering Approach to the Design and Analysis of Burn-in Procedures*. New York: John Wiley and Sons. 1982.

R. Jiang and D.N.P. Murthy, "Modeling Failure Data by Mixture of 2 Weibull Distributions: A Graphical Approach," *IEEE Transactions on Reliability*, Vol. 44, No. 3, 1995. pp. 477-488.

S. Jiang and Dimitri Kececiouglu, “Graphical Representation of Two Mixed Weibull Distributions,” *IEEE Transactions on Reliability*, Vol. 41, No. 2, 1992a. pp. 241-247.

S. Jiang and Dimitri Kececioglu, “Maximum Likelihood Estimates from Censored Data from Mixed-Weibull Distribution,” *IEEE Transactions on Reliability*, Vol. 41, No. 2, 1992b. pp. 248-255.

B.-Y. Jing and A.T.A. Wood, “Exponential Empirical Likelihood Is Not Bartlett Correctable,” *The Annals of Statistics*, Vol. 24, No. 1, 1996. pp. 365-369.

A. R. Kaylan and C. M. Harris, “Efficient algorithms to derive maximum likelihood estimates for finite exponential and Weibull mixtures”, Com*puters and Operations Research*, vol 8, 1981, pp 97-104.

D. Kececioglu and W. Wang, “Parameter Estimation for Mixed-Weibull Distribution,” *Proceedings Annual Reliability and Maintainability Symposium, 1998*. pp. 247-252.

J.H. Kim, “A Bayesian Model for Aggregating Expert Opinions,” PhD Dissertation, University of Maryland, Department of Materials and Nuclear Engineering, College Park, MD, 1991.

K.N. Kim, “Optimal burn-in for minimizing cost and multiobjectives,” *Microelectronics Reliability*, Vol. 38, 1998, pp. 1577-1583.

K.O. Kim and W. Kuo, “Some Considerations on System Burn-in,” *IEEE Transactions on Reliability*, Vol. 54, No. 2, June 2005. pp. 207-214.

L.M. Leemis and M. Beneke, “Burn-in Models and Methods: A Review,” *IIE Transactions*, Vol. 22, No. 2, June 1990. pp. 172-179.

R.A. Maller and S. Zhou, *Survival Analysis with Long-Term Survivors*. New York, NY: Wiley, 1996.

K.V. Mardia, H.R. Southworth, and C.C. Taylor, "On Bias in Maximum Likelihood Estimators," *Journal of Statistical Planning and Inference,* vol. 76, 1999. pp. 31-39.

H. Martz and R. Waller. *Bayesian Reliability Analysis*. John Wiley and Sons. 1982.

W. Q. Meeker, "Limited failure population life tests, application to integrated circuit reliability," *Technometrics,* Vol. 29, 1987. pp. 51-65.

J. Mi, "Minimizing Some Cost Functions Related to Both Burn-in and Field Use," *Operations Research*, Vol. 44, No. 3, 1996. pp. 497-500.

M. Modarres, M. Kaminskiy, and V. Krivtsov. Reliability *Engineering and Risk Analysis: A Practical Guide*. Marcel Dekker, Inc., 1999.

A. Mosleh and G. Apstolakis, "Combining Various Types of Data in Estimating Failure Ragtes," *Transactions of the 1983 Winter Meeting of the American Nuclear Society*, San Francisco, CA, 1983.

S.A. Murphy, "Likelihood Ratio-Based Confidence Intervals in Survival Analysis," *Journal of the American Statistical Association*, Vol. 90, No. 432, 1995. pp. 1399-1405.

G.D. Murray and D.M. Titterington, "Estimation Problems with Data from a Mixture," *Applied Statistics,* Vol. 27, No. 3, 1978. pp. 325-334.

J. Paulsen, E. Reed, and J. Kelley. "Reliability of Tantalum Polymer Capacitors," *Proceedings of CARTS Europe 2004*.

H. Perlstein and J.W. Littlefield, “ESS Quantification for Complex Systems,” *Journal of Environmental Sciences*, 1989. pp. 49-57.

Dror Perlstein, William Jarvis, Thomas Mazzuchi, “Bayesian Calculation of Cost Optimal Burn-in Test Durations for Mixed Exponential Populations,” *Reliability Engineering and System Safety*, Vol. 72, 2001. pp. 265-273.

E. Pohl and D. Dietrich, “Environmental Stress Screening Strategies for Multi-Component Systems with Weibull Failure-Times and Imperfect Failure Detection,” *Proceedings of the Annual Reliability and Maintainability Symposium, 1995*. pp. 223-232.

J. Qin, “Empirical Likelihood Ratio Based Confidence Intervals for Mixture Proportions,” *The Annals of Statistics*, Vol. 27, No. 4, 1999. pp. 1368-1384.

R. Reddy and D. Dietrich, “A Two-Level ESS Model: A Mixed Distribution Approach,” *IEEE Transactions on Reliability*, Vol. 43, 1994. pp. 85-90.

R. Schifani and R. Candela, “A New Algorithm for Mixed Weibull Analysis of Partial Discharge Amplitude Distributions,” *IEEE Transactions on Dielectrics and Electrical Insulation*, Vol. 6, No. 2, 1999. pp. 242-249.

P. de Selding, “Buyers and Insurers Want Satellite Makers to Take on More Financial Risk,” *Space News*, 18 April 2005.

R. Soland. “Bayesian Analysis of the Weibull Process with Unknown Scale and Shape Parameters,” *IEEE Transactions on Reliability*, Vol. R-18 1969, pp. 181-184.

S-H. Sheu and Y-H. Chien, “Minimizing Cost-Functions Related to Both Burn-in and Field-Operation Under a Generalized Model,” *IEEE Transactions on Reliability*, Vol. 53, No. 3, 2004. pp. 435-439.

S. Vander Wiel and W. Meeker, “Accuracy of Approximate Confidence Bounds Using Censored Weibull Regression Data from Accelerated Life Tests,” *IEEE Transactions on Reliability*, Vol. 39, No. 3, 1990. pp. 346-351.

T. Woodrow, “JCAA/JG-PP Lead-Free Solder Project: -20oC to +80oC Thermal Cycle Test,” *The Proceedings of SMTA International Conference*, September 24-28, 2006.

L. Yan, J.R. English, and T.L. Landers, “Modeling Latent and Patent Failures of Electronic Products,” *Microelectronics Reliability*, Vol. 35, No. 12, 1995. pp. 1501-1510.

L. Yan and J.R. English, “Economic Cost Modeling of Environmental-Stress-Screening and Burn-In,” *IEEE Transactions on Reliability*, Vol. 46, No. 2, 1997. pp. 275-282.