JOINT MODELING OF CENSORED LONGITUDINAL AND EVENT TIME DATA

by

Francis Pike

BSc Applied Mathematics, Southern Connecticut State University, 1997

MSc Biostatistics, Columbia University, 2006

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Francis Pike

It was defended on

June 21 2011

and approved by

Joyce Chang, PhD, Associate Professor of Medicine
Biostatistics, and Clinical and Translational Science
Department of Medicine
School of Medicine
University of Pittsburgh

Lan Kong, Ph.D., Assistant Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Mark Unruh, M.D., Associate Professor of Medicine
Renal Electrolyte Division
Department of Medicine
School of Medicine
University of Pittsburgh

Dissertation Director: Lisa Weissfeld, PhD, Professor and Associate Chair
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Lisa Weissfeld, PhD

JOINT MODELING OF CENSORED LONGITUDINAL AND EVENT TIME DATA

Francis Pike, PhD

University of Pittsburgh, 2011

Longitudinal censoring is a common artifact when evaluating biomarkers and an obstacle to overcome when jointly investigating the longitudinal nature of the data and the impact on the survival prognoses of a study population. To fully appreciate the complexity of this scenario one has to devise a modeling strategy that can simultaneously account for (i) longitudinal censoring, (ii) outcome dependent dropout, and potentially (iii) correlated biomarkers. In this thesis we propose a novel joint modeling approach to account for the aforementioned issues by linking together a univariate or multivariate Tobit mixed effects model to a suitable parametric event time distribution. This method is significant to public health research since it enables researchers to evaluate the evolution of the disease process in the presence of complex biomarker data where there may be censoring, correlation, and outcome dependent dropout. This approach allows for the analysis of data in a single unified framework. The performance of the proposed Joint Tobit model will be compared to the commonly used "fill-in" methods for censored longitudinal data in a joint modeling framework. Furthermore, we will show that the implementation of our proposed model is fairly straightforward in commercially available software, thus avoiding the complexity and problem specific nature of the expectation maximization (EM) algorithm.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

The authors would like to thank Dr Derek Angus and the CRISMA laboratory for access to the GenIMS data. We would also like to thank all of the health-care professionals, as well as the patients and their families who supported this trial. A complete list of GenIMS investigators is available at *www.ccm.upmc.edu/genims investigators*. The GenIMS study was funded via grant R01 GM61992 by the National Institute of General Medical Sciences. This manuscript was not prepared in collaboration with the investigators of the GENIMS Study and does not necessarily reflect the opinions or views of the GENIMS Study.

The authors would also like to thank Dr Mark Unruh, for access to the HEMO data, and the HEMO Study investigators and the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) data repository for the data used in this study. The HEMO Study was performed by the HEMO Study investigators and supported by the NIDDK. This manuscript was not prepared in collaboration with the investigators of the HEMO Study and does not necessarily reflect the opinions or views of the HEMO Study or the NIDDK.

We would also like to emphasize that the datasets and results presented in this thesis were used merely to illustrate the proposed methods and were not intended to reflect a formal analysis of the data.

1.0 INTRODUCTION

"A characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathologic processes, or pharmacological responses to a therapeutic intervention", is the current NIH (National Institutes of Health) definition for a biomarker. Biomarkers are frequently used in biomedical studies as standalone predictors or to augment current prediction models. Their precise measurement is then of utmost importance regarding their appropriateness for use as clinical predictors of disease. However due to the sensitivity of the measuring instruments it is common to measure biomarkers subject to a detection limit (DL). This infers that, at best, our precision is limited to an interval above or below such detection limits and is commonly referred to as left or right censoring. Alternatively it may be that the values we seek to use were generated by a biological model and results in impossible or implausible values. In circumstances such as the latter, censoring occurs artificially as investigators typically use their subject matter knowledge of the process to replace these values by a lower or upper limit respectively. For example, censoring, is an issue in all scientific disciplines from the assessment of HIV1 RNA levels or Viral load (Hughes et al, 1999)[1], quantitative measurements of environmental factors (Lubin et al, 2004) [2], to linkage analysis of censored trait data (Epstein et al, 2003) [3].

We encountered data produced by both censoring mechanisms from two different studies. Censoring due to measurement precision was encountered in the Genetic and inflammatory markers of sepsis (GenIMS) study (Kellum et al, 2007) [4]. This was a cohort study investigating the association between severe sepsis, brought on by community acquired pneumonia, and

consequent risk of death. The investigators collected data on plasma necrosis (TNF) levels, IL6 (interleukin6), and IL10 (interleukin10) levels daily for the first week and weekly thereafter with 90 day survival as the primary endpoint. Left and right censored measurements on IL6 and IL10 accounted for 27% and 70% of the data respectively.

Censoring due to modeling was encountered in the HEMO study [5], which was a large randomized clinical trial of dialysis dose and membrane flux carried out at 15 clinical centers between 1995 and 2001. The clearance values were derived from kinetic models and frequently produced negative values and clearances outside of normal biological limits. Oftentimes the impossible / implausible values are replaced by an arbitrary constant or even discarded resulting in biased analyses.

A typical hypothesis relevant to both scenarios is whether these censored longitudinal values are associated, either singly or jointly, with survival. Thus the analytical framework utilized must be able to account for (a) outcome dependent dropout, (b) longitudinal censoring, and potentially correlated biomarkers in one unified model. This thesis will therefore outline the use of a joint model to account for the aforementioned issues but can also be looked at as a novel way to incorporate longitudinally censored covariate(s) into a survival model.

To our knowledge no one has, as yet, accounted for a doubly censored covariate in a survival analysis by joint modeling. Also since Tobit regression models have long been used to model censored data, this thesis will discuss the current methods for survival and longitudinal data analysis and ultimately investigate the feasibility of jointly analyzing survival and doubly censored longitudinal data using a Joint Tobit model. We will also discuss current joint modeling methodology and the implementation of this model in commercially available software.

Specifically, the joint Tobit model we propose will be composed of separate sub-models. The longitudinal sub-model will consist of a univariate or multivariate linear mixed effects Tobit model and is intended to account for the longitudinal censoring. The time to event sub-model can be any parametric form but for ease of exposition will be considered to be an exponential model. We can envisage the issues in a more precise manner by looking at what the separate sub-models were designed to do. The longitudinal sub-model is used to capture all of the information in a longitudinal fashion but assumes MCAR (missing completely at random) or MAR (missing at random) but yields biased results in the presence of outcome dependent dropout. This is where the reason for missing data is due to some feature of the outcome process. The limitation of the survival model is that it is typically designed for a single and fixed covariate value. Time dependent covariates can be handled via a time dependent Cox model but this method unfortunately does not account for longitudinal censoring. Thus better methods can be developed by taking the best of what both of the sub-models have to offer. The question then becomes how we combine these disparate models. We intend to combine or "link" the two models by assuming that the hazard is a function of the longitudinal trajectory and as such share the same set of random effects. These shared random effects and their respective parameters will capture the association between the two sub-models due to the inclusion of a random intercept, slope, and fitted trajectory at the event time. An optional orthogonal frailty can be added to introduce heterogeneity in survival sub-model. The estimation will be done by maximization of the proposed joint Tobit likelihood in SAS PROC NLMIXED which is one of the strengths of this proposal in that we aim to avoid the complexities of programming a problem specific EM algorithm for parameter estimation and turn to a more flexible and commercially available alternative.

2.0 REVIEW OF CLASSICAL MODELS

2.1 SURVIVAL ANALYSIS

2.1.1 Notation

Survival data for the *i*th individual can be conveniently represented by $(T_i, \delta_i, x_i), i = 1 \cdots n$. Here δ_i is a censoring indicator (δ_i=0 if the *i*th individual is censored: δ_i=1 if the *i*th individual failed), T_i is the corresponding event or failure time, and $x_i^{'} = (x_{i1}, x_{i2,} \cdots x_{in})$ is a vector of baseline covariates associated with the *i*th individual.

2.1.2 Likelihood for Right Censored Data

Given that $x_i^{'} = (x_{i1}, x_{i2,} \cdots x_{in})$ and the pairs $(T_i, \delta_i), i = 1 \cdots n$, are independent, the likelihood of the data $(T_i, \delta_i, x_i), i = 1 \cdots n$, conditional on $x_i^{'} = (x_{i1}, x_{i2,} \cdots x_{in})$ can be expressed as:

$$L(\boldsymbol{\theta}) \propto \prod_{i=1}^{n} f(t_i, \boldsymbol{\theta}, \mathbf{x_i})^{\delta_i} S(t_i, \boldsymbol{\theta}, \mathbf{x_i})^{1-\delta_i} \quad 2.1$$

.

The contribution to the likelihood is the probability density function $f(t_i, \theta, x_i)$ for a failure and survival distribution $S(t_i, \theta, x_i)$ for censored time.

The contribution for an individual censored at t_i is $\Pr(T_i > t_i : x_i, \theta)$. It is often convenient to express the likelihood in terms of the hazard. Using basic survival quantities, the likelihood above can be expressed in terms of the hazard as:

$$L(\boldsymbol{\theta}) \propto \prod_{i=1}^{n} h(t_i, \boldsymbol{\theta}, \mathbf{x_i})^{\delta_i} \exp\left(-\int_0^{t_i} h(u)\, du \right) \quad \textbf{2.2}$$

2.1.3 Common Parametric Survival Distributions

If one is willing to accept a distributional form for the generated data, then the only requirement is that the survival distribution be bounded between 0 $(S(\infty) = 0)$ and 1 $(S(0) = 1)$. Some, of the more frequently used, common survival distributions are given below (see Klein and Moeschberger for a more extensive list) [6].

Distribution	Hazard Rate	Survival Function	Density Function
Exponential	λ	$\exp(-\lambda t)$	$\lambda \exp(-\lambda t)$
Weibull	$\alpha\lambda t^{\alpha-1}$	$\exp(-\lambda t^{\alpha})$	$\alpha\lambda t^{\alpha-1} \exp(-\lambda t^{\alpha})$

Figure 1. Common Parametric Survival Distributions

2.1.4 Proportional Hazards Models

Developed by D.R. Cox in 1972 the proportional hazards model is the most widely used analytic tool for survival analysis but is oftentimes misused and / or misunderstood. For example, given a simple treatment and control group, the ratio of their hazards $\Psi = \frac{h_{Trt}(t)}{h_{Ctrl}(t)}$ is assumed to be constant over time but this assumption is too often ignored or even checked prior to analysis. The class of accelerated failure time models is an alternative when the proportionality assumption does not hold and diagnostic tests have been developed to assess this assumption [6][8]. Since this ratio is non-negative, it is current convention to model Ψ , using covariates, as

$$h_i(t) = h_0 \exp\left(\boldsymbol{\beta}' \mathbf{x_i}\right) . \qquad 2.3$$

In this formulation the baseline hazard corresponds to the control group when $x_i = 0$ and to the treatment group when $x_i = 1$. The form of the hazard can be fully parametric, or left unspecified [8]. In order to estimate the parameter vector $\boldsymbol{\beta}$, Cox avoided the need to specify the baseline hazard and maximized his now famous partial likelihood (2.4)

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{D} \frac{\exp\left(\boldsymbol{\beta}' \mathbf{x}_i\right)}{\sum_{j \varepsilon R_j} \exp\left(\boldsymbol{\beta}' \mathbf{x}_j\right)} \qquad 2.4$$

which is now a function only of the covariates and their corresponding regression parameters. The partial likelihood (2.4) above can be extended to include time dependent covariates and reformulated as

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{D} \frac{\exp\left(\boldsymbol{\beta}' \mathbf{x}_i\left(t_j\right)\right)}{\sum_{j \varepsilon R_j} \exp\left(\boldsymbol{\beta}' \mathbf{x}_j\left(t_j\right)\right)}, \qquad 2.5$$

which can be maximized by iterative techniques, such as the Newton Raphson algorithm. This would imply complete knowledge of the covariate at each unique event time. This assumption is an issue when one would wish to include a covariate measured longitudinally over time and examine its effect on mortality. This issue will be revisited in the section on joint modeling.

2.1.5 Frailty Models

Duchateau et al [8] mention that the term 'frailty' was first used in gerontology and indicates that the more frail individuals have a greater risk of morbidity and death. Frailty is also defined as an individual level effect but is often used to describe an effect that is particular to a specific cluster, such as hospitals within a multicenter trial for example. Frailty can also be thought of as a random effect which describes the underlying heterogeneity in the population of interest. Klein et al [6] define frailty s being an "unobservable random effect shared by subjects within a subgroup" and that the most common model is one where the common random effect acts multiplicatively on the hazard rates of the members of the subgroup. This would imply that members with large frailty values experience the event earlier than those with smaller values. Klein et al [6] refer to the common use of the aptly named "shared frailty model", which is an extension of the proportional hazards model. They assume that the hazard for the jth subject in the ith subgroup, given the frailty, is defined by:

$$h_{ij}\left(t\right) = h_0\left(t\right)\exp\left(\sigma w_i + \boldsymbol{\beta}' \mathbf{Z}_{ij}\right), i = 1 \ldots G, j = 1 \ldots n_i, \qquad 2.6$$

where $h_0(t)$ is an arbitrary baseline is hazard rate, $\mathbf{Z}_{ij}$ is a vector of covariates, $\boldsymbol{\beta}$ the vector of coefficients, and $w_1 \ldots w_G$ the frailties. They assume that the frailties are from some distribution with mean zero and variance 1 and 2.6 reduces to the proportional hazards model when $\sigma = 0$.

An-alternative form for 2.6 is $h_{ij}(t) = h_0(t) w_i \exp\left(\boldsymbol{\beta}'\mathbf{Z}_{ij}\right), i = 1 \ldots G,\ j = 1 \ldots n_i$. In this form it is clearly seen that when $w_i > 1$ individuals within a given group tend to fail quicker than those with $w_i < 1$, under the assumed independence model $w_i = 1$. Since the w_i 's are unobserved the joint distribution of the individuals within a group is found by taking the expectation of $\exp\left(-\sum_{j=1}^{n_i} H_{ij}(t)\right)$ and is given by:

$$S\left(x_{i1} \ldots x_{in_i}\right) = P\left(X_{i1} > x_{i1} \ldots X_{in_i} > x_{in_i}\right) = LP\left[\sum_{j=1}^{n_i} H_0\left(x_{ij}\right)\right] \exp\left(\boldsymbol{\beta}'\mathbf{Z}_{\mathbf{ij}}\right) \text{ , 2.7}$$

where LP is the Laplace transform of the frailty. Some common distributional forms for the frailty are the one parameter gamma distribution, the inverse Gaussian distribution, and the log-normal distribution. Estimation of parameters in the semi-parametric model uses an EM algorithm as opposed to maximum likelihood estimation in the parametric models and is very computer intensive in either case. [6]

2.2 LONGITUDINAL MIXED EFFECTS MODELS

Laird and Ware [9] comment that the "defining feature of longitudinal studies is that measurements on the same individual are taken repeatedly through time". They also state that the primary goal of any longitudinal analyses is to capture the change in a response over time. Since repeated measurements are taken on an individual we can also capture the within subject change but this requires any modeling technique to be able to account for the correlation between the measurements. The sampling unit does not necessarily have to be the individual and there may be many levels of clustering. For example we may take measurements on siblings within a family, hospital wards within a hospital, and students within classrooms within schools. In any case we expect that measurements within a cluster are more closely related than measurements between clusters. For the remainder of this discussion we will assume that the sampling unit is the individual and we are interested in average changes, as well as, individual changes over time.

2.2.1 Notation

Let Y_{ij} be the response for the *ith* individual $(i = 1 \cdots N)$ at the jth measurement occasion $(j = 1 \cdots n_i)$. The n_i measurements can be arranged within a $n_i x 1$ vector $Y_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in} \end{bmatrix}$ and denoted by $Y_i' = \begin{bmatrix} Y_{i1} & Y_{i2} & \cdots & Y_{in} \end{bmatrix}'$ for convenience. Each response is associated with a $px1$ vector of

covariates $X_{ij} = \begin{bmatrix} X_{ij1} \\ x_{ij2} \\ \vdots \\ Y_{ijp} \end{bmatrix}$ $i = 1 \cdots N;\ j = 1 \cdots n_i$ at each occasion Y_{ij} and these can be grouped into a

$n_i xp$ covariate matrix $X_i = \begin{pmatrix} X_{i1}^T \\ X_{i2}^T \\ \vdots \\ X_{in_i}^T \end{pmatrix} = \begin{vmatrix} X_{i11} & X_{i12} & \cdots & X_{i1p} \\ X_{i21} & X_{i22} & \cdots & X_{i2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{in_i1} & X_{in_i2} & \cdots & X_{in_ip} \end{vmatrix} i = 1 \cdots N$ .Since our main interest is

in how the mean response changes over time and how these changes depend on covariates, the individual level mean response is denoted by $\mu_{ij} = E\left[Y_{ij}\right]$. This notation will suffice until expanded upon in subsequent sections. (See Laird and Ware [9] for a more detailed description)

2.2.2 Marginal Models

Marginal models, or population averaged models, are *marginal* in that the mean response depends only on the covariates of interest and not on any random effects. This differs from mixed effects models where the dependence is on the covariates plus a vector of random effects. Marginal models are widely used in the biomedical sciences and are very flexible in that they require no distributional assumption for the vector of responses, only a model for the mean response. Avoiding distributional assumptions for the response leads to the estimation methods, proposed by Liang and Zeger in 1986, called *Generalized Estimating equations (GEE)* [10]. They extended the quasi-likelihood score equation:

$$S_k(\beta) = \sum_{i=1}^{k} \frac{\delta \mu_i}{\delta \beta_k} v_i^{-1} (y_i - \mu_i) = 0 \ k = 1 \cdots p, \qquad 2.8$$

to longitudinal data via the multivariate extension of 2.8:

$$\sum_{i=1}^{k} \frac{\delta \boldsymbol{\mu}_i}{\delta \boldsymbol{\beta}_\mathbf{k}} \left(\mathbf{A}_\mathbf{i}^{1/2} \mathbf{R}_\mathbf{i} \left(\boldsymbol{\alpha} \right) \mathbf{A}_\mathbf{i}^{1/2} / \phi \right) \left(\mathbf{y}_i - \boldsymbol{\mu}_\mathbf{i} \right) = \mathbf{0} \quad . \qquad 2.9$$

Here $\mathbf{R}_\mathbf{i}(\boldsymbol{\alpha})$ is a $n_i x n_i$ fully specified "working correlation" matrix with sx1 vector of parameters α, $V =_i \left(\mathbf{A}_\mathbf{i}^{1/2} \mathbf{R}_\mathbf{i} \left(\boldsymbol{\alpha} \right) \mathbf{A}_\mathbf{i}^{1/2} / \phi \right)$ is the "working covariance matrix for y_i , $\mathbf{A}_i$ is a $n_i x n_i$ diagonal matrix consisting of a function of the mean $g\left(\mu_{ij} \right)$ along the main diagonal, and ϕ is a scaling factor. The term "working" is used to imply that the model assumes that the form of the covariance may not be correctly specified. [10]

Even though 2.9 is a function of $\alpha, \beta, and\ \phi$, it can be expressed as a function of β alone by inserting consistent estimates of $\alpha\ and\ \phi$ into 2.9. In the limit $\hat{\beta}_r$ is the consistent estimate of the solution to 2.10 via iteratively weighted least squares (see McCullagh and Neder 1983) [11]

$$\sum_{i=1}^{k} \frac{\delta \boldsymbol{\mu}_i}{\delta \boldsymbol{\beta}_\mathbf{k}} \left(\mathbf{A}_\mathbf{i}^{1/2} \mathbf{R}_\mathbf{i} \left(\hat{\boldsymbol{\alpha}} \right) \mathbf{A}_\mathbf{i}^{1/2} / \hat{\phi} \right) \left(\mathbf{y}_i - \boldsymbol{\mu}_\mathbf{i} \right) = \mathbf{0} \qquad 2.10$$

The strength of the GEE method is that it is robust to the choice of the "working" correlation structure and only requires that the mean response be correctly specified. This robustness property holds if there is a decreasing number of missing data or if the data is missing at random. [12]

2.2.3 Generalized Linear Mixed Effects Models

In section 2.2.2 we introduced the marginal model, via GEE, to account for within subject correlation. It was shown that the marginal model was robust to misspecification of the "working

correlation" structure as the models for the mean and covariance are modeled separately. Another way to account for the within subject correlation is to introduce subject specific random effects. Even though the GEE and random effect approaches both account for the within subject correlation, the interpretation of the estimated coefficients are entirely different. [9] The random effects can be thought of as reflecting heterogeneity in a study population due to unmeasured factors and typically come from a multivariate distribution. Conditional on these "random effects", the repeated measurements within an individual are assumed to be independent observations from a distribution belonging to the exponential family. This is referred to as the "conditional Independence assumption".

