© 2015 IJSRST | Volume 1 | Issue 5 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

An Approach to Survival Analysis when Frailty is Evident

Jude Nwoji Oguejiofor[1], AbubakarBoyi Dalatu[2], KabiruYanusa Gorin Dikko[2]

[1]Department of Computer Science, Waziri Umaru Federal Polytechnic Birnin Kebbi, Nigeria
[2]Department of Statistics, Waziri Umaru Federal Polytechnic Birnin Kebbi, Nigeria

ABSTRACT

The problem with survival estimation using traditional life-table method alone is not only grouping of survival times, but also its insensitivity to existence of frailty in population. In this work, an approach has been proposed for the analysis of time from infection to occurrence of an event vis-à-vis the contributions of frailty along with some covariates. The approach involves the use of nonparametric product limit otherwise known as the Kaplan-Meier (KM) to estimate individual survivals and the use of frailty estimation method to estimate the hazard. In a simulation, the life-table and product limit methods were compared using the standard error estimates and plots of the empirical survivor function. While empirical estimates did not show significant differences, the product limit plot is more informative than the life table plots. The result of hazard estimation shows that event is largely caused by the baseline hazard with significant contribution from frailty and little contribution from age. There is also significant association between individual's events possibly as a result of frailty. This procedure will find wider application in survival estimation in a population-based setting.

Keywords : Survival Estimation, Empirical Survivor Function Plot, Frailty

I. INTRODUCTION

Survival data are summarized through the estimate of survival function and hazard function. Several non-parametric methods which do not require any specific assumption about the underlying distribution of survival time have been discussed by several authors during the past six decades or so Ramadurai et al(2011). Many researchers have written reports about life table. Berkson et al(1950), Gehan, (1969). Petoet al (1976) has published an outstanding review of statistical methods related to clinical trials. The development of the field that have had the most thoughtful impact on clinical trials are the Kaplan-Meier (Product limit) (1958) method for estimating the survival function. A method developed by Mantel (1966) was used to compare two survival patterns in the life table analysis. This method was used by Myers (1969) to analyze the survival experience for male patients with localized cancer of rectum diagnosed in Connecticut from 1935 – 1944 and 1945 – 1954.

2.0 Existing Work

2.1 The Empirical Survivor Function

The survivor function $S(t)$, is the probability that an individual survives for a time greater than or equal to *t*. The function can be estimated

$$\hat{S}(t) = \frac{\text{Number of Individuals with survival times } \geq \text{t}}{\text{Number of Individuals in the data set}} \qquad (2.1)$$

or $\hat{S}(t) = 1 - \hat{F}(t)$, where $\hat{F}$ *(t)* is the empirical cumulative distribution function, that is, the ratio of the total number of individuals alive at time *t* to the total number of individuals in the study. The estimated survivor function $\hat{S}(t)$ is assumed to be a constant between two adjacent death times, and so a plot of $\hat{S}(t)$ against *t* is a step-function. The function which decreases immediately after each observed survival time.

2.2 Life-Table estimate of the survivor function

The life-table estimate of the survivor function is obtained by first dividing the period of observation into series of time intervals. Suppose that in the *j*th interval, the probability of death is estimated by d$_j$/$n_j^{'}$, so that the corresponding survival probability is $(n_j^{'} - d_j)/n_j^{'}$. now consider the probability that an individual survives beyond time $t_k^{'}, k = 1, 2, ..., m,$ that is, until sometime after the start of the *k*th interval. This will be the product of the probabilities that an individual survives beyond the start of the kth interval and through each of the k – 1 preceding intervals, and so the life-table estimate of the survivor function is given by

$$S*(t) = \prod_{j=1}^{k} (\frac{n_j^{'} - d_j}{n_j^{'}})$$

$$t_k^{'} \leq t < t_{k+1}^{'}, k = 1, 2, ..., m.$$

On the assumption that censoring occur uniformly over the intervals. A graphical estimate of the survivor function will then be a step-function with constant values of the function in each time interval. However, life table method requires grouping of survival times which may lead to loss of information and possible biasness. The method only finds the estimates of the survivor function. It does not address such important aspects as randomness and covariate effects.

2.3 Kaplan Meier (product-limit) estimate of the survivor function

To obtain the Kaplan- Meier estimate, a series of time intervals is constructed, as for the life-table estimate. However, each of these intervals is designed to be such that one death time is contained in the interval, and this death time is taken to occur at the start of the interval.

