Case Study

Quality and Reliability Engineering International

(www.interscience.wiley.com) DOI: 10.1002/qre.1091

Published online in Wiley InterScience

Variance-Corrected Proportional Hazard Models for the Analysis of Multiple Failures in Personal Computers

A. A. Sunethra and M. R. Sooriyarachchi*†‡

Proportional hazard (PH) modeling is widely used in several areas of study to estimate the effect of covariates on event timing. In this paper, this model is explored for the analysis of multiple occurrences of hardware failures in personal computers. Multiple failure events consist of correlated data, and thus the assumption of independence among failure times is violated. This study critically describes a class of models known as *variance-corrected PH models* for modeling multiple failure time data, without assuming independence among failure times. The objective of this study is to determine the effect of the computer brand on event timings of hardware failures and to test whether this effect varies over multiple failure occurrences. This study revealed that the computer brand affects hardware failure event timing and that further, this effect of the brand does not change over the multiple failure occurrences.

Keywords: multiple failures; variance-corrected proportional hazard models; information matrix test; modified variance sandwich estimator; risk interval; risk set

1. Introduction

Computers play a prominent role in modern society. Among several types of computers, personal computers (PCs) have appeared to be the most popular over the decades.[1] In line with the rapid growth of PC usage, the industry faces the challenge of maintaining its reliability factor. PCs are repairable and hence may have multiple failures within its usage. Basically, PCs experience hardware failures and software failures. This study only focuses on modeling hardware failures in PCs. Although there are several types of hardware failures that could occur, for simplicity all of these are considered as a single failure type. Thus, in this study 'multiple events' correspond to repeated failures of the same type.

The primary objective of this study is to explore an empirical application of analyzing multiple failure time data without incorporating the assumption of independence among failure events. The application of the study includes failure occurrences in PCs and the effect of the computer brand on the failure event time is evaluated. Thereby, a clear understanding of the brand effect on failure event timing can be achieved.

The methods of analyzing failure type data can be basically classified into parametric and non-parametric methods. The correlated structure of multiple failure time data makes it complicated to use parametric models to model multiple failure time data. Therefore, non-parametric methods in which no distributional assumption is required would be better in this context. The Cox proportional hazard (PH) model developed by Cox[2] is a popular semi-parametric method that has been excessively used to model failure time data and it mainly focuses on evaluating the effect of covariates on event timing. According to Lim *et al.*[3], the conventional analysis used by several authors in the past, to study multiple failure times, consisted of using the usual Cox model; for the time to the first event or for the overall survival time. However, analytic approaches adjusting for the correlations between recurrent failure times within a system are needed to obtain efficient inferences and this requires the modification of the usual Cox model.

The variance-corrected proportional hazards (VCPH) model proposed by the authors is an extension of the PH model, which takes into account the lack of independence between failure times. The VCPH is semi-parametric in the sense that no assumption is made about the distribution of failure times and the only assumption made is that of proportionality of hazards between the levels of the covariates. These models obtain parameter estimates by first fitting a Cox PH model that ignores the dependence

Department of Statistics, University of Colombo, Colombo 03, Sri Lanka
Correspondence to: M. R. Sooriyarachchi, Department of Statistics, University of Colombo, Colombo 03, Sri Lanka.
†*E-mail: roshini@mail.cmb.ac.lk*
‡*Professor.*

structure and then replace the naïve standard errors with estimates from an empirical sandwich variance estimator[4] in order to incorporate the non-independence of times between failures. Analogous to the generalized estimating equations of Liang and Zeger[5] and Zeger and Liang[6], these estimators are asymptotically normal and consistent in sample size *n*, provided that the models are correctly specified.[7–11] This study applies the VCPH models to model multiple occurrences of hardware failures in PCs.

The data were collected from a reputed computer solution company in Sri Lanka, which sells two brands of PCs and maintains a service unit for the PCs sold by the company. The two types of PCs are denoted by *brand 1* and *brand 2* as original brand names cannot be divulged owing to reasons of confidentiality. The date on which each PC was reported for repair is taken as the failure date and the number of days between the sales date of the respective PC and the repair date is taken as the time to failure of that PC. The data consist of repetitive dates of failures and the study is limited to a maximum of five failure events. Thereby, the effect of the brand of PC on failure occurrences is evaluated over multiple failure times.

To assess the goodness of the model fitted, Information Matrix (IM) test of White[12] is used. Finally, this study identifies the most appropriate model that depicts the multiple failure structure of PCs.

VCPH models have previously been used by Wei *et al.*[11] to model bladder cancer data and by Michael *et al.*[13] to model arrest data of a sample of California Youth Authority parolees, although there has been no application of this model in the area of reliability. This study can therefore be regarded as a novel application of VCPH models to reliability data. In addition, in many research papers dealing with multiple failure time data, no goodness-of-fit testing of the model is done. In rare instances when goodness-of-fit is tested, it consists of plots of Cox–Snell residuals, Schoenfeld residuals and Martingale residuals. However, these plots cannot be used with multiple failure times as events are dependent. In our study the goodness-of-fit of the model is tested using White's[12] IM test.

A literature review of data analysis for repairable systems is given in the next section. A detailed description of the VCPH models is presented in Section 3. The model-checking procedure is outlined in Section 4 and the example is illustrated in Section 5. Section 6 gives the conclusions and a brief discussion on the outcome of the study. Section 7 consists of an appendix which includes the SAS programs used for modeling.

2. A literature review of data analysis for repairable systems

In the majority of situations, most of the elements or components constituting a system can be repaired or replaced, so that the system can be restored for satisfactory operation. Such a system is termed as a repairable system.[14–16]. To provide a reliability context for such a system, consider the process of successive failures and repairs of the system. We measure time in terms of operating time, thus ignoring the repair times. Frequently, time between failures will neither be independent nor identically distributed. Usually, interest centers around modeling the probability structure of time between failures and the probability structure of time to failures as a function of system age.

One common failure model is the homogeneous Poisson process (HPP) discussed by Cinlar[17] and Tan *et al.*[18]. Thompson[19] shows that a process is HPP if and only if the times between failure are independent with common exponential distribution. Thompson[19] goes on to discuss that renewal theory[20–22] generalizes the HPP by allowing the times between failure to have other distributions other than exponential. However, it still requires the independence assumption and for repairable systems a renewal process is not capable of modeling a system that is becoming more prone to failure with time. The non-homogeneous poisson[23] is one process that is capable of modeling repairable system aging. Although the non-homogeneous Poisson process (NHPP) model is appealing as a general wearout model, in the past[19] it has been found to have some non-intuitive features; for example the distribution of the first failure determines the entire process. However, this is not the case for most general NHPP models discussed in the recent literature. Ascher and Feingold[24] criticize NHPP models for repairable systems stating that these do not give enough detail or realism for satisfactory results. Zhao and Xie[25], Jeske and Pham[26] and Nayak *et al.*[27] have each found that for all estimators (not limited to MLEs) of parameters of all NHPP models, with the expected number of failures over infinite testing time is finite, are inconsistent. Cook and Lawless[28] explain how to incorporate covariates in the models discussed thus far. The PH model[2, 29] can include the effect of covariates in the reliability function. This model requires two assumptions, namely, the proportionality of hazard rates between the different levels of the covariates and the independence of failure times. *Peña* and Hollander[30] proposed a general class of models for repairable systems, which comprise a general synthesis of several repairable systems models such as the modulated renewal process of Cox[31], extended Cox PH model considered by Prentice *et al.*[32] and the well-known VCPH models of : (i) Anderson–Gill; (ii) Wei, Lin and Weissfeld; and (iii) Prentice, Williams and Peterson as reported in Therneau and Grambsch.[4] Peña *et al.*[33] examined the properties of the estimators of the parameters of the model proposed by Peña and Hollander[30] and the consequences of misspecifying the model and reported that misspecification of the model has unacceptable consequences. There is also no goodness-of-fit and model-validation procedure available for the model. Another drawback is the unavailability of well-known software for fitting these models. Pan and Rigdon[34] consider Bayesian models that are a compromise between NHPP models and Renewal process models. However, covariate adjustment is not considered in their models. Some other less well-known models for repairable systems include Markov models.[35] The problems with this type of model are that it cannot model the repair behavior and it is cumbersome for a large number of states. Another approach to analyzing repairable systems is time-series models.[36] Though this method has some appealing features it requires a large amount of failure data.

