International Journal of Production Research

ISSN: 0020-7543 (Print) 1366-588X (Online) Journal homepage: https://www.tandfonline.com/loi/tprs20

Reliability assessment of high-Quality new products with data scarcity

Cai Wen Zhang, Rong Pan & Thong Ngee Goh

International Journal of Production Research, 2020
https://doi.org/10.1080/00207543.2020.1758355

Reliability assessment of high-Quality new products with data scarcity

Cai Wen Zhang[a*], Rong Pan[b] and Thong Ngee Goh[c]

[a]*Business School, Sun Yat-sen University, Guangzhou, People's Republic of China;* [b]*School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, USA;* [c]*Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore*

(*Received 14 January 2020; accepted 13 April 2020*)

This study concerns the reliability assessment in high-quality new product development in which there is scarcity of data resulting from few or zero failures or the unavailability of failure time information. In such circumstances, traditional reliability assessment methods tend to be inadequate and ineffective. This paper describes a pragmatic approach adopted to address this practical issue. A Bayesian method using reparameterization of the Weibull distribution is proposed, which elicits priors in a meaningful way from technical experts and based on historical data. Unlike existing procedures found in the literature, the method here is developed from the perspective of availability of failure time data. Through a case study from the hard disk drive industry, it is demonstrated that the proposed method can provide an effective and practical solution to the challenging real-life problem. Furthermore, it is shown that failure time information has a significant effect on the inference about the Weibull shape parameter.

Keywords: Weibull distribution; high-quality performance; data scarcity; reliability assessment; hard disk drive

Introduction

An increasing challenge in the modern world is posed by the two ends of the spectrum of data availability. At one end, in such areas as e-commerce, internet search, fintech, social media and urban informatics, we are inundated by the high volume, high variety of data generated at high velocity from various sources – the commonly known Big Data phenomenon (e.g. Kuo and Kusiak 2019; Zhan, Tan, and Huo 2019). At the other end, we are starved of needed purpose-oriented data. For example, during the early stages of the development of high-quality products, very limited failure data are available (e.g. Guida and Pulcini 2002, 2009). Similarly, in high-quality or zero-defect manufacturing processes, failures are rare (Psarommatis et al. 2020). In particular, at the conceptual level in Six Sigma quality management, on the average there are merely 3.4 failures out of a million opportunities. Although abundance of data and scarcity of data seem to be opposite situations, the challenges of making informed data-based decisions are fundamentally similar (Goh 2015).

Our focus in this study is on the latter case, namely using scarce data for a useful reliability assessment. There are several possible reasons that could contribute to data scarcity. First, as a result of advances in design, development and manufacturing technologies, many products are becoming increasingly reliable. Second, fierce competition in the market often results in extreme pressure to reduce time-to-market and allows very limited testing time and resources (e.g. Meeker and Hamada 1995; Olteanu and Freeman 2010; Zhang et al. 2013). Third, in some real-life applications (e.g. electronic devices operating in the field in large quantities), even if failures are available, the retrieval of all or part of information on exact failure times could be prohibitively expensive or technically impossible. The consequence is that either there are few or zero failures or – at least partly – the exact failure times are unavailable. We give two real-life illustrations below.

The first example comes from the hard disk drive (HDD) industry. In reliability tests of HDDs, there is a possibility that the exact failure time of a drive may be lost forever due to corruption of the test log. Moreover, because of the high-quality process, few failures are observed even under accelerated reliability test conditions. On the other hand, it is too expensive to retrieve all the failure times from drives returned from the field owing to their large quantities (e.g. Zhang et al. 2013).

The second example is related to the testing of one-shot or single-use devices, which are usually used to make military weapons, air bags in cars, etc. (e.g. Olwell and Sorell 2001; Fan, Balakrishnan, and Chang 2009; Pintar et al. 2012). The only way to test such devices is to detonate (and destroy) them. If a device fails to explode, it is unknown when it has failed. Thus, only binary data can be observed, i.e. either explosion or failure. Because of their critical role in the success

*Corresponding author. Email: zhangcw@mail.sysu.edu.cn

of military weapons or civilian applications, these devices are usually made highly reliable. Even if an accelerated test is conducted, very few or even zero failures may be observed (e.g. Fan, Balakrishnan, and Chang 2009).

Reliability assessment during new product development is critical to meeting the customer's quality expectations as well as to making decisions on warranty policies. However, with data scarcity, the conventional methods of reliability analysis based on large-sample normal approximation do not work well. Neither do methods based on the worst-case estimation (e.g. Hahn and Meeker 1991, 104), which usually give too conservative estimates to be useful (e.g. Meeker and Escobar 1998, 168; Pintar et al. 2012). These kinds of real-life problems in high-quality product development and testing have posed constant challenges to industrial practitioners and statisticians alike (see, e.g. Guida and Pulcini 2002, 2009; Fan, Balakrishnan, and Chang 2009; Zhang et al. 2013; Li and Meeker 2014). This research has arisen from such challenges.

Manufacturing data such as process parameters and in-process quality characteristics are related to the performance and reliability of the final product. This is a good source of information that can be exploited to improve reliability analysis with data scarcity (e.g. He et al. 2016; He et al. 2019). Another valuable source of information is historical data from past products (e.g. Sanchez and Pan 2011). The Bayesian approach provides an effective means of integrating historical data as well as technical expertise from different sources into the current data to improve the quality of the analysis (e.g. Meeker and Hamada 1995; Hamada et al. 2008). Interesting real-life applications to showcase the power and flexibility of the Bayesian approach over the frequentist or non-Bayesian approach in assessing high-reliability data can be found in Fan, Balakrishnan, and Chang (2009), Guida and Pulcini (2009), Pintar et al. (2012), Li and Meeker (2014), to name a few.

Under the Bayesian framework, historical data are usually incorporated into analysis through the prior distributions for some parameters of concern (e.g. Moala, Rodrigues, and Tomazella 2009). However, where the needed prior distributions should come from is a big concern in Bayesian reliability analysis (Li and Meeker 2014). The choice of priors to capture the useful information contained in the historical data and technical expertise is both art and science. The fundamental argument for adopting the Bayesian approach is to augment the quantity and/or quality of data in order to reduce the uncertainty involved in the decision making. Nonetheless, the use of Bayesian methods will inevitably introduce the hyperparameters of the priors for the parameters of concern. If there is no sensible way of specifying the hyperparameters of the priors, then the benefits gained in Bayesian analysis could become illusive or even outweighed by the additional uncertainty introduced. Unfortunately, it is not uncommon that arbitrary values are assumed for the hyperparameters of Bayesian priors (e.g. Leon et al. 2006; Banerjee and Kundu 2008; Fan, Balakrishnan, and Chang 2009).

