APPLICATION NOTE

A Novel Outlier Statistic in Multivariate Survival Models and its Application to Identify Unusual Under-Five Mortality Sub-Districts in Malawi

Tsirizani M. Kaombe [a] and Samuel O.M. Manda [a,b,c]

[a] Department of Mathematical Sciences, Faculty of Science, University of Malawi, Zomba, Malawi; [b] Department of Statistics, University of Pretoria, Pretoria, South Africa; [c] Biostatistics Research Unit, South African Medical Research Council, Pretoria, South Africa.

ARTICLE HISTORY
Compiled July 8, 2022

ABSTRACT
Although under-five mortality (U5M) rates have declined worldwide, many countries in sub-Saharan Africa still have much higher rates. Detection of subnational areas with unusually higher or lower U5M rates could support targeted high impact child health interventions. We propose a novel group outlier detection statistic for identifying areas with extreme U5M rates under a multivariate survival data model. The performance of the proposed statistic was evaluated through a simulation study. We applied the proposed method to an analysis of child survival data in Malawi to identify sub-districts with unusually higher or lower U5M rates. The simulation study showed that the proposed outlier statistic can detect unusual high or low mortality groups with a high accuracy of at least 90%, for datasets with at least 50 clusters of size 80 or more. In the application, at most 7 U5M outlier sub-districts were identified, based on the best fitting model as measured by the Akaike information criterion (AIC).

KEYWORDS
Clustered data; Multivariate Cox PH model; Outlier statistic; Under-five mortality; Outlying sub-districts.

1. Introduction

The under-five mortality (U5M) rate is the probability of a child dying before the age of five years, expressed per 1,000 live births. The rate is an important indicator of child health and health care, well-being, and socio-economic status of a population [11, 18, 29]. The estimates of U5M rate are critical for monitoring national and global health strategies as well as measuring overall progress towards international goals, such as the Sustainable Development Goals (SDGs). In particular, target 3.2.1 of SDGs is concerned with reducing U5M to less than 25 per 1,000 live births by 2030 [20].

Despite most countries worldwide experiencing substantial reductions in U5M rates, falling by 59 percent from 91 per 1000 live births in 1990 to 38 per 1000 in 2019, the declines have not been uniform. For example, over the same period, the

CONTACT: Tsirizani M. Kaombe. Email: tkaombe@unima.ac.mw

U5M rate in sub-Saharan Africa dropped by 56 percent [9, 11, 29]. The global burden of under-five deaths is also heavily concentrated in sub-Saharan Africa, accounting for 2.8 million (53 percent) of the 5.2 million under-five deaths in 2019. Central and Southern Asia accounted for 1.5 million (29 percent) of the deaths in the same year [11, 33]. Within the sub-Saharan African region, there is considerable variation between countries, with Nigeria, the Democratic Republic of Congo, and Ethiopia accounting for the most burden [3, 5, 26, 33, 34].

The country-level variations in U5M rates mask within-country differences. Several studies have found substantial within-country variations in U5M rates due to differences in socio-economic and environmental factors [8, 26, 30]. Most of these studies have employed spatial statistical models to highlight subnational areas with similar or dissimilar U5M rates, often ignoring possible correlations in the survival times of the children in the same area. Instead of spatial statistical mapping models, we appropriately account for the possible dependence between survival times of children in the same subnational area by using multivariate Cox proportional hazards regression. A standard multivariate Cox regression analysis uses a shared frailty model, which incorporates group (cluster)-specific random effects to account for unmeasured cluster characteristics that impact on the survival outcomes [1, 12, 15, 16, 22, 23].

Even though the analysis of multivariate survival data to account for dependence between event times is implemented in most statistical computer packages, cluster outlier statistics for the models have not been well developed. In this study, we propose a novel method for detecting clusters with unusually larger or smaller survival times based on a multivariate Cox regression model. This is done by extending outlier statistics that have been developed for linear mixed-effects models. We evaluate the performance of the derived outlier statistic through a simulation study. We illustrate the usefulness of our proposed outlier statistic by an application to child survival data, that were collected as part of the 2015-16 Malawi Demographic and Health Survey (2015-2016 MDHS), to identify sub-districts with unusually higher or lower under-five mortality rates. The identified U5M outlier sub-districts could be used by public health policy makers to improve local evidence-informed policy in child health interventions.

The next section presents the shared frailty model for multivariate survival data and the proposed group outlier statistic. This is followed by a simulation study and analysis of child survival data. The paper ends with a conclusion of findings.

2. Methods

2.1. Multivariate survival model

Suppose subjects are nested in one of the M clusters. A Cox proportional hazards (PH) model with mixed effects [1] has the conditional hazard function at time t given as:

$$\lambda_{ij}(t_{ij}|\beta, b_i) = \lambda_0(t)exp(X_{ij}^T\beta + Z_i^T b_i), \tag{1}$$

for the j-th subject, $(j = 1, 2, ..., n_i)$ in cluster i, $(i = 1, 2, ..., M)$, where t_{ij} is the

observed event-time of a survival time variable T_{ij}, X_{ij} is a $p \times 1$ covariate vector for ij-th subject, β is a $p \times 1$ coefficient vector, Z_i is a $q \times 1$ vector of cluster-level covariates, and b_i is a $q \times 1$ vector of random coefficients. The parameter $\lambda_0(t)$ is the baseline hazard at time t. The random effect vectors are assumed to be normally distributed, i.e., $b_i \sim MVN(\mathbf{0}, \mathbf{D})$, with $\mathbf{D}$ as $q \times q$ covariance matrix. In this paper, we assume that $Z_i = 1$, so that $b_i \sim N(0, \sigma^2)$.

Estimates of the unknown parameters in the Cox model (1) are obtained by maximising the partial likelihood as opposed to the usual likelihood [4]. For this, we introduce the notion of a risk set and exposure to experience the event. Suppose $Y_{ij}(t)$ is an indicator function such that $Y_{ij}(t) = 1$ if subject ij is under observation and at risk to the event at time t, and 0 otherwise. Then, the contribution to the conditional partial likelihood by cluster i is given by:

$$L_i(\beta|b_i, t_{ij}, X_{ij}, Z_i) = \prod_{j=1}^{n_i} \left[\frac{exp(X_{ij}^T\beta + b_i)}{\sum_{r_{ij}(t_{ij})} Y_{ij}(t_{ij}) exp(X_{ij}^T\beta + b_i)} \right]^{\delta_{ij}}, \tag{2}$$

where $r_{ij}(t_{ij})$ represents the risk set at time t_{ij} and δ_{ij} is the censoring indicator for subject ij. The full joint partial likelihood function is the product of the conditional likelihood (2) over all clusters and the densities of cluster random effects, given by:

$$\begin{aligned} L(\beta, \sigma^2) &= L_i(\beta|b_i, t_{ij}, X_{ij}, Z_i) \times \prod_{i=1}^{M} f(b_i|\sigma^2) \\ &= \prod_{i=1}^{M} \prod_{j=1}^{n_i} \left[\frac{exp(X_{ij}^T\beta + b_i)}{\sum_{r_{ij}(t_{ij})} Y_{ij}(t_{ij}) exp(X_{ij}^T\beta + b_i)} \right]^{\delta_{ij}} \times \prod_{i=1}^{M} \left[\frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{M} b_i^2 \right) \right]. \end{aligned} \tag{3}$$

Since the random effects are not observed directly, estimation is based either on the marginal partial likelihood after integrating out the random effects [15], the EM algorithm [17], or the penalized joint partial likelihood techniques [6, 23]. Alternatively, the parameters of the conditional Cox PH model are estimated by using Bayesian estimation procedure using Markov Chain Monte Carlo (MCMC) technique [16, 22].