We now make the assumption that the deaths of the individuals in the sample occur independently of one another. Then, the estimated survivor function at any time, *t*, in the *k*th constructed time interval from $t_{(k)}$ *to* t$_{(k+1)}$, *k* = 1, 2, …., r, where $t_{(j+1)}$is defined to be ∞, will be the estimated probability of surviving beyond $t_{(k)}$. This is actually the probability of surviving through the interval from $t_{(k)}$ to

$$t_{(k)} \leq t < t_{(k+1),} k = 1, 2, ..., r, with \hat{S}(t) = 1 \; for \, t < t_{(1)}$$

and all preceding intervals, and leads to the Kaplan-Meier estimate of the survivor function, which is given by

$$\hat{S}^{*}(t) = \prod_{j=1}^{k} \left(\frac{n_j - d_j}{n_j} \right)$$

For $t_{(k)} \leq t < t_{(k+1),} k = 1, 2, ..., r, with \hat{S}(t) = 1 \; for \, t < t_{(1)}$, and where t$_{(r+1)}$ is taken to be ∞. However, this method alone cannot be used to estimate patients survival alongside the influence of covariates and frailty at a particular event time t.

The standard error for the Kaplan-Meier estimate is given by

$$Se\left\{\hat{S}(t)\right\} \approx \hat{S}(t) \left\{ \sum_{j=1}^{k} \frac{d_j}{n_j \left(n_j - d_j\right)} \right\}^{\frac{1}{2}} \text{ for } t_{(k)} \leq t \leq t_{(k+1)} \quad (2.2)$$

(2.4)

While the standard error (se) of the life table estimate is given by

$$Se\left\{\hat{S}(t)\right\} \approx S^{*}(t) \left\{ \sum_{j=1}^{k} \frac{d_i}{n_j' \left(n_j' - d_j\right)} \right\}^{\frac{1}{2}}$$, Collet (2003)

(2.5)

II. METHODS AND MATERIAL

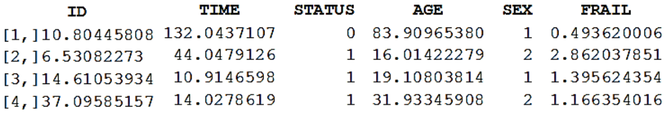
3.0 Methodology

3.1 Simulation study

A situation similar to real population-based is here simulated from which data in appendix B2 is generated; a sample of which is presented below. The covariates used in the simulation are; id, time, status, age, sex, and frail. The id is exponentially distributed with parameter 0.05, time is also exponentially distributed with parameter 0.01, status has a poisson distribution with parameter 0.76, age is exponentially distributed with parameter 0.02, sex has poisson distribution with parameter 1.74, and finally the frail having a gamma distribution with mean fixed to one (1) for identify ability (which is a constraint in frailty estimation, Rotolo et al (2012)) and variance 0.47.

The simulation now gives a survival data set from a population with exponentially distributed baseline hazard (due to the distribution of the survival times) and gamma distributed frailty condition.

Below is a sample from the 500 generated survival data.



```
       ID             TIME       STATUS        AGE        SEX       FRAIL
[1,]10.80445808 132.0437107         0  83.90965380      1  0.493620006
[2,]6.53082273   44.0479126         1  16.01422279      2  2.862037851
[3,]14.61053934  10.9146598         1  19.10803814      1  1.395624354
[4,]37.09585157  14.0278619         1  31.93345908      2  1.166354016
```

Each observation corresponds to α, the variable **id** being the patient’s code. The time (in days) from infection of an event or censoring is stored in time, while status is 1 when event has occurred and 0 for censored observations. Two other covariates may contribute: age, the age of the patient in years, and sex, being 1 for males and 2 for females. The variable frail is the frailty value for each individual which is assumed to have a gamma distribution.

The hazard of event (presented in frailty model form) will be modeled as a function of the patient’s age and sex. An R package **parfm** function is used to do the analysis.

