*American Journal of Applied Mathematics and Statistics, 2018, Vol. 6, No. 2, 72-79*
Available online at http://pubs.sciepub.com/ajams/6/2/6

DOI:10.12691/ajams-6-2-6

# A Joint Model for Exponential Survival Data and Poisson Count Data

**Abeysinghe A Sunethra*, Marina R Sooriyarachchi**

Department of Statistics, university of Colombo, Sri Lanaka
*Corresponding author: sunethra@stat.cmb.ac.lk

**Abstract** This research is motivated by the correlated outcomes of survival times and counts particularly in medical data. Analysis on individual patient data consisting of recurrent events of disease progression and survival times/death times give better results when a bivariate model/joint model of 'survival' and 'count' is used rather than fitting separate univariate models. For this purpose, a joint model for a proportional hazard survival response and a Poisson count responses is presented. The methodological aspects of the proposed model is outlined including parameter estimation. The proposed model was examined on a large scale simulation study and superior performance was demonstrated over the separate univariate models. The model was further showcased on an actual clinical trial data and found out to be capable of capturing the correlation between the survival and count variables found in the data.

***Keywords:*** *bivariate response, joint model, event counts, survival data*

**Cite This Article:** Abeysinghe A Sunethra, and Marina R Sooriyarachchi, "A Joint Model for Exponential Survival Data and Poisson Count Data." *American Journal of Applied Mathematics and Statistics*, vol. 6, no. 2 (2018): 72-79. doi: 10.12691/ajams-6-2-6.

## 1. Introduction

Time-to-event data resembles data scenarios where the duration to the occurrence of a particular event is of interest. This event of interest may be death/survival in medical data, failures/breakdowns in reliability data, marriage/divorce in social science data etc. The distinguishing feature among all such data scenarios is that the time taken for an event of interest to occur is recorded. In medical data, data on patients may be recorded from the onset of the disease until the death/survival of the patient. During this time period, patients usually experience recurrent episodes of the disease such as recurrences of tumors among cancer patients, recurrences of seizures among epileptic patients, recurrences of Myocardial infarctions among cardiovascular patients and etc. It is of paramount importance that the treatments/medication given to such patients should have an impact both on their recurrent episodes of the disease and also on the time to survival/death from the disease. Therefore, in evaluating the treatment procedures undergone by the patients, both the impact of the treatments on the number of recurrent episodes of the disease and the impact of the treatment on survival time are being used in analysis. Such analysis would reveal whether the treatment had reduced the number of disease recurrence and/or increased the survival time. Statistical models deployed for such analysis would either treat time to survival as the response variable leading to survival regression models or use of the number of recurrences as the response which leads to the use of Poisson regression models.

Use of separate models for the two responses of 'survival time' and 'number of recurrences' resembles fitting separate univariate models for the two responses.

However, it is quite possible that these two responses are correlated with each other. That is, the recurrences of the disease has some bearing on the survival time of the patient whether the patients with higher number of recurrence episodes would have a shorter survival time or vice versa. When two or more response variables are found to be correlated with each other, fitting a multivariate model is more preferable than fitting separate univariate models to each response variable. Therefore, rather than fitting a separate Poisson regression model to the number of recurrent events and a survival regression model to the time to survival, it is desirable to fit a bivariate model or a joint model having two response variables of 'survival time' and 'count' because such a model capture the dependency between the responses and hence an improved analyses can be performed. This background manifested the objective of this research to propose a methodology for bivariate/joint modeling of 'survival' and 'count' responses.

## 2. Literature

The intended data scenario of this research consists of a time-to-event/survival variable and a count variable which are correlated. The main particularity here is that both of these variables behave as response variables having common/different explanatory variables associated with each of the two responses. For example, in clinical trials in evaluating the treatment procedures given to patients, both

