Quality Engineering

ISSN: 0898-2112 (Print) 1532-4222 (Online) Journal homepage: http://www.tandfonline.com/loi/lqen20

Choosing a reliability inspection plan for interval censored data

Lu Lu & Christine M. Anderson-Cook

To cite this article: Lu Lu & Christine M. Anderson-Cook (2017) Choosing a reliability inspection plan for interval censored data, Quality Engineering, 29:3, 512-535, DOI: 10.1080/08982112.2017.1321118

To link to this article: http://dx.doi.org/10.1080/08982112.2017.1321118

QUALITY ENGINEERING
2017, VOL. 29, NO. 3, 512–535
https://doi.org/10.1080/08982112.2017.1321118

Taylor & Francis
Taylor & Francis Group

Choosing a reliability inspection plan for interval censored data

Lu Lu[a] and Christine M. Anderson-Cook[b]

[a]Department of Mathematics and Statistics, University of South Florida, Tampa, Florida; [b]Statistical Sciences Group, Los Alamos National Laboratory, Los Alamos, New Mexico

ABSTRACT
Reliability test plans are important for producing precise and accurate assessment of reliability characteristics. This article explores different strategies for choosing between possible inspection plans for interval-censored data given a fixed testing timeframe and budget. A new general cost structure is proposed for guiding precise quantification of total cost in inspection test plan. Multiple summaries of reliability are considered and compared as the criteria for choosing the best plans using an easily adaptable method. Different cost structures and representative true underlying reliability curves demonstrate how to assess different strategies given the logistical constraints and nature of the problem. Results show several general patterns exist across a wide variety of scenarios. Given the fixed total cost, plans that inspect more units with less frequency based on equally spaced time points are favored due to the ease of implementation and consistent good performance across a large number of case study scenarios. Plans with inspection times chosen based on equally spaced probabilities offer improved reliability estimates for the shape of the distribution, mean lifetime, and failure time for a small fraction of population only for applications with high infant mortality rates. This article uses a Monte Carlo simulation-based approach in addition to the commonly used approach based on the asymptotic variance and offers comparison and recommendation for different applications with different objectives. In addition, the article outlines a variety of different reliability metrics to use as criteria for optimization, presents a general method for evaluating different alternatives, as well as provides case study results for different common scenarios.

KEYWORDS
cost ratio; cost structure; equally probability-spaced plan; equally time-spaced plan; inspection frequency; mean time to failure; percentile life; Weibull distributions

Introduction

Statistical analysis of reliability data is broadly used in reliability assessment. Depending on time, budget, and logistical constraints, a variety of types of data can be obtained from different test schemes. The tests are run either for a predetermined duration or end after observing a certain number of failures. For either case, if not all of the units have failed by the end of the test, then the data contain only partial information on the survived units, since it is known that the units survived until the end of the study, but not the exact failure times for those units. These data are referred to as *time-censored* or *failure-censored* data (Meeker and Escobar 1998, Ch. 3).

Exact failure times are observed only if the test units are being monitored individually and continuously, but often this is impractical or not possible. This leads to another type of data commonly encountered in reliability analysis known as *interval censored data*, or also referred to as *grouped data* or *readout data* (Tobias and Trindade 2012, p. 42). For example, when testing electronic components, continuous in situ monitoring of all test units can be too costly for many applications. Instead, the units are inspected at prescheduled time points. At each time point, all units that have survived up until the previous inspection time are examined. The number of failures is recorded and the failed units are not tested further. By conducting the inspections at only the prescheduled time points, the cost can be reduced considerably compared to continuous monitoring. However, this testing scheme only records the number of failures that occur in the time intervals between inspections, and hence there is some loss of precision about the exact failure times. For interval censored data, the failure time is only known to fall within the interval between inspection times, where it passed at the start of the interval time, but failed before the end of the interval.

CONTACT Lu Lu icyemma@gmail.com Department of Mathematics and Statistics, University of South Florida, Tampa, FL.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lqen.

This article focuses on case studies for reliability test plans (Meeker and Escobar 1998, Ch. 10) for interval censored data where there is a fixed total budget constraining the testing options. Meeker (1986) compared multiple inspection schemes (including equally time spaced, equally probability spaced, equally spaced in log time, and constrained theoretical optimum) for interval censored data for Weibull underlying distributions based on evaluating the asymptotic variance of certain interesting percentile lifetime (such as 1%, 10%, and 50% life). Seo and Yum (1991) examined some theoretical and practical test plans for accelerated life test for Weibull distributions based on minimizing the asymptotic variance of certain percentile lifetime under the use condition. Shapiro and Gulati (1996) proposed a two-step monitoring system for testing mean lifetime for Exponential distributions. Kim and Yum (2000) further compared the life test plans for Exponential distributions with strategies based on controlling the producer and the consumer risks. For most of the existing work, the inspection frequency (the planned total number of inspections during the test duration for each test unit) was either pre-specified or chosen arbitrarily and the number of test units were chosen as the smallest number of units required to meet a certain standard based on either minimizing the asymptotic variance or controlling the consumer and producer risks. Considering that cost usually plays an important role in choosing life test plans in real applications, this article takes a different perspective by considering a fixed total cost as the main constraint on the test planning. Given a fixed total cost, the practitioner faces an immediate challenge of choosing between the strategies that test more units and less often versus the strategies that test fewer units and more often. We propose a new general structure (see Eq. [1]) to quantify the total cost comprised of three major components that are commonly relevant in broad applications and explore the impacts of different cost structures associated with different types of tests. We also examine the relative size of different cost components on the estimated reliability summaries to demonstrate a different class of life test strategies that is primarily driven by the cost. The new cost quantification structure helps provide principled guidelines for the practitioners to precisely assessing the cost impact on the inspection test plan. In addition, we explore a simulation-based approach for directly assessing the empirical distribution of the reliability estimates rather than relying completely on the asymptotic assumption for studying the estimate properties. Results from both the empirical studies and the asymptotic approximations are compared and result in general recommendations for use in different applications.

Since applications have different priorities for how the reliability results will be used, we look at how the choice of reliability metric impacts the performance of potential test plans. The reliability engineer in charge of the test should decide on the implementation details of the plan: the number of test units, the overall duration of the test, as well as how often and at what times to conduct the inspections. These decisions should be made based on the goal(s) of the test and any cost or logistical constraints. Historically, some guidance on sampling plans has been given by Lavin (1946) and Goode and Kao (1960), but we revisit this guidance to illustrate differences in performance for a variety of quantities of interest. In general, testing more units gains more information and results in more precise estimates of the reliability. Having more inspections during the course of the test reduces the uncertainty induced by censoring and hence improves the estimation. However, increasing the number of test units or the inspection frequency also increases the overall cost of the inspection plan. In circumstances with a fixed budget, the trade-off between alternatives means that we can choose to test more units less often, or fewer units more often. In addition, the timing of the inspections affects both the cost and the reliability estimates. The earlier the inspections are performed, the fewer failures are expected to be observed, which results in less information gain due to increased degree of censoring and also increases the overall cost as more units need to be inspected multiple times as they survive more inspections. Hence, with a fixed total budget, the number of test units and how often and when to inspect the units can substantially impact the quality of the estimated reliability.

In this article, we explore how different choices of the sample size and inspection timing and frequency impact the reliability assessment through a series of short case studies. Consider a case where the lifetime of a product, T, is assumed to follow a Weibull distribution, *Weibull* (α, β), using the parameterization $f(t) = \frac{\beta t^{\beta-1}}{\alpha^{\beta}} e^{-\left(\frac{t}{\alpha}\right)^{\beta}}$, where α is the characteristic life or scale parameter, and β is the shape parameter (Tobias and Trindade 2012, p. 88). To estimate its lifetime distribution, a number of test units, n, are inspected at m prescheduled time points, $t_1, t_2, \ldots, t_{m-1}, t_m$. At each time point, the unit either passes or fails. Failed units are removed from further testing, and units that pass continue to be tested until the end of the test at time t_m. The obtained interval censored data, denoted as $d_1, d_2, \ldots, d_m$, contain the number of failures d_k observed during the time interval $(t_{k-1}, t_k]$. The number of units that survive (pass) through the final test t_m is denoted by $r = n - \sum_{k=1}^{m} d_k$. For any particular test plan with a selected combination of $\{n; t_1, t_2, \ldots, t_{m-1}, t_m\}$, the parameters of the Weibull distribution, α and β, can be estimated using maximum likelihood estimation (MLE, Tobias and Trindade 2012, p. 98). Details of the data analysis using the MLE approach are given in the section "Reliability estimates using maximum likelihood estimation."

One important initial aspect of setting the objectives of the test is to choose the quantitative reliability summary (or summaries) of interest. For any assumed life distribution, a variety of reliability quantities can be expressed as functions of the unknown parameters. How the results will be used can often guide which summaries are most beneficial and appropriate for the goals of the inspection. Given the estimated model parameters, $\hat{\alpha}$ and $\hat{\beta}$, other reliability summaries are calculated as known functions of the estimated parameters and their uncertainty quantified based on asymptotic theory of MLE (see the next section for more details). The goal is to choose an inspection test plan that gives the most precise estimates of the reliability characteristics of interest. In our case study, we consider several typical reliability quantities of interest:

- the Weibull parameters (the scale parameter α and the shape parameter β);
- the median lifetime T_{50} (the time by which half of the population is expected to fail);
- the mean time to failure $MTTF$ (the expected lifetime for all population units);
- the 10^{th} and 1^{st} percentiles of lifetime, denoted as T_{10} and T_1, respectively (the time by which 10% or 1% of the population is expected to fail); and
- the reliability at the end of the test $R(t_m)$ (the probability that a test unit survives to the end of the test).

Note that for different studies, not all reliability summaries above are of interest or are considered equally important. For example, in warranty studies or accelerated life test plans, early failures for a small fraction of the population under the normal use condition may be of more interest than the mean or median lifetime. For other studies with different goals, different choices of reliability quantities could also be considered. Instead of enumerating all possible reliability summaries, this article focuses on illustrating a general method for interval data test planning given cost constraints through some representative case studies. The general methodology outlined in this article for evaluating the test plan can be easily adapted for any reliability summary of interest chosen for a study.

Given a fixed overall budget, the number of units used and the number of possible tests per unit in the inspection plan is typically determined based on the associated cost. For interval censored data where the inspections are conducted at some preselected time points, there are typically three types of cost involved: (1) the one-time cost of the initial setup for a single test unit, denoted as C_{unit}, (2) the cost of the test setup for the overall inspection at each scheduled time points, denoted as C_{ts}, and (3) the cost of each inspection of a test unit at each time point, denoted as C_{insp}. Recall that units that fail at early time points are removed from testing at all subsequent inspection time points, which means that the actual exact number of inspections for each test unit is unknown before conducting the test. Hence, the cost associated with individual inspections across all test units is also unknown. However, we can calculated the expected total cost, denoted by ETC, as the sum of the three components:

$$ETC = C_{unit} \cdot n + C_{ts} \cdot m + C_{insp} \cdot ETI \qquad [1]$$

where $ETI = n\{m - \sum_{i=1}^{m-1} F(t_i)\}$ is the expected total number of inspections for all n test units, which is calculated based on expecting $n[1 - F(t_{i-1})]$ units to be inspected at time t_i, since all units that survived the previous inspection need to be inspected at the next time point no matter if the units pass or fail the inspection. The Weibull cumulative distribution function (cdf) at ith inspection time t_i, i.e., the expected proportion of failure by time t_i, is given by

$$F(t_i) = 1 - e^{-\left(\frac{t_i}{\alpha}\right)^{\beta}} \qquad [2]$$

Note that the relative contributions of the above three cost components varies considerably for different applications. For example, for assessing the reliability of a stockpile of missiles, the cost of setting up a unit for testing initially, C_{unit} can be negligible, while the main costs may be the test setup at each schedule time (gathering the units for testing and having the equipment set-up and available for the inspection), C_{ts}, and the cost of individual inspections (conducting the non-destructive test of the units), C_{insp}. However, when testing electronic components or devices, the primary costs can be generated from both the initial setup of the test equipment for each device, C_{unit}, and the individual inspections (sending personnel to operate the test equipment and conduct the test), C_{insp}. In this case, the fixed test setup at each inspection time, C_{ts}, might be negligible. Given a fixed total budget, the number of units and how often they are tested are dependent on the relative size of the different costs. In our case studies, we explore the impact of different cost structures and the cost ratios between the different components on the selection of the best inspection plans. In particular, we look at two different cost structures.

(1) Cost structure I considers an application analogous to the stockpile test, where C_{ts} and C_{insp} are the primary costs. Within this scenario, the number of test units allowed primarily depends on the relative size of the two primary costs. If the test setup at each inspection time is of similar magnitude or smaller than the cost for inspecting an individual unit ($C_{ts} \leq C_{insp}$), then more units can be tested and the cost of actual inspections will dominate the total cost. But if the test setup cost is dramatically larger than the individual inspections ($C_{ts} > C_{insp}$), then the number of units can be tested will be reduced considerably due to the large expense on setting up for the inspection at each time point. We explore different possible ratios for these two components of cost, including $CR_1 = \frac{C_{ts}}{C_{insp}} \in \{0.1, 1, 10, 25, 100\}$, which summarize a wide range of possible scenarios typical of a broad variety of applications.

