# Multivariate frailty models using survey weights with applications to twins infant mortality in Ethiopia

YEHENEW G. KIFLE, DING-GENG CHEN*, AND MESFIN T. HAILEYESUS

Several studies have shown that twin birth contributes substantially to infant and child mortality mainly in resource-poor countries. The excess rates among twins call for research in statistical modeling to identify the main causes behind it. In studies involving multiple individuals from the same family, the fundamental independence assumption in the classical statistical modeling is not plausible. In addition, previous studies indicated that ignoring sampling weight while dealing with a dataset collected with complex survey design can introduce serious bias. This study is then aimed to fill these methodological gaps to integrate the dependence from twin birth with an advanced statistical gamma frailty model to correctly identify the determinants of infant mortality among twins in Ethiopia. We compiled all available data from the 2016 Ethiopia Demographic and Health Survey with a total of 908 children (454 pairs of twins) with survey sampling weight incorporated in the analysis. To identify predictors and to assess the presence and significance of frailty, semiparametric univariate, bivariate shared, and correlated gamma frailty models were fitted. The likelihood ratio test was employed to test the significance of frailty term in the model. We found that sex of the child, among twins birth order, preceding birth interval, and succeeding birth interval are significantly associated with twin infant mortality. The results of this study further confirmed the significance of the shared frailty term accounting for the unobserved heterogeneity.

KEYWORDS AND PHRASES: Frailty, Twin, Infant mortality, Sampling weight, Survival.

## 1. INTRODUCTION

In Ethiopia, the rate of infant mortality shows an encouraging decline pattern in the past decade. The 2016 Ethiopia Demographic and Health Survey (EDHS) results showed that infant mortality declined from 97 deaths per 1,000 live births in 2000 to 48 deaths per 1,000 live births in 2016, which is about a 50% reduction in about 16 years. The 2016 figure further indicates that 1 in every 21 children in Ethiopia dies before celebrating their first birthday [5].

*Corresponding author.

Sub-Saharan Africa, not only has the highest rate of twin births in the world [21] but also the world's highest rate of infant and child mortality [24]. Despite the significant improvements, the fate of twins' survival in sub-Saharan Africa is lagging behind that of singletons with an alarming one-fifth of twins in the region die before the age of five years, which is three times the mortality rate among singletons [18]. In Ethiopia, the 2016 EDHS results showed that out of the 914 children of multiple births recorded in the survey (2 triple births and 454 twin births), 285 (31.2%) of them died before they reach their first birthday, indicating that 1 in every 4 children of multiple births in Ethiopia dies before their first birthday [5]. There are a number of biological and environmental contributing factors for this high mortality rate among children of multiple births in Sub-Saharan African countries [8, 12, 21, 18]. For instance, close contact between twin babies increases the chance of cross-infection [12]. Hence, in studies involving multiple individuals from the same family (e.g. twins), it is obvious to expect some sort of association among these twins and the assumption of independence is not plausible unless all-important familial factors were measured and controlled for in the model. Children belonging to the same family share certain unobserved characteristics (heterogeneity), which may not be sufficiently described by the covariates included in the models. Hence, failure to consider such unobserved association may lead to biased parameters estimates [9].

Apart from taking into account the unobserved association among twins, one has to consider also those major elements of the survey design that may have an effect on the model estimates based on survey data, including sampling weights, which is the inverse probability of being included in the sample adjusted for non-response [15, 20]. These weights act to correct sample data for the unequal selection probabilities and failure to include these in the modeling process can lead to estimates that are seriously biased for their corresponding population quantities [17].

The persistence of high levels of infant mortality rates among twins calls for a need to identify the potential determinants of twins' mortality, specifically in resource-poor sub-Saharan countries like Ethiopia. Identifying potential significant determinants of twin birth infant mortality is essential to form policies and strategies to accelerate the re-

duction of infant mortality and also to meet the United Nations Sustainable Development Goals (SDGs). This study, therefore, aims to identify the determinants of infant mortality among twins in Ethiopia, using an advanced frailty modeling approach to incorporate the dependence/correlation from twin mortality so that correct analysis can be performed and appropriate public health recommendations can be made.

## 2. MATERIALS AND METHODS

### 2.1 Data source

This study used data from the 2016 EDHS, where information about the twin's mortality is extracted from the birth history of women included in the survey. All twin births recorded in the survey are included in this study. The 2016 EDHS sample was stratified and selected in two stages. Each region was stratified into urban and rural areas, yielding 21 sampling strata. Samples of EAs were selected independently in each stratum in two stages. Implicit stratification and proportional allocation were achieved at each of the lower administrative levels by sorting the sampling frame within each sampling stratum before sample selection, according to administrative units in different levels, and by using a probability proportional to size selection at the first stage of sampling [5].

### 2.2 Variables of the study

The outcome (response) variable for this study is the survival time of a pair of infant twins measured in days. Among the potential covariates that might have an effect on the survival of twin infants, the following time-invariant factors are included in the study: *categorized mothers age at childbirth* (below 18 years, between 18–35 years and above 35 years), *sex of the child*, *preceding birth interval* (1st born or no preceding child, below 18 months, between 18–24 months and above 24 months), *succeeding birth interval* (below 18 months, between 18–24 months, above 24 months and last born) and among *twins birth order* (firstborn and secondborn). In addition, the less likely-to-be time-variant factor such as *place of residence* (urban, rural) is also included in this study.

### 2.3 Methods of data analysis

With the aim of identifying potential predictors of infant mortality and further assessing the presence and significance of unobserved frailty a semiparametric univariate, shared, and correlated gamma frailty models are fitted to the 2016 EDHS twins survival data. The following discussion is restricted only to bivariate survival data.

#### 2.3.1 Shared gamma frailty model

Let the bivariate random variables $(T_{i1}, T_{i2})$ be the first and second survival times of the two children in the $i^{th}$ cluster (twin) $(i = 1, \ldots, n)$. Assuming that the frailties $Z_i, (i = 1, \ldots, n)$ are acting multiplicatively on the baseline hazard function $h_0(t_{ij})$ and both the survival times of children $T_{i1}$ and $T_{i2}$ are conditionally independent given frailty $Z_i = z_i$. The conditional hazard model for $j^{th}$ child $(j = 1, 2)$ in the $i^{th}$ twin given frailty $Z_i = z_i$ has the form [13]:

$$h_{ij}(t_{ij}|\mathbf{X}_{ij}, Z_i) = z_i h_0(t_{ij}) \exp(\mathbf{X}'_{ij}\boldsymbol{\beta}), \quad j = (1, 2) \tag{1}$$

where the vectors $\mathbf{X}_{ij} = (X_{ij1}, X_{ij2}, \ldots, X_{ijp})'$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$ are covariates and regression parameters, respectively. Those children who possess $z_i > 1$ are more frail for reasons left unexplained by the covariates and will have an increased risk of death. The frailties, $Z_i$, are assumed to be independently and identically distributed random variables. In this study, $Z_i$ are assumed to follow gamma distribution given by:

$$g(z) = \frac{z^{\frac{1}{\sigma^2}-1} \exp(-\frac{z}{\sigma^2})}{\Gamma(\frac{1}{\sigma^2})(\sigma^2)^{\frac{1}{\sigma^2}}} \tag{2}$$