3.2.0 Estimation of Frailty

Estimates of λ, β, ξ are obtained by maximizing the marginal log-likelihood; this can be easily done if one is able to compute higher order derivatives $L^{(q)}(□)$ of the Laplace transformed up to $q = \max\{d_1, ..., d_s\}$

3.2.1 Gamma Frailty

A gamma frailty term is a random variable $z \sim \mathrm{Gam}^*(\theta)$ with probability density function

$$f(z) = \frac{\theta^{\frac{1}{\theta}} z^{\frac{1}{\theta}-1} \exp(-u/\theta)}{\Gamma(1/\theta)}, \quad \theta > 0, \qquad (3.1)$$

where $\Gamma(□)$ is the gamma function. It corresponds to a Gamma distribution $\mathrm{Gam}(\mu, \theta)$ with μ fixed to 1 as stated earlier Collet (2003), Rotolo *et al* (2012) for identifiability. Its variance is then θ. The associated Laplace transform is given by

$$L(s) = (1+\theta s)^{-\frac{1}{\theta}}, \qquad s \geq 0,$$

and it is easy to show that, for $q \geq 1$,

$$L^{(q)}(s) = (-1)^q (1+\theta s)^{-q} \left[\prod_{l=0}^{q-1}(1-l\theta)\right] L(s).$$

(3.3)

Therefore, in equation (1.5), we have (i.e. the log-likelihood)

$$\log\left((-1)^q L^{(q)}(s)\right) = -\left(q + \frac{1}{\theta}\right)\log(1+\theta s) + \sum_{l=0}^{q-1}\log(1+l\theta)$$

, Rotolo *et al* (2012) (3.4)

For the Gamma distribution, the Kendall’s Tau Hougaar *et al* (2000), which measures the association between any two event times from the same cluster in the multivariate case, can be computed as