the number of recurrences of the disease and time to survival are regarded as responses and the effect of the treatment and/or other predictors are tested against each response variables. Therefore, the literature review was focused on researches conducted on data with survival and count data. As per [15], the old practice was to take the survival variable as the response variable while the count variable was introduced to the survival model as a fixed effect covariate. But, treating the count variable as a covariate ignores the randomness in the count variable within individuals [5]. A count variable such as the number of recurrent events (seizures/tumors) experienced by patients would better be treated as a random component of a patients and hence treating it as a fixed covariate is not so desirable. In line with this, [4] suggested to take the count as a covariate measured with some error which can be regarded as an improved approach than treating it as a fixed effect covariate. Surpassing these ideas, [5] indicated through an extensive simulation study performed over correlated survival and count data that joint models provide more precise estimates than standard survival models which take count as a covariate and is obviously better than fitting two separate univariate models as well. Then, the literature review was narrowed down on to instances where joint modelling of survival and count data was performed. Immense literature is available on joint modeling of two or more survival processes such as use of multivariate frailty models [7]. Reference [10] has deployed a frailty proportional hazard model for jointly modeling survival and hospitalization data where data on hospitalizations and survival of the patients was considered. The timing of hospitalizations was also taken as a time-to-event process and time to death was taken as another survival variable. Reference [2] have developed such a joint frailty model for modeling a recurrence process and a survival process and have demonstrated its application for analyzing data on some colon cancer patients. In contrast, this study considered the cumulative count of the recurrent events and overall time to survival/death. Further, the joint model presented in this study can be generalized to any count response not necessarily a recurrent event count, which is perceived to be correlated with the patients' overall survival time such as number of affected organs, number of malignant cells. Reference [6] used a Poisson process approach for jointly modeling a survival and a count variable assuming Poisson distribution for the count variable and an Exponential distribution for the survival variable. Using the connection between Poisson and Exponential distribution and using the memory-less property of Poisson processes, they assumed that the rate of the Exponential distribution assumed for the survival response is similar to the rate of the Poisson distribution assumed for the count variable. Then, assuming the treatments given to patients had affected multiplicatively on the survival, the rate parameter of the Exponential distribution is multiplied by a treatment factor. Their model is adjusted for possible heterogeneity associated with the count variable and also for heterogeneity that may be associated with the survival variables [6]. Some of the shortcoming of this approach highlighted by them were on assuming the rate of the survival variable to be same as the rate of the event count and on assuming the treatment to be effected multiplicatively on the time to survival. The main reason behind these assumptions is to use a Poisson process to resemble the joint model. But, when considering a general situation such as when the count variable resembles some recurrent event of disease progression and when the survival variable is the time to death, this assumption of having the same rate in the Poisson distribution assumed for the count variable with the rate of the Exponential distribution assumed for the time to death is rather unrealistic. In contrast, the proposed methodology of this research is not succumbed to such an assumption and the method proposed here is capable of fitting a Poisson model/negative Binomial model for the count variable selecting the best model for the count variable while the survival variable is free to assume any parametric/non parametric distribution without being constrained on the choice of the distribution made for the count variable.

Another hot topic in joint modelling is on joint modelling of survival and longitudinal data where repeated measurements of patients are reported and the association between these repeated longitudinal measurements and the survival is concerned. However, in some situations a batch of repeated measurements over a particular variable might not be available but a single outcome [9]. This research also focuses on such a situation where joint modelling is required between a survival response and a single Poisson response variable. Generally, joint models of survival and longitudinal data have been developed by specifying two sub models for survival and longitudinal data and then combing the estimates of the sub models to estimate the joint models [12]. These joint models use random effects to resemble the correlation between observations of the same patient while estimates from sub models are combined to represent the dependency between the two responses of longitudinal and survival. In contrast, this research explores the possibility of using random effects to denote the association/dependency between two response variables.

# 3. Methodology

The theoretical aspects of the proposed joint model is outlined in this section. The joint likelihood of the model is derived and parameter estimation is discussed

## 3.1. Joint Model

The development of the joint model for a Poisson distributed Count response and an Exponential survival responses is presented here where it is not assumed that the rate of the Poisson and the mean of the Exponential distribution are identical as in the method of proposed by [6].

The correlation between the survival and count responses is represented by patient specific random effects which are assumed to normally distributed. It is assumed that the censoring mechanism of the survival variable is non-informative. The notations of the model are as follows.

Let $\boldsymbol{Y_i}$ and $\boldsymbol{T_i}$ denotes respectively the count response and the survival response of the $\boldsymbol{i^{th}}$ patient for =

$\mathbf{1}, \mathbf{2}, \ldots\ldots, \boldsymbol{n}$ . The censoring indicator of the $\boldsymbol{i^{th}}$ patient is denoted δi where δi =1 denotes actual survival times and δi =0 denotes censored survival times. Let $\boldsymbol{G_i}$ denotes the covariate vector of the Poisson model assumed for the count responses with a corresponding parameter vector denoted by β. Similarly, $\boldsymbol{W_i}$ denotes the q×1 vector of covariates of the survival model and γ denotes the corresponding q×1 parameter vector.The vector of random effects assumed at patient level for joining the two responses is denoted by $\boldsymbol{u_i}$ where it was assumed $\boldsymbol{u_i \sim Normal(0, \sigma_u^2)}$.

