Camilleri, L. et al. (2022). *Xjenza Online*, Spring Issue: 23–29.

Xjenza Online: Science Journal of the Malta Chamber of Scientists
www.xjenza.org
DOI: 10.7423/XJENZA.2022.1.03

Research Article

Identifying Risk Factors of Aortic Valve Replacement Using Frailty Models

Camilleri, L.[1*], **Grech, L.**[1], **Manche, A.**[2]
[1] *Department of Statistics and Operations Research, University of Malta, Msida, Malta*
[2] *Department of Cardio-thoracic surgery, Mater Dei Hospital, Msida, Malta*

Abstract. Traditional survival modeling techniques, including the Kaplan Meier estimator, Cox regression and parametric survival models assume a fairly homogeneous population, where variation in survival durations can be explained by a small number of observed explanatory variables. However, in the presence of heterogeneity, frailty models are more appropriate to model survival data by introducing random effects that account for the variability generated from unobserved covariates. This paper presents two types of frailty models. The unshared frailty model assumes that different individuals have distinct frailties, while the shared frailty model assumes that the population can be divided into clusters, where members in the same cluster share the same frailty. Due to their nice mathematical properties, the Gamma and the Inverse Gaussian distributions are the most popular choices for the frailty distribution.
These survival models are fitted to a data set using the facilities of STATA. The participants are patients who underwent an aortic valve replacement procedure at a Maltese hospital between 2003 and 2019. The dependent variable is the duration till death or till censored and the eleven predictors provide information about the patients' health condition; surgery operative procedures; and duration of convalesce period. Moreover, in shared frailty models the patients are clustered by their diabetic condition since it is known that diabetic patients are more at risk of dying following aortic surgery.

Keywords: Shared and Unshared Frailty models, Gamma and Inverse Gaussian Distributions, Aortic Valve Replacement

1 Introduction

Survival analysis is a useful statistical tool for problems that deal with survival data, where the outcome variable is the duration for a certain event to occur. Initially survival analysis was used to model survival durations of patients undergoing surgical treatment or rehabilitation therapy; however, this statistical procedure has been extended to several research areas in the last three decades. Survival models are used to estimate the duration till failure of mechanical and electrical devices in engineering, relapse duration to alcohol and drug addiction in criminology. These models are used to evaluate product reliability in market research, measure viability of therapies, instruments and techniques in medicine, estimate life expectancy in demography, model marriage durations before separation/divorce in sociology, evaluate profitability of investment schemes in finance, amongst other applications.

The non-parametric Kaplan-Meier and Nelson-Aalen estimators, the semi-parametric Cox regression models, and the parametric survival models all assume that the members within a population are homogeneous with similar hazards. It is known in survival data, that unobserved heterogeneity exists between members, which cannot be explained directly by observable covariates. This unexplained variability is very common in epidemiological, medical and rehabilitation applications. For example, a specific treatment can have diverse effect on the recovery duration of patients, and a rehabilitation programme can have different impact on the relapse duration of drug abusers. To address this limitation, Vaupel et al. (1979) introduced frailty survival models, where a random effect (frailty) is introduced in the model to have a multiplicative effect on the hazard function of an individual or group of individuals. Frailty models allowed analysts to

Correspondence to: Liberato Camilleri (liberato.camilleri@um.edu.mt)

account for unobserved heterogeneity, which effectively reduces the possibility of inaccurate parameter estimates and biased standard errors. These models assume that the weaker individuals are more likely to succumb earlier than the stronger members. The univariate frailty model proposed by Vaupel et al. (1979) was further extended by Clayton (1978) who applied the technique on multivariate data related to chronic disease incidence in families.