Under the assumption of independence, the conditional survival function in the bivariate case for given frailty $Z_i = z_i$ at time $t_{i1} > 0$ and $t_{i2} > 0$ is,

$$\begin{aligned} S(t_{i1}, t_{i2}|X_{i1}, X_{i2}, Z_i) &= S(t_{i1}|X_{i1}, Z_i) S(t_{i2}|X_{i2}, Z_i) \\ &= e^{-z_i[H(t_{i1})+H(t_{i2})]} \end{aligned} \tag{3}$$

where $H(t_{ij}) = H_0(t_{ij}) \exp(\mathbf{X}'_{ij}\boldsymbol{\beta})$ for $j = (1, 2)$ and $H_0(t)$ denote the cumulative baseline hazard function. Consider the marginal likelihood function:

$$L_{marg} = \int_0^\infty (-1)^{(\delta_1+\delta_2)} \frac{\partial^2}{\partial t_1^{\delta_1} \partial t_2^{\delta_2}} e^{-z(H(t_1)+H(t_2))} g(z) dz \tag{4}$$

where $g(z)$ is the probability density function given in *equation-2*. If a parametric form is not assumed for the baseline hazard, $h_0(.)$, estimates of the model parameters can easily be obtained using either the method of Expectation-Maximization (EM) or penalized partial likelihood (PPL).

Let's consider parameter estimation technique following the modified EM approach for the semiparametric bivariate shared gamma frailty model with incorporating sampling weight. The word 'modified' refers to the modification of the M-step so that the frailty parameter can be estimated from the profile log-likelihood as discussed below. In the univariate gamma frailty model, random effect term is introduced to each infant instead of the twins together. Since, univariate frailty model refers to a shared frailty model when the cluster size is one, estimation of the univariate (individual) frailty approach can be straightforward from the estimation of bivariate shared gamma frailty approach. Consider the full likelihood as if the frailties were observed. The log-likelihood is given by [7]:

$$L(\boldsymbol{\beta}, \sigma^2|Z) = L_1(\boldsymbol{\beta}|Z) L_2(\sigma^2|Z) \tag{5}$$

where

$$L_1(\boldsymbol{\beta}|Z) = \prod_{i=1}^{n} \left[ (z_i h_0(t_{i1}) e^{\mathbf{X}'_{i1}\boldsymbol{\beta}})^{\delta_{i1}} (z_i h_0(t_{i2}) e^{\mathbf{X}'_{i2}\boldsymbol{\beta}})^{\delta_{i2}} \right.$$

$$\left. \times e^{-z_i (H_0(t_{i1}) e^{\mathbf{X}'_{i1}\boldsymbol{\beta}} + H_0(t_{i2}) e^{\mathbf{X}'_{i2}\boldsymbol{\beta}})} \right] \tag{6}$$

and

$$L_2(\sigma^2|Z) = \prod_{i=1}^{n} g(z_i) \tag{7}$$

The complete likelihood representation given in *equation-5* is obtained by taking partial derivatives of *equation-3* with respect to $(t_1, t_2)$ in accordance with censoring information for the twin of children and multiplying the result by the corresponding probability density function given in *equation-2*. Since the data used in this study is from EDHS which uses a sample survey design with weights, accordingly these sampling weights are incorporated in the analysis. Now with sampling weight denoted by $w$ and sampling weight of the individual mother $w_{i1} = w_{i2} = w_i$, *equation-6* become:

$$L_1(\boldsymbol{\beta}|Z) = \prod_{i=1}^{n} \left[ (z_i h_0(t_{i1}) e^{\mathbf{X}'_{i1}\boldsymbol{\beta}})^{\delta_{i1} w_i} (z_i h_0(t_{i2}) e^{\mathbf{X}'_{i2}\boldsymbol{\beta}})^{\delta_{i2} w_i} \right.$$

$$\left. \times e^{-z_i w_i (H_0(t_{i1}) e^{\mathbf{X}'_{i1}\boldsymbol{\beta}} + H_0(t_{i2}) e^{\mathbf{X}'_{i2}\boldsymbol{\beta}})} \right] \tag{8}$$

Denote the weighted version of Nelson–Aalen estimator of $H_0(t)$ [4] by:

$$\hat{H}_0(t) = \sum_{t_{ij} \le t} \frac{d_{w_{pl}}}{\sum_{t_{pl} \ge t_{ij}} w_p e^{\mathbf{X}'_{pl}\boldsymbol{\beta} + \ln(z_p)}} \tag{9}$$

where $d_{w_{pl}} = \sum_{t_{pl}=t_{ij}} \delta_{pl} w_p$ is the number of events at time $t_{ij}$ or number of children died at time $t_{ij}$ (weighted). Let's further substitute the Nelson–Aalen estimator of $H_0(t)$ given in *equation-9* into the logarithm of equation-8 to get:

$$l_1(\boldsymbol{\beta}|Z) = \sum_{i=1}^{n} \sum_{j=1}^{2} \delta_{ij} w_i \left[ \eta_{ij} - \ln \left( \sum_{t_{pl} \ge t_{ij}} w_p e^{\eta_{pl}} \right) \right] \tag{10}$$

where $\eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \ln(z_i)$. Now one can easily see that *equation-10* coincides with Cox's partial likelihood with sampling weight $w$ and $\ln(z)$ as an additional Cox-like covariate with a known regression coefficient equal to one. Further, instead of $Z_i$ one has to use its expected value, $E(Z_i)$. Suppose $\hat{\beta}_q, \hat{H}_{0_q}(t), \hat{\sigma}^2_q$ denote estimates of $\beta, H_0(t)$ and $\sigma^2$ at $q^{th}$ iteration. Estimates of the conditional expectations of $Z_i$ at $q^{th}$ iteration denoted by $\hat{z}_{i_q}, i = 1, \ldots, n$ evaluated at parameter estimates of $q^{th}$ iteration can be obtained using:

$$\hat{z}_{i_q} = \frac{\frac{1}{\hat{\sigma}^2_q} + \sum_{j=1}^{2} w_i \delta_{ij}}{\frac{1}{\hat{\sigma}^2_q} + \sum_{j=1}^{2} w_i \hat{H}_{0_q}(t_{ij}) e^{\mathbf{X}'_{ij}\hat{\boldsymbol{\beta}}_q}} \tag{11}$$

Following profile likelihood approach due to Therneau and Grambsch [22], first express the hazard terms in terms of the expected frailty and the unweighted event. That is for the unweighted gamma frailty, the conditional expectations of the frailty term is $\hat{z}_i = \frac{\theta + d_i}{\theta + [H(t_1) + H(t_2)]}$, implies $[H(t_1) + H(t_2)] = e^{-\ln(\hat{z}_i)}(\theta + d_i) - \theta$, where $\theta = \frac{1}{\sigma^2}$ and the unweighted event is given as $d_i = (\delta_{i1} + \delta_{i2})$. Then substitute this term into the log of the marginal likelihood given in *equation-4* and then, subtracting and adding a penality term $\sum_{i=1}^{n} \frac{-1}{\sigma^2}(\ln(z_i) - z_i)$ results:

$$\begin{aligned} l_{prof}(\sigma^2) = l_{cox} + \sum_{i=1}^{n} \Bigg[ & \frac{1}{\sigma^2}(\log(z_i) - z_i) \\ & + (\frac{1}{\sigma^2} + d_i)(1 - \log(\frac{1}{\sigma^2} + d_i)) \\ & - \frac{1}{\sigma^2}\log(\sigma^2) + \log\Gamma\left(\frac{1}{\sigma^2} + d_i\right) - \log\Gamma\left(\frac{1}{\sigma^2}\right) \Bigg] \end{aligned} \tag{12}$$

where $l_{cox}$ can be obtained using *equation-10* and here $d_i$ is the sum of events (unweighted) in pair $i$. Note that for gamma frailty, both PPL and EM approaches produce the same result [7, 22]. The main feature of *equation-12* is the expression of hazard term in terms of the expected frailty and the unweighted events. i.e., $\sum_{j=1}^{2} w_i H_0(t_{ij}) e^{\mathbf{X}'_{ij}\boldsymbol{\beta}} = \exp(-\ln(z_i))(1/\sigma^2 + d_i) - 1/\sigma^2$, which is not equal to $\exp(-\ln(z_i))(1/\sigma^2 + d_{w_i}) - 1/\sigma^2$ where $d_{w_i}$ is the sum of events (weighted) in the $i^{th}$ twin pair. Now, the algorithm of parameters estimation consists of an inner and outer loops. Let's denote the outer loop iteration by $l$ and the inner loop by $q$.