To improve the quality of reliability analysis of new automotive products, a method of prior elicitation was proposed in Guida and Pulcini (2009) in which objective information from historical failure data is integrated with subjective information on the effectiveness of improvement actions based on available expertise and technical knowledge. The principles reflected in their method is in line with our own experience from industry; such a method is particularly suitable for the evolutionary design of new products.

The Weibull distribution is probably the most widely and successfully applied reliability model. Plenty of literature is available on statistical theory, methodology and applications of the Weibull distribution; see, e.g. Jia et al. (2016), Acitas, Aladag, and Senoglu (2019), Shakhatreh, Lemonte, and Moreno–Arenas (2019). In order to improve the applicability of the Bayesian method proposed, in this study the Weibull distribution is reparameterized so that the new parameters are computationally more stable and the priors for them can be specified in a meaningful way. Furthermore, we address the old problem of Weibull reliability analysis from a new perspective – the availability of failure time information. As a result, new insights are gained into the effect of failure time information on the estimation of Weibull parameters.

The remainder of this paper is organised as follows. Section 2 describes a procedure for eliciting prior information from technical experts and based on historical data. Section 3 presents a method of Bayesian inference on the Weibull distribution without failure time data. This is followed by a Bayesian method proposed for analysing Weibull data with failure time information. A real-life case study from the HDD industry is presented in Section 5. Finally, Section 6 concludes the paper.

Prior elicitation from historical data

When a new product is not revolutionary but the result of improvements on its predecessors, i.e. evolutionary designs, historical data from past products will be a great resource to draw on (e.g. Sanchez and Pan 2011).

Formalisation of historical data

The exploitation of historical data is both art and science. In many real applications, failure data of past-generation products are usually available, but only in the form of number of failures observed during a certain period of time, rather than the exact failure or usage times (e.g. Guida and Pulcini 2002, 2009; Zhang et al. 2013). There are various causes for this deficiency. First, it may be costly or even impractical to track and retrieve the exact failure information. Second, there are various complicating factors involved in the data collecting and reporting, such as erroneous failure times, missing data, delayed reporting, incomplete returns, mixed vintages, different usage conditions, etc.

On the other hand, it is common practice to assign each product failure to one single failure mechanism, which could be the failure of one of its components or other factors such as software bugs or process related defects. Although it is possible that the product might fail for multiple reasons, the primary cause is usually singled out as the chief culprit.

Let $\bar{p}_{0,i}$ represent the fraction of failures out of a reliability test of a certain product due to the i-th failure mechanism estimated based on historical data. During the design of a new model of the product, great effort has been put into the improvements and/or modifications of the old product and processes. But the effectiveness of improvement on a particular failure mechanism is still an uncertain factor. Assume the developer, using their technical expertise and experience, is able to elicit a lower limit $L_{0,i}$ for the probability of the i-th failure mechanism. Following the approach of Guida and Pulcini (2009), the probability of the i-th failure mechanism, denoted by $p_{0,i}$, is assumed to be a uniform random variable distributed over the interval $(L_{0,i}, U_{0,i})$. Here, the upper limit $U_{0,i}$ is given by

$$U_{0,i} = L_{0,i} + \rho_i(\bar{p}_{0,i} - L_{0,i})/4, \tag{1}$$

where ρ_i is the improvement effectiveness factor that takes the values of 1, 2, 3 and 4 according to the following criteria:

1 Significant improvement ($\rho_i = 1$): the improvement involves proven technology and is considered quite effective;
2 Moderate improvement ($\rho_i = 2$): the improvement has been tested and proved to be fairly effective;
3 Minor improvement ($\rho_i = 3$): the improvement involves minor changes with only limited theoretical and/or experimental support;
4 Uncertain or no improvement ($\rho_i = 4$): the improvement involves few or no changes and lacks theoretical and/or experimental support.

If no improvement is made on a certain failure mechanism in the new design, then the failure probability $p_{0,i}$ is assumed to be a constant value of $\bar{p}_{0,i}$.

Now the failure probability of the new product in the reliability test, denoted by p, can be approximated by the sum of the probabilities of individual failure mechanisms as

$$p = \sum_i p_{0,i}. \tag{2}$$

Further assume the random variables $p_{0,i}$ are independent, then the mean and standard deviation of p can be calculated by

$$E[p|\mathrm{PD}] = \sum_i (L_{0,i} + U_{0,i})/2, \tag{3}$$

and

$$\sigma[p|\mathrm{PD}] = \sqrt{\sum_i (U_{0,i} - L_{0,i})^2/12} \tag{4}$$

where the ‘PD’ stands for ‘past data’.

The preceding formalisation of historical data lays the ground for prior specification for the reliability of the new product that follows.

Prior specification for the failure probability of new products

The great popularity and wide success of the Weibull distribution as a reliability model owes as much to its versatility in modelling either increasing or decreasing hazard rates as to the fact that it is one of the limiting distributions of minima, thus providing a good theoretical modelling of the reliability behaviour of a system that consists of many components with independent and identically distributed failure times and that will fail when the first component fails.

The two-parameter Weibull probability density function (pdf) and cumulative distribution function (cdf) can be expressed as

$$f(t;\beta,\lambda) = \beta\lambda t^{\beta-1}e^{-\lambda t^{\beta}}, \tag{5}$$

and

$$F(t;\beta,\lambda) = 1 - e^{-\lambda t^{\beta}}, \quad t > 0, \tag{6}$$

where β is the shape parameter and λ the scale parameter. Note that this parameterisation is slightly different from the conventional one in which the scale parameter is the characteristic life or the approximate 0.632 quantile (e.g. Meeker and Escobar 1998, 85). It is also the method of Weibull specification used in WinBUGS.

For mathematical convenience, conventional Bayesian methods try to make use of conjugate priors for the Weibull parameters. It is well-known that if the shape parameter is unknown, then a continuous joint conjugate prior on both the shape and scale parameters does not exist for the Weibull distribution (e.g. Soland 1966, 1968). If the shape parameter β is assumed known, then a natural conjugate prior for the scale parameter λ is the Gamma distribution.

However, one lacks a sound basis for specifying the hyperparameters of a Gamma prior based on historical data and/or technical expertise because the parameter λ does not have an intuitive meaning or physical interpretation. Consequently, when the Gamma distribution was used as a conjugate prior for the scale parameter, arbitrary values were often assumed for its hyperparameters (e.g. Leon et al. 2006; Kundu 2008; Pan 2009; Kundu and Mitra 2016), and this is problematic.