Two popular frailty distributions in survival models are the Gamma and Inverse Gaussian distributions due their simpler mathematical properties. Hougaard (1984) showed that for a Gamma frailty distribution the relative heterogeneity remains constant, while the inverse Gaussian frailty assumes that this heterogeneity decreases with time. Both Vaupel et al. (1979) and Hougaard (1986) showed that although different individuals may have similar physical health conditions, some may be more susceptible to different threats and frailties. Hougaard (1986) argued that the choice between using an Inverse Gaussian or a Gamma distributed frailty depends entirely on the frailty instability of an individual. While frailty tends to be steady during an individual's life, it tends to deteriorate in later stages. Subsequently, many authors endorsed this frailty concept of a concealed random effect when analyzing survival data.

One of the objectives of frailty survival models is to estimate the variance of unobserved risk among different individuals. There are two approaches how frailty is distributed in the data by using shared and unshared frailty models. Unshared frailty models assume that different individuals have distinct frailties. For example, the recurrence times of machine malfunctions after being fixed is investigated by using unshared frailty models. On the other hand, shared frailty models assume that individuals within a group share frailty; however, this frailty may vary between groups. For example, some individuals may be more susceptible than others to be diagnosed with cancer or heart disease because of some unknown genetic condition, or some countries are more likely to engage in war than others for unknown reasons.

When intragroup correlation exists, shared frailty models may be more appealing than unshared frailty models. The seminal contributions of Clayton (1978) and Clayton et al. (1985) were fundamental in the development of shared frailty models. Identifiability issues related to frailty survival models were first addressed by Elbers et al. (1982), while theoretical proofs were given by Heckman et al. (1984). Other extensions focused on methods of how to measure correlation in bivariate survival data using an arbitrary parametric hazard function. Hougaard (1986) assumed Weibull individual hazards when fitting shared frailty models, while Whitmore et al. (1991) applied an inverse Gaussian shared frailty model by assuming constant individual hazards.

2 Theory of Unshared Frailty Models

The seminal work of Clayton et al. (1985), among other authors, highlighted the utility of frailty models and stressed the benefit of adding frailty to account for unobserved heterogeneity. As described in the introduction, there are two types of frailty models to analyze survival data in the presence of unobserved heterogeneity. In unshared frailty models, the frailty is introduced at the observation level as an unobservable multiplicative effect, on the baseline hazard function $h_0(t)$, given by:

$$h(t|\alpha) = \alpha h_0(t) \tag{1}$$

In this context, α is a non-negative random mixture variable where $E(\alpha) = 1$ and $\text{var}(\alpha) = \sigma^2$. When σ^2 is small, the values of α are located close to 1; however the values of α are more dispersed when σ^2 is large, inducing larger heterogeneity in the individual hazards $\alpha h_0(t)$.

Let $S(t|a)$ denote the survival function of a life conditional on the frailty α and let $\int_0^t h_0(s)ds = M_0(t)$ then

$$S(t|\alpha) = e^{-\int_0^t h(s|\alpha)ds} = e^{-\alpha \int_0^t h_0(s)ds} = e^{-\alpha M_0(t)} \tag{2}$$

If observed covariates, denoted by an *(p x n)* matrix $\mathbf{X}$, are available then the hazard is proportional to the baseline hazard. Moreover, the constant of proportionality is the term $\exp(\boldsymbol{\beta}'\mathbf{X})$, where $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_p)$ and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$ is the vector of regression parameters. So model (1) becomes:

$$h(t|\mathbf{X}, \alpha) = \alpha h_0(t) \exp(\boldsymbol{\beta}'\boldsymbol{X}) \tag{3}$$

The two distributions that are normally used for the probability density function $f(\alpha)$, of α are the gamma and inverse Gaussian distributions.