**Inner loop**

Given a conventional values of $\hat{\boldsymbol{\beta}}_{l,q}, \hat{H}_{0_{l,q}}(t)$ and $\hat{\sigma}^2_l$,

- Step 1: Obtain $\hat{z}_{i_q}$ using equation-11.
- Step 2: Maximize *equation-10* to obtain $\hat{\boldsymbol{\beta}}_{l,q+1}$ and obtain $\hat{H}_{0_{l,q+1}}(t)$ using equation-9. Maximization of *equation-10* given $\hat{z}_{i_q}$ can be carried out using standard Cox PH fit procedure with $\ln(\hat{z}_{i_q})$ as an "offset" term and $w_i$ as weight.
- Step 3: Iterate step 1 and 2 until the maximum absolute differences between successive estimates reach tolerance i.e., $max(|\hat{\boldsymbol{\beta}}_{l,q+1} - \hat{\boldsymbol{\beta}}_{l,q}|) < 10^{-10}$.

**Outer loop**

Given $\hat{\boldsymbol{\beta}}_{l,q+1}, \hat{H}_{0_{l,q+1}}, \hat{z}_{i_{q+1}}$

- Maximize *equation-12* to obtain $\hat{\sigma}^2_{l+1}$.
- Iterate inner and outer loop until convergence. Convergence can be checked using absolute difference of estimated frailty parameters estimated at previous and current outer iterations. i.e., $|\hat{\sigma}^2_{l+1} - \hat{\sigma}^2_l| < 10^{-6}$.

### 2.3.2 Bivariate correlated gamma frailty model

The bivariate correlated gamma frailty model introduced by [26] included univariate and shared frailty models as special cases. If the frailty variances of the two subjects in pairs are zero, then it implies absence of frailty. With non-zero frailty variances, if the correlation between the frailties is zero, then correlated frailty model reduced to univariate frailty model. In addition, if the correlation between the frailties is 1, then correlated frailty model reduced to shared frailty model [10, 11].

Let the survival times of the two subjects (children) be conditionally independent given their frailties $Z_1$ and $Z_2$ and, let $k_0$, $k_1$, $k_2$ be some nonnegative real-valued numbers [26]. Then $Z_1$ and $Z_2$ can be decomposed as $Z_j = Y_0 + Y_j$, $j = (1, 2)$, where $Y_0$, $Y_1$ and $Y_2$ are independent gamma-distributed same scale $\lambda$ random variables with density:

$$g(Y_j) = \frac{\lambda^{k_j} y_j^{k_j-1} e^{-\lambda y_j}}{\Gamma(k_j)}, k_j > 0, \lambda > 0, j = (0, 1, 2) \tag{13}$$

Obviously, $Z_1$ and $Z_2$ are correlated in view of the shared part of frailty $Y_0$ in both $Z_1$ and $Z_2$. Further assuming equal shape parameters ($k_1 = k_2 = k$) for the distributions of $Y_1$ and $Y_2$, forces the frailties $Z_1$ and $Z_2$ to follow gamma-distributed correlated random variables [26] given by:

$$Z_j \sim \Gamma(k_0 + k, \lambda), \quad j = (1, 2) \tag{14}$$

For the standard assumption that the mean frailty of individuals is one (at the beginning of the follow-up), the following holds:

$$E[Z_1] = E[Z_2] = \frac{k_0 + k}{\lambda} = 1 \tag{15}$$

thus, the frailty variances are equal and given by:

$$V(Z) = V(Z_1) = V(Z_2) = \frac{1}{\lambda} = \sigma^2 \tag{16}$$

This leads to the correlation coefficient of $Z_1$ and $Z_2$ given by:

$$\rho = \frac{cov(Z_1, Z_2)}{\sqrt{V(Z_1)V(Z_2)}} = \frac{k_0}{k_0 + k} \tag{17}$$

Let

$$\lambda = \frac{1}{\sigma^2}; k_0 = \frac{\rho}{\sigma^2}; k = \frac{1-\rho}{\sigma^2} \tag{18}$$

then

$$Z_j \sim \Gamma(\frac{1}{\sigma^2}, \frac{1}{\sigma^2}), \quad j = (1, 2) \tag{19}$$

Now it is possible to derive the marginal likelihood function as:

$$L_{marg} = \int_0^\infty \int_0^\infty \int_0^\infty \left( (-1)^{(\delta_1+\delta_2)} \frac{\partial}{\partial t_1^{\delta_1}} e^{-(y_0+y_1)H(t_1)} \right. \tag{20}$$
$$\left. \frac{\partial}{\partial t_2^{\delta_2}} e^{-(y_0+y_2)H(t_2)} \prod_{j=0}^{2} g(y_j) \prod_{j=0}^{2} dy_j \right)$$

where $g(y_j), j = 0, 1, 2$ are the pdf given in *equation-13* with parameters given in equation-18. EM algorithm-based parameter estimation for the above bivariate correlated gamma-frailty model with equal sampling weights has already been developed by Iachine [16] and here we presented short summary of the approach. Iachine [16] derived the complete log-likelihood as if the frailties are observed. Then, showed that the complete log-likelihood structure allows to combine Cox's regression and maximum likelihood techniques to obtain parameter estimates.

**E-step**: Compute the conditional expected values of $y_j$ and $\ln(y_j)$ for $j = 0, 1, 2$ denoted by $B_j = E(y_j|data)$ and $A_j = E(\ln(y_j)|data)$ for $j = 0, 1, 2$ evaluated at the current iteration estimates.

