The Open Public Health Journal

Content list available at: https://openpublichealthjournal.com

RESEARCH ARTICLE

The Importance of the Frailty Effect In Survival Models: For Multidrug-resistant Tuberculosis Data

Sizwe Vincent Mbona[1,*], Henry Mwambi[2] and Shaun Ramroop[2]

[1]*Department of Statistics, Durban University of Technology, Durban, 4001, South Africa*
[2]*School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, King Edward Avenue, Pietermaritzburg, 3209, South Africa*

Abstract:

Background:

Frailty models have been proposed to analyse survival data, considering unobserved covariates (frailty effects). In a shared frailty model, frailties are common (or shared) amongst groups of individuals and are randomly distributed across groups.

Objective:

In this paper, the authors compared the semi-parametric model to shared frailty models by studying the time-to-death of patients with multidrug-resistant tuberculosis (MDR-TB).

Methods:

Secondary data from 1 542 multidrug-resistant tuberculosis patients were used in this study. STATA software was used to analyse frailty models via the *streg* command.

Results:

Of 1 542 patients diagnosed with MDR-TB, 245 (15.9%) died during the study period; 77 (5.0%) had treatment failure; 334 (21.7%) defaulted; 213 (13.8%) completed treatment; 651 (42.2%) were cured of MRD-TB; and 22 (1.4%) were transferred out. The results showed that 797 (51.7%) were females, and the majority were aged 18 – 30 and 31 – 40 years (35.5% and 35.7% respectively). Most of the patients (71.3%) were HIV-positive. The results also showed that most patients (95.7%) had no previous MDR-TB episodes, and 792 (51.4%) had no co-morbidities. The estimate of the variance for the frailty term in the Weibull gamma shared frailty model was 2.83, which is relatively large and therefore suggests the existence of heterogeneity.

Conclusion:

The Laplace transform of the frailty distribution plays a central role in relating the hazards, conditional on the frailty, to the hazards and survival functions observed in a population.

Keywords: Frailty, Hazards, MDR-TB, Risk factors, Survival data.

1. INTRODUCTION

Time-to-event data measure the time elapsed from a given origin to the occurrence of an event of interest. Most commonly, survival data are handled using the proportional hazards (PH) regression model popularized by [1]. Correct inference based on those PH models requires independent and identically distributed samples. The PH assumption states that the ratio of the hazards between any two individuals is constant over time, and a non-parametric "baseline hazard gives the shape of the hazard". Subjects may be exposed to different risk levels, even after controlling for known risk factors, because some relevant covariates are often unavailable to the researcher or are even unknown (univariate case). The study population may also be divided into clusters so that subjects from the same cluster behave more cohesively than subjects from different clusters (multivariate case).

* Address correspondence to this author at the Department of Statistics, Durban University of Technology, 41-43 ML Sultan Road, Mariam BEE Building, Durban, 4001, South Africa; Tel.: +27 31 373 5703;
Fax: +27 31 373 5674; E-mail: sizwem@dut.ac.za

A frailty model introduced by [2] that quickly gained popularity in econometrics [3], demographics [4] and biostatistics [5] is a heterogeneity model where the frailties are assumed to be individual, and the frailty has a multiplicative effect on the baseline hazard function. The estimate of the random effect is assumed to have a unit mean and finite variance. Individuals with a frailty greater than one are said to be frailer and will have an increased risk of failure. When the variance of the frailty term is equal to zero, this indicates that observations from the same group are independent. Therefore, when the standard model fails to account for all the variability in the observed failure times, frailty models provide a useful alternative to a standard survival model. The main assumption of a frailty model is that information about hidden internal or external factors is contained in the shape and structure of the hazard function and the form of the frailty distribution [6].

Frailty models have two broad classes, namely models describing the univariate survival times and multivariate models. Frailty models for univariate data have long been used to account for heterogeneous times-to-failure. The term 'frailty' was first suggested by [2] in the context of mortality studies, and [7] incorporated the frailty concept into a study of the duration of unemployment. The shared frailty model may be considered a random-effects model for survival data because the frailty effect is shared amongst clusters of individuals. Early considerations of these models can be found in [8 - 12].

Although several research papers have been published on frailty models, even in the modelling of infectious diseases [13, 14], to the best of the authors' knowledge, no study has been conducted in KwaZulu-Natal, South Africa, applying shared frailty models to survival data. This paper is intended to demonstrate the analysis of frailty models using secondary data from 1 542 MDR-TB patients who were treated in KwaZulu-Natal, South Africa. This study used the statistical software called STATA version 19 to analyse data.

2. METHODS

2.1. Data Source

The data used in this study are described in [15]. The study protocol was approved by the University of KwaZulu-Natal's Biomedical Research Ethics Committee (Ref: BF052/09) and by the KwaZulu-Natal Department of Health. According to [15], they used data collected by health workers for clinical care. No risks were posed to the patients and informed consent was waived by the ethics committee. The authors report that to protect patient confidentiality and anonymity, the databases were de-identified and access was strictly limited.