Alternatively, the Weibull distribution can be reparameterized in the following way. Note that we are mainly concerned with the type I censoring as it is the most common and realistic case in practice. Under such circumstances, we have the following relationship:

$$p = 1 - e^{-\lambda T^{\beta}}, \tag{7}$$

and

$$\lambda = \frac{-\ln(1-p)}{T^{\beta}}, \tag{8}$$

where T is the censoring time and p is the probability of failure for any one unit or the fraction of all units in the population that will fail by time T.

After substitution, we can reparameterize the Weibull pdf and cdf as follows:

$$f(t;\beta,p) = \frac{-\ln(1-p)\beta t^{\beta-1}}{T^{\beta}}(1-p)^{(t/T)^{\beta}}, \tag{9}$$

and

$$F(t;\beta,p) = 1 - (1-p)^{(t/T)^{\beta}}, \quad t > 0. \tag{10}$$

This reparameterization essentially replaces λ with p, which has advantages in the following aspects. First, the parameter p can be estimated non-parametrically (directly) from historical data. Second, when there is heavy censoring, p and β will be computationally more stable than λ and β (e.g. Li and Meeker 2014). Third, the parameter p has a bounded range of (0, 1), compared to the boundless value of λ, which is conducive to the subsequent numerical and simulation analysis.

Note that p is a parameter of the binomial distribution. It is well-known that a natural conjugate prior for the binomial likelihood is the Beta distribution (e.g. Gelman et al. 2004, 39–40). Therefore, the prior of choice for parameter p is a Beta(a, b) distribution, which has a pdf

$$\pi(p) = \frac{1}{B(a,b)}p^{a-1}(1-p)^{b-1}, \quad 0 < p < 1, \tag{11}$$

where a and b are the hyperparameters; $B(a,b)$ is the Beta function.

The formalised historical data in the form of failure probability in a reliability test provide a sound basis for the specification of a and b. The mean and variance of the Beta random variable p are respectively given by

$$E[p] = a/(a+b), \tag{12}$$

and

$$\sigma[p] = \sqrt{ab/[(a+b)^2(a+b+1)]}. \tag{13}$$

Note that the test from which the $\bar{p}_{0,i}$ data were collected is the same type I reliability test analyzed in the Bayesian method. Therefore, by matching (12) with (3) and (13) with (4) and solving the equations, we can obtain reasonable estimates for a and b based on the historical data and technical expertise (Guida and Pulcini 2009). This practicable prior elicitation procedure improves the specification of hyperparameters in the Bayesian method.

Bayesian inference without failure time data

When failure time data are unavailable or there are zero failures, Weibull data from a type I test essentially reduces to binomial data. Consider a sample of n units of product put on test and r failures are detected by the censoring time T. Let p denote the probability of failure for any one unit at the end of the test and X the number of failures out of n tested units. Assume the failures are independent of each other, then random variable X follows a binomial distribution with a probability mass function (pmf)

$$f(X = x|p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad 0 < p < 1. \tag{14}$$

Given a prior of Beta(a, b), it is easy to show that the posterior of parameter p follows a Beta distribution with parameters $(a + x, b + n - x)$, namely

$$\pi(p|x,n) = \frac{f(x|p)\pi(p)}{\int_0^1 f(x|p)\pi(p)dp} = \text{Beta}(p|a + x, b + n - x). \tag{15}$$

When the shape parameter is known

Remember that either of the two parameters λ and β can be expressed as a function of p and the other parameter. If the shape parameter β is assumed known, then Bayesian inference on the scale parameter λ can be made by using the change of variable method. This method, unlike traditional ones, offers a convenient way of deriving the posteriors in explicit form. First, differentiate λ with respect to p, we get

$$\frac{d\lambda}{dp} = \frac{1}{T^\beta(1-p)}. \tag{16}$$

Given $0 < p < 1$, this derivative is positive, meaning λ is a monotonically increasing function of p. Then the posterior distribution of the scale parameter λ can be derived as

$$\pi(\lambda|\beta,x,n) = \frac{T^\beta}{B(a+x, b+n-x)}(1 - e^{-\lambda T^\beta})^{a+x-1} e^{-\lambda T^\beta (b+n-x)}. \tag{17}$$

Similarly, we can derive the posterior distribution for some commonly used reliability measures. First, consider the mean time to failure (MTTF), denoted by μ, which is given by

$$\mu = \lambda^{-1/\beta}\Gamma(1 + 1/\beta). \tag{18}$$

The posterior distribution of the MTTF is given by

$$\begin{aligned} \pi(\mu|\beta,x,n) = {} & \frac{1}{\Gamma\left(1+\frac{1}{\beta}\right)} \frac{\Gamma(a+b+n)}{\Gamma(a+x)\Gamma(b+n-x)} \left[1 - \exp\left[-\left(\frac{T\Gamma\left(1+\frac{1}{\beta}\right)}{\mu}\right)^\beta\right]\right]^{a+x-1} \\ & \times \exp\left[-\left(\frac{T\Gamma(1+1/\beta)}{\mu}\right)^\beta (b+n-x)\right] \frac{\beta T^\beta}{(\mu/\Gamma(1+1/\beta))^{\beta+1}}. \end{aligned} \tag{19}$$

Likewise, let t_q denote the qth quantile of the Weibull distribution, then the posterior distribution of t_q is given by

$$\begin{aligned} \pi(t_q|\beta,x,n) = {} & \frac{1}{[-\ln(1-q)]^{1/\beta}} \frac{\Gamma(a+b+n)}{\Gamma(a+x)\Gamma(b+n-x)} \left[1 - \exp\left[-\left(\frac{T}{t_q}\right)^\beta (-\ln(1-q))\right]\right]^{a+x-1} \\ & \times \exp\left[-\left(\frac{T}{t_q}\right)^\beta (-\ln(1-q))(b+n-x)\right] \frac{\beta T^\beta [-\ln(1-q)]^{(1+1/\beta)}}{t_q{}^{\beta+1}}. \end{aligned} \tag{20}$$

These posterior distributions are useful for reliability analysis. For example, the posterior mean can minimise the posterior variance with respect to a point estimate of a parameter (Carlin and Louis 2009, 42). So, it is a favourite choice of point estimate, although other measures such as the posterior median or mode are possible choices as well. Interval estimates for the reliability measures can be easily calculated from their individual distributions. By looking at the entire distribution, decision makers can have a bigger picture of the problem of concern and thus can better quantify the uncertainty and risks involved.