After incorporating a shared random effect i.e $\boldsymbol{u_i}$ to both the survival and count model, the Poisson models can be equated as follows.

$$\boldsymbol{f\left(Y_i = y_i \mid u_i\right) = \frac{exp(G_i{}^T \beta + u_i)^{y_i} e^{-exp(G_i{}^T \beta + u_i)}}{y_i!}}. \quad (1)$$

Although the survival distribution can assume any suitable parametric distribution or a Cox proportional hazards model, the theory is explained here is in terms of an Exponential distribution assumed for the survival variable mainly due to the simplicity of explanation and hence the simulation study was also performed on an Exponential distribution for survival data. The density function of Exponential survival model which correspond to the probability of actual survival times (i.e uncensored) is as follows:

$$\boldsymbol{f\left(T_i = t_i \mid u_i\right) = e^{(W_i{}^T \gamma + g^* u_i)} e^{-exp(W_i{}^T \gamma + g^* u_i) t_i}}. \quad (2)$$

The probability of the censored survival times can be calculated from the corresponding survival function of the Exponential distribution:

$$\boldsymbol{S\left(T_i \mid u_i\right) = e^{-exp(W_i{}^T \gamma + g^* u_i) t_i}}. \quad (3)$$

Assuming these marginal models to be independent conditioned on the random effects (i.e $\boldsymbol{u_i}$), the joint density can be specified as:

$$\begin{aligned} &\boldsymbol{f(T_i, Y_i) = \int f(T_i, Y_i \mid u_i) f(u_i) du_i} \\ &\boldsymbol{= \int f(T_i \mid u_i)]^{\delta_i} [S\left(T_i \mid u_i\right)]^{1-\delta_i} f\left(Y_i \mid u_i\right) f(u_i) du_i}. \end{aligned} \quad (4)$$

## 3.2. Likelihood Function

The joint likelihood function of the proposed model can be written as follows.

$$\begin{aligned} &L(\psi; T, Y) = \prod_{i=1}^{n} f(T_i, Y_i \mid \psi) \\ &= \prod_{i=1}^{n} \int f\left(T_i \mid u_i, \psi\right) f\left(Y_i \mid u_i, \psi\right) f\left(u_i\right) du_i \\ &= \prod_{i=1}^{n} \int [e^{(W_i{}^T \gamma + g^* u_i)} e^{-exp\left(W_i{}^T \gamma + g^* u_i\right) t_i}]^{\delta_i} \\ &\times [e^{-\left(W_i{}^T \gamma + g^* u_i\right)}]^{1-\delta_i} [(e^{(G_i{}^T \beta + u_i) y_i}) e^{-exp\left(G_i{}^T \beta + u_i\right)} / y_i!] \\ &\times [1/\sqrt{2\pi\sigma^2}\, e^{-1/2\left(\frac{u_i}{\sigma}\right)^2}] du_i. \end{aligned} \quad (5)$$

A key contribution of the proposed methodology emerged due to the challenge of maximizing the above complex likelihood function which consists of a survival likelihood component and a Poisson likelihood component. In existing standard statistical software, maximization of the survival likelihood and Poisson likelihood cannot be performed within the same routine or package i.e a software for fitting Poisson regression models cannot fit simultaneously a survival regression model and software for survival data cannot fit Poisson regression models. Therefore, the existing practice is to use separate routines for the survival sub model and Poisson sub model and finally combine the results of the two sub models to estimate the joint model which requires additional programming than existing software routines. Avoiding this complexity of using separate software routines for survival and count responses, an indirect method of estimation is used in this research. This indirect approach facilitates estimation of the above likelihood jointly using single software routine particularly software for Generalized Linear Mixed Models (GLMMs).

## 3.3. Parameter Estimation

A greater flexibility in using shared parameter approach for joint modeling is the capability of applying any methodology suited for the respective univariate marginal models which form up the joint model because conditional density of each response over the random effects can take the respective marginal density of each response variable. Therefore, this research deployed an indirect approach for estimating the survival likelihood using any software on GLMMs which was feasible simply due to the equivalence of the log-likelihoods of a survival model and a Poisson model under the assumption of proportional hazard in survival data [17]. Survival data distributed as an Exponential distribution holds the proportional hazard assumption and hence estimating the survival model can be done using Poisson/GLM software. The equivalence of the likelihoods of survival and Poisson data is explained briefly below.

The Log-likelihood of proportional hazard survival data can be derived as (Da Silva & de Lima, 2003):

$$(\beta|b) = \sum_i \delta_i \log(\mu_i) - \mu_i + \delta_i \log\left(\frac{h_o(t_i)}{H_0(t_i)}\right) \quad (6)$$

Where $\mu_i$ denotes the mean survival time, $h_o(t_i)$ denotes the baseline hazard function and $H_o(t_i)$ is the cumulative baseline hazard.

