Taylor & Francis
Taylor & Francis Group

Journal of Quality Technology

A Quarterly Journal of Methods, Applications and Related Topics

ISSN: 0022-4065 (Print) 2575-6230 (Online) Journal homepage: https://www.tandfonline.com/loi/ujqt20

An adaptive two-stage Bayesian model averaging approach to planning and analyzing accelerated life tests under model uncertainty

Xiujie Zhao, Rong Pan, Enrique del Castillo & Min Xie

JOURNAL OF QUALITY TECHNOLOGY
2019, VOL. 51, NO. 2, 181–197
https://doi.org/10.1080/00224065.2019.1571333

RESEARCH ARTICLE

An adaptive two-stage Bayesian model averaging approach to planning and analyzing accelerated life tests under model uncertainty

Xiujie Zhao[a], Rong Pan[b], Enrique del Castillo[c], and Min Xie[a,d]

[a]City University of Hong Kong, Kowloon, Hong Kong; [b]Arizona State University, Tempe, Arizona; [c]The Pennsylvania State University, University Park, Pennsylvania; [d]City University of Hong Kong Shenzhen Research Institute, Kowloon, Hong Kong

ABSTRACT
Accelerated life testing (ALT) is commonly used to predict the lifetime of a product at its use stress by subjecting test units to elevated stress conditions that accelerate the occurrence of failures. For new products, the selection of an acceleration model for planning optimal ALT plans is challenging due to the absence of historical lifetime data. The misspecification of an ALT model can lead to considerable errors when it is used to predict the product's life quantiles. This article proposes a two-stage Bayesian approach to constructing ALT plans and predicting lifetime quantiles. At the first stage, the ALT plan is optimized based on the prior information of candidate models under a modified V-optimality criterion that incorporates both asymptotic prediction variance and squared bias. A Bayesian model averaging (BMA) framework is used to derive the posterior model and the posterior distribution for the life quantile of interest under use stress. If the obtained test data cannot provide satisfactory model selection results, an adaptive second-stage test is conducted based on the posterior information from the first stage. A revisited numerical example demonstrates the efficiency and robustness of the resulting Bayesian ALT plans by comparing with the plans derived from previous methods.

KEYWORDS
design of experiments; lognormal distribution; reliability assessment; robust design; Weibull distribution

1. Introduction

Reliability assessment for highly reliable products has drawn great attention in engineering design and plays a significant role in quality and maintenance management. In reliability studies, the life span of many products is extremely long, making it very difficult to conduct life tests under use conditions. Engineers usually resort to accelerated reliability tests to make inferences of field reliability. Accelerated life testing (ALT) is the experimental process of testing a product by subjecting test units to higher levels of stresses (temperature, voltage, vibration rate, pressure, etc.) than with its use stress, to produce more failures within a limited test duration (Chen, Xu, and Ye 2016).

Belonging to the log-location-scale lifetime distribution families, Weibull and lognormal distributions are widely adopted to model lifetime data. In earlier research of ALT planning, the lifetime prediction for certain products was investigated when either a normal or lognormal lifetime is assumed; see Kielpinski and Nelson (1975) and Nelson and Kielpinski (1976), respectively. Afterward, Weibull and extreme value distributed lifetimes were studied in Nelson and Meeker (1978). In a follow-up work by Meeker (1984), models based on Weibull and lognormal lifetime were compared.

In most previous ALT planning studies, certain acceleration regression models with known parameters were assumed and the optimization of the ALT plan was carried out through minimizing or maximizing a function that involves the Fisher information derived from a given model. There was also some literature that addressed energy or cost-efficient plans (Zhang and Liao 2016). Escobar and Meeker (1986, 1994) gave algorithms to numerically compute the Fisher information of unknown parameters for common log-location-scale lifetime distributions. Recent literature also considered several other aspects of optimization. Monroe et al. (2011) proposed a generalized linear model approach to planning ALTs. Pan and Yang (2014) minimized the variance of life quantile predictor over the entire region of possible use stress and

CONTACT Xiujie Zhao xiujizhao2-c@my.cityu.edu.hk Department of Systems Engineering and Engineering Management, City University of Hong Kong, 88 Tat Chee Rd. Kowloon, Hong Kong.

studied the balance of parameter estimation and life quantile prediction. Freels et al. (2015) compared the quantitative and qualitative ALT and proposed a modified accelerated reliability growth test. More relevant research and the pitfalls of ALTs were summarized in Elsayed (2012) and Meeker and Escobar (1993).

For new products, the pre-assumed values of ALT model parameters can be considerably different from their true values. To deal with parameter uncertainty, Bayesian methods can be employed to plan ALTs with prior information. Polson (1993) evaluated the effects of prior information of the acceleration model on ALT planning. Bayesian approaches to ALT modeling and planning were studied in Müller and Parmigiani (1995) and Erkanli and Soyer (2000). An asymptotic optimal ALT planning method based on the *pre-posterior* information was formulated in Zhang and Meeker (2006). The Bayesian approach to accelerated destructive degradation test planning was introduced in Shi and Meeker (2012). Furthermore, the Bayesian inference approach to step-stress accelerated life tests was studied in Lee and Pan (2012) by assuming exponential lifetime. Sha and Pan (2014) presented a Bayesian analysis for PH model in step-stress ALT under Weibull lifetime assumption.

Even though there existed numerous non-Bayesian or Bayesian approaches to planning reliability tests, very few of them have studied the robustness of these plans to the assumed acceleration model and the assumed lifetime distribution simultaneously. In the literature, Pascual and Montepiedra (2005) compared the ALT plans that minimized asymptotic bias or standard error of the predicted life quantile with Weibull and lognormal candidate models. Sensitivity analysis of the ALT optimal designs with use condition (UC) optimality criterion was carried out in Monroe et al. (2010), based on a generalized linear modeling framework. Yu and Chang (2012) investigated the Bayesian model averaging for different models with the aim of increasing the robustness of ALT plans to model selection. Pascual and Montepiedra (2003) studied model-robust ALT plans based on the weighted asymptotic sample ratio criterion. A simulation-based Bayesian ALT planning method was proposed in Nasir and Pan (2015) for model discrimination. Pan et al. (2015) proposed D and D_s optimal criteria to select the best acceleration model among candidates by employing a generalized linear model. Chen, Tang, and Ye (2016) proposed a robust quantile regression method to analyze heavily censored ALT data. Some older literature that addressed the experimental design problem with respect to model selection and discrimination include Agboto et al. (2010), Atkinson and Fedorov (1975), Dette and Titoff (2009), and Hill (1978). As mentioned in Nelson (2005), the research of robust ALT planning is relatively lacking and needs to be addressed.

To obtain an optimal test plan, we need to know the lifetime distribution and acceleration function. However, it is usually the case that the exact ALT model cannot be given as a priori. For example, when a new product is to be tested, it is very likely that we have no exact information of its lifetime distribution or acceleration model. Therefore, it is important that the selected ALT plan is robust to model uncertainty, and even to lifetime distribution misspecification. In addition, the prediction of the quantity of interest should also address this robustness issue. This motivates us to explore a new way of building a Bayesian framework that provides both ALT planning and lifetime prediction with the consideration of acceleration model and lifetime distribution uncertainties.

In this article, we propose a two-stage ALT planning and life quantile prediction framework from a Bayesian perspective. At the first stage, the plan is optimized based on the prior information of various possible acceleration regression models. The objective is to minimize the asymptotic *pre-posterior* squared error of predicted life quantile of interest. An adaptive second-stage ALT test is planned under a given budget if the test at the first stage cannot give satisfactory posterior results on model selection. To deal with the data from ALT experiments, we use the Bayesian model averaging (BMA) technique to predict the life quantile of interest. Therefore, the robustness of our proposed approach is enhanced at both test planning and data analysis phases.

The remainder of the article is organized as follows. Section 2 gives the model assumptions in the ALT planning problem. The two-stage Bayesian ALT planning and prediction methodology is presented in Section 3. In Section 4, a numerical example is revisited to illustrate the proposed approach and compare the results with those from previous studies through simulation. Section 5 gives sensitivity analysis with respect to prior model probabilities, sample size and adaptive test budget. Section 6 concludes the article and discusses areas for future research.

2. ALT models and assumptions

For a better exposition, we consider an ALT model with only one stress variable z in standardized scale, i.e., $z \in [0, 1]$. It is straightforward to extend our

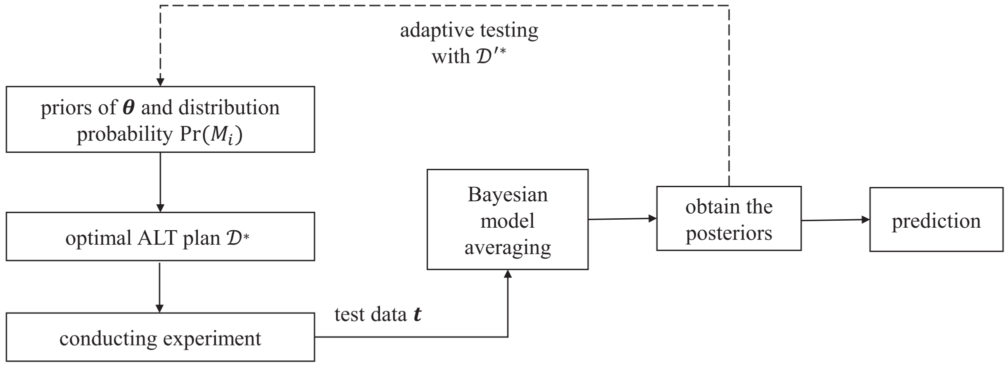

Figure 1. The Bayesian two-stage ALT planning and prediction framework.

proposed framework to multiple stress variables. Let $t_p(z')$ denote the 100pth percentile of the lifetime distribution when the stress level is at $z = z'$.

2.1. Candidate models under Weibull and lognormal distribution

Log-location-scale distributions, such as Weibull, lognormal, and loglogistic distributions, are widely used to model lifetime data in reliability analysis. Without loss of generality, we consider as candidates two log-linear acceleration regression models based on log-location-scale distributions: one assumes that the lifetime follows a Weibull distribution while the other assumes a lognormal distribution. Weibull and lognormal models are denoted by M_1 and M_2, respectively. By assuming a log-linear acceleration regression model, the p quantile of log lifetime at stress level z is as follows:

$$\log\ t_p(z)|M_i = \beta_{0i} + \beta_{1i} z + \sigma_i \Phi_i^{-1}(p), \quad i = 1 \text{ or } 2, \quad [1]$$

where Φ_1^{-1} and Φ_2^{-1} are the quantile functions of standard smallest extreme value (SEV) distribution and standard normal distribution, respectively, and $t_p(z)$ is the life quantile of interest under stress z. In lifetime analysis, we usually set p at a small value, e.g., $p = 0.1$ (i.e., 0.1 quantile, the 10th percentile). In each model, the unknown parameters are denoted by $\boldsymbol{\theta}_i = (\beta_{0i}, \sigma_i, \beta_{1i})'$.

2.2. Acceleration function and the standardization of variables

We standardize the stress levels in the ALT model to make the stress variable range between 0 and 1, where 0 and 1 represent the levels of use stress and highest experimental stress, respectively. In this study, it is assumed that the stress variable is temperature (in °C). The Arrhenius relationship is used to describe the transformation from temperature temp °C to the standardized stress variable z. First, the stress variable ξ of temperature temp °C is given by

$$\xi = \frac{11605}{\text{temp}^\circ\text{C} + 273.15}\ .$$

Let temp_0 be the field use temperature of the product, temp_H as the highest possible temperature in an ALT. ξ is standardized to z as follows:

$$z = \frac{\xi - \xi_0}{\xi_H - \xi_0},$$

where $\xi_H = 11605/(\text{temp}_H{}^\circ\text{C} + 273.15)$ and $\xi_0 = 11605/(\text{temp}_0{}^\circ\text{C} + 273.15)$. By this transformation, the thermal stress variable is coded between 0 and 1.