The linear mixed model can be thought of as an extension to a generalized linear model with fixed and random effects and requires a three part specification [9].

1. The conditional distribution of $\boldsymbol{Y_{ij}}$,given a qx1 vector of random effects $\boldsymbol{b_i}$, belongs to the exponential family, $Var\left(Y_{ij} \mid b_i\right) = v\left\{E\left[Y_{ij} \mid b_i\right]\right\}\phi$ is a function of the conditional mean, and given $\mathbf{b}_i$ the $\mathbf{Y_{ij}}^{\mathbf{s}}$ are independent.

2. The conditional mean depends on fixed covariates and random effects through the linear predictor: $\eta_{ij} = X_{ij}^{'}\beta + Z_{ij}^{'}b_i$ with $g\left\{E\left(Y_{ij} \mid b_i\right)\right\} = \eta_{ij} = X_{ij}^{'}\beta + Z_{ij}^{'}b_i$ for some link function, g (.).

3. The random effects are assumed to have a distribution but are commonly assumed to be multivariate normal with zero mean and qxq covariance matrix G and independent of the covariates.

In vector notation the linear mixed model can be expressed as:

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i, \qquad 2.11$$

where the $\boldsymbol{\beta}$'s are the regression coefficients for the fixed effects (same for all individuals), and the subject specific effects (random effects), b_i, are the deviation from the overall population mean trajectory for an individual. A distinguishing feature of the mixed effects model is that it generates simple expressions for the conditional mean response (subject specific)

$$E\left[Y_i \mid b_i\right] = X_i\beta + Z_i b_i \qquad 2.12$$

and the marginal mean response (averaged over all individuals),

$$E\left[Y_i\right] = X_i\beta. \qquad 2.13$$

The β's therefore have marginal interpretations (population averaged) with respect to their relationship to covariates.

The conditional covariance of the response given the random effects b_i is assumed to be diagonal with:

$$Cov\left(Y_i \mid b_i\right) = Cov\left(e_i\right) = R_i = \sigma^2 I_{n_i} \qquad 2.14$$

The marginal covariance of the response (the covariance of deviations from the ith individual's response from the population mean 2.13) is given by:

$$\begin{aligned} Cov\left(Y_i\right) &= Cov\left(Z_i b_i\right) + Cov\left(e_i\right) \\ &= Z_i G Z_i^T + R_i \\ &= Z_i G Z_i^T + \sigma^2 I_{n_i} \end{aligned} \qquad 2.15$$

This is not a diagonal matrix and illustrates how the introduction of random effects induces correlations, marginally amongst the Yi's. This correlation can also be thought of as a natural consequence of "sharing" a set of underlying random effects.

The betas in generalized linear mixed models have a different interpretation than in marginal models. Marginal models address changes in the mean response and how covariates affect these changes whereas the betas in conditional models address how a "specific" subject's mean response changes and how covariates affect these "specific" changes.

2.2.4 Estimation in Generalized Linear Models

In contrast with marginal models the joint distribution of the vector of responses and the vector of random effects are fully specified so inference and estimation can be based on the likelihood function. The joint distribution of the response Y_i and the random effects, b_i can be written as $f(Y_i b_i) = f(Y_i | b_i) f(b_i)$ where $f(Y_i | b_i) = f(Y_1 | b_1) f(Y_2 | b_2) \cdots f(Y_{n_i} | b_{n_i})$ (conditional independence assumption). In addition, $f(Y_i | b_i)$ is assumed to be from the exponential family of distributions and $f(b_i)$ is assumed to have a multivariate distribution with zero mean and covariance matrix G. In practice the random effects are unobservable so inference about β and G is based on the marginal or integrated likelihood function:

$$L(\beta, \phi, G) = \prod_{i=1}^{N} \int f(Y_i | b_i) f(b_i) db_i, \qquad 2.16$$

The likelihood in 2.16 is obtained by integrating out or averaging over, the unobserved random effects b_i. This results in the marginal likelihood which does not depend on the unobserved b_i's but only on their covariance G and on β and φ. The maximum likelihood

estimates are simply the estimates of G, β and φ that maximize 2.16 but unlike the linear mixed model case has no closed form solution. The likelihood in 2.16 is approximated by a technique known as Gaussian Quadrature which replaces the integral in 2.16 by a weighted sum:

$$L(\beta,\phi,G) \approx \prod_{i=1}^{N}\sum_{k=1}^{K} f\left(Y_i \mid b_i = v_i\right) w_k \qquad 2.17$$

The quadrature points (weights, w_k, and evaluation points w_k, are chosen to give the desired accuracy but there is a tradeoff between accuracy and runtime which has to be taken into account. Increasing the number of quadrature points substantially increases the computational burden and this grows exponentially with the number of random effects in the model. In most cases the time expended collecting data far exceeds the computational time so it is recommended that the quadrature points be increased until there is little or no change in the parameter estimates. Since the random effects are subject specific effects they can be predicted via:

$$\widehat{b_i} = E\left[b_i \mid Y_i, \widehat{\beta}, \widehat{\phi}, \widehat{G}\right] \qquad 2.18$$

This is known as the empirical best linear unbiased estimator or EBLUP with G, β, and φ, replaced by their respective maximum likelihood estimates $\widehat{\beta},\widehat{\phi},\widehat{G}$. Since $\widehat{b_i}$ is a conditional mean it also means that there is no analytic solution and so numerical techniques must also be employed.

For the family of generalized linear mixed models we assumed that the random effects b_i follow a multivariate normal with zero mean and covariance matrix G. In practice however it is very difficult to assess this normality assumption from empirical data. It is known that the predicted random effects in 2.18 are heavily influenced by the normality assumption and therefore cannot be used to assess the normality assumption. On the other hand, the fixed effects

are very robust to misspecification of the random effects but are sensitive to violation of the independence between the random effects and the covariates X.

2.3 TOBIT MODEL

First formulated in 1958 by James Tobin [14], the Tobit model postulates a latent or unobservable variable y^* which depends linearly on a set of covariates X_i by the parameter vector β. The observed variable y_i is equal to y^* if $y^* > \tau_L$ and τ_L otherwise. τ_L can be any real number but in the traditional Tobit model it is usually set to zero. Mathematically this is illustrated as follows:

$$y_i = \begin{cases} y^* \text{ if } y_i^* > 0 \\ 0 \text{ if } y_i^* \le 0 \end{cases} \qquad 2.19$$

The likelihood for the traditional Univariable Tobit model can be written in standard form as:

$$L = \prod_{i=1}^{N} \left[\frac{1}{\sigma} \phi \left(\frac{Y_i - X\beta}{\sigma} \right) \right]^{d_i} \left[1 - \Phi \left(\frac{X\beta}{\sigma} \right) \right]^{1-d_i} \qquad 2.20$$

with the log likelihood function given by

$$\ln L = \sum_{i=1}^{N} \left\{ d_i \left(-\ln \sigma + \ln \phi \left(\frac{y_i - X_i\beta}{\sigma} \right) \right) + (1 - d_i) \ln \left(1 - \Phi \left(\frac{X_i\beta}{\sigma} \right) \right) \right\} \qquad 2.21$$

The first part of the likelihood (2.20) is the regular density function and picks up contributions from uncensored observations (d_i=1). The second part is the cumulative distribution function and picks up contributions from censored observations (d_i=0).

The Tobit model formulation can be extended to arbitrary upper and lower censoring limits as follows:

$$y_i = \begin{cases} y^* & \text{if } y_L \le y_i^* \le y_U \\ y_L & \text{if } y_i^* \le y_L \\ y_U & \text{if } y_i^* \ge y_U \end{cases}$$

With log likelihood function:

$$\ln L = \sum_{i \in (y_L < y_i < y_U)} \ln\left[\frac{1}{\sigma}\phi\left(\frac{y_i - X_i^T\beta}{\sigma}\right)\right] + \sum_{i \in (y_i = y_L)} \ln\left[\Phi\left(\frac{y_L - X_i^T\beta}{\sigma}\right)\right] + \sum_{i \in (y_i = y_U)} \ln\left[1 - \Phi\left(\frac{y_U - X_i^T\beta}{\sigma}\right)\right] \quad 22$$

where ϕ is the standard normal density and Φ is the cumulative standard normal distribution function. The first part of the likelihood in 2.22 represents the contributions from uncensored values while the second and third parts represent the contributions from the left and right censored values, respectively. Even though the Tobit model is commonly used in censored data scenarios the assumptions have to be adhered to. If the error tem is non-normal or heteroscedastic, then the maximum likelihood estimates will be biased. It is also assumed that the underlying data generating process that generated the censored values is the same process that generated the outcome variable.

2.3.1 Tobit Mixed Effects Model

Extending the Tobit model to repeated measures is relatively straightforward. If we assume a random effects regression model and notation as outlined in Lyles et al [15]:

$$Y_{ij} = \alpha + a_i + (\beta + b_i)t_{ij} + e_{ij}$$

Y_{ij} is j^{th} measurement $t_{ij}\left(j = 1\cdots n_i\right)$ on the ith individual $\left(i = 1\cdots k\right)$. The deviations around the slope b_i and the intercept a_i are distributed as $a_i \sim N\left(0,\sigma_a^2\right)$ and $b_i \sim N\left(0,\sigma_b^2\right)$ respectively. The covariance between a_i and b_i is defined as $Cov\left(a_i,b_i\right) = \sigma_{ab}$ and a_i and b_i are independent of the errors $\varepsilon_{ij} \sim N\left(0,\sigma_e^2\right)$. If we assume an upper detection limit as d_1 and a lower detection limit as d_0 then our $Y_{n_i x1}$ vector of observations can be thought to consist of n_{id} detectable values, and $n_{id_0} + n_{id_1}$ undetectable values. The likelihood for the parameter vector $\theta = \left(\alpha,\beta,\sigma_a^2,\sigma_b^2,\sigma_{ab},\sigma_e^2\right)$ can be written, conditioning on the random effects as $L\left(\theta,Y\right) = \prod_{i=1}^{k} f^*\left(Y_i,\theta\right)$ where

$f^*(Y_i;\ \theta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f^*\left(Y_i,a_i,b_i\right)da_i db_i = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f^*\left(Y_i \mid a_i,b_i\right) f\left(a_i,b_i\right) f\left(b_i\right)da_i db_i$. The asterisk denotes that the vector **Y** may contain detectable and non-detectable values. These are non-detectable values falling below the lower detection limit d_0, above the upper detection limit d_1, or detectable values falling in between d_0 and d_1. The n_{id_0} values falling below the detection limit d_0 contribute $F\left(d_{0i} \mid a_i b_i\right)$ to the likelihood. The n_{id_1} Values falling above the detection limit d_1 contribute $1 - F\left(d_{1i} \mid a_i b_i\right)$ and the n_{id_0} values falling in between d_0 and d_1 contribute $f\left(Y_{ij} \mid a_i b_i\right)$. If we assume that the first n_{id} are detectable, the next n_{id_0} are below the lower detection limit, and the last n_{id_1} are above the upper detection limit then the complete likelihood can be written as

$$L(Y_i;\ \theta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left\{ \prod_{j=1}^{n_{id}} f\left(Y_{ij} \mid a_i, b_i\right) \right\} * \left\{ \prod_{j=n_{id}+1}^{n_{id_0}} F\left(d_{i0} \mid a_i, b_i\right) \right\} \left\{ \prod_{j=n_{id_0}+1}^{n_i} 1 - F\left(d_{i1} \mid a_i, b_i\right) \right\} f\left(a_i \mid b_i\right) f\left(b_i\right) da_i db_i \quad 2.23$$

As Lyle et al [15] point out, when expressed this way, the likelihood can be passed to a suitable numerical optimizer after approximating the integrals using adaptive or non-adaptive quadrature.

3.0 JOINT MODELS

3.1 BACKGROUND

The clinical literature is replete with studies that simultaneously measure a covariate longitudinally and investigate its effect on time to event. The separate analyses of longitudinal and survival data is well understood but what is often overlooked is that these models can sometimes be associated and can produce biased estimates of treatment efficacy if this association is not taken into account. Historically joint modeling is a way of accounting for this association.

Tsiatis and Davidian [16], in their overview, mention that, traditionally, the association between the survival and longitudinal data is usually through the proportional hazards model

$$\begin{aligned}\lambda_i(u) &= \lim_{du\to 0} du^{-1} pr\left\{u \le T_i < u + du \mid T_i \ge u, X_i^H(u), Z_i\right\} \\ &= \lambda_0(u)\exp\left\{\gamma X(u)_i + \eta^T Z_i\right\}\end{aligned} \qquad 3.1$$

We can visualize the ideal observed data for each subject $i = 1 \cdots n$ as consisting of $\left\{T_i, Z_i, X_i(u), u > 0\right\}$ where T is the event time; Z is a vector of baseline covariates and $\left\{X_i(u), u > 0\right\}$ is the longitudinal trajectory and $X_i^H(u) = \left(X_i(t), 0 \le t < u\right)$ is the longitudinal covariate history up to time u. X (u) can be thought of as a time dependent covariate and association is through the estimation of γ. The above formulation seems straightforward but

assumes that the covariate history is available at all times u and is measured without error. In practice the covariate (i) may be measured at times $t_{ij} \le T_i$, (ii) we don't observe the true covariate value $X_i\left(t_{ij}\right)$ but rather $W_i\left(t_{ij}\right) = X_i\left(t_{ij}\right) + \varepsilon_{ij}$ and (iii) the event times T are not observed for all subjects. Also making inference on the longitudinal process is hampered by the time-to-event causing informative dropout. With these complications in mind it's apparent that any statistical procedure must take into account the potential relationship between the survival and longitudinal processes. If our interest was on time to event and $X_i\left(u\right)$ was available for all $u \le T_i$, then as Tsiatis et al [16] point out, "the main difficulty would be accounting for the censoring "and Inference, in the presence of censoring, can be achieved by turning to the Cox partial likelihood:

$$\prod_{i=1}^{n}\left[\frac{\exp\left\{\gamma X_i\left(V_i\right)+\eta^T Z_i\right\}}{\sum_{j=1}^{n}\exp\left\{\gamma X_i\left(V_i\right)+\eta^T Z_i\right\} I\left(V_j \ge V_i\right)}\right]^{\Delta_i} \quad 3.2$$

Here C_i is the censoring variable for subject I, $V_i = \min\left(T_i, C_i\right)$, and $\Delta_i = I\left(T_i \le C_i\right)$. As was previously mentioned $X_i\left(u\right)$ is typically not available for all $u \le T_i$ and so some authors turn to imputation. The last value carried forward (LVCF) is a popular, though naive, approach and replaces the covariate at the event time by its last available value but Prentice [17] showed that this leads to biased estimates. Other variations on the LVCF are nearest observed value (NV), Linear Interpolation, (LI) [18] and Last Value Auto Regressed (LVAR) [19].

As Liu and Craig point out [19] the current methods in the literature include jointly modeling the survival and longitudinal components and this is usually done by assuming that the longitudinal

model follows a linear mixed effects model and that the survival model depends on the random effects from this process. Inference is then based on the integrated conditional joint likelihood where the random effects usually follow a multivariate normal distribution.

The other variations on this joint modeling approach try to avoid specifying a distribution for the survival and/or longitudinal process or differ in the estimation process. Faucett and Thomas [20] use a linear mixed model and Bayesian methods for the parameter estimation. Henderson et al [21] use the EM (Expectation Maximization) algorithm, Xu and Zeger [22] use a latent variable approach and implement a Markov Chain Monte Carlo Algortihm for the estimation, and Tsiatis and Davidian [23] derived their conditional score estimator which requires no distributional assumptions on the random effects. De Gruttola and Tu [24] implement a fully parametric joint model by assuming that the survival and longitudinal processes follow a multivariate normal distribution or a 1:1 transformation of a multivariate normal distribution and estimation is via the EM algorithm. Brown et al [25] proposed a flexible B-Spline model when the longitudinal measures could not be well approximated by a multivariate normal distribution and linked the models through the proportional hazards model. Other, perhaps simpler, methods use the two stage approach where the true value of the covariate at the event time is estimated by a linear mixed effects model in the first stage and the best linear unbiased predictors (EBLUP's) are then substituted into the hazards model in the second stage.

From the above examples it is apparent that most of the methods for joint modeling are similar to each other, but differ in the parametric assumptions on the longitudinal models, survival models, or both, and in the estimation methods proposed. Irrespective of the methods employed what is clear is that ignoring the association produces biased estimation of the parameters involved and these methods, though arguably very complex and computationally

intensive, are used to obtain a better estimate of the possible treatment effects in the presence of this association.

3.2 NOTATION

The notation to follow will follow the excellent review of joint modeling outlined by Tsiatis and Davidian [16]. For subject $i, i = 1 \cdots n$, let T_i and C_i denote the event and censoring times respectively. Let Z_i be q-dimensional vector of baseline covariates and let $X_i(u)$ be the longitudinal process at time $u \geq 0$. In practice we do not observe T_i but only $V_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \leq C_i)$ and $X_i(u)$ is measured at intermittent times $t_{ij} \leq V_i, j = 1 \cdots m_i$ for each subject. The observed longitudinal data is, of course, subject to random error and so what we actually observe is $W_i = \left\{ W_i(t_{i1}) \cdots W_i(t_{im_i}) \right\}^T$ which may not equal the true $X_i(t_{ij})$.

