Developing a goodness of fit test for a joint model of clustered survival and count data

A. P. N. G. Adhikari & M. R. Sooriyarachchi

Taylor & Francis
Taylor & Francis Group

Check for updates

Developing a goodness of fit test for a joint model of clustered survival and count data

A. P. N. G. Adhikari and M. R. Sooriyarachchi

Department of Statistics, University of Colombo, Colombo 3, Sri Lanka

ABSTRACT
Mixed responses within clusters are becoming an increasingly popular phenomenon. However, this scenario is a complex one. This research is based on survival and count response data which occur frequently in medical studies. In the literature, there is no satisfactory goodness of fit test for a joint model of these two responses within a cluster setting. Therefore, the main objective of this study is to develop a GOF test for this scenario. Simulation studies were used to examine the type I error rate and the power of the test. An example was used to display the practicality of the test.

1. Introduction

1.1. Background

Clustered data arises when the data from the whole study consists of a number of different groups. The key feature of clustered data is that, observations within a cluster are "more alike" than observations from different clusters (Kreft and de Leeuw 1998). Clustered observations tend to be correlated at each level, because lower-level observations share common groupings at higher levels. Failure to take these correlations into account may lead to biased estimates. Thus, it is essential to adjust not only for the single level variability but also for the variability among clusters too. Multilevel modeling takes in to account the hierarchical structure and can be used as a remedy in order to overcome the above mentioned issues.

Different types of response variables at a particular or different levels are called "mixed" variables. These mixed response variables, in a hierarchical structure require "random effects" modeling. In the case of multivariate responses within an individual these individually identified patients are small clusters which in turn can lie within bigger clusters and then the small and bigger clusters are represented by random effects. In this research we consider the scenario where there are two responses within a patient and these patients are clustered within hierarchies.

Among many approaches available for multilevel models the random intercept model is the simplest model that assumes the lines fitted for different groups have the same slope but different intercepts. In this research random intercept models are considered.

1.1.1. Survival and count responses
Mixed responses occur in many fields of study such as in Medicine, Biology, Social Science and Education etc. Multilevel mixed responses arise when there is more than one response and these

---

are measured within units which are in turn within hierarchical structures. Performing univariate analysis for each response may not be appropriate due to the correlation among responses. Hence it is necessary to fit a joint model which incorporates the hierarchical nature.

This research focuses on a two response scenario where one response is a count while the other is a survival response. In medical studies, survival time of a patient and the count of incidence of the disease are an often encountered phenomenon. These survival and count responses are usually correlated. An example of such a study is in infectious disease epidemiology of a life threatening disease where there is a geographical variation of the intensity of the disease. In such a scenario, it is better to model the survival response and count responses together, within a random effects model framework to take account of the clustering.

1.1.2. Goodness of fit tests

The actual goal of fitting a model is to make inferences on the interested population by predicting the values as close as possible. Therefore, it is crucial to test the goodness of fit of a model before it is used to make statistical inferences. The performance of a model can be simply taken from residual analysis which gives a more immediate and clear illustration of the relationship between fitted values and actual data. However, examining residuals for each and every individual is time consuming. Instead, if it can be done by using a single summary measure it would be easier. That is where the role of a goodness of fit test comes in to play.

Up to now, several GOF tests have been constructed to handle some particular scenarios. It is essential to mention that available GOF tests for models with continuous responses are not suitable for models with categorical responses. Apart from that initial GOF tests for single level data are not appropriate for multilevel data. When it comes to multivariate multilevel data, the GOF test is further complicated.

Multivariate multilevel models for survival response and count response have no suitable GOF test developed as yet though the survival time of a patient and the count of incidence of the disease are an often encountered phenomenon in medical studies. The need for such a test was emphasized by Hapugoda and Sooriyarachchi (2018) who fitted a Joint Model of Discrete Time Hazard with Poisson Regression to an example data set.

1.2. Literature review

Recently, Ranathunga and Sooriyarachchi (2017) discussed about the necessity of a GOF test for multivariate multilevel models. It can be considered as the major encouragement in order to develop a GOF test which can be used to assess multivariate multilevel models. The GOF test developed for univariate binary multilevel models by Perera, Sooriyarachchi, and Wickramasuriya (2016) is based on the Hosmer and Lemesbow (1980) test and the extension of it by Lipsitz, Fitzmaurice, and Molenbergh (1996). This was further extended by adding another binary response to the model by Wickramarachchi & Sooriyarachchi (unpublished B.Sc. thesis, 2017). Further, the scope of this study has been narrowed down to consider the scenario of only two responses when one response is survival while the other is count. Sunethra and Sooriyarachchi (2020) have developed a joint model for a survival response and a count response by considering the hierarchical data structure using a joint model with two random effects, namely within patient and within cluster. Further, Hapugoda and Sooriyarachchi (2018) have simplified this joint model for the case of a discrete survival response using only one random effect, namely within the bigger cluster. The construction of this discrete survival response from continuous survival data is well explained in Hapugoda and Sooriyarachchi (2018) and they show that the discrete survival response is of binary form. The joint model developed by Sunethra and Sooriyarachchi (2020) is complicated. As a result of that, this model with the H-L test cannot be

fitted due to non-convergence. However, Hapugoda and Sooriyarachchi (2018) looked at a simpler model of Sunethra and Sooriyarachchi scenario where they fit only one random effect. This model has no convergence problems when incorporating the H-L test. Thus, the method of Hapugoda and Sooriyarachchi (2018) is used for this research. The lack of a GOF test was emphasized by these studies in the literature. The novel GOF test can be considered as an extension of the above mentioned developed GOF tests and the joint modeling approach of Hapugoda and Sooriyarachchi (2018).

1.3. Objectives of the study

The main objective of this study is to develop a novel GOF test for a joint model of survival response and a count response in the presence of clusters. Apart from that, the secondary objectives are to determine the properties of the developed novel GOF test through observing the type I error rate and power for various combinations of sample sizes, cluster sizes, number of clusters, variances of cluster effects and correlations between responses. This is done by using simulation studies. In addition, another secondary objective was to evaluate the applicability of the newly developed GOF test on a real-life dataset with discretized survival and count responses.

1.4. Outline of the paper

This section presents the problem of developing a GOF test for multivariate multilevel survival data and count data by giving an introduction and some literature related to the research problem.

Sec. 2: The methodologies and theories behind the development of the novel GOF test have been discussed here.

Sec. 3: Results of the simulation studies which were carried out to observe the properties of the developed test are mentioned here.

Sec. 4: Presents the application of the novel GOF test to a real-world dataset.

Sec. 5: Concludes the paper by summarizing the overall content in the preceding chapters while discussing the limitations of the study.

2. Theory and methodology

This section discusses the theoretical background and conceptual framework of the formulation of the novel GOF test for the joint model consisting of the multilevel survival and count data. It mainly explores the distributional properties and models for survival responses and count responses, concepts of the GOF tests establishing the foundation and the development procedure of the novel test.