3. The Bayesian planning and prediction framework

A framework of Bayesian ALT planning and prediction is proposed as shown in Figure 1. To start with, we use the prior information of model selection and parameter settings to plan the ALT experiment by considering both prediction variance and squared bias. Afterward, based on the optimal plan $\mathcal{D}^*$, an ALT experiment is conducted and the testing data are used to proceed to the model averaging step. An adaptive continuous test is planned and carried out if the posterior cannot provide reasonable predictive results. The final model is used to compute the posterior distribution of life quantile of interest by BMA.

Other assumptions are common in the ALT literature and they include:

1. The highest allowed stress level in the test is fixed.
2. Type I right censoring is assumed for the first-stage test. The censoring time is denoted by t_C.

3. The total number of test units N for the first-stage test is predetermined.
4. An arbitrary ALT plan is denoted by $\mathcal{D} = \{(\pi_i, z_i), i = 1, ..., l\}$, where l is the number of stress levels in the plan, π_i is the proportion of test units allocated at stress z_i for $i = 1, ..., l$. We have $\sum_{i=1}^{l} \pi_i = 1$ and $0 < z_i \le 1$ for $i = 1, ..., l$. The elements in $\mathcal{D}$ are the decision variables in the ALT plan optimization.

3.1. Specification of priors

To optimize the ALT plan by Bayesian methods, the prior distribution of $\boldsymbol{\theta}_i$ and the model probability on M_i need to be specified for each candidate model. It is noted that the specification of these priors is analogous to the selection of fixed pilot "planning parameters" in frequentist ALT planning methods. For parameter vector $\boldsymbol{\theta}_i$, the prior information may vary considerably for different products and acceleration models. Specifically, for each log-linear acceleration model in Eq. [1], the slope parameter β_{1i} is determined by the physical or chemical failure mechanism. Therefore, the information on this parameter is usually available from experts and relatively general for products that suffer from similar failures. However, the intercept β_{0i} depends more on the inherent characteristic of certain products, and it may vary significantly among various products although their failure mechanisms are similar, making the information on β_{0i} relatively diffuse.

A test planner needs to specify the priors based on the properties of test units and the information provided by experts and engineers. Available prior information for $\boldsymbol{\theta}_i$ can be quantified in terms of a joint prior distribution of which the density is denoted by $\omega_i(\boldsymbol{\theta}_i)$. The dependency among the elements in $\boldsymbol{\theta}_i$ may be inevitable due to the nature of models. In addition, the quantification of the dependency is usually challenging because such information is not straightforward in engineering senses. Alternatively, a transformed vector $\boldsymbol{\theta}_i^{\diamond}$ can be used to describe several independent sources of prior information. More discussion of this technical point will be given in Section 4.

To describe the prior information on the possible lifetime distributions, we define $\Pr(M_i)$ as the prior probability that model M_i is true, where $\sum \Pr(M_i) = 1$. If $\Pr(M_i) = \Pr(M_j)$ for any $i \neq j$, the prior indicates equal preference to each candidate model.

3.2. Prior-based first-stage ALT planning

The first-stage ALT is planned based on the prior information. To incorporate the uncertainty in the parameters of the acceleration regression model and in the lifetime distribution, we propose a modified V-optimality criterion to optimize the ALT plan from a Bayesian perspective. Based on the priors of $\boldsymbol{\theta}_i$ and given $\Pr(M_i)$ for $i = 1,\ 2$, the optimal ALT plan can be obtained by maximizing the following weighted utility function representing the *pre-posterior* squared error of the predicted life quantile under use stress:

$$U(\mathcal{D}) = -\sum_i \Pr(M_i) \left\{\mathbb{E}_{\boldsymbol{\theta}_i}\left[\hat{C}_i(\mathcal{D})\right] + \mathbb{E}^2_{\boldsymbol{\theta}_i}\left[\text{ABias}\left(\log\ \hat{t}_p(0)\right)|M_i\right]\right\}, \quad [2]$$

for $i = 1$ or 2

$$\hat{C}_i(\mathcal{D}) = \text{AVar}_{\boldsymbol{\theta}_i|\mathbf{t}}\left(\log\ \hat{t}_p(0)|M_i\right) = \mathbf{c}_i'\text{AVar}_{\boldsymbol{\theta}_i|\mathbf{t}}\left(\hat{\boldsymbol{\theta}}_i\right)\mathbf{c}_i,$$

where $\mathbf{c}_i = \left(1, \Phi_i^{-1}(p), 0\right)'$. $C_i(\mathcal{D})$ is the *pre-posterior* asymptotic variance of $\log\ t_p(0)$ under assumption that M_i is true. To compute $\text{AVar}_{\boldsymbol{\theta}_1|\mathbf{t}}(\boldsymbol{\theta}_1)$ and $\text{AVar}_{\boldsymbol{\theta}_2|\mathbf{t}}(\boldsymbol{\theta}_2)$, a large sample approximation is used as in Clyde et al. (1995). Zhang and Meeker (2006) showed that when the sample size was relatively large, a multivariate normal distribution gave a good approximation for the posteriors. Thus, for both models, the following equation holds:

$$\text{AVar}_{\boldsymbol{\theta}_i|\mathbf{t}}\left(\hat{\boldsymbol{\theta}}_i\right) \approx \left[\mathbf{S}_i^{-1} + \ \hat{\mathbf{I}}_{\boldsymbol{\theta}_i}(\mathcal{D})\right]^{-1},\ i = 1 \text{ or } 2, \quad [3]$$

where $\mathbf{S}_i$ is the prior variance-covariance matrix of $\boldsymbol{\theta}_i$, and $\hat{\mathbf{I}}_{\boldsymbol{\theta}_i}(\mathcal{D})$ is the observed information matrix of $\boldsymbol{\theta}_i$ evaluated at its maximum likelihood estimator. By another large sample approximation, we can show that $\hat{C}_i(\mathcal{D})$ can be approximated as $\mathbf{c}_i'\text{AVar}_{\boldsymbol{\theta}_i|\mathbf{t}}(\boldsymbol{\theta}_i)\mathbf{c}_i$ (details are provided in Appendix A). The information matrices $\mathbf{I}_{\boldsymbol{\theta}_i}(\mathcal{D})$ for a given $\mathcal{D}$ and $\boldsymbol{\theta}_i$ are evaluated numerically as described in Appendix B. Afterward, the term $\mathbb{E}_{\boldsymbol{\theta}_i}[C_i(\mathcal{D})]$ is given by

$$\mathbb{E}_{\boldsymbol{\theta}_i}[C_i(\mathcal{D})] = \int C_i(\mathcal{D})\omega_i(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i.$$

To obtain the squared expected bias term $\mathbb{E}^2\left[\text{ABias}\left(\log\hat{t}_p(0)\right)\right]$, we need to decompose it into all misspecification cases. Generally, for $k = 1$ and 2, the squared expected small-sample bias is

$$\mathbb{E}^2_{\boldsymbol{\theta}_i}\left[\text{Bias}\left(\log\hat{t}_p(0)\right)|M_i\right] = \mathbb{E}^2_{\boldsymbol{\theta}_i}\left[\left(\log\hat{t}_p(0)|\textstyle\sum_k \Pr(M_k)M_k\right) - \log t_p(0)|M_i\right], \quad [4]$$

where $\sum_k \Pr(M_k)M_k$ is the assumed weighted model from the prior information, and

$$\log\ \hat{t}_p(0)|\sum_k \Pr(M_k)M_k = \sum_k \Pr(M_k)\left[\log\hat{t}_p(0)|M_k\right].$$

It is noted that $\log\hat{t}_p(0)|\sum_k \Pr(M_k)M_k$ is a random variable because the parameters in M_k are random with assigned prior distributions. The observed bias of the weighted model is expressed as $\left(\log\hat{t}_p(0)\ \ |\sum_k \Pr(M_k)M_k\right) - \log t_p(0)|M_i,$ and the expectation is with respect to both the uncertainty in MLE estimation from ALT data and the model priors. Based on Pascual and Montepiedra (2005), the asymptotic bias can be obtained by

$$\begin{aligned}&\mathbb{E}_{\boldsymbol{\theta}_i}[\text{ABias}(\log\hat{t}_p(0))|M_i]\\ &\quad= \int[(\log t_p(0)|(\sum_k \Pr(M_k)M_k,\ \boldsymbol{\theta}_i, \{\boldsymbol{\theta}_j^*,\ j\neq i\}))\\ &\quad- \log t_p(0)|(M_i,\ \boldsymbol{\theta}_i)]\omega_i(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i.\end{aligned} \quad [5]$$

where $(M_i,\ \boldsymbol{\theta}_i)$ is the assumed true model M_i with parameter $\boldsymbol{\theta}_i$, and $\boldsymbol{\theta}_j^*$ is the expected MLE of $\boldsymbol{\theta}_j$ by assuming that M_j is mis-specified to be the true model. The term $\sum_k \Pr(M_k)M_k, \boldsymbol{\theta}_i, \{\boldsymbol{\theta}_j^*,\ j\neq i\}$ is the asymptotic weighted model with parameters estimated from data under true model $(M_i,\ \boldsymbol{\theta}_i)$. In our problem, by assuming two candidate models M_1 and M_2 to be the Weibull and lognormal linear acceleration regression respectively, the weighted model is $\left(\sum_k \Pr(M_k)M_k,\ \ \boldsymbol{\theta}_i, \boldsymbol{\theta}_{3-i}^*\right)$ for $i = 1$ or 2. The approach to obtaining $\boldsymbol{\theta}^*$ is given in Appendix C.

To evaluate the utility function with a given test plan $\mathcal{D}$, considerable computational effort is needed because of the integration with respect to $\boldsymbol{\theta}_i$ is multi-dimensional in $U(\mathcal{D})$. As an alternative to direct numerical integration, Monte Carlo integration method is employed to compute the utility with a given $\mathcal{D}$. To take expectation on $C_i(\mathcal{D})$ consumes relatively less computation than on $\text{ABias}\left(\log\hat{t}_p(0)\right)$ because in the latter one for every evaluation on a particular true $\boldsymbol{\theta}_i$ a maximum likelihood estimation is called to compute $\boldsymbol{\theta}_j^*$ for $j\neq i$. Therefore, we may have to choose a relatively smaller sample size to conduct Monte Carlo integration for $\text{ABias}\left(\log\hat{t}_p(0)\right)$ and a larger sample for $C_i(\mathcal{D})$.

For simplicity, the utility function in Eq. [2] is rewritten as

$$\begin{aligned}U(\mathcal{D}) &= U_V(\mathcal{D}) + U_B(\mathcal{D}),\\ U_V(\mathcal{D}) &= -\sum_i \Pr(M_i)\mathbb{E}_{\boldsymbol{\theta}_i}[C_i(\mathcal{D})],\\ U_B(\mathcal{D}) &= -\sum_i \Pr(M_i)\mathbb{E}^2_{\boldsymbol{\theta}_i}\left[\text{Bias}\left(\log\hat{t}_p(0)\right)|M_i\right],\end{aligned} \quad [6]$$

where $U_V(\mathcal{D})$ and $U_B(\mathcal{D})$ represents the variance and squared bias part in the utility function. The optimal ALT plan is given by $\mathcal{D}^* = \text{argmax}U(\mathcal{D})$.

