Multi-state reliability demonstration tests

Suiyao Chen, Lu Lu & Mingyang Li

Taylor & Francis
Taylor & Francis Group

Check for updates

Multi-state reliability demonstration tests

Suiyao Chen[a], Lu Lu[b], and Mingyang Li[a]

[a]Department of Industrial and Management Systems Engineering, University of South Florida, Tampa, Florida; [b]Department of Mathematics and Statistics, University of South Florida, Tampa, Florida

ABSTRACT

Reliability demonstration tests have important applications in reliability assurance activities to demonstrate product quality over time and safeguard companies' market positions and competitiveness. With greatly increasing global market competition, conventional binomial reliability demonstration tests based on binary test outcomes (success or failure) at a single time point become insufficient for meeting consumers' diverse requirements. This article proposes multi-state reliability demonstration tests (MSRDTs) for demonstrating reliability at multiple time periods or involving multiple failure modes. The design strategy for MSRDTs employs a Bayesian approach to allow incorporation of prior knowledge, which has the potential to reduce the test sample size. Simultaneous demonstration of multiple objectives can be achieved and critical requirements specified to avoid early/critical failures can be explicitly demonstrated to ensure high customer satisfaction. Two case studies are explored to demonstrate the proposed test plans for different objectives.

Introduction

Reliability of a product is the probability that the product can perform its required function at a given time point. As a time-dependent characteristic, reliability is an important measure of the product quality and safety over time, which has a great impact on the satisfaction of customers and can influence their purchase decisions linked with the revenue of manufacturers. In order to succeed in the market competition, manufacturers need to produce products with high reliability over their expected lifetime. Reliability demonstration tests (RDTs) are often conducted by manufacturers to demonstrate the capability of their products to meet the requirements from customers for achieving good quality and performance over time. Given the budget and time constraints, manufacturers need to determine the number of test units, the time duration of the test, and the maximum number of failures allowed to pass the test. These choices are usually made to ensure the consumer's risk (CR) on having a product that has passed the test but fails to meet the reliability requirement is controlled. Controlling the CR at an acceptable level can take the burden off the customers on bearing a high risk of receiving inferior products which are claimed to

have met the requirements on reliability, and hence can help improve customers' satisfaction.

Different categories of RDTs have been studied in the literature based on different types of reliability data, such as failure counts data (Guo et al., 2011; Li et al., 2016; Lu et al., 2016), failure time data (Guo et al., 2012; McKane et al., 2005) and degradation data (Yang, 2009). Failure counts data report the number of failures that occur during a fixed test period. The RDTs based on failure counts data (Wasserman, 2002, pp. 208–210) are also called binomial RDTs (BRDTs) since failure counts are modeled with binomial distributions.

In a BRDT, within a given testing period, if the number of failures does not exceed the maximum number of allowable failures, the test is passed. The maximum number of allowable failures c and the minimum number of test units n are determined to ensure a certain minimum acceptable reliability requirement, R, is met with the controlled CR at or below β by the end of the test duration. The BRDTs are broadly applied in reliability engineering practices because (i) they require less monitoring efforts in the middle of the test duration; and (ii) they are simple and straightforward to be implemented and analyzed. However, with the

increasing needs from customers, the BRDTs are no longer able to meet all requirements in many applications. For example, customers may have varied requirements on reliability performance over different time periods. It is common that many customers have little tolerance of early failures and hence require high reliability during early lifetime and lower reliability for later time. In this case, a BRDT for demonstrating reliability within a single time period is inadequate to meet all requirements.

Consider a scenario when two companies run BRDTs with the same testing period of 5 years and use the maximum number of allowable failures as $c = 5$. Products from company I had 1 failure in the first two years and 3 failures in the last three years. Products from company II had 3 failures in the first two years and 1 failure in the last three years. Even though the products from both companies can pass the demonstration tests, their underlying reliability performance indicated from the failure counts data can be different. If a customer needs products with high reliability in early lifetime (corresponding to allowing no more than 2 failures during the first two years), the risk of the products from company II failing to meet the requirement can be much higher than that of the product from company I. A typical BRDT with five-year testing period cannot demonstrate the performance over the early two years, and hence raises the CR on accepting an inferior product that fails to meet all requirements.

Another limitation of the BRDTs is that they are often used for pass/failure testing of a product without distinguishing the causes and consequences of different failure modes. A product with a complex system is often composed of multiple key components which may have different failure modes associated with varied consequences. Their failures can have different negative effects on the functionality of the entire product. For instance, the failure of the central processing unit (CPU) of a computer is much more crucial than the failure of a video card. Customers may also have different expectations for different components according to their values or costs of replacement. The cost of replacing a CPU or a motherboard is much higher than replacing a keyboard or a mouse. As a result, customers can have much higher expectation on the reliability of the more valuable and critical parts than the reliability of other parts or accessories. A typical BRDT cannot demonstrate separate reliability requirements for multiple failure modes.

To meet the ever-increasing demands of customers, more versatile RDTs with more tailored plans for testing multiple reliability requirements can better serve the customers with enriched information on product reliability. This article proposes RDT strategies for two categories of reliability demonstration tests over multiple time periods and for multiple failure modes, both of which are referred to as multi-state RDTs (MSRDTs) throughout the rest of the article. Alternative test plans within each category are also explored and compared with the conventional BRDTs for demonstrating multiple reliability requirements. Bayesian analysis is used for quantifying the CR associated with various test plans. The Bayesian method offers more flexibilities on incorporating prior information of product reliability from either subject matter expertise or historical data (Pintar et al., 2012; Weaver et al., 2008; Wilson et al., 2016). The impacts of different test strategies and different prior elicitations on the minimum test sample size (i.e., the number of test units required) will be studied to provide more insights on guiding decisions on demonstration test plans. If there exists historical data which supports higher reliabilities compared to the requirements, then using Bayesian method to incorporate prior information has the potential to reduce the minimum test sample size required for the MSRDTs.

The remaining of the article is organized as follows. In the next section, the conventional BRDT plans are reviewed with discussions of their benefits and limitations. Then the new MSRDTs for demonstrating reliability requirements over multiple time periods are proposed. Two different design strategies are proposed and compared under different prior elicitation settings. In the following section, another category of new MSRDT designs for demonstrating reliability requirements involving multiple failure modes are proposed and their performances are evaluated and compared with the conventional BRDTs. Case studies on two categories of MSRDTs for multiple time periods and multiple failure modes are provided to illustrate the proposed test plans and demonstrate their performances. Conclusions and discussions are provided in the end.

Binomial RDTs

For many single use or "one-shot" product units, the test procedure can be destructive. In this case, binomial RDTs (BRDTs) are the common choices to obtain

the failure counts data at the end of a predetermined test period (Kececioglu et al., 2002, pp. 759–768). Let π denote the probability of failure over the test period, and R denote the minimum acceptable reliability at the end of the test duration. In Bayesian analysis, for a chosen number of test units, n, and a maximum number of allowable failures, c, the CR is measured by the posterior probability of the product failing to meet the reliability requirement given that the product has passed the test, which can be calculated as

$$
\begin{aligned}
\mathrm{CR_{binomial}} &= P(\text{Failure probability fails to meet the} \\
&\quad \text{reliability requirement}|\text{Test is passed}) \\
&= P(\pi > 1 - R | y \leq c) \\
&= 1 - P(\pi \leq 1 - R | y \leq c) \\
&= 1 - \frac{\int_0^{1-R} \left[\sum_{y=0}^{c} \binom{n}{y} \pi^y (1-\pi)^{n-y} \right] p(\pi) d\pi}{\int_0^{1} \left[\sum_{y=0}^{c} \binom{n}{y} \pi^y (1-\pi)^{n-y} \right] p(\pi) d\pi}.
\end{aligned}
$$

$$[1]$$

Note that in Eq. [1], $p(\pi)$ denotes the prior distribution of π which can be specified based on subject matter expert knowledge or historical data and y denotes the number of failures observed in the test period. Let β denote the maximum acceptable value for the CR, then a BRDT is determined by choosing the (n, c) combination such that the corresponding $\mathrm{CR_{binomial}} \leq \beta$. According to (Lu et al., 2016), for any fixed choice of c, $\mathrm{CR_{binomial}}$ increases as the test sample size n increases. We use n_b to denote the minimum test sample size that is required to control the CR within an acceptable range $\mathrm{CR_{binomial}} \leq \beta$.