2.1. Discrete time hazard model for survival data

With the clustered data scenario, when there exists an intra-cluster correlation, a multilevel random intercept discrete time model should be used (Jayanetti and Sooriyarachchi 2015). In this model intercepts are allowed to vary. Therefore, the scores on the dependent variable for each individual observation are predicted by the intercepts that varies across groups. This model assumes that slopes are fixed. Moreover, this model provides information about intra cluster correlations. Here, the intercept consists of two terms: a fixed component β_0 and cluster specific component u_{0j} where j corresponds to the bigger cluster which will be called cluster from now onwards. The model can be expressed as follows:

$$Logit(\pi_{gij}) = \beta_0 + u_{0j} + \sum_{g=1}^{n-1} \alpha_g \ T_{gij} + \ \beta X_{ij} \tag{1}$$

where $u_{0j} \sim N\left(0, \sigma_u^2\right)$.

Here, π_{gij} is the probability of the event happening in time interval g for the ith individual in the jth cluster given that the individual survived interval $g-1$ (hazard at time T_g), X_{ij} are covariates at the cluster level j and individual level i and T_{gij} is the gth time interval status for the *ith individual in the jth* cluster.

The third term on the right hand side of the model given in Equation (1) which represents a piecewise constant baseline hazard function can take the form of a continuous polynomial function (Goldstein 2011).

$$\sum_{g=1}^{n-1} \alpha_g \ T_{gij} = \ \alpha_1 \ T_{1ij} .... + \alpha_{n-1} \ T_{(n-1)ij} \tag{2}$$

Blocking factors can be used to model the baseline hazard function. Blocking factors are a set of dummy variables corresponding to the risk sets where the α's are parameters to be estimated for $g = 1$ (1) n-1. There is a dummy variable corresponding to each risk set. If one risk set is taken as the baseline then there are T_{g-1} dummy variables. For the model to be valid, effects of the covariates should be the same at all time points. This assumption can be tested by including interaction terms between predictors and time factors in the model. . The above mentioned discrete time model needs the proportionality assumption. This assumption is known as the 'proportional odds' assumption when the logit link is used. For it to be valid, effects of the covariates should be the same at all time points. (Stewart 2010). The proportional odds assumption can be tested by including interaction terms between predictors and time in the model. If there is a significant interaction that means that the covariate is time varying and hence should be included in the final model (Yang and Goldstein 2011). The other assumptions in this model are that the hazard rate is assumed constant within the observed time intervals and the random error term at the second level is assumed to be normally distributed.

2.2. Distributional properties and modeling count data

A random variable Y is said to have a Poisson distribution with parameter μ if it takes integer values y = 0, 1, 2, … with probability,

$$P(Y = y) = \frac{e^{-\mu}\mu^y}{y!} \ for \ \mu > 0 \tag{3}$$

The mean and variance of this distribution can be shown to be E(Y) = Var(Y) = μ.

Since the mean is equal to the variance, any factor that affects one will also affect the other. Thus, the usual assumption of homoscedasticity would not be appropriate for Poisson data.

$$Let \ Y_i \sim \ Poisson(\mu_i)$$

Random intercept models can be used to incorporate multilevel data. These types of models include two parts namely the fixed and the random parts. The random part takes the random effect of intercepts of the model. The fitted values of these models are explained by using the explanatory variables and the intercept that varies across groups. The link for the Poisson generalized linear model is the log link. In count data sometimes we deal with rates rather than counts. For this purpose we can use an additional parameter known as the "offset". This offset is set to be equal to the log (base e) of the expected number of cases (which is based on district population for each year). According to the Bristol University, workshop material [http://www.bristol.ac.uk/cmm/software/support/workshops/materials/modeling-count.ppt]. The model can be written as

$$\log(\mu_i) = \log(E_i) + \gamma_{0j} + X_i^{'}\gamma \qquad (4)$$

Where $\gamma_{0j} = \gamma_0 + u_{0j}$ and $u_{0j} \sim N(0, \sigma_{u_0}^2)$

Here, the intercept component of a random intercept consists of two parts: a fixed component β_0 and the random effect u_{0j}. The random effect u_{0j} is assumed to be distributed as normal with zero mean and constant variance $\sigma_{u_0}^2$. Other definitions mentioned in the previous section hold.

2.3. Joint models for survival and count data

2.3.1. Generalized linear mixed models

A generalized linear mixed model (GLMM) is an extension of the generalized linear model (GLM) in which the linear predictor contains random effects in addition to the usual fixed effects. They also allow response variables from different distributions. Hence, GLMMs inherit from GLMs the idea of extending linear mixed models to non-normal data.

Generalized linear mixed models (GLMMs) combine the properties of two statistical frameworks, linear mixed models (which incorporate random effects) and generalized linear models (which handle non normal data by using link functions for exponential family. GLMMs are the best tool for analyzing non normal data that involve random effects. GLMMs provide a broad range of models for the analysis of grouped data, since the differences between groups can be modeled as a random effect. These models are useful in the analysis of many kinds of data, including longitudinal data (Rizopoulos 2012). For using GLMM one has to specify a distribution, link function and structure of the random effects.

The general form of the model can be stated as follows:

$$y = X\beta + Zu + \varepsilon \qquad (5)$$

where y is a $N \times 1$ column vector of the outcome variable; X is a $N \times p$ matrix of the p predictor variables; β is a $p \times 1$ column vector of the fixed-effects regression coefficients (the β's); Z is the $N \times q$ design matrix for the q random effects (the random complement to the fixed X), u is a $q \times 1$ vector of the random effects (the random complement to the fixed β), and ε is a $N \times 1$ column vector of the residuals, that is the part of y that is not explained by the model, $X\beta + Zu$.

Fitting GLMMs via maximum likelihood involves integrating over the random effects. In general, those integrals cannot be expressed in analytical form. Various approximation methods have been developed, but none have good properties for all possible models and data sets (e.g., ungrouped binary data are particularly problematic). For this reason, methods involving numerical quadrature or Markov chain Monte Carlo have increased in use, as increasing computing power and advances in methods have made these more practical. For modeling GLMM few computational software packages are available. For this study SAS PROC GLIMMIX is used.