**M-step**: Obtained new estimates for the next iteration. The derived working log-likelihood function after substituting the unobserved frailty variables by their conditional expected values such as substitute $y_j$ and $\ln(y_j)$ by $B_j$ and $A_j$ respectively and $z_{ij} = B_{i0} + B_{ij}$, for $j = 1, 2$ is given by:

$$ll_C = ll_I(\sigma^2, \rho) + ll_{II}(\beta) \tag{21}$$

where

$$ll_I(\sigma^2, \rho) = n\left(\frac{\rho-2}{\sigma^2}\ln(\sigma^2) - \ln(\Gamma(\frac{\rho}{\sigma^2})) - 2\ln\left(\Gamma(\frac{1-\rho}{\sigma^2})\right)\right) + \sum_{i=1}^{n}\left(\frac{\rho}{\sigma^2}A_{i0} + \frac{1-\rho}{\sigma^2}(A_{i1} + A_{i2}) - \frac{1}{\sigma^2}\sum_{j=0}^{2} B_{ij}\right) \tag{22}$$

$$ll_{II}(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{2}\delta_{ij}\left(\mathbf{x}'_{ij}\boldsymbol{\beta} - \ln\left(\sum_{t_{pl}\geq t_{ij}} e^{\mathbf{x}'_{pl}\boldsymbol{\beta} + \ln(\hat{z}_{pl})}\right)\right) \tag{23}$$

where $\hat{z}_{ij} = (B_{i0} + B_{ij})$. As *equation-23* coincides with Cox's partial likelihood when $\ln(\hat{z}_{ij})$ is considered as an additional Cox-like covariate with a known regression coefficient equal to one, standard procedures for the Cox regression can be used to obtain $\beta$ and $H_0(t)$ and equation-22 is maximized to obtain the new frailty parameters. These two steps are iterated until convergence.

In this paper, we have modified the above bivariate correlated gamma-frailty model estimation procedure in a way that enable us to incorporate the sampling weight. Thus, we have incorporated the sampling weight and developed the parameter estimation procedures. Integrating out the frailty variables in *equation-20* considering the four censoring possibilities i.e., ($\delta_1 = 1, \delta_2 = 1$), ($\delta_1 = 1, \delta_2 = 0$), ($\delta_1 = 0, \delta_2 = 1$) and ($\delta_1 = 0, \delta_2 = 0$) and taking the natural logarithm, the mariginal log-likelihood function can be

given by:

(24)

$$
\begin{aligned}
ll_{marg} = &\sum_{i=1}^{n}\sum_{j=1}^{2}\delta_{ij}\Big(\ln(h_0(t_{ij})) + \mathbf{x}'_{ij}\boldsymbol{\beta}\Big) \\
&- \sum_{i=1}^{n}\Big((k_0+\delta_{i1}+\delta_{i2})\ln(M_0) + (k+\delta_{i1})\ln(M_1) \\
&+ (k+\delta_{i2})\ln(M_2)\Big) + \sum\delta_1\delta_2\ln(L_I) \\
&+ \sum\delta_1(1-\delta_2)\ln(L_{II}) + \sum(1-\delta_1)\delta_2\ln(L_{III})
\end{aligned}
$$

where: $H(t) = H_0(t)e^{\boldsymbol{\beta}'\mathbf{x}}$, $M_0 = 1+\sigma^2(H(t_1)+H(t_2))$, $M_1 = (1+\sigma^2 H(t_1))$, $M_2 = (1+\sigma^2 H(t_2))$,

$$
\begin{aligned}
L_I &= \left[\frac{\rho(\rho+\sigma^2)}{{M_0}^2} + \frac{\rho(1-\rho)}{M_0}\left(\frac{1}{M_1}+\frac{1}{M_2}\right) + \frac{(1-\rho)^2}{M_1M_2}\right] \\
L_{II} &= \left[\frac{\rho}{M_0} + \frac{(1-\rho)}{M_1}\right] \\
L_{III} &= \left[\frac{\rho}{M_0} + \frac{(1-\rho)}{M_2}\right]
\end{aligned}
$$

Denote the equal weight version of the Nelson–Aalen estimator of $H_0(t)$ by:

$$
\hat{H}_0(t) = \sum_{t_{ij}\le t}\frac{d_{pl}}{\sum\limits_{t_{pl}\ge t_{ij}} e^{\mathbf{x}'_{pl}\boldsymbol{\beta}+\ln(z_{pl})}} \tag{25}
$$

where $d_{pl} = \sum\limits_{t_{pl}=t_{ij}}\delta_{pl}$ is the number of events at time $t_{ij}$. The following *equation-26*, which is the first line of *equation-24* can easily be obtained assuming frailty variables ($Z_1$ and $Z_2$) are observed and further by substituting $H_0(t)$ with its Nelson–Aalen estimator, which is given in *equation-25*.

$$
\sum_{i=1}^{n}\sum_{j=1}^{2}\delta_{ij}\left(\mathbf{x}'_{ij}\boldsymbol{\beta} - \ln\left(\sum_{t_{pl}\ge t_{ij}} e^{\mathbf{x}'_{pl}\boldsymbol{\beta}+\ln(z_{pl})}\right)\right) \tag{26}
$$

Now one can easily see that equation-26 coincides with Cox's partial likelihood with $\ln(z_{ij})$ as an additional Cox-like covariate with a known regression coefficient equals to one. This shows that given $z_{ij}$ it is possible to estimate $\boldsymbol{\beta}$; and $H_0(t)$ can be obtained using *equation-25*. However, $z_{ij}$ is not observed and thus needs to be substituted by its expected value. Similar to the mariginal log-likelihood expression, to derive the expression of the conditional expected values of the frailties, the four censoring possibilities needs to be considered. Here, we presented derivation of the expressions of the conditional expected values for the case where both survival times $(t_1, t_2)$ are uncensored i.e., $(\delta_1 = 1, \delta_2 = 1)$. The expressions for the remaining three censoring possibilities can be derived in a similar manner.

The conditional distribution of $y_0$, $y_1$, $y_2$ denoted by $f(y_0, y_1, y_2|(t_{ij}, \delta_{ij}, X_{ij}))$ can be given by:

(27)

$$
\begin{aligned}
f(y_0,y_1,y_2|(.)) &= \frac{\frac{\partial^2}{\partial t_1 \partial t_2}S(t_1,t_2|y_0,y_1,y_2,x_1,x_2)g(y_0,y_1,y_2)}{S(t_1,t_2)} \\
&= C\frac{(y_0^2+y_0y_1+y_0y_2+y_1y_2)g(y_0)g(y_1)g(y_2)}{L_{marg_I}}
\end{aligned}
$$

where $C = h_0(t_1)e^{\boldsymbol{\beta}'\mathbf{x}_1}h_0(t_2)e^{\boldsymbol{\beta}'\mathbf{x}_2}e^{-(y_0+y_1)H(t_1)-(y_0+y_2)H(t_2)}$, $g(y_j), j = 0,1,2$ are the pdf given in *equation-13* with parameters given in *equation-18* and $L_{marg_I}$ is given by:

(28)

$$
L_{marg_I} = h_0(t_1)e^{\boldsymbol{\beta}'\mathbf{x}_1}h_0(t_2)e^{\boldsymbol{\beta}'\mathbf{x}_2}{M_0}^{-k_0}{M_1}^{-k}{M_2}^{-k}\times L_I
$$

where $M_r, r = 0,1,2$ and $L_I$ are expressions given in *equation-24*. Suppose $g(a,b)$ denotes gamma distribution with shape and scale parameters are $a$ and $b$, respectively and let's introduce the following notations:

$$
\begin{aligned}
q_1 &= \frac{\rho(\rho+\sigma^2)}{M_0^2} = \frac{k_0(k_0+1)}{\lambda_0^2} \\
q_2 &= \frac{\rho-\rho^2}{M_0M_1} = \frac{k_0k}{\lambda_0\lambda_1} \\
q_3 &= \frac{\rho-\rho^2}{M_0M_2} = \frac{k_0k}{\lambda_0\lambda_2} \\
q_4 &= \frac{(1-\rho)(1-\rho)}{M_1M_2} = \frac{k_1k}{\lambda_1\lambda_2} \\
S_q &= q_1+q_2+q_3+q_4 \\
p_l &= \frac{q_l}{S_q}, l = 1,2,3,4 \\
g_{d_1}(.) &= g(k_0+2,\lambda_0)g(k,\lambda_1)g(k,\lambda_2) \\
g_{d_2}(.) &= g(k_0+1,\lambda_0)g(k+1,\lambda_1)g(k,\lambda_2) \\
g_{d_3}(.) &= g(k_0+1,\lambda_0)g(k,\lambda_1)g(k+1,\lambda_2) \\
g_{d_4}(.) &= g(k_0,\lambda_0)g(k+1,\lambda_1)g(k+1,\lambda_2)
\end{aligned} \tag{29}
$$

where $\lambda_0 = \frac{1}{\sigma^2}+H(t_1)+H(t_2)$, $\lambda_1 = \frac{1}{\sigma^2}+H(t_1)$, $\lambda_2 = \frac{1}{\sigma^2}+H(t_2)$ and $H(t) = H_0(t)e^{\boldsymbol{\beta}'\mathbf{X}}$. Using theses notations it is possible to re-write the conditional distribution of $y_0$, $y_1$, $y_2$ given in equation-26 as the sum of mixture of gamma distributed variates given below:

(30)

$$
\begin{aligned}
f(y_0,y_1,y_2|(.)) &= \frac{1}{S_q}\Big[q_1g_{d_1}(.) + q_2g_{d_2}(.) \\
&\quad + q_3g_{d_3}(.) + q_4g_{d_4}(.)\Big] \\
&= p_1g_{d_1}(.) + p_2g_{d_2}(.) + p_3g_{d_3}(.) + p_4g_{d_4}(.) \\
&= \sum_{l=1}^{4}p_lg_{d_l}(.)
\end{aligned}
$$

thus, the expected values of $E(y_j)$ denoted by $\hat{y}_j, j = 0, 1, 2$ can be given as follows:

$$\begin{aligned}\hat{y}_0 &= \sum_{l=1}^{4} p_l E_{y_0}(g_{d_l}(.)) \\ &= p_1\frac{k_0+2}{\lambda_0} + p_2\frac{k_0+1}{\lambda_0} + p_3\frac{k_0+1}{\lambda_0} + p_4\frac{k_0}{\lambda_0} \\ &= p_1\frac{(\rho+2\sigma^2)}{M_0} + (p_2+p_3)\frac{(\rho+\sigma^2)}{M_0} + p_4\frac{\rho}{M_0} \\ &= \frac{p_1(\rho+2\sigma^2) + (p_2+p_3)(\rho+\sigma^2) + p_4\rho}{M_0}\end{aligned} \tag{31}$$

$$\begin{aligned}\hat{y}_1 &= \sum_{l=1}^{4} p_l E_{y_1}(g_{d_l}(.)) \\ &= p_1\frac{k}{\lambda_1} + p_2\frac{k+1}{\lambda_1} + p_3\frac{k}{\lambda_1} + p_4\frac{k+1}{\lambda_1} \\ &= \frac{(p_1+p_3)(1-\rho) + (p_2+p_4)(1-\rho+\sigma^2)}{M_1}\end{aligned} \tag{32}$$

$$\begin{aligned}\hat{y}_2 &= \sum_{l=1}^{4} p_l E_{y_2}(g_{d_l}(.)) \\ &= p_1\frac{k}{\lambda_2} + p_2\frac{k}{\lambda_2} + p_3\frac{k+1}{\lambda_2} + p_4\frac{k+1}{\lambda_2} \\ &= \frac{(p_1+p_2)(1-\rho) + (p_3+p_4)(1-\rho+\sigma^2)}{M_2}\end{aligned} \tag{33}$$

Hence, the expected values of $Z_{i1}$ and $Z_{i2}$ denoted by $\hat{z}_{i1}$ and $\hat{z}_{i2}$ respectively can be given by:

$$\hat{z}_{ij} = \hat{y}_{i0} + \hat{y}_{ij}, \quad i = 1, \ldots, n; \quad j = 1, 2 \tag{34}$$

By substituting the expected values $\hat{z}_{ij}$ instead of $z_{ij}$ in equation-26 new estimates of $\hat{\boldsymbol{\beta}}$ can be obtained. Thus, parameter estimation by maximizing the marginal log-likelihood expression can be carried out by the inner and outer loop. Let's denote the outer loop iteration by $l$ and the inner loop by $q$.

**Inner loop**

Given a conventional values of $\hat{\boldsymbol{\beta}}_{l,q}, \hat{H}_{0_{l,q}}(t)$ and $\hat{\sigma}^2_l, \hat{\rho}_l$,

- Step 1: Obtain $\hat{z}_{ij_q}$ using *equation-34*.
- Step 2: Maximize *equation-26* to obtain $\hat{\boldsymbol{\beta}}_{l,q+1}$; and obtain $\hat{H}_{0_{l,q+1}}(t)$ using equation-25. Maximization of equation-26 given $\hat{z}_{ij_q}$ can be carried out using standard Cox PH fit procedure with $\ln(\hat{z}_{ij_q})$ as an "offset" term.
- Step 3: Iterate step 1 and 2 until the maximum absolute differences between successive estimates reach tolerance i.e., $max(|\hat{\boldsymbol{\beta}}_{l,q+1} - \hat{\boldsymbol{\beta}}_{l,q}|) < 10^{-10}$.

**Outer loop**

Given $\hat{\boldsymbol{\beta}}_{l,q+1}, \hat{H}_{0_{l,q+1}}, \hat{z}_{ij_{q+1}}$

- Maximize *equation-24* to obtain $\hat{\sigma}^2_{l+1}, \hat{\rho}_{l+1}$. Here, the first line expression should be substituted by (26).
- Iterate inner and outer loop until convergence. Convergence can be checked using absolute maximum difference of estimated frailty parameters estimated at previous and current outer iterations.

When sampling weights denoted by $w_i = w_{i1} = w_{i2}$ are incorporated, the expressions of the expected value of the frailties are changed. $M_0 = (1+\sigma^2 w(H(t_1)+H(t_2)))$, $M_1 = (1+\sigma^2 wH(t_1))$, $M_2 = (1+\sigma^2 wH(t_2))$ accordingly, $\lambda_0 = \frac{1}{\sigma^2}+w(H(t_1)+H(t_2))$, $\lambda_1 = \frac{1}{\sigma^2}+wH(t_1)$, $\lambda_2 = \frac{1}{\sigma^2}+wH(t_2)$. the expected values $\hat{y}_j, j = 0, 1, 2$ are given as follows:

$$\begin{aligned}\hat{y}_0 &= p_1\frac{k_0+2w}{\lambda_0} + p_2\frac{k_0+w}{\lambda_0} + p_3\frac{k_0+w}{\lambda_0} + p_4\frac{k_0}{\lambda_0} \\ &= p_1\frac{(\rho+2w\sigma^2)}{M_0} + (p_2+p_3)\frac{(\rho+w\sigma^2)}{M_0} + p_4\frac{\rho}{M_0} \\ &= \frac{p_1(\rho+2w\sigma^2) + (p_2+p_3)(\rho+w\sigma^2) + p_4\rho}{M_0}\end{aligned} \tag{35}$$

$$\begin{aligned}\hat{y}_1 &= p_1\frac{k}{\lambda_1} + p_2\frac{k+w}{\lambda_1} + p_3\frac{k}{\lambda_1} + p_4\frac{k+w}{\lambda_1} \\ &= \frac{(p_1+p_3)(1-\rho) + (p_2+p_4)(1-\rho+w\sigma^2)}{M_1}\end{aligned} \tag{36}$$

$$\begin{aligned}\hat{y}_2 &= p_1\frac{k}{\lambda_2} + p_2\frac{k}{\lambda_2} + p_3\frac{k+w}{\lambda_2} + p_4\frac{k+w}{\lambda_2} \\ &= \frac{(p_1+p_2)(1-\rho) + (p_3+p_4)(1-\rho+w\sigma^2)}{M_2}\end{aligned} \tag{37}$$