2.2. Previous work on outliers in linear mixed-effects models

Identification of outliers is mostly assessed by analysing the distribution of post-estimation statistics, such as the model's fitted value or residual [38]. Then in the linear mixed-effects model, $\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}b + \epsilon$, where ϵ is stacked vector of residual (error) vectors from each cluster. Under the standard normal assumption $\epsilon \sim N(\mathbf{0}, \sigma_\epsilon^2 I)$ and $\mathbf{b} = (b_1, b_2, ..., b_M)^T$, outliers are examined by assessing the distribution of residuals $\hat{\mathbf{e}}$, given by:

$$\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\hat{\beta} - \mathbf{Z}\hat{b}, \tag{4}$$

where $\hat{\beta}$ and $\hat{b}$ are estimators of fixed and random effects, respectively, and $\hat{\mathbf{e}} = (\hat{e}_{11}, ..., \hat{e}_{1n_1}, ..., \hat{e}_{M1}, ..., \hat{e}_{Mn_M})^T$. It has been shown that the covariance of residual $\hat{\mathbf{e}}$ in equation (4) is not equal to covariance of the error term ϵ, causing the

residual (4) to have non-normal distribution [35]. The standardised form of (4) is instead used to determine univariate outliers.

In linear mixed-effects model, identification of outlying groups could simply rely on assessing the distribution of standardised residual g_{ij} plotted against clusters [10], given by:

$$g_{ij} = \hat{e}_{ij}/stdev(\hat{e}_{ij}), \tag{5}$$

where $stdev(\hat{e}_{ij})$ is standard deviation of the residual, with j indexing subjects and i clusters. Clusters with highly skewed standardised residuals (5) compared to others, are considered outliers to the linear mixed-effects model [10].

In some cases, plots of the standardised residuals (5) against clusters highly overlap across the clusters. Then, it becomes difficult to determine outright outlier clusters to the mixed model, which is a major setback of method (5). Outlier detection methods that can identify unusual clusters of data are crucial in studying how behaviours of subjects in the clusters affect the modelling. However, the method (5) uses univariate residuals to examine group outliers in the linear mixed-effects model. By applying the above discussed concepts from the linear mixed model, we derive, in the next section, a statistic for assessing outlying groups of observations to the multivariate survival model (1) and apply the method on child mortality data in Malawi.

2.3. Proposed outlier statistic for multivariate survival data

By extending the work of Therneau et al. [27] on residuals for univariate survival data, and considering l-th subject in the risk set r_{il} in i-th cluster, a counterpart residual to (4) for single observations in the survival mixed model (1) is defined as [7]:

$$m(t_{il}) = N(t_{il}) - \int_0^{t_{il}} Y_{il}(s) exp(X_{il}^T \hat{\beta} + Z_i^T \hat{b}_i) d\hat{\Lambda}_0(s), \tag{6}$$

where $N(t_{il})$ is a counting process, indicating number of observed events experienced over time t_{il}; $Y_{il}(t)$ is 0-1 process, indicating whether ij-th subject is at risk at time t_{ij}; and $\hat{\Lambda}_0(s)$ is the cumulative baseline hazard function. The residual (6) is called martingale because of its relation with a counting process. It is interpreted as the difference over $[0, t]$ between the observed and predicted number of events at each time point [27].

Assuming time-independent covariates in model (1), the extended residual (6) sim-

plifies to:

$$m(t_{il}) = \delta_{il} - \hat{\Lambda}_0(t)exp(X_{il}^T\hat{\beta} + Z_i^T\hat{b}_i)$$

$$\Rightarrow \begin{bmatrix} m(t_{11}) \\ \vdots \\ m(t_{1n_1}) \\ m(t_{21}) \\ \vdots \\ m(t_{2n_2}) \\ \vdots \\ m(t_{M1}) \\ \vdots \\ m(t_{Mn_M}) \end{bmatrix} = \begin{bmatrix} \delta_{11} - \hat{\Lambda}_0(t)exp(X_{11}^T\hat{\beta} + Z_1^T\hat{b}_1) \\ \vdots \\ \delta_{1n_1} - \hat{\Lambda}_0(t)exp(X_{1n_1}^T\hat{\beta} + Z_1^T\hat{b}_1) \\ \delta_{21} - \hat{\Lambda}_0(t)exp(X_{21}^T\hat{\beta} + Z_2^T\hat{b}_2) \\ \vdots \\ \delta_{2n_2} - \hat{\Lambda}_0(t)exp(X_{2n_2}^T\hat{\beta} + Z_2^T\hat{b}_2) \\ \vdots \\ \delta_{M1} - \hat{\Lambda}_0(t)exp(X_{M1}^T\hat{\beta} + Z_M^T\hat{b}_M) \\ \vdots \\ \delta_{Mn_M} - \hat{\Lambda}_0(t)exp(X_{Mn_M}^T\hat{\beta} + Z_M^T\hat{b}_M) \end{bmatrix}. \tag{7}$$

Like in univariate survival model, the residual (7) is negatively-skewed, because $\delta_{il} \in [0, 1]$, while $\hat{\Lambda}_0(t)exp(X_{il}^T\hat{\beta} + Z_i^T\hat{b}_i)$ has values in the interval $[0, \infty)$. Hence, it can detect outlying subjects who failed too late, but not those who failed too early [7, 27].

The standardised version of residual (7) is the deviance residual, which measures disagreement between an element of the log-likelihood of the fitted model and the corresponding point that would result if each observation were fitted exactly [25, 27], and it is given by:

$$d_{il} = sgn(m(t_{il}))[-2(m(t_{il}) + \delta_{il}log(\delta_{il} - m(t_{il})))]^{\frac{1}{2}}$$

$$\Rightarrow \begin{bmatrix} d_{11} \\ \vdots \\ d_{1n_1} \\ d_{21} \\ \vdots \\ d_{2n_2} \\ \vdots \\ d_{M1} \\ \vdots \\ d_{Mn_M} \end{bmatrix} = \begin{bmatrix} sgn(m(t_{11}))[-2(m(t_{11}) + \delta_{11}log(\delta_{11} - m(t_{11})))]^{1/2} \\ \vdots \\ sgn(m(t_{1n_1}))[-2(m(t_{1n_1}) + \delta_{1n_1}log(\delta_{1n_1} - m(t_{1n_1})))]^{1/2} \\ sgn(m(t_{21}))[-2(m(t_{21}) + \delta_{21}log(\delta_{21} - m(t_{21})))]^{1/2} \\ \vdots \\ sgn(m(t_{2n_2}))[-2(m(t_{2n_2}) + \delta_{2n_2}log(\delta_{2n_2} - m(t_{2n_2})))]^{1/2} \\ \vdots \\ sgn(m(t_{M1}))[-2(m(t_{M1}) + \delta_{M1}log(\delta_{M1} - m(t_{M1})))]^{1/2} \\ \vdots \\ sgn(m(t_{Mn_M}))[-2(m(t_{Mn_M}) + \delta_{Mn_M}log(\delta_{Mn_M} - m(t_{Mn_M})))]^{1/2} \end{bmatrix}. \tag{8}$$

The logarithm term in (8) inflates values of martingale residual (7) that are close to 1, while the square root contracts large negative values of (7), causing the deviance residual (8) to be symmetric about zero in each cluster [27]. The pattern of values of the deviance residual (8) in each cluster will give a descriptive overview about clusters that require further investigation. That is, clusters with skewed distribution of residual (8), compared to others, will be candidates for outlier investigations, as the standardised residual (5) for linear mixed-effects model is used [10].