2.3.2. Jointly modelling discrete time hazard model with poisson regression model (Hapugoda and Sooriyarachchi 2018)

Though, data on the survival time of a patient and the count (incidence of the disease) are a frequently encountered phenomenon in medical studies, it is not common in the literature to see these two variables considered together as a bivariate response in a joint model. Hence, Hapugoda and Sooriyarachchi (2018) carried out a simulation study in order to develop a joint model of Discrete Time Hazard Model with Poisson Regression Model. This was often further complicated by the data being of a hierarchical nature (in the form of clusters). This simulation study has been carried out in order to study the effect of the sample size and the correlation between the two responses.
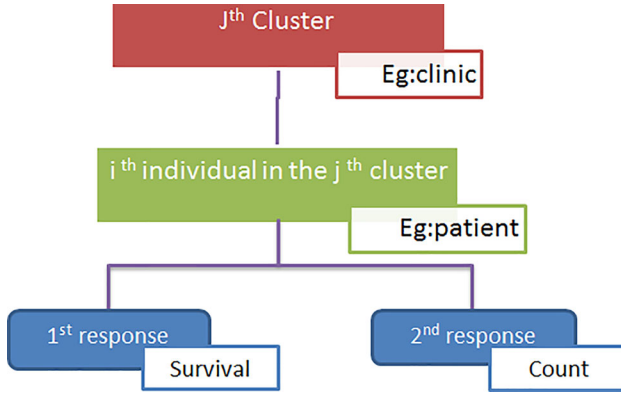
Figure 1. Hierarchical data structure.

The survival variable is continuous in nature with censoring information associated with it, while the count variable is discrete. Moreover, it is sometimes difficult to determine the distribution of the survival variable. To overcome the difficulty about the nature of the two response variables as explained earlier, the survival variable has been converted into a suitable discrete time approach and has been used for modeling survival data (Hapugoda and Sooriyarachchi 2018). This is a common method to analyze survival data and has been strongly advocated in MLwiN (Jayanetti and Sooriyarachchi 2015). Thus, an approach which is based on the data structure used in Discrete Time Hazard Model (DTHM) was carried out, as this will result in forming a binary survival response variable. The other response is counts in discrete form and thus this has can be considered as a Poisson or Negative Binomial distributed variable. Under this situation, a joint model can be constructed using the two responses. When the data are clustered, this joint modeling was constructed via Generalized Linear Mixed Models which perform maximum likelihood estimation based on a Laplace approximation of the marginal log likelihood. For the modeling, PROC GLIMMIX procedure in SAS software version 9.4 has been used.

Hierarchical data structure of the study can be illustrated as in Figure 1.

The model formulation is as follows. A multilevel random intercept discrete time model was used to model survival response.

$$Logit(\pi_{gij}) = \beta_{0j} + \sum_{g=1}^{n} \alpha_g T_{gij} + \beta X_{ij} \tag{6}$$

where

$$\beta_{0j} = \beta_0 + u_{0j}, \ \ u_{0j} \sim N(0, \sigma_u^2)$$

Here π_{gij} is the probability of the event happening in time interval g for the ith individual in the jth cluster given that the patient survived interval g-1 (hazard at time T_g) and X_{ij} are covariates at the cluster level j and individual level i where T_{gij} contains the unit-coded time indicators for the time periods assessed in the study (for an observation at time g, $T_{gij} = 1$ and for a time k where $k \neq g$, $T_k = 0$), X contains the values of the predictors, and α and β contain regression coefficients. Here β_{0j} denotes the random intercept and in order to join the two models random effects for two models were considered as the same (u_{0j}) and having the same variance (σ_u^2).

A two level random intercept Poisson Regression model (without loss of generalizability) was used to model count data and it is given by

$$\log(\mu_{ij}) = \log(E_{ij}) + \gamma_{0j} + \gamma X_{ij} \tag{7}$$

where

$$\gamma_{0j} = \gamma_0 + u_{0j} \ , u_{0j} \sim N(0, \sigma_u^2)$$

y_{ij} is the observed count for the ith observation in level 1, within the jth second level unit where $y_{ij} \sim$ Poisson (μ_{ij}) and μ_{ij} is the mean of the Poisson distribution for the ith observation of level 1, within the jth 2nd level unit. As mentioned previously an offset term is added to the Poisson model. E_{ij} is the Expected count or offset [$E_{ij} = $ Exposed$_{ij}$ * Rate, where Rate $=$ Total Count/ Total Exposed]

To formulate a joint model, Generalized Linear Model (GLM) can be used to form marginal models for each response by considering mean $E(Y_k|X_i)$ and variance $Var(Y_k|X_i)$ where $k = 1,2$. The approach to link the responses is by structuring a covariance matrix $Var(Y_k|X_i)$ to include potential correlations.

Here, the responses of analysis are: Y_{gij1} (binary – survived/not survived) and Y_{ij2} (count - number of patients). $l_1(\mathbf{u})$ is the logit link and $l_2(\mathbf{u})$ is the log link. GLIMMIX is used to estimate two marginal models jointly.

$$l_1(Y_{ij1}) = Logit(\pi_{gij}) = \beta_{0j} + \sum_{g=1}^{n} \alpha_g T_{gij} + \beta X_{ij} \tag{8}$$

$$l_2(Y_{ij2}) = \log(\mu_{ij}) = \log(E_{ij}) + \gamma_{0j} + \gamma X_{ij} \tag{9}$$

The joint model variance-covariance matrix $= \begin{bmatrix} \sigma_u^2 & \rho_{12}\sigma_u^2 \\ \rho_{12}\sigma_u^2 & \sigma_u^2 \end{bmatrix}$. Here $\boldsymbol{\rho}_{12}$ is the correlation between the two responses.

It is important to mention that the developed joint model assumes the exchangeability structure for the correlation matrix. That is the correlation between the same outcome variable within the cluster and the correlation between different responses with the same individual are assumed to be the same. i.e., $corr(Y_{kij1}, Y_{ki'j1}) = corr(Y_{kij1}, Y_{kij2})$ (Hapugoda and Sooriyarachchi 2018).

2.3.3. Joint modeling of survival data and count data for both simple randomized and cluster randomized data (Sunethra and Sooriyarachchi 2020)

The objective of Sunethra and Sooriyarachchi's study was to develop an improved analytical solution for analyzing data with correlated survival and count responses for two common designs of simple randomized and cluster randomized data. This can be a complex scenario of above discussed Hapugoda and Sooriyarachchi (2018) method. This study concerns two random effects (at the individual and cluster levels) whereas Hapugoda and Sooriyarachchi (2018) considered a common random effect. Sunethra and Sooriyarachchi's (2020) example considered epilepsy patients and the time to an attack was the survival variable.

The study initially considered the development of the joint model for simple randomized data and subsequently, the extension of the joint model for cluster randomized data was considered. Since our study is based on a cluster scenario only cluster randomized models are discussed under this section.