Now consider the log likelihood of 'n' independent Poisson random variables $w_i \sim Poisson(\mu_i^*)$ which reduces to [8]:

$$l = \sum_i w_i \log\left(\mu_i^*\right) - \mu_i^* - k. \quad (7)$$

It is noteworthy that these two likelihoods are identical with respect to maximization except for the last term when $\delta_i$ (Censoring indicator) is specified as $\boldsymbol{\delta_i \sim Poisson(\mu_i^*)}$ But, in semi-parametric survival models (Cox-model) and with Exponentially distributed survival data, the last term on the right hand side of (5) does not involve any unknown parameters and hence does not influence the

maximum likelihood estimation [1]. This approach is being extensively used in the literature for estimating fixed effect survival models. But, [8] showed theoretically and with an application that this approach of estimating a survival model using a Poisson model can be used for estimating random effect survival models as well which is the case in this research. In line with this, to estimate the likelihood contribution to the joint likelihood from the conditional survival density $\boldsymbol{f(T_i|U_i)}$ a pseudo conditional Poisson density will be used. Therefore, the joint likelihood of the proposed model can now be estimated as follows:

$$
\begin{aligned}
L(\psi;T,Y) &= \prod_{i=1}^{n} f(T_i,Y_i|\psi) \\
&= \prod_{i=1}^{n}\int f(T_i|u_i,\psi)\, f(Y_i|u_i,\psi)\, f(u_i)\, du_i \\
&\prod_{i=1}^{n}\int \left[ \frac{exp\left(W_i^T\gamma + g^*U_i\right)^{\delta_i} e^{-\exp\left((W_i^T\gamma + g^*U_i\right)}}{\delta_i!} \right] \\
&\times \left[ \frac{exp\left(G_i^T\beta + U_i\right)^{y_i} e^{-\exp\left(G_i^T\beta + U_i\right)}}{y_i!} \right] \\
&\times \left[ 1/\sqrt{2\pi\sigma^2}\, e^{-1/2\left(\frac{u_i}{\sigma}\right)^2} \right] du_i .
\end{aligned}
\tag{8}
$$

Maximizing the above likelihood function is tedious mainly due to the integration over the random effect $\boldsymbol{u_i}$. Therefore, Adaptive Gaussian Quadrature method was used for integral approximation (Pinheiro and Bates, 1995). The SAS procedure Proc NLMIXED was used with careful specifications to tally with the proposed joint model and with proposed estimation method. It was observed that the using 'trust region method' of estimation improved the functionality of the Joint model in Proc NLMIXED.

# 4. Simulation Study

An in-depth simulation study was conducted to evaluate the finite sample properties of the proposed method and also to compare the performance of the proposed joint model with separate univariate models. The sensitivity/adaptability of the model to the level of dependency between the two responses was the main consideration here for which the variance of the random effects ($\boldsymbol{\sigma_u^2}$) and the parameter 'g' were used to represent the dependency between responses were varied. Treatment parameters ($\boldsymbol{\beta_2, \gamma_1}$)in count and survival model respectively were also varied in order to evaluate the effect of the fixed effect parameters on the joint model functionality.

Count data with patient specific random effects can be simulated using a Poisson GLMM. For simplicity only one covariate named 'treatment' ($\boldsymbol{Z_i}$) was simulated from a Bernoulli distribution with a success probability of .5. Patient specific random effect was simulated from a Normal distribution with $\boldsymbol{U \sim N(0, \sigma^2)}$. Therefore, the model used to simulate count variable is of the form:

$$
f(Y_i = y_i|U_i)\frac{(\beta_1 + \beta_2 Z_i + u_i)^{y_i} e^{-(\beta_1+\beta_2 Z_i + u_i)}}{y_i!}. \tag{9}
$$

In simulating survival data, in addition to the random effect $\boldsymbol{U_i}$ and treatment covariate $\boldsymbol{Z_i}$, a parameter to denote the strength of the association between survival and count variable was also incorporated which is denoted by 'g'. To simulate Exponential survival data, Cox frailty model of the following form was used:

$$
\lambda(t|U_i) = \lambda_0(t)\exp(\gamma_1 Z_i + g^* u_i). \tag{10}
$$

The baseline hazard function ($\boldsymbol{\lambda_0(t)}$) was set to 0.15 and treatment covariate ($\boldsymbol{Z_i}$) and random effect ($\boldsymbol{u_i}$) was obtained from those simulated for the count model. Due to the relationship between cumulative density function and cumulative hazard function in proportional hazard survival data, following formula can be used to simulate Exponentially distributed survival times (Bender et al, 2005).

$$
T_i = -\log(Uni_i)/.15\exp(\gamma_1 Z_i + g^* u_i) \tag{11}
$$

Where $\boldsymbol{Uni_i \sim uniform(0,1)}$.

Different specifications for g and Var($u_i$) were used to examine the performance of the joint model while regression coefficients $\boldsymbol{\beta_2}$ **and** $\boldsymbol{\gamma_1}$ were also varied as $\boldsymbol{\beta_2 = 0, 0.5}$ $\boldsymbol{\gamma_1 = 0, 0.5}$. The parameter 'g' which represents the dependency between the two responses were varied as g=-0.8,0, +.8 while variance of the random effect was varied as $\boldsymbol{\sigma_u^2 =, 0.5^2, 1^2, 1.5^2}$ imposing various levels of dependency between the two response. Each parameter combination replicated with 1000 simulations.