According to the researchers [15], their study was a prospective study of 4 rural areas in KwaZulu-Natal (South Africa) between 1 July 2008 and 30 June 2012. In this study, the authors used data from 1 542 MDR-TB patients from five TB centres (four decentralized sites and one centralized hospital). Time-to-death of an MDR-TB patient is the response variable of interest.

2.2. Model Development

In general, frailty models are the equivalent of *random-effects* or *mixed models* in survival analysis. Suppose that X represents a covariate vector, let T be a non-negative random variable representing an individual survival time, with t being a realisation of that random variable, then the Cox proportional hazard (PH) model is:

$$h(t, X) = h_0(t) \times e^{X^T \beta}, \qquad \textbf{(1)}$$

where $X = [X_1, X_1, ..., X_n]$ and $\beta = [\beta_1, \beta_1, ..., \beta_n]$ is a regression parameter vector.

A frailty model introduces the unobserved components represented by a vector denoted as U, and (1) is modified as follows:

$$h(t, X, U) = h_0(t) \times e^{X^T \beta + U}$$
$$= h_0(t) \times e^U \times e^{X^T \beta}, \qquad \textbf{(2)}$$

Let $Z = e^U$ represents the frailty term, then the frailty model is:

$$h(t, X | Z) = Z h_0(t) \times e^{X^T \beta}. \qquad \textbf{(3)}$$

The baseline hazard function $h_0(t)$ can be chosen non-parametrically or parametrically (Weibull, exponential, Gompertz, *etc.*). There are two distinguishable broad classes of frailty models, namely the *univariate frailty model* and the *multivariate frailty model.*

2.3. The Univariate Frailty Model

Suppose that an individual is affected with MDR-TB and has a survival time denoted as t_i, covariate vector X_i, with a frailty term denoted as Z_i [2], stated that the hazard function of the individual i is given as:

$$h_i(t_i, X_i | z_i) = z_i h_0(t_i) \times e^{X_i^T \beta}. \qquad \textbf{(4)}$$

Since Z_i is an unobservable random variable varying over the sample. Those individuals who possess $Z_i > 1$ are said to be more frail and will have an increased risk of failure. Conversely, those individuals with $Z_i < 1$ are less frail and will tend to survive longer.

The model of an individual i can also be represented by its conditional survivor function:

$$S_i(t_i, X_i | z_i) = \exp\left(-z_i e^{X_i^T \beta} \int_0^{t_i} h_0(s, X_i | z_i) \, ds \right)$$
$$= \exp\left(-z_i H_0(t_i) \times e^{X_i^T \beta} \right), \qquad \textbf{(5)}$$

and $H_0(t_i) = \int_0^{t_i} h_0(s) ds$ is the cumulative baseline hazard function.

The model described above is at the individual level, but this individual model is not observable [16]. It is vital to state that this is the reason why the model is considered at a population level. Survival of the total population is the mean of the individual survival functions.

The unconditional survival function of an individual i at

the population level is:

$$S_i(t_i, X_i) = E[S_i(t_i, X_i | z_i)]$$

$$= E\left[\exp\left(-z_i H_0(t_i) \times e^{X_i^T}\right)\right].$$

Knowing the frailty distribution can help one determine an individual's unconditional survival function, and this is the same for the unconditional hazard (average hazard).

2.4. The Distribution of Frailty

The average hazard of an individual i, given the conditional hazard, is:

$$\bar{h}_i(t_i, X_i) = \int_0^\infty z_i h_0(t_i) \times e^{X_i^T \beta} f_z(z_i) dz_i,$$

assume that $h_0(t_i)$ is a constant, then:

$$\bar{h}_i(t_i, X_i) = h_0(t_i) \times e^{X_i^T \beta} \int_0^\infty z_i f_z(z_i) dz_i,$$

$$= \bar{z}_i h_0(t_i) \times e^{X_i^T \beta},$$

where $\bar{z}_i = \int_0^\infty z_i f_z(z_i) dz_i$ is the mean frailty.

Suppose X_i is one covariate representing a variable "study site", which is 0 = Centralized hospital, 1 = Decentralised sites:

$$\bar{h}_i(t_i, X_i) = \bar{z}_i h_0(t_i) \times e^{\beta}. \qquad \textbf{(6)}$$

This shows that the average hazard depends on the frailty distribution.

Many distributions can be chosen for frailty, including gamma frailty; the log-normal distribution; the positive stable frailty model; the inverse Gaussian frailty model; and the compound Poisson frailty model, amongst others [17, 18]. The gamma distribution has been used in most applications and widely applied as a mixture distribution due to the simplicity of the Laplace transformation [2, 8, 19 - 21].

$$L\{f(z)\}(s) = \int_0^\infty f(z)\exp(-zs)dz$$

$$= \int_0^\infty \frac{\beta^\alpha z^{\alpha-1} e^{-\alpha z}}{\Gamma(\alpha)} \exp(-zs)dz$$

$$= \frac{\beta^\alpha}{(s+\beta)},$$

Where the parameter s is a complex number:

$s = a + ib$

with a and b real numbers.