A joint model is made up of two linked sub-models, one for the hypothesized "true" longitudinal process $X_i(t_{ij})$ and one for the failure time T_i. With the necessary specifications and assumptions, this allows a full representation of the joint distribution of the observed data $O_i = \{V_i, \Delta_i, W_i, t_i\}$ where $t_i = \{t_{i1}, \cdots t_{im_i}\}$ where the O_i's are assumed to be independent. It is usual to characterize the longitudinal model in terms of a vector of random effects α_i. The true longitudinal model is often represented as follows:

$$X_i(u) = \alpha_{0i} + \alpha_{1i} u, \quad \alpha_i = (\alpha_{0i}, \alpha_{1i})^T \qquad 3.3$$

where α_{0i} and α_{1i} are the intercept and slope, respectively. It is not necessary to restrict the trajectory to be linear and, in fact, the trajectory can be a flexible polynomial function such as $X_i(u) = \alpha_{0i} + \alpha_{1i}u + \alpha_{2i}u^2 + \cdots + \alpha_{pi}u^p$. This can be written more generally as $X_i(u) = f(u)^T \alpha_i$ where $f(u)^T$ can be any function of time u. The random effects are usually assumed to be normally distributed to reflect inter-subject perturbations around the true longitudinal trajectory.

For the event time process a proportional hazards model is posited as in 3.1 where dependence is on $X_i(u)$ and hence on the random effects $\alpha_i = (\alpha_{0i}, \alpha_{1i})^T$. This then would imply that, given the longitudinal history and any covariates, that the hazard is associated with an assumed constant smooth trend. Since we don't know the true "trend" we assume the observed data are as follows:

$$W_i(t_{ij}) = X_i(t_{ij}) + \varepsilon_i(t_{ij}) \qquad 3.4$$

Where $\varepsilon_i(t_{ij}) \sim N(0, \sigma^2)$ are independent of the random effects α_i and are interpreted as random fluctuations due to measurement error. As Tsiatis and Davidian [16] point out, 3.4, along with 3.3, specify a standard linear mixed model (see Laird and Ware 1982) [27].

Let δ denote the parameters in the multivariate normal distribution $p(\alpha_i, | Z_i; \delta)$ then the usual form of the joint likelihood for the full set of parameters $\Omega = (\lambda_0(.), \gamma, \eta, \sigma^2, \delta)$ given Z_i is

$$\prod_{i=1}^{n} \int \left[\lambda_0(V_i) \exp\left\{ \gamma X_i(V_i) + \eta^T Z_i \right\} \right]^{\Delta_i} \exp\left[-\int_0^{V_i} \lambda_0(u) \exp\left\{ \gamma X_i(u) + \eta^T Z_i \right\} du \right]$$
$$* \frac{1}{(2\pi\sigma^2)^{m_i/2}} \exp\left[-\sum_{j=1}^{m_i} \frac{\left\{ W_i(t_{ij}) - X_i(t_{ij}) \right\}^2}{2\sigma^2} \right] p(\alpha_i \mid Z_i; \delta) \, d\alpha_i \qquad 3.5$$

and is the basis for statistical inference (see [16] for details and derivation).

4.0 ACCOUNTING FOR A DOUBLY CENSORED LONGITUDINAL COVARIATE IN SURVIVAL ANALYSIS VIA JOINT MODELING.

4.1 INTRODUCTION

Joint modeling is becoming increasingly popular in the literature particularly when the hypotheses of interest are in (a) accounting for a time dependent covariate in a survival model that is measured with error, (b) adjusting longitudinal inferences in the presence of outcome-dependent dropout, or (c) understanding the association between a longitudinal and time to event outcome [27]. The naïve imputation methods that were popular prior to Prentice's exposition [17], showing that joint modeling reduces the bias in the estimates, have taken a back seat and emphasis is on making current joint modeling methods more accessible and applicable to a myriad of real life data analysis scenarios.

The basic idea behind joint modeling is to assume that the "true" longitudinal trajectory follows a linear mixed model [27] with random intercept and slope and that the observed data is then modeled as a function of these random effects plus measurement error. The longitudinal and survival processes are then "linked" by assuming that the hazard is a function of this "true" longitudinal trajectory [16].

Carlin et al, in their 2004 paper [26] outline the method proposed by Henderson et al [21] who developed a joint model flexible enough to include fixed effects, random effects, serial correlation, measurement error, and, in the absence of association, returns the same estimates as if separate models were run. Their main idea is to "link" the survival and longitudinal processes via a latent bivariate Gaussian process and to assume that the survival and longitudinal models were independent given this process and available covariates.

Another issue which can further complicate the use of joint modeling methodology is when the longitudinal outcome is subject to censoring. This censoring can be due to natural or somewhat unnatural detection limits. Natural detection limits arise due to the operating characteristics of a measurement instrument whereas unnatural detection limits can arise due to an investigator arbitrarily imposing restrictions on a measurement due to a perceived biological implausibility. The resulting data is then referred to as being right or left censored. This means that data above or below the detection limits are replaced by these lower and upper limits respectively.

Two popular methods of accounting for censoring are the so called "fill in" [2] and Tobit methods [14]. The "fill in" method replaces the censored data by a function of the respective detection limits and then employs regular regression, whereas the Tobit method accounts for the censoring by constructing the appropriate likelihood. This likelihood is a product of a distribution for the censored and uncensored observations respectively. The resulting maximum likelihood estimators are then less biased [2] than when ignoring the censoring as in the "fill in" methods.

Carlin et al (2004) [26] implemented a variation on Henderson et al's joint modeling approach in SAS NLMIXED and WINBUGS. This paper seeks to modify Carlin et al's approach in SAS NLMIXED to account for left and right censoring in the longitudinal model. The censoring will be accounted for by assuming a linear mixed Tobit model for the longitudinal process, and, for ease of exposition, an exponential distribution will be assumed for the event time process. To compare the Tobit based joint model to the "plug in" joint model methods, simulation studies will be conducted with levels of censoring varying from 20-60% and with sample sizes of 100, 250, 500, and 1000.

4.2 NOTATION

The following notation will follow that of Carlin et al [26]. Assume a set of *m* subjects are followed over an interval [0, τ] who possibly contribute partially missing data $\left\{y_{ij},\, j=1\cdots n_i\right\}$ at times $\left\{s_{ij},\, j=1\cdots n_i\right\}$ and a censored survival time t_i. The longitudinal data will be assumed to follow the popular linear mixed model

$$y_{ij} = \mu_{ij}\left(s_{ij}\right) + W_{1i}\left(s_{ij}\right) + \varepsilon_{ij} \qquad 4.1$$

$\mu_{ij}\left(s_{ij}\right) = X_{1i}^{T}\left(s\right)\beta_1$ is the mean response, $W_{1i}\left(s_{ij}\right) = d_{1i}^{T}\left(s\right)U_i$, are subject specific random effects, and $\varepsilon_{ij} \sim N\left(0,\sigma_e^2\right)$ are measurement errors. In this model $W_{1i}\left(s_{ij}\right)$ is assumed to be the true trajectory, $X_{1i}^{T}\left(s\right)\beta_1$ are available covariates which may or may not be time varying, and β are the corresponding coefficients. The U_i's are a set of random effects which are typically a subset of $X_{1i}^{T}\left(s\right)$, and are assumed to be $N\left(0,\Sigma\right)$ iid random variables. This assumption will not

be investigated in this paper since Rizopoulos et al [29] discovered that misspecification of the random effects distribution has little to no effect on parameter estimates and standard errors.

The survival model is typically assumed to be either non parametric or parametric but since Hsieh, Tseng, and Wang [28] discovered that leaving the baseline hazard completely unspecified results in underestimated standard errors of the estimates, in this paper we will assume an exponential distribution such that the hazard, conditional on a set of covariates plus a frailty, can be written as:

$$\lambda_i(t) = \lambda \exp\left(X_{2i}^T(t)\beta_2 + W_{2i}(t) \right), \qquad 4.2$$

where $X_{2i}(t)$ denotes the available covariates and their regression coefficients, respectively and $W_{2i}(t)$ is a frailty and similar in form to W_{1i}.

The joint model can be constructed by assuming that there is dependence between the variables $W_{1i}(t), W_{2i}(t)$ thus linking the longitudinal and survival processes. When association exists, the estimators should be more efficient and less biased. Carlin et al link the survival and longitudinal models as follows:

$$\begin{aligned} &W_{1i}(s) = U_{1i} + U_{2i}s \\ &\textit{and} \\ &W_{2i}(t) = \gamma_1 U_{1i} + \gamma_2 U_{2i} + \gamma_3 \left(U_{1i} + U_{2i}t \right) + U_{3i} \end{aligned} \qquad 4.3$$

$W_{1i}(s)$ can be recognized as the usual Laird and Ware formulation of a random intercept and slope model. The parameters $\{\gamma_1, \gamma_2, \gamma_3\}$ capture the association between the two models by the random intercepts, slopes, and fitted longitudinal value at the event time. The random variables

$\{U_{1i}, U_{2i}\}^T$ are assumed to be zero mean bivariate Gaussian $N(0, \Sigma)$, while U_{3i} are frailty terms orthogonal to $\{U_{1i}, U_{2i}\}^T$ and are iid $N(0, \sigma_3^2)$.

4.3 JOINT TOBIT LIKELIHOOD

Let θ be the unknown parameter vector, and Y^* and N the longitudinal and event time distributions respectively where Y^* indicates that the n_ix1 vector y_i^* contains censored and uncensored values. The likelihood *L=L (θ, Y*, N)* can be constructed by assuming that $W_{2i}(t)$ is proportional to $W_{1i}(t)$ and that the survival and longitudinal models are independent given $W_{2i}(t)$. Then the joint likelihood *L (θ, Y*, N)* can be obtained by factoring the conditional likelihood $L|W_2$ as follows:

$$L(\theta, Y^*, N) = \int_{w_2} L(Y^*, N \mid W_2; \theta) L(W_2)\, dw_2 = \int_{w_2} L(Y^* \mid W_2; \theta) L(N \mid W_2; \theta) L(W_2; \theta)\, dw_2 \quad 4.4$$

$$L(Y^* \mid W_2; \theta) = \prod_j p(y_i^* \mid w_{2i}; \theta) \quad 4.5$$

Hence the contribution of the *ith* subject to the joint log-likelihood can be expressed as:

$$L(\theta, Y^*, N) = \log \int_{w_2} \left[\prod_j p(y_i^* \mid w_{2i}; \theta) \right] L(N \mid W_2; \theta) L(W_2; \theta)\, dw_2 \quad 4.6$$

The Tobit likelihood $\prod_j p(y_i^* \mid w_{2i}; \theta)$ can be written in a linear mixed model form as:

$$L(Y_i \mid W_{2i};\ \theta) = \left\{ \prod_{j=1}^{n_{id}} f(Y_i \mid W_{2i}) \right\} \left\{ \prod_{j=n_{id}+1}^{n_{id_0}} F(d_{i0} \mid W_{2i}) \right\} \left\{ \prod_{j=n_{id_0}+1}^{n_i} 1 - F(d_{i1} \mid W_{2i}) \right\} \quad 4.7$$

The first part of the likelihood in 4.7picks up contributions from the n_{id} uncensored values, and the second and third part pick up contributions from n_{id0} and n_{id1} values censored from below and above respectively. The exponential survival likelihood $L\left(N \mid W_2;\theta\right)$ can be written as a product of the hazard and the cumulative hazard as follows:

$$L\left(N \mid W_2;\theta\right) = \lambda \exp\left(X_2^T\beta + \gamma_1 U_{1i} + \gamma_2 U_{2i} + \gamma_3\left(U_{1i} + U_{2i}s\right) + U_{3i}\right) \quad 4.8$$

and $L\left(W_2;\theta\right)$ is the multivariate normal density for the random effects.

4.4 SIMULATION OUTLINE

This simulation study will compare the Tobit and "Fill-In" joint models to account for a doubly censored covariate in a survival model as in Henderson et al [21]. For the longitudinal and survival sub-models we will focus on the random effects $\gamma 1*u1+\gamma 2*u2 + u3$ as $\gamma 3*(u1+u2*t)$ is the fitted longitudinal value at the event time and thus implies knowledge of the survival times a priori.

The longitudinal model was taken to be:

$$Y_t = \beta_{11} + \beta_{12}t + \beta_{13}X + U_1 + U_2 t + \varepsilon_t$$

X~N (0, 1) and n=6 measurements at t=0, 0.5, 1, 1.5, 2, and 3 units. The true parameter values for the longitudinal data generation were $\left\{\varepsilon = 0.25, \beta_{11} = 0, \beta_{12} = 1, \beta_{13} = 1\right\}$ with covariance matrix $\Sigma_U = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.25 \end{bmatrix}$. The event times were generated from the model:

$$\alpha(t) = \alpha_0\left(t\right)\exp\left(\beta_{21}X + \gamma_1 U_1 + \gamma_2 U_2 + U_3\right)$$

The exponential baseline hazard $\alpha_0(t)$ was λ=0.8 and exponential censoring hazard was λ=0.2, as this gave approximately 40% drop out by time 1, and 70% drop out by time 3. The true parameter values for the survival data generation were $\{\beta_{21}=1, \gamma_1=-1.5, \gamma_2=-1.0\}$.

Each simulation was run N=1000 times with sample sizes {100, 250, 500, 1000) with censoring percentages {20%, 40%, 60%}. The estimates for the joint models were obtained from SAS PROC NLMIXED with options ABSGCONV=1E-05 TECH= TRUREG and QPOINTS=3. Simulations were initially run with a single quadrature point but then increased to three for better parameter estimates and coverage probabilities.

4.5 SIMULATION RESULTS

The Tables, 4.1-4.3, below are a representative sample of the simulations that cover the censoring range of 20-60%. In all cases the joint Tobit model is relatively unbiased, has smaller variance, and exhibits better coverage probabilities than the fill in methods regardless of the censoring proportion. In all cases it appears that the fill in methods exhibit poorer coverage probabilities, larger variances, and biases, as the proportion of censoring increases. The good performance of the Tobit joint model may be attributed to the fact that we generated data conforming to the assumptions for the Tobit model. These assumptions are (i) the data come from a normal distribution and (ii) that the error variances are homogenous. How the Tobit model would perform if these assumptions are violated is to be considered subsequently.

Table 4.1. Simulation results (20% Censoring, N=1000, n=250)

Obs	parameter	true	Mean Joint Tobit	VARJNTTOBIT	bias Tobit	Tobit Coverage	Mean JointDL	VARJNTDL	biasDL	DLCoverage
1	bL0	0.00	-0.00428	0.003305	0.004275	94.95	0.1330	0.00342	0.13303	29.04
2	bL1	1.00	1.0397	0.018665	0.039679	94.85	0.9793	0.01884	0.02071	88.17
3	bL2	1.00	1.0002	0.003141	0.000159	95.16	0.8564	0.00264	0.14360	19.75
4	bs0	0.80	0.8027	0.012777	0.002663	95.36	0.8027	0.01275	0.00275	95.88
5	bs1	1.00	1.0165	0.019386	0.016543	95.36	1.0347	0.02106	0.03471	95.14
6	r1	-1.50	-1.5464	0.061240	0.046354	96.06	-2.1230	0.11286	0.62300	57.72
7	r2	-1.00	-0.9652	0.030670	0.034777	96.06	-0.8822	0.18923	0.11782	96.83
8	s11	0.50	0.4928	0.005748	0.007186	95.66	0.3521	0.00408	0.14795	46.36
9	s22	1.00	0.9513	0.047159	0.048676	94.95	0.4117	0.01247	0.58830	3.48
10	s33	0.25	0.2775	0.033614	0.027531	96.57	0.3146	0.04053	0.06463	94.93
11	se	0.25	0.2536	0.001357	0.003631	94.25	0.3221	0.00170	0.07212	41.50

Obs	Mean JointDL2	VARJNTDL2	biasDL2	DL2Coverage	Mean Joint DLSQRT2	VARJNTDLSQRT2	bias DLSQRT2	DLSQRT2Coverage
1	0.2609	0.00358	0.26092	0.12	0.2059	0.00345	0.20592	3.39
2	0.8155	0.01389	0.18455	56.07	0.8867	0.01687	0.11330	73.95
3	0.7132	0.00234	0.28684	.	0.7759	0.00232	0.22410	0.47
4	0.8125	0.01341	0.01246	96.03	0.8081	0.01318	0.00810	95.56
5	1.0400	0.02261	0.04003	94.86	1.0322	0.02176	0.03223	95.79
6	-3.0844	3.41322	1.58437	2.11	-2.6475	0.15430	1.14746	16.26
7	0.03934	0.93369	1.03934	85.16	-0.5200	0.57734	0.47997	93.69
8	0.2040	0.00216	0.29602	0.70	0.2681	0.00288	0.23194	6.89
9	0.1910	0.00441	0.80905	.	0.2498	0.00638	0.75016	.
10	0.2502	0.02491	0.00016	99.45	0.2822	0.03067	0.03221	98.88
11	0.4912	0.00436	0.24122	.	0.3880	0.00239	0.13800	5.14

Table 4.2: Simulation Results (40% Censoring, N=1000, n=250)

Obs	parameter	true	Mean Joint Tobit	VARJNTTOBIT	bias Tobit	Tobit Coverage	Mean JointDL	VARJNTDL	biasDL	DLCoverage
1	bL0	0.00	-0.00038	0.003442	0.000375	96.24	0.2806	0.00491	0.28064	.
2	bL1	1.00	1.0458	0.021159	0.045796	93.90	0.7512	0.01472	0.24878	29.02
3	bL2	1.00	1.0055	0.003962	0.005494	96.04	0.6745	0.00228	0.32552	.
4	bs0	0.80	0.8054	0.012151	0.005435	95.53	0.8106	0.01186	0.01062	95.80
5	bs1	1.00	1.0055	0.020692	0.005548	95.12	1.0196	0.02307	0.01963	94.10
6	r1	-1.50	-1.5618	0.074284	0.061819	96.13	-2.5813	4.00829	1.08135	22.11
7	r2	-1.00	-0.9639	0.042317	0.036051	96.04	-0.6104	1.01656	0.38961	91.27
8	s11	0.50	0.4831	0.007523	0.016884	96.75	0.2350	0.00269	0.26498	1.25
9	s22	1.00	0.9154	0.059904	0.084598	95.22	0.1815	0.00386	0.81847	.
10	s33	0.25	0.2821	0.037552	0.032058	96.64	0.3513	0.04054	0.10131	96.23
11	se	0.25	0.2619	0.002038	0.011947	95.53	0.2675	0.00137	0.01752	84.81

Obs	Mean JointDL2	VARJNTDL2	biasDL2	DL2Coverage	Mean Joint DLSQRT2	VARJNTDLSQRT2	bias DLSQRT2	DLSQRT2Coverage
1	0.3611	0.00	0.36109	.	0.3234	0.00473	0.32342	.
2	0.5131	0.01	0.48693	0.11	0.6121	0.01028	0.38792	0.99
3	0.4962	0.00	0.50376	.	0.5705	0.00181	0.42948	.
4	0.8195	0.01	0.01946	95.72	0.8156	0.01195	0.01562	96.17
5	1.0351	0.02	0.03514	93.58	1.0184	0.02442	0.01842	93.20
6	-1.7929	1364.67	0.29291	6.83	-3.3118	0.36282	1.81178	8.67
7	2.4449	2.51	3.44491	38.33	0.7157	2.96963	1.71572	73.05
8	0.1089	0.00	0.39110	1.02	0.1602	0.00174	0.33983	0.12
9	0.05943	0.00	0.94057	0.11	0.08887	0.00130	0.91113	.
10	0.2438	0.04	0.00624	99.08	0.3211	0.03030	0.07111	98.24
11	0.2943	0.00	0.04432	84.28	0.2582	0.00145	0.00823	83.31

Table 4.3: Simulation Results (60% Censoring, N=1000, n=250)