3.3. Bayesian model averaging procedure

After the optimal test plan $\mathcal{D}^*$ is found, an ALT experiment is conducted as planned. The experimental data are collected as $\mathbf{t} = (t_1, t_2, ..., t_N)$. For each test unit i for $i = 1,\ 2,\ ...,\ N$, the observed failure time is t_i, and

- z_i denotes the standardized stress level to which the test unit i is allocated.
- δ_i denotes the censoring indicator, where

$$\delta_i = \begin{cases}1 & \text{if observation } i \text{ is not censored}\\ 0 & \text{if observation } i \text{ is right censored}\end{cases}$$

Note that if $\delta_i = 0$, then $t_i = t_C$. Next, we use a Bayesian model averaging (BMA) method to derive the model for life quantile prediction at use stress (for an overview of BMA, see Carlin and Louis 2000). BMA combines the inferences for prediction from different candidate models. For $i = 1$ and 2, after observing data $\mathbf{t}$, BMA gives the posterior distribution of $\log T_P(0)$ as

$$p\left(\log T_p(0)|\mathbf{t}\right) = \sum_i p\left(\log T_p(0)|\mathbf{t}, M_i\right)\Pr(M_i|\mathbf{t}), \quad [7]$$

where $p(\log T_p(0)|\mathbf{t}, M_i)$ is the posterior density of $\log T_p(0)$ by assuming that M_i is the true model; and $\Pr(M_i|\mathbf{t})$ is the posterior probability that M_i is true, which can be computed as follows:

$$\Pr(M_i|\mathbf{t}) \propto \Pr(M_i)p(\mathbf{t}|M_i), \quad [8]$$

where $p(\mathbf{t}|M_i)$ is density of $\mathbf{t}$ under M_i. Since the model parameters are assumed random from the Bayesian perspective, the density is obtained by integrating over the prior of the model parameters:

$$p(\mathbf{t}|M_i) = \int \omega_i(\boldsymbol{\theta}_i)\mathcal{L}(\boldsymbol{\theta}_i|\mathbf{t}, M_i)d\boldsymbol{\theta}. \quad [9]$$

3.4. Evaluation of the posterior prediction

To evaluate the life quantile predictor from Eq. [7], one needs to resort to numerical sampling methods to obtain posteriors as there is no explicit expression for the equation. We need the posterior distributions of $\boldsymbol{\theta}_i$'s to evaluate or $p(\log T_P(0)|\mathbf{t}, M_i)$ in Eq. [7], and the posterior density for $\boldsymbol{\theta}_i$ can be computed as

$$p(\boldsymbol{\theta}_i|\mathbf{t}) = \frac{\mathcal{L}(\boldsymbol{\theta}_i|\mathbf{t}, M_i)\omega_i(\boldsymbol{\theta}_i)}{\int \mathcal{L}(\boldsymbol{\theta}_i|\mathbf{t}, M_i)\omega_i(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i} \quad [10]$$

where $\mathcal{L}(\boldsymbol{\theta}_i|\mathbf{t}, M_i)$ is given by

$$\mathcal{L}(\boldsymbol{\theta}_i|\mathbf{t}, M_i) = \prod_{j=1}^{N}\left[\frac{1}{\sigma_1 z_j}\phi_i\left(\frac{\log t_j - \beta_{01} - \beta_{11} z_i}{\sigma_1}\right)\right]^{\delta_i}$$
$$\left[1 - \Phi_j\left(\frac{\log t_j - \beta_{01} - \beta_{11} z_i}{\sigma_1}\right)\right]^{1-\delta_i}, \quad i = 1 \text{ or } 2, \quad [11]$$

where ϕ_1 and ϕ_2 are the density functions of standard SEV and standard normal distribution, respectively. We use a Markov chain Monte Carlo (MCMC) method to generate random samples from the posterior distribution of $\boldsymbol{\theta}_i$s. Specifically, we use the Metropolis-Hastings (MH) sampling method. Relative details are given in Appendix D.

The BMA posterior probabilities $\Pr(M_i|\mathbf{t})$ in Eq. [8] can be rewritten from Eq. [9] as

$$\Pr(M_i|\boldsymbol{t}) \propto \Pr(M_i)\int \omega_i(\boldsymbol{\theta}_i)\mathcal{L}(\boldsymbol{\theta}_i|\mathbf{t}, M_i)d\boldsymbol{\theta}_i. \quad [12]$$

The integration part $\int \omega_i(\boldsymbol{\theta}_i)\mathcal{L}(\boldsymbol{\theta}_i|\mathbf{t}, M_i)d\boldsymbol{\theta}_i$ represents the expected likelihood of $\boldsymbol{\theta}_i$ with respect to priors under M_i. A straightforward Monte Carlo integration method is used to estimate this part as before. First, generate a large sample of $\boldsymbol{\theta}_{ik}$ from $\omega_i(\boldsymbol{\theta}_i)$ and let the sample size be n_S. Then, the integral part is computed by $(1/n_S)\sum_{k=1}^{n_s}\mathcal{L}(\boldsymbol{\theta}_i|\mathbf{t}, M_i)$.

3.5. The adaptive second-stage planning

The objective of the second-stage planning for the test is to achieve the maximum utility based on the posterior distribution of parameters for a given test budget (TB). The reason for considering the testing cost is that the second-stage experiment is mainly for verification, and test planners are usually unwilling to spend an enormous amount of additional time and money at this stage given that there is already some information from the previous test results.

If the test data from the first-stage test yield a predetermined minimum posterior weight for the preferred model, denoted by α_W, that is,

$$\max_{i=1,\ 2}\left\{\Pr(M_i|\mathbf{t})\right\} \geq \alpha_W. \quad [13]$$

where $\alpha_W > 0.5$, the adaptive second-stage test is exempted. Otherwise, we need to plan the adaptive test under a given budget. In other words, the adaptive test is not necessary if the maximum of the two posterior probabilities is greater than a given threshold α_W, which is chosen by the decision maker based on the subjective tolerance of model uncertainty. In this situation, the test planner can specify another threshold α'_W, which should in general be larger than α_W. If one of the two candidate models has a posterior weight that is greater than α'_W, there is strong evidence to support the model. In the following analysis, it is reasonable to only use the preferred model with posterior weight greater than α'_W to predict the log life quantile. This also applies to the result analysis after the adaptive second-stage test.

Generally, test planners tend to specify a higher α_W if the estimated life quantiles under the two models are believed to differ drastically. For the adaptive ALT plan, the utility function $U'(\mathcal{D}')$ can be expressed as:

$$U'(\mathcal{D}') = U'_V(\mathcal{D}) + U'_B(\mathcal{D}')$$
$$U'_V(\mathcal{D}') = -\sum_i \Pr(M_i|\mathbf{t})\mathbb{E}_{\boldsymbol{\theta}_i|\boldsymbol{t}}\left[C'_i(\mathcal{D}')\right]$$
$$U'_B(\mathcal{D}') = -\sum_i \Pr(M_i|\mathbf{t})\mathbb{E}^2_{\boldsymbol{\theta}_i|\mathbf{t}}\left[\text{ABias}\left(\log \hat{t}_p(0)\right)|M_i\right] \quad [14]$$

To facilitate the testing, the optimal stress levels remain the same as in the prior-based ALT plan, thus the setting of a test chamber does not need to be re-adjusted. At this stage, the decision variables include the censoring time t'_C, the total number of units N' and the optimal test unit allocation to each stress level.

The total test cost is determined by the duration and number of test units in the second-stage test. Let C_I and C_T be the cost per test unit and the cost for running the test for a unit time, respectively. The total cost, denoted by $\text{TC}(N', t'_C)$, is given by $\text{TC}(N', t'_C) = C_I N' + C_T t_C{}'$. The constrained optimization problem is

$$\begin{aligned} \textbf{Minimize} \quad & U'(\mathcal{D}') \\ \textbf{subject to} \quad & \text{TC}\left(N', t'_C\right) \leq \text{TB} \\ & t'_C > 0 \\ & N' \in \mathbb{N}^+ \\ & \mathcal{D}'^* = \left\{\left(\pi'^*_i, z^*_i\right), i = 1, \ldots, l\right\} \end{aligned} \quad [15]$$

The following algorithm is used to determine the optimal adaptive plan:

Algorithm to obtain the optimal adaptive second-stage plan

Step 1: let $N'_U = \lfloor \text{TB}/C_I \rfloor$ be the upper bound of N'.
Step 2: **for** $N' = 1 : N'_U$,

1. set $t'_C = (\text{TB} - C_I N')/C_T$.
2. numerically search for $\mathcal{D}'^*$ that maximizes $U'(\mathcal{D}')$ and calculate $U'(\mathcal{D}'^*)$.
3. if $N' = 1$, let max $U' = U'(\mathcal{D}'^*)$, opt $\mathcal{D}' = \mathcal{D}'^*$, otherwise do the following:

$$\textbf{if } U'(\mathcal{D}'^*) \geq \max U', \text{ let } \max U' = U'(\mathcal{D}').$$

end Step 2
opt $\mathcal{D}'$ is the optimal adaptive plan and the corresponding utility is max U'.

The adaptive second-stage ALT test is carried out based on $\mathcal{D}'^{*}$. If we let $\mathbf{t}'$ be the test data from the adaptive test, then the second-stage posterior of the model parameters can be updated by the same approach as in Section 3.4, thus the test data from both stages are utilized to obtain the posterior prediction. Afterward, the following prediction model is used to make inferences of the log life quantile of the test products,

$$p\left(\log\ T_p(0)|\mathbf{t},\mathbf{t}'\ \right) = \sum_i p\left(\log T_p(0)|\mathbf{t},\mathbf{t}', M_i\right)\Pr\left(M_i|\mathbf{t},\mathbf{t}'\right) \qquad [16]$$

With the above model, the posterior samples of $\log T_p(0)$ can be obtained through MCMC and we can use the samples to build point and interval estimates for $\log T_p(0)$ and therefore $T_p(0)$. As mentioned above, if the test planner find that one model is strongly preferred, i.e., the posterior model weight exceeds α'_W, the other model can be eliminated from the prediction model. By the framework in Figure 1, if the posterior model weights are still unsatisfactory, we can continue to plan more stages of adaptive tests if there is an extra test budget.

From a more general perspective, there could be more than two possible models when planning a reliability test. In the presence of multiple candidate models, it is more difficult to identify the most appropriate one from a test with small sample sizes. In practice, when several models all provide adequate fits to the data, test planners tend to choose the most conservative model in practice, i.e., the one which yields the lowest mean of estimated $t_{0.1}(0)$, to conduct further analysis.

4. Numerical example

The adhesive bond test example from Meeker and Hahn (1985) is revisited to illustrate the proposed method in the article. The objective of the study was to assess the reliability of a type of adhesive bond. The engineers desired to predict the 0.1 life quantile at use temperature 50 °C. The failure process of the adhesive bond was believed to be a simple chemical degradation process, which was well modeled by the Arrhenius relationship. The 0.1 life quantile was expected to be more than several years; therefore, an accelerated life test was needed to predict the quantile. The acceleration regression model is assumed is to be log-linear. After standardization, it is believed that if the lifetime follows a Weibull distribution as in Meeker and Escobar (1998), i.e., M_1, and the pilot fixed parameters for model M_1 are approximately $\boldsymbol{\theta}_1 = (9.36,\ 0.6, -4.65)'$, that is

$$\log t_p(z)|M_1 = 9.36 - 4.65z + 0.6\Phi^{-1}_{SEV}(p) \qquad [17]$$

The previous studies only considered M_1 with fixed $\boldsymbol{\theta}_1$ as a given model and planned the ALT based on this pilot model. To illustrate our approach, suppose lifetimes are actually lognormal distributed but this is unknown to test planners prior to the test, who therefore consider both a Weibull and a lognormal lifetime as possible models. To keep the variability of log life consistent, for the lognormal distribution, the scale parameter is set at $\sigma_2 = 0.77$, because the standard deviation of log life is σ_2 under the lognormal model and $\pi\sigma_1/\sqrt{6}$ under the Weibull model, as shown in Pascual and Montepiedra (2005). It is assumed that there are 300 units available in the test, and the censoring time is six months (183 days). The highest operation temperature in the test is 120 °C.