In examining outlying cluster of observations, the extended deviance residual (8) and its covariance become useful quantities. We, therefore, propose a statistic computed from the ratio of within-cluster variance of the deviance residual (8) to between-cluster variance, to study outlying clusters to the model (1). If observations in the model (1) were independent, the total variation of deviance residual, d_{il} in equation (8) would have been the sum of within-cluster variation and between-cluster variation, given by:

$$\frac{\sum_{i=1}^{M}\sum_{j=1}^{n_i}(d_{ij}-\bar{\bar{d}})^2}{n-1}=\frac{\sum_{i=1}^{M}\sum_{j=1}^{n_i}(d_{ij}-\bar{d}_i)^2}{n-M}+\frac{\sum_{i=1}^{M}n_i(\bar{d}_i-\bar{\bar{d}})^2}{M-1}, \tag{9}$$

where $\bar{\bar{d}}=\frac{\sum_{i=1}^{M}\sum_{j=1}^{n_i}d_{ij}}{n}$ is the grand mean of the deviance residual d_{ij}; $\bar{d}_i=\frac{\sum_{j=1}^{n_i}d_{ij}}{n_i}$ is the mean of d_{ij} for any fixed i; $n=n_1+n_2+...+n_M$ is number of subjects in entire dataset.

However, the within-cluster correlations of observations in model (1) will influence a biased estimate of overall variance of d_{ij} in equation (9) for the entire dataset. Since the clusters are independent, and assuming conditional independence of observations in a cluster due to a shared random effect, the separate within-cluster variance of d_{ij} will be conditionally unbiased estimate of variance for the concerned i-th cluster. These cluster variabilities of d_{ij} will consequently measure how distant the survival times of subjects in each cluster are from the fitted survival curve. Therefore, the proposed group outlier statistic for model (1) is a $M\times 1$ vector, denoted by $\mathbf{k}$, which is the ratio of within-cluster to between-cluster variances of d_{il}, given by:

$$\begin{aligned}\mathbf{k}&=\frac{1}{L}(k_1,...,k_M)^T\\&=\frac{1}{L}\left(\frac{\sum_{j=1}^{n_1}(d_{1j}-\bar{d}_1)^2}{n_1-1},...,\frac{\sum_{j=1}^{n_M}(d_{Mj}-\bar{d}_M)^2}{n_M-1}\right)^T,\end{aligned} \tag{10}$$

where $L=\frac{\sum_{i=1}^{M}n_i(\bar{d}_i-\bar{\bar{d}})^2}{M-1}$ is the between-cluster variance of deviance residual d_{ij}.

Since model (1) is expected to fit observations in all available clusters, the proposed outlier statistic (10) will separate homogeneous clusters that span the survival curve from outlying clusters whose observations are not necessarily close to the survival curve [24]. The small values of the statistic $\mathbf{k}$ will correspond to well-fitted clusters of observations, that is, those units that closely span the fitted survival curve. While large values of (10) will correspond to clusters whose observations have been poorly-fitted by the model (1), and hence outliers.

We explored properties of $k_i=f(K_i,L)=K_i/L$, where K_i is the within-cluster variance component of the proposed statistic (10). Clearly, $k_i\in[0,\infty)$ and it is a non-linear function, since K_i and L, being variances, have support $[0,\infty)$. A common method to estimate the expected value of a ratio estimator is through second order Taylor series expansion about $\mu=(\mu_{k_i},\mu_l)$ [28]. Thus, the expected value of k_i, denoted

by $E(k_i)$, is given by:

$$\begin{aligned} E(k_i) &= E(K_i/L) \\ &\approx f(\mu) + \frac{1}{2}\left[f''_{k_i k_i}(\mu)Var(K_i) + 2f''_{lk_i}(\mu)Cov(L,K_i) + f''_{ll}(\mu)Var(L)\right] \\ &= \frac{\mu_{k_i}}{\mu_l} - \frac{1}{\mu_l^2}Cov(K_i,L) + \frac{\mu_{k_i}}{\mu_l^3}Var(L), \end{aligned} \tag{11}$$

where $f(\mu) = \mu_{k_i}/\mu_l$, $f''_{k_i k_i}(\mu) = 0$, $f''_{lk_i}(\mu) = -1/(\mu_l)^2$, and $f''_{ll}(\mu) = 2\mu_{k_i}/(\mu_l)^3$, since $f(K_i, L) = K_i/L$ and $E(K_i/L) = E(f(K_i, L))$. Also, $E(k_i - \mu_{k_i}) = E(l - \mu_l) = 0$; $Var(K_i) = E(k_i - \mu_{k_i})^2$, and $Cov(K_i, L) = E[(k_i - \mu_{k_i})(l - \mu_l)]$. For variance of k_i, denoted by $Var(k_i)$, it follows from equation (11) of the mean and from first order Taylor series expansion of $f(K_i, L)$ around $\mu = (\mu_{K_i}, \mu_l)$ that:

$$\begin{aligned} Var(k_i) &= Var(K_i/L) \\ &\approx f'^2_{k_i}(\mu)Var(K_i) + 2f'_{k_i}(\mu)f'_l(\mu)Cov(K_i,L) + f'^2_l(\mu)Var(L) \\ &= \frac{1}{\mu_l^2}Var(K_i) - 2\frac{\mu_{k_i}}{\mu_l^3}Cov(K_i,L) + \frac{\mu_{k_i}^2}{\mu_l^4}Var(L). \end{aligned} \tag{12}$$

These properties and others such as estimates of third and fourth moments of the statistic $\mathbf{k}$ can help in characterising the distribution of $\mathbf{k}$, which can in turn provide a basis for formal tests about cluster outliers to model (1). Nonetheless, graphical methods also provide reliable alternative to formal tests of model residuals [32]. Hence, we engaged graphical assessments to analyse outlying clusters to model (1). In practice, relative comparisons of values of a group outlier statistic are enough to examine outlying groups to the mixed model [37].

3. Simulation study

An extensive simulation study was carried out to evaluate performance of the proposed outlier statistic (10). We generated survival times t_{ij} from $T \sim Exponential(1)$ using cumulative hazard inversion method [2] and the shared frailty model:

$$\lambda_{ij}(t|\beta, b_i) = \lambda_0(t)exp(\beta_1 X_{ij1} + \beta_2 X_{ij2} + b_i), \tag{13}$$

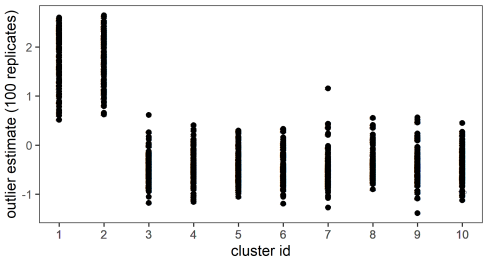
where $\lambda_0(t) = 0.1$; $X_{ij1} \sim Bernoulli(0.7)$ and $\beta_1 = 0.5$, while $X_{ij2} \sim N(0, 1)$ and $\beta_2 = 1$. The random effects b_i were generated from $N(0, 0.4^2)$. The inversion method derives t_{ij} of T from $\Lambda_{ij}^{-1}(-log(S(t_{ij})))$, where $S(t_{ij}) \sim Uniform(0, 1)$ is the survival function at time t_{ij} and hence, making $\Lambda_{ij}(t) = -log(Uniform(0, 1)) \sim Exponential(1)$. The censoring variable was generated from $Bernoulli(0.4)$. The `R package simsurv` [2, 19] was used to set up and draw the data. Samples of size 10, 20, and 50 clusters each were generated. The cluster sizes were 10, 80 and 500, which enabled us to assess effect of sample size on performance of the proposed statistic (10). We additionally simulated data using unbalanced cluster sizes from $Poisson(200)$. We perturbed model parameters in the first two clusters of each simulated data to make them outliers. For each perturbation set, the simulations were replicated 100 and 1000 times in order to observe the effect of simulation size on the performance of the statistic (10) [31].