Under this approach, instead of a single shared random effect defined at cluster level u_{oj}, two separate random effects one defined at patient level (used to represent the correlation between the responses at the patient level) and the other defined at cluster level (leading to the distribution of the patient level to be conditioned on the cluster level random effects) are used. Let v_{1ij}, v_{2j} be two vectors of random effects defined at the ith patient of the jth center where $i = 1, 2, ...n_j$ and $j = 1, 2, ..., m$. Let u_{0j} denote a vector of random effects for the jth cluster, where $j = 1, 2, ., m$. Since patients are nested within clusters, the two levels of random effects should have a nested structure that is (v_{1ij}, v_{2ij}) nested within u_j. Then the joint model for the two responses can be written as,

$$Y_{ij}|v_{1ij}, u_j \sim Poisson \ (\lambda_{ij})$$

$$\log(\lambda_{ij}) = X_{1ij}^T \gamma + v_{1ij} + u_j \tag{10}$$

$$T_{ij}|v_{2ij}, \quad u_j \sim Lognormal(\mu_{ij}, \ \sigma_{ij})$$

$$\mu_{ij} = X_{2ij}{}^T\beta + v_{2ij} + u_j \qquad (11)$$

where $\boldsymbol{v} = (v_{1ij}, v_{2ij}) \sim Normal(0, \Sigma)$

$U_j \sim Normal(0, \sigma_u^2)$ and $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$

where Σ is the variance-covariance matrix and all its elements are scalers. Then, the joint distribution of the two responses was derived as follows:

$$f(t_{ij}, y_{ij}) = \int \int f(t_{ij}, \ y_{ij}|v, u_j)f(\boldsymbol{v}) \ f(u_j)dvdu_j \qquad (12)$$

The parameter estimation of the joint model involves maximizing the above joint likelihood which involves two levels of nested random effects .According to the results of the application of a real life data set, the proposed joint model was compared with the compatible joint model defined with a single level of random effects as well. The performance of the proposed joint model was undoubtedly better than the joint model with single level of random effects. However, it is a more complex model and in some instances there could be convergence problems when compared to the model of Hapugoda and Sooriyarachchi (2018).

2.3.4. Current joint model used for the study

The data was simulated according to Sunethra and Sooriyarachchi's (2020) approach. However, this approach was too complicated for constructing the GOF test due to non-convergence of the models required for the GOF test. Thus the simpler model of Hapugoda and Sooriyarachchi was used in the modeling stage.

According to Hapugoda and Sooriyarachchi (2018) a two-level single random-effect joint model was explained, where i represents an individual and j represents a cluster. For simplicity, models are re-stated here. These two marginal models are considered to fit the joint model and perform the GOF test. Development of the GOF test will be discussed later on.

$$Logit(\pi_{gijk}) = \beta_{01jk} + \sum_{g=1}^{n} \alpha_g \ T_{gij} + \ \beta_1 X_{ijk} \qquad (13)$$

where

$$\beta_{01jk} = \ \beta_{01} + u_{0k}, \ u_{0k} \sim N\left(0, \sigma_u^2\right)$$
$$\log(\mu_{ijk}) = \ \beta_{02jk} + \ \beta_2 X_{ijk} \qquad (14)$$

where

$$\beta_{02jk} = \ \beta_{02} + u_{0k}, \ u_{0k} \sim N\left(0, \sigma_u^2\right)$$

2.4. Previously developed goodness of fit tests used as a basis for this paper

Goodness of fit measures that apply to continuous response models are not always appropriate to check the adequacy of fitted discrete response models. In the presence of one or more continuous explanatory variables in the model, this situation becomes worse. A continuous explanatory variable causes the covariate space to be more complex and expanded, resulting in very small cell frequencies in the contingency table obstructing the chi-square distributional assumption in Pearson's chi-square test and deviance test. Therefore, many researchers have sought various

approaches to overcome this shortcoming through developing customized GOF tests for discrete response models.

In 1980 Hosmer and Lemesbow revealed one such development that is widely used to assess the adequacy of logistic regression models, well-known as the Hosmer–Lemesbow test. Lipsitz, Fitzmaurice, and Molenbergh (1996) have extended the Hosmer – Lemesbow test through a model-based design which has made the procedure easier to implement. Wang Fagerland, Hosmer, and Bofin (2008) have proposed a GOF test for multinomial logistic regression models that is useful owing to the fact that many other types of responses can be converted into multinomial distributions. The GOF test elevation from single level to multilevel context by Perera, Sooriyarachchi, and Wickramasuriya (2016) is a milestone in the progress of GOF tests for various logistic regression models. These four research works have formulated a secure theoretical background for establishing this continuation of the progress into the multivariate framework.

2.5. Development of the novel goodness of fit test

For developing the current GOF test the following steps are used.

Step 1: The two-level random-effects joint model for survival and count data as defined by Hapugoda and Sooriyarachchi (2018) is fitted and the parameters of the model are estimated.

Step 2: The predicted values of the two responses $Expected_{1gjk}$ and $Expected_{2jk}$ are obtained from the fitted joint model for each jth individual clustered within each kth cluster respectively. It is straightforward to obtain these expected values directly in the form of $\widehat{\pi_{1gjk}}$ and $\widehat{\mu_{2jk}}$ from the joint model.

Binary response

$$Predicted\ probability_{1gjk} = \frac{\exp(\widehat{\pi_{1gjk}})}{1 + \exp((\widehat{\pi_{1gjk}})} \tag{15}$$

Count response

$$Expected\ count_{2jk} = \exp(\widehat{\mu_{2jk}}) \tag{16}$$

Step 3: First, Sort predicted values according to $Predicted\ probability_{1gjk}$.

Step 4: Then collapse sorted predicted values into 10 (G) groups within each cluster. Hosmer and Lemesbow (1980) recommend that the optimal value for G is 10. As a general rule G is chosen so that $6 \leq G < n/5K$ where n is the sample size and K is the number of clusters. Then give the ranks $G = 1, 2 \ldots 10$ based on the group that it belongs to. (Perera, Sooriyarachchi, and Wickramasuriya 2016)

Step 5: After the ranks are assigned within each cluster, the observations are re-arranged according to the observation index and the ranks are pooled over the clusters to define in overall 10 groups for the data. This means, all the group 1 observations in each cluster are pooled into a combined single group (Perera, Sooriyarachchi, and Wickramasuriya 2016). Then indicator variables are defined as,

$$I_{G1gij} = \left\{ \begin{array}{l} 1,\ if\ Predicted\ probability_{1gjk}\ is\ in\ group\ G \\ 0,\ otherwise \end{array} \right\}$$

where $G = 2, 3, \ldots, 10$ and the 1st group is taken as the reference group.

Step 6: All the $Expected\ count_{2jk}$ values in the second response are similarly arranged in each cluster and pooled into another combined single group and so on. Nine indicator variables are assigned to capture these 10 groups as follows.

Rearrange the data set by sorting according to the observation index. Repeat step 5 and rank these. Assign indicator variables as discussed above.

$$I_{G2ij} = \left\{ \begin{array}{ll} 1, & \text{if Expected count}_{2jk} \text{ is in group } G \\ 0, & \text{otherwise} \end{array} \right\}$$

Step 7: These indicator variables are introduced to the joint model given by (8) and (9) such that the 2 sets of indicator variables are taken as a covariate in the joint model and the following alternative model is obtained.