4.1. Prior specification and first-stage planning

Prior distributions need to be assigned to parameters to quantify the prior knowledge and credibility in $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. In addition, the prior model probabilities $\Pr(M_1)$ and $\Pr(M_2)$ should be specified as well. As discussed in Zhang and Meeker (2006), it is desirable to use the positive parameters $\boldsymbol{\theta}^{\diamond} = \left(t_{0.001}(0), \sigma, -\beta_1\right)'$, instead of $\boldsymbol{\theta}$, to specify the prior information, because $t_{0.001}(0)$, which is the 0.001 life quantile at the use stress level, is approximately independent of σ. Based on the prior information for the product, engineers believed that about 0.1 percent of the items would fail after 6 months at use temperature 50 °C, i.e., $t_{0.001}(0) \approx$ 183 days, but there is much uncertainty in this value. Because $\log[t_{0.001}(0)]$ is on the same scale of σ_i and $-\beta_{1i}$, we set the prior mean of $\log[t_{0.001}(0)]$ equal to 5.2, making $\exp(5.2) = 183$ days. Moreover, we put a large prior standard deviation on $t_{0.001}(0)$. Therefore, the transformed parameters $\boldsymbol{\theta}_i^{\diamond} = \left(t^{(i)}_{0.001}(0), \sigma_i, -\beta_{1i}\right)$ follow independent lognormal distributions as in Zhang and Meeker (2006), and the prior information is summarized in Table 1.

According to the prior information, the mean value for the prior estimation of $\log t_{0.1}(0)$ under M_1 and

Table 1. Prior specification for $\boldsymbol{\theta}_1^{\diamond}$ and $\boldsymbol{\theta}_2^{\diamond}$.

	M_1			M_2		
	$t^{(1)}_{0.001}(0)$	σ_1	$-\beta_{11}$	$t^{(2)}_{0.001}$	σ_2	$-\beta_{12}$
$\mu_{\theta^{\diamond}}$	5.2038	−0.5635	1.5311	5.2038	−0.2940	1.5311
$\sigma_{\theta^{\diamond}}$	1.4995	0.3246	0.1072	1.4995	0.2555	0.1072
Mean	560	0.60	4.65	560	0.77	4.65
SD	1630	0.20	0.50	1630	0.20	0.50

Table 2. Non-Bayesian optimal plans under M_1 and M_2.

	M_1				M_2			
Condition i	Level in [0,1] z_i	Level °C TEMPC	Proportion π_i	Number n_i	Level in [0,1] z_i	Level °C TEMPC	Proportion π_i	Number n_i
Use	0	50	0	0	0	50	0	0
Low	0.68	94.51	0.707	212	0.36	72.13	0.800	240
High	1	120	0.293	88	1	120	0.200	60

Table 3. Optimal plan $\mathcal{D}^*$ to maximize $U(\mathcal{D})$.

Condition i	Level in [0,1] z_i	Level °C TEMPC	Proportion π_i	Number n_i	Expected Failures
Use	0	50	0	0	0
Low	0.54	84.37	0.553	166	95
High	1	120	0.447	134	130
		$U(\mathcal{D}^*) = -0.1095$			

M_2 is 7.99 and 6.60 respectively, which is significantly different under the log scale. If the scale is in years, these correspond to around 8 and 2 years, respectively. This implies that even under the same prior information, the two models yield significantly different prior estimation for the 0.1 lifetime quantile.

For the prior model probabilities, we assume that the test planner has no evidence to prefer either of the models, thus we set $\Pr(M_1) = \Pr(M_2) = 0.5$. It is assumed that the test planner desires a posterior probability of 0.7 or larger on the preferred model from the test data. Otherwise, an adaptive second-stage test is considered necessary.

In the example, we consider a two-level ALT plan, denoted by $\mathcal{D} = (z_1, z_2, \pi_1, \pi_2)$, where z_1 and z_2 are the standardized lower and higher stress levels and π_1 and π_2 are the proportions of test units that are allocated to z_1 and z_2 respectively. Note that the optimal ALT plan is two-level if and only if the log-linear acceleration regression model is correct.

Prior to planning the ALT tests, non-Bayesian optimal plans to minimize the asymptotic variance of the predicted 0.1 life quantile for M_1 and M_2 are derived; they will be compared with the results in the following analysis. We set the pilot parameters as the mean of the respective priors for M_1 and M_2. The optimal plans for M_1 and M_2 are shown in Table 2. The results show that the non-Bayesian optimal plan under M_1 has a significantly higher z_1 and less allocation on z_1 compared with the optimal plan under M_2.

Next, the ALT plan is optimized based on the proposed criterion in the article. By maximizing the utility function in Eq. [2], the Bayesian optimal plan is obtained by numerical search, as presented in Table 3. The surface plot of utility is shown in Figure 2.

From the results, the optimal ALT plan is $\mathcal{D}^* = (0.54,\ 1,\ 0.\ 553,\ 0.447)$. Compared to the optimal

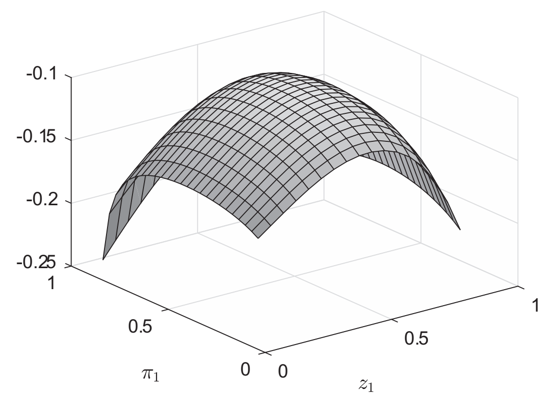

Figure 2. Surface plot of $U(\mathcal{D})$.

plans in non-Bayesian cases in Table 2, one significant difference of the Bayesian optimal plan is that it allocates more test units on the higher stress level and the lower stress z_1 is in between of the two plans in Table 2. The reason behind this allocation scheme is that the available prior information can make the optimal plan extrapolate further from the use condition. One can find more discussions on this in Section 5.1 of Zhang and Meeker (2006). In Figure 3, the contour plots for the integrated utility function $U(\mathcal{D})$ in Eq. [2] as well as the negative weighed *pre-posterior* variance $U_V(\mathcal{D})$ and squared mean bias $U_B(\mathcal{D})$ are given. From the figure, one can see that $U_B(\mathcal{D})$ are much smaller than $U_V(\mathcal{D})$; that is, in this problem, $U_V(\mathcal{D})$ provides the major contribution to the utility function. The figure also shows that $U(\mathcal{D})$ is less sensitive with respect to π_1 because $U_B(\mathcal{D})$ part varies very little with π_1. The change in sample size directly influences the relative importance of $U_V(\mathcal{D})$ and $U_B(\mathcal{D})$, which we will address in a later section.

4.2. Example: Prediction based on the first-stage test

Based on the optimal plan $\mathcal{D}^* = (0.54,\ 1,\ 0.563,\ 0.436)$ with $N = 300$, an accelerated life test is to be conducted. For illustration, Monte Carlo simulation is used to generate the experimental data based on a model that we assume to be true (either M_1 or M_2) with true parameters $\boldsymbol{\theta}_1$ or $\boldsymbol{\theta}_2$.

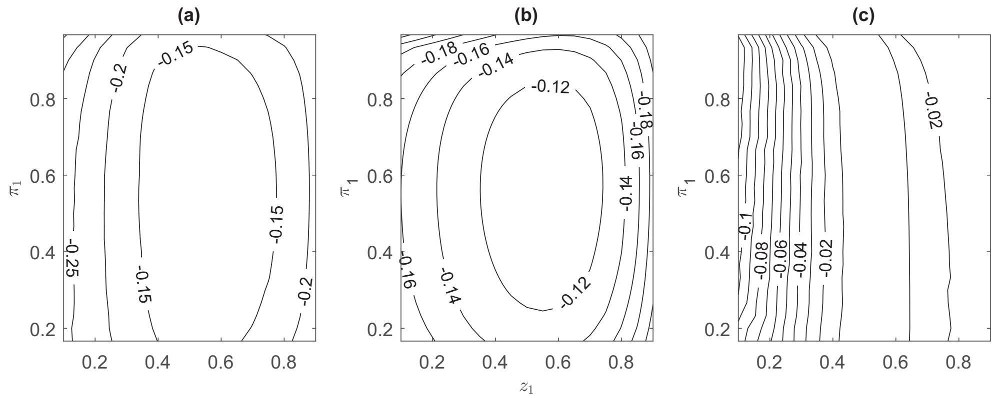

Figure 3. The contour plot for: (a). Integrated utility function $U(\mathcal{D})$ (b). Negative weighted *pre-posterior* variance $U_V(\mathcal{D})$. (c). Negative weighted squared mean bias $U_B(\mathcal{D})$.

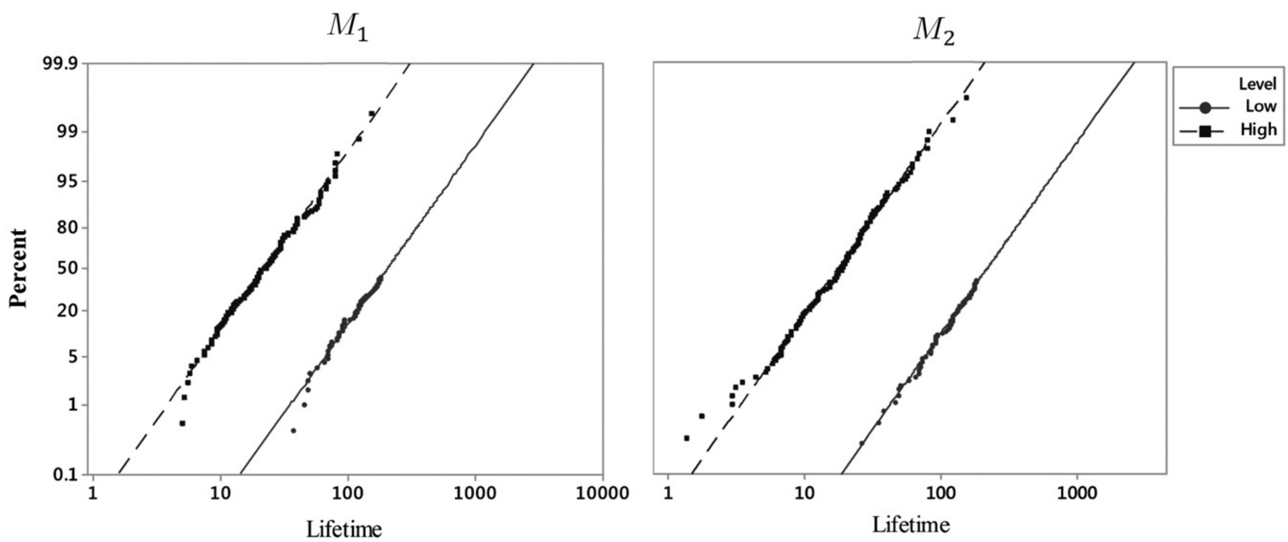

Figure 4. The probability plots for simulated data from $\mathcal{D}^*$ under M_1 and.M_2.

We assume that model M_2 with parameters $\boldsymbol{\theta}_2 = (7.58,\ 0.77,\ -4.65)'$ is the true model and simulate the ALT data from the model. Note that the true model should be unknown to the test planners, even after observing the experiment data $\mathbf{t}$. In other words, we are assuming the true model and parameters that are in a black box that test planners are trying to reveal statistically.

We simulate failure data $\mathbf{t}$ for one time, illustrated as the probability plots in Figure 4. The maximum likelihood estimators under M_1 and M_2 assumptions are given by $\hat{\boldsymbol{\theta}}_1 = (7.945,\ 0.719,\ -4.639)'$ and $\hat{\boldsymbol{\theta}}_2 = (7.821,\ 0.803,\ -4.961)'$.