The perturbations in the first two clusters were set at: $b_1, b_2 \sim \{N(10, 2.5^2), N(15, 5.5^2)\}$, $\beta_1 = \{1.8, 2.7\}$ and $\beta_2 = \{2.0, 2.5\}$. A cluster can

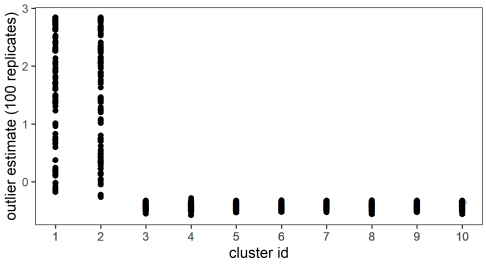
also have unusual observations due to an interplay between fixed and random effects in the model [36]. We thus performed joint perturbations as follows: $\beta_1 = \{1.8, 2.7\}$ jointly with $\beta_2 = \{2.0, 2.5\}$ and then jointly $\beta_1 = \{1.8, 2.7\}$, $\beta_2 = \{2.0, 2.5\}$ and $b_1, b_2 \sim \{N(10, 2.5^2), N(15, 5.5^2)\}$. The performance of the proposed outlier statistic was assessed using the percentage of the outlier clusters that were correctly identified as outliers. This is essentially the number of times for which $k_{1,2} > mean\,[maximum(k_i : i = 3, 4, ..., M)]$ or $k_{1,2} < mean\,[minimum(k_i : i = 3, 4, ..., M)]$ out of 100 or 1000 simulations [7, 31]. The model (13) was fitted to each of the samples, and the statistic (10) was computed.

3.1. Simulation results

The plots in Figure 1, for two cases of simulations that involved fixed effects and with 100 replications, indicate that the proposed statistic had detected clusters 1 and 2 as outliers to the fitted model. It is shown in both Figure 1(a) and 1(b) that the values of the proposed outlier statistic for clusters 1 and 2, where the data-generating model was perturbed, deviated markedly from those of clusters 3 to 10. The rest of the results on the success rates of the proposed statistic (10) are presented in Tables 1, 2 and 3.



(a) Plots of standardised outlier residual for a case of data with perturbed $\beta_1 = 2.7$ in first 2 of 10 clusters sample, each with 500 subjects.



(b) Plots of standardised outlier residual for a case of data with perturbed $\beta_2 = 2.5$ in first 2 of 10 clusters sample, with unbalanced cluster sizes.

Figure 1. Plots of the standardised values of the proposed outlier statistic for two simulation cases in which perturbed models were used to generate data in first two clusters. Source: Researcher.

Table 1 shows the success rates of the proposed statistic in detecting cluster 1 and 2 as outliers based on a cutoff presented before, for simulation cases that involved separate perturbations of model parameters and balanced cluster design. It is shown that the outlier statistic was effective, when the perturbations involved fixed and not random effects. In situations where β_1 or β_2 was perturbed in clusters 1 and 2, the residual generally identified the two clusters as having unusual survival outcomes compared to the rest clusters. But, the values of the proposed statistic were not different across clusters for cases that involved perturbed random effects in clusters 1 and 2. This reflects the fact that subjects from the same cluster share a random effect, which might contribute less to within-cluster variation of the deviance residual that is used in the proposed outlier method, unlike fixed covariates values that vary from subject to subject even within the same cluster.

Furthermore, the results show that performance of the statistic improved with increasing fixed-effect size as well as cluster sample size. In cases of large effect sizes, i.e., $\beta_1 = 2.7$ and $\beta_2 = 2.5$ or cluster sample of 500, the success rates of the statistic steadily converged to very high values of not less than 88.9% at both 100 and 1000 replications. But, the rates were generally low in scenarios of low effect sizes and small cluster sizes of 10, and with high inconsistency between 100 and 1000 simulations. Finally, the outlier statistic performed equally across different number of clusters per data set, holding constant the cluster sample size and fixed effect size.

Table 1. The percentage of the clusters 1 and 2 that were correctly identified as outliers, for different number of clusters and cluster sizes; a case of separate perturbations [1] to β_1, β_2 or $b_1 \sim b_2$

					100 replicates		1000 replicates	
M	n_i	β_1	β_2	$b_1 \sim b_2$	Cluster1	Cluster2	Cluster1	Cluster2
10	10	0.5	1	$N(10, 2.5^2)$	0	0	0	0
	10	0.5	1	$N(15, 5.5^2)$	0	0	0	0
10	80	0.5	1	$N(10, 2.5^2)$	0	0	0	0
	80	0.5	1	$N(15, 5.5^2)$	0	0	0	0
10	500	0.5	1	$N(10, 2.5^2)$	0	0	0	0
	500	0.5	1	$N(15, 5.5^2)$	0	0	0	0
20	10	0.5	1	$N(10, 2.5^2)$	0	0	0	0
	10	0.5	1	$N(15, 5.5^2)$	0	0	0	0
20	80	0.5	1	$N(10, 2.5^2)$	0	0	0	0
	80	0.5	1	$N(15, 5.5^2)$	0	0	0	0
20	500	0.5	1	$N(10, 2.5^2)$	0	0	0	0
	500	0.5	1	$N(15, 5.5^2)$	0	0	0	0
50	10	0.5	1	$N(10, 2.5^2)$	0	0	0	0
	10	0.5	1	$N(15, 5.5^2)$	0	0	0	0
50	80	0.5	1	$N(10, 2.5^2)$	0	0	0	0
	80	0.5	1	$N(15, 5.5^2)$	0	0	0	0
50	500	0.5	1	$N(10, 2.5^2)$	0	0	0	0
	500	0.5	1	$N(15, 5.5^2)$	0	0	0	0
10	10	1.8	1	$N(0, 0.4^2)$	0	0	0	0
	10	2.7	1	$N(0, 0.4^2)$	0	0	0	0
10	80	1.8	1	$N(0, 0.4^2)$	0	0	0	0
	80	2.7	1	$N(0, 0.4^2)$	20	17	22	22
10	500	1.8	1	$N(0, 0.4^2)$	50	50	25.8	24.9
	500	2.7	1	$N(0, 0.4^2)$	100	100	88.9	89.6
20	10	1.8	1	$N(0, 0.4^2)$	0	0	0	0
	10	2.7	1	$N(0, 0.4^2)$	0	0	0	0
20	80	1.8	1	$N(0, 0.4^2)$	3	3	0.6	0.4
	80	2.7	1	$N(0, 0.4^2)$	17	7	2.2	1.7
20	500	1.8	1	$N(0, 0.4^2)$	84	83	17.4	18.9
	500	2.7	1	$N(0, 0.4^2)$	96	100	95.2	95.5
50	10	1.8	1	$N(0, 0.4^2)$	0	0	0	0
	10	2.7	1	$N(0, 0.4^2)$	1	3	0	0
50	80	1.8	1	$N(0, 0.4^2)$	11	15	8.0	8.2
	80	2.7	1	$N(0, 0.4^2)$	6	2	0.7	0.7
50	500	1.8	1	$N(0, 0.4^2)$	57	59	31.7	34.6
	500	2.7	1	$N(0, 0.4^2)$	100	98	98.5	97.5
10	10	0.5	2.0	$N(0, 0.4^2)$	6	6	0	0
	10	0.5	2.5	$N(0, 0.4^2)$	23	25	6.7	6.1
10	80	0.5	2.0	$N(0, 0.4^2)$	95	92	57	59.6
	80	0.5	2.5	$N(0, 0.4^2)$	100	100	92.7	93.8
10	500	0.5	2.0	$N(0, 0.4^2)$	100	100	99.9	99.6
	500	0.5	2.5	$N(0, 0.4^2)$	100	100	100	100
20	10	0.5	2.0	$N(0, 0.4^2)$	29	21	12.8	10.4
	10	0.5	2.5	$N(0, 0.4^2)$	52	44	29.5	32.7
20	80	0.5	2.0	$N(0, 0.4^2)$	82	83	72.8	74.7
	80	0.5	2.5	$N(0, 0.4^2)$	99	98	92.5	93
20	500	0.5	2.0	$N(0, 0.4^2)$	100	100	99.7	100
	500	0.5	2.5	$N(0, 0.4^2)$	100	100	99.7	99.7
50	10	0.5	2.0	$N(0, 0.4^2)$	43	25	23.9	24.1
	10	0.5	2.5	$N(0, 0.4^2)$	53	43	42.4	41.3
50	80	0.5	2.0	$N(0, 0.4^2)$	93	89	79.6	80.4
	80	0.5	2.5	$N(0, 0.4^2)$	98	97	87.2	87.8
50	500	0.5	2.0	$N(0, 0.4^2)$	100	100	99.1	99.1
	500	0.5	2.5	$N(0, 0.4^2)$	100	100	99.9	99.8