Given the simple Laplace transform of the Gamma distribution $\Gamma(\kappa, \lambda)$, it is easy to derive the closed-form expressions of the survival and hazard functions. The exponential distribution is a special case of the Gamma distribution when the shape parameter $\kappa = 1$. If α has a Gamma distribution and $\alpha > 0$, $\lambda > 0$, $\kappa > 0$ its probability density function is given by:

$$f(\alpha) = \frac{\lambda^\kappa}{\Gamma(\kappa)} \alpha^{\kappa-1} e^{-\lambda\alpha} \tag{4}$$

By setting $\kappa = \lambda = 1/\sigma^2$ ensures that the model is identifiable and $E(\alpha) = 1$ and $\text{var}(\alpha) = \sigma^2$. Using Laplace transform, Wienke (2010) derives the unconditional survival and hazard functions, which are given by:

$$S(t) = \mathbf{L}[M_0(t)] = \frac{1}{[1+\sigma^2 M_0(t)]^{1/\sigma^2}} \tag{5}$$

$$h(t) = -h_0(t)\frac{\mathbf{L}[M_0(t)]}{\mathbf{L}'[M_0(t)]} = \frac{h_0(t)}{1+\sigma^2 M_0(t)} \tag{6}$$

Moreover, Wienke (2010) shows that if observed covariates $\mathbf{x}_i$ are available for life i then the mean frailty and frailty variance for a life dying beyond time t are given by:

$$E(\alpha|\mathbf{X}, T > t) = \frac{1}{1+\sigma^2 M_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{X})} \tag{7}$$

$$\text{var}(\alpha|\mathbf{X}, T > t) = \frac{\sigma^2}{[1+\sigma^2 M_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{X})]^2} \tag{8}$$

The Inverse Gaussian distribution is also considered as a frailty distribution because similar to the Gamma distribution, simple closed-form expressions exist for the unconditional survival and hazard functions. If α has an Inverse Gaussian distribution and $\alpha > 0$, $\lambda > 0$, $\mu > 0$ its probability density function is given by:

$$f(\alpha) = \frac{\sqrt{\lambda}}{\sqrt{2\pi\alpha^3}}\exp\left[-\frac{\lambda(\alpha-\mu)^2}{2\mu^2\alpha}\right] \tag{9}$$

Setting $\mu = 1$ and $\lambda = 1/\sigma^2$ ensures that the model is identifiable, where $E(\alpha) = 1$ and $\text{var}(\alpha) = \sigma^2$. The unconditional density function, the unconditional survival and hazard functions are given by:

$$S(t) = \exp\left(\frac{1-\sqrt{1+2\sigma^2 M_0(t)}}{\sigma^2}\right) \tag{10}$$

$$h(t) = \frac{h_0(t)}{\sqrt{1+2\sigma^2 M_0(t)}} \tag{11}$$

If observed covariates $\mathbf{x}_i$ are available for life i then the mean frailty and frailty variance for a life dying beyond time t are given by:

$$E(\alpha|\mathbf{X}, T > t) = \frac{1}{\sqrt{1+\sigma^2 M_0(t)\exp(\boldsymbol{\beta'}\mathbf{X})}} \tag{12}$$

$$\text{var}(\alpha|\mathbf{X}, T > t) = \frac{\sigma^2}{[1+\sigma^2 M_0(t)\exp(\boldsymbol{\beta'}\mathbf{X})]^2} \tag{13}$$

3 Theory of Shared frailty Models

A generalization of the unshared frailty model is the shared frailty model, where the frailty is assumed to be group-specific. Basically shared frailty arises when the heterogeneity impact is common among individuals within a group, yet each set has a distinct random effect, which in turn causes frailties to be interrelated.
Suppose there exist n groups and that group i comprises n_i observations associated with the unobserved frailty α_i for $1 \leq i \leq n$. Their hazard functions are given by:

$$h(t|\alpha_i) = \alpha_i h_0(t) \tag{14}$$

Let $S(t|\alpha_i)$ denote the survival function of a life conditional on the frailty α_i and let $\int_0^{t_{ij}} h_0(s)ds = M_0(t_{ij})$ then

$$S(t_{i1}, ..., t_{in_i}|\alpha_i) = \exp[-\alpha_i \sum_{j=1}^{n_j} M_0(t_{ij})] \tag{15}$$