Hence, $\hat{z}_{i1}$ and $\hat{z}_{i2}$ can be given by:

$$\hat{z}_{ij} = \hat{y}_{i0} + \hat{y}_{ij}, i = 1, 2, ..., n; j = 1, 2 \tag{38}$$

With sampling weight, *equation-26* is changed in to:

$$\sum_{i=1}^{n}\sum_{j=1}^{2}\delta_{ij}w_i\left[\eta_{ij} - \ln\left(\sum_{t_{pl}\geq t_{ij}} w_p e^{\eta_{pl}}\right)\right] \tag{39}$$

where $\eta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \ln(\hat{z}_{ij})$. In order to obtain new estimates of the frailty parameters, using the profile likelihood approach due to [22], the hazard terms are expressed in terms of the expected frailties and the unweighted events and then substitute all to *equation-24*. Thus, similar to the equal weight correlated gamma frailty estimation procedure, the algorithm consists of an inner and outer loops.

## 3. RESULTS

A total of 908 (454 pairs) twin child deliveries were recorded in the 2016 Ethiopia Demographic and Health Survey (EDHS). The overall information on censoring and covariates included in this study are presented in Table 1.

The descriptive summaries in Table 1 shows that mother's age at childbirth, in the first category below 18

*Table 1. Descriptive summaries of variables*

| Covariates | | Child status | |
|---|---|---|---|
| | | Alive | Dead |
| | | Freq (%) | Freq (%) |
| Mother's AGE at child birth | < 18 years | 22 (50.0%) | 22 (50.0%) |
| | 18 – 35 years | 544 (70.1%) | 232 (29.9%) |
| | > 35 years | 60 (68.2%) | 28 (31.8%) |
| Residence | Urban | 117 (76.0%) | 37 (24.0%) |
| | Rural | 509 (67.5%) | 245 (32.5%) |
| Gender | Male | 311 (63.6%) | 178 (36.4%) |
| | Female | 315 (75.2%) | 104 (24.8%) |
| Among twin's birth order | First born | 335 (73.8%) | 119 (26.2%) |
| | Second born | 291 (64.1%) | 163 (35.9%) |
| Preceding birth interval | No precede sibling | 94 (64.4%) | 52 (35.6%) |
| | < 18 months | 53 (51.0%) | 51 (49.0%) |
| | 18 – 24 months | 72 (62.1%) | 44 (37.9%) |
| | > 24 months | 407 (75.1%) | 135 (24.9%) |
| Succeeding birth interval | Last born | 252 (79.2%) | 66 (20.8%) |
| | < 18 months | 53 (45.3%) | 64 (54.7%) |
| | 18 – 24 months | 66 (71.0%) | 27 (29.0%) |
| | > 24 months | 255 (67.1%) | 125 (32.9%) |

years, there were 4.8% of the study population, (of whom 50% died), 9.7% of the children were born from mothers whose age exceeds 35 years, (of whom 31.8% died); and the remaining 85.5% children belong to mothers whose age at birth was between 18 – 35 years (of whom 29.9% of them died). 17.0% of children are in urban areas of which 76% are alive. 53.9% of the total birth are male of whom 36.4% are dead. Among the second-born twin children, 35.9% are dead, while only 26.2% are dead among first-born ones. Regarding preceding child's birth interval, 16.1% of the children were first born of whom 64.4% are alive and 35.6% are dead; 11.4% were born before their older sibling reaches the age of 18 months of whom 49% are dead. Regarding the succeeding birth interval, 35.0% were last born or no child after of whom 79.2% are alive and 20.8% are dead; and 12.9% were having a succeeding birth interval below 18 months of which 54.7% are dead.

The results of the log rank test given in Table 2 indicate that the covariates sex, among twins' birth order, preceding birth interval and succeeding birth interval were found out to be highly significant. However, residence and mother's age at birth were not significant at 5% level of significance.

According to Childs et al. [6], each child in a family has a proper susceptibility to infection, independently of his family members. In addition, inside the common global family behavioral factor, parents may adopt a slightly different prenatal and neonatal attitude from one child to the next in the family. Moreover, it is apparent that twins share a common environmental effect. This is because they are usually growing up in the same household environment and their parents

*Table 2. Results of the log-rank test of covariates*

| Covariates | Test statistic | Df | *p -value* |
|---|---|---|---|
| Age of mother at birth | 1.98 | 2 | 0.370 |
| Residence | 1.06 | 1 | 0.302 |
| Sex | 8.58 | 1 | 0.003* |
| Among twins' birth order | 10.04 | 1 | 0.001* |
| Preceding birth interval | 12.9 | 3 | 0.004* |
| Succeeding birth interval | 18.9 | 3 | <0.001* |

*Table 3. Parameter estimates of UNIVARIATE gamma frailty model with and without sampling weight*

| Covariates | Unweighted | Weighted |
|---|---|---|
| | Coef (SE) | Coef (SE) |
| ***Sex*** | | |
| Male | 0.4660 (0.1248)* | 0.7947 (0.1767)* |
| Female (Ref) | | |
| ***Among twins' birth order*** | | |
| Second born | 0.3475 (0.1206)* | 0.5285 (0.1686)* |
| First born (Ref) | | |
| ***Preceding birth interval*** | | |
| No preced sibling | 0.3341 (0.1647)* | 0.7054 (0.2276)* |
| <18 months | 0.6886 (0.1706)* | 0.8162 (0.2546)* |
| 18–24 months | 0.3155 (0.1761) | 0.1385 (0.2635) |
| >24 months (Ref) | | |
| ***Succeeding birth interval*** | | |
| Last born | −0.4370 (0.1539)* | −0.7201 (0.2037)* |
| < 18 months | 0.5417 (0.1586)* | 0.4226 (0.2407) |
| 18–24 months | −0.0671 (0.2135) | −0.0771 (0.3082) |
| >24 months (Ref) | | |
| $\sigma^2$ | 5e-08 | 1.44* |
| l-loglik | −1836.9 | −1848.2 |

are more likely to adopt similar child care behavior. Thus, it is appropriate to assess and test for the presence of unobserved heterogeneity at individual and pair (twin) levels using statistical models that can take into the presence of correlation and unobserved heterogeneity into account. Hence, in this study, we have used appropriate gamma frailty models sequentially, starting from univariate then shared and finally correlated frailty models. The results are given in Tables 3–5.