$$\tau = \frac{\theta}{\theta+2} \in (0,1) \qquad (3.5)$$

III. RESULT AND DISCUSSION

4.1 Plots of the Empirical Survivor Function Using the Product-Unit (K-M) Method

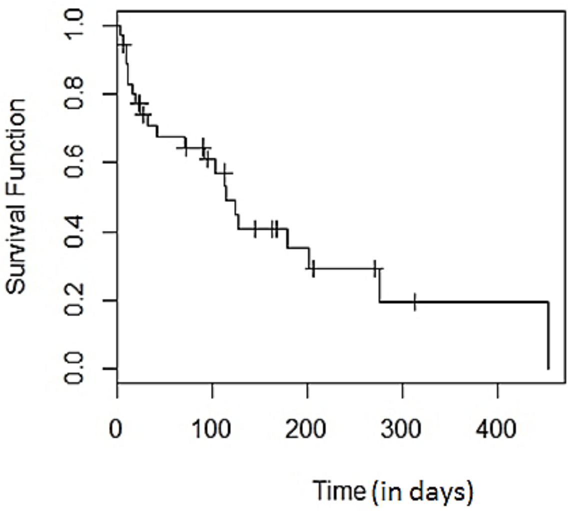


Fig. 4.1: KM Plot of Empirical Survivor Function, when sample size n = 50

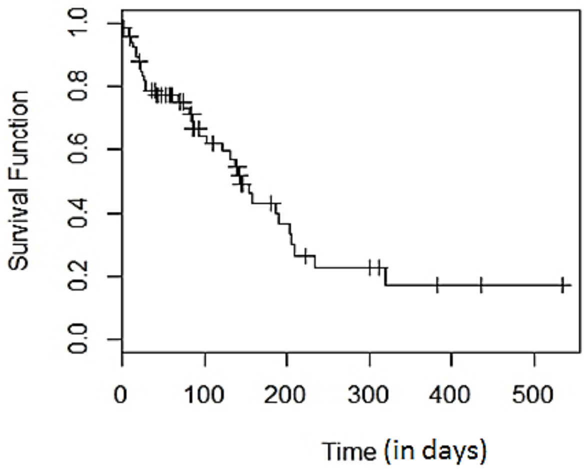


Fig. 4.2: KM Plot of Empirical Survivor Function, when sample size n = 100

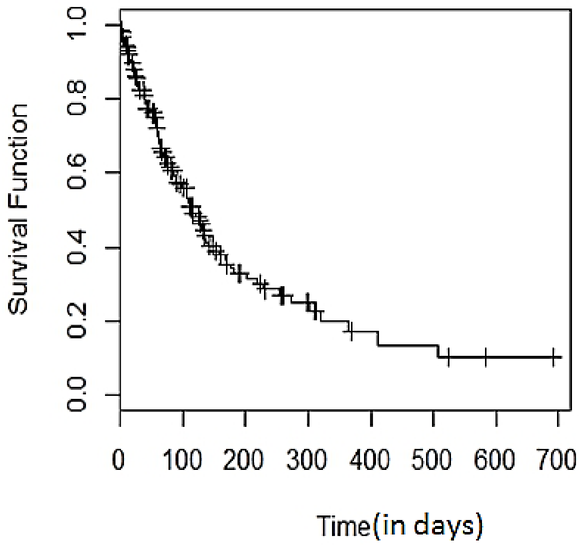


Fig. 4.3: KM Plot of Empirical Survivor Function, when sample size n = 285

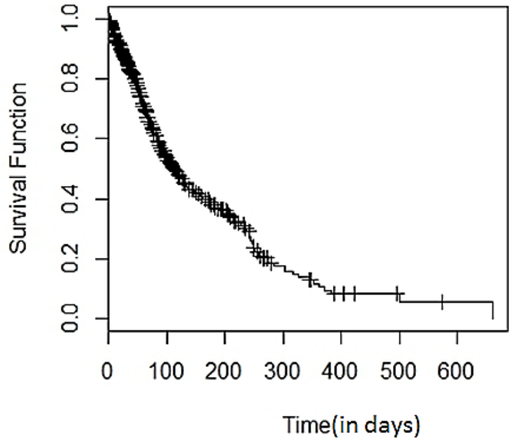


Fig. 4.4: KM Plot of Empirical Survivor Function, when sample size n = 500

4.2 Plots of the Empirical Survivor Function Using the Life-table Method

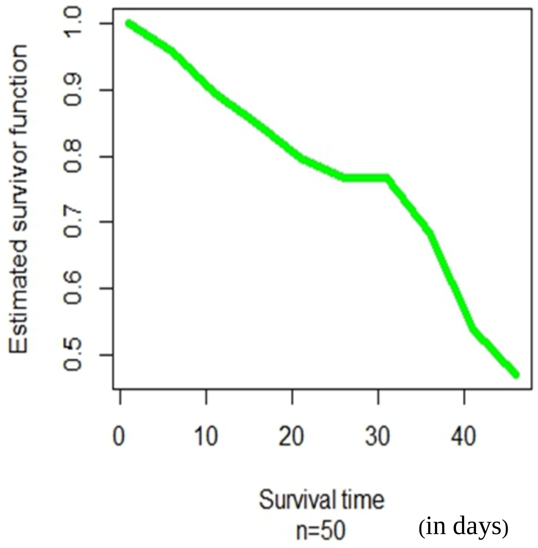


Fig 4.2a. plot of the survivor function; sample size n=50

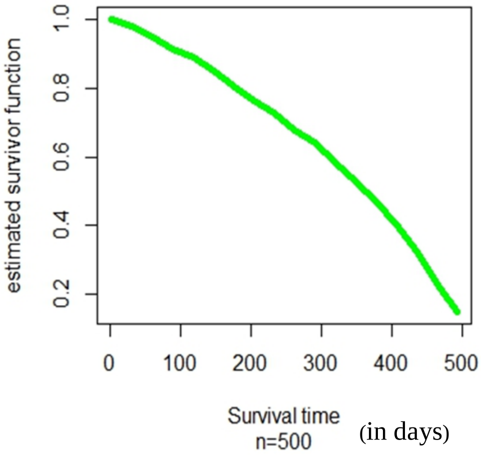


Fig 4.2b ; plot of the survivor function; Sample size n=500; interval length is 29

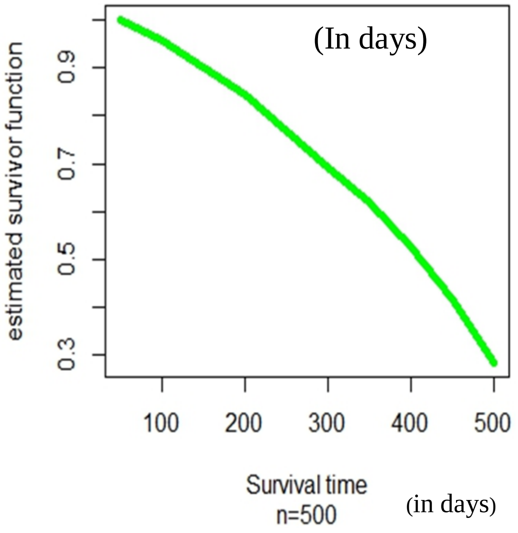


Fig 4.2c. plot of the survivor function, sample size n=500 ; interval length is 50