Next, a BMA framework is used to predict the life quantile from the ALT data, i.e., we compute the BMA weights as well as the parameter posteriors of both models. Computing the posterior model probabilities by Eq. [8], we obtain

$$\Pr(M_1|\mathbf{t}) = \ 0.2693,\ \ \Pr(M_2|\mathbf{t}) = \ 0.7307$$

The posterior probability of 0.7307 on M_2 satisfies the test planner's requirement. Thus, the second-stage test is skipped. By MH sampling with a sample size of 20,000, we calculate the posterior means and variance covariance matrices of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ for each model as follows,

$$\hat{\mathrm{E}}(\boldsymbol{\theta}_1|\mathbf{t}) = (7.9614,\ 0.7158,\ \ -4.6526)$$
$$\hat{\mathrm{E}}(\boldsymbol{\theta}_2|\mathbf{t}) = (7.7895,\ 0.8017,\ \ -4.9163)$$
$$\hat{\mathrm{Cov}}(\boldsymbol{\theta}_1|\mathbf{t}) = \begin{bmatrix} 0.0262 & 0.0014 & -0.0296 \\ 0.0014 & 0.0013 & -0.0021 \\ -0.0296 & -0.0021 & 0.0365 \end{bmatrix}$$
$$\hat{\mathrm{Cov}}(\boldsymbol{\theta}_2|\mathbf{t}) = \begin{bmatrix} 0.0223 & 0.0017 & -0.0254 \\ 0.0017 & 0.0016 & -0.0018 \\ -0.0254 & -0.0018 & 0.0330 \end{bmatrix}$$

In Figure 5, the density-scaled histograms of the simulated posterior samples drawn by MCMC are

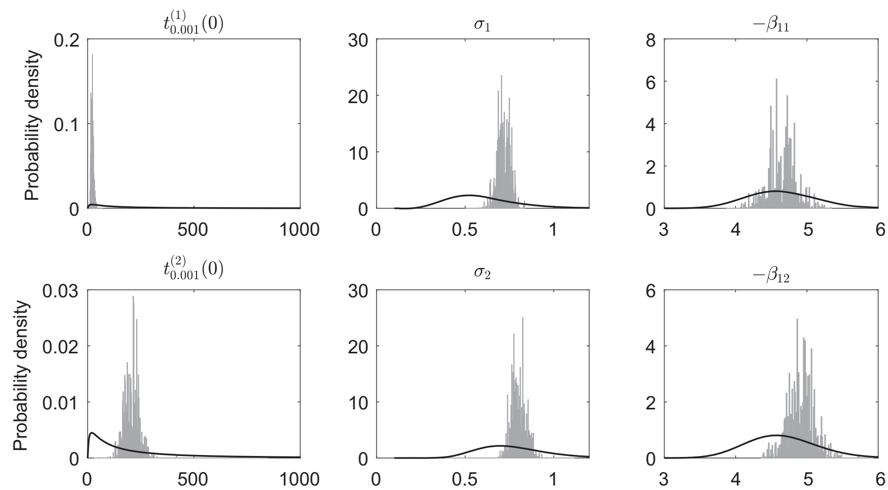

Figure 5. Comparison of posteriors (histograms) and priors (darker density lines) of the transformed parameters in models 1 and 2.

Table 4. Basic statistics of $\log T_{0.1}(0)|\mathbf{t}$ and $T_{0.1}(0)|\mathbf{t}$.

	$\log T_{0.1}(0)\|\mathbf{t}$	$T_{0.1}(0)\|\mathbf{t}$
Average	6.6489	776.7731
Median	6.6476	770.9213
Standard deviation	0.1122	87.4197
Coefficient of variation	0.0169	0.1125
95 percent confidence interval	(6.4166, 6.8581)	(612, 952)

compared with prior densities of each transformed parameters under M_1 and M_2. Note that we use the transformed parameters here for comparison because they are assumed to be independent in the prior information. Generally speaking, the posteriors have lower variability than the priors. For $t^{(i)}_{0.001}$, i.e., the 0.001 life quantile under use stress for each model, the prior is of very large variability, and by comparison the posterior is much more informative. We can notice a significant difference in the posterior $t^{(i)}_{0.001}$ under M_1 and M_2. This is due to the difference in the shape of left tails (lower quantiles) of Weibull and lognormal distribution. For both models, the prior information is the mean of 0.001 life quantile under use stress is about 0.5 years (183 days). However, under the same prior information and the same standard deviation of log life, the 0.1 life quantile $t_{0.1}(0)$ for each model differs significantly. The MLEs give $\hat{t}_{0.1}{}^{(1)}(0) = 559$ days and $\hat{t}_{0.1}{}^{(2)}(0) = 890$ days under M_1 and M_2 respectively, and the ML estimators show a significant difference up to 331 days, which is close to 1 year.

Afterward, we use Eq. [7] to obtain the posterior log 0.1 life quantile at use stress, i.e.,

$\log T_{0.1}(0)|\boldsymbol{t}$, by plugging in the posterior BMA weights and parameter samples. Under a sample size of 20,000, some basic statistics of $\log T_{0.1}(0)|\mathbf{t}$ and $T_{0.1}(0)|\mathbf{t}$ are listed in Table 4. The histograms are shown in Figure 6.

From Table 4 and Figure 6, under a relatively large sample, the posterior distributions of the $\log T_{0.1}(0)|\mathbf{t}$ and $T_{0.1}(0)|\mathbf{t}$ seem to be bell-shaped and unimodal. We can see that the mean and median are very close. The standard deviation is relatively small. For $T_{0.1}(0)|\mathbf{t}$, the value is 87.4, which is less than 3 months. To compare the predicted results of the weighted model with M_1 and M_2, we give the statistics for $\log T_P(0)|\mathbf{t}, M_i,\ i = 1,\ 2$ in Table 5.

Compared to the true log 0.1 life quantile under use stress, $\log t_{0.1}(0) = 6.5932$, i.e., $t_{0.1}(0) = 730.1174$, the weighted model gives the least biased mean and median under the particular dataset. The true value lies in the 95 percent confidence interval for $\log T_{0.1}(0)$ under all three cases. Note that the ALT data should be random for a given ALT plan $\mathcal{D}$, thus to evaluate a plan we need to carry out a simulation study with a large set of ALT data.

4.3. Comparing a Bayesian plan with a plan that does not consider model uncertainty

To evaluate the optimal plan in Table 3, especially the robustness to lifetime distribution and model parameters, the ALT optimal plans for M_1 and M_2 are obtained individually by the method introduced by Zhang and Meeker (2006), where the utility function for M_i is

$$U'_i(\mathcal{D}) = -\int C_i(\mathcal{D})\omega_i(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i$$

This method has considered the parameter uncertainty of the ALT model, but does not consider the

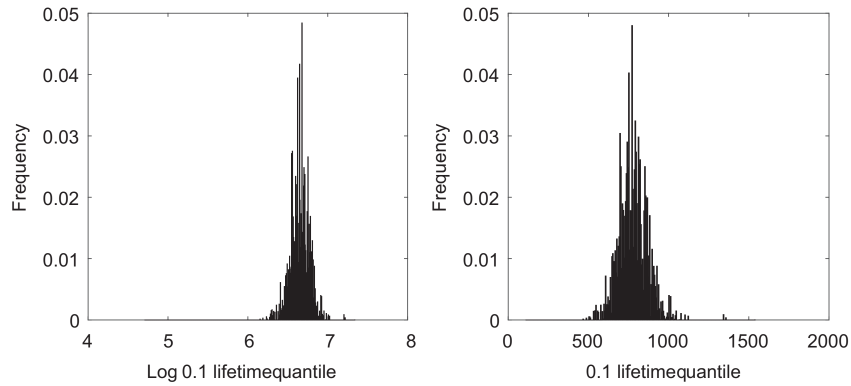

Figure 6. Histograms of $\log T_{0.1}(0)|\mathbf{t}$ and $T_{0.1}(0)|\mathbf{t}$.

Table 5. Basic statistics of $\log T_{0.1}(0)|\mathbf{t}$ under M_1 or M_2 assumption.

	$\log T_P(0)\|\mathbf{t}, M_1$	$\log T_P(0)\|\mathbf{t}, M_2$
Average	6.3507	6.7621
Median	6.3534	6.7705
Standard deviation	0.1627	0.1436
Coefficient of variation	0.0256	0.0212
95 percent confidence interval	(6.0468, 6.6491)	(6.4769, 7.0203)

model selection uncertainty. The same prior information is used for each optimal ALT plan. The optimal plans are denoted by $\mathcal{D}^*_{M_1}$ and $\mathcal{D}^*_{M_2}$ respectively, and are shown in Table 6.

It is assumed that the posterior prediction of $\log t_{0.1}(0)$ for plan $\mathcal{D}^*_{M_1}$ and $\mathcal{D}^*_{M_2}$ is based on M_1 and M_2, respectively. For each simulated ALT data, we use the posterior mode to represent the posterior $\log t_{0.1}(0)$ point estimator by assuming the posterior distribution of $\log T_{0.1}(0)$ is unimodal and symmetric, as in Figure 6. The posterior mode is obtained by maximizing $\mathcal{L}(\boldsymbol{\theta}_i|\mathbf{t}, M_i)\omega(\boldsymbol{\theta}_i)$ for each model. We simulate the ALT data for 2000 times for each optimal plan under the three scenarios. Figure 7 gives the resulting histograms of the predicted $\log T_{0.1}(0)$ under $\mathcal{D}^*$, $\mathcal{D}^*_{M_1}$ and $\mathcal{D}^*_{M_2}$ indicating their average and standard deviation.

In comparison to the true value $\log t_{0.1}(0) = 6.5932$, the optimal plan $\mathcal{D}^*_{M_2}$and the predictions based on M_2 provide the most accurate mode average (6.60). This is as expected since we selected the true model by using $\mathcal{D}^*_{M_2}$. By comparison, the optimal plan $\mathcal{D}^*$ and prediction by our method gives a mode average of 6.52 and standard deviation of 0.144. The mode average has an error rate of 1.06 percent. If the model is mis-specified, i.e., M_1 is used to plan and predict the lifetime quantile, the mode average results in 6.16 and the error rate is 6.58 percent. For the proposed method, the histogram posterior probability on the pre-assumed true model M_2, i.e., $\Pr(M_2|\mathbf{t})$, is given in Figure 8. The figure shows that the model averaging framework introduced in Section 3.3 for the simulated data has shown significant preference for the true model M_2, and the proportion that $\Pr(M_2|\mathbf{t}) \leq 0.5$ is very small.

For a clearer illustration, we calculate the predicted posteriors of $T_{0.1}(0)$ from the three scenarios. The error rates with respect to the true value are 6.75 percent, 35.2 percent and 0.5 percent. This indicates that model mis-specification with a single model is very risky, while the proposed robust planning and prediction framework has a reasonable error rate that is only about 6 percent. It is emphasized that these results are based on diffuse prior information of model selection, i.e., $\Pr(M_1) = \Pr(M_2)$. Furthermore, the standard deviation (SD) of the posterior mode is 0.144 under $\mathcal{D}^*$. This is very close to $\text{SD} = 0.133$ in the true case. $\mathcal{D}^*_{M_1}$ gives a value of 0.211, which is considerably larger than those from $\mathcal{D}^*$ and $\mathcal{D}^*_{M_2}$.

The results show that the incorrect selection of lifetime distribution may lead to a drastic prediction error. The proposed model yields the posterior life quantile weighed by the posterior model probabilities, providing the prediction with reasonable robustness when the lifetime distribution information is diffuse. If more reliable prior information on model selection is given, the weighted model is expected to provide better results. Related discussions are given in Section 5.

4.4. An example of adaptive second-stage planning

In the example described previously, the adaptive second-stage test is skipped. To illustrate the planning of adaptive test, we modify the user-selected desired posterior probability of the preferred model to 0.8. In this situation, the adaptive test is now necessary based on the test data from Section 4.2. It is assumed that a total budget of \$2,000 is given for the adaptive

Table 6. Optimal ALTs under M_1 or M_2 regardless of weights and bias.