For response 1 this yields the model,

$$Logit(\pi_{1gjk}) = \beta_{01} + u_{0k} + \sum_{g=1}^{n-1} \alpha_g \ T_{gjk} + \ \beta_1 X_{ij \ k} + \sum_{G=2}^{10} \gamma_{1G} I_{G1jk} \tag{17}$$

For response 2 this yields the model,

$$\log(\mu_{2jk}) = \log(E) + \beta_{02} + u_{0k} + \ \beta_2 X_{ijk} + \sum_{G=2}^{10} \gamma_{2G} I_{G2jk} \tag{18}$$

Inclusion of the indicator variable has introduced new parameters into the model.
Step 8: If the joint model is correctly specified all these new parameters should simultaneously be equal to zero. Hence, now the interest is in testing the hypothesis,

$$H_0 \quad : \quad \gamma_{1,2} = \ldots = \ \gamma_{1,10} = \ \gamma_{2,2} = \ldots = \ \gamma_{2,10} = 0 \text{ Versus}$$

$$H_1 \quad : \quad \text{At least one one of these parameters is not equal to zero}$$

The test statistic follows a F distribution with $(18, n_2)$ degrees of freedom where $n_{2=}$ (number of groups - no of independent parameters).
Step 9: The calculated F statistic is compared with the critical $F_{18, n2}$ value at 5% significance level. If the calculated F statistic $< F_{18, n2;5\%}$, H_0 is not rejected at the 5% significance level, implying all new parameters are simultaneously equal to zero. Hence, the joint model adequately fits the data. If the calculated F statistic $\geq F_{18, n2; \ 5\%}$, H_0 is rejected at the 5% significance level, implying at least one new parameter is significantly different from zero. Thus, the joint model does not adequately fit the data.

3. Simulation study

The development of the novel GOF test for 2-level survival and count responses was explained step by step in the previous section. Nevertheless, it is important to confirm its main properties, namely, Type I error and power before generalizing the developed GOF test for practical usage. Generally, a well-developed test would have an acceptable type I error and fairly high power. These two properties are examined through simulation studies by considering various combinations of parameters associated with the model. This section explains the approach of carrying out the simulation study and results and conclusions are given.

3.1. Steps in the simulation study

The steps in the simulation study are explained as points in a) to h)

a. Generate 1000 data sets from different sample sizes such that, one response is binary (survival response) and the other is counts (Poisson) using GLIMMIX procedure in SAS version 9.4. Here we simulate data such as from Sunethra and Sooriyarachchi's (2020) model.
- Responses are correlated as these are within one individual
- Consider only one explanatory variable (X)

Table 1. Details of simulation results.

Sample size	Standard deviation of u (response1)	Standard deviation of v (response2)	Type I error rate	Nonconverged data sets	Withinlimits or not?	Power	Nonconverged data sets
100 (5 Clusters of size 20)	1.225	0.866	0.054	2	yes	0.300	3
	1.225	0.922	0.048	2	yes	0.314	3
	1.323	0.866	0.048	3	yes	0.292	3
	1.323	0.922	0.047	2	yes	0.32	3
	1.414	0.866	0.044	2	yes	0.316	3
	1.414	0.922	0.049	1	yes	0.289	3
300 (5 Clusters of size 60)	1.225	0.866	0.046	2	yes	0.643	0
	1.225	0.922	0.051	0	yes	0.659	0
	1.323	0.866	0,051	0	yes	0.627	0
	1.323	0.922	0.045	0	yes	0.655	0
	1.414	0.866	0.047	0	yes	0.621	0
	1.414	0.922	0.049	0	yes	0.581	0
500 (8 Clusters of size unequal (60 or 65)	1.225	0.866	0.056	0	yes	0.779	0
	1.225	0.922	0.066	0	Just outside	0.764	0
	1.323	0.866	0.040	0	yes	0.759	0
	1.323	0.922	0.057	0	yes	0.782	0
	1.414	0.866	0.041	0	yes	0.716	0
	1.414	0.922	0.052	0	yes	0.734	0
750 (10 clusters of size 75)	1.225	0.866	0.052	0	yes	0.803	0
	1.225	0.922	0.057	0	yes	0.857	0
	1.323	0.866	0.043	0	yes	0.768	0
	1.323	0.922	0.053	0	yes	0.813	0
	1.414	0.866	0.044	0	yes	0.740	0
	1.414	0.922	0.041	0	yes	0.786	0
1000 (12 clusters size- 83 and 87 for the last cluster	1.225	0.866	0.052	0	yes	0.827	0
	1.225	0.922	0.087	0	No	0.867	0
	1.323	0.866	0.052	0	yes	0.819	0
	1.323	0.922	0.063	0	yes	0.876	0
	1.414	0.866	0.046	0	yes	0.784	0
	1.414	0.922	0.052	0	yes	0.847	0

b. Fit a joint model by using Discrete Time Hazard Modeling (DTHM) approach to the binary data and Negative Binomial model to the count data as the Poisson has an excess of zeros. Here though we simulated data from Sunethra and Sooriyarachchi's (2020) model due to convergence problems Hapugoda and Sooriyarachchi (2018) approach is used to fit the model.

c. Extend the GOF test by considering the basis of Hosmer Lemesbow test (1980) and Lipsitz, Fitzmaurice, and Molenbergh (1996) for the single level to a two level case.

d. From the model get the predicted values and apply the H-L test within each cluster within each response.

e. Determine the proportion of rejections of the null hypothesis after all the indicator variables from both responses are put together and sorted according to the original order. From this, estimate the Type 1 error.

f. The test will be carried out by changing the standard deviation which will also change the ICC corresponding to the random effect of the cluster levels. Also change the cluster size and the number of clusters.

g. By simulating from a wrong model we determine the power of the test. In this model we take the explanatory variable as $\log_e X^2$.

3.2. Simulation Results - Type1 error rate and power

The type I error and power are given in Table 1 and also includes the sample size, cluster size, the number of clusters, the random variance parameter, the type I error, power and the non-convergent data sets in 1000 simulations.

3.3. Conclusions based on the simulation results

Once the novel test is developed it is essential to examine its performance under varying factors affecting the setting. The performance of such a GOF test is mainly evaluated through type I error and power. Underlying these data generation procedures the results of the simulation studies were carried out to assess these two properties as were outlined in this section. The novel GOF test was able to easily maintain the type I error at most levels and produced higher power when the cluster size is large. Additionally, the percentage of total variance from second level (large cluster) variance has an impact on power and the type I error rate. Finally it can be concluded that, the ideal conditions for the correct usage of this novel GOF test can be considered as designs with small cluster sizes when considering the type I error rate. However, the test performs well for large cluster sizes when considering the power.

4. Example

This section provides a better understanding of the developed novel GOF test by applying it to an example data set. This helps to verify the applicability of the developed test. The developed test is based on a hierarchical nature. Therefore, it is necessary to consider a multilevel structure when applying the test to a real life data set. Multilevel Survival and count responses are encountered often in medical studies. Hence a Dengue dataset, which consists of data from all over Sri Lanka from year 2006 to 2008 was chosen for this purpose.