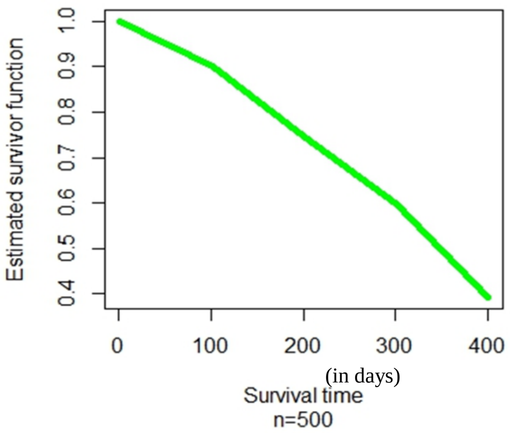


Fig 4.2d. plot of the survivor function, sample size n=500 ; interval length is 100

4.3 Discussion

When the sample size is at $n = 50, n = 100$ the empirical survivor function (esf) plot using K-M method displays good fit(fig 4.1, 4.2, 4.3, and 4.4). It shows decreasing step function behavior with survival function value (1) one at zero (0) survival time. As the survival time increases, the survivor function decreases with constant values between two adjacent death times. The plot of the function display similar behavior and pattern even with large samples as in $n = 500$ and $n = 285$. The plot of the function maintained it defined characteristics i.e. decreasing step function.

On the other hand, when sample size is not large such as n=50, the life-table plot displays the decreasing, step function behavior (fig 4.2a). But when sample size is large as n=500, the life-table plot though displays decreasing behavior, it is less informative about event times, i.e. it does not show the character of being constant between two adjacent event times. Moreover, when sample size is large, if intervals are not long enough, the life-table curve will be less meaningful.

Table 4.1: Comparison of some survival estimates and their standard error (from the results of the simulation) of life-table vs Kaplan-Meier (product limit)

	Nrisk	Surv		Se	
		lifetab	K-M	L/T	K-M
1	437	0.950	0.932	0.009	0.012
2	405	0.913	0.910	0.013	0.013
3	206	0.642	0.606	0.024	0.025
4	176	0.583	0.549	0.025	0.026

A comparative look at the results of the survivor estimates table 4.1 using the two methods (from results of the simulation) shows that when the number of patients at risk are 437, 405, 206 and 176, the respective life table survival estimates are 0.950, 0.913, 0.642 and 0.583; while that of K – M are 0.932, 0.910, 0.606 and 0.549 respectively.

The standard error, which is an essential aid for interpretation of precision, for the said values are; 0.009, 0.013, 0.024 and 0.025 for the life table estimation. While the K–M has 0.012, 0.013, 0.025 and 0.026, which does not show much difference between the survival estimates of the methods and so is the case with the standard error estimates.

4.4 Result of the frailty estimation alongside other covariates

Frailty – Gamma
Baseline hazard – Exponential
Loglikelihood – 1625.279

	Estimates	SE	p-Value
Theta	0.005	0.053	
Lambda	0.008	0.002	
Sex	-0.176	0.124	0.154
Age	-0.002	0.001	0.083

The frailty is estimated alongside age and sex. The result (4.4) shows that age could have a significant impact on the hazard of total blindness given the frailty while it is not affected by sex. The heterogeneity parameter θ i.e. the frailty is estimated at 0.005 leading to a Kendall's tau equal to 0.002, a value which indicates a very close association between two event times (in this case event between two individuals) in the cluster. The conclusion therefore is that event in this study is largely caused by the baseline hazard and frailty. The frailty has a gamma distribution evidently because the generated data is a mixture of exponentially distributed times with Poisson status or censoring indicators. The baseline hazards has exponential distribution due to the fact that the survival times (observations) are generated exponentially.

IV. CONCLUSION

The form of the estimated survivor function and the shape of its plot in life table method is sensitive to the choice of intervals. Though the survival estimates and the standard error estimates obtained using the two methods did not show much differences where the numbers at risk are equal, the empirical survival plot using Kaplan-Meier method is more informative, i.e. it tells more about event times and displays the good feature of being constant between two adjacent event times. Also in the simulated population with exponentially distributed baseline hazard and gamma frailty, the event of interest is found to be largely caused by the baseline hazard with significant contribution of frailty and the covariate age. The covariate sex was not found to have any influence on the occurrence of the event.