If observed covariates $\mathbf{X}_i$ for $1 \leq i \leq n$ are available then the hazard is proportional to the baseline hazard, where the constant of proportionality is the exponential term $\exp(\boldsymbol{\beta'}\mathbf{X})$. Assuming that the survival times in group i are independent, then model (16) becomes:

$$h(t|\mathbf{X}_i, \alpha_i) = \alpha_i h_0(t)\exp(\boldsymbol{\beta'}\mathbf{X}_i) \tag{16}$$

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$ is the vector of regression parameters and $\mathbf{X}_i = (\mathbf{x}_{i1}, ..., \mathbf{x}_{in_i})$ is the covariate matrix of the members in the i^{th} cluster. The conditional survival function on frailty α_i which is shared by all individuals in group i is given by:

$$S(t_{i1}, ..., t_{in_i}|\mathbf{X}_i, \alpha_i) = \exp\left[-\alpha_i \sum_{j=1}^{n_i} M_0(t_{ij})e^{\boldsymbol{\beta}'\mathbf{x}_{ij}}\right] \tag{17}$$

where $M_0(t_{ij})$ is the cumulative baseline hazard function of the j^{th} members in the i^{th} cluster. Averaging (17) with respect to the frailty α_i gives the survival function for the i^{th} cluster,

$$S(t_{i1}, ..., t_{in_i}|\mathbf{X}_i) = \mathbf{L}\left(\sum_{j=1}^{n_i} M_0(t_{ij})\, e^{\boldsymbol{\beta}'\mathbf{x}_{ij}}\right) \tag{18}$$

where $\mathbf{L}$ denotes the Laplace transform of the frailty variable. The univariate unconditional survival function can be expressed by means of the Laplace transform:

$$S(t_{ij}|\mathbf{X}_i) = \mathbf{L}(M_0(t_{ij})e^{\boldsymbol{\beta}'\mathbf{x}_{ij}}) \tag{19}$$

$$M_0(t_{ij})e^{\boldsymbol{\beta}'\mathbf{x}_{ij}} = \mathbf{L}^{-1}[S(t_{ij}|\mathbf{X}_{ij})] \tag{20}$$

where $\mathbf{L}^{-1}$ is the inverse of the Laplace transform $\mathbf{L}$.
The Gamma and Inverse Gaussian frailty models are often used mainly for their nice properties, particularly their

simple Laplace transform. Assuming a Gamma frailty distribution with $E(\alpha) = 1$ and $\text{var}(\alpha) = \sigma^2$, the survival function for the i^{th} cluster is obtained by substituting (5) in (18).

$$S(t_{j1}, ..., t_{in_i}|\mathbf{X}_i) = \left(1 + \sigma^2 \sum_{j=1}^{n_i} M_0(t_{ij}) e^{\boldsymbol{\beta}' \mathbf{x}_{ij}}\right)^{-1/\sigma^2} \tag{21}$$

Moreover, by assuming an inverse Gaussian frailty distribution with $E(\alpha) = 1$ and $\text{var}(\alpha) = \sigma^2$, the survival function for the i^{th} cluster is obtained by substituting (10) in (18).

$$S(t_{i1}, ..., tin_i|\mathbf{X}_i) = \exp\left(\frac{1 - \sqrt{1 + 2\sigma^2 M_0(t) e^{\boldsymbol{\beta}' \mathbf{x}_{ij}}}}{\sigma^2}\right) \tag{22}$$

Popular choices for the baseline hazard include the exponential distribution for constant hazard; the Lognormal and Loglogistic distributions for humped hazards and the Weibull and Gompertz distributions for monotonic increasing hazards.