Many calculations can be performed based on the Laplace transform. The importance of the Laplace transform for these calculations has previously been demonstrated [22]. The mean and variance of the Gamma distribution can be obtained by

using the first and second derivatives of the Laplace transform, respectively.

$$L^{(1)}(s) = \frac{-\alpha\beta^\alpha}{(s+\beta)^{\alpha+1}}$$

$$L^{(2)}(s) = \frac{\alpha(\alpha+1)\beta^\alpha}{(s+\beta)^{\alpha+2}},$$

on evaluating the derivatives at $s = 0$,

$$E(Z) = -L^{(1)}(0) = \frac{\alpha}{\beta}$$

and

$$Var(Z) = L^{(2)}(0) - \left(-L^{(1)}(0)\right)^2 = \frac{\alpha}{\beta^2}.$$

Another reason that this distribution has been used in most applications published to date is that it is a flexible distribution that takes a variety of shapes as α varies. That is:

$$f(z, \alpha, \beta) = \frac{\beta^\alpha z^{\alpha-1}\exp(-\beta z)}{\Gamma(\alpha)}, \alpha > 0 \ \beta > 0 \ \text{and} \ z > 0.$$

Where α and β are the shape and scale parameters, respectively.

The mean is $E(Z) = \frac{\alpha}{\beta}$ and the variance is $V(Z) = \frac{\alpha}{\beta^2}$.

If $\alpha = 1$, it is identical to the well-known exponential distribution. When α is large, it takes a bell-shaped form. The gamma distribution fits very well with failure data because it is easy to derive the closed-form expressions of unconditional survival, cumulative density and hazard function.

2.5. Multivariate Frailty Model

Survival data analysis always assumes that the time-to-event of the individuals considered in the study is independent. However, this may not always be the case because there is a possibility that the survival times of individuals in the same group, for example, in a family or community, are correlated. The correlation between survival time violates the independence assumption, and such data cannot be analysed using the univariate semi-parametric model. According to [23], the data with correlated survival times are known as multivariate survival data, and models developed to analyse such data include the shared frailty model, which was previously introduced [2, 8].

The model is called the shared frailty model because individuals in the same cluster are assumed to share the same frailty [6, 24]. The survival times of individuals within the same group are assumed to be conditionally dependent, while the frailty across the groups is assumed to be independent. However, when a frailty term represents the individual's unmeasured or unobserved covariates after considering the measured covariates, it is called the univariate frailty model.

Let N denote the number of individuals in a given cohort, with each individual in the cohort assigned to a cluster. Let the

total number of clusters be denoted by G such that, given the ith cluster consists of n_i individuals, then:

$$\sum_{i=1}^{G} n_i = N.$$

Given that $\delta = \begin{cases} 1, & \text{Event occurred} \\ 0, & \text{Censored} \end{cases}$

is the censoring indicator. The response variable will be a combination of time-to-event and δ_i, which takes a value 1 if the time-to-event is observed and 0 if censored or the event did not occur.

The hazard function of the jth individual of the ith cluster is given as:

$$h_{ij}(t) = h_0(t) \times e^{X_{ij}^T \beta + u_i}, \tag{7}$$

where X_{ij} is a vector of covariates for individual j in the ith group, u_i is the unobserved covariates and $h_0(t)$ is the baseline hazard function.

Since $z_i = e^{u_i}$, the hazard function can be written as:

$$h_{ij}(t) = z_i h_0(t) \times e^{X_{ij}^T \beta}. \tag{8}$$

Here, the $z_i's$ are independent with an identical probability density function denoted as $f(z)$.

The full likelihood of the shared frailty model is given as follows:

$$L(t_{ij}, X_{ij}, \theta) = \prod_{i=1}^{G} \prod_{j=1}^{n_i} (z_i)^{\delta_i} h_0(t_{ij})^{\delta_i} \times e^{\delta_i X_{ij}^T \beta} [t_{ij}]^{1-\delta_i} \prod_{i=1}^{G} f(z_i). \tag{9}$$

2.6. Parametric Frailty and Shared Frailty

Now consider a parametric survival model which is characterized by its hazard function, $h(t)$. The effect of any covariate is always found in the definitions of all these functions, whether one parameterizes the model as having PH with respect to changes in covariate values or accelerated failure time (AFT) due to the covariates. For instance, in a Weibull PH regression, the hazard function at time t for individual i with covariate vector x_i is:

$$h_i(t) = h(t|x_i) \quad = e^{x_i \beta} \times pt^{p-1} \tag{10}$$

The shape parameter p and regression coefficients β are estimated from the data. The *streg* contains a list of those forms of $h(t)$ currently available in STATA. In the univariate case, a frailty model introduces an unobservable multiplicative effect z on the hazard, so that conditional on the frailty

$$h(t|z) = zh(t) \tag{11}$$

where z is some random positive quantity assumed to have a unit mean equal to one and variance equal to θ.