Although there are minor differences in the parameter estimates of coefficients, the results of all the fitted survival models given in Tables 3, 4 and 5 showed gender, twin's birth order, preceding birth interval, and succeeding birth interval were found to be significantly associated with twin infant mortality at 5% level of significance. Table 3 revealed that incorporation of sampling weight changes the estimate of the individual heterogeneity parameter $\sigma^2$ from zero (5e-08) to 1.44. The likelihood ratio test of the frailty parameter ($H_0 : \sigma^2 = 0$ vs $H_1 : \sigma^2 > 0$) is rejected with p-value

*Table 4. Parameter estimates of SHARED gamma frailty model with and without sampling weight*

| Covariates | Unweighted Coef (SE) | Weighted Coef (SE) |
|---|---|---|
| ***Sex*** | | |
| Male | 0.5930 (0.1659)* | 0.6757 (0.1648)* |
| Female (Ref) | | |
| ***Among twins' birth order*** | | |
| Second born | 0.4209 (0.1251)* | 0.4792 (0.1225)* |
| First born (Ref) | | |
| ***Preceding birth interval*** | | |
| No preced sibling | 0.4262 (0.2781) | 0.6112 (0.2680)* |
| <18 months | 0.8972 (0.3082)* | 0.8408 (0.2986)* |
| 18–24 months | 0.4188 (0.3004) | 0.2413 (0.3006) |
| >24 months (Ref) | | |
| ***Succeeding birth interval*** | | |
| Last born | −0.5748 (0.2393)* | −0.6816 (0.2323)* |
| < 18 months | 0.6608 (0.2941)* | 0.4215 (0.2874) |
| 18–24 months | −0.0429 (0.3360) | −0.0742 (0.3491) |
| >24 months (Ref) | | |
| $\sigma^2$ | 2.33* | 1.38* |
| I-loglik | −1794.9 | −1829.5 |

*Table 5. Parameter estimates of CORRELATED gamma frailty model with and without sampling weight*

| Covariates | Unweighted Coef (SE) | Weighted Coef (SE) |
|---|---|---|
| ***Sex*** | | |
| Male | 0.5930 (0.1655)* | 0.6757 (0.1635)* |
| Female (Ref) | | |
| ***Among twins' birth order*** | | |
| Second born | 0.4209 (0.1247)* | 0.4792 (0.1225)* |
| First born (Ref) | | |
| ***Preceding birth interval*** | | |
| No preced sibling | 0.4262 (0.2773) | 0.6112 (0.2598)* |
| <18 months | 0.8972 (0.3078)* | 0.8408 (0.2868)* |
| 18–24 months | 0.4188 (0.2995) | 0.2413 (0.2910) |
| >24 months (Ref) | | |
| ***Succeeding birth interval*** | | |
| Last born | −0.5748 (0.2380)* | −0.6816 (0.2268)* |
| < 18 months | 0.6608 (0.2935)* | 0.4215 (0.2753) |
| 18–24 months | −0.0432 (0.3357) | −0.0742 (0.3394) |
| >24 months (Ref) | | |
| $\sigma^2$ | 2.33* | 1.38* |
| $\rho$ | 1 | 1 |
| I-loglik | −1794.915 | −1829.38 |

*Table 6. Comparison of weighted Univariate versus Shared gamma frailty model*

| | Univariate | Shared |
|---|---|---|
| I Log-likelihood | −1848.227 | −1829.451 |
| AIC | 3667.8842 | 3596.843 |
| BIC | 4538.261 | 4195.533 |

less than 0.001. This indicates that we cannot ignore the presence of unobserved individual heterogeneity in the study population (at the individual level). However, the presence of heterogeneity at the individual level does not indicate the presence of correlation within groups. As Wienke [25] stated, the estimate of the variance of the frailty from univariate data may have nothing to do with association. Univariate frailty variance is interpreted as a measure of unobserved heterogeneity in the study population.

According to the results in Table 4, there is a reduction in the frailty parameter together with its standard error under the weighted frailty model. The weighted shared frailty model has lower standard error in the frailty variable, which shows that it is indeed a more appropriate choice of model. A similar result has been reported recently by Wang [23]. Moreover, the inclusion of survey weights in the shared frailty model has resulted in the reduction of standard errors of many of the parameters estimates in the model. Due to this, some of the non-significant variables, such as "Preceding birth interval", have become significant under the weighted shared frailty model (Table 4).

The likelihood ratio test of the frailty parameter ($H_0 : \sigma^2 = 0$ vs $H_1 : \sigma^2 > 0$) is rejected with p-value less than 0.001. This indicates that there is significant heterogeneity between pairs and the clustering effect was important in modeling the hazard function. Since both univariate and shared gamma frailty models are the special cases of correlated gamma frailty model, it is possible to test hypotheses about the appropriateness of the models and compare which model fits the data better. Usually correlated gamma frailty model is fitted to assess the genetic effect. To this end, we need the zygosity information that would enable us to compare the correlation between MZ and DZ twins. However, here the fit is used to test hypotheses about the appropriateness of the shared frailty model.

As shown in Table 5, the estimated frailty correlation parameter is 1. As a result, the estimated covariate coefficients are almost equal to the estimated covariate coefficients of shared gamma fit given in Table 4. In addition, the likelihood ratio test statistic for the hypothesis ($H_0 : \rho = 1$ vs $H_1 : \rho < 1$) is insignificant. Thus, we have no evidence to reject that the model is a shared frailty model. Furthermore, Table 6 showed that the shared gamma frailty model has the highest log-likelihood and minimum AIC and BIC values, indicating that this model fits the data better than the univariate gamma model.

The hazard ratio estimates of the shared model given in Table 4 indicated that the estimated hazard ratio of a male twin infant (i.e. an infant of a twin birth) is 1.965 (95% CI: 1.418–2.728) implying that the risk of dying for a male twin infant is 96.5% higher than a female twin infant, controlling for the other covariates in the model.

The estimated hazard ratio of a second-born twin infant is 1.614 (95% CI: 1.271–2.055) implying that the risk of dying for a second-born twin infant is 61.4% more likely than a first-born twin infant (reference group), controlling for the other covariates in the model.

The estimated hazard ratio of an infant of twin who born before the older sibling reaches the age of 18 months is 2.318 (95% CI: 1.112–4.869). Thus, the hazard rate of an infant twin born before the older sibling reaches the age of 18 months is 2.318 times higher than an infant of twin born when the older sibling's age exceeds 24 months (reference group), controlling for other covariates in the model.

The estimated hazard ratio of an infant twin who has no succeeding sibling is 0.505 (95% CI: 0.316–0.807). This implies that an infant of twin who has no succeeding sibling had a 49.5% lower hazard (risk) of death than an infant of twin who has a younger sibling born after 24 months, controlling for the other covariates in the model.

## 4. DISCUSSIONS

Since we aim to investigate the survival of twin infants survey data, it is expected to employ a model that can take the presence of correlation, unobserved heterogeneity and survey weights into account. Thus, we selected the weighted gamma frailty models as the most appropriate model to fit to this type of data. As shown in Table 4, the variance of the random effect estimated from the weighted shared gamma frailty model is significant at 5% level of significance. This indicates that there is significant heterogeneity between pairs and the correlation within pairs cannot be ignored and clustering effect was important in modeling the hazard function.

This study showed that the risk of dying for a male twin infant is higher than for a female child. In agreement with this result, a study on child mortality showed that the risk of dying for a male child is higher than a female child [3]. This study also revealed the significant risk of a first born infant of twin birth. Similar to this result, a child mortality study showed that first-born children experience low survival compared to those children who have higher birth order [1]. However, a study in Burkina Faso showed that the variable is statistically insignificant [2]. Most importantly we have found the significance of among twins' birth order in infant mortality. The risk of dying for a second-born infant of twin birth is statistically higher than a first-born infant of twin birth.

