Cause-Selecting Charts Based on Proportional Hazards and Binary Frailty Models

Sh. Asadzadeh, A. Aghaie * & H. Shahriari

Shervin Asadzadeh, PhD student, Department of Industrial Engineering, K.N.Toosi University of Technology, sh_asadzadeh@dena.kntu.ac.ir
Abdollah Aghaie Professor, Department of Industrial Engineering, K.N.Toosi University of Technology, aaghaie@ kntu.ac.ir
Hamid Shahriari Associate Professor, Department of Industrial Engineering, K.N.Toosi University of Technology, hshahriari@ kntu.ac.ir

KEYWORDS

ABSTRACT

Monitoring the reliability of products in both the manufacturing and service processes is of main concern in today's competitive market. To this end, statistical process control has been widely used to control the reliability-related quality variables. The so-far surveillance schemes have addressed processes with independent quality characteristics. In multistage processes, however, the cascade property must be effectively justified which entails establishing the relationship among quality variables with the purpose of optimal process monitoring. In some cases, measuring the values corresponding to specific covariates is not possible without great financial costs. Subsequently, the unmeasured covariates impose unobserved heterogeneity which decreases the detection power of a control scheme. The complicated picture arises when the presence of a censoring mechanism leads to inaccurate recording of the process response values. Hence, frailty and Cox proportional hazards models are employed and two regression-adjusted monitoring procedures are constructed to effectively account for both the observed and unobserved influential covariates in line with a censoring issue. The simulation-based study reveals that the proposed scheme based on the cumulative sum control chart outperforms its competing procedure with smaller out-of-control average run length values.

1. Introduction

In some multistage manufacturing or service operations, quality characteristics of interest are highly dependent. This implies that a change in an incoming quality variable affects some or all outgoing quality variables. The mentioned attribute is referred to as the cascade property which is the main feature of multistage processes [1]. Thus, the effective procedure is to model the dependency structure among quality characteristics and then monitor a specific quality variable only after it has been adjusted for the effect of all influential upstream variables. Cause selecting charts (CSCs) have been proposed for the sake of monitoring and diagnosing multistage processes with normally distributed quality variables [2]. The underlying models are the linear or multiple linear regression models depending on the number of quality

variables in the process. However, the normality assumption in the use of CSCs makes it a cumbersome monitoring procedure, in that quality characteristics do not always follow this assumption. To relax the normality assumption, some papers have addressed regression adjustment approach based on generalized linear models (GLMs) which include exponential family distributions [3-6].

The above-mentioned monitoring schemes are not fruitful in case the main intention of using a control procedure is to improve the reliability of products (services). Reliability (survival) data have some specific properties which make them quite different. First, they are commonly modeled by a distribution which is a member of Location-Scale or Log-Location-Scale distributions. Second, it is a common practice for their values to be censored as they reach a pre-determined limit selected beforehand to ensure the minimum desirable reliability [7]. There exist numerous quality characteristics which evaluate the product reliability. The tensile strength of adhesive bond between a vinyl fabric and polyvinyl chloride (PVC) foam backing in the interior of a car, the skein strength of spun cotton, the breaking strength of weld and the survival times of patients after performing the surgical operation are among the most commonly used quality variables in manufacturing and service processes. The problem of monitoring quality variables which follow the Location-Scale and Log-Location-Scale distributions in the absence of a censoring mechanism has been addressed largely in the literature (see for example [8-12]).

To account for the censoring issue, Steiner and Mackay [13-15] presented monitoring schemes for Weibull and normally distributed reliability data. Both the fixed and variable censoring levels have been considered respectively. Zhang and Chen [16] proposed exponentially weighted moving average (EWMA) control chart for a process characterized by Weibull distribution in the presence of a fixed censoring level. Furthermore, Olteanu [17] developed some CUSUM control charts for censored reliability data with Weibull distribution. Obviously, these papers have been devoted solely to the processes with no cascade property. However, the optimal multistage process monitoring entails considering the impact of influential quality variables as well.