A multivariate survival model is an extension of the univariate frailty model, where individuals are allowed to share the same frailty value. Sharing a frailty value also generates dependence between those individuals who share frailties,

whereas conditional on the frailty, those individuals are independent. For data consisting of n clusters with the ith cluster comprised of n_i individuals ($i = 1,...,n$), (11) generalizes to

$$h_{ij}(t|z_i) = z_i h_{ij}(t)$$

for $j = 1,2,...,n_i$ with $h_{ij}(t) = h(t|x_{ij})$. That is, for any member of the ith cluster, the standard hazard function is now multiplied by the shared frailty z_i. For instance, in the case of Weibull PH regression, the conditional hazard for an individual is given by

$$\begin{aligned} h_{ij}(t|z_i) &= z_i h_{ij}(t) \\ &= z_i \times e^{x_{ij}\beta} \times pt^{p-1} \end{aligned}$$

and the conditional survival function is

$$\begin{aligned} S_{ij}(t|z_i) &= \{S_{ij}(t)\}^{z_i} \\ &= \exp\{-z_i \times e^{x_{ij}\beta} \times t^p\} \end{aligned}$$

2.7. Statistical Methods

Data analyses were conducted using STATA version 19 and the Statistical Package for the Social Sciences (SPSS version 25). Basic descriptive statistics, such as frequencies and percentages of demographics, were calculated. The proportion of successful treatment is those cured and treatment completed, and those patients who did not finish treatment, such as those loss to follow-up and failed, are considered to have poor treatment outcomes.

Firstly, the authors fit a Cox PH model and then fit a Weibull regression model with gamma-distributed heterogeneity. Since correlation within a "site" might exist for a given patient, one can model this as a shared frailty model where the sharing occurs at the patient level. This is easily done by adding the option *shared (patient)* to the *streg, frailty*.

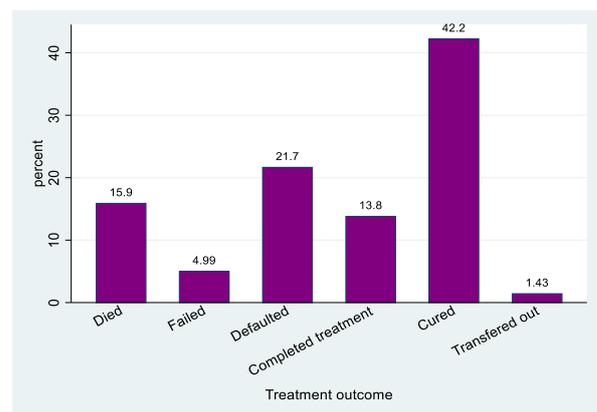**Fig. (1).** Treatment outcomes of MDR-TB patients between 1 July 2008 and 30 June 2012.

3. RESULTS AND DISCUSSION

The study considered an MDR-TB data set consisting of 1 542 patients with time-to-death as the primary outcome of this study. A total of 245 (15.9%) patients with MDR-TB died between 1 July 2008 and 30 June 2012, and 77 (5.0%) had treatment failure. A total of 334 (21.7%) patients were

defaulted, 213 (13.8%) completed treatment; 651 (42.2%) were cured of MRD-TB; and 22 (1.4%) were transferred out (Fig. **1** and Table **1**).

The baseline demographics of the patients treated with MDR-TB showed that 797 (51.7%) were females, and the majority of the patients were aged 18 – 30 years and 31 – 40 years (35.5% and 35.7%, respectively). Most of the patients (71.3%) were HIV-positive. The results also show that most of the patients (95.7%) had no previous MDR-TB episodes, and 792 (51.4%) had no co-morbidities. Furthermore, 1510 (97.9%) of the patients had pulmonary TB (Table **2**). The median follow-up time was 26.8 months.

Table 1. Treatment outcomes of patients with MDR-TB.

Treatment Outcomes	Site 1	Site 2	Site 3	Site 4	All Decentralized Hospitals	Centralized Hospital
	n = 125	n = 142	n = 202	n = 261	n = 730	n = 812
Died	17 (13.6)	21 (14.8)	25 (12.4)	69 (26.4)	132 (18.1)	113 (13.9)
Failed	7 (5.6)	10 (7.0)	12 (5.9)	19 (7.3)	48 (6.6)	29 (3.6)
Defaulted	9 (7.2)	18 (12.7)	50 (24.8)	28 (10.7)	105 (14.4)	229 (28.2)
Completed treatment	12 (9.6)	8 (5.6)	19 (9.4)	15 (5.7)	54 (7.4)	159 (19.6)
Cured	78 (62.4)	79 (55.6)	94 (46.5)	120 (46.0)	371 (50.8)	280 (34.5)
Transferred out	2 (1.6)	6 (4.2)	2 (1.0)	10 (3.8)	20 (2.7)	2 (0.2)

Note: *Data are number (%).