## 4.1. Simulation Results

The estimation of the proposed joint model was performed via Proc NLMIXED by specifying separate conditional likelihood functions for the two response of survival and count. The indirect approach outlined in the methodology which was previously shown to be theoretically and practically viable by [8]was used to estimate the survival likelihood within Proc NLMIXED. To perform the joint model, it is required to convert the data sets into univariate form as outlined in the Appendix. Then, the joint model can be specified via Proc NLMIXED by specifying separate density/likelihood for the two response variables. The separate univariate models i.e a Poisson regression model for the count variable and an Exponential survival regression model for the survival response were also fitted in SAS software.

Several statistics were extracted over the joint model and univariate models fitted for the data simulated namely: Type I Error, Power, Bias, Empirical Standard errors of the parameter estimates (ESE), coverage probability of 95% confidence interval that cover the true parameter value (CP). Type I error of the F-test for significant treatment effect is considered more important here because giving a false conclusion of significant treatment

effect on survival and/or number of recurrences of the disease when there is actually no treatment effect is more serious than power of the test which concludes that there is a significant treatment effect when actually a treatment effect exists.

Table 1 gives the results of the cases where the true values of the treatment parameters are set to '0' i.e $\boldsymbol{\beta_2 = 0}, \boldsymbol{\gamma_1 = 0}$. Here, the rejection of the null hypothesis of the type III F-test gives rise to the Type I Error. In addition to Type I error, properties like Bias, Coverage Probability, Empirical standard errors of the parameter estimates were also used for evaluating and comparing the performance of the joint model.

**Table 1. Simulation Results under '0' treatment effect**

| | | Type I Error | Bias | ESE | CP |
|---|---|---|---|---|---|
| $\sigma_u^2 = .5$, g= 0.8 | | | | | |
| $\gamma_1$ | Univariate | .063, | -.007 | .154 | 0.937 |
| | Joint | .048 | 0.003 | 0.169 | 0.951 |
| $\beta_2$ | univariate | 0.044 | .0017 | .169 | 0.955 |
| | Joint | .024 | -.005 | 0.184 | 0.975 |
| $\sigma_u^2 = .5$, g=- 0.8 | | | | | |
| $\gamma_1$ | Univariate | 0.052 | 0.008 | 0.155 | 0.947 |
| | Joint | 0.045 | -.001 | 0.168 | 0.954 |
| $\beta_2$ | univariate | 0.064 | -.0004 | .169 | 0.936 |
| | Joint | 0.041 | .0003 | 0.183 | 0.958 |
| $\sigma_u^2 = 1$, g= 0.8 | | | | | |
| $\gamma_1$ | Univariate | .083 | .008 | .157 | .92 |
| | Joint | .042 | .008 | .198 | .95 |
| $\beta_2$ | univariate | .187 | .004 | 141 | .80 |
| | Joint | .034 | .002 | .204 | .95 |
| $\sigma_u^2 = 1$, g= -0.8 | | | | | |
| $\gamma_1$ | Univariate | .052 | .008 | .155 | 0.947 |
| | Joint | .034 | -.001 | .198 | 0.966 |
| $\beta_2$ | univariate | 0.162 | -.005 | 141 | .819 |
| | Joint | 0.054 | .003 | .204 | .945 |
| $\sigma_u^2 = 1.5$, g= 0.8 | | | | | |
| $\gamma_1$ | Univariate | 0.089 | 0.0012 | 0.159 | 0.911 |
| | Joint | 0.043 | -0.004 | 0.227 | 0.956 |
| $\beta_2$ | univariate | 0.215 | -.0177 | 0.198 | 0.785 |
| | Joint | 0.035 | -0.0007 | 0.213 | 0.964 |
| $\sigma_u^2 = 1.5$, g= -0.8 | | | | | |
| $\gamma_1$ | Univariate | 0.094 | 0.0015 | 0.159 | 0.906 |
| | Joint | 0.056 | -0.009 | 0.229 | 0.941 |
| $\beta_2$ | univariate | 0.215 | -.0177 | 0.198 | 0.785 |
| | Joint | 0.04 | 0.006 | 0.213 | 0.955 |