When the shape parameter is unknown

It is well-known that the shape parameter of the Weibull distribution is much more informative and important than the scale parameter as it is closely related to the failure mechanisms of a product. Some existent research assumed a prior on the shape parameter β having a range of $(0, \infty)$ (e.g. Kundu 2008; Kundu and Mitra 2016). However, it is well-known that the Weibull shape parameter usually has a narrow range of value for particular products (e.g. Nelson 1985; Abernethy 2006; Pascual 2010; Genschel and Meeker 2010; Zhang, Ye, and Xie 2017). Therefore, without further information on the distributional characteristics, it is reasonable to assume that β is uniformly distributed over the interval $[\beta_L, \beta_U]$ (e.g. Erto and Guida 1985; Guida and Pulcini 2002, 2009; Li and Meeker 2014). That is, the prior density of β is given by

$$\pi(\beta) = \frac{1}{\beta_U - \beta_L}, \quad \beta_L \le \beta \le \beta_U, \tag{21}$$

where β_L and β_U represent the lower and upper bounds, respectively.

Although we may not be able to provide a guideline for specifying the range of the shape parameter value that is generally applicable, in engineering applications this kind of domain knowledge and/or historical data are usually available. For example, bearing life data from the field are usually adequately represented by a Weibull distribution with the shape parameter $\beta = 1.5$ (Nelson 1985). Another example is that the Weibull shape parameter value for a certain class of hard disk drive in a typical reliability demonstration test is usually around 0.7 (Zhang et al. 2013). Li and Meeker (2014) proposed a Bayesian method to analyze the reliability of bearing cages in which the informative prior specified for β has a value range between 1.5 and 3.0 with a 99% confidence. A prior for parameter β uniformly distributed over the interval [1, 5] was used to model the reliability of automobile components in Guida and Pulcini (2009).

We temporarily assume the parameters p and β are prior-independent (e.g. Meeker and Escobar 1998) and will pick up on this point later in Section 5. Then the joint posterior distribution of p and β is given by

$$\pi(p, \beta|\mathrm{AD}) = \frac{\frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1}\binom{n}{x} p^x (1-p)^{n-x} \frac{1}{\beta_U - \beta_L}}{\int_0^1 \int_{\beta_L}^{\beta_U} \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1}\binom{n}{x} p^x (1-p)^{n-x} \frac{1}{\beta_U - \beta_L} d\beta dp} = \frac{p^{a+x-1}(1-p)^{b+n-x-1}}{(\beta_U - \beta_L)B(a+x, b+n-x)}, \tag{22}$$

where the ‘AD’ stands for ‘All Data’.

Then the marginal posterior distribution of p is given by

$$\pi(p|\mathrm{AD}) = \int_{\beta_L}^{\beta_U} (p, \beta|\mathrm{AD}) d\beta = \frac{p^{a+x-1}(1-p)^{b+n-x-1}}{B(a+x, b+n-x)} = \mathrm{Beta}(p|a+x, b+n-x). \tag{23}$$

Note that the posterior distribution of p in equation (23) is the same as the one in equation (15). Similarly, it is easy to show that the posterior of β is the same as the prior, namely $\pi(\beta|\mathrm{AD}) = \pi(\beta)$. As a result, p and β are posterior-independent because $\pi(p, \beta|\mathrm{AD}) = \pi(p|\mathrm{AD})\pi(\beta|\mathrm{AD})$.

The interpretations are twofold. First, failure observations without the exact times are not helpful in updating the inference on the shape parameter β. Second, without the failure time information, the inference on parameter p does not depend on the shape parameter. However, are these interpretations accurate? Remember that we have assumed that p and β are prior-independent. Is this assumption reasonable? We will attempt to answer these questions in Section 5.

Bayesian inference with failure time data

Analysis in the preceding section reveals the importance of failure time information in the Bayesian inference on the shape parameter β. Therefore, failure time data must be incorporated into the analysis in order to draw meaningful inferences on β.

Consider a reliability test with a sample size n and a test duration T. At the end of the test, x_1 failure times are recorded, x_2 observations are left-censored without failure times, and the remaining observations are survivors or right-censored. Then the likelihood of the test data is given by

$$L(p,\beta|\mathrm{TD}) = \prod_{i=1}^{x_1}\left\{\frac{-\ln(1-p)\beta}{T^\beta}t_i^{\beta-1}(1-p)^{(t_i/T)^\beta}\right\}p^{x_2}(1-p)^{n-x_1-x_2}, \tag{24}$$

where t_i denote the exact failure times and the 'TD' stands for 'Test Data'.

Again, assuming p and β are prior-independent, the joint posterior distribution of p and β can be written as

$$\begin{aligned}\pi(p,\beta|\mathrm{AD}) &= \frac{L(p,\beta|\mathrm{TD})\frac{1}{B(a,b)}p^{a-1}(1-p)^{b-1}\frac{1}{\beta_U-\beta_L}}{\int_0^1\int_{\beta_L}^{\beta_U}L(p,\beta|\mathrm{TD})\frac{1}{B(a,b)}p^{a-1}(1-p)^{b-1}\frac{1}{\beta_U-\beta_L}d\beta dp}\\ &= \frac{p^{a-1}(1-p)^{b-1}}{D}L(p,\beta|\mathrm{TD}),\end{aligned} \tag{25}$$

where the normalising constant D is given by

$$D = \int_0^1\int_{\beta_L}^{\beta_U}L(p,\beta|\mathrm{TD})p^{a-1}(1-p)^{b-1}d\beta dp. \tag{26}$$

Then the marginal posterior distributions of p and β can be obtained by integrating the joint posterior density with respect to each other as

$$\pi(p|\mathrm{AD}) = \frac{p^{a-1}(1-p)^{b-1}}{D}\int_{\beta_L}^{\beta_U}L(p,\beta|\mathrm{TD})d\beta, \tag{27}$$

and

$$\pi(\beta|\mathrm{AD}) = \frac{1}{D}\int_0^1 L(p,\beta|\mathrm{TD})p^{a-1}(1-p)^{b-1}dp. \tag{28}$$

Furthermore, let $g(p,\beta)$ denote any function of p and β, which includes the MTTF and the qth quantile t_q as special cases. Then, under *squared error loss*, the Bayesian estimate of $g(p,\beta)$ can be calculated by

$$\hat{g}(p,\beta) = \frac{1}{D}\int_0^1\int_{\beta_L}^{\beta_U}g(p,\beta)L(p,\beta|\mathrm{TD})p^{a-1}(1-p)^{b-1}d\beta dp. \tag{29}$$

Apparently, the posterior analysis of p, β and $\hat{g}$ has to be performed using numerical methods or Markov chain Monte Carlo (MCMC) simulation.