Table 2. Demographic characteristics of the study participants (N=1542).

	n (%)/Median (IQR)						
Variables	**Site 1 (n=125)**	**Site 2 (n=142)**	**Site 3 (n=202)**	**Site 4 (n=261)**	**All decentralized hospitals (n=730)**	**Centralized hospital (n=812)**	**TOTAL (n=1542)**
Baseline weight (kg)	50 (43 - 59)						
Age (in years)	34 (28 - 42)						
Age groups	-						
18 - 30	42 (33.6)	44 (31.0)	74 (36.6)	85 (32.6)	245 (33.6)	303 (37.3)	548 (35.5)
31 - 40	43 (34.4)	55 (38.7)	70 (34.7)	90 (34.5)	258 (35.3)	292 (36.0)	550 (35.7)
41 - 50	30 (24.0)	27 (19.0)	40 (19.8)	56 (21.5)	153 (21.0)	145 (17.9)	298 (19.3)
50+	10 (8.0)	16 (11.3)	18 (8.9)	30 (11.5)	74 (10.1)	72 (8.9)	146 (9.5)
Gender	-						
Male	57 (45.6)	60 (42.2)	104 (51.5)	125 (47.9)	346 (47.4)	399 (49.1)	745 (48.3)
Female	68 (54.4)	82 (57.7)	98 (48.5)	136 (52.1)	384 (52.6)	413 (50.9)	797 (51.7)
HIV status	-						
Positive	96 (76.8)	108 (76.1)	123 (60.9)	197 (75.5)	524 (71.8)	576 (70.9)	1100 (71.3)
Negative	28 (22.4)	30 (21.1)	66 (32.7)	38 (14.6)	162 (22.2)	211 (26.0)	373 (24.2)
Unknown	1 (0.8)	4 (2.8)	13 (6.4)	26 (10.0)	44 (6.0)	25 (3.1)	69 (4.5)
Previous MDR-TB episodes	-						
No previous episodes	119 (95.2)	124 (87.3)	184 (91.1)	246 (94.3)	673 (92.2)	802 (98.8)	1475 (95.7)
One previous episode	5 (4.0)	18 (12.7)	18 (8.9)	14 (5.4)	55 (7.5)	9 (1.1)	64 (4.1)
Two previous episodes	1 (0.8)	0 (0)	0 (0)	1 (0.3)	2 (0.3)	1 (0.1)	3 (0.2)
Comorbidities	-						
No other diseases	6 (40.0)	2 (15.4)	1 (100)	0 (0)	9 (20.5)	780 (97.6)	789 (93.6)
Diabetes	1 (6.7)	1 (7.7)	0 (0)	8 (53.3)	10 (22.7)	10 (1.3)	20 (2.4)
Epilepsy	4 (26.7)	3 (23.1)	0 (0)	1 (6.7)	8 (18.2)	4 (0.5)	12 (1.4)
Hearing loss prior to start of treatment	2 (13.3)	3 (23.1)	0 (0)	5 (33.3)	10 (22.7)	1 (0.1)	11 (1.3)
Renal problems	1 (6.7)	2 (15.4)	0 (0)	0 (0)	3 (6.8)	0 (0)	3 (0.4)
Substance abuse	1 (6.7)	2 (15.4)	0 (0)	1 (6.7)	4 (9.1)	0 (0)	4 (0.5)
Psychiatric problems	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	4 (0.5)	4 (0.5)
Type of TB	-						
Pulmonary	122 (97.6)	131 (92.3)	199 (98.5)	254 (97.3)	706 (96.7)	804 (99.0)	1510 (97.9)
Extra pulmonary	3 (2.4)	11 (7.7)	3 (1.5)	7 (2.7)	24 (3.3)	8 (1.0)	32 (2.1)

Note: *IQR = Interquartile range.

Table 3. Parameter estimates (S. E) in the Cox model and parametric shared frailty models applied to the MDR-TB data.