Obs	parameter	true	Mean Joint Tobit	VARJNTTOBIT	bias Tobit	Tobit Coverage	Mean JointDL	VARJNTDL	biasDL	DLCoverage
1	bL0	0.00	-0.00250	0.00493	0.00250	94.91	0.4308	0.00918	0.43081	.
2	bL1	1.00	1.0696	0.02450	0.06964	93.59	0.5079	0.00817	0.49208	.
3	bL2	1.00	1.0086	0.00606	0.00865	95.02	0.4606	0.00163	0.53941	.
4	bs0	0.80	0.7993	0.01400	0.00075	94.51	0.8066	0.01388	0.00656	94.23
5	bs1	1.00	1.0125	0.01899	0.01248	95.73	1.0190	0.02125	0.01897	95.52
6	r1	-1.50	-1.6348	0.10392	0.13478	96.34	-3.7661	0.67344	2.26606	4.24
7	r2	-1.00	-0.9761	0.07627	0.02392	96.85	0.1137	4.95618	1.11374	80.33
8	s11	0.50	0.4693	0.01043	0.03074	96.44	0.1189	0.00099	0.38110	.
9	s22	1.00	0.8704	0.08309	0.12958	93.18	0.06643	0.00070	0.93357	.
10	s33	0.25	0.2918	0.03707	0.04179	97.38	0.3955	0.04145	0.14545	93.93
11	se	0.25	0.2749	0.00317	0.02488	94.40	0.1692	0.00070	0.08075	9.54

Obs	Mean JointDL2	VARJNTDL2	biasDL2	DL2Coverage	Mean Joint DLSQRT2	VARJNTDLSQRT2	bias DLSQRT2	DLSQRT2Coverage
1	0.3887	0.0062	0.38872	.	0.4070	0.0072	0.40701	.
2	0.2897	0.0029	0.71031	.	0.3854	0.0044	0.61459	.
3	0.3047	0.0011	0.69526	.	0.3717	0.0012	0.62832	.
4	0.8077	0.0163	0.00768	94.79	0.8089	0.0143	0.00888	94.05
5	1.0323	0.0209	0.03233	95.70	1.0246	0.0218	0.02460	94.87
6	-4.8227	73.0442	3.32274	12.79	-4.5804	1.3670	3.08043	5.72
7	6.4934	11.3790	7.49338	26.05	3.1655	11.8545	4.16545	55.54
8	0.05009	0.0003	0.44991	0.23	0.07589	0.0005	0.42411	0.12
9	0.01717	0.0001	0.98283	0.23	0.02901	0.0002	0.97099	0.12
10	0.2662	0.5642	0.01617	99.59	0.3409	0.0344	0.09088	97.85
11	0.1418	0.0008	0.10822	1.13	0.1379	0.0006	0.11213	0.35

4.6 SENSITIVITY

As was mentioned earlier the simulation results show that the joint Tobit linear mixed model seemed to perform better than the "fill-in" methods irrespective of the proportion of censoring. The results to follow investigate how robust the Tobit model is to the assumptions of normality and homogenous variance. For the normality assumption we generated from a mixture of multivariate normal distributions and a multivariate T distribution. For the normal mixture distribution a Bernoulli (0, 1) random variable was generated with probability *p*. The data from the true distribution of the response variable was polluted with *p%* from a secondary multivariate normal distribution (MVN) with mean $\boldsymbol{\beta}$+c and common covariance Σ.

We also generated data from a multivariate T distribution with degrees of freedom ranging from 1-5, with the same mean, and covariance matrix as the true distribution of Y. This was to see what would happen to our estimates if we assumed a normal distribution with some mean and covariance when the data actually came from a T distribution with those same parameters. The assumption of non-constant variance was assessed by adding a small constant to one or more of the diagonal elements of the true covariance matrix, Σ, then fitting a model which assumes homogenous variances.

Tables 4.4 and 4.5 below show that with 20% pollution, and a mean deviation of $\boldsymbol{\beta}$+**0.5** from the truth does not impact the estimates or the coverage probabilities to any serious extent. However in tables 4.6 and 4.7 we can notice that as the proportion of pollution increases the bias tends to increase and coverage probabilities decrease.

Table 4.4. Mixture of Normal's with 20% pollution, 20% Censoring and c=0.5

Obs	parameter	true	Mean Joint Tobit	VARJNTTOBIT	bias Tobit	Tobit Coverage
1	bL0	0.00	0.03901	0.003185	0.039007	89.73
2	bL1	1.00	1.0668	0.019042	0.066819	93.96
3	bL2	1.00	1.0461	0.003568	0.046059	87.71
4	bs0	0.80	0.8109	0.013609	0.010861	94.96
5	bs1	1.00	0.9937	0.018782	0.006269	95.77
6	r1	-1.50	-1.5064	0.062573	0.006377	95.17
7	r2	-1.00	-0.9653	0.031469	0.034702	96.68
8	s11	0.50	0.5053	0.005954	0.005259	95.57
9	s22	1.00	0.9617	0.045584	0.038278	96.58
10	s33	0.25	0.2982	0.033595	0.048226	96.29
11	se	0.25	0.2545	0.001407	0.004548	94.76

Table 4.5. Mixture of Normal's with 20% pollution, 60% Censoring and c=0.5

Obs	parameter	true	Mean Joint Tobit	VARJNTTOBIT	bias Tobit	Tobit Coverage
1	bL0	0.00	0.03029	0.004786	0.03029	92.35
2	bL1	1.00	1.1033	0.024561	0.10325	92.25
3	bL2	1.00	1.0487	0.006116	0.04868	92.15
4	bs0	0.80	0.8130	0.013224	0.01302	95.47
5	bs1	1.00	0.9999	0.022231	0.00006	94.36
6	r1	-1.50	-1.5665	0.091605	0.06649	97.08
7	r2	-1.00	-0.9597	0.067260	0.04026	96.58
8	s11	0.50	0.4862	0.010270	0.01380	97.38
9	s22	1.00	0.8941	0.097054	0.10586	93.25
10	s33	0.25	0.3038	0.035798	0.05378	97.26
11	se	0.25	0.2726	0.002952	0.02264	94.86

Table 4.6. Mixture of Normal's with 40% pollution, 20% Censoring and c=0.5

Obs	parameter	true	Mean Joint Tobit	VARJNTTOBIT	bias Tobit	Tobit Coverage
1	bL0	0.00	0.1664	0.00621	0.16639	43.27
2	bL1	1.00	1.2500	0.04255	0.24998	79.18
3	bL2	1.00	1.1823	0.01047	0.18225	57.14
4	bs0	0.80	0.8070	0.01119	0.00697	96.73
5	bs1	1.00	0.9679	0.02213	0.03212	93.06
6	r1	-1.50	-1.3837	0.11705	0.11630	90.20
7	r2	-1.00	-0.8589	0.26812	0.14106	93.47
8	s11	0.50	0.5369	0.01676	0.03688	95.92
9	s22	1.00	0.9076	0.15627	0.09244	93.06
10	s33	0.25	0.4263	0.04473	0.17626	89.87
11	se	0.25	0.2975	0.00467	0.04748	92.65

Table 4.7. Mixture of Normal's with 40% pollution, 60% Censoring and c=0.5

Obs	parameter	true	Mean Joint Tobit	VARJNTTOBIT	bias Tobit	Tobit Coverage
1	bL0	0.00	0.1912	0.003171	0.19115	10.25
2	bL1	1.00	1.2358	0.026971	0.23583	68.44
3	bL2	1.00	1.2145	0.004641	0.21451	8.20
4	bs0	0.80	0.8130	0.014545	0.01298	94.67
5	bs1	1.00	0.9790	0.022521	0.02097	93.03
6	r1	-1.50	-1.2144	0.045389	0.28564	69.67
7	r2	-1.00	-0.8912	0.047475	0.10876	93.03
8	s11	0.50	0.6240	0.007852	0.12403	68.03
9	s22	1.00	1.0337	0.056073	0.03368	94.67
10	s33	0.25	0.5162	0.056427	0.26622	75.00
11	se	0.25	0.2576	0.001497	0.00763	93.85

Tables 4.8 and 4.9 below are a sample of the results from when we generated data from a T distribution with varying degrees of freedom and longitudinal censoring with an assumed normal model. We tended to see most of the distortion in parameter estimates, biases, and coverage probabilities when the degrees of freedom were low but as the degrees of freedom increased these estimates tended to move towards the true values and nominal coverage. This is to be expected since as the degrees of freedom increase the T distribution tends towards a normal distribution and we should see results similar to our main proposed model simulations.

Table 4.8. T Distribution with 1df, 20%Censoring, and c=0.5

Obs	parameter	true	Mean Joint Tobit	biastobit	VARJNTTOBIT	tobit coverage
1	bL0	0.00	0.05937	0.05937	0.01243	93.06
2	bL1	1.00	0.9961	0.00394	0.05805	93.98
3	bL2	1.00	0.9783	0.02170	0.01147	94.30
4	bs0	0.80	0.8102	0.01024	0.01381	94.81
5	bs1	1.00	1.0027	0.00271	0.02198	93.69
6	r1	-1.50	-1.4429	0.05707	0.56272	76.62
7	r2	-1.00	-0.8925	0.10746	0.15958	93.26
8	s11	0.50	0.6881	0.18809	0.17316	77.70
9	s22	1.00	1.0497	0.04967	0.17007	93.68
10	s33	0.25	0.3823	0.13228	0.05877	91.90
11	se	0.25	2.4657	2.21571	0.48795	.

Table 4.9. T Distribution with 15df, 20%Censoring, and c=0.5

Obs	parameter	true	Mean Joint Tobit	VARJNTTOBIT	bias Tobit	Tobit Coverage
1	bL0	0.00	-0.00586	0.00459	0.00586	96.00
2	bL1	1.00	1.0705	0.02662	0.07047	93.56
3	bL2	1.00	1.0080	0.00641	0.00800	95.11
4	bs0	0.80	0.8053	0.01108	0.00533	96.44
5	bs1	1.00	1.0257	0.02285	0.02574	94.44
6	r1	-1.50	-1.6079	0.10583	0.10790	95.78
7	r2	-1.00	-0.9705	0.07850	0.02948	98.22
8	s11	0.50	0.4787	0.01190	0.02128	95.33
9	s22	1.00	0.8847	0.09311	0.11533	94.00
10	s33	0.25	0.2898	0.03256	0.03977	97.33
11	se	0.25	0.3044	0.00423	0.05444	86.00

Tables 4.10 and 4.11 below are results from simulations in which we generated data from a model with a diagonal covariance matrix but perturbed the [1, 1], [3, 1], and [6, 6] elements by adding a constant C. This was done to violate the homogenous variance assumption.

We can see that when 3 elements of the covariance matrix differ from the truth by 0.1 (40%) the standard error estimate increases by about 20% and its coverage drops to ~75%. Most other estimates seemed stable except the s33 component which measures the variance of the orthogonal frailty. This seems to follow the general pattern of the proposed model simulation as this component tended to be estimated well when the sample size was large. We see similar results to those already discussed when we longitudinally censor the data at 60% (Table 4.11).

Table 4.10. Normal Distribution with unequal variances, and C=0.1.

Obs	parameter	true	Mean Joint Tobit	VARJNTTOBIT	Bias Tobit	Tobit Coverage
1	bL0	0.00	-0.00297	0.003514	0.002970	94.78
2	bL1	1.00	1.0275	0.016382	0.027525	97.59
3	bL2	1.00	1.0068	0.003766	0.006842	95.18
4	bs0	0.80	0.8181	0.012841	0.018141	93.17
5	bs1	1.00	0.9941	0.018938	0.005897	93.57
6	r1	-1.50	-1.4182	0.060804	0.081841	91.16
7	r2	-1.00	-0.9735	0.029858	0.026527	96.79
8	s11	0.50	0.5345	0.007594	0.034495	91.97
9	s22	1.00	0.9364	0.036332	0.063636	95.98
10	s33	0.25	0.2904	0.032689	0.040412	97.37
11	se	0.25	0.3012	0.001735	0.051150	75.10

Table 4.11. Normal Distribution with unequal variances, C=0.1 and 60% censoring

Obs	parameter	true	Mean Joint Tobit	VARJNTTOBIT	bias Tobit	Tobit Coverage
1	bL0	0.00	-0.01178	0.00550	0.01178	93.95
2	bL1	1.00	1.0698	0.03184	0.06983	92.34
3	bL2	1.00	1.0055	0.00724	0.00546	94.76
4	bs0	0.80	0.8000	0.01228	0.00005	95.97
5	bs1	1.00	1.0069	0.02173	0.00693	93.95
6	r1	-1.50	-1.4787	0.08709	0.02135	94.76
7	r2	-1.00	-0.9768	0.10763	0.02319	95.16
8	s11	0.50	0.5144	0.01136	0.01435	98.39
9	s22	1.00	0.8855	0.09200	0.11455	95.56
10	s33	0.25	0.3030	0.04217	0.05301	96.09
11	se	0.25	0.3251	0.00445	0.07509	77.02

4.7 ILLUSTRATIVE EXAMPLE

The HEMO study [5] was a 2x2 factorial randomized trial performed to assess the relationship between dialysis dose and flux with mortality and morbidity in patients undergoing thrice weekly maintenance dialysis. The study enrolled 1846 subjects, aged 18 to 80, currently undergoing thrice weekly dialysis and randomized them to high or low flux dialyzers along with standard or high dose dialysis. The primary outcome was all-cause mortality. Secondary outcomes included hospitalization rates (not related to vascular access) and cause specific mortality rates.

The data presented here include 1815 subjects with either bimonthly clearance measurements for subjects assigned to high flux dialyzers or six-monthly measurements for subjects assigned to the low flux group. Clearances less than 5 were censored at 5 and clearances greater than 60 were censored at 60. In this population of 1815 subjects, with a total of 19065 measurements, 3908 had clearances below 5 (20.05%), 945 had clearances above 60, (4.96%) and 855 died over the course of the study (47.13%). The aim of this analysis is to: (1) examine the effect of doubly censored longitudinal beta2 clearance values on all-cause mortality adjusting for informative dropout, and (2) examine the consequences of ignoring the censoring.

The joint models considered here will include the basic demographic variables, age, sex, and race, along with the study interventions flux and dialysis dose assignment. In figure 4.12 we can see a tendency to underestimate the parameters in the fill-in method and that ignoring the censoring results in a non-significant association parameter estimate of -0.24, with 95% CI (-0.55, 0.08), and p-value=0.14. This would indicate no apparent association between the longitudinally measured clearance values and all-cause mortality. When we account for the censoring we see a statistically significant association estimate of -0.31, with 95% CI (-0.55, -0.08) and p-value=0.01. Since beta2 clearances were log transformed this would suggest that

increasing clearances lead to better survival. Also since clearance is a function of the dialyzer this suggests improvements in survival may be realized by the use of high flux dialyzers.

Table 4.12. Joint Tobit versus Joint "Fill-in" Models: Effect of Beta2-Clearance on Survival

	Joint Fill-In Model (Ignoring Censoring)						Joint Tobit Model (accounting for censoring)					
	Longitudinal Sub Model			Survival Sub-Model			Longitudinal Sub Model			Survival Sub-Model		
Parameter	Estimate	95%CI	Pvalue	Estimate	95% CI	Pvalue	Estimate	95%CI	Pvalue	Estimate	95% CI	Pvalue
Intercept	1.83	(1.75, 1.90)	<0.001	2.80	(2.48, 3.11)	<0.001	1.4	(1.30, 1.51)	<0.001	2.80	(2.49,3.12	<0.001
Age	-4.00E-05	(-1E-3,1E-3)	0.94	-0.02	(-0.03, -0.02)	<0.001	5.20E-04	(-1E-3,2E-3)	0.51	-0.02	(-0.03, -	<0.001
Sex	0.01	(-0.02. 0.04)	0.57	-0.01	(-0.12.0.10)	0.82	0.02	(-0.02.0.06)	0.39	-0.01	(-0.12	0.83
Race	0.03	(-0.01, 0.06)	0.13	0.24	(0.13, 0.35)	<0.001	0.03	(-0.02,0.07)	0.24	0.24	(0.13,	<0.001
Flux Grp	1.48	(1.45, 1.51)	<0.001	0.06	(-0.05,0.17)	0.27	1.87	(1.83,1.92)	<0.001	0.06	(-0.05,	0.28
Ktv	-0.01	(-0.03 0.001)	0.07	0.03	(-0.02, 0.08)	0.26	-0.02	(-0.04 ,0.003)	0.10	0.03	(-0.02,	0.25
Time	0.02	(-0.02,0.08)	<0.001	*	*	*	0.02	(0.01, 0.03)	<0.001	*	*	*
Σ_{11}	0.17	(0.15,0.19)	<0.001	*	*	*	0.32	(0.28,0.36)	<0.001	*	*	*
Σ_{12}	-0.02	(-0.03,-0.02)	<0.001	*	*	*	-0.05	(-0.06,-0.04)	<0.001	*	*	*
Σ_{22}	0.01	(0.00,0.01)	<0.001	*	*	*	0.01	(0.009,0.001)	<0.001	*	*	*
se	0.23	(0.23,0.24)	<0.001	*	*	*	0.35	(0.34, 0.36)	<0.001	*	*	*
Model Association	*	*	*	-0.24	(-0.55,0.08)	0.14	*	*	*	-0.31	0.55, -0.08)	0.01

4.8 DISCUSSION

Censoring is a common artifact in all biomedical disciplines where measurements are truncated due to the natural limitations of a measuring instrument. Censoring also occur, somewhat unnaturally, when truncated measurements are the result of an investigator imposing restrictions due to some perceived biological implausibility. If one is unable to avoid collecting data that is censored it is clear that appropriate statistical methods be applied to account for such data. Doubly censored data is fairly common in the modeling of biomarkers but less so in the Renal literature. Double censoring was the result of kinetic models producing implausible measurements in the HEMO study and was the motivation for this paper. It is common practice to only use complete data or ignore the censoring by imputing the censored measurement by the limit of detection.

The proposed joint Tobit method is a way of assessing the impact of a longitudinally measured doubly censored covariate on the survival prognoses of a study population. The strength of this approach is that it avoids the complexities of problem specific EM algorithms by turning to commercially available software for the maximization of the joint likelihood. The likelihood is user defined so this Tobit method can be applied to other parametric distributions for the longitudinal and survival components to reflect the problem at hand.

The simulation studies presented indicate an improvement over the standard replacement strategy and the example show how misleading inferences and estimation are dependent on the method used. This method will be extended in a future paper to the multivariate setting where the interest is in the joint effect of doubly censored covariates on survival.

One of the major criticisms of joint modeling is the inherent numerical complexity but with the ever increasing power of personal computers and more efficient maximization algorithms this is rapidly becoming less of an issue. Another common criticism is the interpretation of the linking or association parameters can become muddied when the longitudinal outcome is distributed other than normally. In most instances the linking parameters are interpreted as an association with the event time rather than actual magnitudes of risk.