The full conditional distributions of p and β can be derived in closed form from (25) as follows:

$$\pi(p|\beta,\mathrm{AD}) \propto [-\ln(1-p)]^{x_1}p^{x_2}(1-p)^{n-x_1-x_2+\sum_{i=1}^{x_1}(t_i/T)^\beta}, \tag{30}$$

and

$$\pi(\beta|p,\mathrm{AD}) \propto \frac{\beta^{x_1}}{T^{\beta x_1}}(1-p)^{\sum_{i=1}^{x_1}(t_i/T)^\beta}\prod_{i=1}^{x_1}t_i^{\beta-1}. \tag{31}$$

Given the full conditionals of p and β, the MCMC algorithm of choice is the Gibbs sampler. With the WinBUGS package available for free, the MCMC simulation can be carried out relatively easily.

Application case study

As mentioned, this research has been motivated by challenges faced in the real practice of new product development. One industry in which we have first-hand experience is the HDD industry. It is a widely accepted fact that the reliability of HDDs is well modelled by the Weibull distribution (e.g. Yang and Sun 1999; Elerath and Shah 2004; Sun and Zhang 2007; Zhang et al. 2013; Zhang, Ye, and Xie 2017). During the reliability demonstration test (RDT) of new products, few failures were often observed, which left quality analysts in a dilemma. On the one hand, they could not afford testing more HDDs or for a longer time for various practical reasons. On the other hand, they had to take a decision on the product's reliability and

Table 1. Historical data on failure mechanisms of HDD product A.

Category	Failure Mechanism	$p_{0,i}$	ρ_i	$L_{0,i}$	$U_{0,i}$
M1	Scratch	0.00213	2	0.001	0.00157
M2	Erasure	0.00213	2	0.001	0.00157
M3	Disk defect	0.00045	3	0.0004	0.000437
M4	Degraded head	0.00034	2	0.0003	0.000318
M5	Poor writing	0.00022	1	0.0001	0.000131
M6	Motor	0.00022	4	0.0002	0.000224
M7	Head related	0.00045	4	0.0004	0.000449
M8	Card	0.00022	4	0.0002	0.000224
M9	Others	0.00034	4	0.0003	0.000337

Table 2. MCMC simulation results for the posterior β in the six scenarios.

Scenario	Failure Data	Mean	Std. Dev.	5% CL	Median	95% CL	MLE
S1	All 5 failures left-censored	1.14	0.27	0.62	1.20	1.47	0.42
S2	1 time {86} and 4 failures left-censored	1.04	0.31	0.46	1.08	1.47	0.41
S3	2 times {86, 286} and 3 failures left-censored	1.09	0.27	0.60	1.14	1.46	0.54
S4	3 times {86, 286, 341} and 2 failures left-censored	1.09	0.26	0.63	1.11	1.46	0.63
S5	4 times {86, 286, 341, 446} and 1 failure left-censored	0.99	0.27	0.54	0.99	1.43	0.72
S6	All 5 failure times {86, 286, 341, 446, 536}	0.88	0.28	0.43	0.87	1.37	0.81

evaluated the risks involved. As a rule of thumb, it is quite unreliable to estimate the Weibull parameters using the classical maximum likelihood estimation (MLE) if the number of failures in the data is less than five. It typically requires at least 10 failures for the MLE to generate accurate estimates of the Weibull parameters (Tobias and Trindade 1995, 95).

From a practitioner's perspective, the only reasonable way to solve or at least mitigate the problem is to leverage historical data and technical expertise. The Bayesian method proposed can serve such a purpose, which involves subjective judgment about the effectiveness of improvement actions.

In Table 1 we present a set of historical data on the failure mechanisms of a model of HDD, named product A, discovered in a certain reliability test. The failures of product A could be caused by component failures, process defects, contamination, etc., which were classified into nine categories. The estimates of the fraction failing $p_{0,i}$ in the test were calculated from the historical data.

Because the historical data were from the same test as the RDT, one can directly match equations (12) and (13) with (3) and (4), respectively. Solving the equations yields the values of $a = 385.6$ and $b = 83873.6$. In other words, these values have captured, more or less, the information elicited from technical experts and based on the historical data.

During the RDT of product B, a sample of 1,200 HDDs were put on test for 1,000 h. At the end of the test, five failures were observed at the times: 86, 286, 341, 446, and 536 in hours. The remaining 1,195 HDDs were right-censored.

Effect of failure time information

In order to examine the effect of failure time information on Bayesian inference about the Weibull parameters, we have designed six scenarios as described in Table 2. In Scenario 1 (S1 for short), the failure data consist of five failures that are assumed to have all lost their exact times, i.e. left-censored. In S2, the failure data consist of five failures that are assumed to have only one failure time recorded with the other four left-censored. Likewise, in S3, two failure times are recorded and the other three left-censored. In S4, three failure times are recorded and the other two left-censored. In S5, four failure times are recorded and only one left-censored. In S6, all five failure times are available.

Based on the historical data from similar products, the value of parameter β is expected to be around 0.7 (e.g. Zhang et al. 2013; Zhang, Ye, and Xie 2017). Hence, a uniform distribution over the interval [0.3, 1.5] is considered to be an informative prior for parameter β. MCMC simulations are implemented in WinBUGS 1.4. Each simulation runs two chains with different initial values and each chain runs for 500,000 iterations with 10,000 burn-in samples. The statistics of the posterior β are presented in Table 2. The posterior distributions of β in the six scenarios are plotted in Figure 1.

The density plot in S1 shows the posterior distribution of β when no failure time information is available in the data. Recall that the prior distribution of β used in the simulation is a uniform distribution over the interval [0.3, 1.5]. Obviously, the posterior distribution is a bit far from the prior one, contrasting with the result obtained in Section 3, namely $\pi(\beta|\mathrm{AD}) = \pi(\beta)$. The reason behind this contradiction is that the assumption that p and β are prior-independent does not