For instance, in an industrial process, the tensile strength of adhesive bond between vinyl fabric and foam backing is affected by the amount of the dominant plasticizer called DOP (Dioctyl Phthalate). The more DOP dampens the adhesive bond strength between the two surfaces. But, it gives more flexibility and durability to the mentioned plastics. Moreover, the skein strength of the spun cotton is affected by the fiber length and the fiber strength which can be referred to as influential quality variables. Besides, in a service organization such as a hospital, the survival times of patients are definitely dependent on the unique risk factors arising from each patient's health history. Sego et al. [18] elaborated on the case in which the survival times of cardiac patients are adjusted for the effect of Parsonnet score which is a weighted composite of factors associated with the risk of mortality for adults with heart disease.

However, it is not always straightforward to incorporate all covariates into the model to describe their influences since measuring their values is inevitable without great financial cost and time effort or possibly no prior information is available on their values. The neglect of such covariates leads to the unobserved heterogeneity which dampens the detection ability of control schemes. Consequently, the purpose of this paper is to present monitoring procedures in multistage processes with both the observed and unobserved covariates. The construction of the proposed control schemes is discussed in the presence and absence of a censoring mechanism.

2. Process Modeling

Consider a cascade process in which the outgoing quality characteristic is affected by the incoming quality variables. The output quality variable of interest, denoted by *Y*, is selected in a way to evaluate and reflect the reliability of products (services). Without loss of generality, it is assumed that the number of influential variables (*X*) is equal to 2 in this multistage process monitoring which introduce heterogeneity to the process output. However, including both of these covariates in the analysis is not possible due to the complexity of measurement or some other time effort and budget restrictions. Apparently, the application of the so-far monitoring methods, discussed in Introduction, is useless because they do not justify the cascade property, unobserved heterogeneity and a censoring issue simultaneously. Therefore, a novel control procedure is indeed needed to tackle the mentioned problems.

Survival analysis regression models have been devised to explain the occurrence of reliability data taking various covariates into account. The two main approaches are the Cox proportional hazard (PH) and the accelerated failure time (AFT) models [7]. In this paper, the PH model is implemented to form the hazard and survival functions of output quality variable. Due to the fact that only one observable covariate has been assumed in this multistage process monitoring, the hazard and the survival functions are represented as:

$$h(y \mid x) = h_0(y)\exp(\beta x)$$
$$S(y \mid x) = S_0(y)^{\exp(\beta x)}$$

(1)

where $h_0(y)$ and $S_0(y)$ are the baseline hazard and baseline survival functions respectively and β is the regression parameter. Without loss of generality, the underlying distribution for the baseline hazard and

survival functions is assumed to be the Weibull distribution which is the most commonly used one:

$$h_0(y) = \frac{\kappa}{\eta}(\frac{y}{\eta})^{\kappa-1}$$

$$S_0(y) = \exp\left(-\left(\frac{y}{\eta}\right)^{\kappa}\right) \tag{2}$$

where $\kappa > 0$ is a shape parameter and $\eta > 0$ is a scale parameter.

In addition, the frailty models are used which offer a convenient way to introduce the heterogeneity imposed by unobserved covariate [19]. It specifies that:

$$h(y \mid \gamma) = \gamma h_0(y) \tag{3}$$

To successfully account for both the observed and unobserved covariates, the hazard function takes the form of:

$$h(y \mid x, \gamma) = \gamma h_0(y) \exp(\beta x) \tag{4}$$

Referring to our industrial case, consider that the tensile strength of an adhesive bond between the vinyl fabric and PVC foam backing in the interior of a car were to be evaluated. The DOP is considered as an observable incoming quality variable. However, measuring such values is not always straightforward without great time effort and financial cost. Hence, this quality variable can be considered as an unobserved covariate as well which makes the use of frailty models indispensable.

In addition, various suppliers from whom the vinyl fabric and PVC foam backing are provided may be contemplated as another example of unobserved covariate. The products provided by each supplier may have different levels of quality which affect the strength of the adhesive bond. However, in a process, it may be indistinguishable which manufacturer a product has come from and measuring the quality of such products is complicated.