The first two cases of Table 1 represented a lower level of dependency between the two responses ($\boldsymbol{\sigma_u^2 = .5}$) with positive (g=.8) and negative (g=-.8) directions. Here, the Type I error rate of both the joint model and the Separate Models were within [0.036, 0.064] which is the 95% confidence Interval for 5% error rate in 1000 simulations except for the Type I Error resulted for the treatment parameter for the count response in the joint model with positive dependency (case I). When the strength of the dependency between the responses was higher ($\boldsymbol{\sigma_u^2 = 1, 1.5}$), both for the positive and negative dependency, the type I Errors of the separate models were Inflated whereas the Type I Errors of the Joint models were all well within the 95% confidence interval. Thus, the joint model showcased superior performance than separate univariate models with respect to type I error. With respect to Bias of the parameter estimates, both modelling approaches resulted in small Bias out of which the lower was incurred in the Joint model. Marginally high values were observed for the empirical standard errors of joint models' parameter estimates in all the cases which were perceived to be due to the non-convergence of the model in Proc NLMIXED for which improving/changing the parameter estimation procedure would be desirable such as the use of EM algorithm for model estimation or use of Pseudo-Adaptive Gaussian Quadrature Estimation for estimating the Joint model [12,13]. It is noteworthy that the coverage probability of the joint model was higher in all the cases than that with the case of separate univariate models i.e the joint model was capable of resulting in a confidence interval which contained the true parameter most of the time than that with separate univariate models.

Table 2 gives the results of the simulations done for the case with true values of the treatment parameter was not equal to '0' and hence the rejection of the null hypothesis of the Type II F-test gives rise to the Power of the test.

**Table 2. Simulation Results under non-zero treatment effect**

| | | Power | Bias | ESE | CP |
|---|---|---|---|---|---|
| $\sigma_u^2 = .5$, g= 0.8 | | | | | |
| $\gamma_1$ | Univariate | .972 | .014 | .169 | .94 |
| | Joint | .972 | -.004 | .184 | .95 |
| $\beta_2$ | univariate | .925 | -.001 | .208 | .95 |
| | Joint | ..902 | .001 | .22 | .963 |
| $\sigma_u^2 = .5$, g=- 0.8 | | | | | |
| $\gamma_1$ | Univariate | .974 | .016 | .169 | .94 |
| | Joint | ..972 | -.002 | .181 | .96 |
| $\beta_2$ | univariate | .914 | .003 | .209 | .94 |
| | Joint | .9 | -.004 | .22 | .96 |
| $\sigma_u^2 = 1$, g= 0.8 | | | | | |
| $\gamma_1$ | Univariate | .926 | .055 | .169 | .89 |
| | Joint | .915 | -.0001 | .21 | .95 |
| $\beta_2$ | univariate | .933 | .001 | .172 | .86 |
| | Joint | .855 | .002 | .228 | .96 |
| $\sigma_u^2 = 1$, g= -0.8 | | | | | |
| $\gamma_1$ | Univariate | .942 | .047 | .169 | .91 |
| | Joint | .933 | -.008 | .209 | .96 |
| $\beta_2$ | univariate | .93 | .006 | .173 | .84 |
| | Joint | .85 | .004 | .230 | .95 |
| $\sigma_u^2 = 1.5$, g= 0.8 | | | | | |
| $\gamma_1$ | Univariate | .84 | .029 | .17 | .92 |
| | Joint | .84 | -.004 | .238 | .95 |
| $\beta_2$ | univariate | .823 | .005 | .206 | .786 |
| | Joint | .863 | .002 | .221 | .94 |
| $\sigma_u^2 = 1.5$, g= -0.8 | | | | | |
| $\gamma_1$ | Univariate | .85 | .09 | .17 | .81 |
| | Joint | .69 | -.004 | .241 | .95 |
| $\beta_2$ | univariate | .812 | -.001 | .206 | .80 |
| | Joint | .68 | -.001 | .220 | .95 |

As per the results given in Table 2, Power was not always higher in the joint model in all the cases. It should be noted that the assumed value for the parameter i.e .5 in the cases simulated above also impact on the rejection of the null hypothesis of the type III F-test. Hence, changing

the value of the $\gamma_1$ and/or $\beta_2$ may improve the Power of the test. But, the coverage probability is not sensitive to the assumed value of the true parameter because it gives the percentage of confidence interval which contained the true parameter. So, when considered the coverage probability, it can be seen that the Joint model outperformed the separate univariate models in all the cases in large amounts. Moreover, the bias of the parameter estimate were also reduced in the joint model especially when the dependency between the two responses were higher ($\boldsymbol{\sigma_{u=}^2 1^2, 1.5^2}$).

In summary, the power is comparable for the joint and univariate models for smaller $\boldsymbol{\sigma_u^2}$ however, when $\boldsymbol{\sigma_u^2}$ becomes larger the power of the joint model is lower especially for the case when g=-0.8. As under the null hypothesis here too the bias is lower and coverage probability is higher for the joint model when compared to the univarate models. Also, the standard errors of the two parameters are slightly higher for the joint model. Summing up the results observed by the simulation study, the joint model proposed in this research is better than fitting separate univariate models when two responses of survival and count are correlated with each other particularly with respect to Type I error, bias and coverage probability and.