In Bayesian analysis, the $\mathrm{CR_{binomial}}$ in Eq. [1] can be calculated using Monte Carlo integration (Robert et al., 2004, pp. 71–131), where samples of π with a large size $M = 15000$ are generated from the specified prior distribution $p(\pi)$, and $\mathrm{CR_{binomial}}$ is calculated approximately by

$$
\mathrm{CR_{binomial}} \approx 1 - \frac{\sum_{j=1}^{M} \left[\sum_{y=0}^{c} \binom{n}{y} (\pi^{(j)})^y (1-\pi^{(j)})^{n-y} \right] I(\pi^{(j)} \leq 1 - R)}{\sum_{j=1}^{M} \left[\sum_{y=0}^{c} \binom{n}{y} (\pi^{(j)})^y (1-\pi^{(j)})^{n-y} \right]},
$$

$$[2]$$

where $\pi^{(j)}$ is the jth generated sample of failure probability for the specified prior distribution.

Table 1 shows an example of BRDT plans with different choices of prior distributions of π. The mean and standard deviation (i.e., the square of variance) values are provided to give some intuitions about the center

Table 1. Minimum sample sizes required by BRDTs with different choices on c and prior distributions of π.

$\pi \sim$ Beta	(1, 1)	(2, 18)	(4, 16)	(10, 15)	(10, 10)
Mean(π)	0.5	0.1	0.2	0.4	0.5
SD(π)	0.2893	0.0647	0.0873	0.0965	0.1086
c			n_b		
0	13	4	18	45	53
1	18	7	23	51	58
2	24	11	28	57	64
3	29	15	34	62	69
4	34	19	39	68	74
5	39	22	44	74	80
6	44	26	49	80	85

Settings: $M = 15000$, $R = 0.8$, $\beta = 0.05$

and the spread of the prior distributions. For example, $\pi \sim \mathrm{Beta}(1, 1)$ is centered at 0.5 but has large standard deviation at 0.2893. While $\pi \sim \mathrm{Beta}(2, 18)$ has the mean failure probability of 0.1 but much smaller standard deviation (0.0647) around its mean. The minimum acceptable reliability from the consumers requirement was set at $R = 0.8$ and the maximum tolerable CR is chosen to be $\beta = 0.05$. When no historical data or prior information is available, a non-informative prior $\pi \sim \mathrm{Beta}(1, 1)$ can be used. For any assumed prior distribution of π, manufacturers can choose a test plan determined by (n_b, c) using the minimum sample size n_b for any chosen maximum number of allowable failures c. For instance, when $c = 0$ and a non-informative prior $\pi \sim \mathrm{Beta}(1, 1)$ is assumed, the minimum sample size which can ensure the CR calculated in Eq. [2] to be no more than $\beta = 0.05$ is calculated as $n_b = 13$. Hence, at least 13 units need to be tested if the test can only be passed when no failure is observed. However, as larger maximum number of allowable failures being set for passing the test, the CR increases as it becomes easier to pass the test for a given sample size n. Hence, to control the CR at or below $\beta = 0.05$, more units need to be tested as more failures are allowed to pass the test.

When more informative priors are available from historical data or expertise, they can affect the selection of test plans. Table 1 has explored the impacts of different prior distributions $p(\pi)$ on the selected test plan for different tolerances on the maximum number of allowable failures, c. Figure 1 shows the five prior
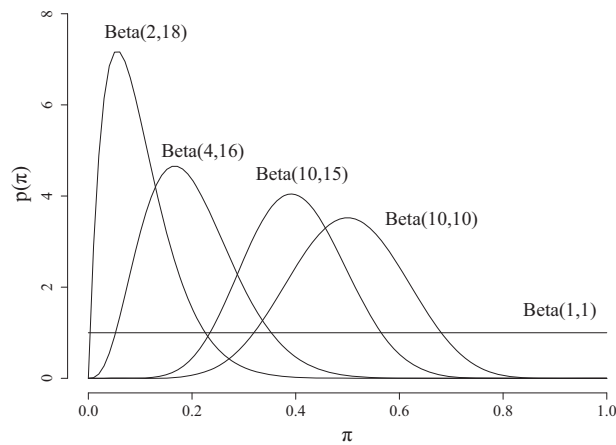
Figure 1. Density curves of different prior distributions explored in Table 1.

distributions explored in Table 1. The flat density curve corresponds to the non-informative prior Beta$(1, 1)$ which assumes that all possible values for $\pi \in (0, 1)$ have equal probability. Other prior distributions from Beta$(10, 10)$ to Beta$(2, 18)$ become more informative with reduced spread (corresponding to smaller standard deviation in Table 1) and provide stronger support for smaller failure probability π. For any given c, the minimum sample size required can be reduced if the prior distribution from historical data supports the reliability requirement. For example, when a prior distribution $\pi \sim$ Beta$(2, 18)$ is used, which supports high reliability around $1 - 2/(2 + 18) = 0.9 > R = 0.8$, fewer units need to be tested to demonstrate the reliability requirement (e.g., $4 < 13$ when $c = 0$). However, if the specified prior distribution is not in favor of the reliability requirement, as illustrated with prior distributions Beta$(4, 16)$, Beta$(10, 15)$, and Beta$(10, 10)$, which favor incrementally lower reliability, more units are required to be tested to demonstrate the same reliability requirement.

On the other hand, Table 2 demonstrates the impact of different requirements on reliability. For a given choice on the prior distribution, as R decreases correspondingly to reduced requirement on reliability, the minimum sample size, n_b, decreases for a fixed choice on c. This matches our intuition that fewer units need to be tested for demonstrating lower requirement on reliability.

The BRDTs are useful for demonstrating reliability requirements for binary tests. For example, a test plan $(n_b = 81, c = 5)$ for a predetermined test period of 5 years can demonstrate no less than 0.9 reliability in 5 years with the CR controlled by 0.05. However, it

Table 2. Minimum sample sizes required by BRDTs with different choices on c and reliability requirements.

	n_b		
c	$R = 0.9$	$R = 0.8$	$R = 0.6$
0	28	13	5
1	39	18	8
2	50	23	11
3	61	28	13
4	71	33	16
5	81	38	18
6	91	43	20

Settings: $M = 15000$, $\beta = 0.05$
$\pi \sim$ Beta$(1, 1)$

offers no capability of demonstrating reliability at any time before the end of the test period. For example, if the customers are particularly concerned about the reliability in the first two years in addition to the reliability by the end of the five years, the conventional BRDTs are unable to demonstrate all requirements over multiple time periods. In addition, BRDTs are unable to differentiate and demonstrate reliability requirements involving multiple failure modes associated with different consequences. In the next two sections, two categories of new MSRDTs are proposed to demonstrate reliability requirements over multiple time periods and for multiple failure modes, respectively. Alternative designs are also proposed and their performances are evaluated and compared under different prior elicitations.

MSRDTs over multiple time periods

Conventional BRDTs often demonstrate the product reliability within a single time period, such as during the mission time or the service life, to meet with the customers' requirements. However, customers' satisfaction in different time periods may differ. For instance, upon the purchase of products, customers may expect higher reliability during the early lifetime. The occurrence of early failures may have stronger negative impact on customers' satisfaction and company's reputation than failures occurred in the later stage of the service period. To explicitly demonstrate different product reliability requirements over multiple time periods rather than a single time period, the strategies of MSRDTs, i.e., multi-state RDTs, are proposed in this section to meet customers' expectation on reliability over multiple time periods.