In a repairable system prone to multiple failures, times between failures will rarely be independent and/or identically distributed. Thus the HPP model, the Renewal process model and the PH model will generally not suffice in the case of repairable systems. The NHPP model though appealing as a general wearout model has been found to have some drawbacks. This leads us to search for other alternatives. The VCPH model introduced in Section 1 and explained in more detail in Section 3 is an appealing alternative to the NHPP model for modeling repairable systems.

3. VCPH models

The Cox PH models[2] discussed in this paper are hybrid versions of the single-event Cox model and are specified in one of the two ways:

$$h_{ik}(t) = h_0(t)\exp(\beta' Z_{ik}) \tag{1}$$

$$h_{ik}(t) = h_{0k}(t)\exp(\beta_k' Z_{ik}) \tag{2}$$

In these specifications, Z_{ik} is a p- dimensional vector of measured covariates $(j=1,2,\ldots,p)$ for the kth event, and β and β_k are vectors of regression parameters to be estimated. In Equation (1), $h_0(t)$ is a non-negative baseline hazard function that is an arbitrary function of time and common to all events (i.e. $h_{0k}(t)=h_0(t)$ for $k=1,2,\ldots,K$). In the second specification, the baseline hazard function $h_{0k}(t)$ is allowed to vary over each of the events as an arbitrary function of time. $h_{ik}(t)$ refers to the hazard function of the ith subject on the kth failure event at time t. Model (2) is known as a stratified *Cox PH model* and in the models discussed in this paper, the stratification is over k failure events. The stratification is important because it allows the baseline hazard function to vary over each of the k events.

Another two key components that systematically differentiate these models are the way the risk intervals are defined with reference to the starting point and the compilation of risk sets at each distinct failure time. Risk intervals refer to the time scales used to define when a unit is at risk of experiencing a specific event[37]. There are three possible ways of defining a risk interval and each of these describes a different substantive type of risk process. These three intervals are namely, *gap time, total time and counting process* risk intervals.

The models used in this study use the partial likelihood function; they differ in the composition of the risk sets at each observed failure time[38]. The risk set at time t is composed of all individuals currently at risk (i.e. under observation) of a failure at the observed failure time t. The composition of the risk set used in the partial likelihood function at the time of each failure is a direct result of both whether the model is 'stratified' by the event number and the type of risk interval specified[37]. As Kelly *et al.*[37] discuss, there are three possible types of risks set: *unrestricted, restricted and semi-restricted*.

The vector of parameter estimates for both the models is calculated via Cox's[2] partial likelihood function. The partial likelihood is based only on the rank ordering of the failure times (i.e it ignores the precise timing of failures) and, in essence reduces to a series of conditional logistic regression analysis of the conditional probability of observing the covariate pattern of the individuals that fail at time t, given the weighted covariate pattern of the individuals that were at risk of failing[39].

For the remainder of this paper, we denote the true failure times and censoring times of the ith subject for the kth event $(k=1,2,3,\ldots,K)$ as random variables T_{ik} and C_{ik}, respectively, and assume that, conditional on the possibly time-dependent covariate vector, $Z_{ik}(t)$, C_{ik} is independent of T_{ik} for all i and k.[11] The observed random failure time is defined as $x_{ik}=\min(T_{ik}, C_{ik})$ and the censoring indicator is defined as $\delta_{ik}=I(T_{ik}<C_{ik})$, where $I(\cdot)$ is the indicator function. The risk process denoted by $Y_i(t)$ is equal to 1 if the ith subject is under observation just prior to time t and if it is at risk of a failure event just prior to time t , and is equal to 0 if it not under observation or not at the risk of a failure just prior to time t.

Assuming that there are no tied failure times, the *partial likelihood function* for Equation (1) is,

$$PL(\beta)=\prod_{i=1}^{n}\prod_{k=1}^{K}\left[\frac{\exp(\beta' Z_{ik}(x_{ik}))}{\sum_{j=1}^{j=n}\sum_{i=1}^{i=K} Y_{ji}(x_{ik})\exp(\beta' Z_{ji}(x_{ik}))}\right]^{\delta_{ik}} \tag{3}$$

The *score function* that corresponds to Equation (3) is:

$$U(\beta)=\sum_{i=1}^{n}\sum_{k=1}^{K}\delta_{ik}\left[Z_{ik}(x_{ik})-\frac{\sum_{k=1}^{K}\sum_{j=1}^{n} Y_{jk}(x_{ik})\exp(\beta' Z_{jk}(x_{ik})Z_{jk}(x_{ik}))}{\sum_{k=1}^{K}\sum_{j=1}^{n} Y_{jk}(x_{ik})\exp(\beta' Z_{jk}(x_{ik}))}\right] \tag{4}$$

and the maximum partial likelihood estimates $\hat{\beta}$ are obtained by taking the partial derivatives of the log-partial likelihood with respect to β,

$$U(\hat{\beta})=\frac{\partial \ln PL(\hat{\beta})}{\partial \hat{\beta}} \tag{5}$$

For stratified Cox model such as Equation (2), the partial likelihood is

$$PL(\beta)=\prod_{i=1}^{n}\prod_{k=1}^{K}\left[\frac{\exp(\beta_k' Z_{ik}(x_{ik}))}{\sum_{j=1}^{j=n}\sum_{i=1}^{i=K} Y_{ji}(x_{ik})\exp(\beta_k' Z_{ji}(x_{ik}))}\right]^{\delta_{ik}} \tag{6}$$

and the corresponding score function is:

$$U(\beta)=\sum_{i=1}^{n}\sum_{k=1}^{K}\delta_{ik}\left[Z_{ik}(x_{ik})-\frac{\sum_{k=1}^{K}\sum_{j=1}^{n} Y_{ik}(x_{ik})\exp\beta_k' Z_{jk}(x_{ik})Z_{jk}(x_{ik})}{\sum_{k=1}^{K}\sum_{j=1}^{n} Y_{jk}(x_{ik})\exp(\beta_k' Z_{jk}(x_{ik}))}\right] \tag{7}$$

The overall log-partial likelihood in a stratified model is actually a sum of the log-partial likelihoods obtained within each of the distinct k strata

$$\log PL(\beta)=\sum_{k=1}^{K}\log PL_k(\beta) \tag{8}$$

The corresponding score function is also a sum of the score functions from each stratum; the derivatives of these score functions are found by summing the derivatives across each of the K stratums and maximum partial likelihood estimates are the numerical estimates that solve the score vector according to Equation (5).[40] In essence, a stratified model fits K sub models and only subjects who are both observed at time t and in the same kth strata as the subject who failed at time t contributes information to the stratum-specific likelihood function. Stratification of the Cox PH model is one way to handle the problem of dependence through the formulation of a 'conditional' partial likelihood function, whereby a subject is only included in the partial likelihood of higher-order events upon the occurrence of each of the lower-order events.[39, 41]