Finally, the surface characteristic of the discussed products has severe effects on the bond strength. Measuring such values as an incoming covariate is impossible without using technical equipment and spending a lot of cost. These bring about the situations in which the effect of both the observed and unobserved covariate must be justified. Assuming two different suppliers or two types of surface characteristic, a binary frailty model is used. Let the proportion of products from the first supplier with frailty value γ_1 is π. To fulfill the standardization condition ($\mathbf{E}(\gamma)=1$) the following relation yields the frailty value corresponding to the second supplier

$$\gamma_2 = \frac{1 - \pi\gamma_1}{1 - \pi} \tag{5}$$

Denote $S(y \mid x, \gamma)$ to be the survival function of the output response given the observed covariate x and frailty γ, the unconditional survival function is obtained as follows:

$$S(y) = \mathbf{E}(S(y \mid x, \gamma)) = \mathbf{E}(e^{-\int_0^y h(y|x,\gamma)dy}) = \mathbf{E}(e^{-\int_0^y \gamma h_0(y)\exp(\beta x)dy})$$
$$= \mathbf{E}(e^{-\gamma H_0(y)\exp(\beta x)}) = \mathbf{L}(H_0(y)\exp(\beta x)) \tag{6}$$

where H_0 is the cumulative baseline hazard function and $\mathbf{L}$ is the Laplace transform of the frailty γ. The density and hazard functions are then characterized by the above Laplace transform and its derivatives:

$$f(y) = -h_0(y)\exp(\beta x)\mathbf{L}'(H_0(y)\exp(\beta x))$$
$$h(y) = -h_0(y)\exp(\beta x)\frac{\mathbf{L}'(H_0(y)\exp(\beta x))}{\mathbf{L}(H_0(y)\exp(\beta x))} \tag{7}$$

Hence, the unconditional functions for the output quality variable of the discussed process are given by the following equations:

$$S(y) = \pi e^{-\gamma_1 H_0(y)\exp(\beta x)} + (1-\pi)e^{-\gamma_2 H_0(y)\exp(\beta x)}$$
$$f(y) = (\pi\gamma_1 e^{-\gamma_1 H_0(y)\exp(\beta x)} + (1-\pi)\gamma_2 e^{-\gamma_2 H_0(y)\exp(\beta x)})h_0(y)\exp(\beta x) \tag{8}$$
$$h(y) = \frac{\pi\gamma_1 e^{-\gamma_1 H_0(y)\exp(\beta x)} + (1-\pi)\gamma_2 e^{-\gamma_2 H_0(y)\exp(\beta x)}}{\pi e^{-\gamma_1 H_0(y)\exp(\beta x)} + (1-\pi)e^{-\gamma_2 H_0(y)\exp(\beta x)}}h_0(y)\exp(\beta x)$$

Next section elaborates on the control strategies based on the PH and frailty models.

3. Monitoring Procedures

In this section, two control procedures are developed for the sake of detecting out-of-control situations. The proposed control charts are one-sided since detecting deterioration in the product reliability is of main interest.

To this end, a coefficient υ' is used to shift the nominal in-control value corresponding to the scale parameter of the Weibull distribution since changing the scale parameter has the similar interpretation as changing the Weibull mean. Note that two different

scenarios including the presence and the absence of a censoring mechanism are considered respectively.
Control charts based on probability limits are the most straightforward monitoring procedure. The lower control limit (*LCL*) of this monitoring scheme is obtained as follows:

$$\alpha = \int_x \int_0^{LCL} ((\pi\gamma_1 e^{-\gamma_1 H_0(y)\exp(\beta x)} + (1-\pi)\gamma_2 e^{-\gamma_2 H_0(y)\exp(\beta x)})$$

$$h_0(y)\exp(\beta x))f(x)dydx \tag{9}$$

$$\alpha = \int_x (1 - (\pi e^{-\gamma_1 H_0(LCL)\exp(\beta x)} + (1-\pi)e^{-\gamma_2 H_0(LCL)\exp(\beta x)}))$$

$$f(x)dx$$

$$CEV_i =$$

$$\begin{cases} y_i & \text{if} \quad y_i \le c \\ (\int_c^\infty (\pi\gamma_1 e^{-\gamma_1 H_0(y)\exp(\beta x_i)} + (1-\pi)\gamma_2 e^{-\gamma_2 H_0(y)\exp(\beta x_i)})h_0(y)\exp(\beta x_i)\,y\,dy)\,/ \\ \pi e^{-\gamma_1 H_0(c)\exp(\beta x_i)} + (1-\pi)e^{-\gamma_2 H_0(c)\exp(\beta x_i)} & \text{otherwise} \end{cases} \tag{10}$$