Variables	Cox Regression Model			Weibull Regression Model with Gamma Shared Frailty			Log-normal Regression Model with Gamma Shared Frailty		
	Coef. (95% CI)	S.E	P-value	Coef. (95% CI)	S.E	P-value	Coef. (95% CI)	S.E	P-value
Baseline weight (kg)	-0.03 (-0.04 to -0.02)	0.01	0.00	0.07 (0.04 to 0.09)	0.01	0.00	0.06 (0.04 to 0.09)	0.01	0.00
Study sites									
Centralized hospital (Ref)	-			-			-		
Decentralized sites	0.45 (-0.51 to 1.41)	0.49	0.36	-1.15 (-4.01 to 1.72)	1.46	0.43	-1.25 (-4.12 to 1.63)	1.47	0.40
Age group (years)									
18-30 (Ref)	-			-			-		
31-40	0.53 (0.02 to 0.85)	0.26	0.04	-1.48 (-2.68 to -0.28)	0.61	0.02	-1.42 (-2.60 to -0.23)	0.61	0.02
41-50	0.60 (-0.01 to 1.21)	0.31	0.05	-1.19 (-2.56 to 1.19)	0.70	0.09	-1.13 (-2.56 to 1.29)	0.73	0.12
>-51	0.48 (-0.27 to 1.23)	0.38	0.21	-1.44 (-1.18 to 2.31)	0.89	0.11	-1.27 (-2.99 to 1.45)	0.88	0.15
Gender									
Male (Ref)	-			-			-		
Female	0.52 (0.09 to 0.96)	0.22	0.02	-1.19 (-2.17 to -0.22)	0.50	0.02	-1.32 (-2.32 to -0.33)	0.51	0.01
HIV status									
Positive (Ref)	-			-			-		
Negative	-0.06 (-0.56 to 1.45)	0.26	0.82	0.15 (-0.96 to 1.26)	0.57	0.79	0.17 (-0.98 to 1.31)	0.58	0.78
Type of TB									
Pulmonary (Ref)	-			-			-		
Extrapulmonary	0.25 (-1.17 to 1.67)	0.72	0.73	-0.78 (-4.33 to 2.77)	1.81	0.67	-0.72 (-4.17 to 2.73)	1.76	0.68
Comorbidities conditions									
No (Ref)	-			-			-		
Yes	-0.33 (-1.42 to 1.76)	0.56	0.55	0.75 (-2.29 to 3.79)	1.55	0.63	0.85 (-2.18 to 3.89)	1.55	0.58
Previous MDR-TB episodes									
No (Ref)	-			-			-		
Yes	-0.19 (-1.62 to 1.24)	0.73	0.80	0.49 (-3.13 to 4.10)	1.85	0.79	0.47 (-3.16 to 4.11)	1.85	0.80
Likelihood ratio test on 10 df	42.68, p-value < 0.01			40.57, p-value < 0.01			39.99, p-value < 0.01		
θ (S.E)	-			2.83 (1.89)			0 (0.09)		
Likelihood ratio test of è	-			chibar2(01) = 4.19, p-value = 0.02			chibar2(01) = 0, p-value = 1.00		

Note: • S. E = Standard error; CI = Confidence interval.

The results in Table **3** show the Cox regression parameter estimates, Weibull shared frailty parameter estimates and log-normal frailty parameter estimates. The results also show the chi-square test statistic ($X^2 = 4.19$) for a Weibull frailty model with a p-value of 0.02. The estimate of the variance for the frailty term in the Weibull gamma shared frailty model is 2.83, which is different from zero. This implies that unobserved heterogeneity was present at the site level. Therefore, one can use the Weibull frailty shared model because the results suggest that patients at some sites were associated with a higher risk of dying than the other sites and there is a difference in the conclusions drawn about the dataset.

Further evidence is the differences in the coefficient estimates between the Weibull frailty model and the ordinary Cox regression model. Hazard ratios now have an interpretation that is conditional on frailty. Note that the standard deviation increases in the gamma frailty and log-normal models. The log-normal model results in a slightly lower likelihood ratio test. If one looks at the results produced by the log-normal frailty model, the frailty variance is zero and is not significant. This implies that all the differences amongst the mortality rates of the MDR-TB patients are explained by the observed fixed covariates stated in the model. That is, the log-normal model indicates that there is no unobserved heterogeneity amongst groups (sites), but the estimate of the standard errors increased.

When comparing the estimates of the standard Cox PH model and those of the Weibull gamma frailty model, an increase in the estimates was observed after correcting for frailty (Table **3**). Factors strongly associated with mortality

rates were identified by examining their confidence intervals (CIs) and p-values. Factors whose CIs include 1 implied that these factors were insignificant and these results were confirmed by p-values > 0.05 significance level. Baseline weight, age group (31- 40) and gender were found to be strongly associated with mortality rates. The results show that patients treated in decentralized sites had a lower death rate than those treated in centralized sites, but this was insignificant.

4. DISCUSSION

Frailty models are mixture of models within survival analysis. The estimation in this general frailty model framework is performed by using a simple two-step procedure where the fixed effects and the individual frailty terms are estimated while keeping the frailty variance parameter fixed. This procedure leads to simple estimation equations but results in an underestimation of the estimated variances of the fixed effects parameters because the variance estimation does not take into account the variability in the estimation of $\hat{e}$.

The objective of this paper was to assess the effect of omitting unmeasurable variables in the modelling exercise. The researchers began by fitting a semi-parametric stratified Cox regression model with random effects and then fitted a Weibull parametric stratified model with random effects. This was an extension of the stratified Cox regression model with the site variable as the stratifying variable. The results of the exercise suggested that there exists significant unobserved heterogeneity. When accounting for site effects, the confidence intervals are wider in the Weibull frailty model, which implies that there is much more heterogeneity at decentralized sites than in the centralized hospital.

The other factors included in the analysis, namely HIV status, type of TB, previous MDR-TB episodes and co-morbidities, affect on the survival time of MDR-TB patients. Still, there was not enough evidence from the data to confirm their association with this survival time.