# 5. Example

To examine the performance of the proposed model on actual data, an individual patient data from a clinical trial on Epilepsy was used. Owing to reasons of confidentiality, source of the data and actual names of some of the variables are not included here. The data on patient's age, sex, type of seizure, treatment/drug were the predictor variables of the models fitted. The outcome variables were the number of seizures prior to randomization to a treatment and time to first seizure since randomization which will be the count and survival Reponses respectively. Time to first post-randomisation seizure, is an internationally agreed outcome (ILAE Commission on Antiepileptic Drugs, 1998) in evaluating the treatment given to Epileptic patients. The final dataset consisted of 924 patients after data pre-processing was carried out. Table 3 gives basic information regarding the variables and it can be seen that the data is balanced fairly over the predictor variables.

**Table 3. Details of the Variables in the Example Data**

| Variable | | |
|---|---|---|
| Type of Epilepsy | Type I | 538(58.2%) |
| | Type II | 372 (40.3%) |
| Drug | Drug A | 482 (52.2%) |
| | Drug B | 442 (47.8%) |
| Sex | Male | 538 (58.2%) |
| | Female | 386 (41.8%) |

## 5.1. Fitting Univariate Models

The results of the two univariate models fitted for the two response variables are presented here. A Poisson regression model was fitted for the seizure count and an Exponential survival model was fitted for the survival model which considered the time to first seizure. The results of the selected models for each response are given in Table 4.

**Table 4. Results of the Univariate Models Fitted to Example Data**

| Response | Model Variables | Parameter Estimates | Std. Error | P-value |
|---|---|---|---|---|
| Time to first Seizure | | | | |
| | Intercept | 7.75 | .01 | <.0001 |
| | Drug | -0.26 | .16 | .11 |
| | Type | -0.02 | .139 | .87 |
| | Type*Drug | 0.53 | .21 | 0.0199 |
| No. of Seizures | | | | |
| | Intercept | 1.01 | .04 | <.0001 |
| | Sex | .35 | .04 | <.0001 |
| | Type | -.53 | .06 | <.0001 |

When considered the survival model, only the interaction term of 'drug' and 'type' were significant. Hence, 'drug' and 'type' were also kept in the model as main effects though they were non-significant at the 5% level where a marginal significance was found for the variable 'drug' at 10% level of significance. In the Poisson regression model fitted for the number of seizures, the sex and Type of Epilepsy were significant.

## 5.2. Fitting the Joint Model to Example Data

Then, a joint model as proposed by this research was fitted for the joint response of 'No.of seizures' and 'time to first seizure' of which the chosen model is given in Table 5.

**Table 5. Results of the Joint Model**

| Response | Model Variables | Parameter Estimates | Std. Error | P-value |
|---|---|---|---|---|
| Time to first Seizure | | | | |
| | Intercept | 7.94 | .11 | <.0001 |
| | Drug | -.62 | .17 | .0004 |
| | Type | -.15 | .15 | .3799 |
| | Type*Drug | .85 | .23 | .0002 |
| No. of Seizures | | | | |
| | Intercept | .61 | .06 | <.0001 |
| | Sex | .23 | .07 | .0004 |
| | Type | -.44 | .07 | <.0001 |
| Covariance parameters | $\sigma_u^2$ | 0.4407 | 0.03 | <.0001 |

Among the parameter estimates of the joint model, the estimates obtained for random effect used to represent the dependency between the two responses, is of greater importance and it can be seen that the variance of the random effect has resulted in a high significance. Thus, it can be confirmed that the two responses are correlated and hence a joint model is desirable.

With respect to fixed effects of the survival model, the variable 'drug' has resulted to be a significant variable to the time to first seizure which was not significant in the

univariate model. That is when adjusted for the random effect with patient, the effect of the drug is significant on timing of the first seizure. Further, the combined effect of drug and type of epilepsy has a significant effect on time to first seizure of the Epilepsy patients. When considering the joint model estimates for the count model, the variables 'drug' and interaction of 'drug' and 'type', which were significant in the univariate count model, was no longer significant in the joint model. This indicted that the sex and the type of epilepsy could significantly explain the variation in the number of seizures among the patients when the random effect of patient is incorporated into the model.

## 6. Discussion

The main objective of this paper was to propose a bivariate model for a joint response of 'survival' and 'count' variables. The proposed model was developed deploying random effects to obtain the joint distribution of the two responses, which is a popular approach in deriving joint distributions of correlated variables [12]. This approach is being widely used to develop joint models for longitudinal and survival data. A key feature in such models is that a bunch of repeated measurement are available for each patients. However, the literature has highlighted the need of joint models for a single repeated measurement and survival data [9]. A handful of researches were found where the focus on joint modeling of survival and count data [5,6,14]. A major limitation in the existing models was that survival model can only be fitted with a Cox proportional hazard model and/or strong assumptions were imposed for the model parameters of the marginal models used for survival and count such as rate parameter of the Poisson models used for the count variable is similar to the rate parameter of the Exponential model used for the survival model [6]. The speciality of the proposed joint model is that it can assume any suitable distribution for the count variable and any suitable survival distribution or a Cox model for the survival variable. Another distinguishing feature of the proposed model is that usually when fitting joint models, specialized software is needed (e.g- JM package in R). However, we proposed an indirect approach as outlined in the methodology section to estimate the joint model within any general statistical software with a routine for generalized linear mixed models whereas the SAS procedure Proc NLMIXED was used here.