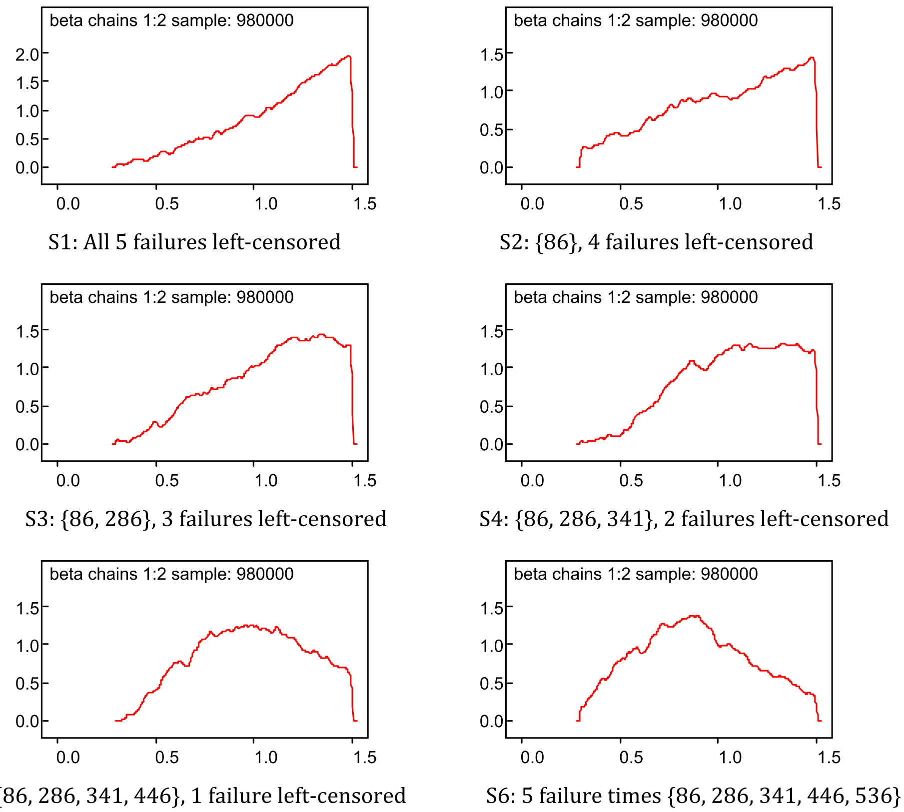

Figure 1. Posterior density plots of β in the six scenarios.

Table 3. MCMC simulation results for posterior p in the 6 scenarios.

Scenario	Failure data	Mean	Std. Dev.	5% CL	Median	95% CL	MLE
S1	All 5 failures left-censored	0.00458	0.00023	0.00420	0.00457	0.00496	0.00417
S2	1 time {86} and 4 failures left-censored	0.00458	0.00023	0.00420	0.00457	0.00496	0.00417
S3	2 times {86, 286} and 3 failures left-censored	0.00458	0.00023	0.00421	0.00458	0.00496	0.00417
S4	3 times {86, 286, 341} and 2 failures left-censored	0.00458	0.00023	0.00420	0.00457	0.00496	0.00417
S5	4 times {86, 286, 341, 446} and 1 failure left-censored	0.00458	0.00023	0.00420	0.00457	0.00496	0.00417
S6	All 5 failure times {86, 286, 341, 446, 536}	0.00457	0.00023	0.00420	0.00457	0.00496	0.00417

hold here. This analysis sheds light on the extent to which the independence assumption may hold. Nonetheless, as there is no easy way to quantify the correlation between the two parameters, such an assumption might still be considered acceptable for mathematical convenience (e.g. Erto and Guida 1985; Meeker and Escobar 1998, 348; Li and Meeker 2014).

On the other hand, by sequentially adding the failure time information into the data set, the posterior distribution of β has been updated step by step from S2 to S6. It is manifest that the failure time information has a major effect on the posterior distribution of β. It can be seen that as the amount of failure time information increases, the Bayesian inference on β becomes more credible and accurate. Particularly, the Bayes estimate of β in S6 is 0.88, which is a credible value matching the expectation based on historical data and experience. For comparison, the last column of Table 2 also presents the ML estimates of β in all six scenarios. It is also clear that the failure time information has a significant effect on the ML estimate of β. In other words, if the failure time data are totally unavailable like the case in Fan, Balakrishnan, and Chang (2009), it may not be possible to make a useful estimate of parameter β. This is probably the reason why an exponential lifetime model was used therein.

The configuration of parameter p is much simpler than that of parameter β because the range of p is (0, 1) by nature and no tweaks are necessary. The statistics of the posterior p are presented in Table 3. The posterior densities of p are almost identical in all six scenarios and thus only one of them is shown in Figure 2 for brevity.

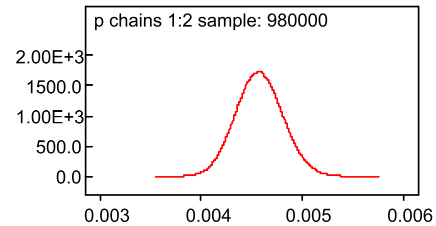

Figure 2. Posterior density plot of p in one of the six scenarios.

Table 4. MCMC simulation results for posterior β with different priors.

Prior	β Range	Mean	Std. Dev.	5% CL	Median	95% CL
Informative prior	[0.3, 1.5]	0.88	0.28	0.43	0.87	1.37
Diffuse prior	[0.1, 4.0]	0.95	0.39	0.37	0.91	1.67
Diffuse prior	[0.05, 4.0]	0.96	0.36	0.44	0.92	1.62
Diffuse prior	[0.1, 8.0]	0.93	0.40	0.39	0.86	1.65

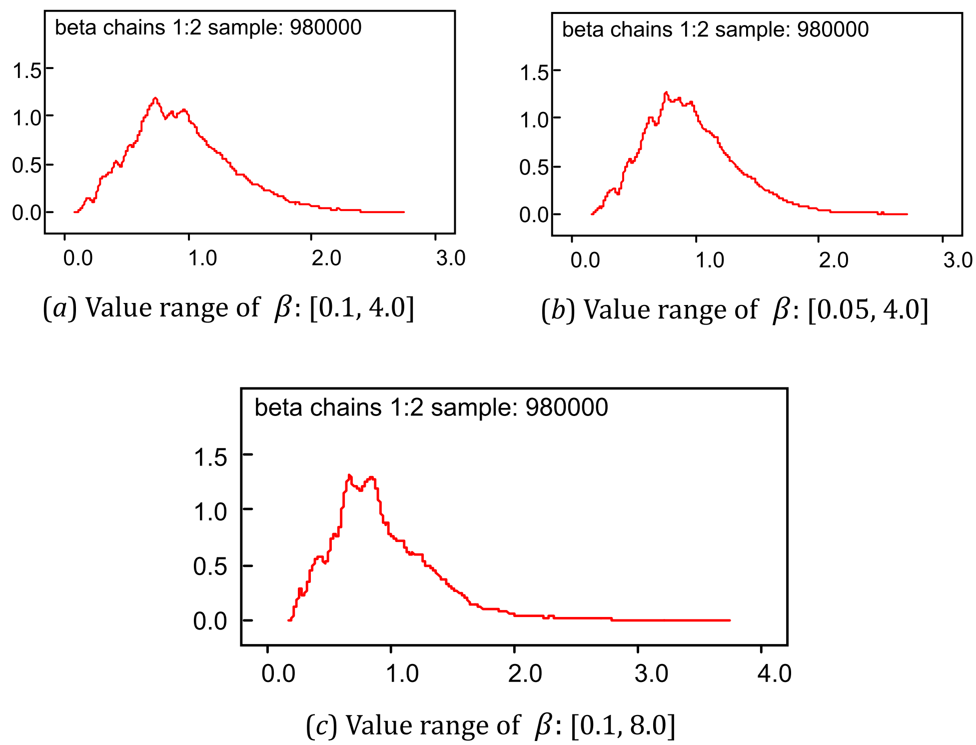

Figure 3. Posterior density plots of β given different priors.