Unlike a study that showed children residing in urban areas have a better chance of survival than those residing in rural areas [1], this study found out that the place of residence is not significant. In addition, the result of this study disagreed with studies that showed children born from women at youngest and oldest age are subject to high risk of death [16, 19], this study found out the variable is not significantly associated with survival of infant twins.

The current study also showed that infants of twin birth who were born before the older sibling reaches the age of 18 months experience low survival compared to those infants of twin birth who have more than 24 months spacing. Several studies conducted on determinants of child mortality found similar results [2, 14, 16].

## 5. CONCLUSIONS

Integrating the dependence from twin birth with the advanced statistical weighted gamma frailty models, the results of this study showed that the main significant factors associated with twin infant mortality in Ethiopia are gender, twin's birth order, preceding birth interval, and succeeding birth interval. The significant effect of the birth spacing of the previous and succeeding sibling on the survival chance of infants of twin birth indicates that efforts have to be exerted to educate the public about family planning and birth spacing, mainly in resource-poor countries like Ethiopia. Further, male twins have a lower chance of survival compared to female twin children. Those twins who were born before their sibling reaches 18 months experience low survival compared to those twins who have more than 24 months spacing. The risk of infant mortality for a second-born twin infant is higher than a first-born twin infant. This study also showed the presence of significant heterogeneity between pairs. Overall, the different modeling approaches in this study revealed the significance of including a random effect in the model mainly to take into account the correlation of event times among twins. The inclusion of survey weights in the shared frailty model has improved the precision of model parameter estimates.

## CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## AUTHOR'S CONTRIBUTIONS

All Authors contribute equally.


## FUNDING

This work is partially based upon research supported by the National Research Foundation, South Africa (South Africa DST-NRF-SAMRC SARChI Research Chair in Biostatistics, Grant number 114613). Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF.

## ACKNOWLEDGEMENT

We thank the two anonymous reviewers for their careful reading and insightful comments and suggestions. Our sincere thanks are also due to all the staff members who are involved in the data collection of the 2016 Ethiopia Demographic and Health Survey.

*Received 2 August 2021*

## REFERENCES

[1] Balk, D., Pullum, T., Storeygard, A., Greenwell, F., and Neuman, M. (2004). A spatial analysis of childhood mortality in west africa. *Population, Space and Place*, 10(3):175–216.

[2] Becher, H., Müller, O., Jahn, A., Gbangou, A., Kynast-Wolf, G., and Kouyaté, B. (2004). Risk factors of infant and child mortality in rural burkina faso. *Bulletin of the World Health Organization*, 82:265–273.

[3] Bereka, S. G., Habtewold, F. G., and Nebi, T. D. (2017). Under-five mortality of children and its determinants in ethiopian somali regional state, eastern ethiopia. *Health Science Journal*, 11(3):1.

[4] Borgan, Ø. (2014). N elson–a alen estimator. *Wiley StatsRef: Statistics Reference Online*.

[5] Central Statistical Agency (CSA) and ICF. (2016). Ethiopia demographic and health survey 2016. *Addis Ababa, Ethiopia, and Rockville, Maryland, USA*.

[6] Childs, B., Moxon, E. R., and Winkelstein, J. A. (2002). *Genetics and Infectious Diseases: In The Genetic Basis of Common Diseases, eds. R. A. King, J. I. Rotter, and A. G. Motulsky*. Oxford University Press, Oxford.

[7] Duchateau, L. and Janssen, P. (2008). *The frailty model*. Springer Science & Business Media. MR2723929

[8] Guo, G. and Grummer-Strawn, L. M. (1993). Child mortality among twins in less developed countries. *Population Studies*, 47(3):495–510.

[9] Guo, G. and Rodriguez, G. (1992). Estimating a multivariate proportional hazards model for clustered data using the em algorithm, with an application to child survival in guatemala. *Journal of the American Statistical Association*, 87(420):969–976.

[10] Hanagal, D. D. and Pandey, A. (2015). Gamma frailty models for bivariate survival data. *Journal of Statistical Computation and Simulation*, 85(15):3172–3189. MR3369612

[11] Hanagal, D. D. and Sharma, R. (2013). Modeling heterogeneity for bivariate survival data by shared gamma frailty regression model. *Model Assisted Statistics and Applications*, 8(2):85–102. MR3369612

[12] Justesen, A. and Kunst, A. (2000). Postneonatal and child mortality among twins in southern and eastern africa. *International journal of epidemiology*, 29(4):678–683.

[13] Kifle, Y., Janssen, P., Yewhalaw, D., Speybroeck, N., and Duchateau, L. (2013). Coping with time and space in modelling malaria incidence: a comparison of survival and count regression models. *Statistics in medicine*, 32(18):3224–3233. MR3073795

[14] Koissi, M.-C. and Högnäs, G. (2005). Using winbugs to study family frailty in child mortality, with an application to child survival in ivory coast. *African population studies*, 20(1).

[15] Korn, E. L. and Graubard, B. I. (1995). Examples of differing weighted and unweighted estimates from a sample survey. *The American Statistician*, 49(3):291–295.

[16] Lachine, I. A. (1995). *Parameter estimation in the bivariate correlated frailty model with observed covariates via the EM-algorithm*. Odense Universitet, Center for Helsetjenesteforskning og Socialpolitik.

[17] Madise, N., Stephenson, R., Holmes, D., and Matthews, Z. (2003). Impact of estimation techniques on regression analysis: an application to survey data on child nutritional status in five african countries.

[18] Monden, C. W. and Smits, J. (2017). Mortality among twins and singletons in sub-saharan africa between 1995 and 2014: a pooled analysis of data from 90 demographic and health surveys in 30 countries. *The Lancet Global Health*, 5(7):e673–e679.

[19] Sastry, N. (1997). A nested frailty model for survival data, with an application to the study of child survival in northeast brazil. *Journal of the American Statistical Association*, 92(438):426–435. MR3186029

[20] Skinner, C. J., Holt, D., and Smith, T. F. (1989). *Analysis of complex surveys*. Wiley. MR1049386

[21] Smits, J. and Monden, C. (2011). Twinning across the developing world. *PloS one*, 6(9):e25239.

[22] Therneau, T. M. and Grambsch, P. M. (2000). The cox model. In *Modeling survival data: extending the Cox model*, pages 39–77. Springer. MR1774977

[23] Wang, J. (2019). Weighted estimation for multivariate shared frailty models for complex surveys. *Lifetime data analysis*, 25(3):469–479. MR3960835

[24] WHO (2016). Under five mortality rates.

[25] Wienke, A. (2010). *Frailty models in survival analysis*. CRC press. MR2682965

[26] Yashin, A. I., Vaupel, J. W., and Iachine, I. A. (1995). Correlated individual frailty: an advantageous approach to survival analysis of bivariate data. *Mathematical population studies*, 5(2):145–159.

Yehenew G. Kifle
Department of Mathematics and Statistics
University of Maryland Baltimore County
USA
E-mail address: yehenew@umbc.edu

Ding-Geng Chen
College of Health Solutions
Arizona State University
Phoenix, AZ
USA
Department of Statistics
University of Pretoria
South Africa
E-mail address: dinchen@asu.edu

Mesfin T. Haileyesus
Department of Statistics
Dire Dawa University
Ethiopia
E-mail address: mesfin.tsegaye@ddu.edu.et