4 Application

The dataset consists of 480 patients who underwent an aortic valve replacement at the cardiothoracic centre in a Maltese hospital. This data was collected by a cardio-vascular surgeon over a period of 16 years, ranging between 2003 and 2019. Patients who had missing information were excluded from the dataset. Most of the patients who underwent this treatment were aged over 60 years, which is expected since the prevalence of heart disease increases drastically with age. After surgery, all patients had follow-up appointments. The time of death of patients who died before the end of the investigation period (2019) was recorded and the survival duration was computed. Patients who were still alive after the end of the investigation period were right censored.

The dataset includes a number of patient-related explanatory variables, together with other information related to the patients' health conditions in pre-operative and the post-operative periods. In this study, the dependent variable is *Time*, which measures the survival duration between the surgery and the time of death/end of the investigation period. *Status* indicates whether the patient is dead or alive at the end of the investigation period and will be used as a censoring variable. *BMI* provides the ratio of the patient's weight (kilograms) to the patient's height squared (m^2). The *Parsonnet* score measures the risk of death of a patient after undergoing heart surgery, where the larger the score the higher is the risk. *HDU* and *ITU* record the duration (days) of the patient's recovery in the High Dependency Unit and the Intensive Therapy Unit respectively. *Hypertension* indicates the patient's presence or absence of high blood pressure and *Transfusion* indicates whether the patient required/not required blood transfusion directly from another individual. *Ventilation* measures the duration (hours) that the patient spent on a life-assisting mechanical ventilator following the surgery. *Creatinine* indicates the presence/absence of waste product in the blood that normally passes through the kidneys and is eliminated through urine. *Dialysis* indicates whether or not the patient has kidney problems and is receiving dialysis treatment. *Blood* measures the blood volume (millilitres) that was provided to the patient during or after surgery and *IABP* indicates whether or not the patient required an intra-aortic balloon pump during heart surgery. *Diabetes* indicates whether the patient is diabetic or normal and will be used as a clustering variable in shared frailty models.

Of the 480 patients participating in the study, 22.8% died before the end of the investigation period and the rest were right censored. The mean Parsonnet score (6.24) indicates that the risk of mortality is fair and that there is a 5% predicted mortality rate. All the patients undergoing heart surgery spend one night in ITU and are retained in this unit if health condition is critical. If the patients' health condition is not life-threatening, they are transferred to the HDU for a convalescence period. The mean duration of patients requiring support of a ventilator was 5.24 hours and the mean blood volume transfused was 565.66 millilitres; however, these values were considerably larger for high risk patients. The mean BMI (29.44 kg/m^2) is larger than average indicating that the majority of the patients were overweight or obese. 29.7% of the patients were diabetic; 50.9% suffered from high blood pressure; 1.7% were on dialysis, 2.6% required the use of an intra-aortic balloon pump during surgery; 35% required blood transfusion and 3.2% of the patients had high levels of creatinine.

5 Results

It is known that the Gompertz distribution provides a remarkable close fit to adult mortality in contemporary developed countries. For this reason, all fitted models were implemented assuming a Gompertz baseline hazard function given by:

$$h_0(t) = \lambda_j e^{\gamma t_j} \tag{23}$$

where $\lambda_j = \exp(\beta_0 + ... + \beta_p x_p)$ and γ is an ancillary parameter. Table 1 displays the hazard ratios, standard errors and p-values of the non-frailty model. Since a number of the predictors were not significant, a backward procedure

was used to identify the parsimonious model.

Parameter	H.R	S.E	Z	P> |z|
Constant	0.000	0.000	-24.4	0.000
BMI	1.005	0.012	0.42	0.675
Hypertension	0.865	0.158	-0.79	0.428
Parsonnet	1.096	0.010	9.57	0.000
ITU	0.799	0.152	-1.17	0.241
HDU	1.056	0.012	4.81	0.000
Ventilation	0.995	0.011	-0.46	0.647
Blood	0.999	0.001	-0.09	0.924
IABP	1.423	0.489	1.03	0.305
Dialysis	3.523	1.410	3.15	0.002
Creatinie	1.539	0.411	1.62	0.106
Transfusion	0.978	0.317	-0.16	0.874
γ	0.000	0.000	9.47	0.000
Log-Likelihood				-914.31