Lin and Wei[42] discuss the robust inference for the Cox PH model. When the usual Cox model is used to model multiple failures, which results in non-independent failure times, then the model is misspecified. Struthers and Kalbfleisch[43] show that in this case the estimates $\hat{\beta}$ converge to a well-defined constant vector β^*. They explain that $n^{1/2}(\hat{\beta}-\beta^*)$ is asymptotically normal with zero mean and covariance matrix estimated by $\hat{V}(\hat{\beta})$ the so-called sandwich estimator. This is the intuitive reason for using the sandwich method for variance correction to adjust the standard errors of the parameter estimates in the Cox model in order to incorporate the non-independence of times between failures. For this misspecified model they show that $\beta^*=0$ under the null hypothesis (H_0) that there is no linear effect of covariates on failure time. Thus, it is possible to construct a valid Wald test based on $\hat{\beta}$ for testing H_0. If the model is correctly specified $\beta^*=\beta$, but in the case of a misspecified model β^* is generally unequal to β. However, Wei *et al.*[11] and Lin and Wei[42] indicate that for large samples the difference between β^* and β is negligible (that is asymptotically $\hat{\beta}$ is unbiased for β) and thus $\hat{\beta}$ is appropriate for practical use, under a wide range of misspecified Cox models. Thus, β parameters are estimated without considering data dependence in the VCPH model.

3.1. Modified sandwich variance estimator

In a standard Cox model with independent observations, the variances of estimated parameters are calculated by inverting the observed information matrix (IM):

$$V(\beta)=I^{-1}=-\left[\frac{\partial^2 \ln PL(\hat{\beta})}{\partial\hat{\beta}\partial\hat{\beta}'}\right]^{-1} \tag{9}$$

The standard errors of the parameter estimates are the square roots of the diagonal elements of the model-based variance covariance matrix resulted from Equation (9); these standard errors are often referred to as naïve standard errors. These estimates of the variances of the parameter estimates are not consistent estimators of the true asymptotic variances when the data contain correlated observations.[11, 44, 45] In such cases, the method of calculating standard errors from Equation (9) overestimates the number of independent data points in the sample. For example, two random observations from the same subject over time are more likely to be similar in both covariate pattern and event timing than the two randomly sampled individuals from the population. Thus, much of the within-subject information is redundant; the end result is that if the event times are positively correlated, then the computed standard errors will be biased downward. This will produce inflated tests of significance, especially for fixed covariates.[46, 47]

To correct the variance for dependence among the observations, an 'empirical' estimate of the variance of the parameter estimates is obtained using a 'modified sandwich' variance estimator that allows for within-subject correlation. This estimate is based on the robust variance estimator of Lin and Wei[42], who derived their estimator for the independent failure data by applying the sandwich variance estimator of Huber[48] and White[49]. The Lin and Wei[42] robust estimator for single, independent event data is

$$V_R(\hat{\beta})=I^{-1}(U'U)I^{-1} \tag{10}$$

where U is a $n\times p$ matrix of coefficient score residuals with elements

$$u_{ij}=\int_0^t [Z_{ij}(s)-\bar{Z}_j(s)]\,d\hat{M}_i(s) \tag{11}$$

$\bar{Z}_j(s)$ is the mean of covariate $j(j=1,2,\ldots,p)$ at time s defined as

$$\bar{Z}_j(s)=\frac{\sum Y_i(s)Z_i(s)\exp(\hat{\beta}'(Z_i(s))}{\sum Y_i(s)\exp(\hat{\beta}'Z_i(s))} \tag{12}$$

and $d\hat{M}_i(s)$ is the change in the martingale residual at time s for the ith subject.

This estimator is often referred to as the sandwich estimator because the data-based 'correction' or 'adjustment' factor (i.e. the $U'U$ component) is 'sandwiched' between the two naïve variance estimates obtained from Equation (9).

As derived by Lin and Wei[42], the sandwich estimator in Equation (10) assumes that there are n independent observations that are used to calculate the elements in the U matrix. In case of multiple failure time data, there are not n independent observations, but rather there are m independent clusters $(c_1, c_2, \ldots, c_m)$, n_i correlated observations within the ith cluster. Assuming that the observations are independent across the clusters, the overall efficient score residuals are calculated by summing the efficient score residuals within each of the m independent clusters prior to being entered in the calculation of the sandwich standard errors.[44, 50] That is, U is now a collapsed $m\times p$ matrix of efficient score residuals that consist of the u_{ijk}'s of Equation (11) (here suffix k refers to the kth observation in the ith cluster) summed within each of the m clusters

$$u_{ij}=\sum_{k=1}^{n_i} u_{ijk}; \quad (i=1,\ldots,m) \text{ and } (j=1,\ldots,p) \tag{13}$$

By summing the efficient score residuals within each of the clusters in Equation (12), the U matrix now takes into account the fact that the score contributions are correlated within kth subject but are independent across the m clusters.[44, 51] Hardin and Hilbe[44] refer to this type of variance estimator as the 'modified sandwich' variance estimator (MSVE). Wei *et al.*[11] are credited with deriving the MSVE for the analysis of multivariate survival data. Wei *et al.*[11] found that the MSVE to be a consistent estimator of the variance of the parameter estimates even under the misspecification of the dependence structure (see also Lee *et al.*[8]).

3.2. *Wei, Lin and Weissfeld* (*WLW*) *marginal model and its variants*

3.2.1. Wei Lin Weissfeld Marginal Model (*WLW Model*). The WLW marginal model is specified with an event-specific baseline hazard function in which it allows the baseline hazard function to vary over each of the separate failure events and a semi-restricted risk set in which it allows the given individual to contribute information to the partial likelihood function of the kth event at time t as long as they have not experienced the kth event prior to time t and are still under observation at time t. Stated in another way, subjects are considered at the risk of kth event prior to experiencing the $(k-1)$th event. This is the risk set of choice for the analysis of unordered data where the risks are developing simultaneously.[9]

Therefore, WLW models can be specified as follows:

$$h_{ik}(t)=h_{0k}(t)\exp(\beta_k' Z_{ik}) \tag{14}$$

(Note that Equation (14) is the same as Equation (2)).

Models using the semi-restricted risk sets are often referred to as *marginal* or *population-averaged* models because all individuals are at risk of experiencing all of the events from the time of the initial onset of risk.[37]

This is achieved by incorporating a total time risk along with an event-specific baseline hazard function. The risk set for this model is $Y_{ik}(t)=I(x_{ik}\geq t)$. In this model, $Y_{ik}(t)$ is equal to 1 for the kth event for the entire time that the individual is under observation and only upon the occurrence of kth event it become equal to 0.

This model focusses on the distribution of each response as a distinct 'marginal' distribution and that the model is 'averaging' over all of the event times from the previous $(k-1)$th events.[52] As Therneau and Hamilton[53] note, the use of the term 'marginal' is to indicate 'what would result if the data recorder ignored all information except the given event type'.