(2) Cost structure II examines cases similar to testing electronic devices where C_{unit} and C_{insp} are the dominant costs. For this case, we also consider a wide range of scenarios including $CR_2 = \frac{C_{unit}}{C_{insp}} \in \{0.1, 1, 5, 10, 25\}$. Small ratios correspond to a very small initial cost for setting up the test for each unit compared to an individual inspection, while the larger ratio values indicate the initial cost dominating the cost for testing each unit at multiple time points. When the initial cost per unit is relatively small, the individual inspections dominate the total cost; but if as the initial cost per unit becomes substantially larger than the individual inspection cost, then the number of units possible within budget mainly depend on the initial cost per unit, which does not change much with how often the inspections are conducted during the test.

In addition to the cost structure, the shape of the underlying lifetime distribution also affects the expected total cost (ETC) and the relative performance of the test plans. If the lifetime curve starts with high reliability and drops slowly throughout the duration of the test, then fewer units are expected to fail and more total inspections (ETI) are anticipated across all test units. With a fixed total budget, fewer units can then be tested and heavier censoring at the conclusion of the inspection plan is expected. However, if the lifetime curve decreases quickly over time, then more units are likely to fail, leading to fewer inspections expected for each test unit and allowing for more initial units to be considered at the start of the test. This case can lead to improvements in the precision of estimation. Therefore, since the true underlying reliability curve is generally not precisely known at the time of planning the test, exploring different possible shapes of the reliability curve allows us to choose a test plan with robust performance for different possible underlying life distributions. We consider four different Weibull reliability curves with differently shaped hazard functions. The selected curves reflect common characteristics from our experiences with applications of real systems and products. In general, the products start with very high overall reliability, and interest in testing involves covering the early stages of their degrading reliability. Often when the units start to have reliability that drops below 50–70%, the manufacturer is less interested in characterizing their performance. Since the hazard function often varies between increasing, stable, or decreasing over time, we explore each of these scenarios. While individual testing scenarios from real examples are proprietary, the examples chosen are indicative of common problems. The overall test duration was chosen to be $t_m = 3000$ hr to have about 40–70% censoring rate across the four cases explored. Note that this choice can be easily adapted to meet the constraints and budget for other studies. Since the scale parameter can be considered essentially as a multiplying factor on the chosen time scale,

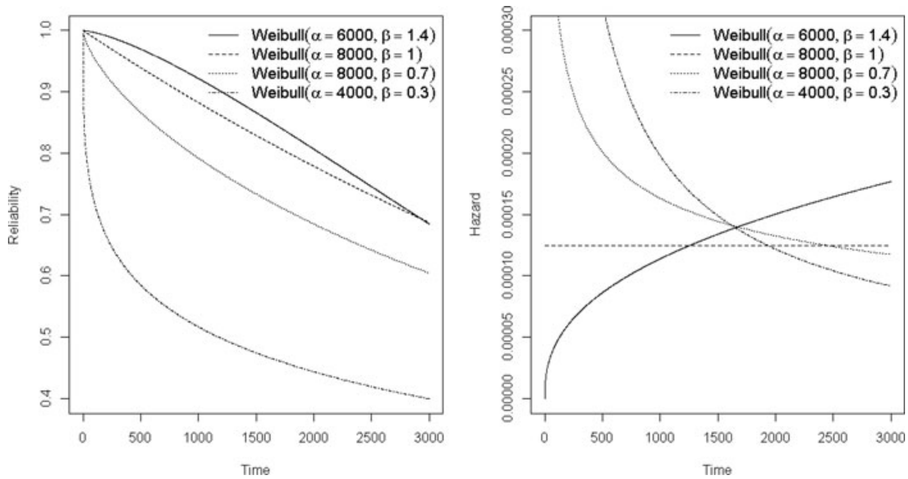


Figure 1. (a) The reliability/survival functions and (b) the hazard functions for the four different Weibull distributions considered.

the choice of the scale parameter should not affect the relative performance of the test plans or the general conclusions. Figure 1 shows the reliability/survival functions and the hazard functions for the four Weibull distributions. Curve 1 is a $Weibull(\alpha = 6000, \beta = 1.4)$ distribution with an increasing hazard function with time. Curve 2 has a constant failure rate over time (exponential or $Weibull(\alpha = 8000, \beta = 1)$ distribution). Curve 3 is a $Weibull(\alpha = 8000, \beta = 0.7)$ distribution with has a decreasing hazard function over time. Finally Curve 4 has a sharply decreasing hazard function ($Weibull(\alpha = 4000, \beta = 0.3)$). Note that Curves 1 and 2 have similar reliability of around 0.68 by the end of the test duration. Curve 3 has a lower reliability with $R(t_m) \approx 0.63$, and Curve 4 has much lower reliability of $R(t_m) \approx 0.4$ by the end of the test. The scale parameters are chosen to control the censoring rate within the test duration to be around the (1/3,2/3) window.

To explore different choices of inspection times $\{t_1, t_2, \ldots, t_m\}$, we consider two common strategies that are either probability-based or time-based (Meeker 1986). The probability-based strategy chooses inspection times with equally spaced reliabilities (i.e., the change of reliability between adjacent time points are approximately equal). The time-based strategy chooses equally spaced inspection times within the test duration. Note that the time-based strategy is straightforward to use regardless of the underlying true lifetime curve, while the probability-based strategy requires a good guess or estimate of reliability curve based on subject matter expertise or historical data. If the probability-based strategy shows no obvious advantage in performance, then the time-based strategy is likely a better choice because of its ease of implementation. In our case study, we compare six inspection plans with different inspection frequencies based on using both probability-based and time-based plans. For the simplicity of notation, we refer to the probability-based plans as P1, P2, and P3 with 8, 6, and 4 inspection times, respectively, and we use T1, T2, and T3 to denote the time-based plans with 8, 6, and 4 inspection times. Figure 2 shows the six inspection plans for the first Weibull distribution (Curve 1) with the top row containing the probability-based choices. Note that the probabilities between inspection points are equally spaced on the y-axis. The bottom row of Figure 2 shows the time-based choices with equal spacing on the x-axis between inspections.

The remainder of this article is organized as follows. In the next section, we provide more details about the planned analysis using maximum likelihood estimation and how potential quantities of interest can be estimated from the Weibull parameterization. The following section compares the different reliability estimates for the six inspection plans for different cost ratios under cost structure I ($CR_1 = \frac{C_{ts}}{C_{insp}}$) for different underlying reliability curves. Then, the next section considers similar comparisons for the second cost structure II ($CR_2 = \frac{C_{unit}}{C_{insp}}$) across different cost ratios. The final section contains some conclusions and discussion.

Reliability estimates using maximum likelihood estimation (MLE)

Since interval censored data have incomplete information on the failure times of the test units in the sample, the likelihood of each observed failure is captured by the probability that the unit failed in its observed

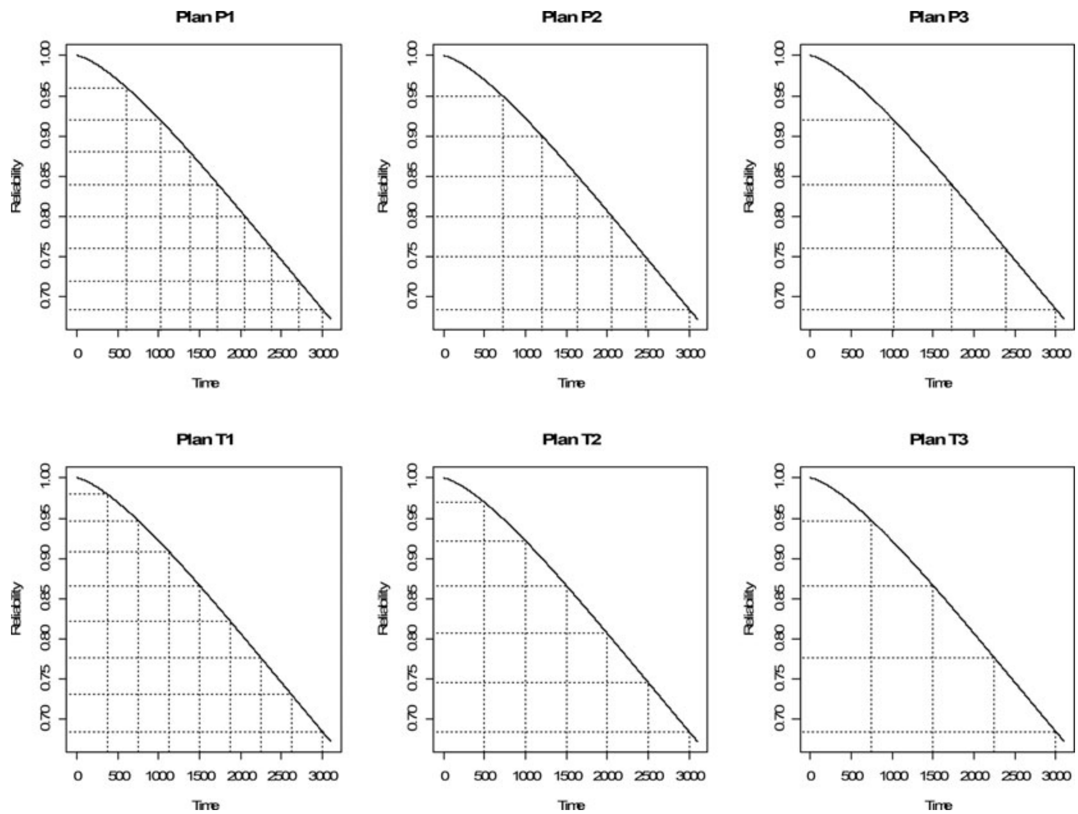


Figure 2. Six inspection plans combining the probability-based (first row) or time-based (second row) plans with the number of inspections across the test duration (8, 6, and 4 times corresponding to left, middle, and right columns, respectively) for Curve 1 with the *Weibull*(6000, 1.4) distribution.

time interval $(t_{i-1}, t_i]$, calculated by $F(t_i) - F(t_{i-1})$, where $F(\cdot)$ is the Weibull cdf defined in Eq. [2]. The likelihood for each unit still working at the end of the test (right censored at the end of the test, t_m) is given by $1 - F(t_m)$. For a dataset with $d_1, d_2, \ldots, d_m$ observed failures at times $t_1, t_2, \ldots, t_m$, respectively, and $n - \sum_{i=1}^{m} d_i$ survivors from n total test units, the likelihood function is given by

$$L(\alpha, \beta) \propto F(t_1)^{d1} \left\{ \prod_{i=2}^{m} [F(t_i) - F(t_{i-1})]^{d_i} \right\} \times [1 - F(t_m)]^{n - \sum_{i=1}^{m} d_i}.$$

The MLEs of the model parameters are obtained by minimizing the negative log-likelihood

$$-\log(L(\alpha, \beta)) = C + d_1 logF(t_1) + \sum_{i=2}^{m} d_i log[F(t_i) - F(t_{i-1})] + \left(n - \sum_{i=1}^{m} d_i\right) log[1 - F(t_m)],$$

where C is a constant. A closed form analytical solution for the MLEs that minimize the negative log-likelihood is generally not available for interval censored data. Instead, a numerical search algorithm, such as the Newton-based approaches (Deuflhard 2004), can be used to find the MLEs for α and β.

Since numerical search algorithms, especially the Newton-based approaches, can be sensitive to the initial values specified, we use a combination of reliability probability plotting and least squares estimation to provide sensible initial values to accelerate the search and avoid spurious convergence. For a Weibull distribution, we have $\ln(t) = \ln(\alpha) + \frac{1}{\beta}\ln\{-ln[1 - F(t)]\}$, and hence a linear regression model fitted between $\ln(t_i)$ and $ln\{-ln[1 - \hat{F}(t_i)]\}$ provides guidance on sensible starting values for model parameters α and β. $\hat{F}(t_i)$ is estimated with the empirical cdf, $\hat{F}(t_i) = \sum_{j=1}^{i} d_j / \sum_{j=1}^{m} d_j$ at all t_i's where $d_i > 0$. Least squares (LS) methods can be used to estimate the intercept $ln(\alpha)$ and slope $1/\beta$, which are then used as initial values in the Newton-based numerical search algorithm. The *nlm* function in R (https://stat.ethz.ch/R-manual/R-devel/library/stats/html/nlm.html) was used to find the MLEs. The code for implementing the method is available upon request from the authors.