Table 1: Non-frailty model

Table 2 displays the hazard ratios, standard errors and *p*-values of the parsimonious non-frailty model. This survival model assuming a Gompertz baseline hazard function identifies three significant predictors of survival duration, where the Parsonnet score is the best predictor, followed by recovery duration in HDU and dialysis treatment. The hazard of death for patients on dialysis is 2.878 times than those who have no kidney problems. Moreover, for every additional treatment day in the High Dependency Unit, the hazard of death increases by 4.9% and for every 1 unit increase in the Parsonnet score the risk of death increases by 9.1%, given that other effects are kept constant.

Parameter	H.R	S.E	Z	P> |z|
Constant	0.000	0.000	-75.45	0.000
Parsonnet	1.091	0.010	9.49	0.000
HDU	1.049	0.010	5.00	0.000
Dialysis	2.878	1.108	2.75	0.006
γ	0.000	0.000	10.19	0.000
Log-Lokelihood				-919.40

Table 2: Parsimonious non-frailty model

Table 3 and table 4 display the hazard ratios and corresponding standard errors of the parsimonious unshared frailty models assuming a Gamma and Inverse Gaussian distribution and a Gompertz baseline hazard function. The likelihood ratio statistics (3.72 and 2.91) yield p-values (0.027 and 0.044), which are less than the 0.05 level of significance. This implies that the frailty variance is significantly positive.

Parameter	H.R	S.E	Z	P> |z|
Constant	0.000	0.000	-51.07	0.000
Parsonnet	1.113	0.018	6.68	0.000
HDU	1.060	0.013	4.84	0.000
Dialysis	3.346	1.704	2.37	0.018
γ	0.000	0.000	6.47	0.000
Log-Lokelihood				-917.54

Table 3: Parsimonious unshared Gamma frailty model
LR test of $\sigma^2 \text{var}(\alpha) = 0$: Chibar2(01) = 3.72, p = 0.027

Parameter	H.R	S.E	Z	P> |z|
Constant	0.000	0.000	-50.16	0.000
Parsonnet	1.111	0.019	6.06	0.000
HDU	1.059	0.013	4.84	0.000
Dialysis	3.347	1.660	2.44	0.015
γ	0.000	0.000	5.33	0.000
Log-Lokelihood				-917.95

Table 4: Parsimonious unshared Inverse Gaussian frailty model
LR test of $\sigma^2 \text{var}(\alpha) = 0$: Chibar2(01) = 2.91, p = 0.044

Table 4 and table 5 display the hazard ratios and corresponding standard errors of the parsimonious shared frailty models assuming a Gamma and Inverse Gaussian distribution and a Gompertz baseline hazard function. The likelihood ratio statistics (approx. 0) yield p-values (approx. 1), which exceed the 0.05 level of significance. This implies that frailty vanishes completely when the patients are grouped by their diabetic condition.

Parameter	H.R	S.E	Z	P> |z|
Constant	0.000	0.000	-75.45	0.000
Parsonnet	1.091	0.010	9.49	0.000
HDU	1.049	0.010	5.0	0.000
Dialysis	2.879	1.108	2.75	0.006
γ	0.000	0.000	10.19	0.000
Log-Lokelihood				-919.40

Table 5: Parsimonious shared Gamma frailty model
LR test of $\sigma^2 \text{var}(\alpha) = 0$: Chibar2(01) = 0.00, p = 1.000

Parameter	H.R	S.E	Z	P> |z|
Constant	0.000	0.000	-75.45	0.000
Parsonnet	1.091	0.010	9.49	0.000
HDU	1.049	0.010	5.0	0.000
Dialysis	2.879	1.107	2.75	0.006
γ	0.000	0.000	10.19	0.000
Log-Lokelihood				-919.40

