Design of Reliability Test Specification for a New Product with Unknown Failure Modes

Neda Shafiei [1], Jeffrey W. Herrmann [1], Aaron Krive [2], Maxim P. Nikiforov [2], Guneet Sethi [2], and Mohammad Modarres [1]

1. University of Maryland, College Park, MD, U.S.
2. Amazon Lab 126, CA, U.S.

Abstract

This study proposes an approach for designing test specifications for a new product with unknown failure modes. This approach determines the sample size, and test conditions for reliability testing based on the target reliability, its confidence level, and manufacturer's constraints including the test duration and maximum sample size. The approach obtains various use conditions (i.e., stress profiles and frequencies of use by the owners) from the same class of devices, such as earlier versions of the product already in use, by relying on customer survey data. Anticipated users (customers) applied stresses are summarized by introducing a normalized metric called stress-index (SI). The SI values and frequencies of customer use are clustered into various groups. Then, (1) a frequency-accelerated test is designed that applies the grouped use stresses to test samples but increases their frequencies, and (2) a stress-accelerated test is designed that uses a stress-life model to convert the actual customer use frequencies into higher stress reliability test frequencies, such that the same amount of damage as the use condition is accumulated. The number of samples for the frequency-accelerated and stress accelerated tests is calculated using the binomial distribution representing reliability of the products. A point estimate reliability and its lower bound can be estimated based on the test outcome. The application of the approach is illustrated for designing a reliability test plan using a simulated survey dataset for a hypothetical electronic device.

Key Words: test specification, frequency-accelerated test, stress-accelerated test, reliability estimation, user survey.

Definitions

Stress	An agent that causes damage to the product.
Stress adjustor	A variable that increases or decreases the stress magnitude or stress absorption.
Stress profile	A combination of stresses and stress adjustors.

1. Introduction

Manufacturers assess the reliability of their products using laboratory tests to understand whether they meet or exceed their minimum reliability requirement. A reliability test follows specifications including sample size, test conditions (i.e., stress levels and their frequencies of occurrence), the analytical procedures for assessing reliability, and the acceptance criteria (e.g., numerical limits) that are used to evaluate the reliability of a product. If the test results of the sample are within the acceptance criteria, the product can be released for mass production. Otherwise, a root cause analysis is needed to understand the reason for the poor test results [1, 2].

Yang [3] divided the reliability demonstration test methods into bogey test (or zero-failure test), life test, and degradation test. In a bogey test, the sample size, test time, and level of stresses are predetermined; and the target reliability is achieved if no units fail during the test. Yang argued that the reliability cannot be estimated if any samples fail in a bogey test. Life test methods include sequential life tests (e.g., step-stress test) and conventional life tests (e.g., failure terminated test); the reliability estimate is based on the number of failures. Degradation tests are performed on products with performance characteristics that degrade over time, which leads to failure. Reliability can be estimated by measuring the performance characteristics at different times during testing.

Yang [3] developed a degradation model that models a product's performance characteristics with the Weibull distribution and estimates the parameters of the model using Maximum Likelihood Estimation (MLE). He also estimated the test cost as a function of test time, number of samples, and cost of measurement. He estimated the number of samples and the test time by minimizing the cost function.

Gerokostopoulos et al. [4] overviewed the methods for determining the sample size for a reliability test. These methods were divided into (1) methods that use the theory of confidence interval (known as estimation approach) and (2) methods that control the type I and type II errors (known as risk control approach). For a life test, the sample size can be determined either analytically or using simulation given the time-to-failure (TTF) distribution.

Designing an accelerated life test (ALT) requires determining the total number of samples, the appropriate stress levels, and the number of samples tested at each stress level. Due to the complexity of simulation approaches, the analytical methods can be used for determining the number of samples for ALT. [4, 5].

In the risk control approach, the sample size is determined by controlling type I risk (the probability that the product meets the reliability requirement but does not pass the demonstration test), type II risk (the probability that the product does not meet the reliability requirement but passes the demonstration test), or both [4].

Some studies used immature test specifications that were arbitrarily selected without considering the actual use conditions. Yang et al. [6] designed an accelerated degradation test for predicting the reliability of smart electricity meters (SEMs) using a mix of analytical methods and arbitrary choices for stress levels. Chang et al. [7] converted field use data of automotive headlamps to a laboratory bench test specification using the theory of fatigue damage equivalence between time domain and frequency domain data but arbitrarily selected the number of samples.

In previous work, we discussed designing a user survey and introduced an analytical procedure for assessing reliability using user survey and reliability test data [8, 9]. That approach utilized a stress-life model to convert the applicable stress profiles of the surveyed and tested devices into the mean usage time (or cycles). The mean usage times (or cycles) were then converted into the mean equivalent times (or cycles) under a reference stress profile. Assuming the mean equivalent cycles follow a parametric life distribution model, the parameters of the model were estimated in a Bayesian analysis.

This paper describes a procedure for designing test specifications for a new product with unknown failure modes. The proposed test design method is an extension of the bogey test [3], but this approach can estimate the reliability and its confidence intervals even if failures occur. The approach determines the sample size, stress levels, and their frequencies for reliability testing based on the target reliability, confidence level, manufacturer's constraints including test duration and maximum sample size, and the actual use conditions of similar products collected by user survey. The similarity between the reliability of the new product and the similar products depends on the extent to which the new product maintains a similar structure, material(s), and components [10].

The proposed test design approach uses some of the rules explained in our previous works like scoring the stress and stress adjustors and combining them through an additive or multiplicative stress-index (S.I.) model [8, 9]. Although our previous work used data about known failure modes, the approach proposed here does not consider a specific failure mode for the product. But the stress, stress adjustors, and frequencies of drops are similar to the previous studies.