The next surveillance procedure is based on the cumulative sum (CUSUM) control chart. The CUSUM statistic is given by:

$$s_i = \min(0, s_{i-1} - w_i) \quad i = 1, 2, \dots$$
$$s_0 = 0 \tag{11}$$

where w_i is the CUSUM score computed as follows conditional that the observations are recorded genuinely:

$$w_i = -\kappa \log\upsilon + \log\left(\pi(\gamma_1 e^{-\gamma_1 (\frac{y_i}{\upsilon\eta})^\kappa \exp(\beta x_i)} - \gamma_2 e^{-\gamma_2 (\frac{y_i}{\upsilon\eta})^\kappa \exp(\beta x_i)}) + \right.$$

$$\left. \gamma_2 e^{-\gamma_2 (\frac{y_i}{\upsilon\eta})^\kappa \exp(\beta x_i)} \right) - \log\left(\pi(\gamma_1 e^{-\gamma_1 (\frac{y_i}{\eta})^\kappa \exp(\beta x_i)} - \gamma_2 e^{-\gamma_2 (\frac{y_i}{\eta})^\kappa \exp(\beta x_i)}) + \right. \tag{12}$$

$$\left. \gamma_2 e^{-\gamma_2 (\frac{y_i}{\eta})^\kappa \exp(\beta x_i)} \right)$$

The constructed control procedure generates a signal as soon as the control chart statistic falls below the *LCL*. It should be noted that in case of having a censoring mechanism with a fixed censoring level, denoted by *c*, all censored observations can be replaced with their conditional expected values (CEVs). Doing so, the control chart statistic takes the form ofl:

in which υ is a pre-specified shift that the CUSUM chart is designed for optimal detection. Thus, the CUSUM chart triggers a signal when the value statistic is less than its *LCL*.
The development of the CUSUM chart undergoes a thorough modification as the presence of a censoring mechanism leads to inaccurate recording of a fraction of data. For the sake of simplicity, the following variable is defined as:

$$z_i = \min(y_i, c)$$

$$\delta_i = \begin{cases} 1 & \text{if} \quad y_i \le c \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

Subsequently, the CUSUM score is obtained via the below equation:

$$w_i = \delta_i \left(-\kappa \log\upsilon + \log\left(\pi(\gamma_1 e^{-\gamma_1 (\frac{z_i}{\upsilon\eta})^\kappa \exp(\beta x_i)} - \gamma_2 e^{-\gamma_2 (\frac{z_i}{\upsilon\eta})^\kappa \exp(\beta x_i)}) + \gamma_2 e^{-\gamma_2 (\frac{z_i}{\upsilon\eta})^\kappa \exp(\beta x_i)} \right) \right.$$

$$-\log\left(\pi(e^{-\gamma_1 (\frac{z_i}{\upsilon\eta})^\kappa \exp(\beta x_i)} - e^{-\gamma_2 (\frac{z_i}{\upsilon\eta})^\kappa \exp(\beta x_i)}) + e^{-\gamma_2 (\frac{z_i}{\upsilon\eta})^\kappa \exp(\beta x_i)} \right)$$

$$-\log\left(\pi(\gamma_1 e^{-\gamma_1 (\frac{z_i}{\eta})^\kappa \exp(\beta x_i)} - \gamma_2 e^{-\gamma_2 (\frac{z_i}{\eta})^\kappa \exp(\beta x_i)}) + +\gamma_2 e^{-\gamma_2 (\frac{z_i}{\eta})^\kappa \exp(\beta x_i)} \right) \tag{14}$$

$$\left. +\log\left(\pi(e^{-\gamma_1 (\frac{z_i}{\eta})^\kappa \exp(\beta x_i)} - e^{-\gamma_2 (\frac{z_i}{\eta})^\kappa \exp(\beta x_i)}) + e^{-\gamma_2 (\frac{z_i}{\eta})^\kappa \exp(\beta x_i)} \right) \right)$$

$$+\log\left(\pi(e^{-\gamma_1(\frac{z_i}{\upsilon\eta})^\kappa \exp(\beta x_i)} - e^{-\gamma_2(\frac{z_i}{\upsilon\eta})^\kappa \exp(\beta x_i)}) + e^{-\gamma_2(\frac{z_i}{\upsilon\eta})^\kappa \exp(\beta x_i)} \right)$$

$$-\log\left(\pi(e^{-\gamma_1(\frac{z_i}{\eta})^\kappa \exp(\beta x_i)} - e^{-\gamma_2(\frac{z_i}{\eta})^\kappa \exp(\beta x_i)}) + e^{-\gamma_2(\frac{z_i}{\eta})^\kappa \exp(\beta x_i)} \right)$$

Obtaining the CUSUM score, the plotted statistic is calculated as shown by equation (11).