By examining the statistics and density plots, it is clear that the failure time information almost has no effect on the Bayesian inference about the parameter p. This point is not so straightforward by just looking at the equation (27). It is interesting to note that the failure time information almost has no effect on the ML estimate of p either.

Effect of different priors

We can further examine the effect of the choice of prior on the posterior of β. We have used a uniform distribution over the interval [0.3, 1.5] as an informative prior for parameter β. For comparison, a uniform distribution over the interval [0.1, 4] is considered to be a diffuse prior. From industrial experience, a value of β outside this range would not be expected. Two other choices of diffuse prior, namely uniform distributions over the intervals [0.05, 4] and [0.1, 8], are also considered for the purpose of sensitivity analysis. With all five failure times available, MCMC simulations are performed in the same manner as above. The results are summarised in Table 4 and displayed in Figure 3. Note that the statistics in the first row of Table 4 are the same as those in the last row of Table 2.

By comparing the posterior densities of β generated with the three different priors, it can be seen that a uniform distribution over the interval [0.1, 4.0] is a reasonable choice of diffuse prior for parameter β. When extending the range of prior to either [0.05, 4.0] or [0.1, 8.0], it has little effect on the posterior distribution of β. In particular, the likelihood beyond 3.0 is all but negligible.

On the other hand, a diffuse prior for p is the Beta(1, 1) distribution, which can be interpreted as having observed 0 failures and 0 survivors previously (e.g. Pintar et al. 2012). Theoretically, when a diffuse or non-informative joint prior is used, the Bayes estimates of β and p should agree with their respective ML estimates. As suggested in Li and Meeker (2014), under such circumstances, the median of the marginal posterior distribution is less affected by the long tail of a skewed posterior distribution. As a result, we will use the median, instead of the mean, of the posterior β or p as its Bayes estimate in this case. When a uniform distribution over the interval [0.1, 4] is used as a diffuse prior for β and Beta(1, 1) as a diffuse prior for p, the Bayes estimates (i.e. the medians) of β and p are 0.885 and 0.00473, respectively. In comparison, their ML estimates are 0.806 and 0.00417, respectively.

In summary, given the informative priors, for the RDT of product B the Bayes estimates of β and p are 0.880 and 0.00457, respectively. The resultant estimate of the MTTF is 484,383. This result matched the judgment of the quality analysts based on their experience and technical analysis. In the end, they were satisfied with the method proposed.

Conclusions

Two issues arising from the development of high-quality new products have motivated this research, namely data scarcity and Bayesian inference on the Weibull distribution. Data scarcity refers to either few or zero failures or the unavailability of failure time information. The second issue is well-known to be a difficult proposition because there does not exist a continuous joint conjugate prior for both the shape and scale parameters of the Weibull distribution. With data scarcity, quality analysts often find themselves in a dilemma. On the one hand, reliability analysis using traditional methods are likely to be unreliable. On the other hand, they do not have the luxury of testing more products or for a longer time to augment the data. A practical and effective Bayesian method is proposed in this study to address these challenges. Quality practitioners may choose to tweak some of the criteria settings to suit their own needs and contexts.

The Bayesian method is developed from the perspective of availability of failure time information. When the failure time information is unavailable and the shape parameter is assumed known, it is shown that Bayesian inference on the Weibull parameters along with other reliability measures can be derived in closed form.

A real-life case study is used to demonstrate the applicability of the method proposed. Through the case study, it is revealed that the failure time information has a significant effect on the inference about the Weibull shape parameter, which in turn affects the estimation of the scale parameter. Therefore, if the failure time data are totally unavailable, it may not be possible to make a precise estimate of the Weibull parameters. It is also demonstrated that the parameters β and p are not prior-independent, although they might be assumed independent for mathematical convenience in a Bayesian analysis and MCMC simulation.

Finally, as a word of caution, although Bayesian methods offer a convenient way of incorporating historical data and technical knowledge into a formal statistical analysis, their applications also involve potential higher risks of being misused. In particular, it is important to refrain from mistaking wishful thinking for prior information, as has been pointed out before (Li and Meeker 2014).

Acknowledgements

The authors would like to thank the Associate Editor and two anonymous reviewers for their valuable comments and suggestions that have significantly improved the paper. Dr. Zhang's work is supported by the National Natural Science Foundation of China under Grant No. 71771221.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Abernethy, R. B. 2006. *The New Weibull Handbook*. 5th ed. North Palm Beach, FL: R. B. Abernethy.

Acitas, S., C. H. Aladag, and B. Senoglu. 2019. "A New Approach for Estimating the Parameters of Weibull Distribution via Particle Swarm Optimization: An Application to the Strengths of Glass Fibre Data." *Reliability Engineering and Systems Safety* 183: 116–127.

Banerjee, A., and D. Kundu. 2008. "Inference Based on Type-II Hybrid Censored Data from a Weibull Distribution." *IEEE Transactions on Reliability* 57 (2): 369–378.

Carlin, B. P., and T. A. Louis. 2009. *Bayesian Methods for Data Analysis*. 3rd ed. Boca Raton, FL: Chapman & Hall/CRC.

Elerath, J. G., and S. Shah. 2004. "Server Class Disk Drives: how Reliable are They?" In *Proceedings of the Annual Reliability and Maintainability Symposium*, 151–156. IEEE.

Erto, P., and M. Guida. 1985. "Estimation of Weibull Reliability from Few Life Test." *Quality and Reliability Engineering International* 1: 161–164.

Fan, T. H., N. Balakrishnan, and C. C. Chang. 2009. "The Bayesian Approach for Highly Reliable Electro-explosive Devices Using One-shot Device Testing." *Journal of Statistical Computation and Simulation* 79 (9): 1143–1154.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. *Bayesian Data Analysis*. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC.