Two test specifications are designed in this study: (1) a frequency accelerated test, and (2) a stress-accelerated test. For both tests, the number of samples is determined by assuming a binomial distribution; and the use stress profiles of similar products are grouped through a multiplicative S.I. model as proposed in [8, 9]. The multiplicative S.I. model multiplies the quantitative stress and stress adjustors of a stress profile together and delivers an S.I. value for each applicable stress profile. The S.I. values and the use frequencies of similar products are grouped using clustering methods. For the frequency-accelerated test, the grouped use frequencies are converted to the test frequencies and are applied to the samples during the test. In the stress-accelerated test, the grouped use conditions are replaced by some accelerated stress profiles which their frequencies are determined using a known stress-life model. The accelerated stress profiles are applied to the samples during the test.

The rest of this paper proceeds as follows. Section 2 explains the process of summarizing the use conditions (S.I. values and their frequencies) using three clustering methods. Section 3 describes the method for creating a table of all possible stress profiles and their S.I. values. Section 4 presents the method for designing the test specification for a frequency-accelerated and a stress-accelerated test. Section 5 illustrates the application of the approach using a simulated dataset for an electronic device that is accidentally dropped by users. Section 6 concludes the paper.

2. Summarize Use Stress Profiles

Our proposed approach utilizes the use conditions (i.e., the way owners are expected to use the product that leads to possible damaging stresses) of similar devices collected by user surveys to determine the test specification. As users may have many different use conditions, summarizing them into a small number of use condition groups will simplify the reliability test plan. We considered three clustering methods for this step: (1) K-means clustering, (2) Gaussian mixture model (GMM), and (3) SI-cycle graph to group and summarize the use conditions.

The input data is a set of data points, where each data point has two values: (1) an S.I. (stress-index) value for one stress profile and (2) the frequency of occurrence (how frequently a user's device experienced that stress profile). The S.I. value is obtained through an additive or a multiplicative S.I. model which combines all scored stresses and stress adjustors of a stress profile [9]. The clustering approach yields a set of clusters, and the centroids of the clusters are taken as the grouped use conditions. (That is, there is one grouped use condition for each cluster.)

2.1. K-Means Clustering

K-means is a clustering method that allows finding groups of similar use conditions. K-means is computationally very efficient compared to the other clustering algorithm, but it does not have any mechanism to handle the uncertainties [11, 12]. The K-means algorithm performs clustering as follows. It first specifies K centroids and initializes their coordinates randomly. Then, it calculates the distance between the data points and the centroids to assigns the data points to their nearest centroids. Finally, it updates the coordinates of each centroid to the mean of the data points in the centroid's cluster. The elbow graph which plots the distortion (i.e., the average of the squared distances from the cluster centers) or inertia (i.e., the sum of squared distances of samples to their closest cluster center) versus the possible number of clusters is then used to assign K [11]. The centroids (geometric means) of the K clusters are known as the grouped use conditions.

2.2. Gaussian Mixture Model

The GMM is a clustering technique that uses a probabilistic assignment of data points to clusters and unlike the K-means algorithm considers uncertainties in clustering. The GMM algorithm performs clustering as follows. It, first, specifies K multivariate Gaussian models (clusters) and initializes their means and variances randomly. Then, it calculates the probability density function (PDF) of each data point using the existing Gaussian models and assigns the data point to the cluster with the highest PDF value. Finally, it updates the mean and variance of each cluster to the mean and variance of all data points assigned to that cluster. The trend of Akaike information criterion (AIC) or Bayesian information criterion (BIC) over the number of clusters are then used to determine the number of groups, K, representing the number of multivariate models in the GMM. The optimum K is on the elbow of the graph. The centroids of the K clusters are known as the grouped use conditions.

2.3. SI-Cycle Graph

The SI-cycle graph is a two-dimensional graph that shows the S.I. values on one axis and the frequencies of the S.I. values (i.e., frequencies of the stress profiles) on the other axis. The area on the plot is divided into N equal elements. Each element contains some data points. The number and shape of the elements are updated based on the optimized number of clusters (K) obtained by the K-means and GMM algorithm. The N-K elements with the least number of data points on the graph (scarcely occupied elements) are combined with their nearest neighbors. The nearest neighbor is defined as the element that has the closest boundary to the data points of the scarcely occupied element. This combination reduces the number of elements into K. The centroids of the K clusters are known as the grouped use conditions.

3. Table of Possible Stress Profiles and S.I. Values

Each cluster's centroid is associated with an S.I. value and the frequency of its occurrence. We call these S.I. values and frequencies the "group S.I. values" and "grouped frequencies." The next step in our approach creates a table of all possible stress profiles and their corresponding S.I. values. Then, each group S.I. value is compared with the entries in the table, and the stress profile associated with the next higher S.I. value in the table is known as the "grouped stress profile". Thus, this step "translates" each group S.I. value to an appropriate stress profile that can be used to specify conditions for the reliability test.

4. Design Test Specification

This section designs test specifications for a new product with unknown failure modes. A frequency-accelerated and a stress-accelerated reliability test are proposed. Details about the tests and approaches for assigning their specification are discussed in Section 4.1 and Section 4.2

4.1. Design Test Specification for a Frequency-Accelerated Reliability Test

A frequency-accelerated reliability test applies the group usage stress profiles into the device but accelerates the group use frequencies. The test specification is determined based on the manufacturer's reliability requirements including the desired warranty time, reliability level, confidence level, maximum number of units for the test, and test duration. The use conditions of similar devices are also needed to determine the test stress profiles and their frequencies of occurrence.