3.2.2. LWA variant of the WLW model. Lee *et al.*[8] proposed the LWA variant of the WLW model.[8] This model is defined with a total time risk interval, an unrestricted risk set and with a baseline hazard function common to all k events. Thus the model can be specified as follows:

$$h_{ik}(t)=h_0(t)\exp(\beta' Z_{ik}) \tag{15}$$

4. Assessment of the fit of the WLW model

The most frequently applied methods for assessing the goodness of PH models require examining the Cox–Snell residual[54] plots. Crouchley and Pickles[55] explain that in the case of multivariate survival data (e.g. multiple failure data), the Cox–Snell residuals are no longer unit exponential. Consequently, standard residual plots based on the unit exponential distribution are no longer appropriate. They go on to explain that inspite of these problems visual inspection of such plots remains the standard check of model specification. Crouchley and Pickles[55] provide a clear example of a case where the Kaplan–Meier plot of Cox–Snell residuals contradicts the IM test suggested here.

This paper applies the IM test of White[12] for the checking of the PH model specification with multivariate failure time data. White's[12] paper on inference from misspecified models presented the IM test as a test for correct model specification. The IM test is based on the sum of the mean of the cross-product of the first derivatives of the log-likelihood and the mean of the second derivatives. Both these terms are calculated at the estimated parameter values. If the model is correct, the sum of these two alternative measures should be asymptotically zero.

With multivariate failure time data, three different tests can be formed.[55] The different tests arise from the treatment given to the dependence between the univariate margins and the over-dispersion within them. According to Crouhcley and Pickels[55], a test that examines homogeneity in the effect of the covariate β, in the univariate margins and can also be used for checking model specification, takes the form

$$\bar{D}_C(\hat{\beta}) = n^{-1}\left[\sum_i \sum_k [-\hat{h}_{ik}(t)] + \sum_i \sum_k [\delta_{ik} - \hat{h}_{ik}(t)][\delta_{ik} - \hat{h}_{ik}(t)]\right]$$

where $\bar{D}_C(\hat{\beta})$ follows a chi-square distribution with one degree of freedom under the correct model specification.

Here, the subscript k distinguishes each of the failure times from sample unit i.

5. Example

5.1. Description

We now present an empirical application to investigate which brand of PCs is preferred based on the impact of the brand on the timing of hardware failures. The data used in this study are from two groups of PCs sold from 4 April 2004 to 29 December 2006 on which a data set of 17 438 sales of PCs was obtained. Their failure times were observed until 25 May 2007 and only the hardware failures (which were assumed to be of a single type) occurring within this period were considered in this study. Together with the observed and censored observations over five failure events, 87 190 observations (5376 actual observations and 81 814 censored observations) were used. The only explanatory variable considered is the brand of the PC.

5.2. Descriptive statistics

For the 17 438 PCs, an overall 80% of the PCs survived without any report of a hardware failure, about 13% had only one failure, 4% had two failures and about 1% failed three times within the study period.

Figure 1(a)–(e) depicts the variation in the failure pattern between each brand for each of the five failure events, respectively, taking into account the time at which failures occur.

When Figure 1 is examined, a clear distinction between the failure patterns of the two brands can be observed. For the first two failures (1(a)–(b)), brand 1 shows a larger percentage of failures in each of the time points. This distinction is not so clear in the third and fourth failures (1(c)–(d)), where the curves are quite closer together. When considering the fifth failure event (1(e)), the failure curves cross each other indicating that brand 1 has a slightly lower percentage of failures at early time points of the study period while this pattern is reversed later on in the study period. These indications elucidated by Figure 1 suggest that there is a distinct difference among the failure curves of the two brands in each failure event, and further it can be seen that the failure pattern of a particular brand is not similar in all the failure events. From this arises the need for investigating the effect of the computer brand and whether this effect is the same in multiple failure occurrences of hardware failures.

5.3. Univariate tests

Under the methods explained for modeling multiple failure times, different approaches were discussed. The significant differences among these models were the way in which the risk sets were constructed and the way the risk intervals were defined. As stated by Hannan *et al.*[56], 'the programming to rearrange the data for the correct analysis requires a clear understanding of the implicit definition of the separate risk sets'. A procedure for identifying the proper structure of the risk interval suitable for the data set considered in this study is suggested by using log-cumulative hazard (LCH) plots based on Kaplan–Meier estimates for different types of risk intervals. The type of risk interval, which does not show a significant departure from the proportionality of the hazards, will be considered as the suitable data structure for the modeling session. When the suitable risk interval is identified in this way, the allocation of varied risk sets and baseline hazard functions would be done within the modeling session.

The models that are being used in this study are multivariate generalizations of the Cox PH model. The only assumption in Cox PH models is that the hazards are proportional over the strata variables. In this study, the only strata variable involved is the brand of PCs. Thus before applying PH models, the assumption of proportionality of hazards will be checked over the two brands of PCs for the five separate failure events.

This graphical test is a univariate approach and is not so reliable in the case of multivariate failure times since the dependence among the failure events cannot be grasped very clearly by this method. Thus, the indications illustrated by the LCH plots will be regarded as tentative and the validity of assumption will further be assessed in the modeling session.