Consider a finite testing period with the start time at t_0 and the end time at t_K. The testing time duration $(t_0, t_K]$ is exclusively partitioned into K
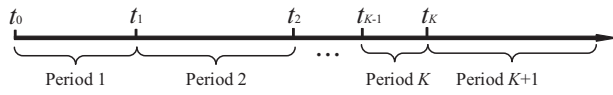
Figure 2. Illustration of the multiple time periods in a K-period MSDRT between $(t_0, t_K]$.

non-overlapping time periods, $(t_{i-1}, t_i]$, $i = 1, \ldots, K$, as illustrated in Figure 2. Let π_i and y_i denote the probability of failure and the number of observed failures within the ith time period $(t_{i-1}, t_i]$, respectively. Then the number of units that survive the entire test duration (right-censored at the end of the test duration t_K) can be expressed as $n - \sum_{i=1}^{K} y_i$, where n is the total number of test units. The probability of surviving the test is given by $\pi_{K+1} = 1 - \sum_{i=1}^{K} \pi_i$. The objective of a MSRDT over multiple time periods is to simultaneously demonstrate the product reliability at multiple time points satisfying a set of lower reliability requirements, R_i, $i = 1, \ldots, K$, with the assurance level controlled at $(1 - \beta)$. Here, R_i is the minimum acceptable reliability in the first i cumulative time periods, $(t_0, t_i]$, β is the maximum acceptable consumer's risk and assurance level can be explained as the minimum probability that the reliability requirements are not met all given the test is passed (Hamada et al., 2008, pp. 343–347). Two different scenarios of acceptance criteria are proposed as follows.

Scenario I. The MSRDT will be passed if the cumulative number of observed failures $\sum_{k=1}^{i} y_k$ at each cumulative time period $(t_0, t_i]$ is no more than its corresponding cumulative maximum number of allowable failures $\sum_{k=1}^{i} c_k$ for all cumulative time periods $(t_0, t_i]$, at $i = 1, \ldots, K$. For example, consider a two-period MSRDT with tests conducted at the end of the second and fifth year. For 100 test units, the MSRDT will be passed if the number of observed failures in first two years do not exceed 1 and the number of observed failures at the end of the fifth year do not exceed 5.

Scenario II. The MSRDT will be passed if the number of observed failures y_i at each non-overlapping time period $(t_{i-1}, t_i]$ is no greater than its corresponding maximum number of allowable failures c_i for all time periods $(t_{i-1}, t_i]$, at $i = 1, \ldots, K$. For the same two-period test, the MSRDT will be passed if the number of observed failures in first two years do not exceed 1 and the number of observed failures in the next three years do not exceed 4. It is noticed that the

major difference between the two scenarios is that Scenario II plans the tests for non-overlapping time periods while Scenario I considers the cumulative time-periods instead.

For each acceptance criterion, the design of MSRDT over multiple time periods aims to determine (i) the minimum sample size, denoted by n_{I} and n_{II} for Scenarios I and II, respectively, and (ii) the cumulative maximum number of allowable failures at time t_i, $\sum_{k=1}^{i} c_k$, for Scenario I and the maximum number of allowable failures within i^{th} time period, c_i, $i = 1, \ldots, K$ for Scenario II. For either scenario, the MSRDT is selected by choosing the test plans which control the CR at or below β. It is noticed that the proposed MSRDT strategies are suitable for demonstration tests that generate failure counts data (Li et al., 2016; Guo et al., 2011) over multiple time periods, and do not make any assumptions on the failure time distribution. The advantages of the proposed methods are to fulfill the reliability requirements of customers over different testing periods (e.g., either cumulative time periods from Scenario I or the non-overlapping periods from Scenario II) simultaneously and provide different testing strategies that require different minimum test sample sizes based on different maximum numbers of allowable failures. Assuming a certain failure time distribution over multiple time periods or for multiple failure modes may limit the use of the proposed strategies because the lifetime distribution assumption has to be valid for the whole test period and only the expected number of failures can be obtained, which is not commensurate with the objectives of proposed strategies as mentioned above. Alternative RDT designs such as Weibull testing which is more suitable for failure time data, is out of the scope of this article, but is of interest for future work.

To illustrate the proposed MSRDTs over multiple time periods and further investigate the difference between two scenarios of acceptance criteria, the MSRDTs over two time periods (i.e., $K = 2$) are considered without loss of generality. Let R_1 and R_2 denote the minimum acceptable reliabilities over the time periods $(t_0, t_1]$ and $(t_0, t_2]$ with $R_2 < R_1$. The probabilities of failure for each cumulative time period meet the requirements if $\pi_1 \leq 1 - R_1$ and $\pi_1 + \pi_2 \leq 1 - R_2$. For acceptance criterion in Scenario I, the test of MSRDT is to determine $\{n_{\mathrm{I}}, c_1, c_1 + c_2\}$, and the probability of accepting the test for any given (π_1, π_2),

denoted by $H_I(n, c_1, c_2)$, can be explicitly written as

$$H_I(n, c_1, c_2) = \sum_{y_1=0}^{c_1} \sum_{y_2=0}^{c_1+c_2-y_1} \left[\left(\frac{n!}{y_1! y_2! (n - y_1 - y_2)!} \right) \right.$$
$$\left. \times \pi_1^{y_1} \pi_2^{y_2} (1 - \pi_1 - \pi_2)^{n-y_1-y_2} \right]$$

and the corresponding CR_I is controlled at or below β by

$$CR_I = 1 - \frac{\int_0^{1-R_1} \int_0^{1-R_2-\pi_1} H_I(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1}{\int_0^1 \int_0^1 H_I(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1} \le \beta,$$

[3]

where $p(\pi_1, \pi_2)$ denotes the joint prior distribution of $(\pi_1, \pi_2, 1 - \pi_1 - \pi_2)$.

For the acceptance criterion in Scenario II, the MSRDT plan can be determined by specifying $\{n_{II}, c_1, c_2\}$, and the probability of accepting the test for any combination of (π_1, π_2), denoted by $H_{II}(n, c_1, c_2)$ is given by

$$H_{II}(n, c_1, c_2) = \sum_{y_1=0}^{c_1} \sum_{y_2=0}^{c_2} \left[\left(\frac{n!}{y_1! y_2! (n - y_1 - y_2)!} \right) \right.$$
$$\left. \times \pi_1^{y_1} \pi_2^{y_2} (1 - \pi_1 - \pi_2)^{n-y_1-y_2} \right],$$

and the corresponding CR_{II} is controlled by

$$CR_{II}$$
$$= 1 - \frac{\int_0^{1-R_1} \int_0^{1-R_2-\pi_1} H_{II}(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1}{\int_0^1 \int_0^1 H_{II}(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1} \le \beta.$$

[4]

A case study is shown below for illustrating the proposed MSRDT strategies for a two-period test. The reliability requirements are set as $R_1 = 0.8$ and $R_2 = 0.6$ over the time periods $(t_0, t_1]$ and $(t_0, t_2]$ with $t_2 < 2t_1$, which indicates longer time interval of $(t_0, t_1]$ than $(t_1, t_2]$. A higher reliability requirement R_1 is desired for the early cumulative time period $(t_0, t_1]$ because the customers are averse to early failures. The CR is controlled at $\beta = 0.05$, indicating that the probability of accepting the test when the actual reliability requirements are not met is controlled at or below 0.05. To evaluate the complex integration in either Eq. [3] or Eq. [4], Monte Carlo sampling is performed with the sample size of $M = 15,000$ to maintain the evaluation accuracy. The Dirichlet distribution, denoted by Dirichlet $(\alpha_1, \alpha_2, \alpha_3)$, is used as

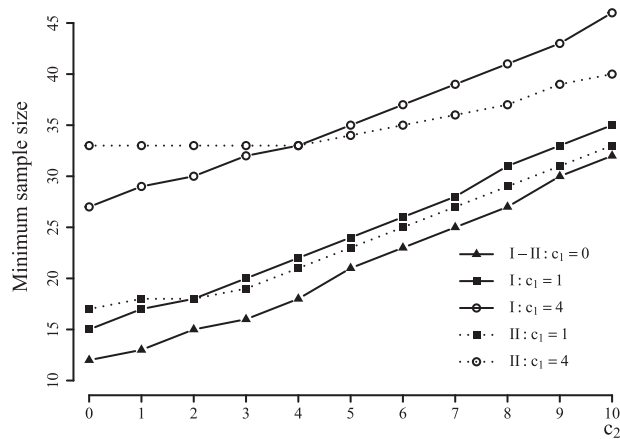**Table 3.** Comparison between Scenarios I and II and BRDT, with non-informative prior.