4.1. Description of the data set

This study has been selected for illustrating the developed method since it is based on multilevel survival and count responses. Therefore, the application of the test is aimed at a dataset with these responses within a three-level hierarchy to maintain the simplicity. This data set is appropriate to apply the developed test since it includes the above characteristics. This same data set has been used by Hapugoda and Sooriyarachchi (2018) for fitting a joint model.

The original data set comprises data from 22 districts in Sri Lanka along with survival data and climate data. It includes 13 explanatory variables, a count response and a survival response and consists of 528 observations. The survival response is the time to death of a patient discretized into 3 intervals. If the patient does not die the patient is censored. The count variable is the incidence of dengue within district within month within year that corresponds to the patient. Explanatory variables are taken as the 3 climate variables rainfall, temperature and humidity. Moreover, these climate attributes have been recorded under time lags. That is climate data include rainfall, temperature, humidity and their first and second lags for each district, in each month and in each year separately. The survival times were computed for individual records of patients for each district, each month in each year. The hierarchical structure is made up of three levels where the first level consists of the two responses; count and survival, the second level is a patient and the third level is the district. The original data set had to be rearranged such that it matches with the theories discussed in chapter 2.

The original data set consists of complete cases, ignoring missing values. Hence no missing value imputation method was used. A district represents a cluster according to this study. The problem encountered was that some districts consist of only a few observations. Consequently, the cluster sizes of those districts become low. According to the simulation study, the smallest cluster size was 20. Hence it is not appropriate to use the small cluster sizes below 20. After removing these small district records, 423 observations were remaining. In the simulation study of the GOF test, the maximum cluster size was 87. However, in the example every district

consists of a lower number of records than 87. Hence there is no issue about the upper limit of the cluster sizes.

It is important to note that the "outcome" variable consists of 3 levels discharged, transferred and dead. Since there were only a few observations in the transferred category, they have not been included in the data set. The "Discharged" category is encoded as 0 and "dead" category is encoded as 1. This is taken to be the censoring indicator for the survival response. The survival time is continuous as usual. As discussed in Sec. 2 continuous survival time was discretized into pre-determined time intervals according to Kalayanarooj (2011) such that the survival time of a dengue patient can be categorized into 3 phases. Those are febrile phase (<7 days), critical or leakage phase (7–9 days) and convalescence phase (> 9 days). Then the outcome variable can be linked with these 3 categories such that the survival response is converted into a binary form. This discretizing process was done according to the theory explained in Hapugoda and Sooriyarachchi (2018).

Climate data has been recorded for each month. However, for the simplicity and the higher performance of the model building the month attribute was converted into quarters.

For model fitting log values of rainfall variables and expected counts variable were considered in order to get over convergence problems. For the count model, an offset as explained previously was fitted.

4.2. Model fitting

It is necessary to identify the variables that have a significant impact on the responses. After data pre-processing, as the next phase univariate random intercept models for each response were fitted first in order to compare the performance with the joint model. Then the bivariate random intercept joint model was fitted and the model adequacy is assessed using the novel GOF test. The forward selection method was used to accomplish the task of selecting the explanatory variables for the most appropriate model. Inclusion or exclusion of an explanatory variable was decided on the basis of F statistics performed at each stage. Initially, separate models were fitted for each explanatory variable and the most significant explanatory variable was included to the model first.

In each step of the forward selection the variable that has the highest F statistic among the significant F-test (p-value$< =0.05$) was selected. Since forward selection method has been used to select variables, once a variable is included in the model it will not be removed from the model. Univariate multilevel model fitting has been discussed by Perera, Sooriyarachchi, and Wickramasuriya (2016) while joint multilevel model fitting was explained in this paper previously.

Model fit statistics for the survival response is shown in the following output.

AIC and BIC values of the fitted logit model are

AIC-4109.26; BIC- 4109.5

The univariate log model is fitted for the count response using Negative Binomial distribution. Now time indicator variables should not be used as covariates/factors since they are not taken into account for the count response. Each and every explanatory variable is included into the model, one at a time as mentioned in model fitting. LnExp attribute has been set to offset. The final model fit statistics for the log model are represented as follows.

AIC and BIC scores are used to check the performance of the fitted log model. According to this model

AIC = 21130.48; BIC= 21137.56

As the next phase, a bivariate random-intercept joint model was fitted, of which the model adequacy is assessed using the novel GOF test. The forward selection method was used to accomplish the task of selecting the explanatory variables for the most appropriate model. Inclusion or

Table 2. Final joint model coefficients.

Factor	Response	Category	$\hat{\beta}$ (std error($\hat{\beta}$))	t-value	p-value
Intercept	2	–	−9.7993(0.971)	−10.09	<.0001*
t_1	1	0	−2.5161 (0.243)	−10.34	<.0001*
t_2	1	0	−2.0832 (0.247)	−8.45	<.0001*
Temp_lag21	2	–	0.2711 (0.035)	7.72	<.0001*
Qtr	2	1	−0.8422 (0.080)	−10.53	<.0001*
		2	−0.5514 (0.065)	−8.44	<.0001*
		3	−0.2821 (0.068)	−4.17	<.0001*
lnRainfall_lag21	12	–	−0.1014(0.05)	−2.046.84	0.0415*
			0.1615 (0.024)		<.0001*
Year1	1	1	−2.0002 (0.140)	−14.30	<.0001*
		2	−1.8964 (0.147)	−12.90	<.0001*
	2	1	−0.5178 (0.063)	−8.18	<.0001*
		2	−0.4976 (0.066)	−7.52	<.0001*

exclusion of an explanatory variable was decided on the basis of F statistic of significant variables at each stage. The procedure ended in 6 steps. LnExp was taken as an offset variable.

It is essential to note that the null model of the negative binary part included the constant term and in addition the binomial part included time indicators t_1, t_2. The final joint model summary is represented in Table 2.

Table 2 shows that for the discrete time survival component of the model both time intervals t_1 and t_2 are highly significant with negative coefficients, indicating that time intervals 1 and 2 have less deaths compared to time interval 3. The odds ratios for the two intervals are 0.081 and 0.125 respectively. For this response the log of rainfall lag2 is also significant and negative, indicating that as this increases the number of deaths decreases. The odds ratio for a unit increase in log rainfall lag2 is 0.904. Also the year is significant with negative coefficients for years 1 and 2 compared to year3. This indicates that there are less deaths in years 1 and 2 compared to year3. The odds ratios are 0.135 and 0.15, respectively.

For the count response as temperature lag2 increase the counts of dengue increases and the odds ratio of two values of temperature lag2 one unit apart is 1.31.The quarter is also important for response 2. The first 3 quarters having less dengue cases than the 4th quarter. The odds ratios are respectively 0.43, 0.576 and 0.754. When Log rainfall lag2 increases the dengue counts increases and the odds are 1.18. As for response 1 for response 2 as well there are less dengue cases in years 1 and 2 compared to year3. The odds being 0.596 and 0.608, respectively.