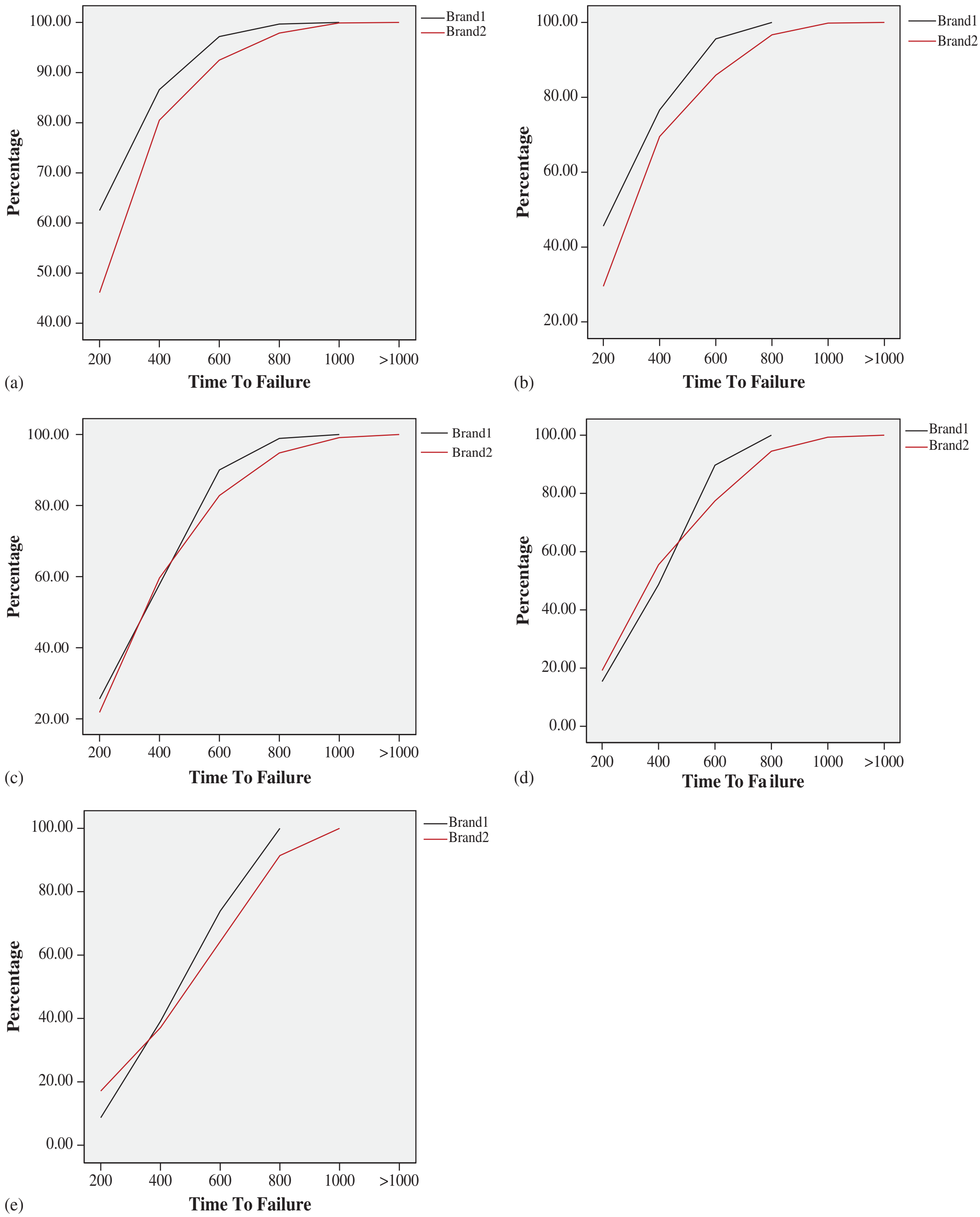

Figure 1. Percentage of failures for each failure event: (a) first failure; (b) second failure; (c) third failure; (d) fourth failure; and (e) fifth failure.

The two types of risk intervals used by the PH models explained earlier are total time risk intervals and gap time risk intervals, in which the kth total time considers the duration between the kth failure date and the installation date of the particular PC and the kth gap time risk interval considers the duration between kth failure date and the repair completion date of the $(k-1)$th failure as the risk interval for each failure. Therefore, an investigation will be done by using LCH plots to look for the risk set that comparatively secures the PH assumption. Figure 2(a)–(c) gives LCH plots for the first three gap time risk intervals, respectively.

When Figure 2(a) is examined, it is seen that the LCH curves for the first failure do not depart too much from being parallel. Thus, it can be concluded that the PH assumption is secured for the first failure when the risk interval is formed as gap time

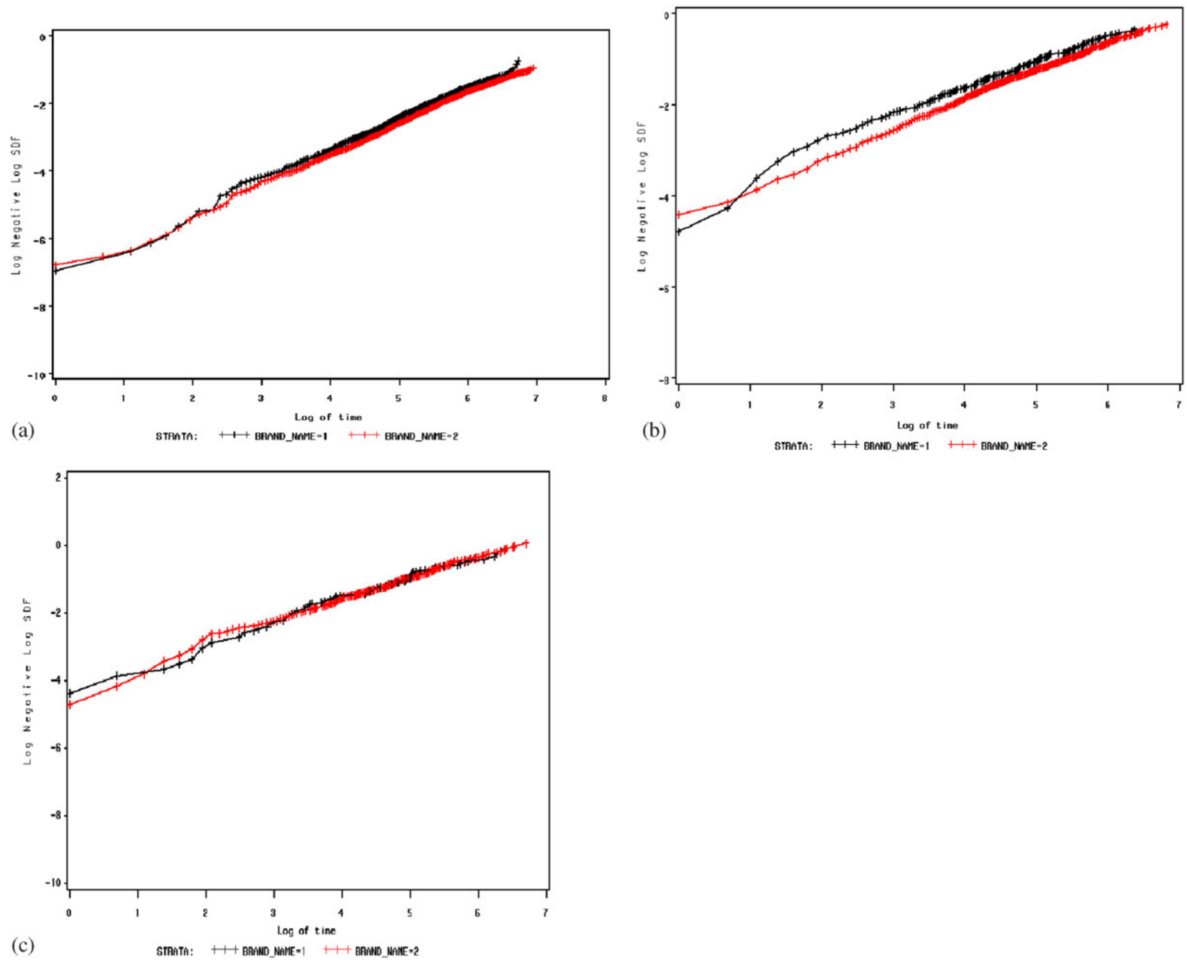

Figure 2. LCH plots for the gap time risk interval: (a) 1st failure; (b) 2nd failure; and (c) 3rd failure.

risk interval. However, it should be noted that both the gap time risk interval and the total time risk interval are the same for the first failure as both take the duration between the first failure date and the installation date of the particular PC. Therefore, it can be concluded that for the first failure event both types of risk intervals secure the PH assumption.

In case of the second failure (Figure 2(b)), it can be noted that the assumption of proportionality of hazards is slightly violated since the LCH curves of the two brands show one crossing. In the case of the third failure (Figure 2(c)), this assumption is fairly violated as the two LCH curves cross each other several times. Based on this investigation of Figure 2, further examination of gap time risk interval will be halted and the validity of the total time risk interval with respect to the PH assumption will be checked.

Figure 3(a)–(d) illustrates the LCH plots for the separate brands over the 2nd, 3rd, 4th and 5th total time risk intervals, respectively.

In Figure 2(a), it was identified that for the first failure both gap and total time risk intervals secure the PH assumption and from Figure 3(a) it can be seen that in case of second failure too the total time risk interval secures the PH assumption somewhat better than the gap time risk interval. When the third failure is concerned, here too the total time appears to be preferred over the gap time risk interval as the violation of PH assumption is slight. The same fact that the violation of PH assumption is not severe is elucidated in case of 4th and 5th failures also when the risk interval is constructed as the total time risk interval. These graphical tests thus point out that the total time risk interval is preferred to the gap time risk interval. Further, the appropriateness of the total time risk interval is checked using the Schoenfeld global test for PH.[57] The *p*-Values obtained for this test are listed in Table I along with the results of the log-rank test[58], which was done to test whether there is a significant difference among the performance of the two brands at each failure.