V. REFERENCES

[1] Ardino, V., De Angelis, R., Francis, S. and Grande, E. (2007).Methodology for Estimation of Cancer Incidence Survival and Prevalence in Italian Regions. Tumore 93: 337 – 344.

[2] Berkson, J. and Gage, R. P. (1952).Survival Curve for Cancer Patients Following Treatment. Journal of The American Statistical Association, 47: 501 – 515.

[3] Baili, P., Micheli, A., Angelis, R. D., Weir, H. K., Francis, S., Santaguilic, M., Hakulinen, T., Quaresma, M., Coleman, M. P., and the Concord working group (2008). Life Tables for Worldwide Comparison of Relative Survival for Cancer (Concord Study).Tumori, 94: 658 – 668.

[4] Collet D (2003). Modeling Survival Data in Medical Research. Chapman & Hall/CRC

[5] Cox DR (1972). “Regression Models and Lifetables” Journal of the Royal Statistical Society Series B (Methodological) 34(2) 187 – 220.

[6] Dickman, P. W. (2010). An Introduction and Some Recent Development in Statistical Methods for Population-Based Cancer Survival Analysis. Statistical Methods for Population-Based Cancer Survival Analysis. Milan.

[7] Dickman PW, Hakilinen I (2001). Population Based Cancer Survival Analysis. Chapman & Hall.

[8] Duchateau L, Janssen P (2008). The Frailty Model. Springer-Verlag

[9] Duchateau L, Janssen P, Lindsey P, Legrand C, Nguti R, Sylvester R (2002). “The Shared Frailty Model and the Power of Heterogeneity Tests in Multicenter Trials.” Computational Statistics and Data Analysis, 40(3), 603 – 620. doi:10.1016/S0167-9473(02)00057 – 9.

[10] Geham, E. H. (1969). Estimating Survival Functions From the Life Table. Journal of Chronic Diseases, 21: 629 – 694.

[11] Hanagal D (2011). Modelling Survival Data using frailty models. Chapman & Hall/CRC Press. Taylor and Francis Group, LCC.

[12] Hougaard P (1995). “Frailty Models for Survival Data.” Lifetime Data Analysis,1(3), 255 – 273. doi:10.1007/BF00985760.

[13] Hougaard P (2000). Analysis of Multivariate Survival Data. Springer-Verlag.

[14] Kaplan, E. L., Meier, P. (1958). Nonparametric Estimation From Incomplete Observations. J Am. Stat. Association.55: 457 – 81.

[15] Mantel N (1966). Evaluation of Survival Data and Two New Rank Order Statistics Arising in its Consideration. Cancer Chemotherapy Report, 50, 163 – 170.

[16] Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J. and Smith, P. G. (1976). Design and Analysis of Randomised Critical Trials Requiring Prolonged Observation of Each Patient. Introduction and Design. British Journal of Cancer. 34: 585 – 612.

[17] Ramadurai, M. and Ponnuraja, C. (2011).Nonparametric Estimation of the Survival Probability of Children Affected by TB Meningitis. Journal of Arts and Science and Commerce. 2231 – 4172.

[18] Ravichandran, K., AlHandam, N., Al Dyab, A. (2006). Asian Pacific Journal of Cancer Prevention. Vol. 6.

[19] Rotolo F, Munda M (2012). Parfm; Parametric Frailty Models. R package version 0.66, URL http://CRAN.R-project.org/package=parfm

[20] Therneau TM, Grambsch PM (2000). Modeling Survival Data:Extending the Cox Model. Springer-Verlag.

[21] Vaupel JW, Manton KG, Stallard E (1979). “The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality.” Demography, 16(3), 439 – 454.

[22] Wang ST, Klein JP, Moeschberger ML (1995). “Semi-parametric Estimation of Covariate Effects Using the Positive Stable Frailty Model.” Applied stochastic models and data analysis, 11(2), 121 – 133.

[23] Wienke A (2010). Frailty Models in Survival Analysis. Chapman & Hall CRC Biostatistics Series. Taylor and Francis. doi: 10.1201/9781420073911.