Table 6: Parsimonious shared Inverse Gamma frailty model
LR test of $\sigma^2 \text{var}(\alpha) = 0$: Chibar2(01) = 0.00, p = 1.000

6 Conclusion and Recommendations

All five models identify three significant predictors and all models highlight that the hazard of death is higher for patients who are on dialysis and increases with an increase in the Parsonnet score and an increase in the treatment duration in HDU. Table 7 shows the AIC and BIC values of the five fitted models. The fact that these values vary marginally between the five model fits indicate, that for this data set, shared and unshared frailty models did not provide a considerably improvement in goodness of fit compared to non-frailty models. This is a clear example where a more complex model does not always provide more predictive power than a simpler model. However, addressing heterogeneity due to unobserved covariates is highly recommended to obtain robust estimates of the hazard ratios and standard errors.

Frailty Distribution	AIC	BIC
No frailty assumed	1848.8	1869.7
Unshared Gamma	1847.1	1872.1
Unshared Inverse Gaussian	1847.9	1872.9
Shared Gamma	1848.8	1875.8
Shared Inverse Gaussian	1848.8	1875.8

Table 7: AIC and BIC measures for goodness of fit

The Gamma and Inverse Gaussian distributions have been used extensively as frailty distributions, mainly because of their simple Laplace transform. Another suggestion is to use the log-normal distribution for frailty, particularly when the random effects are assumed to be normally distributed. This allows more flexibility especially in modelling multivariate correlation structures. The development of new statistical software in enhancing computational power and the development of user-friendly estimation procedures such as the MCMC adaptive quadrature techniques make it possible to accommodate normal distributed random effects.

Another recommendation is to use semi-parametric frailty models, which extends the proportional hazards Cox model by introducing random effects to account for unobserved heterogeneity in the data. This semi-parametric approach is available both for shared and unshared Gamma frailty. Two approaches can be used to estimate parameters in semi-parametric frailty models. The first approach is the EM algorithm, which iterates between the Expectation and the Maximization steps. The second approach is the penalized partial likelihood (PPL) method, where estimation is based on Laplace approximation of the likelihood function.

Another recommendation is to use accelerated failure time (AFT) survival models instead of proportional hazards (PH) models. AFT models assume that the effect of a covariate is to decelerate or accelerate the survival outcome by a constant. This differs from PH models which assume that the effect of a covariate multiplies the hazard by a constant. Survival distributions that accommodate AFT survival models in STATA are the Exponential, Weibull, Log-normal and Log-logistic distributions.

Another recommendation is to use correlated frailty models, which are mixture models that assume that the frailty for each individual is random. These models assume that event times in a cluster are independent, given the frailties of the individuals. In other words, frailty variables are allowed to be correlated but may not necessarily be common to all individuals in a cluster, implying dependence between event times. The shared frailty models are special cases of the correlated frailty models by setting the correlations between the frailties to be equal to 1.

Another recommendation is to use multilevel survival models. Traditional survival models assume that individuals are independent of each other. However, individuals who are nested within higher level structures are more likely to have correlated outcomes, thus violating the assumption of independence. The homogeneity within clusters may be caused by cluster characteristics that are difficult to measure, such as practices that vary between hospitals. However, through multilevel survival models it is possible to accommodate the multilevel structure while accounting for the grouping of lower level units within higher level units.

One final recommendation is to use Copula models to model clustered data; however a requirement for these models is that the sample size for each cluster is the same. Copula models are fitted by using a two-stage procedure. The marginal survival functions are estimated, in the first step, ignoring the groupings within the data. This can be carried out by using a parametric, semi-parametric or non-parametric approach. The copula parameters are then estimated in the second step. Alternatively, one can use a one-step procedure by maximizing the likelihood.

References

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of statistics*, 701–726.

Akaike, H. (1974). A innovative look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Camilleri, L., Caruana, R. & Manche, A. (2017). Modeling survival times using frailty models, 428–432.