The procedure of designing test specifications for a frequency-accelerated test is shown in Figure 1. The various use conditions are grouped into *K* conditions using the clustering methods discussed in Section 2 and their associated stress profiles are estimated using the table of possible stress profiles and S.I. values introduced in Section 3. If the manufacturer desires to run the test under a smaller number of stress profiles than *K*, it may replace some of the stress profiles with the harsher profiles in the list of *K* profiles. This results in a rigorous reliability estimate. When the grouped stress profiles are determined, their corresponding frequencies of occurrence are multiplied by the ratio between the usage time window that the use frequency was calculated from it (e.g., 1 year) and the test duration (e.g., 1 week) to determine the frequencies of the stress profiles during the test (i.e., test frequencies). This is mathematically shown in Eq. (1).

$$\text{Test frequency}\left(\frac{\text{time(or cycle)}}{\text{test duration}}\right) = \text{usage frequency} \times \left(\frac{\text{time (or cycle)}}{\text{usage time window}}\right) \times \frac{\text{usage time window}}{\text{test duration}} \quad \text{Eq. (1)}$$

The number of test samples is determined using Eq. (2), where *m* is the number of samples, R_l is the lower-bound reliability determined by the manufacturer, *1-α* is the confidence level, and *l* is the desired maximum number of failures when the test is complete [13].

$$1\text{-}\alpha=\sum_{i=0}^{l} \frac{m!}{i!(m-i)!}(1-R_l)^i \, R_l^{(m-i)} \quad \text{Eq. (2)}$$

The reliability test is performed on *m* samples under the *K* (or smaller) stress profiles with the test frequencies. The test outcome is the number of failed (*f*) and right-censored (*r*) samples. If the number of failures is greater than *l*, the actual reliability is less than the target reliability of the manufacturer. The point estimate of reliability is obtained from Eq. (3) where $\hat{R}$ is the point estimate of reliability, *f* is the number of failed samples when the test is complete, and *m* is the number of tested samples. The lower-bound reliability is calculated using the regularized incomplete Beta function, as shown in Eq. (4), where R_l is the lower-bound reliability, I_R is the regularized incomplete Beta function, *m* is the total number of tested samples, *f* is the number of failed samples, and $1-\alpha$ is the confidence level.

$$\hat{R} = 1 - \frac{f}{m} \quad \text{Eq. (3)}$$

$$R_l = I_R(m-f,\ f+1) \leq \alpha \quad \text{Eq. (4)}$$

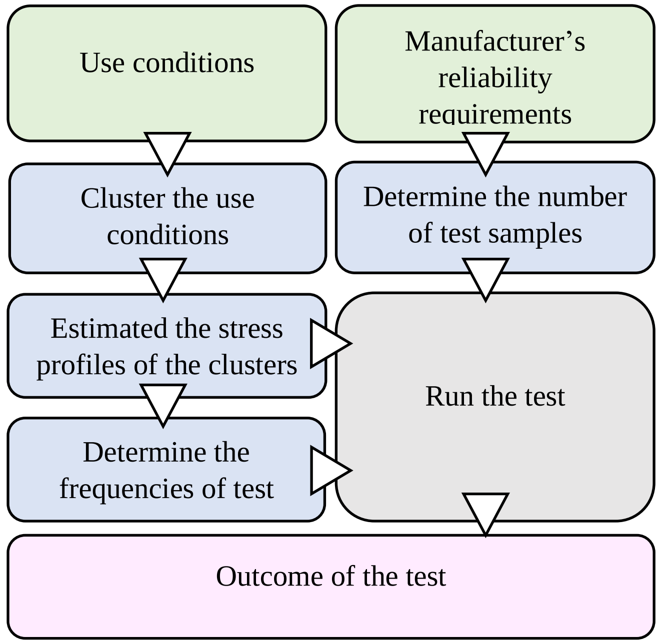

Figure 1 procedure of designing test specification for a frequency-accelerated test.

4.2. Design Test Specification for a Stress-Accelerated Reliability Test

The stress-accelerated reliability test is performed at stress levels higher than the use stress levels. This test requires an additional input compared to the frequency-accelerated test which is the underlying stress-life model of the class of products. The manufacturer selects some harsher stress profiles (i.e., accelerated stress profiles) than the grouped stress profiles to run the test. The stress-life model is then used to convert the frequency of the grouped stress profiles into the frequency of the accelerated stress profiles.

The procedure of designing test specifications for a stress-accelerated test is shown in Figure 2. Similar to the frequency-accelerated test, the use conditions are grouped using the clustering methods discussed in Section 2. The centroids of the clusters show the group S.I. values and the grouped frequencies. The grouped stress profiles are determined using the table of possible stress profiles and S.I. values introduced in Section 3. The manufacturer then selects several accelerated stress profiles which are harsher than the grouped stress profiles. The grouped frequencies are converted into the equivalent frequencies (frequencies of the accelerated stress profiles) using the underlying stress-life model, as shown in Eq. (5), where P is the cumulative density function (CDF), t_s is the grouped frequency, SI_s is the S.I. value of the grouped stress profile, ν_a is the accelerated frequency, and SI_a is the S.I. value of the accelerated stress profile. The accelerated frequencies are then converted into the test frequencies using Eq. (6).

$$P(t_s, SI_s) = P(\nu_a, SI_a) \quad \text{Eq. (5)}$$

$$\text{Test frequency}\left(\frac{\text{time(or cycle)}}{\text{test duration}}\right) = \text{accelerated frequency} \times \left(\frac{\text{time (or cycle)}}{\text{accelerated time window}}\right) \times \frac{\text{accelerated time window}}{\text{test duration}} \quad \text{Eq. (6)}$$

The number of samples for the stress-accelerated test (*m*) is determined using Eq. (2). Then, the stress-accelerated test is performed on *m* samples under the accelerated stress profiles with their associated test frequencies. The outcome of the test is the number of failed (*f*) and right-censored (*r*) units. The point estimate reliability and the lower-bound reliability are estimated using Eq. (3) and Eq. (4).