With the MLEs, $\hat{\alpha}$ and $\hat{\beta}$, for the Weibull parameters, other reliability quantities of interest can be calculated that characterize specific aspects of the distribution. For example, the median lifetime, T_{50}, the mean time to failure, $MTTF$, the 10th and 1st percentile of lifetime, T_{10} and T_1, and the reliability at time t_m, $R(t_m)$, can be estimated by

$$\begin{aligned}\hat{T}_{50} &= \hat{\alpha}\ (-\ln(0.5))^{1/\hat{\beta}} = \hat{\alpha}(\ln(2))^{1/\hat{\beta}}\\ \widehat{MTTF} &= \hat{\alpha}\ \Gamma\left(1+1/\hat{\beta}\right)\\ \hat{T}_{10} &= \hat{\alpha}\ (-\ln(0.9))^{1/\hat{\beta}}\\ \hat{T}_{1} &= \hat{\alpha}\ (-\ln(0.99))^{1/\hat{\beta}}\\ \hat{R}\ (t_m) &= e^{-\left(\frac{t_m}{\hat{\alpha}}\right)^{\hat{\beta}}},\end{aligned}$$

respectively. Here, $\Gamma\ (x) = \int_0^\infty t^{x-1}e^{-t}dt$ is the Gamma function. Note, the low percentile lifetime summaries are useful for estimating the time by which a small fraction of the population units will fail. This quantification of early failure time is often of interest in reliability demonstration and assurance activities, where the impact of even a small fraction of failures is potentially of high consequence. For example, a new design can be accepted if the 10% life (by which 10% of the units fail) is expected to exceed 2,000 hr with a certain confidence level. To quantify the uncertainty of the estimated reliability quantities, the asymptotic properties of the MLEs are often used to estimate approximate confidence intervals. Let $G(\theta)$ denote a function of the model parameters, θ, then the asymptotic variance, AV, of $G(\hat{\theta})$ can be obtained using the Delta method (Casella and Berger 2002) as given by

$$AV\left[G\left(\hat{\theta}\right)\right] = \left[\frac{\partial G(\theta)}{\partial\theta}\right]_{\theta=\hat{\theta}} AV\left(\hat{\theta}\right)\left[\frac{\partial G(\theta)}{\partial\theta}\right]^T_{\theta=\hat{\theta}}. \quad [3]$$

In our application, we use the parameterization $\theta = (\ln(\alpha),\ 1/\beta)$ to obtain the MLEs and the asymptotic variance of $\hat{\theta}$ can be obtained using the inverse matrix of the Hessian matrix obtained from the R output. Then the asymptotic variances of $\hat{T}_{50}$, $\widehat{MTTF}$, $\hat{T}_{10}$, $\hat{T}_1$, and $\hat{R}(t_m)$ can be obtained using the formulas in Eq. [3]. Note that the asymptotic approximation works well for large samples. For our case studies, we also examine the empirical confidence intervals obtained using the Monte Carlo method from the simulations. In the next section, we compare the relative performance of the Monte Carlo and asymptotic confidence intervals from different inspection plans.

Comparisons under cost structure I

In this section, we compare different inspection plans based on a variety of cost ratios under cost structure I ($CR_1 = \frac{C_{ts}}{C_{insp}}$), where the one-time set-up cost for each unit, C_{unit}, is assumed to be negligible. For this problem, we assume a fixed total budget of 1,200 cost units. Note that like the scale parameters, the total cost also serves as a multiplying factor on the cost unit, and hence has little impact on the patterns of reliability performance or the general conclusions. A practitioner could easily adapt these constraints to match their problem, and the general methodology adapts easily to other fixed costs. To explore the impact of different reliability curves, we simulate $M = 10,000$ data sets from each of the four reliability curves in Figure 1, and use the methods described previously to analyze the data and obtain uncertainty summaries of the reliability quantities of interest for all combinations of the six inspection plans and five possible cost ratios, $CR_1 \in \{0.1, 1, 10, 25, 100\}$.

Reliability curve 1 – Weibull distribution with an increasing hazard function

Consider the $Weibull(\alpha = 6000, \beta = 1.4)$ distribution with an increasing hazard function, shown with the solid curve in Figure 1b. For a test duration of 3,000 hr, Figure 2 shows the inspection times for the six plans chosen with the probability-based (top row) or time-based (bottom row) strategy for 8, 6, or 4 inspections (left to right columns). Recall from the introduction that the probability-based plans are equally spaced on the y-axis, while the time-based plans are equally spaced on the x-axis. The time-based strategies (plans T1, T2, and T3) inspect every 375, 500, and 750 hr for a total of 8, 6, and 4 inspections, respectively. Compared to the evenly time spaced strategies, the probability-based strategies (plans P1, P2, and P3) tend to have longer gaps between inspections at early times (e.g., the first inspection is done around 600, 750, and 1,000 hr for 8, 6, and 4 inspections, respectively) and more frequent inspections towards the end of the test. This is because the reliability reduces more quickly for older

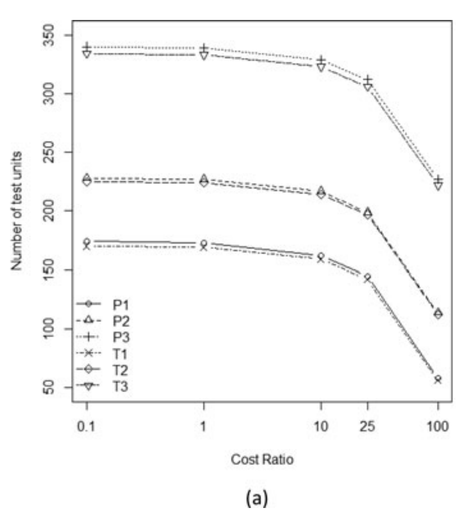
Table 1. The number of test units and the number of expected total inspections for all combinations of the six inspection plans and five cost ratios ($CR_1 = C_{ts}/C_{insp}$) under cost structure 1.

	Number of Units (N)						Number of Expected Total Inspections					
CR_1	P1	P2	P3	T1	T2	T3	P1	P2	P3	T1	T2	T3
0.1	174	228	340	170	225	334	1197	1197	1197	1195	1195	1199
1	173	227	339	169	224	333	1190	1192	1193	1188	1189	1195
10	162	217	329	159	214	323	1115	1139	1158	1118	1136	1159
25	145	199	312	142	197	306	998	1045	1098	998	1046	1098
100	58	114	227	56	112	222	399	599	799	394	595	797

ages and hence the time interval between inspections becomes shorter for equal changes in reliability.

Given the total cost of 1,200, we consider five different cost ratios where $CR_1 \in \{0.1, 1, 10, 25, 100\}$. By ignoring the first term in Eq. [1] (since it is assumed to be negligible), we calculate the largest number of test units possible that stays within the total budget. Figure 3a shows a plot of the number of test units for different cost ratios. The cost ratio on the x-axis is plotted on a log scale, which allows easier interpolation between the examined cost ratios. Different inspection plans are shown with different line types and symbols. When the cost ratio is relatively small (below 1), there is hardly any difference in the number of units within a given inspection paradigm. However, as the cost ratio becomes larger, the initial sample size decreases considerably for each test plan. Comparing across the plans, there are substantial differences between different inspection frequencies. The more often we conduct the inspections, the fewer units we can afford to test. However, there are negligible differences between probability-based and time-based strategies for the same number of inspections (e.g., T1 vs. P1). Figure 3b shows the expected total number of inspections (ETI) across all individual test units for the six plans. When the cost ratio is smaller than 1, there is little difference between the ETIs for all six plans. But as the test setup cost (C_{ts}) becomes considerably larger than the cost for individual inspections (C_{insp}), fewer total units can be tested for more frequent inspection plans, as more of the budget is spent on test setup at each testing time point. The probability-based and the time-based strategies show negligible differences in terms of the ETIs. Table 1 gives the numerical values for both quantities shown in Figure 3.

To evaluate the performance of the testing plans, we are interested in both the accuracy and precision of the estimated quantities. Hence, we compare the different test plans using the relative root of mean squared error (RMSE). Figure 4 shows this quantity for α, β, T_{50}, $MTTF$, T_{10}, T_1, and $R(t_m)$, and the real cost from the actual tests, summarized over the 10,000 simulations. Note that the relative RMSE reports the ratio of the RMSE and the true value, and hence provides a standardized (unitless) measurement for the overall discrepancy between the estimates and the true value



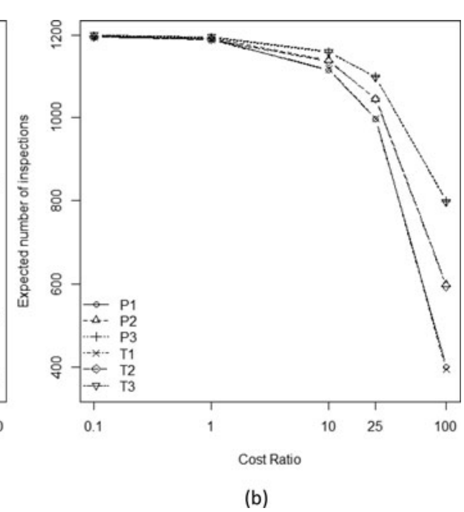


Figure 3. (a) The number of test units and (b) the expected number of inspections (ETIs) for the six inspection plans across five different cost ratios ($CR_1 = C_{ts}/C_{insp}$) under cost structure I for Curve 1 with the $Weibull(6000, 1.4)$ distribution.

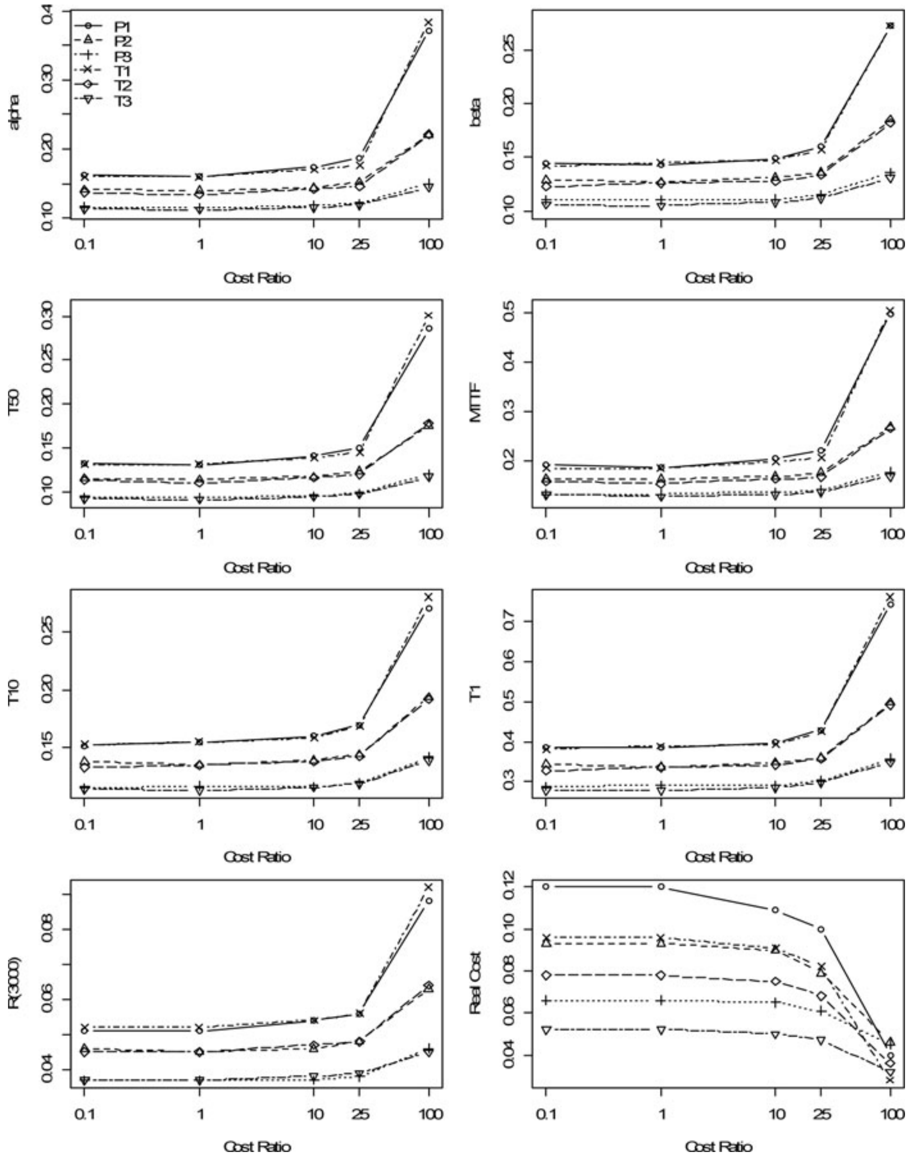


Figure 4. The *relative root of mean squared error* (RMSE relative to the true value) for the seven reliability summaries and the real cost for the six inspection plans with five cost ratios under cost structure I for Curve 1 with the *Weibull* (6000, 1.4) distribution.

measured in the squared distance across a larger number of simulations. The summary can be partitioned into components for the variance and squared bias, which quantify both precision and accuracy of the estimates. The real cost relative RMSE measures how much difference there is in the total cost of the inspection plans once the actual number of observed failures for a particular test is taken into account at each time point. Generally, the reliability estimates have increased relative RMSE for plans where more units are inspected less often. The probability-based and time-based strategies have similar performance across all reliability summaries. The real cost also has smaller relative RMSE as we inspect more units and less often. However, the probability-based strategies generally have considerably larger relative RMSE than the time-based, which indicates the time-based strategy tends to have more consistent real cost relative to the expected cost.

If we look at the reliability estimates across different cost ratios between C_{ts} and C_{insp}, we see that when the test setup costs no more than the individual inspections ($CR_1 < 1$), there are only small differences between the inspection paradigms for all summaries. As the cost ratio increases, the relative RMSE increases slowly until the test setup becomes more than 25 times the cost of individual inspections, then the relative RMSE increases more rapidly. The probability-based and time-based strategies still have similar performance. However, the relative RMSE inflates more quickly for plans that inspect fewer units more

Table 2. The RMSE of the seven reliability quantities and the real cost for all combinations of the six inspection plans and five cost ratios under cost structure 1. Best values (with ties) are shown in bold and second choices are shown in italics for each combination.