The *p*-Values for the log-rank test are all significant at 5% for each of the failures, indicating that the performance of the two brands has a significant difference at each of the five failure events. Further, the *p*-Values of the test for PH assumption appeared to be non-significant at the 5% level in each of the five failures. Thus, it can be concluded that the assumption of proportionality of hazards is prevalent in each of the failure instances under the total time risk interval.

With the findings of the univariate tests, it was decided to use the total time risk interval as the preferred structure of data for the data set of this study. But, as these univariate tests take into account each failure event separately, further validation will be done in the modeling session.

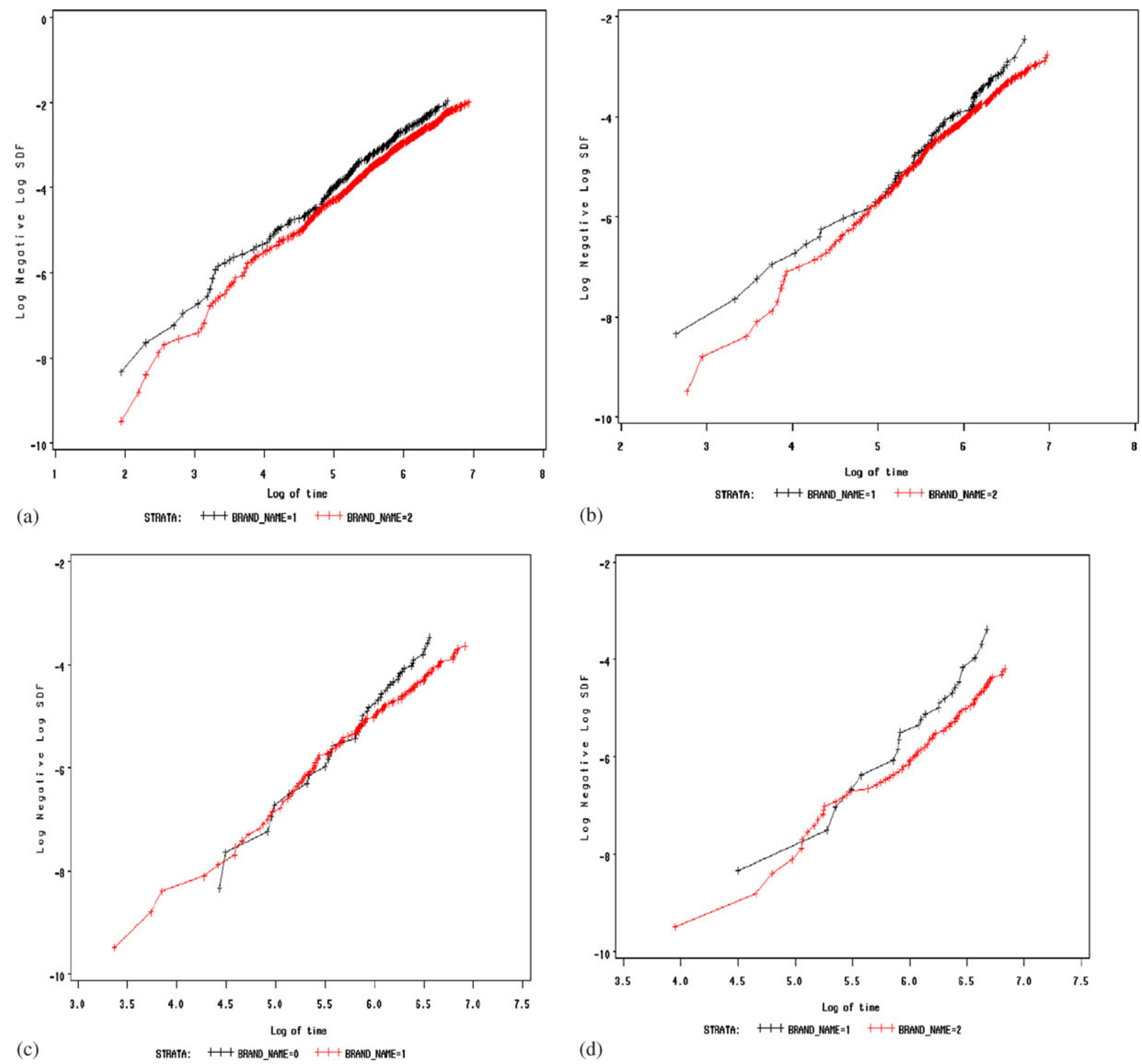

Figure 3. LCH plots for the total time risk interval: (a) 2nd failure; (b) 3rd failure; (c) 4th failure; and (d) 5th failure.

Table I. Summary of log-rank tests and the test for PH assumption over total time risk interval

Failure event	Log-rank statistic (LR)	df	*p*-Value for LR test	*p*-Value of test for PH	Proportionality of the hazards
First	8.513	1	0.0035	0.5989	Good
Second	14.7035	1	0.0001	0.6608	Good
Third	5.2116	1	0.0224	0.3108	Good
Fourth	4.0579	1	0.044	0.1698	Slight departure
Fifth	7.7012	1	0.0055	0.1794	Slight departure

5.4. Model fitting

Among the VCPH models described, the main differences were the definition of risk sets, definition of risk intervals and the form of the baseline hazard function. Rather than fitting all the VCPH models, it was considered sensible to identify a model (or models) that would be comparatively better than the rest of the models for the given data set and apply only the identified models. For this purpose, LCH plots for each failure event using both the gap time risk interval and the total time risk interval were drawn and presented in Section 5.2. These plots showed that the total time risk interval is preferred over the gap time risk interval as the PH assumption was approximately valid for the former but not for the latter. Therefore, only the models that are defined with a total time risk interval are applied in this study.

The WLW marginal model and the LWA variant of the WLW models are specified with a total time risk interval. The only difference among these two models is that the WLW model is defined with an event-specific baseline hazard and a semi-restricted

Table II. Details of the fitted models

Model	Parameter estimates ($\hat{\beta}$)	Hazard ratio $\exp(\hat{\beta})$	Std. error of $\hat{\beta}$	95% Conf. limits of the hazard ratio $\exp(\beta)$		P-Value
Model-1	−0.18590	0.83	0.03676	0.773	0.892	<0.0001
Model-2	−0.12409	0.883	0.04233	0.813	0.96	0.0002
	−0.27732	0.758	0.07222	0.658	0.873	0.0001
	−0.27294	0.761	0.12022	0.601	0.963	0.0231
	−0.36738	0.693	0.18331	0.484	0.992	0.0451
	−0.6773	0.508	0.2524	0.31	0.833	0.0073

risk set, whereas the LWA marginal model is defined with a common baseline hazard and an unrestricted risk set. Initially, the LWA model (*Model 1*), which assumes a common baseline hazard for all the failure events is fitted to the data set. Then, a WLW marginal model (*Model 2*) which assumes five different baseline hazards for the five failure events, is examined. Table II gives the details of the fitted models.

The p-Value of the Wald test for model 1 indicates that the effect of the brand is highly significant (p-Value $<.0001$). A hazard ratio of 0.83 indicates that the instantaneous failure rate of brand 2 PCs is less than that of brand 1 PCs and that the failure rate of brand 2 PCs is about 17% less than the failure rate of the brand 1 PCs. However, further assessment of the fit of the model is required before concluding the final model to model the multiple failure occurrences of PCs. The IM test is carried out on *model 1* to find out the goodness of the model specification.