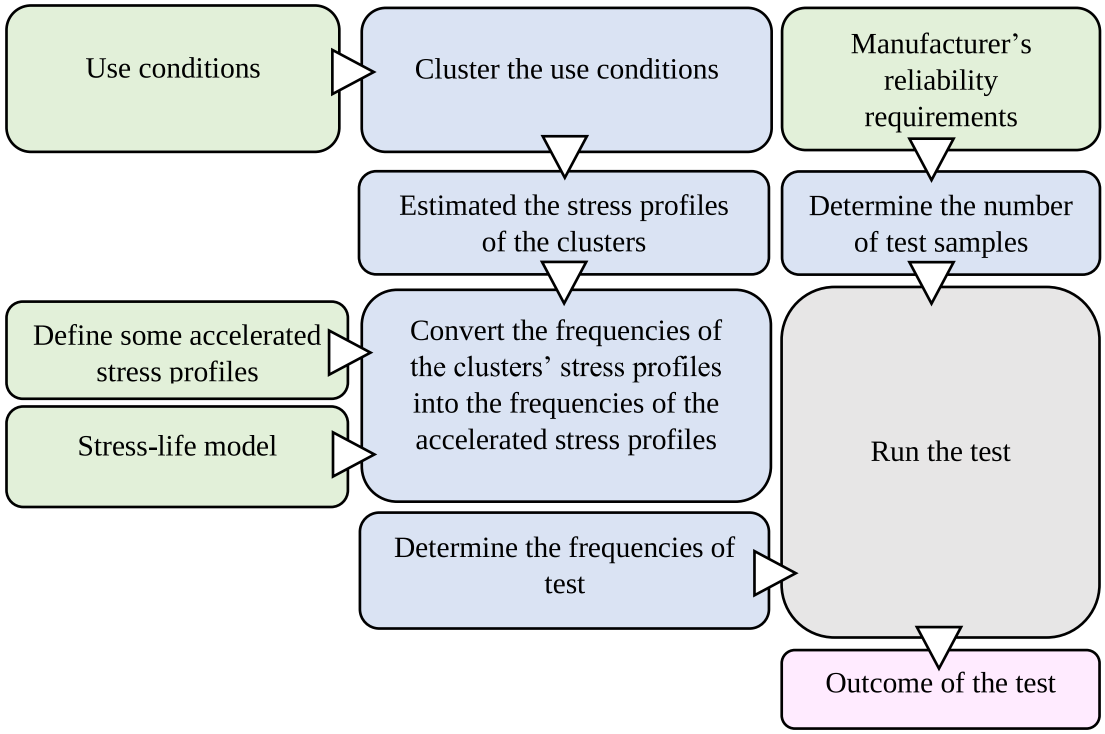

Figure 2 procedure of designing test specification for a stress-accelerated test.

5. Case Study

This section illustrates the application of the proposed approach using a simulated user survey dataset for an electronic device that is accidentally dropped by users. For this product, the applied stress is a drop during use. Three stress adjustors that can describe the relevant quantitative stress levels are (a) the drop height, (b) the type of surface on which the device dropped, and (c) the type of user's activity when the drop occurred. Weather conditions are not considered as a significant stress adjustor for this product [9]. The process of simulating the user survey dataset is explained in detail in our previous study [9]. The dataset contains 1000 users (24% young, 63% middle-aged, and 13% senior users). The fraction of age groups and the height of users in the groups are consistent with the US population data [14, 15, 16, 17]. The users dropped their devices under various use conditions (i.e., stress profiles) for many different times during their ownership times. The ownership times were randomly drawn from a discrete distribution that contained 400 ownership times of 1 year, 300 ownership times of 2 years, 200 ownership times of 3 years, and 100 ownership times of 4 years.

The qualitative drop heights in the user survey were knee height, waist height, chest height, and head or higher height; the qualitative surface types were soft, semi-soft, semi-hard, and hard surface; and the qualitative activities were benign and harsh activity. The stress adjustors were scored between 0 and 100 using the method explained in [9]. The quantitative stress adjustors were then combined through a multiplicative S.I. model as shown in Eq. (7) to estimate an S.I. value for each stress profile.

$$SI = \prod_{i=1}^{3} s_i \qquad \text{Eq. (7)}$$

5.1. Summarize Use Conditions

The use conditions were grouped using K-means clustering, GMM, and SI-cycle graph. For K-means clustering, the elbow graph, as shown in Figure 3 (a), along with the Kneedle algorithm [18] was used to determine the best (minimum) number of clusters (the number of clusters should not be less than the minimum but a manufacturer may select more). The elbow graph shows the trend of distortion versus the number of clusters. The best number of clusters is at the knee point of the elbow graph and the Kneedle algorithm estimates the location of the knee point. Using this method, the best number of clusters was estimated as 4. Then, using the K-means algorithm the use conditions were divided into 4 clusters, as shown in Figure 3 (b). The centroids of the clusters (i.e., the geometric mean of the S.I. values and frequencies) represent the grouped use conditions.

For GMM, the trend of BIC (or AIC) versus the number of clusters, as shown in Figure 4 (a), along with the Kneedle algorithm was used to determine the best number of clusters. This analysis resulted in 4 clusters. The GMM was then used to divide the data into 4 clusters, as shown in Figure 4 (b). Each cluster in Figure 4 (b) has three shaded parts which show the six-sigma region of the cluster's Gaussian mixture distribution.