Scenario I			Scenario II			BRDT	
c_1	$c_1 + c_2$	n_I	c_1	c_2	n_{II}	c	n_b
0	0	12	0	0	12	0	5
0	1	13	0	1	13	1	8
1	1	15	1	0	17		
0	2	14	0	2	14	2	11
1	2	17	1	1	18		
2	2	19	2	0	22		
0	5	20	0	5	20	5	18
1	5	22	1	4	21		
2	5	24	2	3	23		
3	5	26	3	2	28		
4	5	28	4	1	33		
5	5	30	5	0	37		
0	6	22	0	6	22	6	20
1	6	24	1	5	23		
2	6	26	2	4	24		
3	6	28	3	3	28		
4	6	30	4	2	33		
5	6	32	5	1	37		
6	6	34	6	0	42		

Settings: $p(\pi_1, \pi_2) \sim$ Dirichlet$(1, 1, 1)$
$R_1 = 0.8, R_2 = 0.6, M = 15000, \beta = 0.05$

the prior distribution for $(\pi_1, \pi_2, 1 - \pi_1 - \pi_2)$, where $\alpha_1, \alpha_2, \alpha_3$ are hyper-parameters to be elicited based on the prior knowledge. The Dirichlet distribution is a family of continuous multivariate probability distribution parametrized by the vector of positive hyper-parameters $\alpha_i, i = 1, \ldots, K$ for K categories of outcomes. The advantage of using Dirichlet distribution is two folded. First of all, it is the conjugate prior for the multinomial distribution, and hence allows an easy update of knowledge as new data are observed because the posterior distribution of the failure probabilities also follow a Dirichlet distribution. Second, the hyper-parameters in the Dirichlet distribution are associated with more intuitive practical implications as they are directly connected with the failure probabilities for each outcome category based on the prior knowledge in the form of $\alpha_i / \sum_{i=1}^K \alpha_i$. A few different settings of hyper-parameters will be explored later to investigate the impact of prior knowledge on the performance of the proposed test plan.

When no prior information is available, a non-informative prior distribution, given by $(\pi_1, \pi_2, 1 - \pi_1 - \pi_2) \sim$ Dirichlet$(1, 1, 1)$ can be used for indicating the lack of prior knowledge. The selected test plans under the acceptance criteria of two scenarios with different choices on the maximum number of allowable failures are illustrated in Table 3. The test plans are grouped based on the total number of failures allowed during the entire test duration. Several features are

Figure 3. Comparison between Scenarios I and II based on the minimum sample size as c_2 increases for some fixed c_1 values.

observed. First of all, under both Scenarios I and II, given a fixed choice of c_2, the minimum sample size n_I or n_{II} increases as c_1 increases. Similarly, given a fixed c_1, n_I and n_{II} also increase with c_2. As for a given fixed number of test units, allowing more failures (i.e., increasing c) can make it easier to pass the test and thus increase the CR. Hence, it requires to test more units to control the CR at a predetermined maximum acceptable level. The patterns of minimum sample sizes can be observed more clearly in Figures 3 and 4.

Figure 3 shows the change in the minimum sample size as c_2 increases for a few selected c_1 values under both scenarios. Solid lines are used for showing Scenario I and dash lines are used for Scenario II. Different symbols are used for displaying different c_1 values. For a fixed c_1 value, the minimum sample sizes under both Scenarios I and II increase as c_2 increases. For example, when $c_1 = 0$, two scenarios are essentially the same in terms of the acceptance criteria. Hence, the same minimum sample size is required
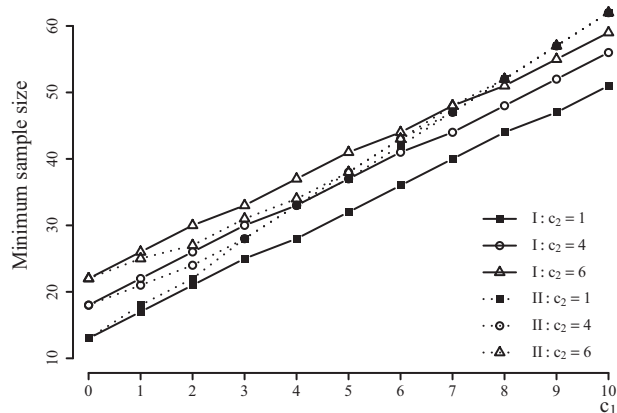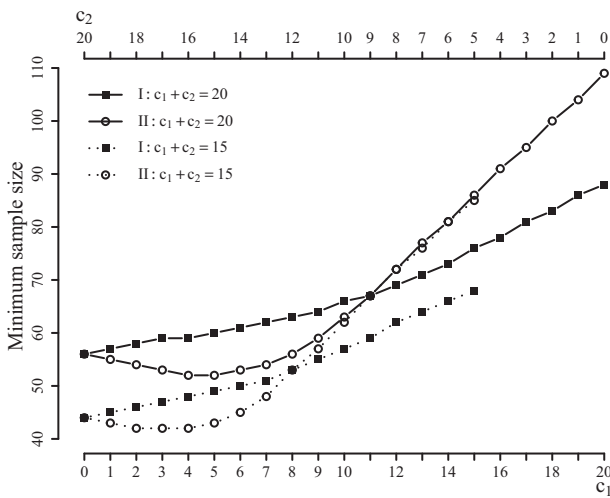

Figure 4. Comparison between Scenarios I and II based on the minimum sample size as c_1 increases for some fixed c_2 values.

for both scenarios, which is shown with the solid line with the triangles and increases as c_2 increases. When $c_1 > 0$, the minimum sample size still generally increases as c_2 increases. However, the trend is slightly different between the two scenarios. The n_I increases monotonically with c_2, while the n_{II} starts off with similar sample sizes for small c_2 values to a certain point and then starts to increase more quickly as c_2 increases. For example, when $c_1 = 4$, the minimum sample size for Scenario II (shown with a dotted line with the open circles) is relatively flat for $c_2 \leq 4$ and then increases for $c_2 > 4$. This is because under Scenario II, the maximum number of allowable failures for the two non-overlapping periods determines their corresponding minimum required test units, which then jointly determine the overall minimum sample size for the entire test. Therefore, the overall sample size can be dominated by the maximum number of allowable failures for one of the test periods if one of the c_i is considerably larger compared to its failure probability under the reliability requirements to be demonstrated. Thus, when c_2 is small, c_1 plays an dominating role in determining the overall sample size for the entire test, which corresponds to the flat portion of the minimum sample size curve for $c_1 = 4$. However, as c_2 becomes larger than c_1, the overall minimum sample size is dominated by the requirement from period 2 and hence resumes an increasing pattern as c_2 increases. To compare the two scenarios, it appears that n_I is usually larger than n_{II} for small c_2 values, but becomes smaller than n_{II} when c_2 becomes larger than a certain value. This is because for the same required c_i values, the test plans in Scenario I generally can allow larger maximum number of allowable failures for period 2 (when the maximum number of allowable failure is not reached during period 1) and hence request to test more units when c_2 has dominating impact on the overall minimum sample size.

Figure 4 shows how the minimum sample size changes with c_1 for fixed c_2 values under both scenarios. Generally, for any fixed c_2, the minimum sample size increases as c_1 increases under Scenario I. Also, a larger c_2 value requires to test more units and the difference in n_I among different c_2 values are similar across different c_1 values, which is evidenced by the almost parallel lines observed for Scenario I in Figure 4. However, for Scenario II, even though n_{II} increases monotonically with c_1, there are diminishing differences in n_{II} at different c_2 values as c_1 increases. This is because under Scenario II, increasing c_1 will affect n_{II} by increasing the minimum sample size

Figure 5. Comparison between Scenarios I and II based on the minimum sample size for fixed $c_1 + c_2$ values.

needed to demonstrate the reliability requirement in period 1 and hence leads to a dominating effect on the size of n_{II} (which is equivalent to a diminishing impact of the difference in c_2 values). While under Scenario I, increasing c_1 will result in increases in the minimum sample sizes needed for demonstrating both reliability requirements at the end of the two non-overlapping time periods, and hence has a consistent impact on the overall minimum sample size n_I.