[1] No perturbations were done to data in other clusters except clusters 1 and 2. The remaining clusters had $\beta_1 = 0.5$, $\beta_2 = 1$, and $b_i \sim N(0, 0.4^2)$.

The results in Table 2 show the success rates of the proposed statistic from simulation cases with jointly perturbed parameters. It is shown that the statistic correctly detected clusters 1 and 2 as outliers in comparison with the other clusters. As was the case with separate perturbations, the random effects had no contribution in making clusters 1 and 2 outliers, the success rates of the statistic from joint perturbations of β_1 and β_2 were generally not different from those of jointly perturbed β_1, β_2 and b_i.

Similar to separate perturbations, the results in Table 2 show that the success rates of the outlier statistic improved with sample size and fixed effect size. In addition, the joint effect sizes for β_1 and β_2 combined with large sample size of 500 yielded even higher success rates of not less than 95.2%. Thus, proper mix of the fixed effect covariates in the fitted model is another important factor to consider in the application of the proposed statistic. Again, the statistic performed equally between different number of clusters per dataset, controlling for cluster sample size and fixed effect size.

Table 2. The percentage of the clusters 1 and 2 that were correctly identified as outliers, for different number of clusters and cluster sizes; a case of joint perturbations [1] among β_1, β_2, and $b_1 \sim b_2$

					100 replicates		1000 replicates	
M	n_i	β_1	β_2	$b_1 \sim b_2$	Cluster1	Cluster2	Cluster1	Cluster2
10	10	1.8	2.0	$N(0, 0.4^2)$	12	10	0	0
	10	2.7	2.5	$N(0, 0.4^2)$	22	14	0	0
10	80	1.8	2.0	$N(0, 0.4^2)$	55	68	39.1	41.8
	80	2.7	2.5	$N(0, 0.4^2)$	90	89	74.6	72.7
10	500	1.8	2.0	$N(0, 0.4^2)$	100	100	99.4	99.1
	500	2.7	2.5	$N(0, 0.4^2)$	100	100	100	100
20	10	1.8	2.0	$N(0, 0.4^2)$	5	10	7	6
	10	2.7	2.5	$N(0, 0.4^2)$	18	21	15.3	12.3
20	80	1.8	2.0	$N(0, 0.4^2)$	58	58	43.4	44.8
	80	2.7	2.5	$N(0, 0.4^2)$	86	88	70.2	69.6
20	500	1.8	2.0	$N(0, 0.4^2)$	99	99	98.9	98.6
	500	2.7	2.5	$N(0, 0.4^2)$	100	100	100	100
50	10	1.8	2.0	$N(0, 0.4^2)$	8	15	7.1	7.9
	10	2.7	2.5	$N(0, 0.4^2)$	32	26	21.4	20.4
50	80	1.8	2.0	$N(0, 0.4^2)$	59	54	41.2	39.7
	80	2.7	2.5	$N(0, 0.4^2)$	86	87	69	66.8
50	500	1.8	2.0	$N(0, 0.4^2)$	100	100	99	98.5
	500	2.7	2.5	$N(0, 0.4^2)$	100	100	97.7	98
10	10	1.8	2.0	$N(10, 2.5^2)$	6	7	0	0
	10	2.7	2.5	$N(15, 5.5^2)$	9	12	0	0
10	80	1.8	2.0	$N(10, 2.5^2)$	75	74	38.7	40.9
	80	2.7	2.5	$N(15, 5.5^2)$	85	86	74.6	72.6
10	500	1.8	2.0	$N(10, 2.5^2)$	100	100	99.3	99.2
	500	2.7	2.5	$N(15, 5.5^2)$	100	100	100	100
20	10	1.8	2.0	$N(10, 2.5^2)$	9	17	4.1	4.1
	10	2.7	2.5	$N(15, 5.5^2)$	19	25	15.5	15.9
20	80	1.8	2.0	$N(10, 2.5^2)$	48	54	38.3	35.4
	80	2.7	2.5	$N(15, 5.5^2)$	82	76	73.4	73.3
20	500	1.8	2.0	$N(1, 2.5^2)$	100	100	97.9	98.4
	500	2.7	2.5	$N(15, 5.5^2)$	100	100	99.5	99.3
50	10	1.8	2.0	$N(10, 2.5^2)$	15	15	8.2	9.8
	10	2.7	2.5	$N(15, 5.5^2)$	30	25	19.2	19.8
50	80	1.8	2.0	$N(10, 2.5^2)$	65	51	41.6	40.8
	80	2.7	2.5	$N(15, 5.5^2)$	79	80	68.9	71.8
50	500	1.8	2.0	$N(10, 2.5^2)$	100	99	96.1	95.2
	500	2.7	2.5	$N(15, 5.5^2)$	100	100	98.7	99.4

[1] No perturbations were done to data in other clusters except clusters 1 and 2. The remaining clusters had $\beta_1 = 0.5$, $\beta_2 = 1$, and $b_i \sim N(0, 0.4^2)$.

Table 3 presents results of performance of the statistic from unbalanced cluster designs of the simulations, that involved both separate and joint perturbations of fixed effects. It is shown that the proposed statistic was equally effective in the unbalanced cluster design, as in the balanced cluster sizes. Like in balanced cluster sizes, the performance of the statistic was higher in cases where the effect size was high. Again, the proposed statistic performed equally across different number of clusters per sample, controlling for effect size.