For the SI-cycle graph, first, the data were arbitrarily divided into 9 identical regions, as shown in Figure 5 (a). This resulted in four scarcely occupied regions (i.e., regions 2, 3, 6, and 9 in Figure 5 (a)). These regions were combined with their nearest neighbors and the number of regions reduced from 9 to 4, as shown in Figure 5 (b). The final number of clusters is consistent with the number of clusters for the K-means and GMM algorithm.

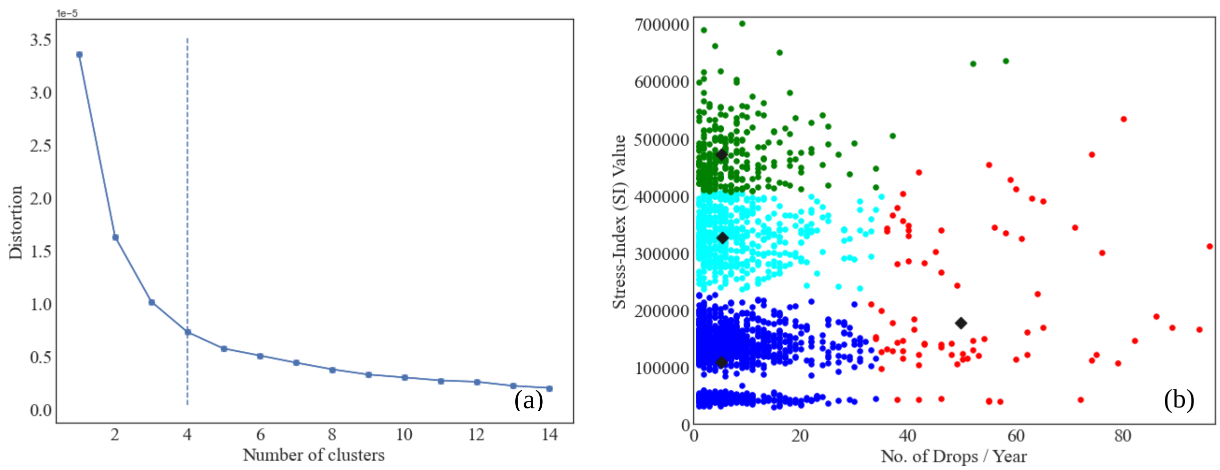

Figure 3 (a) the elbow graph, (b) the result of clustering using the K-means algorithm.

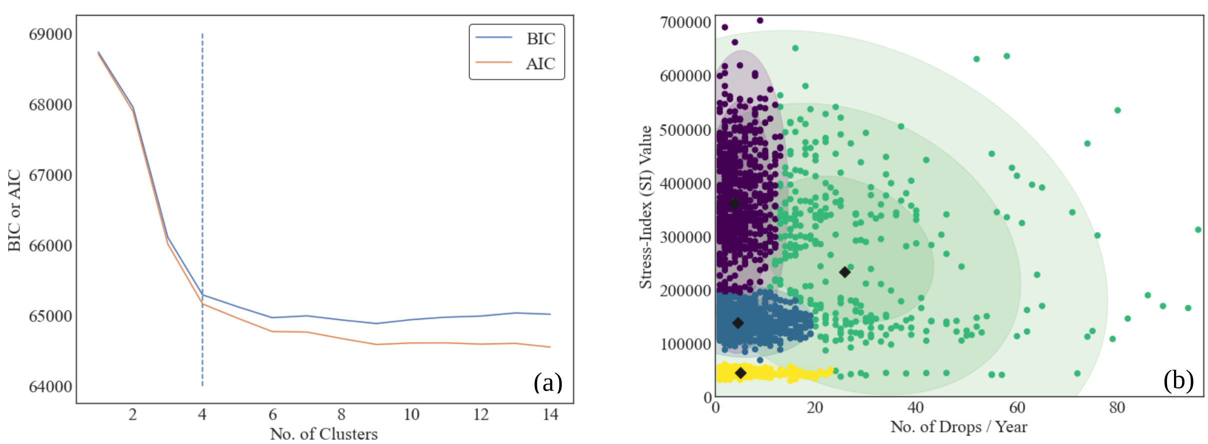

Figure 4 (a) trend of BIC and AIC vs. the number of clusters, (b) the result of clustering using GMM algorithm.

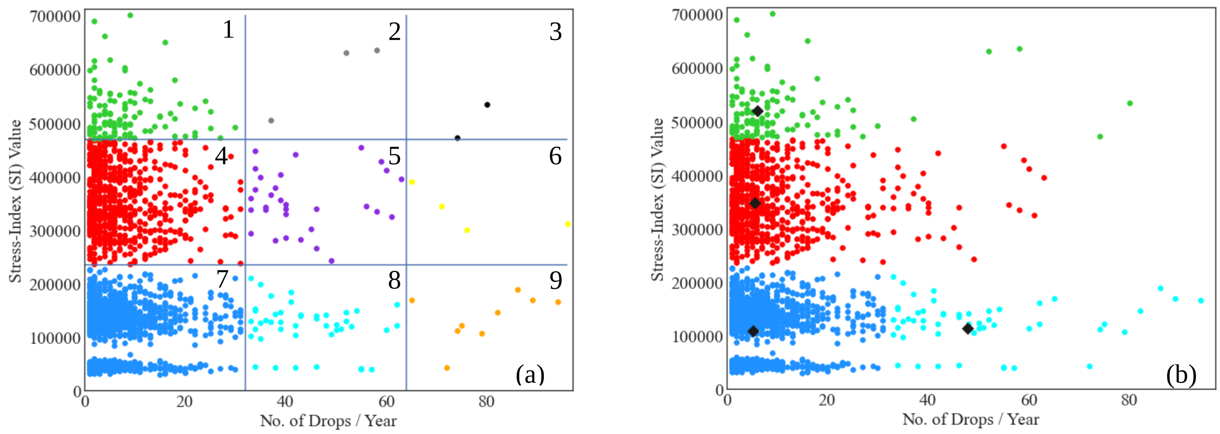

Figure 5 (a) SI-cycle graph with arbitrary divisions, (b) SI-cycle graph with 4 clusters.