4. ARL Study

Simulation-based studies are conducted in this section to investigate and compare the performance of the proposed control charts. To this end, the average run length (ARL) criterion is used as the performance measure. Note that the ARLs are calculated accurately for the first monitoring scheme as follows:

$$ARL =$$
$$1/\int_x (1-(\pi e^{-\gamma_1(\frac{LCL}{\upsilon'\eta})^\kappa \exp(\beta x)} + (1-\pi)e^{-\gamma_2(\frac{LCL}{\upsilon'\eta})^\kappa \exp(\beta x)}))f(x)dx \quad (15)$$

while simulation studies with 10000 replications are used to obtain the ARLs of the CUSUM chart. The LCLs of the both monitoring procedures are set in way to reach the in-control ARL of approximately 200. Concentrating on the no-censoring scenario, the two competing control charts are compared to detect decreasing mean shifts of size 2.5%, 5%, 10%, 20% and 30%.

The results are provided in Table I. It is remarkable that the CUSUM control chart far outweighs its competing counterpart because the out-of-control ARL values are much smaller. To exemplify, we may refer to the out-of-control ARLs when there exists a 10 percent reduction in the mean. It is noticeable that the proposed CUSUM control chart generates a signal 93 points sooner than the other control chart, in that the corresponding ARL values are 77 and:

Tab. 1. The performance of the monitoring procedures in the absence of censoring

υ'	CUSUM control chart		Control chart with probability limit
	ARL	Standard Error	ARL
1	200.1058	1.6526	200
0.975	155.7662	1.2117	192.5674
0.95	122.6633	0.9040	185.2308
0.9	77.0165	0.5386	170.8466
0.8	41.5014	0.2624	143.2725
0.7	25.0265	0.1484	117.3724

170 respectively. This indicates that the detection power of the CUSUM control chart is much greater.

Provided that the CUSUM chart is superior, its performance is studied in the presence of censoring. As a result, low, moderate and high censoring rates of 20%, 50% and 80% are considered. For easier comparison, the ARL curves of censoring and no-censoring scenarios are illustrated in Figure 1. The outcomes show a considerable increase in the values of out-of-control ARLs as the censoring rate becomes higher. This implies that the presence of the
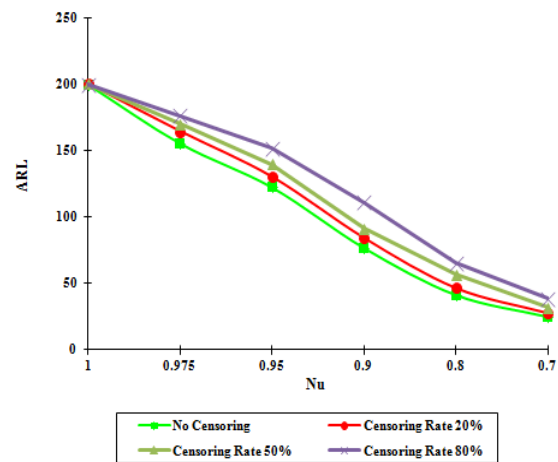

Fig. 1. ARL curves of the CUSUM control chart

censoring mechanism imposes detrimental effect on the performance of the monitoring procedure. This problem arises due to the great loss of information associated with the quality characteristic of interest. Thus, it is advisable to set the censoring level in a way not to have a large censoring proportion in the population.

5. Conclusion

In this paper, two control schemes have been proposed to effectively monitor the reliability-related quality characteristic. The monitoring procedures were adjusted for the effect of both the observed and unobserved heterogeneities caused by the incoming quality variables. The proportional hazards and the frailty models were employed and a one-sided control chart based on a probability limit along with a one-sided CUSUM control chart has been constructed. Two different scenarios including the presence and the

absence of a censoring mechanism were considered and discussed respectively. Subsequently, simulation-based studies were conducted to investigate and compare the detection power of the monitoring procedures.