The results in this paper do not deviate much from the findings obtained by [27], who concluded that frailty is a strong predictor of mortality, as has been shown by previous systematic reviews [28 - 30]. Two of these reviews systematically conducted studies that used different definitions of frailty, including the frailty phenotype by [31] and the frailty index (FI) by [32], and demonstrated that frailty consistently increased the risk of death in most studies [28, 29].

When comparing the Cox PH model to the frailty models, it was found that the estimates of the standard error for the fixed effects increased. In some cases, adding a frailty term can render a result insignificant in the frailty model even though it was significant in the Cox PH model [33].

A follow-up study may be a good alternative to MDR-TB surveys. In this case, a cohort of patients may be followed from different cities. In this paper, the researchers considered only models with gamma or log-normal frailty. This approach can be easily extended to other frailty distributions available, such as inverse Gaussian distributions. However, the approach for the positive stable distribution [19], which is expressed as a Laplace transform, would be interesting future work.

CONCLUSION

This study shows that the gamma frailty model provides a better fit to the MDR-TB data than the standard Cox model. Although further research must be conducted about the models, an initial investigation suggests that the models will serve as an enhancement to the field of survival analysis. The authors conclude that frailty in a survival model is an important consideration and is especially useful in situations where clustering needs to be accounted for.

LIST OF ABBREVIATIONS

STATA	=	Statistics and data
MDR-TB	=	Multidrugresistant tuberculosis
TB	=	Tuberculosis
FI	=	Frailty index

LIMITATIONS

The findings should be interpreted with caution due to a few limitations. Some covariates that were included in the model had missing information. The major problem with this is that the quality of the results might decrease due to less completeness of the data. Furthermore, some patients who were included in the study were lost to follow-up.

AUTHORS' CONTRIBUTIONS

All the authors made contributions to this paper. SVM planned the study, wrote the paper's initial draft, and did the analysis. HW and SR revised and edited the paper. All authors read and approved the final work of this manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The study protocol was approved by the University of KwaZulu-Natal's Biomedical Research Ethics Committee (Ref: BF052/09) and by the KwaZulu-Natal Department of Health. Only secondary data, the data routinely collected by health workers for clinical care, was used in this study.

HUMAN AND ANIMAL RIGHTS

No animals were used in this research. All procedures performed in studies involving human participants were in accordance with the ethical standards of institutional and/or research committees and with the 1975 Declaration of Helsinki, as revised in 2013.

CONSENT FOR PUBLICATION

Informed consent was waived by the ethics committee since all patient data used were previously collected during the course of routine medical care and did not pose any additional risks to the patients.

AVAILABILITY OF DATA AND MATERIALS

Data will be made available upon request but will be controlled. To protect patient confidentiality and anonymity, the data bases were de-identified and access strictly limited.

FUNDING

None Declared.

CONFLICT OF INTEREST

All authors report no potential conflict of interest.

REFERENCES

[1] Cox DR. Regression models and life‐tables. J R Stat Soc B 1972; 34(2): 187-202.
[http://dx.doi.org/10.1111/j.2517-6161.1972.tb00899.x]

[2] Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography 1979; 16(3): 439-54.
[http://dx.doi.org/10.2307/2061224] [PMID: 510638]

[3] Heckman J, Singer B. A method for minimizing the impact of distributional assumptions in econometric models for duration data. Econometrica 1984; 52(2): 271-320.
[http://dx.doi.org/10.2307/1911491]

[4] Vaupel JW, Yashin AI. Heterogeneity's ruses: Some surprising effects of selection on population dynamics. Am Stat 1985; 39(3): 176-85.
[PMID: 12267300]

[5] Aalen OO. Heterogeneity in survival analysis. Stat Med 1988; 7(11): 1121-37.
[http://dx.doi.org/10.1002/sim.4780071105] [PMID: 3201038]

[6] Hanagal DD. Modeling survival data using frailty models. Boca Raton: Chapman & Hall/CRC 2011.
[http://dx.doi.org/10.1201/b10510]

[7] Lancaster T. Econometric methods for the duration of unemployment. Econometrica 1979; 47(4): 939-56.
[http://dx.doi.org/10.2307/1914140]

[8] Clayton DG. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 1978; 65(1): 141-51.
[http://dx.doi.org/10.1093/biomet/65.1.141]

[9] Clayton D, Cuzick J. Multivariate generalizations of the proportional hazards model. J R Stat Soc 1985; 148(2): 82-108.
[http://dx.doi.org/10.2307/2981943]

[10] Hougaard P. Survival models for heterogeneous populations derived from stable distributions. Biometrika 1986; 73(2): 387-96.
[http://dx.doi.org/10.1093/biomet/73.2.387]

[11] Whitmore GA, Lee MLT. A multivariate survival distribution generated by an inverse Gaussian mixture of exponentials. Technometrics 1991; 33(1): 39-50.
[http://dx.doi.org/10.1080/00401706.1991.10484768]

[12] Sahu SK, Dey DK, Aslanidou H, Sinha D. A Weibull regression model with gamma frailties for multivariate survival data. Lifetime Data Anal 1997; 3(2): 123-37.
[http://dx.doi.org/10.1023/A:1009605117713] [PMID: 9384618]