It is also interesting to compare the two scenarios given the same total maximum number of allowable failures $c_1 + c_2$ in the entire test duration. Figure 5 compares the minimum sample sizes for both scenarios given a fixed total maximum number of allowable failures $c_1 + c_2$. Two cases with $c_1 + c_2 = 15$ and $c_1 + c_2 = 20$ are investigated, which are shown in Figure 5 with the solid and dotted lines, respectively. The bottom and the top axes display all combinations of c_1 and c_2 values. A few patterns can be observed.

First, both n_I and n_{II} increase as $c_1 + c_2$ increases. This matches with the pattern for the conventional BRDTs in that it generally requires to test more units to ensure the same assurance level if a more relaxed criterion has been used for passing the test by allowing more failures to be observed during the entire test duration. Second, increasing c_1 (at the same time reducing c_2) will consistently increase n_I but reduce n_{II} first for small c_2 values and then increase n_{II} after c_2 reaches a certain value. Third, in terms of the relative performance of the two strategies, Scenario II is associated with smaller overall minimum sample size for large c_1 and small c_2 values. As c_2 increasing to about the same size as c_1,

Scenario I starts to have a smaller minimum sample size and the difference becomes larger as c_1 continues to increase. This can be evidenced by the crossing pattern between the monotonically increasing line with the squares for Scenario I and the U-shaped curve with the open circles for Scenario II. Brief analytical explanations can also be found in the Appendix to improve the understanding of the observed differences between two scenarios.

Under the same maximum number of allowable failures $c_1 + c_2$ for the entire test duration, Scenario II is expected to have more strict requirements ($y_1 \leq c_1, y_2 \leq c_2$) than Scenario I ($y_1 \leq c_1, y_1 + y_2 \leq c_1 + c_2$), meaning that any tests that pass in Scenario II will also pass in Scenario I. Intuitively, Scenario II will be preferred if minimizing the CR is the only criterion of interest, which on the other hand generally requires larger minimum sample size. However, smaller test sample is also generally preferred in RDT plan from the manufacturer's point of view. Hence, the tests with minimum sample size after controlling the CR are generally preferred. As illustrated in Figure 5, the two test scenarios may have varied performance in the required minimum sample size for different settings and Scenario II does not consistently outperform Scenario I based on the minimum sample size. It is also noticed in Table 3 that the difference between the two scenarios when c_1 is small becomes smaller for small $c_1 + c_2$ values, and is almost negligible for $c_1 + c_2 \leq 6$. On the other hand, Scenario I can be preferred for relatively large c_1 values when $c_1 + c_2$ is large or when only small $c_1 + c_2$ is allowed. It is also noticed that for tests using more strict passing conditions, they are generally associated with smaller probabilities of passing the test (i.e., low acceptance probability) and often higher probabilities for manufacturers to reject the products that actually have met the reliability requirements (Lu et al., 2016). Hence, a decision on the selection of scenarios should be catered for a particular application to meet the objectives of a specific demonstration test.

In addition, Table 3 also shows the comparison between the MSRDT strategies over two time periods with the conventional BRDTs when non-informative prior is used. The last two columns in Table 3 give the maximum number of allowable failures and the minimum sample size for demonstrating the reliability requirement at the end of test duration (i.e., the end of period 2). For any given total maximum number of

allowable failures over the entire test duration, $c = c_1 + c_2$, the conventional BRDTs require to test fewer units for demonstrating only a single reliability at the end of the test. The MSRDTs, on the other hand, gain the capability of demonstrating multiple reliability requirements at different time points at the expense of testing a few more units. However, as $c = c_1 + c_2$ increases, fewer extra units are required to be tested for demonstrating more reliability requirements at multiple time points. For example, for $c = 5$, the conventional BRDT requires to test 18 units to demonstrate reliability at the end of the two-year period as 0.6. To demonstrate an additional higher reliability at the end of the first year at 0.8, both MSRDT strategies require to test at least 20 units with no failure allowed to be observed during the first year. More units need to be tested if more failures are allowed to be observed during the first year.

It is well known that incorporating different prior information may have large impacts on the results in Bayesian analysis. Next, we explore the impact of different prior elicitations on the selected MSRDT plans under both scenarios. Tables 4 and 5 summarize the required minimum sample sizes for the MSRDT plans over two test periods with different choices of prior distributions for Scenarios I and II, respectively. Seven different prior distributions including the non-informative prior, Dirichlet$(1, 1, 1)$, are explored. The patterns are rather consistent across Tables 4 and 5. Under both scenarios, when the prior distribution supports higher reliabilities than the minimum requirements, such as Dirichlet$(3, 3, 24)$ shown in the fourth column in both tables, the minimum sample size can be substantially reduced for any given combination of c_1 and c_2 values than using the non-informative prior (shown in the third column in both tables).

On the other hand, if the prior distribution supports reliabilities at or below the requirements, more units need to be tested to demonstrate the requirements than using the non-informative prior. This can be observed in Figures 6 and 7 which show the minimum sample size for fixed $c_1 + c_2$ under Scenario I and II, respectively. In both figures, the solid lines with triangles represent the sample sizes for different (c_1, c_2) combinations using a non-informative prior. The dash lines with squares show the sample sizes for a prior distribution Dirichlet$(3, 3, 24)$ that supports higher reliabilities than the requirements, which are consistently below the line of non-informative prior. All other prior distributions support reliabilities at or below the requirements, and hence all require to test more units with the corresponding lines located above the line of non-informative prior. The farther the specified prior distribution is to the reliability requirements, the more test units are needed in the MSRDTs over multiple time periods. One special case is the dash line with open circles observed in Figure 7 for a prior distribution Dirichlet$(3, 12, 15)$,

Table 4. Minimum sample sizes required by the two-period MSRDT using the acceptance criterion in Scenario I for different prior distributions.

Dirichlet		$(1, 1, 1)$	$(3, 3, 24)$	$(6, 6, 18)$	$(12, 3, 15)$	$(3, 12, 15)$	$(6, 12, 12)$	$(12, 6, 12)$
c_1	$c_1 + c_2$				n_I			
0	0	12	1	22	58	22	30	67
0	1	13	1	22	58	25	32	66
1	1	15	2	24	62	25	33	69
0	2	14	2	23	58	28	35	65
1	2	17	3	26	62	28	35	69
2	2	19	4	27	66	28	36	72
0	5	20	5	27	60	35	42	66
1	5	22	5	29	65	36	43	69
2	5	24	5	32	69	36	43	73
3	5	26	8	34	73	36	44	76
4	5	28	11	36	76	36	44	79
5	5	30	13	37	79	37	44	81
0	6	22	6	28	60	38	45	67
1	6	24	6	31	65	39	45	70
2	6	26	6	33	70	39	46	73
3	6	28	8	36	74	39	46	77
4	6	30	11	38	77	39	47	80
5	6	32	14	39	81	39	47	82
6	6	34	16	40	83	39	47	84

Settings: $M = 15000$, $R_1 = 0.8$, $R_2 = 0.6$, $\beta = 0.05$

Table 5. Minimum sample sizes required by the two-period MSRDT using the acceptance criterion in Scenario II for different prior distributions.

Dirichlet		(1, 1, 1)	(3, 3, 24)	(6, 6, 18)	(12, 3, 15)	(3, 12, 15)	(6, 12, 12)	(12, 6, 12)
c_1	c_2				n_{II}			
0	0	12	1	22	58	22	30	67
0	1	13	1	22	58	25	32	66
1	0	17	2	27	64	24	33	73
0	2	14	2	23	58	28	35	65
1	1	18	2	27	63	27	35	71
2	0	22	5	32	69	27	37	79
0	5	20	5	27	60	35	42	66
1	4	21	5	29	65	35	42	70
2	3	23	5	33	69	34	43	74
3	2	28	7	38	74	34	44	80
4	1	33	10	43	79	34	45	87
5	0	37	14	48	86	34	47	97
0	6	22	6	28	60	38	45	67
1	5	23	6	30	65	37	45	71
2	4	24	6	34	70	37	45	74
3	3	28	7	38	74	36	46	79
4	2	33	9	43	79	36	47	85
5	1	37	13	48	84	36	49	93
6	0	42	18	54	92	36	51	104

Settings: $M = 15000$, $R_1 = 0.8$, $R_2 = 0.6$, $\beta = 0.05$

which is consistently below the non-informative line indicating smaller minimum sample sizes are required for all (c_1, c_2) combinations. Since the prior distribution regarding period 1 supports higher reliabilities than the requirements, while the prior distribution regarding period 2 supports reliabilities below the requirements, the effects of sample size reduction from period 1 and sample size increase from period 2 may jointly determine the overall minimum sample size, and hence lead to slightly different pattern than what has been observed for other prior distributions.