	$\mathcal{D}^*_{M_1}$				$\mathcal{D}^*_{M_2}$			
Condition i	Level in [0,1] z_i	Level °C TEMPC	Proportion π_i	Number n_i	Level in [0,1] z_i	Level °C TEMPC	Proportion π_i	Number n_i
Use	0	50	0	0	0	50	0	0
Low	0.67	94	0.50	150	0.51	82.27	0.70	210
High	1	120	0.50	150	1	120	0.30	90
$U'(\mathcal{D}^*)$		−0.1501				−0.0550		

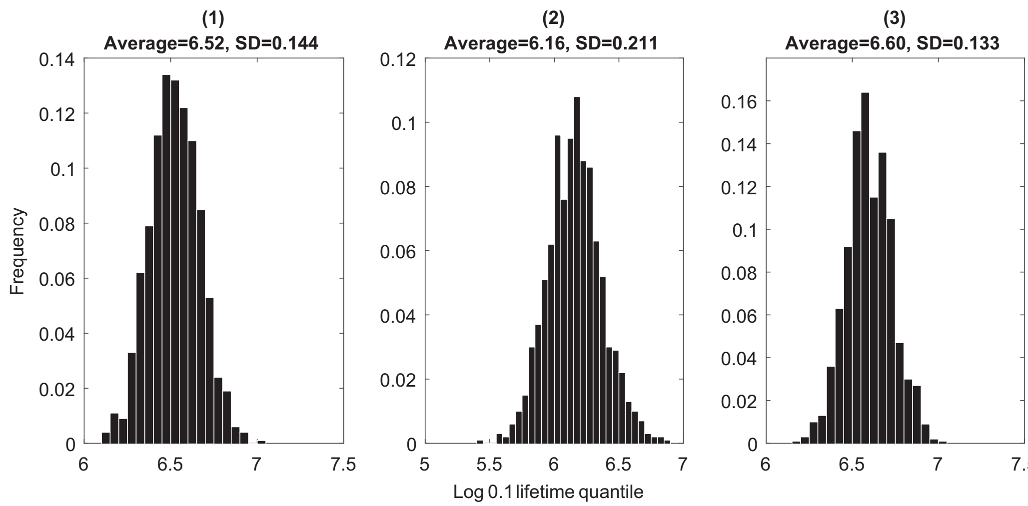

Figure 7. Histograms for posterior mode under plans (1) $\mathcal{D}^*$, (2) $\mathcal{D}^*_{M_1}$, and (3).$\mathcal{D}^*_{M_2}$.

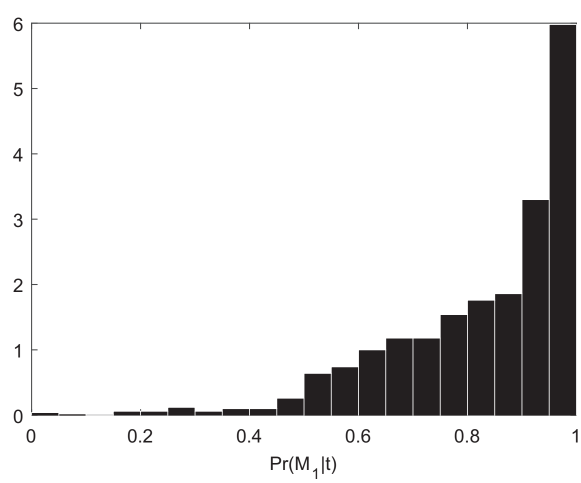

Figure 8. Histogram of.$\Pr(M_2|\mathbf{t})$.

Table 7. Optimal plan $\mathcal{D}'*$ to maximize $U'(\mathcal{D}')$.

Condition i	Level in [0,1] z_i	Level °C TEMPC	Proportion π_i	Number n_i	Expected Failures
Use	0	50	0	0	0
Low	0.54	84.37	0.489	128	16
High	1	120	0.511	134	127
$t'_c = 69$ days				$U'(\mathcal{D}'^*) = -0.0187$	

second-stage planning. Each test unit costs \$5 and test-running cost per day is \$10, i.e., $C_I = 5$, $C_T = 10$.

Under the budgetary constraint, the optimal adaptive plan indicates using 262 test units in total (Table 7) and suggests a censoring time of 69 days, which is considerably shorter than 183 days for the first stage. In contrast with the utility of -0.1095 for the previous plan, the adaptive plan gives a much greater utility of -0.0187.

Through a single-time simulation of test data, we obtain the result $\Pr(M_1|\mathbf{t},\mathbf{t}') = 0.053$, which makes $\Pr(M_2|\mathbf{t},\mathbf{t}') = 0.947$. By the model in Eq. [16], the predicted log 0.1 life quantile is 6.617 with standard deviation 0.1015. The predicted life quantile is closer to the true value with a smaller standard deviation. We can see that the results from both stages of test has shown much stronger preference to M_2. The test planner may wish to eliminate M_1 for the following the analysis.

5. Sensitivity analysis

This section discusses the sensitivity of the test plans and life quantile predictions given by our proposed framework with respect to the prior probability of candidate models, the sample size, and the available budget for an adaptive second-stage test.

Table 8. Summary of results under different prior probabilities of candidate models.

$\Pr(M_1)$	Optimal ALT plan $\mathcal{D}^*$ π_1 (n_1)	z_1 (TEMP$_1$°C)	$U(\mathcal{D}^*)$	Mean $(\Pr(M_1\|\mathbf{t}))$
0	0.71 (213)	0.51 (82.27)	−0.0550	0.000
0.1	0.68 (204)	0.54 (84.37)	−0.0669	0.034
0.2	0.65 (196)	0.54 (84.37)	−0.0781	0.054
0.3	0.63 (189)	0.54 (84.37)	−0.0890	0.080
0.4	0.61 (184)	0.54 (84.37)	−0.0995	0.101
0.5	0.55 (166)	0.54 (84.37)	−0.1095	0.108
0.6	0.54 (162)	0.54 (84.37)	−0.1190	0.126
0.7	0.53 (159)	0.53 (84.37)	−0.1283	0.175
0.8	0.51 (154)	0.54 (84.37)	−0.1372	0.185
0.9	0.49 (148)	0.55 (85.08)	−0.1456	0.189
1.0	0.48 (143)	0.55 (85.08)	−0.1535	1.000

5.1. Influence of prior probability of candidate models on the prior-based ALT planning

The prior probability of M_i, i.e., $\Pr(M_i)$, describes how the test planners prefer each model prior to conducting an ALT. Based on the utility function, the change in value of $\Pr(M_i)$ has an influence on the optimal plan $\mathcal{D}^*$ and the following prediction of lifetime quantiles. To investigate the influence in this example, we set $\Pr(M_1)$ from 0 to 1 with increase step 0.1, making $\Pr(M_2)$ vary from 1 to 0, and obtain the optimal plan, the corresponding optimal utility value, and the average value of $\Pr(M_1|\mathbf{t})$ by simulating the ALT data 2000 times for each case. The results are summarized in Table 8.

As $\Pr(M_1|\mathbf{t})$ increases, the proportion of units allocated to lower stress decreases, while the level of lower stress does not vary much with only less than 4 °C increase when $\Pr(M_1|\mathbf{t})$ varies from 0 to 1. The reason for this behavior is that if the test planner believes that M_2 is much more plausible than M_1 by setting a small prior probability to $\Pr(M_1)$, the optimal plan would be nearer to plan $\mathcal{D}^*_{M_2}$, which calls for a higher π_1 and a lower z_1. As $\Pr(M_1)$ increases, the plan becomes closer to $\mathcal{D}^*_{M_1}$. Note that the utility function is not a linear combination of asymptotic posterior variance and squared bias weighted by $\Pr(M_i)$ because the bias function is dependent on $\Pr(M_i)$. However, in this example, under the sample size of 300, the utility value seems to be dominated by the $U_V(\mathcal{D})$ part, thus the optimal plans under different $\Pr(M_i)$ are close to the linear combination of $\mathcal{D}^*_{M_1}$ and $\mathcal{D}^*_{M_2}$ weighted by $\Pr(M_i)$.

We are also interested in the posterior model probability $\Pr(M_i|\mathbf{t})$ under different $\Pr(M_i)$. If we set $\Pr(M_1)$ as 0.5 or smaller, the average $\Pr(M_1|\mathbf{t})$ is smaller than 0.108, making $\Pr(M_2|\mathbf{t})$ greater than 0.8. The data simulated in Section 4.2 yields $\Pr(M_1|\mathbf{t}) \geq 0.26$ if $\Pr(M_1) = 0.5$, which is much larger than the average value 0.108 shown in Table 8, but the results are still satisfactory. Therefore, the Bayesian model averaging procedure works well to give reasonable prediction by providing posterior model probabilities that combing the prior information with data.

Under extreme cases where $\Pr(M_1) = 0$ or 1, the BMA framework will result in the same model as in the prior. This will only occur, however, in the unlikely case where the test planner has extremely strong confidence that one of the models is true, which is not common in practice because tested products are usually new and the prior probability for a certain $\Pr(M_i)$ should not be set to 1. To indicate high preference, one can make $\Pr(M_i)$ close to 1, for instance, as in Table 8 we can set $\Pr(M_1) = 0.9$. The model selection is still of high efficiency when seeking for a better model that is close to the true one by taking advantage of the BMA framework.

Table 9. Summary of results under different sample size N.

N	Optimal ALT plan $\mathcal{D}^*$ π_1 (n_1)	z_1 (TEMP$_1$°C)	$U(\mathcal{D}^*)$	Average $(\Pr(M_1\|\mathbf{t}))$
30	0.44 (13)	0.52 (80.2)	−0.2612	0.313
100	0.52 (52)	0.52 (83.0)	−0.1659	0.228
200	0.54 (108)	0.53 (83.7)	−0.1275	0.157
300	0.55 (166)	0.54 (84.4)	−0.1095	0.108
400	0.56 (222)	0.54 (84.4)	−0.0985	0.062
500	0.56 (282)	0.54 (84.4)	−0.0908	0.048
1,000	0.62 (580)	0.54 (85.1)	−0.0707	0.000

5.2. Influence of sample size on the prior-based ALT test

In non-Bayesian ALT planning problems, the sample size does not affect the planning of ALT, but the sample size should be relatively large to assume the estimated parameters are asymptotically normally distributed. However, the total number of test units N, relative to the amount of prior information, has an influence on both the optimization of Bayesian ALT planning and on inference. A larger sample size results in a decrease of influence of prior information on the prediction. Another consideration is that the expected squared bias part $U_B(\mathcal{D})$ term is not affected by sample size, while the variance part $U_V(\mathcal{D})$ decreases as N increases under the same $\mathcal{D}$. By fixing $\Pr(M_1) = \Pr(M_2) = 0.5$, the optimal ALT plans with corresponding utility values and the model selection results under varied N are given in Table 9. The results are based on 2,000 times simulation of ALT data.

The optimal π_1^* increases as the sample size becomes larger, while z_1^* does not change much. A larger sample size decreases the relative information of priors; thus, the optimal plan gives more allocation at lower stress to make the data less extrapolated. We notice that the utility value increases fast when N changes from very small (30) to relatively large (300), and the increasing rate becomes very low when N is larger than 300. The reason is that, with a small sample size, the variance term dominates the utility function, in which the increase of sample size will reduce the *pre-posterior* variance fast. However, as N becomes even larger, the squared bias term in the utility does not change and the variance term is already very low, thus the improvement of utility will diminish. In addition, the squared bias $U_B(\mathcal{D})$ will reach its highest value when all test units are allocated to the low stress level. This is another reason that π_1^* is higher with large sample size.

For the model averaging, a larger N is expected to provide a better result. When the sample size is extremely large, e.g., $N = 1,000$, the model selection procedure almost eliminates the wrong model every time in 2000 simulation runs. In real engineering cases, the sample size is usually limited by the available number of products or cost. Under the proposed analysis for this example, the sample size 300 seems reasonable in the tradeoff as the utility has improved substantially from $N = 200$ to $N = 300$. However, the test is planned before it is conducted, thus the only criterion is the utility value for each optimal plan, which is influenced by prior information. Nevertheless, test planners can resort to simulation techniques to analyze the influence of changing sample size and prior information under several anticipated scenarios.