The "User beware" is a useful adage to describe this authors experience with NLMIXED. We noticed that even though the procedure appears to run and converges, the user still needs to look in the log to see if there were any issues. We noticed that with sample sizes ranging from 100-1000 convergence issues arose in approximately 1-2% of Tobit runs and 3-15% with the fill-in method. We also found that not including starting values for the procedure dramatically affect the stability of the procedure. In some cases the procedure won't even begin to iterate. We found that including reasonable starting values dramatically improves stability and reduces run-time. Since there is no automatic method for choosing starting values we recommend the following procedure. (a) run Proc Mixed with random intercept and slope model and output longitudinal parameter estimates (b) run Proc Lifereg and output event time parameter estimates (c) combine the estimates into a single file (c) use these model based estimates as starting values in Proc NLMIXED.

Joint modeling has long been known to reduce bias and improve precision. The proposed joint Tobit model discussed herein was shown to handle large censoring proportions compared to the fill-in method. We would therefore recommend the use of the joint Tobit method when the interest is in accounting for a censored longitudinally measured covariate and its association with survival where the collection of censored longitudinal data is unavoidable.

5.0 BIVARIATE JOINT MODELING OF CENSORED LONGITUDINAL AND EVENT TIME DATA

5.1 INTRODUCTION

Longitudinally censored data is unfortunately a common occurrence in biomedical studies, especially when one is interested in the prognostic use and evaluation of biomarkers. In a previous paper by Pike and Weissfeld [32] we constructed a joint model which coupled a linear mixed effects Tobit model and an exponential event time distribution. We showed that this 'Joint Tobit model" outperformed a joint model comprised of the more naïve 'fill-in [3] methods which does not model the censoring but replaces data beyond detection limits by a constant function of the detection limits (DL) such as DL, DL/2 or DL/SQRT(2). We found that this was an effective method to account for a doubly censored longitudinal covariate and its effect on survival.

The GENIMS [4] study was no exception to the issues of longitudinal censoring as it was designed to investigate the association between genetic and inflammatory markers and consequent development of community acquired pneumonia and mortality. The assays used to measure these biomarkers were not sensitive enough and resulted in a vast amount of left and some right censored measurements. The study collected data on 2320 subjects from 28 different centers all across the United States. Biomarker data were collected daily for the first seven days and weekly thereafter. Since one hypothesis of interest was whether higher levels of these

biomarkers were associated with higher risk of death it is of utmost importance to account for the heavy censoring in the analysis and outcome dependent dropout. IL10 is a pleiotropic cytokine which is hypothesized to have important anti-inflammatory properties and represses pro-inflammatory cytokines such as TNF-ALPHA (Tumor Necrosis Factor-Alpha), IL6 (Interleukin-6), and IL-1 (Interleukin-1). Since there is a biological connection between these cytokines it would appear that they affect the immune system in a complementary manner and therefore it makes sense to model their effects on the disease process and survival in a multivariate or joint manner. Any analyses must therefore be able to account for (i) correlated biomarkers, (ii) outcome dependent dropout, and (iii) longitudinal censoring.

Laird and Ware [9] mixed effects models are effective in modeling the longitudinal nature of biomarker data and are a common tool for this type of data as they can account for all available information. Carlin et al [26] effectively coupled a Laird and Ware model to a parametric survival distribution by assuming that the hazard was a function of random effects common to both longitudinal and survival sub-models [21]. They then showed that the resulting joint likelihood could be easily maximized in SAS' Proc NLMIXED [33] thus avoiding the complexity and problem specific nature of the expectation maximization (EM) algorithm [8].

In this paper we extended Carlin's approach in NLMIXED to the bivariate case with censored longitudinal data in both outcomes by assuming that the hazard is a function of both longitudinal trajectories. We linked a bivariate linear mixed effects Tobit [14] model to an exponential survival model by a set of shared random effects whose covariance matrix accounts for correlations within and between outcomes. In section 2 we will present the notation and methods for the presentation of the bivariate joint Tobit model. In section 3 we discuss the

simulation studies and results. We give an example based on data collected as part of the GENIMS study described previously in section 4 and close our discussion in section 5.

5.2 NOTATION AND METHODS

5.2.1 Longitudinal Model

The general notation for a bivariate linear mixed model, excluding covariates, which includes a random intercept, slope, and independent error, is as follows from Thiebaut et al [10]:

Let $Y_i = \begin{bmatrix} Y_i^1 & Y_i^2 \end{bmatrix}^T$ be the response vector for subject i=1...N, with Y_i^k the n_i^k vector of repeated measurements on the outcome k (k=1, 2) with $n_i^1 = n_i^2 = n_i$ where the number of measurements may be different between markers and within subjects.

$$Y_i = X_i\beta + Z_i\gamma_i + \varepsilon_i \text{ with } \begin{cases} \varepsilon_i \sim N(0, \Sigma_i) \\ \gamma_i \sim N(0, G) \end{cases},$$

and

$$\beta = \begin{bmatrix} \beta^1 \\ \beta^2 \end{bmatrix},\ X_i = \begin{bmatrix} X_i^1 & 0 \\ 0 & X_i^2 \end{bmatrix},\ \gamma_i = \begin{bmatrix} \gamma_i^1 \\ \gamma_i^2 \end{bmatrix},\ Z_i = \begin{bmatrix} Z_i^1 & 0 \\ 0 & Z_i^2 \end{bmatrix},\ \varepsilon_i = \begin{bmatrix} \varepsilon_i^1 \\ \varepsilon_i^2 \end{bmatrix}.$$

X_i^k is an n_i x p^k design matrix of covariates for maker k, β^k is a p^k-vector of fixed effects, Z_i^k is an n_i x q^k design matrix and is usually a subset of X_i^k, γ_i^k is a q^k-vector of subject specific random effects with $q^k \leq p^k$. The covariance matrix of measurement errors is a diagonal matrix containing $\sigma_{\varepsilon^1}^2$ and $\sigma_{\varepsilon 2}^2$ denoted by Σ_i and represents the measurement error for each marker.

The covariance matrix of random effects is the matrix $G = \begin{bmatrix} G^1 & G^{12} \\ G^{21} & G^2 \end{bmatrix}$. The elements of G are

G^1, the covariance matrix of random effects for the first marker, G^2, the covariance matrix of random effects for the second marker, and $G^{21}=G^{12}$, the matrix of covariance's between the random effects from each marker. It is through G where the correlation between the two markers is accounted for. If $G_{21}=G_{12}=\mathbf{0}$ then the two markers are assumed to be independent.

5.2.2 Event Time Distribution

In order to link the longitudinal and survival sub-models we assume the hazard to be a function of both of the individual longitudinal trajectories and as such share the same set of random effects. For the bivariate case the hazard can be written as

$\lambda_i(t) = \lambda_0 \exp\left(X_{2i}^T(t)\beta_2 + r_1 a_i^1 + r_2 b_i^1 + r_3 a_i^2 + r_4 b_i^2 \right)$ where $\gamma_i^k = \begin{bmatrix} a_i^k \\ b_i^k \end{bmatrix}$ *k=1, 2*, are random

intercepts and slopes for the individual biomarkers and $X_i^T(t)\beta_2$ are possibly time dependent covariates and their respective regression coefficients. The baseline hazard can be any parametric form but for simplicity, in this paper, we assume the baseline hazard to be exponential. The random effects need not be linear and should ideally be chosen to suit the problem at hand In Henderson's and Carlin's papers [5][6] they created a univariate joint model and chose to link the univariate longitudinal and event time sub models by random intercepts, slopes, fitted longitudinal value at the death time, and an added frailty orthogonal to the random effects (*a, b*) as $\lambda_i(t) = \lambda \exp\left(X_{2i}^T(t)\beta_2 + r_1 a + r_2 b + r_3 (a + bt_i) + u_i \right)$.

We extended Carlin's methodology to the longitudinally censored bivariate case in NLMIXED by including parameters in the hazard that capture the association due to adding random intercepts, slopes, and fitted longitudinal trajectory at the event time, from both outcome trajectories. The full complex hazard linking the sub-models can be written as:

$$\lambda_i(t) = \lambda \exp\left(X_{2i}^T(t)\beta_2 + r_1 a_i^1 + r_2 b_i^1 + r_3 a_i^2 + r_4 b_i^2 + r_5\left(a_i^1 + b_i^1 t_i\right) + r_6\left(a_i^2 + b_i^2 t_i\right) + u_i \right)$$

The joint likelihood $L=L\,(Y, N; \theta)$ can be constructed by assuming that the longitudinal and event time sub-distributions are independent given a set of shared random effects γ_i. The joint likelihood can then be factored into a product of conditionally independent distributions as follows:

$$L(Y, N; \theta) = \prod_{i=1}^{N} \int_{\gamma_i} \prod_{k=1}^{2} L(Y_{ik} \mid \gamma_i; \theta) L(N \mid \gamma_i; \theta) L(\gamma_i; \theta)\, d\gamma_i \quad . \tag{5.1}$$

The first components of (5.1) are the conditional distributions of the longitudinal data, the survival data, and the random effects respectively. We can envisage the vector Y to contain censored and uncensored data respectively and denote this by Y* and then rewrite the longitudinal component in the joint likelihood in (1) as:

$$\prod_{k=1}^{2} \left\{ \prod_{j \in n_d} f\left(y_{ijk} \mid \mathbf{u_i}, \sigma_e^2\right) \prod_{j \in n_{d0}} \Phi\left(y_{ijk}^*\right) \left(\prod_{j \in n_{d1}} \left(1 - \Phi\left(y_{ijk}^*\right)\right) \right) \right\}, \tag{5.2}$$

where the contributions to the likelihood are from left, right, and uncensored values respectively. The components of the likelihood in (5.2) accounts for, left censoring by use of the cumulative distribution (CDF), right censoring by use of 1-CDF, and uncensored values from the probability distribution function (pdf).

The Survival component of (5.1) is denoted by:

$$f\left(t_i,\Delta_i \mid a_i,b_i,\lambda_0,\beta\right)=\left[\lambda_0\left(t_i\right)\exp\left\{X_{2i}^T\left(t\right)\beta_2+r_1a_i^1+r_2b_i^1+r_3a_i^2+r_4b_i^2+r_5\left(a_i^1+b_i^1t_i\right)+r_6\left(a_i^2+b_i^2t_i\right)+u_i\right\}\right]^{\Delta_i}$$
$$*\exp\left[-\int_0^{t_i}\lambda_0\exp\left\{X_{2i}^T\left(t\right)\beta_2+r_1a_i^1+r_2b_i^1+r_3a_i^2+r_4b_i^2+r_5\left(a_i^1+b_i^1t_i\right)+r_6\left(a_i^2+b_i^2t_i\right)+u_i\right\}dt_i\right] \quad . \quad (5.3)$$

where $\lambda_0\left(t_i\right)$ is the baseline hazard, a_i and b_i are random effects linking the two distributions, and $X_i^T\left(t\right)\beta$ are possibly time dependent covariates and their respective regression coefficients. The distribution of the random effects is MVN(0,G) and can be denoted by:

$$f\left(\gamma_i \mid \boldsymbol{\gamma},\mathbf{G}\right)=\left(2\pi\mid\mathbf{G}\mid\right)^{-1/2}\exp\left\{-\left(\boldsymbol{\gamma}_\mathbf{i}-\boldsymbol{\mu}\right)^T\mathbf{G}^{-1}\left(\boldsymbol{\gamma}_\mathbf{i}-\boldsymbol{\mu}\right)/2\right\} . \quad (5.4)$$

In this paper we will focus on a bivariate joint model with longitudinal sub-models conditional on the random effects as:

$$\begin{cases} E\left(Y_{ij}^1 \mid a_i^1,b_i^1\right)=\left(\alpha^1+a_i^1\right)+\left(\beta^1+b_i^1\right)t_{ij}^1 \\ E\left(Y_{ij}^2 \mid a_i^2,b_i^2\right)=\left(\alpha^2+a_i^2\right)+\left(\beta^2+b_i^2\right)t_{ij}^2 \end{cases}, \quad (5.5)$$

and sub hazard:

$$\lambda_i\left(t\right)=\lambda\exp\left(X_i^T\left(t\right)+r_1a_i^1+r_2b_i^1+r_3a_i^2+r_4b_i^2\right). \quad (5.6)$$

The components of the sub-hazard (r1, r2, r3, r4) in (5.6) account for association between the longitudinal and survival models due to inclusion of random intercepts and slopes in both outcome trajectories. When association exists between the sub-models the use of a joint model allows one to obtain less biased estimates and more efficient inferences. The joint likelihood in (5.1) can be passed to a suitable non-linear optimizer but we chose to turn to SAS PROC NLMIXED as it afforded a great deal of flexibility and cutting edge optimization algorithms with a plethora of fine tuning options to aid in convergence.

5.3 SIMULATION OUTLINE

As mentioned previously we generated data from a longitudinal sub-model with random and fixed intercept (***a_i*** **,** ***α***), random and fixed slope (***bi, β***) , and covariance matrix of random effects ***G*** .The survival data was generated from an exponential survival distribution as in (5.6).

In order to use NLMIXED, and simplify the coding, the dataset needs to be created in such a way as to stack the outcome vectorsY1 and Y2 into a single outcome vector Y. The basic idea is to distinguish between the two outcomes via an indicator variable in the dataset. We also need to create a longitudinal censoring indicator which distinguishes between outcomes censored from, below, above, and uncensored values. This indicator is not to be confused with the survival censoring indicator. The specifics for creating the dataset can be found in [34].

The longitudinal data will be generated from:

$$\begin{cases} E\left(Y_{ij}^1 \mid a_i^1, b_i^1\right) = \left(\alpha^1 + a_i^1\right) + \left(\beta^1 + b_i^1\right) t_{ij}^1 : L1 \\ E\left(Y_{ij}^2 \mid a_i^2, b_i^2\right) = \left(\alpha^2 + a_i^2\right) + \left(\beta^2 + b_i^2\right) t_{ij}^2 : L2 \end{cases}$$

With true values:

$$\beta^1 = \begin{bmatrix} 6 \\ -0.3 \end{bmatrix},\ \beta^2 = \begin{bmatrix} 23 \\ -0.01 \end{bmatrix}$$

$$\gamma_i \sim MVN\left(0, \begin{bmatrix} G_1 & G_{12} \\ G_{12} & G_2 \end{bmatrix}\right) = MVN\left(0, \begin{bmatrix} \begin{bmatrix} \sigma_{11}^2 & \rho\sigma_{22}\sigma_{11} \\ \rho\sigma_{22}\sigma_{11} & \sigma_{22}^2 \end{bmatrix} & \begin{bmatrix} \rho\sigma_{11}\sigma_{33} & \rho\sigma_{11}\sigma_{44} \\ \rho\sigma_{22}\sigma_{33} & \rho\sigma_{22}\sigma_{44} \end{bmatrix} \\ \begin{bmatrix} \rho\sigma_{33}\sigma_{11} & \rho\sigma_{33}\sigma_{22} \\ \rho\sigma_{44}\sigma_{11} & \rho\sigma_{44}\sigma_{22} \end{bmatrix} & \begin{bmatrix} \sigma_{33}^2 & \rho\sigma_{33}\sigma_{44} \\ \rho\sigma_{44}\sigma_{33} & \sigma_{44}^2 \end{bmatrix} \end{bmatrix}\right)$$

$$\begin{bmatrix} \sigma_{11}^2 \\ \sigma_{22}^2 \\ \sigma_{33}^2 \\ \sigma_{44}^2 \end{bmatrix} = \begin{bmatrix} 0.7 \\ 0.03 \\ 1.3 \\ 0.3 \end{bmatrix},\ \rho = \left(0, -0.1, -0.3\right),\ \begin{bmatrix} \varepsilon_i^1 \\ \varepsilon_i^2 \end{bmatrix} = \begin{bmatrix} 0.7 \\ 1.1 \end{bmatrix},\quad and\ t_{ij}^k \sim UNIF\left(0, 2\right)$$

The event time data will be generated from the hazard:

$$\lambda_i(t) = \lambda \exp\left(X_i^T(t)\beta_2 + r_1 a_i^1 + r_2 b_i^1 + r_3 a_i^2 + r_4 b_i^2 \right)$$

With true values:

$$\beta_2 = -0.05,\ \lambda_0 = 0.8, \lambda_c = 0.4 \ \ and \begin{bmatrix} r1 \\ r2 \\ r3 \\ r4 \end{bmatrix} = \begin{bmatrix} -0.6 \\ -2.5 \\ -1.7 \\ -0.9 \end{bmatrix}$$

We ran a series of simulations to determine the models adequacy and stability. We first determined if the model returned the true values when the two outcomes were independent. This meant generating data after setting the covariance ρ to zero. We then added longitudinal censoring at 20% and 60% and reported the results in Table 5.1. In the next series of simulations we generated data with ρ set to -0.1 and -0.3 which induces correlation within and between the two markers. We then added longitudinal censoring at 20% and 60% and reported the results in Table 5.2 and Table 5.3.

5.4 SIMULATION RESULTS

The results in Table 5.1 are from generating independent bivariate data by setting ρ=0. This creates a covariance matrix ***G*** of random effects with diagonal sub-matrices $G_1 = \begin{bmatrix} 0.7 & 0 \\ 0 & 0.03 \end{bmatrix}$, $G_2 = \begin{bmatrix} 1.3 & 0 \\ 0 & 0.3 \end{bmatrix}$, and $G_{12} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. It is the G_{12} matrix which accounts for the correlated outcomes so setting all elements to zero implies independence. The independence model returns unbiased estimates with small variance and nominal coverage regardless of the proportion of

censoring. This is consistent with previous results in the univariate Tobit joint model by Pike and Weissfeld (2011) [32]. The large variance for r2 was due to a single extreme positive estimate but since the gradient was small enough to indicate a stationary point it was left in the final dataset. This is interesting since we noticed that the impact of such outliers seems to worsen as the percentage of censoring and degree of correlation increases. Since this behavior is due to an outlying estimate in one or more generated datasets leads us to question whether it's due to some anomaly in the generated datasets or inaccuracies in the estimation procedure due to the Laplace approximation. This behavior may be an avenue for future research to see how such models behave in the presence of extreme outliers with varying number of quadrature points.

The results in Tables 5.2 and 5.3 are from setting ρ= -0.1 and -0.3 respectively. This induces mild and moderate negative correlation within and between the random effects. When we induced correlations with and between random effects we see that, on the whole, the estimates are still unbiased but the coverage's are inflated for isolated estimates. This may be due in part to fitting the models with a Laplace approximation for purposes of efficiency. In the case of ρ=-0.3 and 60 percent censoring we were only to get approximately 37% of the runs to converge which may be the reason for the inflated coverage's and slight bias in some of the estimates.