	α						β					
CR_1	P1	P2	P3	T1	T2	T3	P1	P2	P3	T1	T2	T3
0.1	0.163	0.141	*0.116*	0.16	0.137	**0.113**	0.145	0.129	*0.11*	0.142	0.123	**0.106**
1	0.16	0.14	*0.115*	0.16	0.134	**0.112**	0.143	0.127	*0.11*	0.145	0.126	**0.105**
10	0.174	0.144	*0.118*	0.17	0.142	**0.115**	0.149	0.131	*0.111*	0.147	0.128	**0.108**
25	0.187	0.152	*0.121*	0.177	0.146	**0.119**	0.16	0.136	*0.115*	0.157	0.134	**0.112**
100	0.372	0.22	*0.15*	0.384	0.221	**0.145**	0.273	0.185	*0.136*	0.273	0.182	**0.131**
	T_{50}						$MTTF$					
0.1	0.132	0.115	*0.093*	0.131	0.113	**0.092**	0.192	0.164	*0.133*	0.185	0.158	**0.131**
1	0.13	0.114	*0.093*	0.131	0.11	**0.091**	0.186	0.163	*0.133*	0.185	0.154	**0.128**
10	0.141	0.117	*0.095*	0.139	0.116	**0.094**	0.205	0.168	*0.137*	0.198	0.163	**0.132**
25	0.15	0.123	*0.098*	0.145	0.12	**0.097**	0.222	0.176	*0.141*	0.207	0.168	**0.137**
100	0.286	0.175	*0.12*	0.301	0.178	**0.117**	0.498	0.268	*0.177*	0.503	0.266	**0.169**
	T_{10}						T_1					
0.1	0.152	0.138	*0.115*	0.153	0.133	**0.114**	0.386	0.343	*0.288*	0.381	0.328	**0.279**
1	0.155	0.135	*0.117*	0.155	0.135	**0.113**	0.384	0.337	*0.292*	0.389	0.336	**0.278**
10	0.16	0.139	*0.116*	0.159	0.138	**0.116**	0.397	0.348	*0.291*	0.393	0.342	**0.285**
25	0.17	0.144	*0.12*	0.169	0.143	**0.119**	0.429	0.36	*0.302*	0.425	0.358	**0.298**
100	0.271	0.194	*0.142*	0.28	0.192	**0.139**	0.743	0.497	*0.358*	0.759	0.492	**0.348**
	$R(t_m)$						Real Cost					
0.1	0.051	0.046	**0.037**	0.052	0.045	**0.037**	0.12	0.093	*0.066*	0.096	0.078	**0.052**
1	0.051	0.045	**0.037**	0.052	0.045	**0.037**	0.12	0.093	*0.066*	0.096	0.078	**0.052**
10	0.054	0.046	**0.037**	0.054	0.047	*0.038*	0.109	0.09	*0.065*	0.091	0.075	**0.05**
25	0.056	0.048	**0.038**	0.056	0.048	*0.039*	0.1	0.079	*0.061*	0.082	0.068	**0.047**
100	0.088	0.063	*0.046*	0.092	0.064	**0.045**	0.04	0.046	*0.045*	0.028	0.036	**0.032**

frequently. In contrast, the relative RMSE for real cost reduces as the cost ratio gets larger between C_{ts} and C_{insp}. As CR_1 approaches 100, the difference between the real costs between the 6 inspection plans decrease and results look quite similar. The corresponding numerical values of relative RMSE for all six plans with different cost ratios are shown in Table 2.

The RMSE provides a combined summary of both bias and variance of the estimates. To understand the relative contribution from the two aspects, Table A1 in the Appendix reports the fraction of variance out of the total mean squared error (MSE) across the 10,000 simulations for the 5 reliability summaries and the real cost. The variance consistently dominates the total MSE for all reliability summaries with a lowest fraction of 97.9% for all combinations of scenarios. Therefore, the estimated biases of the reliability summaries are minimal for all reliability summaries. In contrast, the variance of real cost is generally small and counts for no more than 16% of the MSE across all scenarios. The real cost is generally larger than the expected cost by 2–12% across different cost ratios and inspection plans. This indicates that the cost tend to be consistently higher than the expected cost with relatively small variability across simulations. However, the difference gets smaller for largest cost ratios since the test setup cost takes a larger proportion of the total budget and is not subject to any uncertainty. Also, the time-based strategies generally have smaller bias than the probability-based strategies.

Figure 5 shows the relative width of the 95% empirical confidence interval (CI) for the reliability summaries and the real cost for all six inspection plans across different cost ratios. Again, for improved interpretability we standardize the CI width by dividing by the true value. The patterns are consistent with the relative RMSE in Figure 4 for all reliability summaries. Across all reliability summaries, the $MTTF$ is estimated least precisely, while $R(t_m)$ is most precisely estimated relative to its size. The main differences in the precision of the CIs are caused by the different inspection frequencies. Generally, more precise estimation (narrower CI) is obtained by inspecting more units and less often. The probability-based and the time-based strategies have similar performance. As the cost ratio increases for more expensive test setups, we gradually lose precision until the test setup becomes very expensive (more than 25 times of the cost for individual inspections) and the width of the empirical CI is substantially inflated. On the other hand, the plans that inspect fewer units more often usually have a smaller empirical CI width for real cost, which translates into

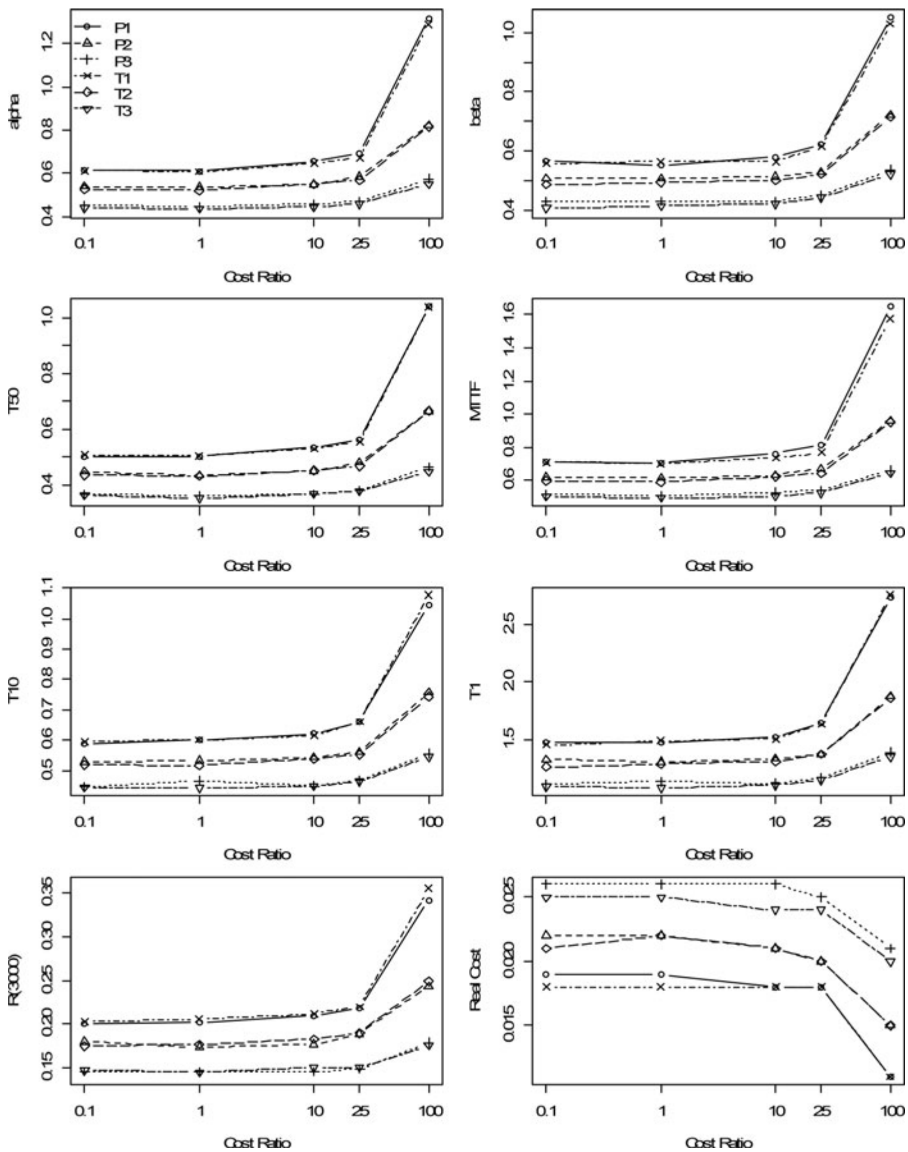


Figure 5. The relative width of the 95% empirical confidence intervals relative to the true value for the seven reliability summaries and the real cost for the six inspection plans with five cost ratios under cost structure I for Curve 1 with the *Weibull*$(6000, 1.4)$ distribution.

smaller variation of the realized cost. This is because more frequent inspections tend to result in less variability in the number of failures observed within each time interval, and hence less variation in the realized cost. The numerical values of the relative empirical CI widths from Figure 5 are shown in Table 3.

Note that the average empirical CI width summary in Figure 5 was based on summarizing over 10,000 simulations. An alternative summary frequently used in many real applications is the asymptotic confidence interval, which can be computed from a single set of data. Measures of uncertainty based on the width of the 95% asymptotic confidence intervals for four reliability summaries T_{50}, $MTTF$, T_{10}, and $R(t_m)$ are provided in Figure A1 in the Appendix, with the actual numerical values included in Table 4. The four plots in the first columns of Figure A1 shows the average relative asymptotic CI width over 10,000 simulations (the average width of the calculated CIs relative to the true value of the quantity of interest), while the second column shows the upper 95% asymptotic CI width across all the simulations. These two summaries provide typical and worst case scenario summaries for the uncertainty measured based on the asymptotic results. Note that the average asymptotic CI width for the four reliability summaries on T_{50}, $MTTF$, T_{10}, and $R(t_m)$ generally match with the empirical CI width summary in Figure 5, except that the average asymptotic CI width for $MTTF$ was a bit larger than the empirical CI width. This can be manifested in some cases that we have

Table 3. The relative width of the empirical 95% confidence intervals summarized across 10,000 simulations (relative to the size of the quantity) for the seven reliability quantities and the real cost for all combinations of the six inspection plans and five cost ratios under cost structure 1. Best values (with ties) are shown in bold and second choices are shown in italics for each combination.

	α						β					
CR_1	P1	P2	P3	T1	T2	T3	P1	P2	P3	-T1	T2	T3
0.1	0.615	0.539	0.451	0.612	0.527	**0.441**	0.564	0.507	*0.429*	0.557	0.486	**0.407**
1	0.611	0.537	0.445	0.608	0.521	**0.436**	0.55	0.507	*0.429*	0.564	0.493	**0.414**
10	0.656	0.549	0.457	0.647	0.548	**0.444**	0.579	0.514	*0.429*	0.564	0.5	**0.421**
25	0.692	0.585	0.471	0.673	0.569	**0.459**	0.622	0.529	*0.45*	0.614	0.521	**0.443**
100	1.317	0.821	0.575	1.287	0.815	**0.552**	1.05	0.721	*0.536*	1.029	0.714	**0.521**
	T_{50}						$MTTF$					
0.1	0.502	0.447	*0.364*	0.507	0.434	**0.363**	0.714	0.62	*0.514*	0.71	0.597	**0.504**
1	0.502	0.433	*0.361*	0.504	0.431	**0.352**	0.706	0.615	*0.506*	0.7	0.591	**0.496**
10	0.534	0.451	**0.366**	0.529	0.452	*0.367*	0.763	0.631	*0.527*	0.736	0.624	**0.506**
25	0.563	0.48	*0.38*	0.555	0.467	**0.377**	0.812	0.67	*0.54*	0.769	0.643	**0.525**
100	1.04	0.665	*0.464*	1.038	0.665	**0.447**	1.647	0.96	*0.661*	1.576	0.951	**0.645**
	T_{10}						T_1					
0.1	0.588	0.53	*0.448*	0.595	0.52	**0.446**	1.476	1.323	*1.108*	1.453	1.263	**1.09**
1	0.601	0.533	*0.467*	0.602	0.518	**0.443**	1.469	1.306	*1.133*	1.488	1.287	**1.077**
10	0.621	0.544	**0.45**	0.616	0.538	**0.45**	1.518	1.332	*1.117*	1.504	1.307	**1.101**
25	0.662	0.562	*0.469*	0.661	0.554	**0.464**	1.647	1.368	*1.164*	1.632	1.372	**1.149**
100	1.045	0.758	*0.556*	1.076	0.744	**0.546**	2.73	1.873	*1.386*	2.749	1.862	**1.35**
	$R(t_m)$						Real Cost					
0.1	0.2	0.18	**0.146**	0.203	0.175	*0.147*	*0.019*	0.022	0.026	**0.018**	0.021	0.025
1	0.202	0.174	*0.146*	0.206	0.177	**0.145**	*0.019*	0.022	0.026	**0.018**	0.022	0.025
10	0.21	0.177	**0.146**	0.212	0.183	*0.15*	**0.018**	0.021	0.026	**0.018**	0.021	0.024
25	0.219	0.189	**0.149**	0.22	0.19	*0.151*	**0.018**	0.02	0.025	**0.018**	0.02	0.024
100	0.342	0.243	*0.179*	0.355	0.249	**0.176**	**0.011**	0.015	0.021	**0.011**	0.015	0.02

evaluated using other reliability curves and/or cost structures (shown in the SM) and it results from having a few exceptional cases where the numerical search for the MLEs fail to converge to sensible estimates, which led to large estimates of $MTTF$ and its associated asymptotic confidence intervals. We used 10 times the true values as the threshold to exclude the 4 extreme cases (out of 10,000 simulations) for this case. With this wide acceptable range, the average asymptotic CI width is still inflated by a few extreme estimates less than 10 times the true values. For some other cases shown in the SM, the upper 95% asymptotic CI width can be more robust to the extreme MLEs from the numerical optimization. Comparing Figures 5 and A1, we find the 95% empirical CI width and the average 95% asymptotic CI width are quite comparable for the T_{50} and $R(3000)$ summaries. However, the asymptotic CI summaries are wider for $MTTF$ compared to the empirical CI width. This is caused by a poor approximation of the sampling distribution of $MTTF$ based on a normal distribution from the asymptotic inference. However, the asymptotic CI width for the 10^{th} and 1^{st} percentiles of lifetime (Table 4) are substantially smaller than the empirical CI width obtained from the simulations (Table 3). The smaller percentiles have a more dramatic difference between the asymptotic and simulation results (the 95% empirical CI width is about 10 times wider than the average asymptotic CI width). This indicates the asymptotic approximation does not work well for (extremely) small percentile lifetime summaries, and using the asymptotic results for early failure times tend to substantially underestimate the true associated uncertainty.