The $\bar{D}_C(\beta)$ statistic for the LWA model, which assumes a common β for all the five failure events, is 1.72052 on 1 degree of freedom. A p-Value of 0.1896 indicates that the model is correctly specified. Therefore, it is reasonable to assume that the ratio between the instantaneous failure rates of the two brands of PCs is common in all the five failure events. That is, the chance of a failure in brand 1 PCs is higher than that of the brand 2 PCs is consistent over all five failure events.

Model 2 given in Table II gives the details of the WLW model fitted to the data set. The p-Values of the Wald test statistics given indicates that the brand effect is significant over all event times. The 95% confidence intervals given separately for the five failure events appear to be overlapping, thereby further indicating that assuming the baseline is changing over the five failure events is unnecessary. Thus, it can be concluded that the LWA model fits the data set well.

The parameter estimates obtained for the LWA model indicate that the chance of a hardware failure in brand 2 Pcs is less than the chance of a failure in Brand 1 PCs and it remains the same in all five failure events.

6. Conclusion

The analysis carried out in this study showed that the computer brand affects failure event timings. Further it was identified that this effect of the computer brand remains constant over the multiple failure occurrences. The failure rate of brand 2 PCs is approximately 17% less than that of brand 1 PCs.

The VCPH models discussed in this paper give an appropriate analysis of multiple failure occurrences in which dependence between the failure events is captured. The conventional method of checking the goodness-of-fit of a Cox proportional model (Cox–Snell residuals, martingale residuals, etc) is not appropriate in the event of multiple failure occurrences due to the dependence among failure events.

In contrast to the usual methods that have been frequently used to check the goodness-of-fit of PH models, the IM test proposed by White[12], which is preferred in the presence of multiple failures is used in this study.

Lin and Wei[42] have performed extensive simulation studies to evaluate the sensitivity of the PH assumption to sample size, in both the usual Cox model and in the VCPH Cox model. Their study indicates that both the model-based Wald and score tests will exceed the nominal level when the true model has non-PH irrespective of the sample size. However, the robust Wald and score tests maintain their size near the nominal level for large samples. This indicates that the validity of the PH assumption is vital for the usual Cox model irrespective of sample size; however, the VCPH model is fairly robust to departures from PH for large samples (as is the case in this example). Lawless and Thiagarajah[59] use a simulation study to show that for HPP and renewal process models the size of both Wald and score tests exceed the nominal value and do not vary systematically as the sample size increases when applied to repairable systems. They also show that the Laplace test indicates that the NHPP process performs better in the sense that the size of the test is closer to the nominal value; however, the size does not vary systematically with respect to the sample size when applied to repairable systems.

In a repairable system prone to multiple failures, times between failures will usually not be independent nor identically distributed ruling out the use of HPP and renewal process models. The VCPH model proposed here avoids both these assumptions as does the NHPP model. The only assumption in the VCPH model is that the hazards between levels of the covariates are proportional. While this assumption should be checked, the model is fairly robust to departures from this assumption for large samples. The NHPP model on the other hand has been found to have inconsistent estimators and there is some criticism about its use for repairable systems. Thus, the VCPH model is seen as preferable to the NHPP model.

Appendix A

```
/*Read the data set into SAS*/

PROC IMPORT OUT= WORK.REPAIR
DATAFILE= 'E:\project\final_data.csv'
        DBMS=CSV REPLACE;
   GETNAMES=YES;
   DATAROW=2;
RUN;
/* Re-arrange the data set in a suitable way for modeling*/
data com_repairs;
set repair;
keep id tstart tstop status brand_name visit;
array ff F1-F5;
array cc C1-C5;
infile datalines missover;
ff[1]= first;
ff[2]= second;
ff[3]= third;
ff[4]= fourth;
ff[5]= fifth;
cc[1]= first_out;
cc[2]= second_out;
cc[3]= third_out;
cc[4]= fourth_out;
cc[5]= fifth_out;
id+1;
count=0;
tstart=0;
k=0;
do i=1 to 5;
k=k+1;
visit=i;
   if ff[i]= . then do;
      count=count+1;
      tstop=follow_up_time;
      status=0;
   end;

   else do;
      tstop=ff[i];
      status=1;
   end;
   output;

   if count>0 then do;
      tstart=follow_up_time;
   end;
      else do;
         tstart=cc[k];
      end;
   end;

   if(tstart<follow_up_time) then delete;
   run;
/*Fitting the LWA model */
proc phreg data=com_repairs;
         model tstop*status(0)= brand_name ;
         strata visit;
```

```
run;
/* construct a data set that consist of interaction terms of brand and failure event*/

data com_repairs2;
set com_repairs;
        if visit < 6;
        failure1= brand_name * (visit=1);
        failure2= brand_name * (visit=2);
        failure3= brand_name * (visit=3);
        failure4= brand_name * (visit=4);
        failure5= brand_name * (visit=5);
run;
/* Fitting the WLW model*/

proc phreg data=com_repairs2;
        model tstop*status(0)= failure1-failure5;
        strata visit;
        id id;
run;
```