5.3. Influence of test budget on the adaptive test

To explore the planning of the adaptive second-stage test, we modify the total budget for the test from \$500 to \$10,000 and obtain the optimal plans by the algorithm described in Section 3.5, and the results are summarized in Table 10. Under very limited budgets, e.g., \$500, the optimal sample size is very small to ensure the censoring time to be reasonably adequate. Meanwhile, most test units are allocated to higher stress to decrease the chance of censoring. If the budget is moderate, e.g., \$1,000–\$5,000, the optimal censoring time does not vary much, and the improvement in utility mainly benefits from the increase of sample size. On the contrary, when the budget is even larger, e.g., \$10,000, the optimal censoring time becomes much longer and, therefore, the proportion of test units allocated to the lower stress is high because the test duration is long enough to produce enough failures. Generally, the planning of the adaptive test with a given budget faces the tradeoff between sample size and test duration. As the budget increases, the emphasis of planning activity shifts from censoring time to sample size determination.

Table 10. Optimal adaptive ALT plan $\mathcal{D}'*$ under different total budgets.

	Optimal adaptive ALT plan $\mathcal{D}'*$			
TB (in \$)	N'	t_c'	π_1' (n_1')	$U'(\mathcal{D}'^*)$
500	4	48	0.25 (1)	−0.0331
1,000	58	71	0.24 (17)	−0.0221
2,000	262	69	0.49 (128)	−0.0187
5,000	858	71	0.51 (434)	−0.0157
10,000	1200	400	0.93 (1120)	−0.0124

6. Conclusions and areas for future research

This article presents a systematic two-stage planning and prediction approach to accelerated life tests under model uncertainty from a Bayesian perspective. A novel modified V-optimality criterion that simultaneously considers *pre-posterior* variance and squared bias is used to optimize the first-stage ALT plan. Afterward, we use the Bayesian model averaging technique to obtain the posterior prediction of life quantile of interest under use stress. An adaptive second-stage test can be conducted if the first-stage ALT data are not sufficient for model differentiation. The adaptive second-stage ALT is planned based on the results obtained from the first-stage test and with a budget constraint. In most reliability testing applications, the true lifetime distribution and acceleration model are unknown, no matter before or after ALT experiments. Our proposed approach instills robustness to both ALT planning and reliability prediction to counter model uncertainties. This holistic approach has not been explored before.

The adhesive bond test example is revisited to illustrate the proposed approach and make comparisons by simulation with plans from previous Bayesian ALT studies. The comparison shows that our approach yields a closer result to the true model. Meanwhile, it compromises very little statistical efficiency in comparison to the model misspecification error even when the prior information on lifetime distribution is relatively diffuse. The sensitivity analysis with respect to prior probabilities of candidate models and sample size is addressed for the first-stage planning, and the sensitivity analysis with respect to test budget is addressed for the second-stage planning.

The proposed framework can be extended to ALT planning and reliability prediction problems with multiple stress factors with more stress levels (Huang and Wu 2017; Wu and Huang 2017), and it can also be extended for more general acceleration regression models and different lifetime distributions (Fan and Yu 2013; Abdel Ghaly et al. 2016). In these future studies, the efficiency of numerical optimization is challenging due to the increasing number of decision variables. In addition, it is noted that the planning methods in the article is based on large-sample approximation, for ALT problems with small samples of test units, the similar idea is of interest for exploration and validation. As mentioned in Zhao, Xu, and Liu (2017) and Hong and Ye (2017), for more reliable products, degradation-based reliability test plans that are robust to model uncertainty may also be addressed in future research.

About the authors

Mr. Zhao received his PhD degree from the Department of Systems Engineering and Engineering Management. He is the corresponding author. His email address is xiujizhao2-c@my.cityu.edu.hk

Dr. Pan is associate professor in the School of Computing, Informatics & Decision Systems Engineering. His email address is rong.pan@asu.edu

Dr. del Castillo is distinguished professor of Industrial and Manufacturing Engineering and professor of Statistics. His email address is exd13@psu.edu

Dr. Xie is chair professor in the Department of Systems Engineering and Engineering Management. His email address is minxie@cityu.edu.hk

Acknowledgment

We are grateful to the editor and two referees for their insightful comments to earlier versions of the article.

Funding

Xie and Zhao were supported in part by the Research Grants Council of Hong Kong under Grant T32-101/15-R and CityU 11203815, and in part by the National Natural Science Foundation of China (71532008). Pan was partially supported by National Science Foundation CMMI 1726445.

ORCID

Xiujie Zhao http://orcid.org/0000-0003-3450-5480

References

Abdel Ghaly, A. A., H. M. Aly, and R. N. Salah. 2016. Different estimation methods for constant stress accelerated life test under the family of the exponentiated distributions. *Quality and Reliability Engineering International* 32 (3):1095–108. doi: 10.1002/qre.1818.

Agboto, V., W. Li, and C. Nachtsheim. 2010. Screening designs for model discrimination. *Journal of Statistical Planning and Inference* 140 (3):766–780. doi: 10.1016/j.jspi.2009.09.005.

Atkinson, A. C., and V. V. Fedorov. 1975. Optimal design: Experiments for discriminating between several models. *Biometrika* 62 (2):289–303. doi: 10.2307/2335364.

Carlin, B. P., and T. A. Louis. 2000. *Bayes and empirical Bayes methods for data analysis*. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC.

Chen, N., Y. Tang, and Z. Ye. 2016. Robust quantile analysis for accelerated life test data. *IEEE Transactions on Reliability* 65 (2):901–13. doi: 10.1109/TR.2015.2500368.

Chen, P., A. Xu, and Z. S. Ye. 2016. Generalized fiducial inference for accelerated life tests with Weibull distribution and progressively Type-II censoring. *IEEE Transactions on Reliability* 65 (4):1737–44. doi: 10.1109/TR.2016.2604298.

Clyde, M., P. Müller, and G. Parmigiani. 1995. Optimal design for heart defibrillators. In *Case studies in Bayesian statistics*. Volume II, 278–92. New York, NY: Springer New York.

Dette, H., and S. Titoff. 2009. Optimal discrimination designs. *The Annals of Statistics* 37 (4):2056–82. doi: 10.1214/08-AOS635.

Elsayed, E. A. 2012. Overview of reliability testing. *IEEE Transactions on Reliability* 61 (2):282–91. doi: 10.1109/TR.2012.2194190.

Erkanli, A., and R. Soyer. 2000. Simulation-based designs for accelerated life tests. *Journal of Statistical Planning and Inference* 90 (2):335–48. doi: 10.1016/S0378-3758(00)00126-9.

Escobar, L. A., and W. Q. Meeker. 1986. Algorithm as 218: Elements of the Fisher information matrix for the smallest extreme value distribution and censored data. *Applied Statistics* 35 (1):80–6. doi: 10.2307/2347873.

Escobar, L. A., and W. Q. Meeker. 1994. Algorithm as 292: Fisher information matrix for the extreme value, normal and logistic distributions and censored data. *Applied Statistics* 43 (3):533–40. doi: 10.2307/2986276.

Fan, T. H., and C. H. Yu. 2013. Statistical inference on constant stress accelerated life tests under generalized gamma lifetime distributions. *Quality and Reliability Engineering International* 29 (5):631–8. doi: 10.1002/qre.1412.

Freels, J. K., J. J. Pignatiello, R. L. Warr, and R. R. Hill. 2015. Bridging the gap between quantitative and qualitative accelerated life tests. *Quality and Reliability Engineering International* 31 (5):789–800. doi: 10.1002/qre.1636.

Hill, P. D. H. 1978. A review of experimental design procedures for regression model discrimination. *Technometrics* 20(1):15–21. doi: 10.1080/00401706.1978.10489611.

Hong, L. Q., and Z. S. Ye. 2017. When is acceleration unnecessary in a degradation test? *Statistica Sinica* 27 (3): 1461–83.

Huang, S.-R., and S.-J. Wu. 2017. Optimal sample size allocation for accelerated life test under progressive Type-II censoring with competing risks. *Journal of Statistical Computation and Simulation* 87 (1):1–16. doi: 10.1080/00949655.2016.1187608.

Kielpinski, T. J., and W. Nelson. 1975. Optimum censored accelerated life tests for normal and lognormal life

distributions. *IEEE Transactions on Reliability* R-24 (5): 310–20. doi: 10.1109/TR.1975.5214920.
Lee, J., and R. Pan. 2012. Bayesian analysis of step-stress accelerated life test with exponential distribution. *Quality and Reliability Engineering International* 28 (3):353–61. doi: 10.1002/qre.1251.
Meeker, W. Q. 1984. A comparison of accelerated life test plans for Weibull and lognormal distributions and type I censoring. *Technometrics* 26 (2):157–71. doi: 10.1080/00401706.1984.10487941.
Meeker, W. Q., and L. A. Escobar. 1993. A review of recent research and current issues in accelerated testing. *International Statistical Review* 61 (1):147–68. doi: 10.2307/1403600.
Meeker, W. Q., and L. A. Escobar. 1998. *Statistical methods for reliability data*. New York, NY: John Wiley & Sons.
Meeker, W. Q., and G. J. Hahn. 1985. *How to plan an accelerated life Test - Some practical guidelines (Vol. 10 of the ASQC basic references in quality control: Statistical techniques)*. Milwaukee, WI: American Society for Quality Control.
Monroe, E. M., R. Pan, C. M. Anderson-Cook, D. C. Montgomery, and C. M. Borror. 2011. A generalized linear model approach to designing accelerated life test experiments. *Quality and Reliability Engineering International* 27 (4):595–607. doi: 10.1002/qre.1143.
Monroe, E. M., R. Pan, C. M. Anderson-Cook, D. C. Montgomery, and C. M. Borror. 2010. Sensitivity analysis of optimal designs for accelerated life testing. *Journal of Quality Technology* 42 (2):121–35. doi: 10.1080/00224065.2010.11917811.
Müller, P., and G. Parmigiani. 1995. Optimal design via curve fitting of Monte Carlo experiments. *Journal of the American Statistical Association* 90 (432):1322–30. doi: 10.1080/01621459.1995.10476636.
Nasir, E. A., and R. Pan. 2015. Simulation-Based Bayesian optimal ALT designs for model discrimination. *Reliability Engineering & System Safety* 134:1–9. doi: 10.1016/j.ress.2014.10.002.
Nelson, W. B. 2005. A bibliography of accelerated test plans. *IEEE Transactions on Reliability* 54 (2):194–7. doi: 10.1109/TR.2005.847247.
Nelson, W., and T. J. Kielpinski. 1976. Theory for optimum censored accelerated life tests for normal and lognormal life distributions. *Technometrics* 18 (1):105–14. doi: 10.2307/1267923.
Nelson, W., and W. Q. Meeker. 1978. Theory for optimum accelerated censored life tests for Weibull and extreme value distributions. *Technometrics* 20 (2):171–7. doi: 10.1080/00401706.1978.10489643.
Pan, R., and T. Yang. 2014. Design and evaluation of accelerated life testing plans with dual objectives. *Journal of Quality Technology* 46 (2):114–26. doi: 10.1080/00224065.2014.11917957.
Pan, R., T. Yang, and K. Seo. 2015. Planning constant-stress accelerated life tests for acceleration model selection. *IEEE Transactions on Reliability* 64 (4):1356–66. doi: 10.1109/TR.2015.2421514.
Pascual, F. G., and G. Montepiedra. 2003. Model-robust test plans with applications in accelerated life testing. *Technometrics* 45 (1):47–57. doi: 10.1198/004017002188618699.
Pascual, F. G., and G. Montepiedra. 2005. Lognormal and Weibull accelerated life test plans under distribution misspecification. *IEEE Transactions on Reliability* 54 (1): 43–52. doi: 10.1109/TR.2004.837316.
Polson, N. G. 1993. A Bayesian perspective on the design of accelerated life tests. *Advances in Reliability*:321–30.
Sha, N., and R. Pan. 2014. Bayesian analysis for step-stress accelerated life testing using Weibull proportional hazard model. *Statistical Papers* 55 (3):715–26. doi: 10.1007/s00362-013-0521-2.
Shi, Y., and W. Q. Meeker. 2012. Bayesian methods for accelerated destructive degradation test planning. *IEEE Transactions on Reliability* 61 (1):245–53. doi: 10.1109/TR.2011.2170115.
Wu, S.-J., and S.-R. Huang. 2017. Planning two or more level constant-stress accelerated life tests with competing risks. *Reliability Engineering & System Safety* 158:1–8. doi: 10.1016/j.ress.2016.09.007.
Yu, I.-T., and C.-L. Chang. 2012. Applying Bayesian model averaging for quantile estimation in accelerated life tests. *IEEE Transactions on Reliability* 61 (1):74–83. doi: 10.1109/TR.2012.2182814.
Zhang, D., and H. Liao. 2016. A fully integrated double-loop approach to the design of statistically and energy efficient accelerated life tests. *IIE Transactions* 48 (4): 371–88. doi: 10.1080/0740817X.2015.1109738.
Zhang, Y., and W. Q. Meeker. 2006. Bayesian methods for planning accelerated life tests. *Technometrics* 48 (1): 49–60. doi: 10.1198/004017005000000373.
Zhao, X., J. Xu, and B. Liu. 2017. Accelerated degradation tests planning with competing failure modes. *IEEE Transactions on Reliability* 67 (1):142–55. doi: 10.1109/TR.2017.2761025.