The simulation study revealed that the joint model is superior to the univariate marginal models in the sense that its parameters had lower bias and higher coverage probability when compared to the univariate models. The type one error is often within probability limits for the Joint model while some cases the type one error is inflated for the univariate models especially when the variables were highly correlated. What is gained by the type one error in the joint model is that rejection of the null hypothesis was within stipulated limits and lost by the univariate models with a high proportion of rejections of the null hypothesis is illustrated under the alternative hypothesis too and this is the reason for the higher power in the univariate models. It can be argued that it is better to have a type one error within limits and have a somewhat lower power justifying the use of the joint model in place of the univariate models. When it comes to the standard errors of the parameters these are almost comparable between the two types of models.

The proposed model was applied to an actual dataset which had a count and a survival outcome correlated to each other. Since the main purpose for the use of actual data to evaluate the performance to the proposed joint mode, an in-depth analysis over the example data was not included and due to the confidentiality of the data some limitations were imposed into use of actual names of the variables and etc. However, the joint model fitted to the example data showed comparable results with respective univariate models, significant estimates for the random effects confirmed that the model has captured the correlation between the two outcomes, showed different combinations of significant fixed effects for each response separately. Both the simulated data and actual data fitted only a Poisson distribution for the count variable and an Exponential distribution to the survival variable. Therefore, extending the situation study over other candidate distributions for count and survival variables is desirable. Further, validation of the proposed model is appealing which was unable to be done due to the non-availability and limitations on the use of actual data.

## Acknowledgments

The authors acknowledge the postgraduate research scholarship offered by the University of Colombo, Sri Lanka for this study.

## References

[1] Aitkin, M., & Clayton, D. (1980). The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM. Applied Statistics, 156-163.

[2] Belot, A., Rondeau, V., & Remontet, L. (2014). A joint frailty model to estimate the recurrence process and the disease-specific mortality process without needing the cause of death. Statistics in Medicine.

[3] Bender, R., Augustin, T., & Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models. Statistics in medicine, 24(11), 1713-1723.

[4] Carroll, R., Ruppert, D., Stefanski, L., & Crainiceanu, C. (2006). Measurement error in nonlinear models: a modern perspective. CRC press.

[5] Cowling, B. J. (2003). Survival models for censored point processes. Doctoral dissertation, University of Warwick.

[6] Cowling, B. J., Hutton, J. L., & Shaw, J. E. (2006). Joint modelling of event counts and survival times. Journal of the Royal Statistical Society: Series C (Applied Statistics).

[7] Crowder, M. J. (2012). Multivariate Survival Analysis and Competing Risks. CRC Press.

[8] Da Silva, G. T., & De Lima, A. C. (2003). Using SAS software for Multilevel Models in Survival Analysis. *PharmaSUG SAS Users Group Conference*. Miami, Florida.

[9] Lai, TL, Lavori, PW, Shih, MC (2012). Sequential design of phase II-III cancer trials. Stat Med, 31, 18:1944-60.

[10] Liu, L., Wolfe, R. A., & Huang, X. (2004). Shared frailty models for recurrent events and a terminal event. Biometrics, 60(3), 747-756.

[11] Pinheiro, J.C. and Bates, D.M. (1995), "Approximations to the Log-likelihood Function in the Nonlinear Mixed-effects Model," Journal of Computational and Graphical Statistics, 4, 12 -35.

[12] Rizopoulos, D. (2012). Joint Models for Longitudinal and Time-to - Event data with Applications in R . CRC Press.

[13] Rizopoulos, D. (2012). Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. Computational Statistics & Data Analysis, 56(3), 491-501.

[14] Rogers, J. K., & Hutton, J. L. (2013). Joint modelling of pre-randomisation event counts and multiple post-randomisation survival times with cure rates: application to data for early epilepsy and single seizures. Journal of Applied Statistics, 40(3), 546-562.

[15] Verity, C. M., Hosking, G., & Easter, D. J. (1995). A multicentrecomparative trial of sodium valproate and carbamazepine in paediatric epilepsy.Developmental Medicine & Child Neurology, 37(2), 97-108.

[16] Vonesh, E. F., Greene, T., & Schluchter, M. D. (2006). Shared parameter models for the joint analysis of longitudinal data and event times. Statistics in medicine, 25(1), 143-163.

[17] Whitehead, J. (1980). Fitting Cox's regression model to survival data using GLIM .Applied Statistics, 268-275.