MSRDTs for multiple failure modes

In the previous section, the MSRDT strategies consider each time period as an individual state for demonstrating specific reliability requirement within the time period. This section proposes a different category of MSRDTs which considers different failure modes as individual states that are often associated with different consequences of failures and different costs of replacement. The conventional BRDTs report dichotomous outcomes (i.e., success and failure) for each test unit,
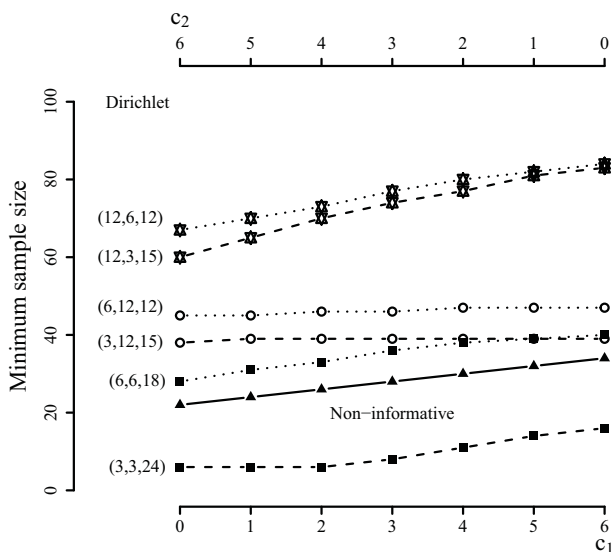
Figure 6. Minimum sample sizes required in Scenario I with fixed $c_1 + c_2 = 6$ for different prior distributions.
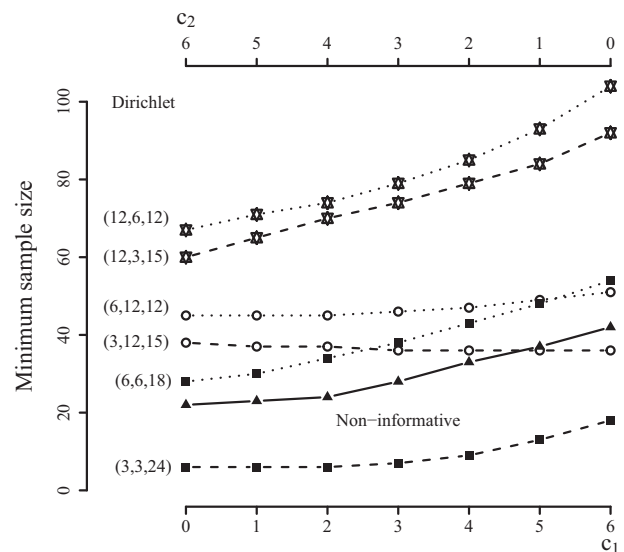
Figure 7. Minimum sample sizes required in Scenario II with fixed $c_1 + c_2 = 6$ for different prior distributions.

in which case different failure modes of the product are not differentiated and the severity levels of different consequences associated with different failures modes are overlooked. In real applications, a product often has multiple failure modes in varied levels of severity, which can lead to different impacts on customers' dissatisfaction.

For instance, the failure of a CPU or a hard drive of a computer system is much more critical than the failures of some accessory parts such as a keyboard or a microphone, since the former can lead to a complete break down of the system, a loss of valuable information and/or a high repair/replacement cost while the latter usually only results in system under-performance and a low repair/replacement cost. Consequently, the failures of critical or valuable parts will lead to stronger dissatisfaction of customers, and hence result in higher expectation on reliability for these components. It is desirable to develop test strategies that allow demonstrating separate reliability requirements for multiple failure modes.

The product with J independent failure modes is considered. For each test unit, it will either have failed in mode j, $j = 1, \ldots, J$ or remain working by the end of the testing period. Let π_j and y_j denote the probability of failure and the number of observed failures in failure mode j within the test period (or an equivalent mission time period), respectively. Then, $\pi_{J+1} = 1 - \sum_{j=1}^{J} \pi_j$ and $n - \sum_{j=1}^{J} y_j$ denote the probability of success and the number of survived units by the end of the test. The MSRDTs for multiple failure modes aim to demonstrate at an assurance level at $(1 - \beta)$ that the product reliability will meet multiple minimum reliability requirements for each of the different failure modes, denoted by R_j, $j = 1, \ldots, J$. Here, β is the CR on having a product that has passed the demonstration test but fails to meet all reliability requirements for different failure modes. Note that all failure modes are defined in the same test period. For any specified reliability requirements R_j's and the maximum acceptable CR controlled at or below β, the MSRDTs for multiple failure modes are designed to determine the minimum sample size n_m as well as the maximum number of allowable failures c_j in the jth failure mode for $j = 1, \ldots, J$.

Without loss of generality, considering two failure modes with $J = 2$ for illustrating the proposed MSRDT strategy. Let R_1 and R_2 denote the minimum acceptable reliabilities for failure modes 1 and 2, respectively.

The test is passed if the number of observed failures y_j is less or equal to the maximum number of allowable failures c_j for both failure modes, and the test plan is to determine the choice on $\{n_m, c_1, c_2\}$. For independent failure modes, the acceptance probability $H_m(n, c_1, c_2)$ for certain (π_1, π_2) values can be written as

$$H_m(n, c_1, c_2) = \sum_{y_1=0}^{c_1} \left[\left(\frac{n!}{y_1!(n - y_1)!} \right) \pi_1^{y_1} (1 - \pi_1)^{n-y_1} \right]$$
$$\times \sum_{y_2=0}^{c_2} \left[\left(\frac{n!}{y_2!(n - y_2)!} \right) \pi_2^{y_2} (1 - \pi_2)^{n-y_2} \right]$$

and the corresponding CR, denoted by CR_m, is calculated by

$$CR_m = 1 - \frac{\int_0^{(1-R_1)} \int_0^{(1-R_2)} H_m(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1}{\int_0^1 \int_0^1 H_m(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1},$$
[5]

where $p(\pi_1, \pi_2)$ is the joint prior distribution of (π_1, π_2). For independent failure modes, there is $p(\pi_1, \pi_2) = p(\pi_1) p(\pi_2)$. The minimum sample size is determined by controlling the CR_m obtained in Eq. [5] to be at or below β. Simulation case studies are conducted for exploring different reliability requirements, maximum numbers of allowable failures for different failure modes, as well as different prior elicitations and their impacts on the required minimum sample size for the MSRDTs for two failure modes. The results are summarized in Tables 6 and 7 for two cases with similar or different reliability requirements for the two failure modes. In Table 6, identical minimum reliability requirements are assumed for the two failure modes, where $R_1 = R_2 = 0.8$ indicates that the customers have the same expectation on reliability for both failure modes. Table 7 assumes different reliability requirements with $R_1 = 0.8$ and $R_2 = 0.6$. Here, failure mode 1 is considered more critical and/or have more severe consequences associated with its failure, and hence is required for a higher reliability. The CR_m is still controlled at $\beta = 0.05$ and the sample size for Monte Carlo sampling is chosen as $M = 15000$ to maintain the simulation accuracy. Beta distributions are used for specifying the prior distributions for both π_1 and π_2 for the two failure modes.

When two failure modes have the same reliability requirements at $R_1 = R_2 = 0.8$, Table 6 summarizes the minimum sample size with different choices of the maximum number of allowable failures and different prior settings. When no prior information is available,

Table 6. Multiple failure modes with the same reliability requirements for different prior distributions.