Table 5.1. Bivariate simulations with ρ=0 and variable censoring rates

Bivariate Joint Tobit and Event Time Model Simulation Results
Independent Outcomes
250 Iterations and sample size 500

		No Longitudinal Censoring				**20 % Longitudinal Censoring**				**60 % Longitudinal Censoring**			
Parameter	TRUE	Mean	Variance	Bias	95% Coverage	Mean	Variance	Bias	95% Coverage	Mean	Variance	Bias	95% Coverage
β_{L01}	**6**	6.002	0.003	0.002	94.38	6.003	0.003	0.003	94.80	6.003	0.004	0.003	96.40
β_{L02}	**23**	22.999	0.006	0.001	95.18	22.998	0.006	0.002	96.40	22.994	0.007	0.006	93.60
β_{L11}	**-0.3**	-0.301	0.000	0.001	97.19	-0.301	0.000	0.001	96.40	-0.301	0.000	0.001	96.00
β_{L12}	**-0.1**	-0.099	0.001	0.001	94.38	-0.098	0.001	0.002	96.00	-0.101	0.001	0.001	94.80
λ_0	**0.8**	0.838	0.090	0.038	93.57	0.837	0.090	0.037	94.00	0.833	0.093	0.033	94.40
β_{21}	**-0.05**	-0.051	0.000	0.001	95.18	-0.051	0.000	0.001	95.20	-0.051	0.000	0.001	95.60
r1	**-0.6**	-0.619	0.040	0.019	94.38	-0.610	0.042	0.010	94.40	-0.605	0.058	0.005	92.40
r2	**-2.5**	-2.444	1.725	0.056	94.38	-2.566	2.123	0.066	93.20	-2.574	20.233	0.074	94.80
r3	**-1.7**	-1.763	0.022	0.063	95.18	-1.766	0.023	0.066	96.00	-1.808	0.045	0.108	94.40
r4	**-0.9**	-0.870	0.046	0.030	95.58	-0.862	0.047	0.038	96.80	-1.043	0.060	0.143	90.40
g11	**0.7**	0.704	0.006	0.004	93.98	0.706	0.007	0.006	96.40	0.700	0.012	0.000	92.00
g22	**0.03**	0.030	0.000	0.000	93.98	0.029	0.000	0.001	94.00	0.028	0.000	0.002	92.00
g33	**1.3**	1.269	0.026	0.031	93.17	1.266	0.027	0.034	94.40	1.168	0.051	0.132	85.20
g44	**0.3**	0.300	0.001	0.000	95.58	0.299	0.001	0.001	94.40	0.321	0.003	0.021	95.20
se1	**0.7**	0.701	0.001	0.001	95.18	0.702	0.001	0.002	96.40	0.707	0.002	0.007	96.00
se2	**1.1**	1.103	0.002	0.003	95.98	1.104	0.002	0.004	96.00	1.105	0.005	0.005	96.00

Table 5.2. Bivariate simulations with ρ= -0.1 and variable censoring rates

Bivariate Joint Tobit and Event Time Model Simulation Results
Correlated Outcomes ρ= -0.10
250 Iterations, sample size 500,

		No Longitudinal Censoring				**20 % Longitudinal Censoring**				**60 % Longitudinal Censoring**			
Parameter	TRUE	Mean	Variance	Bias	95% Coverage	Mean	Variance	Bias	95% Coverage	Mean	Variance	Bias	95% Coverage
β_{L01}	**6**	6.001	0.003	0.001	95.14	6.002	0.003	0.002	95.95	6.002	0.004	0.002	97.02
β_{L02}	**23**	23.001	0.005	0.001	95.14	23.002	0.006	0.002	95.14	22.999	0.007	0.001	94.02
β_{L11}	**-0.3**	-0.301	0.000	0.001	97.17	-0.301	0.000	0.001	96.76	-0.301	0.000	0.001	95.28
β_{L12}	**-0.1**	-0.099	0.001	0.001	94.74	-0.100	0.001	0.000	95.14	-0.102	0.001	0.002	95.71
λ_0	**0.8**	0.835	0.091	0.035	95.55	0.834	0.091	0.034	95.55	0.831	0.100	0.031	94.85
β_{21}	**-0.05**	-0.051	0.000	0.001	94.98	-0.051	0.000	0.001	95.36	-0.051	0.000	0.001	95.35
r1	**-0.6**	-0.632	0.107	0.032	96.77	-0.634	0.135	0.034	97.98	-0.626	0.178	0.026	99.57
r2	**-2.5**	-2.594	6.992	0.094	93.95	-2.589	7.424	0.089	95.56	-2.934	12.411	0.434	97.88
r3	**-1.7**	-1.792	0.028	0.092	95.14	-1.804	0.032	0.104	95.95	-1.881	0.047	0.181	96.14
r4	**-0.9**	-0.861	0.079	0.039	96.76	-0.850	0.102	0.050	97.58	-0.917	0.195	0.017	94.92
g11	**0.7**	0.708	0.014	0.008	93.12	0.700	0.016	0.000	94.33	0.666	0.025	0.034	93.13
g21	**-0.014**	-0.016	0.001	0.002	94.26	0.299	0.001	0.001	94.76	-0.007	0.001	0.007	93.42
g22	**0.03**	0.030	0.000	0.000	93.42	-0.014	0.001	0.001	93.42	0.028	0.000	0.002	94.71
g31	**-0.095**	-0.105	0.010	0.010	95.55	0.029	0.000	0.001	94.71	-0.102	0.012	0.007	96.57
g32	**-0.020**	-0.018	0.001	0.001	97.13	-0.107	0.010	0.011	96.34	-0.016	0.001	0.004	96.89
g33	**1.3**	1.285	0.035	0.015	94.33	-0.018	0.001	0.002	96.30	1.171	0.054	0.129	88.94
g41	**-0.046**	-0.043	0.002	0.002	94.74	1.268	0.040	0.032	92.31	-0.048	0.003	0.003	93.56
g42	**-0.009**	-0.009	0.000	0.001	95.44	-0.043	0.002	0.003	92.71	-0.010	0.000	0.000	96.04
g43	**-0.062**	-0.068	0.003	0.005	95.14	-0.009	0.000	0.000	95.49	-0.041	0.005	0.022	93.13
g44	**0.3**	0.301	0.001	0.001	94.76	-0.060	0.004	0.002	94.33	0.306	0.003	0.006	93.59
se1	**0.7**	0.700	0.001	0.000	95.16	0.703	0.001	0.003	95.97	0.711	0.002	0.011	95.34
se2	**1.1**	1.101	0.002	0.001	98.38	1.104	0.002	0.004	98.38	1.109	0.005	0.009	96.17

Table 5.3. Bivariate simulation with ρ= -0.3 and variable censoring rates

Bivariate Joint Tobit and Event Time Model Simulation Results
Correlated Outcomes ρ= -0.3
250 Iterations, sample size 500,

		No Longitudinal Censoring				20 % Longitudinal Censoring				60 % Longitudinal Censoring			
Parameter	TRUE	Mean	Variance	Bias	95% Coverage	Mean	Variance	Bias	95% Coverage	Mean	Variance	Bias	95% Coverage
β_{L01}	**6**	6.001	0.003	0.001	95.04	6.003	0.003	0.003	94.34	6.003	0.005	0.003	94.57
β_{L02}	**23**	23.001	0.005	0.001	96.28	22.999	0.005	0.001	96.21	22.993	0.006	0.007	96.67
β_{L11}	**-0.3**	-0.301	0.000	0.001	97.52	-0.301	0.000	0.001	98.58	-0.302	0.000	0.002	95.56
β_{L12}	**-0.1**	-0.098	0.001	0.002	95.45	-0.098	0.001	0.002	95.26	-0.096	0.001	0.004	97.80
λ_0	**0.8**	0.819	0.072	0.019	96.28	0.823	0.075	0.023	95.73	0.832	0.084	0.032	95.56
β_{21}	**-0.05**	-0.052	0.000	0.002	96.93	-0.052	0.000	0.002	95.54	-0.052	0.000	0.002	93.83
r1	**-0.6**	-0.818	0.218	0.218	97.11	-0.834	0.283	0.234	97.64	-0.887	0.345	0.287	97.83
r2	**-2.5**	-3.585	17.883	1.085	96.69	-3.977	54.116	1.477	96.7	-4.665	31.876	2.165	97.83
r3	**-1.7**	-1.957	0.224	0.257	96.69	-1.960	0.174	0.260	96.68	-2.085	0.269	0.385	94.44
r4	**-0.9**	-1.128	0.725	0.228	97.52	-1.155	1.838	0.255	97.64	-1.200	0.884	0.300	96.67
g11	**0.7**	0.708	0.013	0.008	94.21	0.703	0.015	0.003	94.29	0.673	0.024	0.027	93.26
g21	**-0.014**	-0.046	0.001	0.002	93.89	-0.045	0.001	0.002	94.55	-0.045	0.001	0.001	93.98
g22	**0.03**	0.031	0.000	0.001	96.45	0.031	0.000	0.001	96.24	0.033	0.000	0.003	97.47
g31	**-0.095**	-0.298	0.009	0.012	95.00	-0.299	0.011	0.013	93.27	-0.256	0.012	0.030	97.70
g32	**-0.020**	-0.057	0.001	0.002	96.96	-0.058	0.001	0.001	97.52	-0.068	0.001	0.009	95.18
g33	**1.3**	1.283	0.033	0.017	93.78	1.271	0.039	0.029	91.39	1.224	0.041	0.076	90.91
g41	**-0.046**	-0.138	0.002	0.000	94.51	-0.136	0.002	0.002	96.63	-0.151	0.003	0.013	94.25
g42	**-0.009**	-0.028	0.000	0.001	92.63	-0.028	0.000	0.000	95.05	-0.027	0.000	0.002	96.47
g43	**-0.062**	-0.186	0.004	0.002	93.70	-0.178	0.005	0.009	90.38	-0.167	0.006	0.021	93.10
g44	**0.3**	0.300	0.001	0.000	93.31	0.297	0.001	0.003	93.3	0.299	0.003	0.001	93.33
se1	**0.7**	0.700	0.001	0.000	95.04	0.703	0.001	0.003	95.28	0.710	0.002	0.010	100.00
se2	**1.1**	1.104	0.002	0.004	97.52	1.104	0.002	0.004	95.73	1.119	0.007	0.019	90.00

Table 5.4. Bivariate simulation with ρ= 0.5 and variable censoring rates

Bivariate Joint Tobit and Event Time Model Simulation Results
Correlated Outcomes ρ= 0.5
250 Iterations, sample size 500,

	No Longitudinal Censoring					20 % Longitudinal Censoring				60 % Longitudinal Censoring			
Parameter	TRUE	Mean	Variance	Bias	95% Coverage	Mean	Variance	Bias	95% Coverage	Mean	Variance	Bias	95% Coverage
β_{L01}	**6**	6.001	0.003	0.001	94.55	6.003	0.003	0.003	96.33	6.005	0.004	0.005	95.38
β_{L02}	**23**	23.000	0.006	0.000	95.45	22.997	0.006	0.003	97.24	22.993	0.007	0.007	96.53
β_{L11}	**-0.3**	-0.301	0.000	0.001	94.55	-0.302	0.000	0.002	95.37	-0.302	0.000	0.002	97.06
β_{L12}	**-0.1**	-0.101	0.001	0.001	93.18	-0.101	0.001	0.001	93.98	-0.097	0.001	0.003	95.88
λ_0	**0.8**	0.865	0.092	0.065	95.45	0.853	0.090	0.053	94.01	0.863	0.099	0.063	95.38
β_{21}	**-0.05**	-0.052	0.000	0.002	95.90	-0.051	0.000	0.001	94.65	-0.051	0.000	0.001	96.69
r1	**-0.6**	-0.664	0.351	0.064	95.02	-0.646	0.515	0.046	96.80	-0.623	0.459	0.023	98.27
r2	**-2.5**	-2.290	12.614	0.210	98.20	-2.397	15.169	0.103	99.09	-2.553	14.869	0.053	98.84
r3	**-1.7**	-1.801	0.079	0.101	96.36	-1.801	0.099	0.101	97.70	-1.831	0.142	0.131	98.26
r4	**-0.9**	-0.776	0.226	0.124	97.29	-0.740	0.273	0.160	98.17	-0.844	0.402	0.056	96.53
g11	**0.7**	0.726	0.011	0.026	96.79	0.728	0.014	0.028	96.63	0.722	0.025	0.022	97.62
g21	**0.072**	0.063	0.000	0.009	95.52	0.063	0.001	0.009	95.90	0.058	0.001	0.014	96.82
g22	**0.03**	0.033	0.000	0.003	94.21	0.032	0.000	0.002	95.43	0.034	0.000	0.004	97.22
g31	**0.477**	0.467	0.010	0.010	97.24	0.467	0.011	0.010	96.17	0.434	0.014	0.043	92.26
g32	**0.099**	0.099	0.001	0.000	93.97	0.099	0.001	0.000	95.96	0.094	0.001	0.005	95.57
g33	**1.3**	1.296	0.038	0.004	93.67	1.302	0.043	0.002	96.23	1.210	0.062	0.090	92.40
g41	**0.229**	0.235	0.002	0.005	94.34	0.234	0.002	0.005	95.15	0.242	0.003	0.013	96.47
g42	**0.047**	0.048	0.000	0.000	96.43	0.047	0.000	0.000	98.42	0.050	0.000	0.003	94.08
g43	**0.312**	0.304	0.003	0.008	94.88	0.303	0.003	0.009	96.14	0.296	0.005	0.016	94.67
g44	**0.3**	0.304	0.001	0.004	95.77	0.304	0.001	0.004	96.63	0.333	0.004	0.033	90.59
se1	**0.7**	0.696	0.001	0.004	95.02	0.698	0.001	0.002	96.35	0.700	0.002	0.000	97.11
se2	**1.1**	1.098	0.002	0.002	97.27	1.096	0.002	0.004	96.33	1.090	0.005	0.010	91.91

5.5 ILLUSTRATIVE EXAMPLE

The GENIMS study was perceived to investigate the relationship between genetic and inflammatory markers on sepsis after a diagnosis of community acquired pneumonia. The study collected data on 2320 subjects from 28 US based centers. Measurements on several cytokines were taken daily for the first week then every subsequent week thereafter. One of the hypotheses was whether higher values of cytokines lead to an increased risk of death. The data considered herein consists of 8245 measurements on anti-inflammatory and pro-inflammatory markers Interleukin-6 (IL6) and Interleukin-10 (IL10), Age, Sex and Charlson co-morbidity index which gives a patient a score to reflect whether they had preexisting co-morbid conditions prior to hospitalization. The cytokines were measured using a chemiluminescent immunoassay using an automated analyzer. The lower limit of detection for this analyzer was 2 and 5 for IL6 and 5 for IL10. In this longitudinal dataset there were 2224 (27.2 percent) and 5837 (70.79 percent) measurements below the detectable limit of 2 and 5 for IL6 and IL10 respectively. The censoring tended to increase over the study period from 12 percent to 35percent for IL-6 and from 46 percent to 76 percent for IL-10. Within this dataset 211 (11.6 percent) died within 90days and 87 people died with both IL6 and IL10 measurements above the lower limit of detection (LOD).

The following analysis was run to investigate whether the bivariate longitudinal IL-6 and IL-10 measurements are associated with survival. One of the hypotheses of interest was whether high initial levels of both of these biomarkers that tend to increase over time are associated with a poorer survival prognosis. To test this hypothesis we constructed a bivariate joint Tobit model by linking a bivariate linear mixed effects Tobit model to a Weibull survival distribution. This linkage was accomplished by assuming that the longitudinal trajectories and Weibull hazard were both functions of the same set of random effects. In this particular case we assumed that the Weibull hazard was a function of the random intercepts and slopes from both of the longitudinal trajectories. The choice of the Weibull distribution is arbitrary but was chosen because of the flexible form of the hazard. If we look at figure 5.5 (table 4) we can see all of the association parameters r1, r2, r3, and r4 are statistically significant and negative which implies that, on average, patients with high initial values of both markers that tend to increase over time are associated with an increased hazard of death. One question worth noting is what is to be gained from fitting a bivariate model. This can be answered by focusing our attention on the parameters of the random effects covariance matrix $G = \begin{bmatrix} G^{1}_{2x2} & G^{12}_{2x2} \\ G^{21}_{2x2} & G^{2}_{2x2} \end{bmatrix}$.We can see that the elements of G^{21}_{2x2} are all significantly different from zero which implies that modeling these biomarkers jointly, with correlation, was justified.

Table 5.5. Bivariate Joint Tobit Model of IL6 and IL10 and association with 90 day Survival

Bivariate Joint Tobit abd Event Time Model
IL6 IL10 and 90 Day Survival

Parameter	Estimate	SE	t Value	Pr > |t|	Parameter	Estimate	SE	t Value	Pr > |t|
IL6					**IL10**				
β_{Int}	3.278	0.164	19.988	<0.001	β_{Int}	2.187	0.055	39.498	<0.001
β_{Time}	-0.453	0.014	-31.583	<0.001	β_{Time}	-0.097	0.005	-20.747	<0.001
β_{Age}	0.009	0.002	4.222	<0.001	β_{Age}	0.001	0.001	0.866	0.387
$\beta_{Charlson}$	-0.179	0.083	-2.160	0.031	$\beta_{Charlson}$	-0.030	0.027	-1.111	0.267
β_{Sex}	0.308	0.072	4.275	<0.001	β_{Sex}	0.022	0.023	0.931	0.352
se1	0.642	0.016	40.806	<0.001	**se2**	0.169	0.003	50.359	<0.001
Survival Model					**Random Effects Covariance Estimates**				
β_{Int}	26.303	1.795	14.653	<0.001	Σ_{11}	4.124	0.182	22.670	<0.001
β_{Age}	-0.148	0.019	-7.977	<0.001	Σ_{21}	-0.636	0.040	-15.864	<0.001
$\beta_{Charlson}$	-1.788	0.573	-3.123	0.002	Σ_{22}	0.202	0.012	16.472	<0.001
β_{Sex}	-1.106	0.417	-2.656	0.008	Σ_{31}	0.792	0.055	14.417	<0.001
gamma	0.362	0.023	15.463	<0.001	Σ_{32}	-0.158	0.014	-11.623	<0.001
Association Parameters					Σ_{33}	0.739	0.031	23.729	<0.001
r1	-1.405	0.169	-8.300	<0.001	Σ_{41}	-0.158	0.011	-13.836	<0.001
r2	-9.273	1.555	-5.961	<0.001	Σ_{42}	0.042	0.003	14.043	<0.001
r3	-1.649	0.534	-3.088	0.002	Σ_{43}	-0.113	0.006	-18.963	<0.001
r4	-7.386	5.449	-1.356	0.018	Σ_{44}	0.023	0.001	16.458	<0.001

5.6 DISCUSSION

Censoring is unfortunately a common artifact when one is interested in the evaluation of a biomarker. The analytical scenario in which biomarkers are evaluated can become quite complex and may include issues such as (a) longitudinal censoring, due to the sensitivity of the assay used, (b) outcome dependent dropout, and (c) correlation between biomarkers. Therefore it is readily apparent that the chosen analytical procedure must have the potential to simultaneously account for all of the aforementioned issues depending upon the complexity of the scenario encountered.

All of the aforementioned issues were encountered in the GENIMS study [4] and was the primary motivation for this paper. It is common practice to discard all censored data (complete case analyses) or impute the censored values via some function of the detection limit (DL, DL/2. DL /√2). The proposed bivariate joint Tobit method is a way of assessing the joint impact of two correlated and longitudinally censored biomarkers on the survival prognoses of a study population. The beauty and strength of this analytical framework is that censoring, outcome dependent dropout, and correlation is handled in one unified model and is easily implemented in commercially available software.