The graphical summaries shown in Figures 3–5 as well as Figure A1 in the Appendix allow the pattern of change across different cost structures to be visualized and provide a simple method for approximately interpolating to other intermediate cost structures. Generally, inspecting less often with more units results in the best possible performance with more precise estimates of reliability summaries. The difference becomes larger as the test setup cost becomes more expensive than inspecting individual units. When the cost ratio is smaller than 25, the uncertainty of reliability estimates is quite robust to change in cost ratio, which is evidenced by the relatively flat curves in Figures 4 and 5. Also, the probability-based and time-based strategies are very similar (almost overlapping lines) for cases with moderate cost ratios ($CR_1 < 25$). Only P1 and T1 show some differences in the reliability summaries for larger cost ratios ($CR_1 > 25$). The difference in the real cost differs more between inspection plans, but the

Table 4. The average (left columns) and 95% (right columns) relative width for the 95% asymptotic confidence intervals across simulations for T_{50}, $MTTF$, $R(t_m)$, T_{10}, and T_1 for all combinations of the six inspection plans and five cost ratios under cost structure 1. Best values (with ties) are shown in bold and second choices are shown in italics for each combination.

	Average Asymptotic CI Width						Upper 95% Asymptotic CI Width					
CR_1	P1	P2	P3	T1	T2	T3	P1	P2	P3	T1	T2	T3
			T_{50}						T_{50}			
0.1	0.504	0.434	*0.357*	0.499	0.433	**0.354**	0.827	0.682	*0.525*	0.824	0.676	**0.514**
1	0.507	0.436	*0.359*	0.504	0.435	**0.355**	0.847	0.687	*0.53*	0.838	0.68	**0.516**
10	0.525	0.445	*0.366*	0.522	0.445	**0.359**	0.89	0.712	*0.544*	0.875	0.698	**0.527**
25	0.564	0.47	*0.376*	0.556	0.463	**0.369**	1	0.761	*0.564*	0.954	0.737	**0.543**
100	1.102	0.661	*0.447*	1.126	0.652	**0.444**	2.612	1.27	*0.726*	2.603	1.224	**0.706**
			$MTTF$						$MTTF$			
0.1	0.715	0.61	*0.508*	0.695	0.601	**0.494**	1.34	1.078	*0.831*	1.289	1.045	**0.797**
1	0.716	0.614	*0.512*	0.703	0.604	**0.496**	1.34	1.075	*0.842*	1.315	1.052	**0.795**
10	0.747	0.627	*0.522*	0.728	0.618	**0.501**	1.444	1.121	*0.863*	1.378	1.074	**0.806**
25	0.811	0.668	*0.536*	0.783	0.644	**0.515**	1.63	1.219	*0.9*	1.533	1.148	**0.841**
100	2.21	0.987	*0.647*	2.246	0.951	**0.63**	5.375	2.184	*1.187*	5	2.047	**1.128**
			$R(t_m)$						$R(t_m)$			
0.1	0.201	0.176	**0.144**	0.204	0.177	*0.145*	0.21	0.183	**0.149**	0.213	0.184	*0.15*
1	0.202	0.176	**0.144**	0.204	0.177	*0.146*	0.211	0.183	**0.149**	0.214	0.184	*0.151*
10	0.209	0.18	**0.147**	0.21	0.182	*0.148*	0.218	0.188	**0.151**	0.22	0.189	*0.153*
25	0.22	0.188	**0.15**	0.223	0.189	*0.152*	0.231	0.196	**0.156**	0.234	0.197	*0.157*
100	0.348	0.249	**0.176**	0.353	0.251	*0.178*	0.374	0.262	**0.183**	0.38	0.265	*0.185*
			T_{10}						T_{10}			
0.1	0.809	0.71	*0.601*	0.8	0.696	**0.584**	1.143	0.961	*0.773*	1.137	0.945	**0.749**
1	0.811	0.708	*0.603*	0.802	0.699	**0.583**	1.149	0.951	*0.773*	1.142	0.941	**0.748**
10	0.839	0.728	*0.612*	0.83	0.718	**0.596**	1.205	0.99	*0.79*	1.195	0.977	**0.773**
25	0.9	0.764	*0.627*	0.887	0.746	**0.612**	1.317	1.059	*0.813*	1.294	1.028	**0.795**
100	1.602	1.048	*0.748*	1.613	1.035	**0.726**	2.98	1.61	*1.025*	3.038	1.594	**0.989**
			T_1						T_1			
0.1	0.157	0.136	**0.111**	0.158	0.136	**0.111**	0.234	0.197	*0.153*	0.236	0.195	**0.152**
1	0.156	0.135	**0.111**	0.159	0.137	**0.111**	0.235	0.195	*0.154*	0.239	0.197	**0.153**
	0.162	0.139	**0.112**	0.163	0.14	*0.113*	0.245	0.202	*0.157*	0.246	0.203	**0.156**
25	0.173	0.146	**0.116**	0.174	0.147	*0.117*	0.269	0.214	*0.164*	0.272	0.214	**0.163**
100	0.299	0.198	**0.137**	0.306	0.2	**0.137**	0.586	0.328	*0.203*	0.594	0.328	**0.201**

largest relative RMSE is still moderate as no more than 12% of the total cost for all cases. For large cost ratios, the RMSE of the cost is within 5% of the total cost.

We now summarize the patterns seen for the other Weibull curves (Curves 2, 3, and 4). Details in the corresponding figures and tables are given in the online supplementary information (SI) A.2, A.3, and A.4, respectively. For Curve 2 with a $Weibull(8000, 1)$ distribution and a constant failure rate, the patterns shown in SM A.2 looks very similar to Curve 1, except the relative RMSE for T_{50} and $MTTF$ increase slightly. As Curve 2 is even closer to a straight line compared to Curve 1 (see Figure 1a), there are even smaller differences between probability-based and the time-based reliability, which make the two strategy types have almost identical testing intervals, and hence very similar performance for reliability summaries. Note that the difference for real cost between the probability- and time-based strategies is smaller compared to Curve 1 results. Inspection frequency is still the main driving factor for differences in estimation precision. Small cost ratios tend to have little impact on the reliability estimates. But as the test setup becomes dramatically more expensive than the individual examination, fewer test units can be evaluated, which results in less precise estimation of reliability summaries. In contrast, the real cost becomes more consistent when the fixed test setup cost increases with higher cost ratio.

For Curve 3 based on a $Weibull(8000, 0.7)$ distribution and a decreasing failure rate, a few more units can be tested with the slightly lower reliability at the end of the test interval (hence more failures and fewer ETIs during the course of the test). Results are shown in the SI A.3. Compared with previous cases especially Curve 2, the reliability decreases faster early and has a bit less linear shape over time, hence there are slightly increased differences between the probability-based and time-based strategies. The probability-based strategies achieve a bit more precision for estimating the model parameter β, as well as

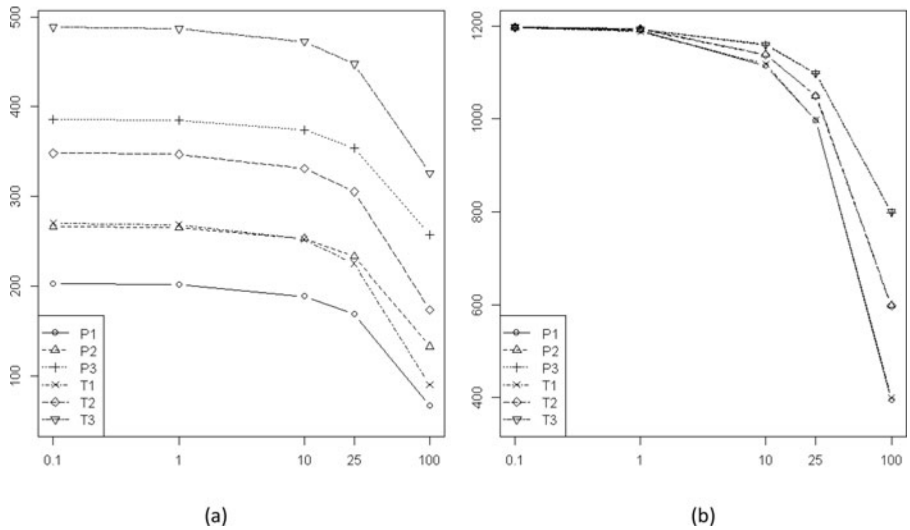


Figure 6. (a) The number of test units and (b) the expected total inspections (ETIs) for the six inspection plans across the five different cost ratios ($CR_1 = C_{ts}/C_{insp}$) under cost structure I for Curve 4 with the *Weibull*(4000, 0.3) distribution.

$MTTF$. The improvement is substantial for estimating $MTTF$ when the cost ratio is larger than 25. The probability-based strategies also have more consistency in the realized cost. Although the time-based strategies seem to estimate $R(3000)$ slightly better for all cost ratios, the differences are very small.

As the reliability curve gets further away from a straight line as shown in Curve 4 with a *Weibull*(4000, 0.3) distribution and a quickly decreasing hazard function, higher failure rates are expected throughout the test period and hence considerably more units are allowed to be tested given the fixed total budget, as shown in Figure 6 (with numerical values available in Table A.4.1 in the SI). However, due to the quickly reducing reliability during early inspection time, the probability-based strategies have very frequent inspections during the first 200 hr and very few inspections after 500 hr (as shown in Figure A2 in the Appendix). As a result, the interval data have fewer failures in the early short inspection time intervals but more failures during the later long inspection intervals, which may result in less precise estimation of reliability shape parameter β. Figure 7 shows the relative RMSE for the seven reliability summaries and the real cost for this scenario. A few observations can be made. First, inspection frequency still dominates the performance for estimating the scale parameter α, T_{50}, and $R(3000)$. However, with the same inspection frequency, the time-based strategies consistently have smaller RMSE than the corresponding probability-based strategies. The difference becomes more substantial as we inspect more often, as more precision is lost with the smaller numbers of failures during early inspections. On the other hand, having more inspections during the early times when the reliability changes very quickly tend to give a more precise estimation of the shape parameter β, the $MTTF$ as well as the early percentiles T_{10} and T_1, which even dominates the impact from the inspection frequency for cases without extremely large cost ratios. This is evidenced by the probability-based strategies with different inspection frequencies all have consistently smaller RMSE than the time-based strategies for $CR_1 < 50$. The patterns are less pronounced for the $MTTF$ plot as the RMSE for plan T1 was substantially inflated due to a few extreme estimates obtained in the simulations. Also the probability-based strategies have consistently smaller RMSE for real cost due to the small uncertainty associated with the failures during the early inspection time intervals. Numerical summaries are available in Table A.4.2 in the SI. Figure 8 shows the relative width of the 95% empirical CIs for Curve 4. Similar patterns can be observed for the precision of all reliability summaries and the real cost. The asymptotic results are available in Section A.4 in the SI, which shows the same general performance across the different inspection paradigms.