Beta	π_1 π_2	(1, 1) (1, 1)	(2, 18) (2, 18)	(4, 16) (4, 16)	(10, 15) (10, 15)	(2, 18) (4, 16)	(2, 18) (10, 15)	(4, 16) (10, 15)
c_1	c_2				n_m			
0	0	16	7	22	70	18	48	60
0	1	20	9	25	75	23	55	66
1	0	20	9	25	71	19	45	57
0	2	25	12	29	81	29	61	73
1	1	22	11	28	75	24	52	62
2	0	25	12	30	74	20	45	56
0	5	40	22	43	103	44	81	94
1	4	35	19	39	93	39	70	79
2	3	31	18	36	84	34	62	72
3	2	31	19	36	80	30	56	67
4	1	35	20	40	82	27	50	61
5	0	40	23	45	88	27	45	57
0	6	45	25	47	112	49	88	102
1	5	40	22	43	100	44	76	85
2	4	35	21	40	90	39	68	77
3	3	34	21	39	84	34	61	72
4	2	35	21	40	83	31	55	66
5	1	40	23	45	86	29	50	62
6	0	45	27	50	94	30	46	59

Settings: $M = 15000$, $R_1 = 0.8$, $R_2 = 0.8$, $\beta = 0.05$

a non-informative prior distribution of Beta$(1, 1)$ is assigned for both π_1 and π_2. Similar patterns can be observed as for the MSRDTs over multiple time periods. When c_1 is fixed, the minimum sample size n_m increases as c_2 increases; when c_2 is fixed, n_m increases with c_1. This is intuitive as having more allowable failures makes it easier to pass the test and thus increases

Table 7. Multiple failure modes with different reliability requirements for different prior distributions.

| Beta	π_1 π_2	(1, 1) (1, 1)	(2, 18) (2, 18)	(10, 10) (10, 10)	(4, 16) (10, 15)	(2, 18) (10, 10)	(10, 10) (2, 18)
c_1	c_2			n_m			
0	0	13	3	105	19	17	59
0	1	14	3	101	20	20	63
1	0	19	7	116	25	18	64
0	2	14	3	97	21	23	66
1	1	19	7	111	24	20	68
2	0	24	10	126	31	18	68
0	5	19	5	85	26	31	70
1	4	20	7	99	27	28	74
2	3	24	11	113	30	26	77
3	2	29	14	128	35	24	79
4	1	34	18	143	44	23	80
5	0	38	21	158	56	22	80
0	6	21	6	82	28	33	69
1	5	21	7	95	28	31	75
2	4	24	11	109	30	28	79
3	3	29	14	123	35	26	82
4	2	34	18	138	42	25	84
5	1	39	22	154	52	24	85
6	0	43	25	169	65	24	84

Settings: $M = 15000$, $R_1 = 0.8$, $R_2 = 0.6$, $\beta = 0.05$

the CR. To control a reasonable CR, a larger number of units need to be tested by allowing more failures to be observed during the test. When $c_1 + c_2$ is fixed, the minimum sample size n_m exhibits a symmetric pattern under the non-informative prior setting due to the identical reliability requirements for both failure modes. For example, when $c_1 + c_2 = 6$, the minimum sample sizes for $c_1 = 0, c_2 = 6$, and $c_1 = 6, c_2 = 0$ are identical. In addition, when c_1 and c_2 become more similar in size (e.g., $c_1 = 2, c_2 = 4$ compared to $c_1 = 0, c_2 = 6$), it requires smaller minimum sample size to remain the same assurance level for demonstrating the requirements on both failure modes. This makes sense as when the maximum number of allowable failures is considerably larger for one failure mode given the same reliability requirement, it requires to test more units for demonstrating the requirement for this failure mode, which then inflates the overall minimum sample size needed in the MSRDT for demonstrating reliability requirements for both failure modes.

Different prior elicitations also have large impacts on the selected test plan, as shown in Table 6. When prior knowledge supports higher reliabilities than the requirements to be demonstrated, fewer units need to be tested and vice versa. For instance, the prior distributions of $\pi_1 \sim$ Beta$(2, 18)$ and $\pi_2 \sim$ Beta$(2, 18)$ indicate that there is a strong belief of lower failure probabilities than the requirements within the test period for both failure modes. Thus, the corresponding minimum sample size is smaller than what is needed for using the non-informative prior. On the other hand, when the prior distributions of $\pi_1 \sim$ Beta$(10, 15)$ and $\pi_2 \sim$ Beta$(10, 15)$ are used, which indicates a moderately strong belief in larger failure probabilities than the requirements for both failure modes, more units need to be tested to demonstrate the higher reliability requirements compared to what is needed when no prior information is available.

When $c_1 + c_2$ is fixed, the required minimum sample size is also sensitive to the specified prior distribution. Figure 8 illustrates the change in the n_m for different (c_1, c_2) combinations given fixed $c_1 + c_2 = 6$. When the non-informative priors are assumed, the curve for n_m (the solid line with the triangles) shows a symmetric pattern with the minimum sample size achieved at $c_1 = c_2 = 3$. When informative priors indicating lower failure probabilities than requirements for both failure modes (such as $\pi_1 \sim$ Beta$(2, 18)$, $\pi_2 \sim$
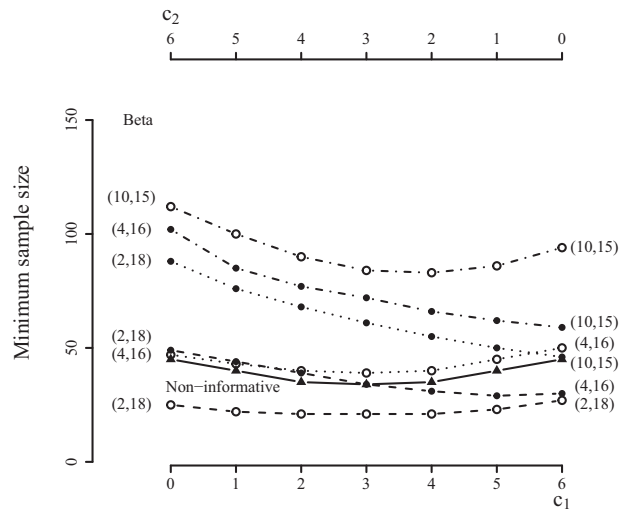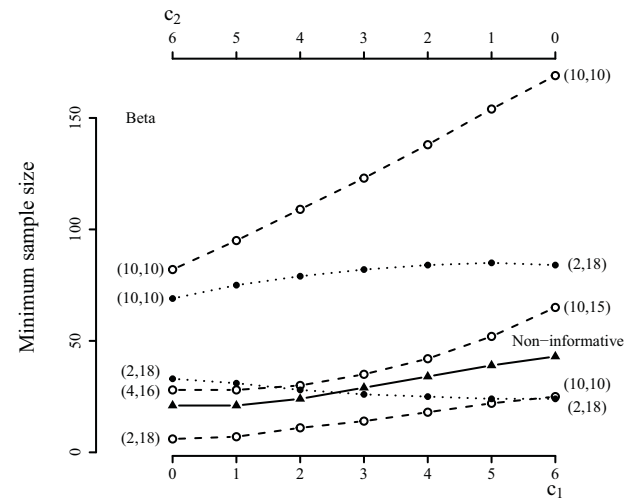
Figure 8. Multiple failure modes with the same reliability requirements for fixed $c_1 + c_2$ and different prior distributions.

Figure 9. Multiple failure modes with different reliability requirements for fixed $c_1 + c_2$ and different prior distributions.

Beta(2, 18) corresponding to the dash line with the open circles) are assumed, the minimum sample size curve is below the non-informative curve. As the prior belief indicates higher failure probability for at least one of the failure modes (such as $\pi_1 \sim$ Beta(2, 18), $\pi_2 \sim$ Beta(10, 15) corresponding to the dotted line with the solid circles or $\pi_1 \sim$ Beta(10, 15), $\pi_2 \sim$ Beta(10, 15) corresponding to the dash-dotted line with the open circles), the corresponding minimum sample size curve shifts upwards on at least one side of tails or on both sides.