Table 3. The percentage of the clusters 1 and 2 that were correctly identified as outliers, for different number of clusters and cluster sizes; a case of unbalanced cluster sample sizes with average size of 200

				100 replicates		1000 replicates	
M	β_1	β_2	b_i	Cluster1	Cluster2	Cluster1	Cluster2
10	1.8	1	$N(0, 0.4^2)$	22	19	3.9	3.9
	2.7	1	$N(0, 0.4^2)$	73	68	33.6	32.5
20	1.8	1	$N(0, 0.4^2)$	15	24	5.7	5.4
	2.7	1	$N(0, 0.4^2)$	80	78	46.1	47.8
50	1.8	1	$N(0, 0.4^2)$	17	21	6.5	6.2
	2.7	1	$N(0, 0.4^2)$	63	66	28.3	27.2
10	0.5	2.0	$N(0, 0.4^2)$	97	95	90.5	89.7
	0.5	2.5	$N(0, 0.4^2)$	100	100	99.3	99.0
20	0.5	2.0	$N(0, 0.4^2)$	99	100	95.2	94.5
	0.5	2.5	$N(0, 0.4^2)$	99	100	99.3	99.0
50	0.5	2.0	$N(0, 0.4^2)$	96	97	95.8	96.8
	0.5	2.5	$N(0, 0.4^2)$	100	100	98.8	99.3
10	1.8	2.0	$N(0, 0.4^2)$	96	98	83.1	82.7
	2.7	2.5	$N(0, 0.4^2)$	100	100	98.1	97.9
20	1.8	2.0	$N(0, 0.4^2)$	95	97	75.6	79.5
	2.7	2.5	$N(0, 0.4^2)$	97	99	94.8	95.8
50	1.8	2.0	$N(0, 0.4^2)$	85	84	76.1	77.1
	2.7	2.5	$N(0, 0.4^2)$	96	97	91.3	93.9

No perturbations were done to data in other clusters except clusters 1 and 2. The remaining clusters had $\beta_1 = 0.5$, $\beta_2 = 1$.

4. Application to child clustered survival data from Malawi

4.1. Data description

We applied the proposed outlier statistic along with the standard method of visual inspection of standardised residuals [10] on child survival data from the 2015-16 MDHS. The MDHS was conducted between 19 October 2015 and 18 February 2016, and it collected child survival data from women respondents aged between 15 and 49 years, who provided birth histories. The survey used two-stage stratified sampling, with emuneration areas and households as primary and secondary sampling units, respectively [13]. We studied a total of 17,286 children who were born during the 5 years preceding implementation of the survey. These children were clustered at the 28 districts in Malawi, and each district was further split into rural or urban areas. Thus, for this study, we used the 56 subdistricts to illustrate the usefulness of our proposed multivariate survival data outlier detection method. The Demographic and Health Surveys (DHS) program provides free and publicly available datasets that can be accessed upon submitting a request at `https://dhsprogram.com/data/new-user-registration.cfm`.

The main outcome was the survival time in months of a child, from birth to either death (event) or not yet dead (censored) before his or her fifth birthday. For the predictor variables, we pre-selected some of them based on the existing literature on the determinants of child mortality in the region. For our study, we included birth order, sex of child, weight of the child at birth, mother's age at childbirth, mother's education, place of residence, place of delivery and household wealth index [8, 15]. In addition to controlling for the fixed effects variables, a sub-district was considered as a random effect in the shared frailty Cox regression model for the analysis of child

survival in Malawi. Due to missing data in some variables, our analysis was based on complete case of 14,645 children, of whom 539 (3.68%) had died before age 60 months.

Table 4 shows the distribution of deaths and characteristics of the 14,645 children that were analysed. About 49.5% of the children were females. Nearly, 82.6% of the children were residing in rural areas. The majority (97.5%) of the babies were delivered at a health facility. The mean birth weight (SD) was 3,241.63 (777.12); the smallest and largest birth weights were 500 and 6000 grams, respectively. About 15.0% of the babies weighed <2500g. The log-rank test showed that there was a significant difference in survival of the children across levels of most variables except mother's education level (p-value = 0.3), place of delivery (p-value = 0.8), and household wealth index (p-value = 0.2).

Table 4. Distribution of deaths by socio-demographic characteristics of children, 2015-16 MDHS (n = 14,645).

Child's characteristics	Children, n (%)	Deaths, n (%)	Log-rank Test (p-value)
Overall sample	14,645 (100)	539 (3.68)	
Sex of child			
Male	7,393 (50.5)	298 (4.03)	5.6 (0.02)
Female	7,252 (49.5)	241 (3.32)	
Birth weight			
$\geq$ 2500gms	12,465 (85.1)	430 (3.45)	12.9 (< 0.001)
< 2500gms	2,180 (14.9)	109 (5.00)	
Mean (SD)	3,241.63 (777.12)		
Mother's age at child birth			
12−19 yrs	2,121 (14.5)	109 (5.14)	18.7 (< 0.0001)
20−34 yrs	10,531 (71.9)	346 (3.29)	
35−49 yrs	1,993 (13.6)	84 (4.21)	
Mother's education level			
No education	1,588 (10.8)	55 (3.46)	2.2 (0.3)
Primary	9,611 (65.6)	370 (3.85)	
Secondary or above	3,446 (23.5)	114 (3.31)	
Place of delivery			
Home and other	369 (2.50)	15 (4.07)	0.1 (0.8)
Health facility	14,276 (97.5)	524 (3.67)	
Household wealth index			
Poor	6,166 (42.1)	245 (3.97)	2.8 (0.2)
Middle	2,823 (19.3)	100 (3.54)	
Rich	5,656 (38.6)	194 (3.43)	
Place of residence			
Urban	2,544 (17.4)	77 (3.03)	3.5 (0.06)
Rural	12,101 (82.6)	462 (3.82)	
Birth order			
1	3,769 (25.7)	169 (4.48)	11.7 (0.003)
2-3	5,480 (37.2)	175 (3.19)	
$\geq$ 4	5,396 (36.9)	195 (3.61)	
Mean (SD)	3.16 (2.04)		

SD=standard deviation; n = sample size.

4.2. Model estimates results

We considered four models with Model 0 being a standard Cox model without accounting for clustering of survival times, but included all the covariates from Table 4, namely: sex of child, birth weight, mother's age at child birth, mother's education, household wealth index, place of residence, place of delivery and birth order. The other three were shared frailty models, where Model 1 included all covariates as in Model 0, Model 2 included all covariates in Model 1 except mother education, household wealth index and place of delivery. Model 3 included all covariates in Model 2 except place of residence. Model comparison was done through the Akaike information criterion (AIC) [21].

The results in Table 5 show that Models 2 and 3 had lowest AIC. These two models were subsequently used in the detection of U5M outlying subdistricts. Female gender, mother's age-at-child birth of 20 years and above, and higher birth order were associated with with a reduced risk of U5M. Birth weights of less than 2500 grams and residing in rural areas were associated with increased risk of U5M. The coefficients of birth order and squared birth order imply a U-shaped relationship between birth order and U5M. The results are consistent with previous findings [8, 14].

Table 5. Regression parameter estimates from fitting Cox PH frailty model on Malawi child survival data, 2015-16 MDHS.