Even though, in this paper, we have discussed the bivariate case the proposed method is of course readily extendable to the multivariate case in which there are multiple biomarker trajectories. We evaluated the bivariate case which leads to four random effects and thus a 4x4 covariance matrix of random effects with ten parameters to be estimated. We can see that as more and more markers are added these models quickly become unwieldy and the number of

parameters to be estimated increases almost exponentially. We would therefore recommend restricting the use of these models to the bivariate case until the case of n>2 biomarkers are thoroughly evaluated. The proposed bivariate joint Tobit likelihood is user defined so this method can be applied to other parametric distributions for the longitudinal and survival components to reflect the problem at hand. An interesting sub-model would be where the survival time was unavailable. In this scenario one would merely replace the survival distribution by a logistic distribution and link the logistic distribution to the longitudinal sub-model by a set of shared random effects as follows:

$$\log\left(\pi_i \Big/ 1-\pi_i\right) = \beta_0 + \beta_1 t_{ij} + a_{1i} + b_{1i} t_{ij}, \qquad (7)$$

where the probability of death π is modeled. A question worth noting here is whether information with respect to time is ever really absent. Even though we don't have the event time we do know that the event happened after the last longitudinal measurement. So can we gain efficiency by including what we know about time rather than exclude it altogether? This is an interesting question and will be pursued by the authors in a subsequent paper.

The simulation studies presented indicate that in general these models perform adequately with 4 random effects. It is also interesting to note that as the percentage of censoring increases the more impact outlying values have on the estimation of the variance of the parameters. This is not unexpected since the imputation of a large number of values with a fixed constant will underestimate the variance so even a single outlying value can have dramatic effects. We also noticed that when the covariance was increased to ρ= -0.3 with 60 percent censoring we were only able to get 37% of the simulations to converge. What needs to be taken into account is that we are maximizing over a multidimensional surface and no procedure can guarantee convergence to a stationary point. We must also be mindful that the complexity of

these maximization problems require the starting values to be chosen wisely. What this means is that simulation in general is difficult as these models need to be massaged in order to get a single model to run. In a real world scenario this is not an issue because we are only dealing with a single optimization and we can try different maximization algorithms, different starting values, and convergence tolerances etc. This type of experimentation cannot be performed in a simulation environment but is a necessary part of any complete and thorough analysis.

One of the major criticisms of joint modeling is the inherent numerical complexity and more so when the joint model is extended to the proposed multivariate setting. The authors feel that with the ever increasing power of personal computers and more efficient maximization algorithms this is rapidly becoming less of an issue. Even so, at present, the proposed bivariate model can have excessive run times depending on the size of the dataset, the number of quadrature points fitted, and the complexity of the model chosen. We would therefore recommend balancing accuracy with runtime and feel that in most cases using five quadrature points is an effective compromise. One has to keep in mind though that the time taken to undertake the analyses is of little consequence compared to the time and resources expended in collecting the data. another viewpoint is to keep increasing the number of quadrature points until the desired level of accuracy is achieved Another issue worth noting is that the interpretation of linking or association parameters can become muddied when the longitudinal outcome is distributed other than normally. In most instances the linking parameters are interpreted as an association with the event time rather than actual magnitudes of risk.

In a previous paper by Pike and Weissfeld [32] we discussed some of the pitfalls and issues to be kept in mind when utilizing NLMIXED in the univariate joint modeling case. We feel that these issues are worth reiterating here due to the complexity of the proposed

multivariate joint Tobit model. We noticed that even though the procedure appears to run and converges, the user still needs to look in the log to see if there were any issues. We need to keep in mind that no procedure can guarantee convergence to a global maximum. This is immaterial in simulation since we know the true values of the parameters. In practice what needs to be done is to choose varied sets of starting parameters and ensure they converge to the same estimates by checking that the gradient is sufficiently small so as to indicate a stationary point. But the major question is how to choose effective starting values since NLMIXED has no automatic way to assign them. What the authors found effective was the following. Run a random effects model in PROC MIXED, ignoring the censoring, and output the parameter estimates. Run an accelerated failure time model in PROC LIFEREG and output these results. Take both sets of estimates and combine them into a single file and then use these as starting values for the bivariate joint Tobit model.

Joint modeling has long been known to reduce bias and improve precision. The proposed bivariate joint Tobit model discussed herein was shown to handle censoring proportions in excess of twenty percent with correlation. We would therefore recommend the use of the bivariate joint Tobit method when the interest is in accounting for the joint effect of censored longitudinally measured covariates and their combined association with survival where the collection of correlated censored longitudinal data is unavoidable.

6.0 CONCLUDING DISCUSSION

From the discussion presented herein it is apparent that longitudinal censoring is an unfortunate artifact of data gathering particularly when one is interested in the development of biomarkers and their subsequent effect on survival. We showed how to account for a longitudinally censored covariate in a survival model via the formation of a Joint Tobit Model. We have shown the limitations of the separate longitudinal and survival sub-models that made up the proposed Joint Tobit Model and discussed how to extract the best of what these sub-models have to offer. Explicitly we "glued" or "linked" these disparate models together by assuming that the hazard and the longitudinal trajectories were a function of the same set or shared random effects. These shared random effects are typically a random intercept and slope and when part of the hazard they capture association between the two sub-models. When no association exists these parameters are excluded from the model and there is nothing to be gained from the inclusion of a longitudinally measured covariate in a survival model. However when association does exist joint modeling tends to produce estimates with less bias, more precision etc.

We then went on to extend the univariate Joint Tobit Model to the multivariate case where it is of primary interest to account for more than one longitudinally censored trajectory and their combined or joint effect on the survival experiences of a study population. In this case we developed a multivariate Joint Tobit Model to account for longitudinal censoring in several

trajectories and linked these trajectories to the survival distribution by assuming that the hazard is a function of more than one single longitudinal trajectory.

In the "illustrative examples" we implemented the univariate and multivariate Joint Tobit Models with data from the GENIMS and HEMO studies. In the HEMO study example we discussed the impact of ignoring the censoring whilst in the GENIMS study we were primarily interested in the association between the survival and multiple longitudinal sub-models when the longitudinally measured biomarkers were censored and correlated. In both instances we implemented the proposed models conveniently in SAS PROC NLMIXED thus avoiding the inherent programming complexity and problem specific nature of the expectation maximization algorithm (EM) and discussed some of the strengths and weaknesses of the implementation.

We would therefore recommend the proposed Joint Tobit model in its univariate or multivariate forms to (i) account for correlated or independent longitudinally censored covariates in a survival model or (ii) investigate the patterns of association between longitudinally censored, correlated or independent biomarkers and survival.

6.1 FUTURE WORK

6.1.1 Motivation

Joint modeling is an increasingly popular topic in the biomedical literature and extensions on the basic principles fall into one of two categories. (1) Complicate the longitudinal sub-model or (2) complicate the survival sub-model. One possible extension to the work presented herein is to account for competing risks in the survival component (2). Oftentimes it is of interest to focus on a particular event time whilst accounting for other "competing" event times. In particular we propose to investigate extending the joint models proposed herein by including a parametric form for the competing risk model and to link the longitudinal and competing risk sub-models by a set of common random effects. The resulting joint model may then be maximized conveniently in SAS PROC NLMIXED.

Even though it is agreed that joint modeling is a useful tool in certain scenarios we feel that one of the biggest issues is the reluctance to implement these models because of the inherent numerical and programming complexity. What we feel is lacking is a basic tutorial on how to implement the common types of joint models in a user friendly environment. Therefore the author is at present working on a tutorial which will basically instruct the reader how to fit the common parametric joint models in SAS Proc NLMIXED to account for longitudinal censoring. We intend to present the reader with all the required SAS code to implement the proposed univariate and bivariate Joint Tobit models presented herein with most of the common parametric forms the user is likely to encounter.

APPENDIX: UNIVARIATE JOINT MODEL SIMULATION RESULTS

The tables to follow are results for all of the univariate simulations we ran but not included in the main paper. The longitudinal censoring ranged from 20-60 percent with sample sizes ranging from 100 – 1000. In the material presented previously herein we chose to include a representative sample of the simulation results to convey the main results.

(Quadrature Points=3, n=1000, 60% Censoring, N=1000) Time: 400+ hrs

Obs	parameter	true	Mean Joint Tobit	VARJNTTOBIT	Tobit Coverage	Mean Joint DL	VARJNTDL	biasDL	DLCoverage
1	bL0	0.00	0.001672	0.001108	0.001672	94.35	0.6692	0.02176	0.66924	.
2	bL1	1.00	1.0037	0.003344	0.003713	93.15	0.4522	0.00189	0.54776	.
3	bL2	1.00	1.0009	0.001224	0.000871	94.46	0.4749	0.00056	0.52506	.
4	bs0	0.80	0.7977	0.002979	0.002297	95.16	0.8022	0.00297	0.00220	94.79
5	bs1	1.00	1.0051	0.004470	0.005066	94.25	0.9649	0.00593	0.03510	89.15
6	r1	-1.50	-1.5420	0.016183	0.041967	94.46	-2.9432	0.33226	1.44317	.
7	r2	-1.00	-1.0132	0.005165	0.013177	94.25	-1.2694	0.92040	0.26937	22.55
8	s11	0.50	0.4850	0.001954	0.015008	94.46	0.1789	0.00166	0.32108	.
9	s22	1.00	0.9969	0.013929	0.003053	89.11	0.1024	0.00062	0.89761	.
10	s33	0.25	0.2031	0.010198	0.046866	98.99	0.4104	0.01091	0.16040	64.57
11	se	0.25	0.2562	0.000371	0.006218	93.65	0.1715	0.00007	0.07850	.

Obs	Mean Joint DL2	VARJNTDL2	biasDL2	DL2Coverage	Mean Joint DLSQRT2	VARJNTDLSQRT2	bias DLSQRT2	DLSQRT2Coverage
1	0.6086	0.0194	0.60862	.	0.6374	0.01970	0.63737	.
2	0.2223	0.0016	0.77767	.	0.3188	0.00188	0.68115	.
3	0.3288	0.0005	0.67123	.	0.3997	0.00063	0.60028	.
4	0.8055	0.0030	0.00550	95.70	0.8047	0.00306	0.00473	95.02
5	0.9556	0.0078	0.04438	86.25	0.9340	0.00799	0.06604	77.88
6	-4.6374	0.6009	3.13738	0.10	-3.7027	0.42586	2.20267	.
7	0.9074	11.7444	1.90743	69.05	-0.5494	5.10460	0.45062	18.28
8	0.08874	0.0007	0.41126	.	0.1279	0.00114	0.37214	.
9	0.03401	0.0001	0.96599	.	0.05394	0.00027	0.94606	.
10	0.3421	0.0211	0.09209	79.68	0.3827	0.01407	0.13272	73.00
11	0.1953	0.0013	0.05474	2.41	0.1590	0.00024	0.09101	.

(Quadrature Points=3, n=1000, 40% Censoring, N=1000) Time: 377 hrs

Obs	parameter	true	Mean Joint Tobit	VARJNTTOBIT	bias Tobit	Tobit Coverage	Mean Joint DL	VARJNTDL	biasDL	DLCoverage
1	bL0	0.00	0.002448	0.000937	0.002448	94.25	0.2848	0.00122	0.28480	.
2	bL1	1.00	1.0334	0.004849	0.033436	94.55	0.7544	0.00370	0.24564	1.09
3	bL2	1.00	1.0024	0.001059	0.002379	95.26	0.6747	0.00064	0.32529	.
4	bs0	0.80	0.7994	0.003285	0.000559	95.16	0.8029	0.00323	0.00293	95.19
5	bs1	1.00	1.0039	0.004777	0.003940	95.66	1.0173	0.00564	0.01728	93.66
6	r1	-1.50	-1.5562	0.019408	0.056213	94.95	-2.7249	0.05398	1.22493	.
7	r2	-1.00	-0.9752	0.007769	0.024844	95.16	-0.6484	0.22743	0.35157	91.48
8	s11	0.50	0.4864	0.001900	0.013617	94.35	0.2322	0.00071	0.26782	.
9	s22	1.00	0.9327	0.015838	0.067305	91.42	0.1778	0.00096	0.82216	.
10	s33	0.25	0.2772	0.010958	0.027164	96.37	0.3926	0.01463	0.14265	78.69
11	se	0.25	0.2588	0.000446	0.008819	94.45	0.2714	0.00034	0.02135	62.19

Obs	Mean Joint DL2	VARJNTDL2	biasDL2	DL2Coverage	Mean Joint DLSQRT2	VARJNTDLSQRT2	bias DLSQRT2	DLSQRT2Coverage
1	0.3647	0.00120	0.36467	.	0.3312	0.00119	0.33122	.
2	0.5158	0.00140	0.48421	.	0.6233	0.00242	0.37666	.
3	0.4975	0.00047	0.50248	.	0.5733	0.00051	0.42666	.
4	0.8125	0.00339	0.01251	94.28	0.8088	0.00333	0.00883	94.40
5	1.0318	0.00550	0.03184	92.75	1.0177	0.00555	0.01770	94.76
6	-3.9752	3.35546	2.47521	.	-3.3745	0.07910	1.87448	.
7	2.5288	0.43307	3.52877	3.78	0.7882	0.71039	1.78823	46.60
8	0.1080	0.00025	0.39204	.	0.1581	0.00042	0.34186	.
9	0.05866	0.00014	0.94134	.	0.08586	0.00030	0.91414	.
10	0.1504	0.00788	0.09955	99.90	0.3557	0.00939	0.10569	92.25
11	0.2931	0.00045	0.04308	17.98	0.2621	0.00034	0.01214	73.06

(Quadrature Points=3, n=1000, 20% Censoring,N=1000) Time: 407 hrs

Obs	parameter	true	Mean Joint Tobit	VARJNTTOBIT	bias Tobit	Tobit Coverage	Mean Joint DL	VARJNTDL	biasDL	DLCoverage
1	bL0	0.00	-0.00164	0.000756	0.001640	95.97	0.1349	0.000835	0.13490	0.21
2	bL1	1.00	1.0287	0.004649	0.028681	93.35	0.9749	0.004448	0.02514	89.68
3	bL2	1.00	1.0008	0.000814	0.000831	95.57	0.8567	0.000702	0.14329	0.10
4	bs0	0.80	0.8000	0.003243	0.000048	94.66	0.7996	0.003257	0.00038	94.12
5	bs1	1.00	1.0000	0.004664	0.000029	95.87	1.0216	0.005109	0.02164	93.91
6	r1	-1.50	-1.5262	0.015154	0.026168	94.86	-2.1155	0.029539	0.61550	2.37
7	r2	-1.00	-0.9736	0.006368	0.026404	95.27	-0.9092	0.036525	0.09083	96.90
8	s11	0.50	0.4940	0.001569	0.006009	95.27	0.3516	0.001086	0.14841	0.52
9	s22	1.00	0.9568	0.011299	0.043197	94.86	0.4071	0.003203	0.59291	.
10	s33	0.25	0.2708	0.010153	0.020849	96.68	0.2972	0.014916	0.04716	93.09
11	se	0.25	0.2531	0.000323	0.003109	95.57	0.3250	0.000368	0.07501	1.03

Obs	Mean Joint DL2	VARJNTDL2	biasDL2	DL2Coverage	Mean Joint DLSQRT2	VARJNTDLSQRT2	bias DLSQRT2	DLSQRT2Coverage
1	0.2641	0.00091	0.26407	.	0.2084	0.00084	0.20839	.
2	0.8216	0.00319	0.17841	8.94	0.8926	0.00392	0.10744	52.25
3	0.7137	0.00061	0.28626	.	0.7749	0.00063	0.22508	.
4	0.8108	0.00334	0.01075	94.74	0.8058	0.00321	0.00584	94.67
5	1.0260	0.00549	0.02601	94.12	1.0205	0.00536	0.02052	94.19
6	-3.1248	0.04868	1.62478	.	-2.6530	0.03917	1.15297	.
7	0.1037	0.23745	1.10367	39.53	-0.5023	0.14503	0.49770	78.44
8	0.2036	0.00060	0.29644	0.12	0.2657	0.00073	0.23431	.
9	0.1828	0.00101	0.81724	.	0.2402	0.00164	0.75979	.
10	0.2759	0.00669	0.02589	99.27	0.2917	0.01116	0.04170	97.04
11	0.4879	0.00104	0.23791	.	0.3878	0.00054	0.13776	.

(Quadrature Points=3, n=500, 60% Censoring, N=1000) Time: 382+hrs

Obs	parameter	true	Mean Joint Tobit	VARJNTTOBIT	bias Tobit	Tobit Coverage	Mean Joint DL	VARJNTDL	biasDL	DLCoverage
1	bL0	0.00	-0.00289	0.002228	0.00289	95.66	0.4221	0.00399	0.42210	.
2	bL1	1.00	1.0539	0.012406	0.05390	93.64	0.5067	0.00448	0.49330	.
3	bL2	1.00	1.0033	0.003247	0.00335	94.14	0.4603	0.00091	0.53971	.
4	bs0	0.80	0.8059	0.005968	0.00595	94.95	0.8121	0.00581	0.01205	95.24
5	bs1	1.00	1.0023	0.010341	0.00229	95.25	1.0124	0.01151	0.01236	94.08
6	r1	-1.50	-1.6064	0.047617	0.10637	95.76	-3.7699	0.43019	2.26994	0.23
7	r2	-1.00	-0.9836	0.022923	0.01637	95.86	0.06729	2.63450	1.06729	81.53
8	s11	0.50	0.4693	0.004529	0.03074	97.27	0.1171	0.00049	0.38292	0.12
9	s22	1.00	0.8801	0.040528	0.11993	92.63	0.06419	0.00038	0.93581	.
10	s33	0.25	0.2696	0.021872	0.01955	97.52	0.4292	0.02251	0.17917	86.33
11	se	0.25	0.2738	0.001500	0.02380	92.83	0.1694	0.00033	0.08060	0.35

Obs	Mean Joint DL2	VARJNTDL2	biasDL2	DL2Coverage	Mean Joint DLSQRT2	VARJNTDLSQRT2	bias DLSQRT2	DLSQRT2Coverage
1	0.3826	0.00293	0.38257	.	0.3978	0.00341	0.39777	.
2	0.2849	0.00125	0.71510	.	0.3836	0.00229	0.61644	.
3	0.3020	0.00058	0.69796	.	0.3704	0.00065	0.62956	.
4	0.8133	0.00574	0.01335	95.36	0.8167	0.00596	0.01672	94.34
5	1.0241	0.01141	0.02413	94.28	1.0156	0.01181	0.01562	93.53
6	-5.5178	2.91239	4.01780	1.41	-4.6382	0.60498	3.13816	0.23
7	6.4622	3.96951	7.46219	6.70	3.0712	5.45014	4.07117	43.58
8	0.04833	0.00013	0.45167	0.11	0.07416	0.00025	0.42584	0.12
9	0.01571	0.00003	0.98429	.	0.02734	0.00009	0.97266	.
10	0.1981	0.01237	0.05189	98.90	0.3297	0.02372	0.07971	95.76
11	0.1398	0.00037	0.11023	.	0.1369	0.00028	0.11307	.