AIC and BIC values of the fitted joint model are
AIC − 12819.19
BIC- 12834.77
Model performance
Total AIC value for univariate two models= 4109.26 + 21130.48 = 25239.74
AIC value for the joint model= 12819.19
Total BIC value for separate two models= 4109.5 + 21137.56 = 25247.06
BIC value for the joint model= 12834.77
AIC and BIC values are much lower for the joint model than for the total of separate models. Hence the fitted joint model performs better than the two univariate models.

4.3. Testing the GOF test

After the multilevel joint model has been fitted, as the next phase, the developed test should be applied to examine the adequacy of the fitted joint model. In count data sometimes we deal with rates rather than counts. For this purpose, we can use additional parameter known as "offset". It is important to mention that "LnExp" is set to be offset in GLIMMIX procedure when

Table 3. GOF test results.

Effect	Num DF	Den DF	F Value	Pr > F
Intercept	2	58	13.04	<.0001
t_1	2	58	1.21	0.3070
t_2	2	58	1.69	0.1930
Temp_lag21	2	58	9.15	0.0004
Qtr	6	58	0.90	0.4979
lnRainfall_lag21	2	58	1.19	0.3120
Year1	4	58	2.20	0.0803
I	18	58	0.71	0.7881

performing the GOF test and all model fittings since Exposed$_{jk}$ * Rate where Rate = Total Count/ Total Exposed.

According to this data set, the multilevel structure is made up based on districts. After data cleaning, 15 districts were remaining in the data set. These 15 districts represent 15 clusters. Since the number of patients within a district is not equal in each district it results unequal cluster size scenario. However, in the simulation study of developing the novel GOF test, both equal and unequal cases were included.

By using the fitted joint model expected values for each individual are taken for each response. By considering one response at a time, expected values are sorted and collapsed within each cluster into G = 10 groups and indicator variables were defined. In the example data set, most of the clusters consisted of observations that were not divisible by 10. Therefore, the allocation was done as in Fernando and Sooriyarachchi (2020).

Then the GOF test was performed under the methodology discussed in chapter 2.

Given the partitioning of the data, the goodness of fit statistic is formulated by defining 9 group indicator variables for each cluster for the two responses as follows.

$$I_{G1gjk} = \left\{ \begin{array}{c} 1, \quad if \quad Predicted \ probability_{1gjk} \quad is \ in \ group \ G1 \\ 0, \ otherwise \end{array} \right\}$$

Where $Expected_{1gjk}$ is the expected value for the first response jth patient in the kth district given that he/she has survived up to the g-1th time interval.

$$I_{G2jk} = \left\{ \begin{array}{c} 1, \quad if \quad Expected \ count_{2jk} \quad is \ in \ group \ G2 \\ 0, \ otherwise \end{array} \right\}$$

Where $Expected_{2jk}$ is the expected value for the second response jth patient in the kth district. For the goodness of fit test the null and the alternative hypotheses are as follows.

H$_0$: The joint model fits the data well

H$_1$: The joint model does not fit the data well

Then for assessing the GOF test of the final joint model the indicator variables are constructed and added to the final selected model.

If the joint model is correctly specified, the null hypothesis is not rejected. It indicates that,

$$\gamma_{1,1} = \cdots = \gamma_{1,9} = \gamma_{2,1} = \cdots = \gamma_{2,9} = 0$$

SAS 9.4 is used to get the F statistic of the indicator variable for the joint model.

Summary of the Type III tests of the fixed effects are given in Table 3.

By using the SAS software the p–value corresponding to the indicator variable can be used to check the GOF of the test. The type III tests of the fixed effects show that the F value of indicator variables (I) is 0.71 with a p-value of 0.7881.

Since there are 18 indicator variables for the joint model the degrees of freedom of the numerator is 18. The degrees of freedom of the denominator is 58 (number of groups - no of independent parameters).According to the p-value it can be seen that the indicator term is not at all significant since it is much greater than 0.05. As well, the critical value of $F_{18,58;5\%} = 1.78$ which

is greater than 0.71. According to the results it can be concluded that the null hypothesis,

$$H_0: \quad \gamma_{11} = ... = \quad \gamma_{1,9} = \quad \gamma_{21} = ... = \quad \gamma_{2,9} = 0,$$

is not rejected at 5% level of significance.

Hence the joint model without indicator variables fits the data well

5. Discussion and conclusion

This section provides an overall discussion of the developed novel goodness of fit test.

5.1. Goodness of fit test for multilevel joint model of survival and count data

After constructing the model and estimating the model parameters it is necessary to assess the model adequacy before it is used to make inferences. A well fitted model explains data properly and it can be used to make valid inferences for the entire population by generalizing the results. In order to address this issue goodness of fit tests come in to play.

This research mainly discusses multilevel joint survival and count data. There are many developed GOF tests for single level univariate models of count and survival data. In this study for modeling survival data Discrete time Hazard modeling approach is used. This results in a logistic regression model. Logistic regression models are used in many areas of scientific research (Liu 2007). The widely used Hosmer–Lemesbow GOF test is a test developed by Hosmer and Lemesbow (1980) based on two grouping strategies for predicted probabilities of the fitted models for single-level binary data.

To model multilevel binary data random intercept logistic regression models are used. A recent work related to developing GOF test for multilevel binary data has been published by Perera, Sooriyarachchi, and Wickramasuriya (2016) based on Hosmer-Lemesbow test and the extension of that by Lipsitz et al .(1996). Lipsitz, Fitzmaurice, and Molenbergh (1996) extended the Hosmer-Lemesbow test into a model-based approach by introducing indicator variables to denote the deciles of risk.

However, the above-mentioned proposed GOF tests can be used only for univariate models. When it comes to joint models developed GOF tests are not very common in literature. Survival responses and count responses can most often occur in medical studies. It is more interesting when data is hierarchical.

As Sunethra and Sooriyarachchi (2020) found that their test was superior to Hapugoda and Sooriyarachchi's (2018) test our initial objective was to develop a GOF test for Sunethra and Sooriyarachchi's test. Thus, the simulations were done under this scenario. However, this model did not converge very often when the indicator variables were added. Thus, we had to revert back to the simpler model of Hapugoda and Sooriyarachchi (2018) for which the model with indicator variables converged well. This also showed a very important result which is if the data structure was of a complex nature such as Sunethra and Sooriyarachchi (2020), a simple model such as Hapugoda and Sooriyarachchi (2018) fitted well.

Hapugoda and Sooriyarachchi (2018) has developed a joint model for such hierarchical survival and count data using common random effect for both responses. However, there is no satisfactory goodness of fit test for multilevel joint model of survival and count data.

Hence the main objective of this study was to develop a novel goodness of fit test to assess the model adequacy of a multilevel joint model of survival and count data (Hapugoda and Sooriyarachchi 2018). The secondary objective of this study was to identify the properties of the developed novel GOF test. In order to achieve the second objective a simulation study was carried out by changing sample sizes, cluster sizes, number of clusters and different standard deviation of random effects. Then the type 1 error rate and the power of the novel test were determined for

each combination. It provides how the test performance varies for each scenario. Finally the developed GOF test was applied to a real world data set in order to check the applicability of the developed GOF test.