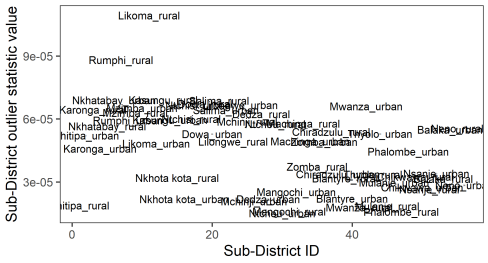
Fixed effects predictors	Estimate (SD)	Estimate (SD)	Estimate (SD)	Estimate (SD)
	[1] Model 0	Model 1	Model 2	Model 3
Sex of child				
Male	0	0	0	0
Female	-0.231 (0.087)	-0.232 (0.087)	-0.233 (0.087)	-0.233 (0.087)
Birth weight				
≥2500gms	0	0	0	0
<2500gms	0.603 (0.108)	0.599 (0.108)	0.370 (0.108)	0.607 (0.108)
Mother's age at child birth				
12−19 yrs	0	0	0	0
20−34 yrs	-0.307 (0.134)	-0.308 (0.134)	-0.318 (0.132)	-0.334 (0.131)
35−49 yrs	-0.056 (0.216)	-0.061 (0.216)	-0.089 (0.213)	-0.116 (0.212)
Mother's education				
No education	0	0		
Primary	0.173 (0.150)	0.158 (0.151)		
Secondary or above	0.128 (0.184)	0.114 (0.185)		
HH wealth index				
Poor	0	0		
Middle	-0.104 (0.119)	-0.100 (0.119)		
Rich	-0.053 (0.110)	-0.040 (0.110)		
Place of residence				
Urban	0	0	0	
Rural	0.191 (0.137)	0.197 (0.138)	0.208 (0.125)	
Place of delivery				
Home and other	0	0		
Health facility	-0.014 (0.263)	-0.015 (0.263)		
Birth order of child	-0.115 (0.075)	-0.114 (0.075)	-0.126 (0.062)	-0.103 (0.074)
Birth order squared	0.012 (0.008)	0.012 (0.008)	0.013 (0.006)	0.011 (0.008)
Sub-district random effect SD		0.149	0.155	0.158
AIC	10,132.79	10,129.28	10,121.01	10,121.89

[1] Model 0 is the standard univariate Cox PH model without cluster-specific random effects term.

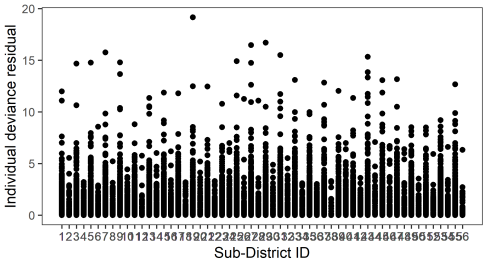
4.3. Outlying Sub-Districts to child mortality

Detection of U5M outlier sub-districts was based on Models 2 and 3 as they had smallest AIC values. We used a national U5M rate of 63 deaths per 1000 live births as the baseline hazard [13]. We also used the visual inspection method for standardised residuals in the detection of outlier sub-districts [10].

For Model 2, the estimates of the proposed statistic in Figure 2(a) show that most subdistricts were well-fitted by the model, as the residual values were close to zero. The outlier statistic detected two sub-districts, namely: *Likoma* rural and *Rumphi* rural as U5M outlier sub-districts. This means that the U5M rates in the two sub-districts were uniquely higher or lower compared to the rest sub-districts. Considering the individual observation outliers in Figure 2(b), we see that the deviance residuals were concentrated towards zero, with few children having larger values. Some of the sub-districts had several single observation outliers. However, the visual results in Figure 2(b) could not help in determining whether any of the affected sub-districts is an outlier, since the ranges of the plots in different sub-districts are overlapping.



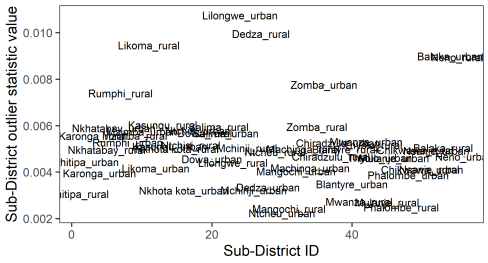
(a) Estimates of proposed outlier statistic per sub-district based on Model 2.



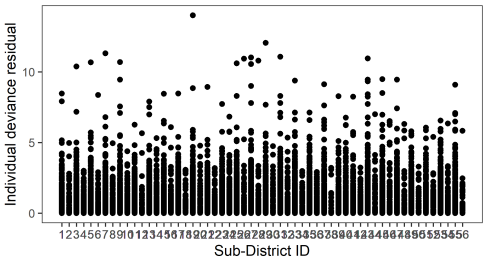
(b) Plots of deviance residuals for children in each subdistrict based on Model 2.

Figure 2. Outlier results for the proposed statistic compared with visual inspection of residuals using Model 2

For Model 3, the outlier results in Figure 3(a) show that the majority of the sub-districts were also well-fitted by the model. Fitting Model 3 resulted in more sub-district outliers than Model 2. Now, *Lilongwe* urban, *Dedza* rural, *Likoma* rural, *Balaka* urban, *Neno* rural, *Zomba* urban and *Rumphi* rural were detected as U5M outlier sub-districts, two of which were also outliers under Model 2. Using single observation outliers in Figure 3(b), one can not conclusively identify outlier sub-districts, because of the overlaps in the range of the plots across the sub-districts.



(a) Estimates of proposed outlier statistic per sub-district based on Model 3.



(b) Plots of deviance residuals for children in each sub-district based on Model 3.

Figure 3. Outlier results for the proposed statistic compared with visual inspection of residuals using Model 3

5. Conclusion

In statistical modelling, the presence of extreme data points that deviate from the other observations in the dataset could bias and influence estimates. The basic assumptions of regression as well as other statistical models could also be impacted by the presence of outlier observations. Much of the work has been devoted to the detection of univariate outliers in linear models. In linear mixed models, limited work has been done on multivariate outliers. Detection of multivariate outliers for multivariate survival models, that are now commonly used to model correlated survival data, has received relatively scant research attention. This study was set out to derive a novel group outlier detection statistic for multivariate survival data. Using a broad simulation study, the proposed outlier statistic performed exceedingly well in detecting groups with extremely low or high survival times at a moderate number of clusters and cluster size as well as with appropriate combination of fixed covariates.

For a well-fitted multivariate survival model, the outlying groups may provide information about unique patterns of survival outcomes that are prevalent in the outlier groups. The unusual values in a data set could be due to typos or unique measurements in one of the covariates or the response variable worth investigating [36]. In the application of child survival with 56 sub-districts, our outlier group detection tool identified between 2 and 7 under-five mortality outlier sub-districts, depending on the fit of the multivariate survival model. The detected outlying sub-districts could be further investigated by child health policy makers to understand the risk factors that are adversely impacting child health. In this way, appropriate child health impact interventions could be implemented or enhanced in the sub-districts. Or, it could be that the unusual child survival data observed in the outlying sub-districts relate to errors in data management, which could also be investigated. Our proposed group outlier detection method was more accurate than the usual visual inspection of univariate outliers [10] in detecting the sub-districts with unusually high or low under-five mortality rates in Malawi.

We believe that our proposed group outlier detection statistic could be of use in the context of finding clusters (e.g. subnational areas and hospitals) with much lower than normal survival rates under a multivariate survival model. We relied on time-constant covariates and the Cox proportional hazards frailty regression model in this study. Future studies may develop outlier statistics for various formulations of the multivariate survival model, such as stratified and time-dependent survival models.

Disclosure statement

The authors have declared no conflict of interest.