(Quadrature Points=3, n=500, 40% Censoring, N=1000) Time: 380+ hrs

Obs	parameter	true	Mean Joint Tobit	VARJNTTOBIT	bias Tobit	Tobit Coverage	Mean Joint DL	VARJNTDL	biasDL	DLCoverage
1	bL0	0.00	-0.00121	0.0018	0.00121	95.17	0.2839	0.00250	0.28391	.
2	bL1	1.00	1.0373	0.0108	0.03726	93.96	0.7542	0.00748	0.24580	9.60
3	bL2	1.00	1.0021	0.0020	0.00207	95.17	0.6739	0.00123	0.32607	.
4	bs0	0.80	1.4730	0.5262	0.67304	50.05	1.4737	0.52604	0.67367	50.51
5	bs1	1.00	1.0048	0.0109	0.00484	94.36	1.0198	0.01192	0.01982	92.32
6	r1	-1.50	-1.5533	0.0351	0.05331	95.47	-2.6856	0.09264	1.18563	1.13
7	r2	-1.00	-0.9708	0.0155	0.02923	96.07	-0.6697	0.47114	0.33031	93.22
8	s11	0.50	1.5102	1.2228	1.01021	50.55	0.8766	0.47260	0.37658	.
9	s22	1.00	3.7860	12.1999	2.78601	49.24	0.7760	0.41290	0.22401	.
10	s33	0.25	0.9621	0.6693	0.71210	51.30	1.2004	1.03194	0.95039	47.63
11	se	0.25	0.7479	0.2698	0.49786	49.04	0.7592	0.27120	0.50915	40.11

Obs	Mean Joint DL2	VARJNTDL2	biasDL2	DL2Coverage	Mean Joint DLSQRT2	VARJNTDLSQRT2	bias DLSQRT2	DLSQRT2Coverage
1	0.3649	0.002	0.36489	.	0.3290	0.00234	0.32901	.
2	0.5175	0.003	0.48252	.	0.6208	0.00520	0.37917	.
3	0.4987	0.001	0.50133	.	0.5736	0.00100	0.42635	.
4	1.4863	0.539	0.68631	50.80	1.4774	0.53951	0.67740	51.21
5	1.0325	0.012	0.03253	92.33	1.0197	0.01228	0.01973	92.38
6	-3.2181	263.267	1.71806	0.11	-3.3250	0.14156	1.82505	.
7	2.4949	1.012	3.49486	19.81	0.7265	1.50620	1.72655	66.58
8	0.6407	0.324	0.14067	0.11	0.7234	0.37118	0.22341	.
9	0.5562	0.285	0.44382	.	0.5975	0.30730	0.40253	.
10	0.7816	0.720	0.53161	56.50	1.1213	0.85583	0.87129	51.99
11	0.7858	0.276	0.53583	18.53	0.7413	0.27059	0.49127	41.55

(Quadrature Points=3, n=500, 20% Censoring, N=1000) Time: 203+ Hrs

Obs	parameter	true	Mean Joint Tobit	VARJNTTOBIT	bias Tobit	Tobit Coverage	Mean Joint DL	VARJNTDL	bias DL	DL Coverage
1	bL0	0.00	-0.00289	0.002228	0.00289	95.66	0.4221	0.00399	0.42210	.
2	bL1	1.00	1.0539	0.012406	0.05390	93.64	0.5067	0.00448	0.49330	.
3	bL2	1.00	1.0033	0.003247	0.00335	94.14	0.4603	0.00091	0.53971	.
4	bs0	0.80	0.8059	0.005968	0.00595	94.95	0.8121	0.00581	0.01205	95.24
5	bs1	1.00	1.0023	0.010341	0.00229	95.25	1.0124	0.01151	0.01236	94.08
6	r1	-1.50	-1.6064	0.047617	0.10637	95.76	-3.7699	0.43019	2.26994	0.23
7	r2	-1.00	-0.9836	0.022923	0.01637	95.86	0.06729	2.63450	1.06729	81.53
8	s11	0.50	0.4693	0.004529	0.03074	97.27	0.1171	0.00049	0.38292	0.12
9	s22	1.00	0.8801	0.040528	0.11993	92.63	0.06419	0.00038	0.93581	.
10	s33	0.25	0.2696	0.021872	0.01955	97.52	0.4292	0.02251	0.17917	86.33
11	se	0.25	0.2738	0.001500	0.02380	92.83	0.1694	0.00033	0.08060	0.35

Obs	Mean Joint DL2	VARJNTDL2	bias DL2	DL2 Coverage	Mean Joint DLSQRT2	VARJNTDLSQRT2	bias DLSQRT2	DLSQRT2 Coverage
1	0.3826	0.00293	0.38257	.	0.3978	0.00341	0.39777	.
2	0.2849	0.00125	0.71510	.	0.3836	0.00229	0.61644	.
3	0.3020	0.00058	0.69796	.	0.3704	0.00065	0.62956	.
4	0.8133	0.00574	0.01335	95.36	0.8167	0.00596	0.01672	94.34
5	1.0241	0.01141	0.02413	94.28	1.0156	0.01181	0.01562	93.53
6	-5.5178	2.91239	4.01780	1.41	-4.6382	0.60498	3.13816	0.23
7	6.4622	3.96951	7.46219	6.70	3.0712	5.45014	4.07117	43.58
8	0.04833	0.00013	0.45167	0.11	0.07416	0.00025	0.42584	0.12
9	0.01571	0.00003	0.98429	.	0.02734	0.00009	0.97266	.
10	0.1981	0.01237	0.05189	98.90	0.3297	0.02372	0.07971	95.76
11	0.1398	0.00037	0.11023	.	0.1369	0.00028	0.11307	.

(Quadrature Points=3, n=100, 60% Censoring, N=1000) Time: 13+ Hrs

Obs	parameter	true	Mean Joint Tobit	VARJNTTOBIT	bias Tobit	Tobit Coverage	Mean Joint DL	VARJNTDL	bias DL	DL Coverage
1	bL0	0.00	-0.01413	0.012	0.01413	94.77	0.4196	0.02	0.41958	1.25
2	bL1	1.00	1.1208	0.075	0.12082	91.79	0.5058	0.02	0.49422	2.17
3	bL2	1.00	1.0196	0.016	0.01965	95.08	0.4600	0.00	0.54000	.
4	bs0	0.80	0.8262	0.034	0.02617	94.15	0.8321	0.03	0.03215	94.30
5	bs1	1.00	1.0089	0.051	0.00890	96.31	1.0226	0.06	0.02261	94.75
6	r1	-1.50	-1.6277	0.250	0.12771	97.03	-3.1321	83.42	1.63214	52.23
7	r2	-1.00	-0.4910	111.769	0.50898	93.70	1.8191	1606.72	2.81906	85.35
8	s11	0.50	0.4813	0.024	0.01874	98.05	0.1250	0.00	0.37501	.
9	s22	1.00	0.8531	0.232	0.14694	95.56	0.06984	0.00	0.93016	2.07
10	s33	0.25	0.3062	0.055	0.05617	99.30	0.3525	0.05	0.10247	98.91
11	se	0.25	0.2726	0.008	0.02256	95.79	0.1640	0.00	0.08597	32.95

Obs	Mean Joint DL2	VARJNTDL2	bias DL2	DL2 Coverage	Mean Joint DLSQRT2	VARJNTDLSQRT2	bias DLSQRT2	DLSQRT2 Coverage
1	0.3775	0.01	0.37751	0.59	0.3954	0.016	0.39543	0.69
2	0.2883	0.01	0.71174	.	0.3757	0.011	0.62434	.
3	0.3029	0.00	0.69713	.	0.3705	0.003	0.62951	.
4	0.8364	0.04	0.03640	94.13	0.8339	0.037	0.03388	94.81
5	1.0239	0.06	0.02390	95.67	1.0149	0.056	0.01491	95.74
6	-0.3580	3166.39	1.14204	56.97	-3.8041	73.605	2.30413	51.75
7	6.3043	64.24	7.30430	69.05	3.2587	115.072	4.25868	77.38
8	0.05634	0.00	0.44366	1.14	0.08184	0.001	0.41816	0.71
9	0.02110	0.00	0.97890	5.04	0.03478	0.001	0.96522	3.17
10	0.3446	0.09	0.09457	98.51	0.3472	0.142	0.09715	98.17
11	0.1416	0.00	0.10841	12.88	0.1348	0.001	0.11522	8.53

(Quadrature Points=3, n=100, 40% Censoring, N=1000) Time: 32+hrs

Obs	parameter	true	Mean Joint Tobit	VARJNTTOBIT	bias Tobit	Tobit Coverage	Mean Joint DL	VARJNTDL	bias DL	DL Coverage
1	bL0	0.00	-0.00437	0.0094	0.00437	94.83	0.2809	0.01243	0.28089	10.29
2	bL1	1.00	1.0935	0.0608	0.09354	90.47	0.7737	0.03192	0.22627	57.96
3	bL2	1.00	1.0114	0.0110	0.01141	95.13	0.6749	0.00608	0.32509	1.22
4	bs0	0.80	0.8237	0.0329	0.02374	95.74	0.8289	0.03298	0.02891	95.69
5	bs1	1.00	1.0159	0.0536	0.01590	93.61	1.0237	0.05932	0.02371	93.81
6	r1	-1.50	-1.4615	14.7587	0.03846	95.93	-2.6788	0.49640	1.17879	72.39
7	r2	-1.00	-0.8460	4.6796	0.15404	95.74	-0.3457	4.18342	0.65435	89.01
8	s11	0.50	0.4855	0.0186	0.01454	97.56	0.2341	0.00624	0.26588	22.81
9	s22	1.00	0.9085	0.1735	0.09146	95.13	0.1828	0.01042	0.81724	4.11
10	s33	0.25	0.3247	0.0635	0.07473	98.03	0.3477	0.05808	0.09770	99.20
11	se	0.25	0.2632	0.0054	0.01318	94.42	0.2667	0.00332	0.01668	85.07

Obs	Mean Joint DL2	VARJNTDL2	bias DL2	DL2 Coverage	Mean Joint DLSQRT2	VARJNTDLSQRT2	bias DLSQRT2	DLSQRT2 Coverage
1	0.3585	0.013	0.35847	1.58	0.3246	0.012	0.32461	3.64
2	0.5143	0.017	0.48569	4.51	0.6185	0.022	0.38147	15.60
3	0.4954	0.005	0.50457	.	0.5717	0.005	0.42826	.
4	0.8373	0.036	0.03734	95.60	0.8360	0.036	0.03595	95.56
5	1.0352	0.061	0.03525	93.80	1.0286	0.062	0.02865	93.85
6	-1.9283	412.051	0.42832	53.03	-3.4559	87.161	1.95594	57.19
7	2.4011	10.724	3.40108	68.37	1.5147	431.806	2.51469	80.78
8	0.1148	0.002	0.38521	1.96	0.1641	0.004	0.33590	3.56
9	0.06753	0.003	0.93247	2.99	0.09611	0.004	0.90389	3.11
10	0.3044	0.380	0.05438	98.94	0.3500	0.237	0.10005	98.50
11	0.2961	0.006	0.04611	64.49	0.2596	0.004	0.00964	79.95

(Quadrature Points=3, n=100, 20% Censoring, N=1000) Time:35+ Hrs

Obs	parameter	true	Mean Joint Tobit	VARJNTTOBIT	bias Tobit	Tobit Coverage	Mean Joint DL	VARJNTDL	bias DL	DLCoverage
1	bL0	0.00	-0.00077	0.00751	0.000768	96.30	0.1382	0.00822	0.13822	58.29
2	bL1	1.00	1.0815	0.04720	0.081491	94.14	1.0081	0.04340	0.00806	89.95
3	bL2	1.00	1.0039	0.00938	0.003945	93.83	0.8548	0.00815	0.14516	54.76
4	bs0	0.80	0.8151	0.03597	0.015144	94.14	0.8170	0.03600	0.01697	93.80
5	bs1	1.00	1.0215	0.05109	0.021463	95.27	1.0407	0.05633	0.04073	94.97
6	r1	-1.50	-1.5656	0.40853	0.065644	96.30	-2.0811	2.61390	0.58109	88.96
7	r2	-1.00	-0.9467	0.12437	0.053319	97.43	-0.8004	0.66490	0.19962	94.11
8	s11	0.50	0.4821	0.01490	0.017933	97.12	0.3472	0.01047	0.15278	78.56
9	s22	1.00	0.9252	0.11690	0.074794	96.09	0.4044	0.02980	0.59556	38.54
10	s33	0.25	0.3073	0.06265	0.057294	97.28	0.3261	0.05425	0.07612	98.39
11	se	0.25	0.2568	0.00341	0.006764	94.35	0.3177	0.00383	0.06767	68.24

Obs	Mean Joint DL2	VARJNTDL2	biasDL2	DL2Coverage	Mean Joint DLSQRT2	VARJNTDLSQRT2	bias DLSQRT2	DLSQRT2Coverage
1	0.2648	0.0097	0.26478	14.41	0.2100	0.00860	0.21001	29.39
2	0.8194	0.0283	0.18055	75.50	0.9073	0.03495	0.09275	84.80
3	0.7142	0.0069	0.28578	7.82	0.7737	0.00701	0.22625	21.73
4	0.8278	0.0372	0.02778	94.19	0.8185	0.03593	0.01850	94.26
5	1.0418	0.0583	0.04177	94.75	1.0408	0.05810	0.04083	94.71
6	-2.6840	67.4879	1.18401	61.49	-2.6239	0.41653	1.12387	70.99
7	0.3208	13.5554	1.32080	83.41	-0.3773	1.74383	0.62274	91.89
8	0.2070	0.0060	0.29298	30.02	0.2686	0.00784	0.23141	52.31
9	0.1902	0.0114	0.80978	5.51	0.2432	0.01533	0.75684	10.37
10	0.2578	0.0512	0.00776	98.65	0.2800	0.04164	0.03000	99.09
11	0.4976	0.0121	0.24763	6.93	0.3888	0.00640	0.13875	28.04

BIBLIOGRAPHY

[1] Hughes, J. P. (1999). Mixed effects models with censored data with application to HIV RNA Levels.Biometrics55, 625-629.

[2] Lubin, J.H.,Colt,J.S,et al. (2004) Epidemiologic Evaluation of Measurement Data in the Presence of DetectionLimits. Environmental health Perspectives; Vol112, Num 17.

[3] Epstein, M.P., Lin, X, Boehnke,M.(2003) A Tobit Variance-Component Method for Linkage Analysis of Censored Trait Data. Am. J. Hum. Genet; 72:611-620

[4] Kellum, J. A., Kong, L., Fink, M. P., Weissfeld, L. A., Yealy, D. M., Pinsky, M.R.; GenIMS Investigators (2007). Understanding the inflammatory cytokine response in pneumonia and sepsis: results of the Genetic and Inammatory Markers of Sepsis (GenIMS) Study. Arch. Intern. Med. 167:1655-63.

[5] Eknoyan, G, Beck, G.J., Cheung, A. K., Daugirdas, J.T., Green, T. et al (2002). Effect of dialysis dose and membrane flux in maintenance hemodialysis. N Engl J Med. 347(25):2010-9

[6] Klein, J. and Moeschberger, M. Survival Analysis: Statistics for biology and health, 2nd ed. Springer Science+Business Media Inc.: New York. (2003).

[7] Kaplan, E.L. and Meyer, P. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 53(282):457-481, 1958

[8] Duchateau, L, and Janssen, P, The Frailty Model: Statistics for Biology and Health. Springer Science + Business Media Inc.: New York [2008].

[9] Fitzmaurice, G, M, Laird, N, and Ware, J, Applied Longitudinal Analysis, Wiley series in Probability and Statistics, John Wiley & Sons, Inc, Hoboken, New Jersey [2004]

[10] Zeger, S, L, Liang, K, Longitudinal Data Analysis for Discrete and Continuous Outcomes, *Biometrics*, Vol. 42, No. 1. (Mar, 1986), pp. 121-130.

[11] McCullagh,P. and Nelder, J, A. (1983). Quasi-likelihood functions. Annals of Statistics 11, 59-67.

[12] Rubin, D, B. (1976) Inference and missing data (with Discussion). Biometrika 63, 581-592.

[13] Verbeke, G, and Molenberghs, Linear Mixed Models for Longitudinal Data, Springer Series in Statistics, Springer Science + Business Media Inc.: New York [200].

[14] Tobin, James (1958), "Estimation of relationships for limited dependent variables", *Econometrica* (The Econometric Society) **26** (1): 24–36.

[15] Lyles, R. H., Lyless, C.M, Taylor, D.J. (2000), Random Regression Models for Human Immunodeficiency Virus Ribonucleic Acid Data Subject to Left Censoring and Informative Drop-Outs, Journal of the Royal Statistical Society. Series C (Applied Statistics), Vol. 49, No. 4 (2000), pp. 485-497

[16] Tsiatis, A.A, and Davidian, M: Joint Modeling of longitudinal and time-to-event data: An overview. Stat Sin 14:808-834, 2004.

[17] Prentice, R. (1982). Covariate measurement errors and parameter estimates in a failure time regression model. Biometrika 69, 331-342.

[18] Collett, D. Modelling Survival Data in Medical Research. Chapman & Hall: London 1999.

[19] Liu,Y,and Craig,A,B: Incorporating time –dependent covariates in survival analysis using the LVAR method: Statistics in Medicine. 2006: 25:1729-1740.

[20] Faucett CL, Thomas, DC. Simultaneously modeling censored survival data and repeatedly measured covariates a Gibbs sampling approach. Statistics in medicine 1996: 15:1663-1685

[21] Henderson R, Diggle P, Dobson A. Joint modeling of longitudinal measurements and event time data. Biostatistics 2000: 1(4):465-480.

[22] Xu J, Zeger SL. Joint analysis of longitudinal data comprising repeated measures and times to events. Applied Statistics 2001; 50(3):375-387.

[23] Tsiatis AA, Davidian M. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. Biometrika 2001; 88(2):447-458.

[24] De Gruttola V, Tu X. Modeling progression of CD$-lymphocyte count and its relationship to survival time. Biometrics 1994; 50(4):1003-1014.

[25] Brown E, R, Ibrahim J, G, and DeGruttola V. A Flexible B-Spline Model for Multiple Longitudinal Biomarkers and Survival. Biometrics 2005; 61, 64-73.

[26] Guo, X. and Carlin, B.P. (2004), Separate and Joint Modeling of Longitudinal and Event Time Data Using Standard Computer Paxckages. *The American Statistician*, 58, 16--24.

[27] Laird, N, M. and Ware, J. H. (1982) Random effects models for longitudinal data. Biometrics 38, 963-974.

[28] Hsieh F, Tseng YK, Wang JL (2006). "Joint Modeling of Survival and Longitudinal data: Likelihood Approach Revisited." Biometrics, 62, 1037-1043.

[29] Rizopoulos D, Verbeke G, and Molenberghs G, (2008). "Shared Parameter Models under Random Effects Misspecification." Biometrika, 95, 63-74.

[30] Wulfsohn M, and Tsiatis A, . "A Joint Model for Survival and Longitudinal Data Measured with Error."Biometrics, Vol. 53, No.1 (Mar. 1997), pp 330-339.

[31] DeGruttola, V. and Tu, X. M. (1994). Modeling progression of CD-4 Lymphocyte count and its relationship to survival time. Biometrics 50, 1003-1014.

[32] Pike, F., Weissfeld, L. (2011). Accounting for a Doubly Censored Longitudinal Covariate in Survival Analysis via Joint Modeling. (in review)

[33] SAS Institute Inc. 2008. SAS/STAT® 9.2

[34] Thiebaut, R, Jacqmin-Gadda, H, Chene, G, Leport, C, D. Commenges, Bivariate linear mixed models using SAS® Proc Mixed, Comput. Methods Programs Biomed. 69 (2002) 249-256.

[35] Rizopoulos, D. Fully Exponential Laplace approximations for the joint modeling of survival and longitudinal data. J.R. Statist. Soc B (2008) 71, part3, pp 637-654.