Figure 6 compares the grouped use conditions (the centroids of the clusters) obtained by the three methods. The centroids were put into 4 groups which are shown by black ellipses in the figure. In 3 out of the 4 groups (i.e., groups 1, 2, and 3), the centroids estimated by the GMM had the lowest S.I. values. Besides, in all four groups, the GMM resulted in the smallest number of drops per year. Therefore, the grouped use conditions obtained by GMM are optimistic. In 3 out of the 4 groups (i.e., groups 1, 2, and 3), the centroids estimated by the SI-cycle graph had the highest S.I. values and number of drops per year. Thus, the SI-cycle graph results in pessimistic grouped use conditions. The grouped use conditions obtained by K-means are moderate because the S.I. value or/and the number of drops estimated by K-means clustering are usually between the values estimated by the other two methods.

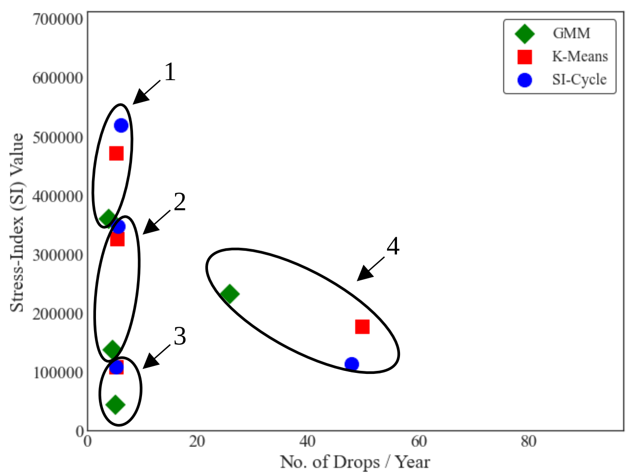

Figure 6 The grouped use conditions estimated by K-means clustering, GMM, and SI-cycle graph.

The pessimistic grouped use conditions obtained by the SI-cycle graph were used to infer the group use stress profiles because they resulted in the most rigorous reliability estimate. To infer the grouped stress profiles, we built the table of all possible stress profiles and their S.I. values, as shown in Table 1. Because there were four height choices (i.e., knee, waist, chest, and head or higher), four surface choices (i.e., soft, semi-soft, semi-hard, and hard), and two activity choices (i.e., benign, and harsh) in the user survey, in total there were 32 different possible combinations of them ($\binom{4}{1} \times \binom{4}{1} \times \binom{2}{1} = 32$). Each combination is a possible stress profile that may be observed by a user in the field. The list of all 32 combinations is shown in Table 1. To calculate the S.I. values in the table, the scores for knee, waist, chest, and head (or higher) height were assumed as 25, 50, 75, 100, for soft, semi-soft, semi-hard, and hard surfaces were assumed as 25, 50, 75, 100, and for benign and harsh activity were assumed as 50 and 100. The height scores are consistent with the scores used for the middle-aged group in the user survey dataset. As the

Table 1 Table of all possible stress profiles and their SI values.

Stress Profile	S.I. Value	Stress Profile	S.I. Value
Knee- soft- benign	31,250	Waist- hard- benign	250,000
Knee- semisoft- benign	62,500	Waist- semisoft- harsh	250,000
Knee- soft- harsh	62,500	Head- semisoft- benign	250,000
Waist- soft- benign	62,500	Head- soft- harsh	250,000
Knee- semihard- benign	93,750	Chest- semihard- benign	281,250
Chest- soft- benign	93,750	Chest- hard- benign	375,000
Knee- hard- benign	125,000	Chest- semisoft- harsh	375,000
Knee- semisoft- harsh	125,000	Head- semihard- benign	375,000
Waist- semisoft- benign	125,000	Waist- semihard- harsh	375,500
Waist- soft- harsh	125,000	Head- hard- benign	500,000
Head- soft- benign	125,000	Head- semisoft- harsh	500,000
Knee- semihard- harsh	187,500	Waist- hard- harsh	510,000
Waist- semihard- benign	187,500	Chest- semihard- harsh	562,500
Chest- semisoft- benign	187,500	Chest- hard- harsh	750,000
Chest- soft- harsh	187,500	Head (or higher)- semihard- harsh	750,000
Knee- hard- harsh	250,000	Head (or higher)- hard- harsh	1,000,000

middle-aged group has the highest scores in the dataset, these scores result in a rigorous reliability estimate. The scores associated with the stress adjustors of the 32 stress profiles were combined through a multiplicative S.I model, Eq. (7), to find the S.I. values in Table 1. In Table 1, there are some stress profiles with similar S.I. values. It is assumed that the device is equally damaged under the stress profiles with the same S.I. value.