Genschel, U., and W. Q. Meeker. 2010. "A Comparison of Maximum Likelihood and Median-rank Regression for Weibull Estimation (with Discussions)." *Quality Engineering* 22 (4): 236–255.

Goh, T. N. 2015. "Emerging Megatrends in Quality Engineering and the 'New 5S' Response." *Quality Engineering* 27 (4): 450–460.

Guida, M., and G. Pulcini. 2002. "Automotive Reliability Inference Based on Past Data and Technical Knowledge." *Reliability Engineering and Systems Safety* 76: 129–137.

Guida, M., and G. Pulcini. 2009. "Bayesian Reliability Inference on Innovated Automotive Components." In *Statistics for Innovation*, edited by P. Erto, 193–211. Milano: Springer-Verlag Italia.

Hahn, G., and W. Meeker. 1991. *Statistical Intervals: A Guide for Practitioners*. Hoboken, NJ: John Wiley & Sons.

Hamada, M. S., A. G. Wilson, C. S. Reese, and H. F. Martz. 2008. *Bayesian Reliability*. New York, NY: Springer.

He, Y., L. Wang, Z. He, and X. Xiao. 2016. "Modelling Infant Failure Rate of Electromechanical Products with Multilayered Quality Variations From Manufacturing Process." *International Journal of Production Research* 54 (21): 6594–6612.

He, Y., Y. Zhao, Y. Wei, A. Zhang, and D. Zhou. 2019. "Modeling and Detecting Approach for the Change Point of Electronic Product Infant Failure Rate." *Microelectronics Reliability* 99: 222–231.

Jia, X., D. Wang, P. Jiang, and B. Guo. 2016. "Inference on the Reliability of Weibull Distribution with Multiply Type-I Censored Data." *Reliability Engineering and Systems Safety* 150: 171–181.

Kundu, D. 2008. "Bayesian Inference and Life Testing Plan for Weibull Distribution in Presence of Progressive Censoring." *Technometrics* 50: 144–154.

Kundu, D., and D. Mitra. 2016. "Bayesian Inference of Weibull Distribution Based on Left Truncated and Right Censored Data." *Computational Statistics and Data Analysis* 99: 38–50.

Kuo, Y. H., and A. Kusiak. 2019. "From Data to big Data in Production Research: the Past and Future Trends." *International Journal of Production Research* 57 (15–16): 4828–4853.

Leon, R. V., R. Ramachandran, A. J. Ashby, and J. Thyagarajan. 2006. "Bayesian Modeling of Accelerated Life Tests with Random Effects." *Journal of Quality Technology* 39 (1): 1–14.

Li, M., and W. Q. Meeker. 2014. "Application of Bayesian Methods in Reliability Data Analysis." *Journal of Quality Technology* 46 (1): 1–23.

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

Meeker, W. Q., and M. Hamada. 1995. "Statistical Tools for the Rapid Development & Evaluation of High-reliability Products." *IEEE Transactions on Reliability* 44 (2): 187–198.

Moala, F. A., J. Rodrigues, and V. L. D. Tomazella. 2009. "A Note on the Prior Distributions of Weibull Parameters for the Reliability Function." *Communications in Statistics—Theory and Methods* 38 (7): 1041–1054.

Nelson, W. 1985. "Weibull Analysis of Reliability Data with Few or No Failures." *Journal of Quality Technology* 17 (3): 140–146.

Olteanu, D., and L. Freeman. 2010. "The Evaluation of Median-rank Regression and Maximum Likelihood Estimation Techniques for a Two-parameter Weibull Distribution (with Discussions)." *Quality Engineering* 22 (4): 256–272.

Olwell, D. H., and A. A. Sorell. 2001. "Warranty Calculations for Missiles with Only Current-status Data, Using Bayesian Methods." In *Proceedings of Annual reliability and Maintainability Symposium*, 133–138.

Pan, R. 2009. "A Bayes Approach to Reliability Prediction Utilizing Data from Accelerated Life Tests and Field Failure Observations." *Quality and Reliability Engineering International* 25: 229–240.

Pascual, F. 2010. "EWMA Charts for the Weibull Shape Parameter." *Journal of Quality Technology* 42 (4): 400–416.

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: 473–485.

Psarommatis, F., G. May, P. A. Dreyfus, and D. Kiritsis. 2020. "Zero Defect Manufacturing: State-of-the-art Review, Shortcomings and Future Directions in Research." *International Journal of Production Research* 58 (1): 1–17.

Sanchez, L. M., and R. Pan. 2011. "An Enhanced Parenting Process: Predicting Reliability in Product's Design Phase." *Quality Engineering* 23 (4): 378–387.

Shakhatreh, M. K., A. J. Lemonte, and G. Moreno–Arenas. 2019. "The log-Normal Modified Weibull Distribution and its Reliability Implications." *Reliability Engineering and Systems Safety* 188: 6–22.

Soland, R. M. 1966. *Use of the Weibull distribution in Bayesian decision theory*. Research Analysis Corporation, McLean, VA, Report no. RAC-TP-225.

Soland, R. M. 1968. “Bayesian Analysis of the Weibull Process with Unknown Scale Parameter and its Application to Acceptance Sampling.” *IEEE Transactions on Reliability* R-17 (2): 84–90.

Sun, F. B., and S. Zhang. 2007. “Does Hard Disk Drive Failure Rate Enter Steady-state After one Year?” In *Proceedings of the Annual reliability and Maintainability Symposium*, 356–361. Orlando, FL, USA, IEEE.

Tobias, P. A., and D. C. Trindade. 1995. *Applied Reliability*. 2nd ed. New York: Chapman & Hall/CRC.

Yang, J., and F. Sun. 1999. “A Comprehensive Review of Hard-Disk Drive Reliability.” In *Proceedings of the Annual reliability and Maintainability Symposium*, 403–409. IEEE.

Zhan, Y., K. H. Tan, and B. Huo. 2019. “Bridging Customer Knowledge to Innovative Product Development: A Data Mining Approach.” *International Journal of Production Research* 57 (20): 6335–6350.

Zhang, C. W., Z. S. Ye, and M. Xie. 2017. “Monitoring the Shape Parameter of a Weibull Renewal Process.” *IISE Transactions* 49 (8): 800–813.

Zhang, C. W., T. Zhang, D. Xu, and M. Xie. 2013. “Analyzing Highly Censored Reliability Data Without Exact Failure Times: An Efficient Tool for Practitioners.” *Quality Engineering* 25 (4): 392–400.