The results revealed that the proposed CUSUM control chart far outweighs the competing control chart with smaller out-of-control ARL values. Moreover, the careful investigation of having low, moderate and high censoring rate indicated that the performance of the control chart deteriorates as the censoring proportion increases. Finally, it should be noted that the proposed surveillance procedures can also be applied to healthcare systems such as a hospital where patients are heterogeneous due to their measured and unmeasured unique risk factors prior to the surgery operation.

References

[1] Asadzadeh, Sh., Aghaie, A., *Improving the Product Reliability in Multistage Manufacturing and Service Operations. Quality and Reliability Engineering International,* 28, 2012, pp. 397-407.

[2] Wade, M.R., Woodall, W.H., *A Review and Analysis of Cause-selecting Control Charts. Journal of Quality Technology*, 25, 1993, pp. 161-169.

[3] Jearkpaporn, D., Montgomery, D.C., Runger, G.C., Borror, C.M., *Process Monitoring for Correlated Gamma-distributed Data using Generalized-linear Model-based Control Charts. Quality and Reliability Engineering International,* 19, 2003, pp. 477-491.

[4] Skinner, K.R., Montgomery, D.C., & Runger, G.C., *Process Monitoring for Multiple Count Data using Generalized-Linear Model-Based Control Charts. International Journal of Production Research*, 41, 2003, pp. 167-1180.

[5] Skinner, K.R., Montgomery, D.C., Runger, G.C., *Generalized-Linear Model-Based Control Charts for Discrete Semiconductor Process Data.* Quality and Reliability Engineering International , 20, 2004, pp. 777-786.

[6] Jearkpaporn, D., Borror, C.M., Runger, G.C., Montgomery, D.C., *Process Monitoring for Mean Shifts for Multiple Stage Processes.* International Journal of Production Research, 45, 2007, pp. 5547-5570.

[7] Lawless, J.F., *Statistical Models and Methods for Lifetime Data.* Wiley, New York. 2003.

[8] Xie, M., Goh, T.N., Ranjan, P., *Some Effective Control Chart Procedures for Reliability Monitoring. Reliability Engineering and System Safety*, 77, 2002, pp.143-150.

[9] Nichols, M.D., Padgett , W.J., *A Bootstrap Control Chart for Weibull Percentiles. Quality and Reliability Engineering International*, 22, 2006, pp. 141-151.

[10] Erto, P., Pallotta, G., & Park, S.H., *An Example of Data Technology Product: A Control Chart for Weibull. Processe*s. *International Statistical Review*, 76, 2008, pp. 157-166.

[11] Surucu, B., & Sazak, H.S., *Monitoring Reliability for a Three-parameter Weibull Distribution.* Reliability Engineering and System Safety, 94, 2009, pp. 503-508.

[12] Zhang, H.Y., Shamsuzzaman, M., Xie, M., Goh, T.N., *Design and Application of Exponential Chart for Monitoring Time-Between-Events Data under Random Process Shifts.* International Journal of Advanced Manufacturing Technology, 57: 2011, pp. 849-857.

[13] Steiner, S.H., MacKay, R.J., *Monitoring Processes with Highly Censored Data.* Journal of Quality Technology, 32, 2000, pp. 199-208.

[14] Steiner, S.H., MacKay, R.J., *Detecting Changes in the Mean from Censored Lifetime Data.* Frontiers in Statistical Quality Control 6, H.J. Lenz and P.T. Wilrich editors, Physica-Verlag, 2001a , pp. 275-289.

[15] Steiner, S.H., MacKay, R.J., *Monitoring Processes with Data Censored owing to Competing Risks by using Exponentially Weighted Moving Average Control Charts.* Applied Statistics, 50, 2001b , pp. 293-302.

[16] Zhang, L., Chen, G., *EWMA Charts for Monitoring the Mean of Censored Weibull Lifetimes*, Journal of Quality Technology, 36, 2004, pp. 321-328.

[17] Olteanu, D.A., *Cumulative Sum Control Charts for Censored Reliability Data. Ph.D. Thesis*, Virginia Polytechnic Institute and State University, Blacksburg, VA, 2010.

[18] Sego, L.H., Reynolds, M.R., Woodall, W.H., *Risk-Adjusted Monitoring of Survival Times.* Statistics in Medicine, 28, 2009, pp. 1386-1401.

[19] Wienke, A., *Frailty Models in Survival Analysis.* Taylor and Francis Group, Boca Raton, FL, U.S, 2011.