We compared each group S.I. value with all S.I. values in Table 1 and selected the stress profile with the next higher S.I. value as the grouped stress profile. For instance, in Figure 7, the next higher S.I. value to the S.I. value of centroid 1 was 562,500 which belonged to a drop from chest height on a semihard surface during a harsh activity. The other grouped stress profiles were estimated using the same scenario and were listed in Table 2. In the cases where the next higher S.I. value belongs to several stress profiles, each of them can be selected as the grouped stress profile because it is assumed that the device is equally damaged under all those stress profiles. The grouped stress profiles along with the grouped frequencies are used in Section 5.2 and Section 5.3 to design the test specification for a frequency-accelerated and a stress-accelerated test.

Table 2 The grouped stress profiles.

Centroid No. in Figure 7	Grouped stress profile	Grouped Frequency of Use
1	Chest- semihard- harsh	6 drops in 1 year
2	Chest- hard- benign Chest- semisoft- harsh Head- semihard- benign Waist- semihard- harsh	6 drops in 1 year
3	Knee- hard- benign Knee- semisoft- harsh Waist- semisoft- benign Waist- soft- harsh Head- soft- benign	6 drops in 1 year
4	Knee- hard- benign Knee- semisoft- harsh Waist- semisoft- benign Waist- soft- harsh Head- soft- benign	49 drops in 1 year

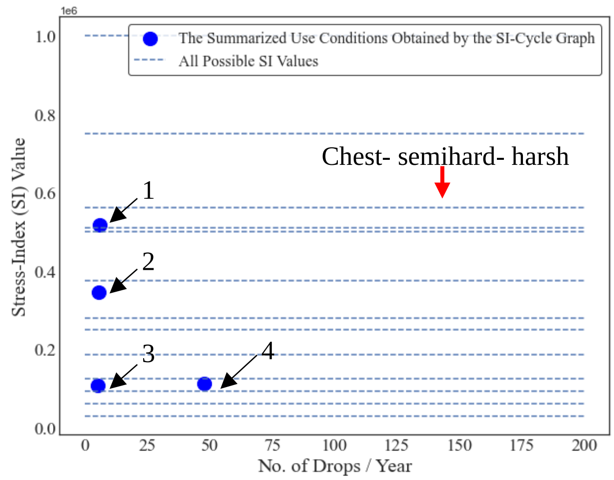

Figure 7 The grouped use conditions and all possible S.I. values.

5.2. Design Test Specification for a Frequency-Accelerated Reliability Test

A frequency-accelerated test requires three elements which are (1) the group use stress profiles, (2) the test frequencies, and (3) the number of test samples (or the allowed number of failures if the maximum number of samples has been decided by the manufacturer). The first element was determined in Section 5.1 and listed in Table 2. The test frequencies were calculated using Eq. (1) and by assuming the test duration of 1 week. These frequencies are 6, 6, 6, and 49 drops in one week for the grouped stress profile 1 to 4 listed in Table 2, respectively.

It was assumed that the manufacturer wanted to achieve at least 95% reliability with 90% confidence after 1 year of warranty, and the maximum number of test samples was 100. By substituting these values in Eq. (2), the allowed number of failures was estimated as 2 samples. If the number of failures after completing the frequency-accelerated test is less than 2, the product meets the target reliability. Otherwise, a root cause analysis is needed to understand and resolve the reason for out-of-specification. For instance, if the number of failures is 3, the point estimate reliability is 97% and the lower-bound reliability is 93.44% which is less than the minimum desired reliability of the manufacturer. Therefore, a root cause analysis should be conducted to understand the reason for out-of-specification and appropriate actions should be performed to improve the product's reliability.

5.3. Design Test Specification for a Stress-Accelerated Reliability Test

A stress-accelerated test requires three elements which are (1) the accelerated stress profiles, (2) the equivalent frequencies (i.e., the frequencies of the accelerated stress profiles), and (3) the number of test samples (or the allowed number of failures if the maximum number of samples has been decided by the manufacturer). The accelerated stress profiles are harsher than the grouped stress profiles and are decided by the manufacturer. For instance, the accelerated stress profiles for this case study can be the profiles shown in Table 3. These accelerated stress profiles have at least one harsher stress adjustor than their corresponding grouped stress profiles listed in Table 2.

The next step is to calculate the equivalent frequencies for the accelerated stress profiles such that the accelerated test conditions cause the same amount of damage as the grouped use conditions. To calculate the equivalent frequencies, the underlying stress-life model of the class of products is needed. An inverse power law (IPL) stress-life model with known parameters was assumed. This model is shown in Eq. (8), where N and SI represent the frequency of drops and the S.I. value, respectively. The Eq. (9) was obtained from the ratio between the frequency of an accelerated stress profile and the frequency of a grouped stress profile where ν and SI_a are the equivalent frequency and the S.I. value of the accelerated stress profile, and d and SI_a are the frequency and S.I. value of the grouped stress profile. The SI_a and SI_s were calculated using Eq. (7). These frequencies are smaller than the grouped frequencies and thus reduce the testing time. The test frequencies were estimated using Eq. (6) and are listed in Table 3. The allowed number of failures, the point estimate reliability, and the lower-bound reliability for the stress- accelerated test are estimated using the same equations and the same scenario explained for the frequency accelerated test.

$$N = 300. SI^{-1.2} \quad \text{Eq. (8)}$$

$$\nu = d. (\frac{SI_s}{SI_a})^{1.2} \quad \text{Eq. (9)}$$

Table 3 The test conditions of the stress-accelerated test.