Table 7 shows the test plans when different reliability requirements are used for the two failure modes with $R_1 = 0.8$ and $R_2 = 0.6$. When the non-informative priors are used, the symmetric pattern is no longer observed due to different requirements on reliability for the two failure modes. Particularly, n_m is larger when c_1 is large since more units need to be tested to demonstrate higher reliability requirement for failure model 1 while allowing more failures to be observed during the test period. Also, for the same c_1 and c_2 settings, the minimum sample size for demonstrating $R_1 = R_2 = 0.8$ is smaller than what is required for demonstrating $R_1 = 0.8$ and $R_2 = 0.6$ since fewer units can be tested to demonstrate a lower reliability requirement for failure mode 2. When more informative priors are used, similar patterns are observed from both Table 7 and Figure 9. A potential sample size reduction can be achieved when the prior knowledge supports higher reliability than what is required to be demonstrated by the MSRDT.

Conclusions

Conventional binomial RDTs, which focus on demonstrating a single reliability requirement within a single test period, have limited use when multiple reliability requirements need to be met. This article proposes two types of RDTs for demonstrating reliabilities over multiple time periods and for multiple failure modes. These RDTs with multiple reliability requirements are all referred to as multi-state RDTs (MSRDTs).

In the MSRDTs over multiple time periods, every time period of interest is treated as a state, and the joint distribution of failure counts over the non-overlapping time periods can be modeled by a multinomial distribution. Two different test strategies are proposed for demonstrating multiple requirements over different time periods. One strategy uses the cumulative failure counts at the end of each cumulative time period periods as the criteria for passing the test; while the other uses separate failure counts over non-overlapping time intervals as the criteria for passing the test. Simulation studies were conducted for comparing the two strategies by considering two-period MSRDTs. It was founded that the strategy based on cumulative failure counts (Scenario I) is generally preferred for cases that allow fewer total failure counts over all time periods or when a larger maximum number of allowable failures is allowed for the early cumulative time period. The strategy using separate failure counts (Scenario II) is only preferred for requiring smaller minimum sample size when a smaller maximum number of allowable failures is allowed for the early separate time period.

In the MSRDTs for multiple failure modes, each failure mode is treated as a state and all reliability requirements for the multiple failure modes that may be associated with different consequences in varied levels of severity and/or costs of repair/replacement can be simultaneously demonstrated. The required minimum sample size is usually determined mainly by the failure mode that has the highest reliability requirement and/or least stringent criterion for passing the test (i.e., allowing a larger maximum number of allowable failures for a particular failure mode).

The impacts of incorporating different prior distributions are also explored for both categories of MSRDTs. The patterns are consistent regardless of which test strategy is considered. When the prior knowledge supports higher reliabilities than the requirements to be demonstrated, fewer units can be tested compared to using the non-informative priors for demonstrating the same reliability requirements. However, if the historical data supports lower reliabilities than what are required to be demonstrated, then more units need to be tested to override the effects of the prior distribution for demonstrating higher reliabilities than what has been indicated from existing data. For future work, it is expected to develop thorough mathematical justifications with theoritical formulations and derivations to validate the discussed patterns using both non-informative and informative prior distributions.

About the authors

Suiyao Chen is a Ph.D. student in the Department of Industrial and Management Systems Engineering at University of South Florida. He received his B.S. degree (2014) in Economics from Huazhong University of Science and Technology and M.A. degree (2016) in Statistics from Columbia University. His research focus is on statistical reliability data analysis, demonstration tests design, and data analytics.

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 associate 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.

Mingyang Li is an assistant Professor in the Department of Industrial & Management Systems Engineering at the University of South Florida. He received his Ph.D. in systems & industrial engineering and his M.S. in statistics from the University of Arizona in 2015 and 2013, respectively. He also received his M.S. in mechanical & industrial engineering from the University of Iowa in 2010 and his B.S. in control science & engineering from Huazhong University of Science and Technology in 2008. His research interests include reliability and quality assurance, Bayesian data analytics and system informatics. Dr. Li is a member of INFORMS, IISE, and ASQ.

Funding

References

Guo, H., T. Jin, and A. Mettas 2011. Designing reliability demonstration tests for one-shot systems under zero component failures. *IEEE Transactions on Reliability* 60 (1):286–294.

Guo, H., and H. Liao 2012. Methods of reliability demonstration testing and their relationships. *IEEE Transactions on Reliability* 61 (1):231–237.

Hamada, M. S., A. Wilson, C. S. Reese, and H. Martz 2008. *Bayesian reliability*. Springer Science & Business Media.

Kececioglu, D. 2002. *Reliability and life testing handbook*. Vol. 2. DEStech Publications, Inc.

Li, M., W. Zhang, Q. Hu, H. Guo, and J. Liu 2016. Design and risk evaluation of reliability demonstration test for hierarchical systems with multilevel information aggregation. *IEEE Transactions on Reliability* 66 (1):135–147.

Lu, L., M. Li, and C. M. Anderson-Cook 2016. Multiple objective optimization in reliability demonstration tests. *Journal of Quality Technology* 48 (4):303–326.

McKane, S. W., L. A. Escobar, and W. Q. Meeker 2005. Sample size and number of failure requirements for demonstration tests with log-location-scale distributions and failure censoring. *Technometrics* 47 (2):182–190.

Pintar, A., L. Lu, C. M. Anderson-Cook, and G. L. Silver 2012. Bayesian estimation of reliability for batches of high reliability single-use parts. *Quality Engineering* 24 (4):473–485.

Robert, C., and G. Casella 2004. *Monte Carlo statistical methods*. 2nd ed. Springer Science & Business Media.

Wasserman, G. 2002. *Reliability verification, testing, and analysis in engineering design*. CRC Press.

Weaver, B. P., and M. S. Hamada 2008. A Bayesian approach to the analysis of industrial experiments: An illustration with binomial count data. *Quality Engineering* 20 (3):269–280.

Wilson, A. G., and K. M. Fronczyk 2016. Bayesian reliability: combining information. *Quality Engineering* 29 (1):119–129.

Yang, G. 2009. Reliability demonstration through degradation bogey testing. *IEEE Transactions on Reliability* 58 (4):604–610.

Appendix

To analytically show the difference between Scenarios I and II in the proposed MSRDTs over multiple time periods when $c_1 + c_2$ is fixed, let $\Delta H(n, c_1, c_2) = H_I(n, c_1, c_2) - H_{II}(n, c_1, c_2)$, which can be explicitly written as

$$\Delta H(n, c_1, c_2) = \sum_{y_1=0}^{c_1} \sum_{y_2=c_2+1}^{c_1-y_1} \left[\left(\frac{n!}{y_1! y_2! (n - y_1 - y_2)!} \right) \times \pi_1^{y_1} \pi_2^{y_2} (1 - \pi_1 - \pi_2)^{n-y_1-y_2} \right].$$

When $c_1 = 0$, $\Delta H(n, c_1, c_2) = 0$ and both scenarios become equivalent, as shown in Tables 3–5. When $c_1 > 0$, $\Delta H(n, c_1, c_2) > 0$, which indicates that the probability of accepting test plan under Scenario II is always smaller than the probability calculated under Scenario I. However, this finding does not imply that for a fixed n, one scenario will always give a consistently higher/lower CR than the other. To justify this, let $A = \int_0^{1-R_1} \int_0^{1-R_2-\pi_1} H_{II}(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1$ and $B = \int_0^1 \int_0^1 H_{II}(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1$, CR_{II} and CR_I can be written as

$$CR_{II} = 1 - \frac{A}{B},$$

$$CR_I = 1 - \frac{A + \Delta A}{B + \Delta B},$$

where $\Delta A = \int_0^{1-R_1} \int_0^{1-R_2-\pi_1} \Delta H(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1$ and $\Delta B = \int_0^1 \int_0^1 \Delta H(n, c_1, c_2) p(\pi_1, \pi_2) d\pi_2 d\pi_1$. Then $CR_{II} - CR_I$ is given by

$$CR_{II} - CR_I = \frac{B\Delta A - A\Delta B}{B(B + \Delta B)}.$$

Although $B > A$, as n, c_1 and c_2 vary, ΔA can be larger/smaller than ΔB. Thus, for a fixed sample size n, neither $CR_{II} > CR_I$ nor $CR_{II} < CR_I$ will hold consistently. It also explains results in Figure 5, and Tables 4 and 5 that when controlling CR, one scenario cannot give a consistently larger/smaller minimum sample size than the other scenario.