References

1. Wikipedia 2008. Wikipedia Foundation, Inc. California. Available at: http://en.wikipedia.org/wiki/Personal Computer [1 November 2008].
2. Cox DR. Regression models and life tables. *Journal of the Royal Statistical Society* 1972; **34**:187–220.
3. Lim HJ, Liu J, Melzer-Lange M. Comparison of methods for analyzing recurrent event data: Application to the emergency department visits of the pediatric firearm victims. *Accident Analysis and Prevention* 2007; **39**(2):290–299.
4. Therneau TM, Patricia MG. *Modeling Survival Data*: *Extending the Cox Model*. Springer: New York, 2000.
5. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**:13–22.
6. Zeger SL, Liang K-Y. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986; **42**:121–130.
7. Hannan PJ, Xiao O, Shu DW, Goldman A. Analysis of failure times for multiple infections following bone marrow transplantation: An application of multiple failure time proportional hazards model. *Statistics in Medicine* 1998; **17**:2371–2380.
8. Lee EW, Wei LJ, Amato DA. Cox-type regression analyses for large numbers of small groups of correlated failure time observations. *Survival Analysis*: *State of the Art*, Klein JP, Goel PK (eds.). Kluwer Academic Publishers: Derdrest, Netherlands, 1992; 237–247.
9. Lin DY. Cox regression analysis of multivariate failure time data: The marginal approach. *Statistics in Medicine* 1994; **13**:2233–2247.
10. Lin DY. Multivariate failure time data. *Recent Advances in Clinical Trial Design and Analysis*, Thall PF (ed.). Kluwer Academic Publishers: Boston, MA, 1995; 73–93.
11. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* 1989; **84**:1065–1073.
12. White H. Maximum likelihood estimation of misspecified models. *Econometrica* 1982; **50**:1–25.
13. Ezzell ME, Land KC, Cohen LE. Modeling multiple failure time data: A survey of variance corrected proportional hazard models with empirical application to arrest data. *Sociological Methodology* 2003; **33**:111–167.
14. Ebeling CE. *An Introduction to Reliability and Maintainability Engineering*. Tata McGraw-Hill: New Delhi, 1997.
15. O' Connor PDT. *Practical Reliability Engineering*. Wiley: India, 2002.
16. Srinath LS. *Reliability Engineering*. East-West Press: New Delhi, 2005.
17. Cinlar E. *Introduction to Stochastic Processes*. Prentice-Hall: Englewood Cliffs, NJ, 1975.
18. Tan F, Jiang Z, Bai T. Reliability analysis of repairable systems using stochastic point processes. *Journal of Shanghai Jiaotong University* (*Sci*) 2008; **13**(3):366–369.
19. Thompson Jr WA. On the foundations of reliability. *Technometrics* 1981; **23**:1–13.
20. Barlow RE, Proschan E, Hunter LC. *Mathematical Theory of Reliability*. Wiley: New York, 1967.
21. Goldberg H. *Extending the Limits of Reliability Theory*. Wiley: New York, 1981.
22. Yanez M, Joglar F, Modarres M. Generalized renewal processes for analysis of repairable systems with limited failure experience. *Reliability Engineering and System Safety* 2002; **77**(2):167–180.
23. Yu J, Tian G, Tang M. Predictive analyses for non-homogeneous Poisson processes with power law using Bayesian approach. *Computational Statistics and Data Analysis* 2007; **51**(9):4254–4268.
24. Ascher H, Feingold H. *Repairable Systems Reliability*. Marcel Dekker: New York, 1984.
25. Zhao M, Xie M. On maximum likelihood estimation for a general non-homogeneous Poisson process. *Scandinavian Journal of Statistics* 1996; **23**:597–607.
26. Jeske DR, Pham H. On the maximum likelihood estimates for the Goel–Okunoto software reliability model. *American Statistician* 2001; **55**:219–222.
27. Nayak TK, Bose S, Kundu S. On inconsistency of estimators of parameters of non-homogeneous Poisson process models for software reliability. *Statistics and Probability Letters* 2008; **78**(14):2217–2221.
28. Cook RJ, Lawless JF. *The Statistical Analysis of Recurrent Events*. Springer Science: New York, 2007.
29. Prasad PVN, Rao KRM. Reliability models of repairable systems considering the effect of operating conditions. *Proceedings of the Annual Reliability and Maintainability Symposium*, Seattle, WA, 2002; 503–510.
30. Peña E, Hollander M. Models for recurrent events in reliability and survival analysis. *Mathematical Reliability*: *An Expository Perspective*, Chapter 6, Soyer R, Mazzuchi T, Singpurwella N (eds.). Kluwer Academic Publishers: Dordrecht, 2004; 105–123.
31. Cox DR. The statistical analysis of dependencies in point processes. *Stochastic Point Processes*: *Statistical Analysis, Theory, and Applications* (Conference, IBM Research Center, Yorktown Heights, NY, 1971). Wiley-Interscience: New York, 1972; 55–66.
32. Prentice R, Williams B, Peterson R. On the regression analysis of multivariate failure time data. *Biometrika* 1981; **68**:373–379.

33. Peña E, Slate E, Gonzales J. Semiparametric inference for general class of models for recurrent events. *Journal of Statistical Planning and Inference* 2007; **137**:1727–1747.
34. Pan R, Rigdon S. Bayes inference for general repairable systems. *Journal of Quality Technology* 2009; **41**(1):82–94.
35. Dhillon BS, Yang N. Reliability analysis of a repairable robot system. *Journal of Quality Maintenance Engineering* 1996; **2**(2):30–37.
36. Xie M, Ho SL. Analysis of repairable system failure data using time series models. *Journal of Quality and Maintenance Engineering* 1999; **5**(1):50–61.
37. Kelly PJ, Lynette L-Y, Lim C. Survival analysis for recurrent event data: An application to childhood infectious diseases. *Statistics in Medicine* 1999; **19**:13–33.
38. Finkelstein DM, Schoenfeld DA, Stamenovic E. Analysis of multiple failure time data from an aids clinical trial. *Statistics in Medicine* 1997; **16**:951–961.
39. Clayton D. The analysis of event history data: A review of progress and outstanding problems. *Statistics in Medicine* 1988; **7**:819–841.
40. Klein JP, Moeschberger ML. *Survival Analysis*: *Techniques for Censored and Truncated Data*. Springer: New York, 1997.
41. Wei LJ, Glidden DV. An overview of statistical methods for multiple failure time data in clinical trials. *Statistics in Medicine* 1997; **16**:833–839.
42. Lin DY, Wei LJ. The robust inference for the Cox proportional hazard model. *Journal of the American Statistical Association* 1989; **84**:1074–1078.
43. Struthers CA, Kalbfleisch JD. Misspecified proportional hazard models. *Biometrika* 1986; **71**:291–298.
44. Hardin J, Hilbe J. *Generalized Linear Models and Extensions*. Stata Press: College Station, TX, 2001.
45. Liang K-Y, Self SG, Chang Y-C. Modelling marginal hazards in multivariate failure time data. *Journal of the Royal Statistical Society B* 1993; **55**:441–453.
46. Burton P, Gurrin L, Sly P. Extending the simple linear regression model to account for correlated responses: An introduction to generalized estimating equations and multi-level modeling. *Statistics in Medicine* 1998; **17**:1261–1291.
47. Zeger SL, Liang K-Y. An overview of methods for the analysis of longitudinal data. *Statistics in Medicine* 1992; **11**:249–273.
48. Huber PJ. The behavior of maximum likelihood estimates under non-standard conditions. *Proceedings of the Fifth Symposium on Mathematical Statistics and Probability*, Berkeley, CA, vol. 1, 1967; 221–233.
49. White H. A heteroscedasticity—Consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrika* 1980; **48**:317–330.
50. Cleeves M. Analysis of multiple failure time data with stata. *Stata Technical Bulletin* 1999; **49**:30–39.
51. Barlow WE. Robust variance estimation for the case-cohort design. *Biometrics* 1994; **50**:1064–1072.
52. Cook RJ, Lawless JF. Discussion of paper by Wei and Glidden. *Statistics in Medicine* 1997; **16**:841–842.
53. Therneau TM, Hamilton SA. RhDNase as an example of recurrent event analysis. *Statistics in Medicine* 1997; **16**:2029–2047.
54. Cox DR, Snell EJ. A general definition of residuals (with Discussion). *Journal of the Royal Statistical Society, Section B* 1968; **30**:248–275.
55. Crouchley R, Pickels A. A specification test for univariate and mulitvariate proportional hazard models. *Biometrics* 1993; **49**:1067–1076.
56. Hannan PJ, Xiao OS, Daniel W, Goldman A. Analysis of failure times for multiple infections following bone marrow transplantation: An application of multiple failure time proportional hazards model. *Statistics in Medicine* 1988; **17**:2371–2380.
57. Abeysekera WWM, Sooriyarachchi MR. Use of Schoenfeld's global test to test the proportional hazards assumption in the Cox proportional hazards model: An application to a clinical study. *Journal of the National Science Foundation of Sri Lanka* 2009; **37**(1):43–53.
58. Collett D. *Modeling Survival Data in Medical Research*. Chapman & Hall: London, 1994.
59. Lawless JF, Thiagarajah K. A point-process model incorporating renewals and time trends, with application to repairable systems. *Technometrics* 1996; **38**(2):131–138.

Authors' biographies

A. A. Sunethra has a BSc (Statistics Special) degree with Second Class Upper Division from the University of Colombo in Sri Lanka. She is a Temporary Assistant Lecturer in the Department of Statistics, University of Colombo and she lectures on Statistical Reliability to undergraduates following the BSc (Statistics Special) programme.

M. R. Sooriyarachchi is a Professor of Statistics in the Department of Statistics, University of Colombo in Sri Lanka. She has a PhD in Applied Statistics from the Department of Applied Statistics, University of Reading, U.K. She has several papers to her credit in both international and local journals.