Centroid No. in Figure 7 and Table 2	Accelerated Stress Profile	Test Frequency
1	Head- hard- harsh	3 drops in 1 week
2	Head- hard- benign	4 drops in 1 week
3	Chest- hard- harsh	1 drop in 1 week
4	Chest- hard- benign	13 drops in 1 week

6. Conclusions

This study showed that the reliability test specification of a new product with unknown failure modes can be designed based on the usage conditions of similar products collected using a reliability-informed user survey. A test specification that is based on the user data allows revealing the failure modes observed in the field. The user survey is a cost-effective and quick way to collect the use conditions. A frequency-accelerated and a stress-accelerated test were proposed which determined the test specification including the test stress profiles, their frequencies, and the allowed number of failures based on the manufacturer's reliability requirements including the warranty time, desired reliability, its confidence level, test duration, and the maximum number of test samples. The various use conditions collected by the user survey data were grouped through the K-means clustering, GMM, and SI-cycle graph and were used to determine the test frequencies and the test stress profiles. The stress-accelerated test required an additional input compared to the frequency-accelerated test but delivered a shorter testing time. The process of designing test specifications was illustrated using a simulated dataset for an electronic device that was accidentally dropped by users in the field. Our case study showed that the SI-cycle graph resulted in the most pessimistic grouped use conditions and thus delivered the most rigorous test specification and reliability estimate.

Acknowledgment

This research was funded under the CRR-Amazon company's Lab 126 collaboration in advancing reliability engineering research and education. Amazon company's funding is highly appreciated.

References

[1] X. Dong, Y. Tsong and M. Shen, "Statistical Considerations in Setting Product Specifications," *Journal of Biopharmaceutical Statistics*, vol. 25, pp. 280-294, 2015.

[2] Y.-L. Lee, S. Makam, S. McKelvey and M.-W. Lu, "Methods, Durability Reliability Demonstration Test," *Procedia Engineering*, vol. 133, pp. 31-59, 2015.

[3] G. Yang, "Heuristic Degradation Test Plans for Reliability Demonstration," *IEEE Transactions on Reliability*, vol. 62, no. 1, pp. 305-311, 2013.

[4] A. H. G. a. E. P. Gerokostopoulos, "Determining the Right Sample Size for Your Test: Theory and Application," in *Annual Reliability and Maintainability Symposium Tutorial Notes*, 2015.

[5] S. Woo, "Sample Size Equation of Mechanical System for Parametric Accelerated Life Testing," in *Soft Computing Methods for System Dependability*, IGI Global, 2020, pp. 234-248.

[6] Z. Yang, Y. Chen, Y. Li, E. Zio and R. Kang, "Smart electricity meter reliability prediction based on accelerated degradation testing and modeling," *Electrical Power and Energy Systems*, vol. 56, pp. 209-219, 2014.

[7] W.-L. Chang, K.-Y. Lin, C.-D. Hsueh and J.-M. Chang, "Vibration Test Specification Design and Reliability Analysis," *International Journal of Materials and Manufacturing,* vol. 4, no. 1, pp. 675-685, 2011.

[8] N. Shafiei, J. W. Herrmann, A. Krive, G. Sethi and M. Modarres, "Designing Reliability-Informed Customer Surveys," in *European Safety and Reliability Conference (ESREL)*, Angres, 2021.

[9] N. Shafiei, J. W. Herrmann, A. Krive and M. Modarres, "Estimating the Reliability of Consumer Electronics from User Survey Data and Test Data," in *Reliability and Maintainability Symposium (RAMS)*, Tuscon, USA, 2022.

[10] S. Woo, Reliability Design of Mechanical Systems: A Guide for Mechanical and Civil Engineers, Singapore: Springer , 2020.

[11] S. Raschca, Python Machine Learning, Packt Publishing Itd, 2015.

[12] E. Patel and D. S. Kushwaha, "Clustering cloud workloads: k-means vs gaussian mixture model," in *Procedia Computer Science*, 2020.

[13] M. Modarres, M. Kaminskiy and V. Krivtsov, Reliability engineering and risk analysis : a practical guide, CRC press, 2016.

[14] C. D. Fryar, D. Kruszon-Moran, Q. Gu and C. L. Ogden, "Mean body weight, weight, waist circumference, and body mass index among adults: United States 1999–2000 through 2015–2016.," National Health Statistics Reports, Hyattsville, MD, USA, 2018.

[15] S. C. Savva, Y. Kourides, M. Tornaritis, M. Epiphaniou-Savva, P. Tafouna and A. Kafatos, "Reference growth curves for Cypriot children 6 to 17 years of age," *Obesity research,* vol. 9, no. 12, pp. 745-762, 2001.

[16] C. C. Gordon, T. Churchill, C. E. Clauser, B. Bradtmiller, J. T. McConville and I. Tebbets, "Anthropometric Survey of US Personnel: Summary Statistics Interim Report," US Research Development and Engineering Center, 1989.

[17] L. M. Howden and J. A. Meyer, "Age and sex composition, 2010," US Department of Commerce, Economics and Statistics Administration, US Census Bureau, 2011.

[18] V. Satopa, J. Albrecht, D. Irwin and B. Raghavan, "Finding a “Kneedle” in a Haystack: Detecting Knee Points in System Behavior," in *International Conference on Distributed Computing Systems Workshops (ICDCSW)*, 2011.

[19] I. H. T. Guideline, "Specifications: test procedures and acceptance criteria for new drug substances and new drug products: chemical substances," in *International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use*, Geneva, Switzerland, 1999.

[20] D. N. P. Murthy, T. Østerås and M. Rausand, "Component Reliability Specification," *Reliability Engineering & System Safety,* vol. 94, no. 10, pp. 1609-1617, 2009.

[21] Q. Guan and Y.-c. T. , "Bayesian Planning of Optimal Step-stress Accelerated Life Test for Log-location-scale Distributions," *Acta Mathematicae Applicatae Sinica,* vol. 34, no. 1, pp. 51-64, 2018.

[22] N. Shafiei, J. W. Herrmann, A. Krive and M. Modarres.