[13] Hens N, Wienke A, Aerts M, Molenberghs G. The correlated and shared gamma frailty model for bivariate current status data: An illustration for cross-sectional serological data. Stat Med 2009; 28(22): 2785-800.
[http://dx.doi.org/10.1002/sim.3660] [PMID: 19591117]

[14] Unkel S, Farrington CP, Whitaker HJ, Pebody R. Time varying frailty models and the estimation of heterogeneities in transmission of infectious diseases. Appl Stat 2014; 63(1): 141-58.
[http://dx.doi.org/10.1111/rssc.12033]

[15] Loveday M, Padayatchi N, Wallengren K, *et al.* Association between health systems performance and treatment outcomes in patients co-infected with MDR-TB and HIV in KwaZulu-Natal, South Africa: Implications for TB programmes. PLoS One 2014; 9(4): e94016.
[http://dx.doi.org/10.1371/journal.pone.0094016] [PMID: 24718306]

[16] Wienke A. Frailty models in survival analysis. Boca Raton: Chapman and Hall/CRC 2010.
[http://dx.doi.org/10.1201/9781420073911]

[17] Duchateau L, Janssen P. The frailty model. New York: Springer Verlag 2008.

[18] Munda M, Rotolo F, Legrand C. parfm: Parametric frailty models in R. J Stat Softw 2012; 51(11): 1-20.
[http://dx.doi.org/10.18637/jss.v051.i11]

[19] Hougaard P, Hougaard P. Shared frailty models. In: Analysis of Multivariate Survival Data Statistics for Biology and Health. New York, NY: Springer 2000.
[http://dx.doi.org/10.1007/978-1-4612-1304-8_7]

[20] Oakes D. A model for association in bivariate survival data. J R Stat Soc B 1982; 44(3): 414-22.
[http://dx.doi.org/10.1111/j.2517-6161.1982.tb01222.x]

[21] Yashin AI, Iachine IA. Genetic analysis of durations: Correlated frailty model applied to survival of Danish twins. Genet Epidemiol 1995; 12(5): 529-38.
[http://dx.doi.org/10.1002/gepi.1370120510] [PMID: 8557185]

[22] Hougaard P. Life table methods for heterogeneous populations: Distributions describing the heterogeneity. Biometrika 1984; 71(1): 75-83.
[http://dx.doi.org/10.1093/biomet/71.1.75]

[23] Phipson B, Mwambi H. Incorporating frailty effects in the Cox proportional hazards model using two independent methods in independent data sets: Theory and methods. S Afr Stat J 2010; 44(1): 61-81.

[24] Shih JH, Louis TA. Assessing gamma frailty models for clustered failure time data. In: Lifetime Data: Models in Reliability and Survival Analysis. Boston, MA: Springer 1996.
[http://dx.doi.org/10.1007/978-1-4757-5654-8_39]

[25] Laserson KF, Thorpe LE, Leimane V, *et al.* Speaking the same language: Treatment outcome definitions for multidrug-resistant tuberculosis. Int J Tuberc Lung Dis 2005; 9(6): 640-5.
[PMID: 15971391]

[26] World Health Organisation. Guidelines for the programmatic management of drug-resistant tuberculosis Emergency Update 2008 WHO/HTM/TB/2008402. Geneva: World Health Organisation 2008.

[27] Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. Lancet 2013; 381(9868): 752-62.
[http://dx.doi.org/10.1016/S0140-6736(12)62167-9] [PMID: 23395245]

[28] Kane RL, Shamliyan T, Talley K, Pacala J. The association between geriatric syndromes and survival. J Am Geriatr Soc 2012; 60(5): 896-904.
[http://dx.doi.org/10.1111/j.1532-5415.2012.03942.x] [PMID: 22568483]

[29] Shamliyan T, Talley KMC, Ramakrishnan R, Kane RL. Association of frailty with survival: A systematic literature review. Ageing Res Rev 2013; 12(2): 719-36.
[http://dx.doi.org/10.1016/j.arr.2012.03.001] [PMID: 22426304]

[30] Chang SF, Lin PL. Frail phenotype and mortality prediction: A systematic review and meta-analysis of prospective cohort studies. Int J Nurs Stud 2015; 52(8): 1362-74.
[http://dx.doi.org/10.1016/j.ijnurstu.2015.04.005] [PMID: 25986959]

[31] Fried LP, Tangen CM, Walston J, *et al.* Frailty in older adults: Evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001; 56(3): M146-57.
[http://dx.doi.org/10.1093/gerona/56.3.M146] [PMID: 11253156]

[32] Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. ScientificWorldJournal 2001; 1: 323-36.
[http://dx.doi.org/10.1100/tsw.2001.58] [PMID: 12806071]

[33] Phipson B. Analysis of time-to-event data including frailty modeling (Doctoral dissertation). South Africa: University of Kwazulu-Natal Pietermaritzburg 2006.