In summary, under cost structure I when the test setup and the individual inspections dominate the estimated total cost in Eq. [1], the plans that inspect more units less often tend to offer more precise estimates of reliability quantities. The probability-based and time-based strategies have very similar performance when

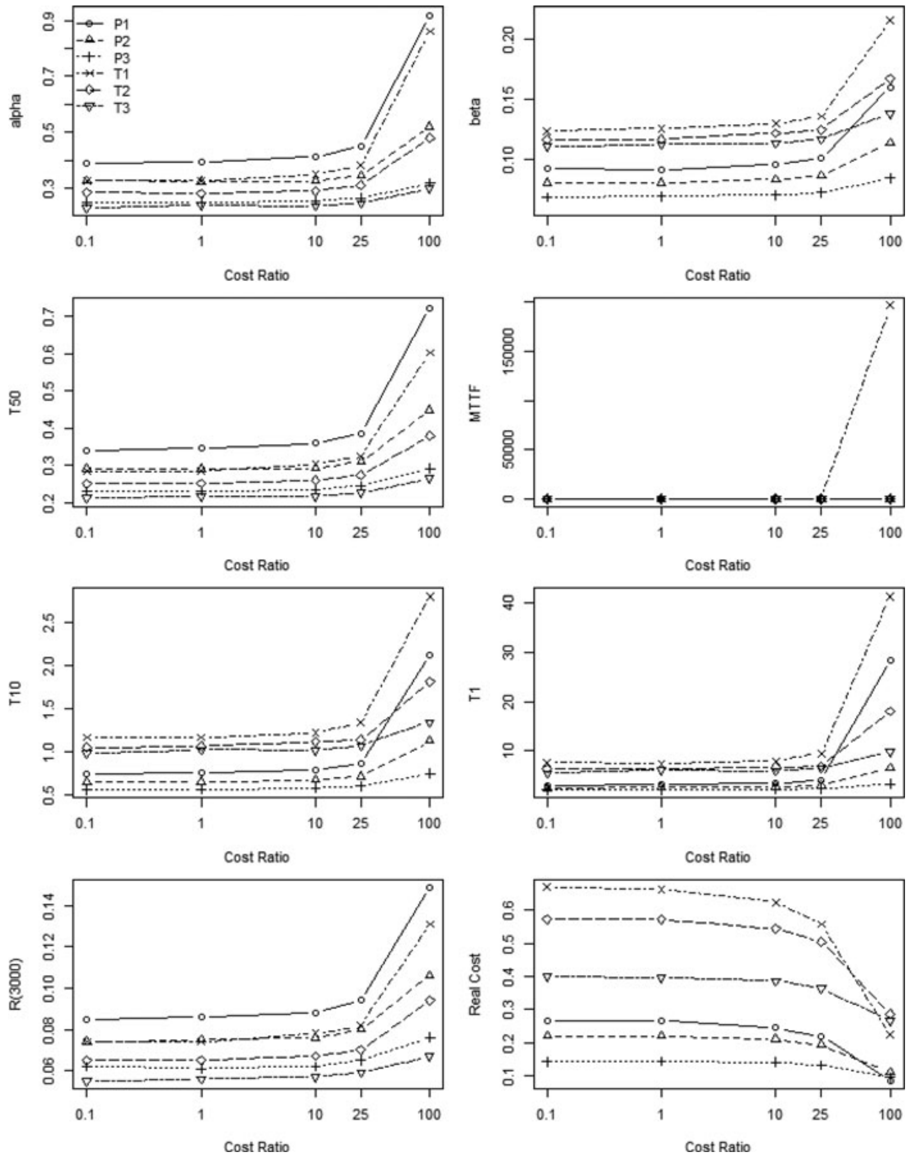


Figure 7. The RMSE (relative to the true value) for the seven reliability summaries and the real cost for the six inspection plans with five cost ratios under cost structure I for Curve 4 with *Weibull*(4000, 0.3) distribution.

the true reliability curve does not differ much from a linear function of time. In this case, the time-based strategy is likely the best choice due to its simplicity of implementation. However, as the reliability curve becomes farther away from a straight line with a sharp decrease in early lifetime, the time-based strategy is still best for estimating the scale parameter α, median lifetime T_{50}, and the reliability at the end of inspection period $R(t_m)$ with generally smaller relative RMSE and width of the empirical and asymptotic CIs for all possible cost ratios. However, the probability-based strategies generally provide better estimates of the shape parameter β, the mean lifetime $MTTF$, and the early percentiles of lifetime such as T_{10} and T_1. This effect is even more important than the inspection frequency when $CR_1 < 50$. As to the variability in the real cost, the time-based plans which inspect more units less often also tend to have smaller relative RMSE for most of the cost ratios. Therefore, using a time-based strategy with a larger number of test units and fewer testing points could be an easily implemented robust choice for most of the scenarios except for a population with very high infant failure rate when there is a focus on the shape parameter β or the average lifetime, $MTTF$, or any early percentile of lifetime such as T_{10} or T_1.

Comparisons under cost structure II

Cost structure II considers a scenario when the test setup C_{ts} is considered negligible and the initial set up cost per unit C_{unit} and the individual inspection

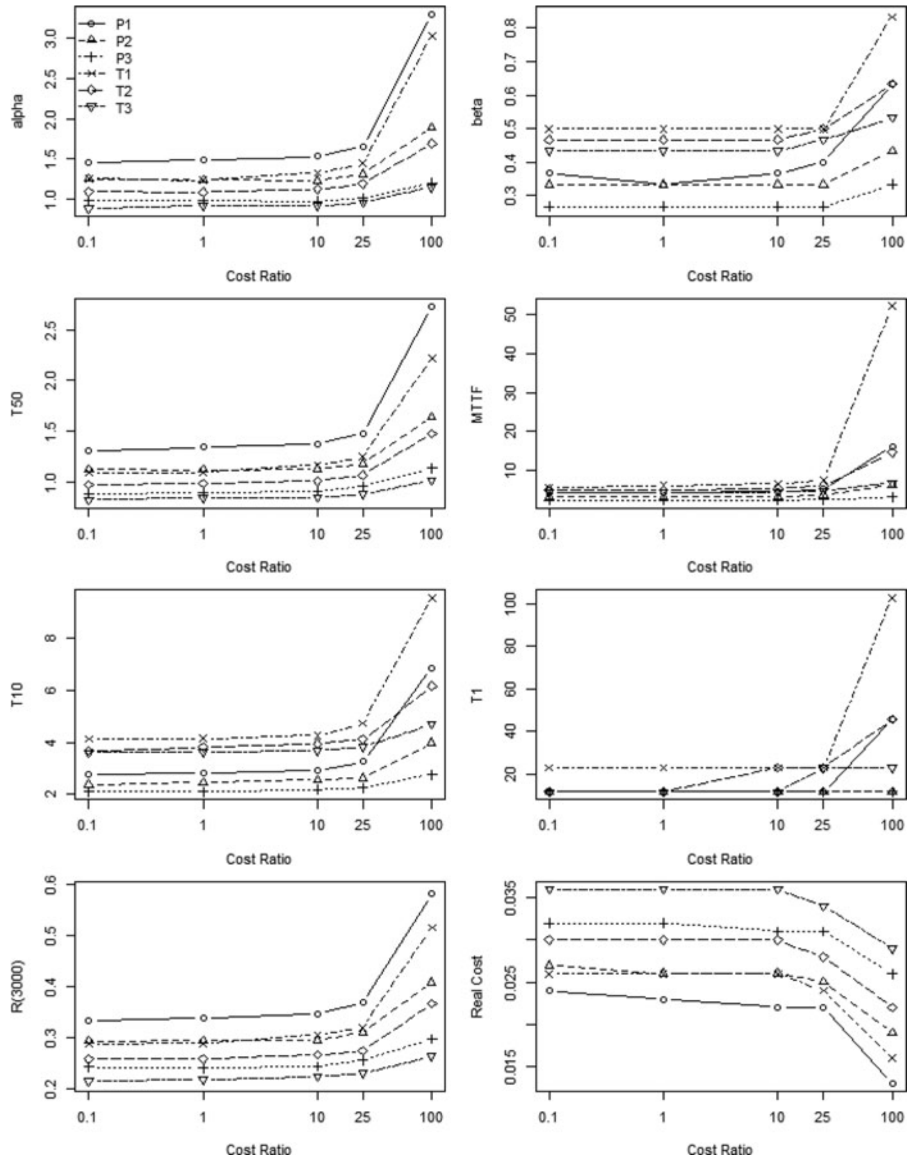


Figure 8. The relative width of the 95% empirical confidence intervals relative to the true value for the seven reliability quantities and the real cost for the six inspection plans with five cost ratios under cost structure I for Curve 4 with *Weibull*(4000, 0.3) distribution.

at each time point, C_{insp}, are the major components contributing to the total cost. Under this cost structure, the total cost for each unit is $C_{unit} + s_i \times C_{insp}$, where s_i represents the actual number of inspections performed for ith unit. As the set up cost per unit, C_{unit}, becomes substantially larger than the cost of inspection at each time point, C_{insp}, the fixed cost per unit will dominate the total cost for each test unit regardless how often the inspections are conducted throughout the test. Hence as the ratio $CR_2 = C_{unit}/C_{insp}$ increases, there are fewer distinctions between the different inspection plans regardless of how often the units are tested. We explore a wide range of possible cost ratios, $CR_2 \in \{0.1, 1, 5, 10, 25\}$. Figure 9a shows the number of test units and the total expected inspections for the six inspection plans for Curve 1 based on the *Weibull*(6000, 1.4) distribution. We can see that when the set up cost per unit is relatively small compared to the individual inspections, inspecting less often allows the experimenter to test a lot more units. However, as the cost ratio increases, the difference in the number of test units possible becomes smaller until there are almost no differences between the six inspection plans. On the other hand, for small CR_2 the total number of inspections starts off very similarly for all six plans, since the cost of the individual inspections dominates the total cost. As the cost ratio increases, the total cost associated with each test unit increases, and hence fewer units can be tested which leads to less precision of reliability estimates. As the fixed cost per unit

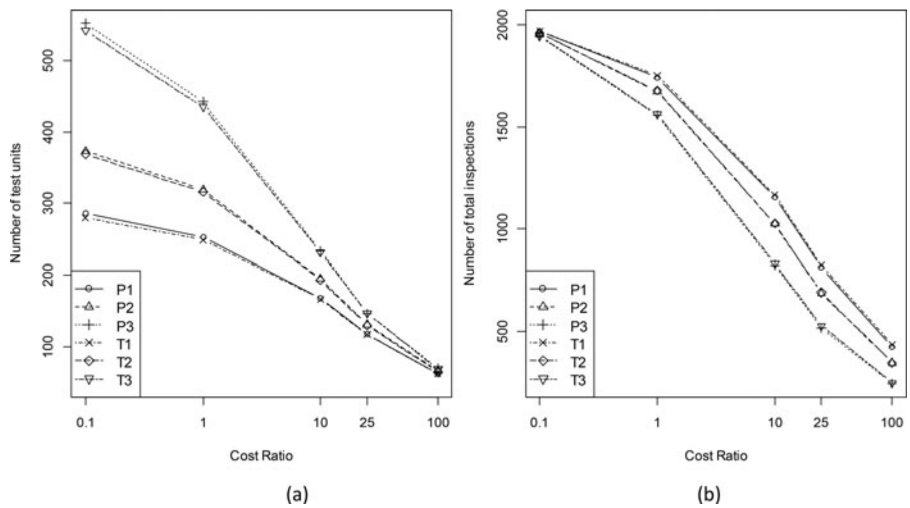


Figure 9. (a) The number of test units and (b) the total expected inspections for the six inspection plans and the five cost ratios under cost structure II for Curve 1 with a $Weibull(6000, 1.4)$ distribution.

dominates the total cost on each test unit, all the plans have essentially similar numbers of test units. As a result, the plans that test the units more often have more precise estimation.

Figure 10 shows the relative RMSE for each of the seven reliability summaries and the real cost for Curve 1. When the cost ratio is small to moderate ($CR_2 \leq 10$), similar to cost structure I, the inspection frequency is the main factor driving changes in the precision of reliability estimates. The plans that inspect more units less often have smaller relative RMSE, and there is little difference between the probability-based and time-based strategies. As the cost ratio increases, estimation precision is lost due to being able to test fewer units. However, as the cost ratio becomes so large that all the plans have similar numbers of test units, then the plans that test more often have more precise estimation. The probability-based plans have similar performance as the time-based plans for true reliability curves that are close to linear during the time of the testing.

In addition, the plans that inspect more units less often tend to have smaller RMSE for the real cost, while the time-based strategies generally have a more consistent real cost than the probability-based strategies with the same inspection frequency. Also, for small to moderate cost ratios, the real cost is generally higher than the expected cost, but the miss from the expected cost is typically smaller as cost ratio increases, which can be explained by the larger fraction of variance relative to MSE for bigger cost ratios (see Table B.1.3 in the SI). For large cost ratios ($CR_2 \approx 25$), the real cost is within 1.5% of the anticipated cost.

Figure 11 shows the relative width of the 95% empirical CIs, which provides an alternate means of quantifying the variability of the quantities. Similar to cost structure I, the variance of all reliability estimates dominate the bias and take about 97–100% of the MSE for all scenarios (see Table B.1.3 in the SI). The empirical CI width also favors inspecting more units less often for small to moderate cost ratios. On the other hand, the plans that inspect less often have more variation for the real cost. The differences diminish as the cost ratio becomes large (close to 25). The probability-based and time-based strategies with the same inspection frequency have similar performance. More details for this scenario are given in the SI.