Appendix A

Approximation of $U_1(\mathcal{D})$

Recall Eq. [2], the first part representing the *pre-posterior* variance,

$$\hat{U}_V(D) = -\sum_i \Pr(M_i)\int C_i(D)\omega_i(\hat{\boldsymbol{\theta}}_i)d\hat{\boldsymbol{\theta}}_i,$$

where for $i = 1$ or 2

$$C_i(D) = \mathrm{AVar}_{\boldsymbol{\theta}_i|\mathbf{t}}(\log t_p(0)|M_i) = \mathbf{c}_i'\mathrm{AVar}_{\boldsymbol{\theta}_i|\mathbf{t}}(\hat{\boldsymbol{\theta}}_1)\mathbf{c}_i,$$

By Clyde *et. al* (1995) 's large sample approximation,

$$\mathrm{AVar}_{\boldsymbol{\theta}_i|\mathbf{t}}(\hat{\boldsymbol{\theta}}_1) = \left[\mathbf{S}_i^{-1} + \hat{\mathbf{I}}_{\boldsymbol{\theta}_i}(D)\right]^{-1} \approx \left[\mathbf{S}_i^{-1} + \mathbf{I}_{\boldsymbol{\theta}_i}(\mathcal{D})\right]^{-1}. \quad \text{[A1]}$$

For $\omega_i(\hat{\boldsymbol{\theta}}_i)$, the density could be obtained by integrating the sampling distribution of the conditional variables as

$$\omega_i(\hat{\boldsymbol{\theta}}_i) = \int \omega_i\left(\hat{\boldsymbol{\theta}}_i|\boldsymbol{\theta}_i\right)\omega_i(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i.$$

When the sample size becomes larger the conditional variables converge to the distribution of $\boldsymbol{\theta}_i$, most variability originates from the prior, therefore, $\omega_i(\hat{\boldsymbol{\theta}}_i) \approx \omega_i(\boldsymbol{\theta}_i)$, and by combining the approximations with the those in Eq. [A1], the $U_V(\mathcal{D})$ is approximated as in Eq. [8].

Appendix B

Fisher information of unknown parameters under two models

Taking $\mathbf{I}_{\boldsymbol{\theta}_1}(\mathcal{D})$ as an example, for $\mathcal{D} = \{(\pi_i, z_i), i = 1, ..., l\}$, then

$$\mathbf{I}_{\boldsymbol{\theta}_1}(\mathcal{D}) = \frac{N}{\sigma^2}\sum_{i=1}^{l}\pi_i F_i,$$

where F_i is the scaled Fisher information matrix (Escobar and Meeker 1994):

$$F_i = \begin{bmatrix} f_{11}(\zeta_i) & f_{12}(\zeta_i) & f_{11}(\zeta_i)z_i \\ f_{12}(\zeta_i) & f_{22}(\zeta_i) & f_{12}(\zeta_i)z_i \\ f_{11}(\zeta_i)z_i & f_{12}(\zeta_i)z_i & f_{11}(\zeta_i)z_i^2 \end{bmatrix},$$

where

$$\zeta_i = \frac{\log(t_C) - \beta_{01} - \beta_{11}}{\sigma}.$$

Then the elements $f_{11}(\zeta_i)$, $f_{12}(\zeta_i)$, $f_{22}(\zeta_i)$ can be calculated as

$$\begin{aligned} f_{11}(\zeta_i) &= \Psi_0(\zeta_i) + \eta(\zeta_0), \\ f_{12}(\zeta_i) &= \Psi_1(\zeta_i) + \zeta_0\eta(\zeta_0), \\ f_{21}(\zeta_i) &= \Psi_2(\zeta_i) + \zeta_0^2\eta(\zeta_0), \end{aligned}$$

where $\zeta_0 = (-\beta_{01} - \beta_{11}z_i)/\sigma$, and

$$\Psi(\zeta_i) = \int_{\zeta_0}^{\zeta_i}[1 + xH(x)]^i H(x)^{2-i}g(x)dx, \quad i = 0, \ 1, \ 2,$$

$$H(x) = \frac{g'(x)}{g(x)} + \frac{g(x)}{1 - G(x)},$$

$$\eta(x) = \frac{g(x)}{[1 - G(x)]G(x)}.$$

Here, $g(x)$ and $G(x)$ are the PDF and CDF of SEV distribution, $g(x) = \exp(x - \exp(x))$, $G(x) = 1 - \exp(-\exp(x))$. To derive $\mathbf{I}_{\boldsymbol{\theta}_2}(\mathcal{D})$, simply replace $\boldsymbol{\theta}_1$ with $\boldsymbol{\theta}_2$ and set that $g(x)$ and $G(x)$ is the PDF and CDF of standard normal distribution, respectively.

Appendix C

Approach to obtaining θ_i^*

Based on the result in Pascual and Montepiedra (2005), $\boldsymbol{\theta}_i^*$ can be obtained by minimizing the expected value with respect to M_i of the negative M_{3-i} loglikelihood under test plan $\mathcal{D} = \{(\pi_i, z_i), i = 1, ..., l\}$. If M_1 with $\boldsymbol{\theta}_1$ is assumed to be true, then $\boldsymbol{\theta}_2^*$ can be obtained by minimizing the following equation:

$$\begin{aligned} &\frac{1}{N}\mathbb{E}_{M_1}[-\log\mathcal{L}_2(\boldsymbol{\theta}_2, \mathcal{D})] \\ &= \log(\sqrt{2\pi}\sigma_2) - \sum_{n=1}^{l}\pi_n\{\log\{1 - \Phi_2[\zeta_2(z_n)]\} \\ &\quad + \log(\sqrt{2\pi}\sigma_2)\}\exp\{-\exp[\zeta_1(z_n)]\} \\ &+ \frac{1}{2}\sum_{n=1}^{l}\pi_n\int_{-\infty}^{\zeta_1(z_n)}[\frac{\sigma_1}{\sigma_2}y - \Delta_{21}(z_n)]^2\exp[y - \exp(y)]dy, \end{aligned} \quad [C1]$$

where $\Delta_{21}(z_n) = [\beta_{02} + \beta_{12}z_n - \beta_{01} - \beta_{11}z_n]/\sigma_2$ and $\zeta_i(z_n) = [t_C - \beta_{0i} - \beta_{1i}z_n]/\sigma_i$ for $i = 1$ and 2. If instead M_2 with $\boldsymbol{\theta}_2$ is assumed to be true, $\boldsymbol{\theta}_1^*$ is obtained by minimizing

$$\begin{aligned} &\frac{1}{N}\mathbb{E}_{M_2}[-\log\mathcal{L}_1(\boldsymbol{\theta}_1, \mathcal{D})] \\ &= \sum_{n=1}^{l}\pi_n\{\log\sigma_1 + \Delta_{12}(z_n) - \exp[\zeta_1(z_n)]\} \\ &\quad \Phi_2[\zeta_2(z_n)] + \frac{\sigma_2}{\sigma_1}\sum_{n=1}^{l}\pi_n\phi_2[\zeta_2(z_n)] \\ &\quad + \sum_{n=1}^{l}\pi_n\left\{\exp\left[\frac{\sigma_2^2}{2\sigma_1^2} - \Delta_{12}(z_n)\right]\right. \\ &\quad \left.\Phi_2\left[\zeta_2(z_n) - \frac{\sigma_2}{\sigma_1}\right] + \exp[\zeta_1(z_n)]\right\}, \end{aligned} \quad [C2]$$

where $\Delta_{12}(z_n) = [\beta_{01} + \beta_{11}z_n - \beta_{02} - \beta_{12}z_n]/\sigma_1$.

Appendix D

MCMC methods to get a posterior sample of θ_1 and θ_2

Because the joint posterior in Eq. [10] is not a regular known distribution, the Metropolis-Hastings algorithm is employed to generate the posterior samples of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. Taking $\boldsymbol{\theta}_1$ as the example, the sampling approach is as follows.

1. First, an initial sample of $\boldsymbol{\theta}_1$, denoted by $\boldsymbol{\theta}_1^{(0)}$ is sampled from the joint proposal distribution $q(\boldsymbol{\theta}_1)$, which is different from the posterior distribution of $\boldsymbol{\theta}_1$ but is believed to be relatively close to $p(\boldsymbol{\theta}_1|\boldsymbol{t})$ and easy to sample from.
2. For iteration $i = 1, 2, ...$, do the following steps,
   a. Propose a candidate sample randomly from distribution $q(\boldsymbol{\theta}_1)$, denoted by $\boldsymbol{\theta}_1^{\text{cand}}$.
   b. Compute the acceptance probability:

   $$\alpha\left(\boldsymbol{\theta}_1^{\text{cand}}\boldsymbol{\theta}_1^{(i-1)}\right) = \min\left\{1, \frac{q\left(\boldsymbol{\theta}_1^{(i-1)}|\boldsymbol{\theta}_1^{\text{cand}}\right)p\left(\boldsymbol{\theta}_1^{\text{cand}}|\boldsymbol{t}\right)}{q\left(\boldsymbol{\theta}_1^{\text{cand}}|\boldsymbol{\theta}_1^{(i-1)}\right)p\left(\boldsymbol{\theta}_1^{(i-1)}|\boldsymbol{t}\right)}\right\}. \quad [D1]$$

   c. Draw a random number $u \sim \text{Uniform}(0, 1)$.
   d. If $u < \alpha$, then accept the proposed candidate: $\boldsymbol{\theta}_1^{(i)} = \boldsymbol{\theta}_1^{\text{cand}}$; otherwise reject the candidate and use $\boldsymbol{\theta}_1^{(i-1)}$ as $\boldsymbol{\theta}_1^{(i)} : \boldsymbol{\theta}_1^{(i)} = \boldsymbol{\theta}_1^{(i-1)}$.
3. Stop if i reaches the desired sample size.

The independent Metropolis-Hastings algorithm is used, where the proposal distribution $q(\boldsymbol{\theta}_1)$ does not depend on $\boldsymbol{\theta}_1^{(i-1)}$, which is efficient when the proposal distribution is close to the posterior. To obtain a posterior $\boldsymbol{\theta}_2$ sample, i.e., a sample from $\boldsymbol{\theta}_2|\mathbf{t}$, the same algorithm is implemented. The prior joint distributions $\omega_i(\boldsymbol{\theta}_i)$ are used as the proposal distribution for sampling each $\boldsymbol{\theta}_i$.