Funding

Tsirizani Kaombe was partly funded by the World Bank's Malawi Skills Development Project (SDP), grant number P131660 that was implemented at Chancellor College, University of Malawi and partly by SSACAB (Grant:107754/Z/15/Z) through the DELTAS Africa Initiative of the Welcome Trust. Samuel Manda was supported by the

South African Medical Research Council. The views expressed in this publication are those of the authors and not necessarily the funders.

References

[1] Abrahantes, J. and Burzykowski, T. (2005). A version of the EM algorithm for proportional hazard model with random effects. *Biometrical Journal*, 47(6):847–862.

[2] Brilleman, S., Rory, W., Moreno-Betancur, M., and Crowther, M. (2018). simsurv: A Package for simulating simple or complex survival data. In *UseR! Conference 2018, Brisbane, Australia*. Monash University.

[3] Burke, M., Heft-Neal, S., and Bendavid, E. (2016). Sources of variation in under-5 mortality across Sub-Saharan Africa: A spatial analysis. *The Lancet Global Health*, 4(12):e936–e945.

[4] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.

[5] Gayawan, E., Adarabioyo, M. I., Okewole, D. M., Fashoto, S. G., and Ukaegbu, J. C. (2016). Geographical variations in infant and child mortality in West Africa: A geo-additive discrete-time survival modelling. *Genus*, 72(1):1–20.

[6] Goeman, J. (2010). L1 penalized estimation in the Cox proportional hazards model. *Biometrical Journal*, 52(1):70–84.

[7] Kaombe, T. M. and Manda, S. O. (2021). Detecting influential data in multivariate survival models. *Communications in Statistics-Theory and Methods*, 0(0):1–17.

[8] Kazembe, L. N. and Mpeketula, P. M. (2010). Quantifying spatial disparities in neonatal mortality using a structured additive regression model. *PloS One*, 5(6):e11180.

[9] Khodaee, G. H., Khademi, G., and Saeidi, M. (2015). Under-five Mortality in the World (1900-2015). *International Journal of Pediatrics*, 3(6.1):1093–1095.

[10] Langford, I. and Lewis, T. (1998). Outliers in multilevel data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 161(2):121–160.

[11] Li, Z., Karlsson, O., Kim, R., and Subramanian, S. (2021). Distribution of under-5 deaths in the neonatal, postneonatal, and childhood periods: A multicountry analysis in 64 low-and middle-income countries. *International journal for equity in health*, 20(1):1–11.

[12] Liang, K. and Zeger, S. L. (1993). Regression analysis for correlated data. *Annual Review of Public Health*, 14(1):43–68.

[13] Malawi National Statistical Office (NSO) and ICF (2017). 2015-16 Malawi Demographic and Health Survey: Key Findings. *Zomba, Malawi, and Rockville, Maryland, USA: Author*.

[14] Manda, S. (1999). Birth intervals, breastfeeding and determinants of childhood mortality in Malawi. *Social Science & Medicine*, 48(3):301–312.

[15] Manda, S. (2001). A comparison of methods for analysing a nested frailty model to child survival in Malawi. *Australian & New Zealand Journal of Statistics*, 43(1):7–16.

[16] Manda, S. (2011). A nonparametric frailty model for clustered survival data. *Communications in Statistics-Theory and Methods*, 40(5):863–875.

[17] Manda, S. and Meyer, R. (2005). Age at first marriage in Malawi: a Bayesian multilevel analysis using a discrete time-to-event model. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(2):439–455.

[18] Masuy-Stroobant, G. and Gourbin, C. (1995). Infant health and mortality indicators. *European Journal of Population/Revue européenne de démographie*, 11(1):63–84.

[19] Moriña, D. and Navarro, A. (2014). The R package survsim for the simulation of simple and complex survival data. *Journal of Statistical Software*, 59(2):1–20.

[20] Morton, S., Pencheon, D., and Squires, N. (2017). Sustainable Development Goals (SDGs), and their implementation: A national global framework for health, development and equity needs a systems approach at every level. *British medical bulletin*, pages 1–10.

[21] Portet, S. (2020). A primer on model selection using the Akaike Information Criterion. *Infectious Disease Modelling*, 5:111–128.

[22] Ripatti, S., Larsen, K., and Palmgren, J. (2002). Maximum likelihood inference for mul-

tivariate frailty models using an automated Monte Carlo EM algorithm. *Lifetime Data Analysis*, 8(4):349–360.

[23] Ripatti, S. and Palmgren, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics*, 56(4):1016–1022.

[24] Rousseeuw, P. J. and Hubert, M. (2011). Robust statistics for outlier detection. *Wiley interdisciplinary reviews: Data mining and knowledge discovery*, 1(1):73–79.

[25] Sarkar, S. K., Midi, H., and Rana, S. (2011). Detection of outliers and influential observations in binary logistic regression: An empirical study. *Journal of Applied Sciences*, 11(1):26–35.

[26] Storeygard, A., Balk, D., Levy, M., and Deane, G. (2008). The global distribution of infant mortality: A subnational spatial view. *Population, space and place*, 14(3):209–229.

[27] Therneau, T., Grambsch, P., and Fleming, T. (1990). Martingale-based residuals for survival models. *Biometrika*, 77(1):147–160.

[28] Van Kempen, G. and Van Vliet, L. (2000). Mean and variance of ratio estimators used in fluorescence ratio imaging. *Cytometry: The Journal of the International Society for Analytical Cytology*, 39(4):300–305.

[29] Van Malderen, C., Amouzou, A., Barros, A. J., Masquelier, B., Van Oyen, H., and Speybroeck, N. (2019). Socioeconomic factors contributing to under-five mortality in Sub-Saharan Africa: A decomposition analysis. *BMC Public Health*, 19(1):1–19.

[30] Wakefield, J., Fuglstad, G.-A., Riebler, A., Godwin, J., Wilson, K., and Clark, S. J. (2019). Estimating under-five mortality in space and time in a developing world context. *Statistical methods in medical research*, 28(9):2614–2634.

[31] Xiang, L., Tse, S.-K., and Lee, A. H. (2002). Influence diagnostics for generalized linear mixed models: Applications to clustered data. *Computational Statistics & Data Analysis*, 40(4):759–774.

[32] Yang, H. (2012). Visual assessment of residual plots in multiple linear regression: A model-based simulation perspective. *Multiple Linear Regression Viewpoints*, 38(2):24–37.

[33] Yaya, S., Uthman, O. A., Okonofua, F., and Bishwajit, G. (2019). Decomposing the rural-urban gap in the factors of under-five mortality in sub-Saharan Africa? Evidence from 35 countries. *BMC public health*, 19(1):1–10.

[34] Yourkavitch, J., Burgert-Brucker, C., Assaf, S., and Delgado, S. (2018). Using geographical analysis to identify child health inequality in sub-Saharan Africa. *PLoS One*, 13(8):e0201870.

[35] Zewotir, T. and Galpin, J. (2005). Influence diagnostics for linear mixed models. *Journal of data science*, 3(2):153–177.

[36] Zewotir, T. and Galpin, J. S. (2006). Evaluation of linear mixed model case deletion diagnostic tools by Monte Carlo simulation. *Communications in Statistics-Simulation and Computation*, 35(3):645–682.

[37] Zewotir, T. and Galpin, J. S. (2007). A unified approach on residuals, leverages and outliers in the linear mixed model. *Test*, 16(1):58–75.

[38] Zhang, Z. (2016). Residuals and regression diagnostics: focusing on logistic regression. *Annals of Translational Medicine*, 4(10):1–8.