5.2. Discussion of the results under the simulation study

It is essential to evaluate the performance of a test before it is used. The performance of the test is appraised through type I error and power. Thirty different combinations of sample sizes and variances of random effects were used in the simulation study. For each combination 1000 data sets were simulated to estimate the type I error rate.

Type I error rates were maintained at nominal level for 28 of all the combinations used in the simulation study. Only 1 case just deviated and another case severely deviated from the acceptable limits (particularly exceeding the upper limit). However, the test performs well for 28 cases out of 30 which is 93.33%. Moreover, for small sample sizes 100 and 300 the Type I error rate of the novel GOF test remains within the 95% probability interval. Type I error is slightly outside for one combination out of 6 for the moderate sample size 500. All cases for sample size 750 are within limits. Nevertheless, type I error rate for one combination of 1000 sample size is quite a lot outside from the probability interval. It is explainable according to Musca et al. (2011). The inflation in Type-I error is more pronounced if the number of higher- order units is low, the number of lower-order of units is high, and the ICC is high.

In general, it could be concluded that the novel GOF test performs well for small to moderate cluster sizes. When the total variance of the second level is high the performance of the developed GOF test gets lower.

The power of the developed test was evaluated through simulating the data under the alternative hypothesis and checking what proportion rejects the null hypothesis and 1000 data sets were generated under each combination.

None of the combinations were able to produce 100% power. However, the novel test could produce satisfactory powers having 87.6% as the highest. Power for sample 100 is quite low while it is at a moderate level for 300 sample size. Test produces higher powers for 500,750 and 1000 combinations.

Finally, it can be concluded that the power exponentially increased when sample size and the cluster size increase. However, it is important to note that cluster size but not the sample size that determines the power of the test.

Furthermore, power simulations exhibited interesting relationships between the power of the test and the total variance. It revealed that the lowest total variance combination ($\sigma_u = 1.414$, $\sigma_v = 0.866$) has lower powers with compare to other variance combinations. This is the vice versa of the results of type I error rate.

5.3. Application to a real-world data set

The developed GOF test can be applied even if there are one or more explanatory variables and the cluster sizes are unbalanced. Initially two separate models were constructed for the two responses. Then a joint model of multilevel survival data and count data was fitted using the approach of Hapugoda and Sooriyarachchi (2018). According to BIC and AIC values the joint model performs better when compared to two separate models. As the next phase the developed novel GOF test was used to check the model adequacy of the joint model. According to the result it is indicated that the fitted joint model is adequate. It is important to mention that the fitted model is quite interpretable.

5.4. Limitations of the study

There were a few limitations when carrying out this study.

The ideal model to fit is the nested model of patient within a clinic/district. However, this model of indicator variables gives results only for a few simulations. This gets worse with increasing sample size. This is because one patient has only 2 observations. Thus, the convergence is difficult. Therefore, it was compulsory to use only the upper level cluster for the subject effect. i.e., the study is considering only the big cluster even if two random effects were used in simulating data. Another limitation is that the power is low for small clusters. Also for joint modeling Negative Binomial distribution was used instead of Poisson since the data consists of excess of zeros.

References

Fernando, G., and Sooriyarachchi, R. 2020. The development of a goodness-of-fit test for high level binary multi-level models. *Communication in Statistics – Simulation and Computation*. Advance online publication. doi:10.1080/03610918.2019.1700275.

Goldstein, H. 2011. *Multilevel statistical models*. 4th ed. UK: Wiley Series in Probability and Statistics.

Hapugoda, J. C., and M. R. Sooriyarachchi. 2018. Joint modeling of discrete time hazard model with poisson regression model: A simulation study. Jaffna University International Research Conference, Jaffna, Sri Lanka.

Hosmer, D. W., and S. Lemesbow. 1980. Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics – Theory and Methods* 9 (10):1043–69. doi:10.1080/03610928008827941.

http://www.bristol.ac.uk/cmm/software/support/workshops/materials/modelling-count.ppt. December, 2013.

Jayanetti, W., and R. Sooriyarachchi. 2015. A multilevel study of dengue epidemiology in Sri Lanka: Modeling survival of dengue patients. *International Journal of Mosquito Research* 2 (3):114–21.

Kalayanarooj, S. 2011. Clinical manifestations and management of dengue/DHF/DSS. *Tropical Medicine and Health* 39 (4 Suppl):83–87. doi:10.2149/tmh.2011-S10.

Kreft, I., and J. de Leeuw. 1998. *Introducing statistical methods. Introducing multivlevel modeling*. London: SAGE Publications, Inc.

Lipsitz, S., G. M. Fitzmaurice, and G. Molenbergh. 1996. Goodness-of-fit tests for ordinal response regression models. *Applied Statistics* 45 (2):175–90. doi:10.2307/2986153.

Liu, Y. 2007. On goodness-of-fit of logistic regression model [Unpublished Dissertation]. https://krex.k-state.edu/dspace/bitstream/handle/2097/530/YingLiu2007.pdf?sequence=1

Musca, S.C., Kamiejski, R., Nugier, A., Meot, A., Er-Ratify, A., and Brauer, M. 2011. Data with hierarchical structure: Impact of intraclass correlation and sample size on Type-I error. *Frontiers in Psychology* 2 :74. doi:10.3389/fpsyg.2011.00074.

Perera, A. A. P. N. M., M. R. Sooriyarachchi, and S. L. Wickramasuriya. 2016. A goodness of fit test for the multi-level logistic model. *Communications in Statistics - Simulation and Computation* 45 (2):643–59. doi:10.1080/03610918.2013.868906.

Ranathunga, N. O., and R. Sooriyarachchi. 2017. Multivariate multilevel modeling of age related diseases. *Journal of Modern Applied Statistical Methods* 16 (1):498–517. doi:10.22237/jmasm/1493598540.

Rizopoulos, D. 2012. *Joint models for longitudinal and time-to-event data with applications in R*. Glasgow: CRC Press.

Stewart, C. H. 2010. Multilevel modelling of event history data: Comparing methods appropriate for large datasets. Doctoral dissertation. University of Glasgow.

Sunethra, A. A., and M. R. Sooriyarachchi. 2020. A novel method for joint modeling of survival data and count data for both simple randomized and cluster randomized data. *Communications in Statistics – Theory and Methods*. Advance online publication. doi:10.1080/03610926.2020.1713366

Wang Fagerland, M., Hosmer, D., and Bofin, A. M. 2008. Multinomial goodness-of-fit tests for logistic regression models. *Statistics in Medicine* 27 (21):4238–53.

Yang, M., and H. Goldstein. 2011. *Modelling survival data in MLwiN 1.20*. London: Institute of Education University of London.