The results for Curves 2–4 are available in Sections B.2–B.4 in the SI, respectively. The overall patterns are generally consistent with Curve 1 for the relative performance of the different testing strategies, with a few exceptions. As the reliability curves become considerably different from a straight line, the probability-based plans tend to have a slight advantage with more precisely capturing the shape of the distribution (β) compared to the time-based plans. Another difference between the different Weibull distributions is that the probability-based and time-based strategies show more differences for Curve 4 from the other three curves. The time-based strategies can provide slightly more precise estimates for the scale parameter α, median lifetime T_{50}, and the $R(3000)$ given a particular inspection frequency, while the probability-based strategies tend to provide better estimates of the shape parameter β, $MTTF$, T_{10}, and T_1 regardless of the inspection frequency. This effect becomes even more

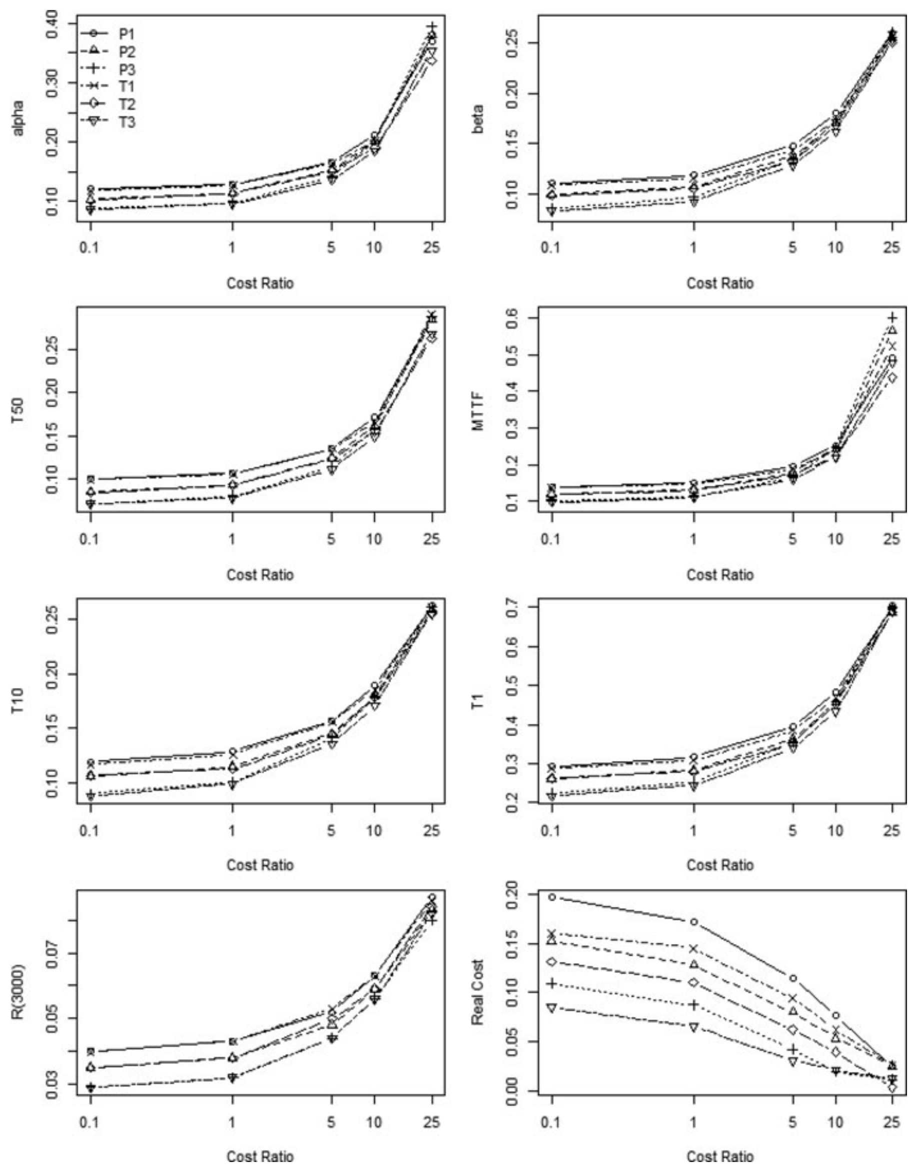


Figure 10. The *relative root of mean squared error* (RMSE relative to the true value) for the seven reliability summaries and the real cost for the six inspection plans with five cost ratios under cost structure II for Curve 1 with a *Weibull*(6000, 1.4) distribution.

important than the inspection frequency for all cost ratios and becomes more prominent for larger cost ratios.

In summary, under cost structure II when the cost per unit and the individual inspections dominate the anticipated total cost in Eq. [1], the plans that inspect more units less often generally offer more precise estimates of the reliability quantities described in this article. The probability-based and time-based strategies have similar performance when the true reliability curve does not differ much from a linear function of time. The time-based strategies are again likely to be preferred for their simplicity of implementation. Note that as the cost ratio increases, fewer units can be tested as the cost per unit takes a larger fraction of the total cost, and hence more precision is lost for all of the reliability estimates across all inspection plans. However, with the diminishing difference in the number of test units across different plans with the dominant cost per unit, the difference in the estimated reliabilities from different inspection plans also becomes very small. The overall difference between the two strategies for inspection for a fixed number of inspection points is generally not large, and if multiple quantities are of interest, there is generally not a severe trade-off between alternatives. Recall, that the probability-based strategies are dependent on an educated guess about the shape of the true underlying lifetime curve. Therefore, using a time-based strategy with more test units

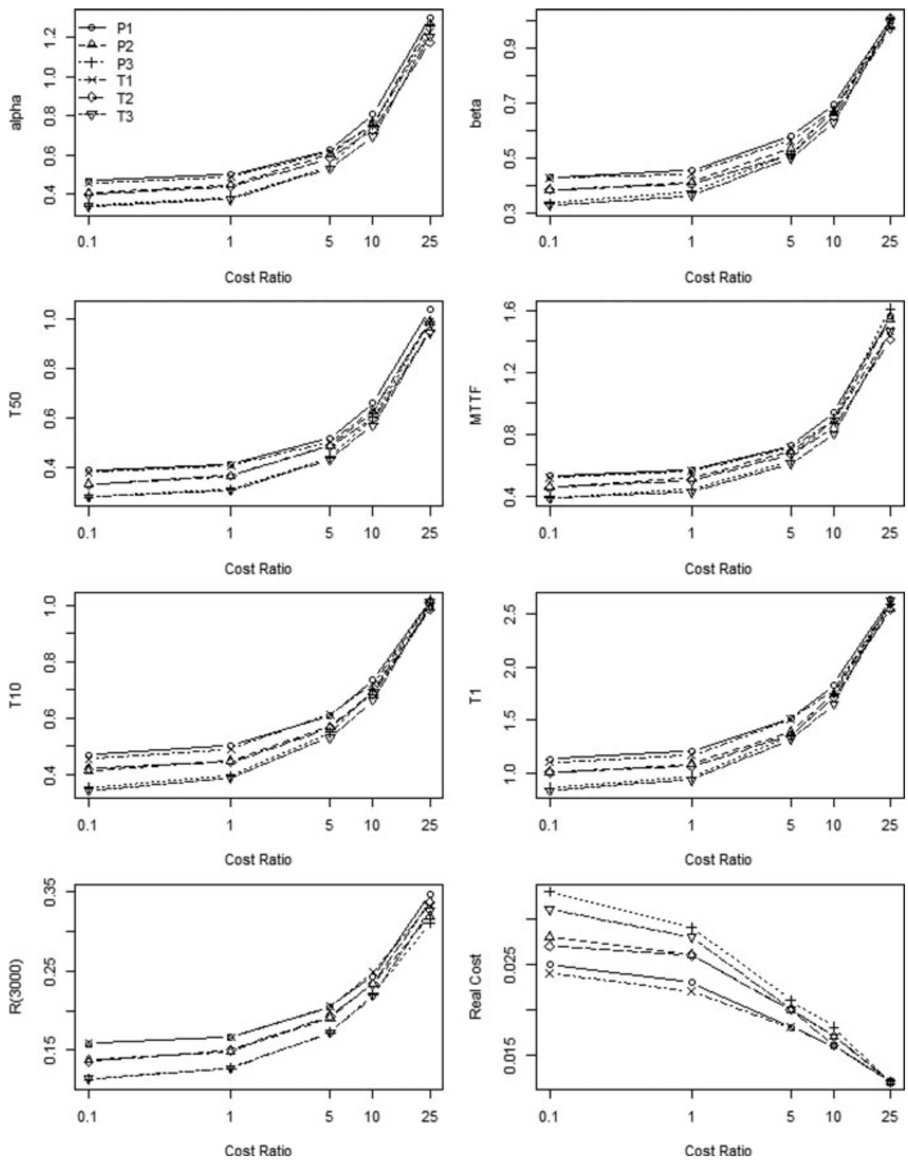


Figure 11. The relative width of the 95% empirical confidence intervals relative to the true value for the seven reliability summaries and the real cost for the six inspection plans with five cost ratios under cost structure II for Curve 1 with a $Weibull(6000, 1.4)$ distribution.

and less frequent inspections could be a convenient and robust choice for most scenarios except for distributions with high infant mortality rates and when the experimenter is primarily interested in good estimation of the shape parameter β, or the average lifetime, $MTTF$, or the early failure times such as T_{10} or T_1.

Discussion and conclusions

This article considers the selection of inspection plans for interval censored data to achieve good reliability estimates subject to a constraint on the total cost. A new general cost structure was proposed to guide the precise quantification of total cost in inspection test planning. The article outlines a variety of different reliability metrics to use as potential criteria for optimization, presents a flexible method for evaluating different alternatives, as well as provides case study results for a variety of different common scenarios. Under this framework, a practitioner should be able to adapt the approach to their particular scenario and find tailored results. The characteristics of interest that were considered were the Weibull scale and shape parameters, median lifetime, mean time to failure, the 10th and the 1st percentiles of lifetime, and estimated reliability at the conclusion of the test. Several main factors that drive change in the estimated reliability for multiple summaries of interest have been evaluated. These include the number of test units, the inspection frequency, and the choice of inspection time points based on achieving evenly spaced probabilities or times. Considering the potentially different impacts from the relative cost of different contributions to the total cost for different applications, we explored two representative case studies considering fixed total budgets with very different cost structures. The main contributors to cost structure I are the cost for test setup at each inspection time point and the cost for individual inspections. Alternately, the main contributors to cost structure II are the cost for the initial setup for each test unit in addition to each individual inspections. A wide variety of cost ratios between the two primary cost components are evaluated for each case study to understand the potentially different impacts on reliability estimates. In addition, the roles of the underlying reliability mechanism, captured by different shapes of reliability curves, are studied by exploring four different Weibull reliability curves with different hazard rates. Also, the assessment of the reliability estimates are conducted based on simulation studies in addition to the asymptotic approximation that was commonly employed in existing work.

Across the diverse combinations of different impact factors, the patterns have many similarities. For both cost structures, when the reliability curves are not too different from a straight line, the dominant factor driving changes in the MSE of the reliability estimates and the real cost for the test is the inspection frequency. Table 5 shows the general results across a variety of

Table 5. Summary of funding with general recommendations.

		Top Inspection Plans				
		Cost Structure I For $CR_1 \in \{0.1, 1, 10, 25, 100\}$*		Cost Structure II For $CR_2 \in \{0.1, 1, 5, 10, 25\}$*		
Reliability Curve	Resp./Metric	RMSE	ECIW	RMSE	ECIW	Comments
Curve 1: * Increasing hazard * Censor rate at 0.68	β $MTTF$ T_{50} T_{10} T_1	Best: T3; Second: P3				1. T3&P3 perform similarly. 2. T3 is preferred for being slightly better for most scenarios and easy to implement.
	R	Best: T3 ≈ P3				
	C	Best: T3 Second: P3	Best: T1≈P1	Best: T3 Second: P3	Best: T1≈P1	
Curve 2: * Constant hazard * Censor rate at 0.68	β $MTTF$ T_{50} T_{10} T_1	Best: P3; Second: T3				1. Despite P3 is slightly better for a lot of scenarios, T3 is generally preferred due to very similar perform with easier implementation
	R	Best: T3≈P3				
	C	Best: T3 Second: P3	Best: T1≈P1 Second: P3	Best: T3	Best: T1≈P1	
Curve 3: * Decreasing hazard * Censor rate at 0.62	β $MTTF$ T_{50} T_{10} T_1	Best: P3; Second: T3				1. T3&P3 still perform similarly, except P3 shows slightly more advantage when estimating β and $MTTF$ 2. T3 is still recommended for easy implementation
	R	Best: T3≈P3				
	C	Best: T3 Second: P3	Best: T1≈P1 Second: P3	Best: T3 Second: T1	Best: P1	
Curve 4: * Fast decreasing hazard * Censor rate at 0.4	β $MTTF$ T_1 T_{10}	Best: P3 Second: P2				1. If T_{50} or R is of interest, T3 is preferred for most precise estimation and easy implementation 2. If β or $MTTF$ or an earlier percentile (T_{10} or T_1) is of more interest, P3 is preferred for most precise estimates
	T_{50} R	Best: T3 Second: P3				
	C	Best: P3 Second: P2	Best: P1 Second: T1	Best: P3 Second: P2	Best: P1 Second: T1	

$R = R(t_m)$, C = Real Cost, $ECIW$ = 95% (relative) empirical confidence interval width.
*Some isolated exceptions.

scenarios considered, and highlights the top inspection plans across the different metrics and cost ratios under each of the cost structures. In general, testing more units less often given a fixed total budget results in more precisely estimated reliability on the multiple characteristics considered. Increasing the cost for each individual unit relative to the other contributors to cost under either cost structure is associated with less precise reliability estimates, but more consistency in the real cost of implementation relative to the expected cost. A pattern for the cost structure I, which resembles some typical stockpile reliability surveillance plans, has the following characteristics: The difference between plans with different inspection frequencies increases as the test setup at multiple inspection times becomes to dominate the total cost. For cost structure II, which is similar to testing electronic parts, there is a diminishing difference between plans as initial setup cost per unit becomes dominating the total cost. Across both cost structures considered, there is not much difference between probability-based and time-based plans for reliability curves that are close to straight line. But for cases with high infant mortality rates, probability-based strategies tend to give more precise estimates of the shape of the distribution, the mean lifetime, and the early failure times (corresponding to small percentiles of lifetime).