Clayton, D. G. (1978). A model for association in bivariate life tables and its applications in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, *65*, 141–151.

Clayton, D. & Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model. *Royal Statistical Society*, *48*, 82–117.

Congdon, P. (1995). Modelling frailty in area mortality. *Statistics in Medicine*, *14*, 1859–1874.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society*, *34*, 187–220.

Dempster, A. P., Laird, N. M. & Rubin, D. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, *39*(1), 1–38.

Duchateau, L. & Janssen, P. (2008). *The frailty model*. Springer.

Duchateau, P. & Janssen, P. (2007). *The frailty model*. Springer.

Elbers, C. & Ridder, G. (1982). True and spurious duration dependence: The identifiability of the proportional hazard model. *The Review of Economic Studies*, *49*, 403–409.

Flinn, C. & Heckman, J. (1982). New methods for analyzing structural models of labour force dynamics. *Journal of Econometrics*, *18*, 115–168.

Gill, R. & Schumacher, M. (1987). A test of proportional hazards assumptions. *Biometrika*, *74*(2), 289–300.

Gupta, R. & Gupta, R. D. (2009). General frailty model and stochastic orderings. *Journal of Statistical Planning and Inference*, *139*, 3277–3287.

Gutierrez, R. G. (2002). Parametric frailty and shared frailty models. *The Stata Journal*, *2*(1), 22–44.

Hanagal, D. D. (2011). *Modeling survival data using frailty models*. Chapman; Hall/CRC.

Heckman, J. & Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *22*(2), 271–320.

Horowitz, J. L. (1999). Semiparametric estimation of a proportional hazard model with unobserved hetrogeneity. *Econometrica*, *67*(5), 1001–1028.

Hougaard, P. (1984). Life tables for heterogeneous populations distributions describing the heterogeneity. *Biometrika*, *71*(1), 75–83.

Hougaard, P. (1986). Survival model for heterogeneous populations derived from positive stable distributions. *Biometrika*, *71*, 75–83.

Kalbfleisch, J. D. (1974). Some efficiency calculations for survival distributions. *Biometrika*, *61*, 31–38.

Kaplan, E. L. & Meier, P. (1958). Nonparametric estimation from incompleteo obsevations. *Journal of the American Statistical Association*, *53*, 457–481.

Lam, K., Fong, D. & Tang, O. (2005). Estimating the proportion of cured patients in a censored sample. *Statistics in Medicine*, *24*, 1865–1879.

Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica*, *47*, 939–956.

Nelson, W. (1972). Theory and applications of hazards plotting for censored failure data. *Technometrics*, *14*(4), 945–966.

Ripatti, S. & Palmgren, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, *56*, 1016–1022.

Schwarz, G. (1978). Estimating the dimension of models. *The Annals of Statistics*, *6*, 461–464.

Therneau, T. M., Grambsch, P. M. & Pankratz, V. S. (2003). Penalized survival models and frailty. *Computational and Graphical Statistics Journal*, *12*(1), 156–175.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society*, *38*, 290–295.

Vaupel, J. W. (1988). Inherited frailty and longevity. *Demography*, *25*(2), 277–287.

Vaupel, J. W., Manton, K. G. & Stallard, E. (1979). The impact in individual frailty on the dynamics of mortality. *Demography*, *16*(3), 439–454.

Vaupel, J. W. & Yashin, A. I. (1985). Heterogeneity's ruses: Some surprising effects of selection on population dynamics. *The American Statistician*, *39*, 176–185.

Whitmore, G. A. & Lee, M. L. T. (1991). A multivariate survival distribution generated by an inverse Gaussian mixture of exponentials. *Technometrics*, *33*, 39–50.

Wienke, A. (2010). *Frailty models in survival analysis*. Chapman; Hall/CRC.

Xue, X. & Brookmeyer, R. (1996). Bivariate frailty model for the analysis of multivariate survival time. *Lifetime Data Analysis*, *2*, 277–290.