Since the probability-based strategies rely on having a good prior assessment of the reliability distribution, but generally do not demonstrate substantial improvement in the precision or accuracy of the reliability estimates (except in cases with high infant mortality rates), the time-based strategies are preferred due to their ease of implementation and good performance for multiple aspects. The time-based plans show good robustness across various model specifications. Therefore, in general, when optimizing over a fixed total budget, we recommend time-based test plans that inspect more units less frequently. The probability-based strategies are only recommended for cases with strong evidence from historical data that the underlying reliability has a high infant mortality rate. In addition, the asymptotic approximation is generally not recommended for quantifying uncertainty for early failure times and small failure proportions as it tends to substantially underestimate the true uncertainty. In this case, the simulation approach illustrated in this article provides more accurate quantification for uncertainty.

About the authors

Lu Lu is an Assistant Professor of Statistics in the Department of Mathematics and Statistics at the University of South Florida in Tampa. She was a postdoctoral research associated in the Statistics Sciences Group at Los Alamos National Laboratory. She earned a doctorate in statistics from Iowa State University in Ames, IA. Her research interests include reliability analysis, design of experiments, response surface methodology, survey sampling, and multiple objective/response optimization.

Christine M. Anderson-Cook is a Research Scientist at Los Alamos National Laboratory in the Statistical Sciences Group. Her research areas include design of experiments, response surface methodology, reliability, multiple criteria optimization, and statistical engineering. She is a Fellow of the American Statistical Association and the American Society for Quality.

References

Casella, G., and R. L. Berger. 2002. *Statistical Inference*. 2nd ed. Belmont, California: Duxbury Press.

Deuflhard, P. 2004. *Newton methods for nonlinear problems. Affine invariance and adaptive algorithms*. Springer Series in Computational Mathematics, Vol. 23. Berlin: Springer.

Goode, H. P., and J. H. Kao. 1960. *"Sampling Plans based on the Weibull Distribution" (No. TR-1)*. Ithaca, NY: Cornell University.

Kim, S. H., and B. J. Yum. 2000. Comparisons of exponential life test plans with intermittent inspections. *Journal of Quality Technology* 32:217–30.

Lavin, M. 1946. Inspection efficiency and sampling inspection plans, *Journal of the American Statistical Association* 41: 432–8.

Meeker, W. Q. 1986. Planning life tests in which units are inspected for failures. *IEEE Transactions on Reliability* 35:571–8.

Meeker, W. Q., and L. A. Escobar. 1998. *Statistical methods for reliability data*. New York: Wiley.

Seo, S. K., and B. J. Yum. 1991. Accelerated life test plan under intermittent inspection and type-I censoring: The case of Weibull failure distribution. *Naval Research Logistics* 38:1–22.

Shapiro, S. S., and S. Gulati. 1996. Selecting failure monitoring times for an exponential life distribution". *Journal of Quality Technology* 28:429–38.

Tobias, P. A., and D. C. Trindade. 2012. *Applied Reliability*. 3rd ed. Boca Raton, FL: CRC Press.

Appendix

Table A1. The fraction of variance relative to the mean squared error for all reliability summaries and real cost for the six inspection plans with five cost ratios for Curve 1 under cost structure I.

	α						β					
CR_1	P1	P2	P3	T1	T2	T3	P1	P2	P3	T1	T2	T3
0.1	0.993	0.993	0.994	0.993	0.993	0.995	0.994	0.995	0.998	0.994	0.997	0.998
1	0.992	0.994	0.994	0.994	0.995	0.996	0.996	0.997	0.998	0.993	0.994	0.998
10	0.992	0.993	0.993	0.992	0.994	0.994	0.996	0.996	0.999	0.995	0.994	0.998
25	0.99	0.993	0.996	0.991	0.996	0.995	0.994	0.995	0.997	0.993	0.993	0.996
100	0.983	0.987	0.99	0.985	0.99	0.991	0.983	0.993	0.997	0.98	0.991	0.998
	T_{50}						$MTTF$					
0.1	0.994	0.994	0.995	0.994	0.995	0.996	0.989	0.99	0.991	0.99	0.99	0.992
1	0.993	0.995	0.995	0.994	0.996	0.997	0.988	0.991	0.992	0.99	0.993	0.994
10	0.993	0.994	0.994	0.994	0.994	0.995	0.987	0.989	0.99	0.989	0.991	0.991
25	0.991	0.994	0.997	0.992	0.997	0.996	0.986	0.989	0.993	0.987	0.994	0.993
100	0.985	0.989	0.992	0.986	0.991	0.993	0.978	0.982	0.986	0.981	0.985	0.987
	T_{10}						T_1					
0.1	0.996	0.997	0.999	0.996	0.998	0.999	0.987	0.989	0.995	0.987	0.992	0.995
1	0.998	0.999	0.999	0.996	0.996	0.999	0.99	0.992	0.994	0.986	0.988	0.995
10	0.999	0.998	0.999	0.997	0.996	0.998	0.991	0.991	0.995	0.988	0.988	0.993
25	0.995	0.997	0.999	0.996	0.996	0.998	0.985	0.989	0.993	0.986	0.987	0.992
100	0.99	0.996	0.999	0.987	0.994	0.999	0.967	0.985	0.992	0.965	0.981	0.994
	$R(t_m)$						Real Cost					
0.1	1	1	1	1	1	1	0.041	0.059	0.1	0.05	0.069	0.123
1	1	1	1	1	1	1	0.041	0.058	0.101	0.05	0.07	0.123
10	1	1	1	1	1	1	0.044	0.058	0.099	0.052	0.071	0.125
25	1	1	1	1	1	1	0.045	0.064	0.101	0.054	0.075	0.13
100	1	1	1	1	1	1	0.071	0.083	0.12	0.102	0.105	0.159

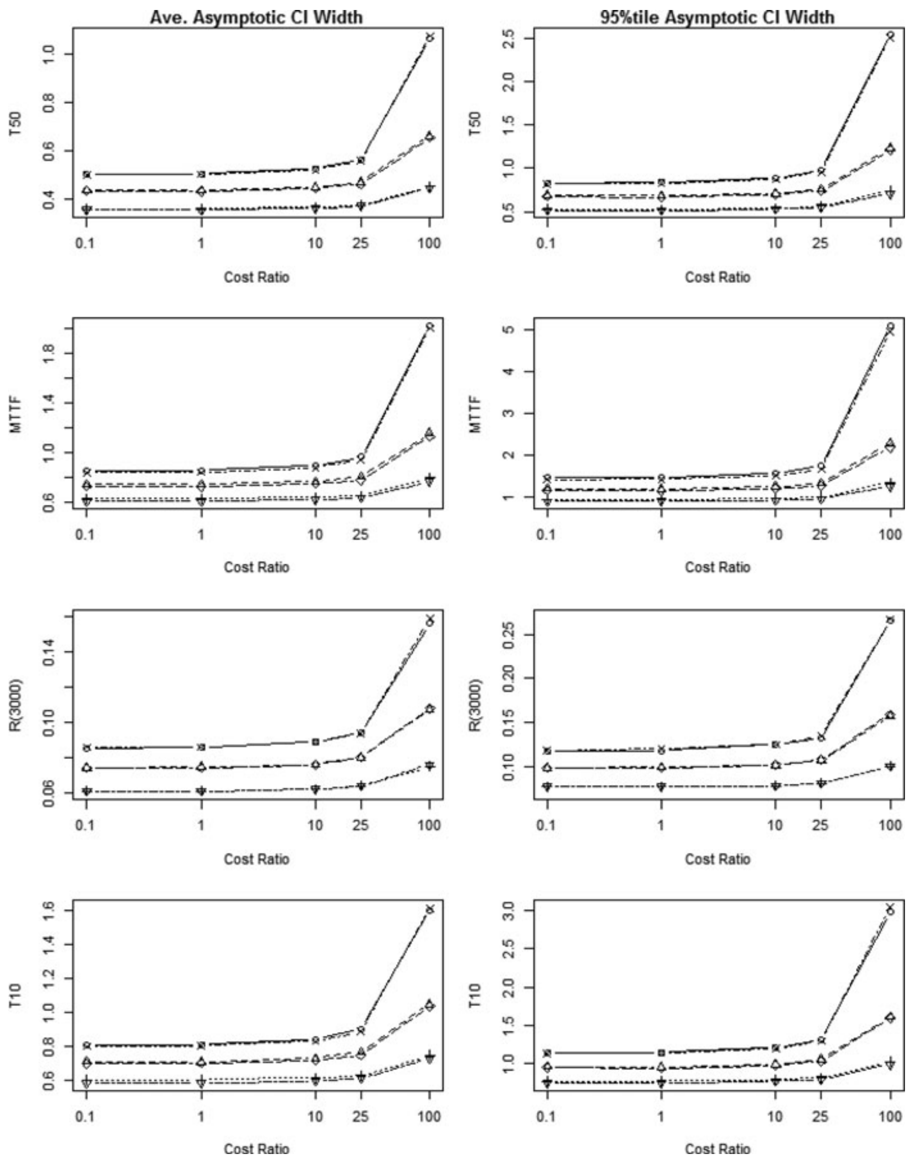


Figure A1. The mean (left column) and 95% (right column) relative width for the 95% asymptotic confidence intervals for T_{50}, $MTTF$, $R(t_m)$, and T_{10} for Curve 1 under cost structure I.

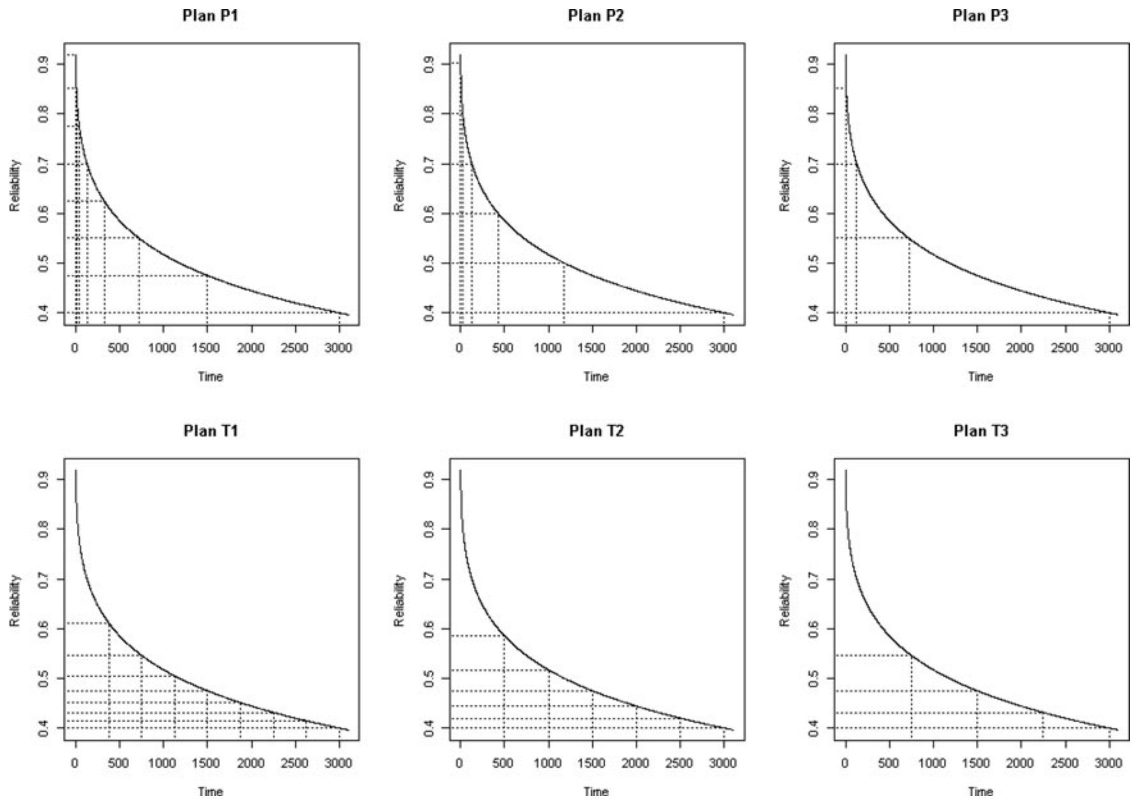


Figure A2. The six inspection plans with different inspection frequencies (8, 6, and 4 times from left to right columns) using the probability-based (first row) or time-based (second row) plans for Curve 4.