Published in final edited form as:

Comput Stat Data Anal. 2015 May 1; 85: 37–53. doi:10.1016/j.csda.2014.11.011.

A Fast EM Algorithm for Fitting Joint Models of a Binary Response and Multiple Longitudinal Covariates Subject to Detection Limits

Paul W. Bernhardt^{a,*}, Daowen Zhang^b, and Huixia Judy Wang^c

^aDepartment of Mathematics and Statistics, Villanova University, Villanova, PA, USA

^bDepartment of Statistics, North Carolina State University, Raleigh, NC, USA

^cDepartment of Statistics, George Washington University, Washington, DC, USA

Abstract

Joint modeling techniques have become a popular strategy for studying the association between a response and one or more longitudinal covariates. Motivated by the GenIMS study, where it is of interest to model the event of survival using censored longitudinal biomarkers, a joint model is proposed for describing the relationship between a binary outcome and multiple longitudinal covariates subject to detection limits. A fast, approximate EM algorithm is developed that reduces the dimension of integration in the E-step of the algorithm to one, regardless of the number of random effects in the joint model. Numerical studies demonstrate that the proposed approximate EM algorithm leads to satisfactory parameter and variance estimates in situations with and without censoring on the longitudinal covariates. The approximate EM algorithm is applied to analyze the GenIMS data set.

Keywords

Detection limit; EM algorithm; Joint model; Logistic regression; Multiple longitudinal covariates; Normal approximation

1. Introduction

In biomedical studies, biomarkers are commonly measured repeatedly over time in order to capture changes within patients over the progression of a certain disease. When it is of

© 2014 Elsevier B.V. All rights reserved.

*Corresponding author: Tel: 610 519 7343; Fax: 610 519 6928, paul.bernhardt@villanova.edu (Paul W. Bernhardt), URL: <http://www.homepage.villanova.edu/paul.bernhardt> (Paul W. Bernhardt).

Supplementary Material

The reader is referred to the online Supplementary Material for technical appendices and to <http://www.homepage.villanova.edu/paul.bernhardt/Resources3.htm> for simulation code, application code, and data similar to that analyzed in the application.

Conflict of Interest: None declared.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

interest to use these longitudinal biomarkers as covariates for the purpose of describing a response, joint models have become a popular modeling strategy.

A joint model is a type of model comprised of two or more submodels that are related through a common set of latent variables. Generally, linear or no-linear mixed models are proposed for describing covariates that are measured over time and an outcome model is proposed for the response. The random effects in the longitudinal mixed models are then incorporated into the outcome model in order to capture the effect of the covariates on the response. By including the random effects in the outcome model rather than the actual longitudinally-observed covariate values, joint models also naturally account for measurement error on the longitudinal covariates. Most commonly, the response of interest for joint modeling is a time-to-event variable, which may be described by a survival model such as the Cox proportional hazards model. The standard approach for fitting joint models for a survival outcome was developed by Wulfsohn and Tsiatis (1997), who suggested using an expectation-maximization (EM) algorithm in which Gauss-Hermite quadrature is used to compute integrals in the E-step. Unfortunately, this approach for obtaining maximum likelihood estimates can be very computationally expensive for joint models with a large number of random effects.

In the motivating GenIMS study, biological measurements were collected daily on patients admitted to the hospital for community acquired pneumonia, generally over a one to eight day period. One of the main purposes of the study was to investigate the relationship between the event of survival after 90 days and three cytokine biomarkers, which were longitudinally measured and subject to lower detection limits (DLs). Since the response of interest, survival after 90 days, is binary, we consider fitting a logistic submodel rather than a survival submodel. This challenge, together with the fact that there are multiple longitudinal covariates of interest and these covariates are subject to censoring at DLs, makes the GenIMS data set somewhat unique in the joint modeling literature.

Several papers have considered extensions of the standard joint model which individually address one of these three data challenges. With respect to handling multiple longitudinal covariates, Lin et al. (2002) suggested a maximum likelihood method based on using an EM algorithm where expectations in the E-step are computed using Monte Carlo integration. However, since the number of random effects in the model – corresponding to the dimension of integration in the joint model likelihood – increases with more longitudinal covariates, obtaining parameter estimates may be computationally cumbersome. For this reason, Ye et al. (2008) and Rizopoulos et al. (2009) proposed using Laplace approximations to calculate the expectations in the E-step of the EM algorithm while Proust-Lima et al. (2009) suggested a different modeling strategy altogether, where the longitudinal covariates are described by one latent process but with a single additional random effect specific to each covariate.

A few researchers have considered fitting joint models with a logistic submodel for the response variable. Wang et al. (2000) suggested both regression calibration and estimating equation approaches, and Li et al. (2004) later used similar estimating equation approaches based on sufficient and conditional scores that can be applied to any generalized linear

model with subject-specific random effects. This estimating equation strategy was extended by Li et al. (2007a) to a model with multiple longitudinal covariates. Li et al. (2007b) proposed a semiparametric version of the joint model in which normality of the random effects was not assumed, and then suggested using an EM algorithm or two-stage pseudo-likelihood approach to estimate the parameters in the model. Recently, Hwang et al. (2011) extended the EM algorithm approach developed by Wulfsohn and Tsiatis (1997) to the context of a logistic regression submodel for the response.

There exists limited work for joint modeling with longitudinal covariates subject to DLs. Wu (2002) considered a joint model for a censored longitudinal outcome described by a nonlinear mixed effects model with covariates measured with error. May (2011) modeled survival time for AIDS patients using CD4 counts that were repeatedly measured. Though the CD4 counts were not censored, the longitudinal model for these CD4 counts included viral loads, which were subject to lower DLs. May (2011) proposed a Bayesian joint model that fits a truncated prior to the viral loads that were observed below the DL. Su et al. (2009) described semicontinuous longitudinal data via a two-part mixed model with correlated random effects. This model can be applied to cases with censoring due to DLs but considers the actual event of censoring to be of interest rather than a model for the unobserved outcome. Pike (2013) also recently considered a joint model with a single longitudinal covariate subject to a lower or upper DL in the context of a survival outcome. Several others have considered joint inference for modeling longitudinal response data subject to missingness or dropout together with the missingness mechanism (e.g. Wu and Carroll, 1988; Have et al., 1998, 2000; Wu et al., 2008; Yuan and Little, 2009; Sattar et al., 2011). However, none of these authors considered the longitudinal data as covariates to a separate response and only Sattar et al. (2011) incorporated left-censoring. Additionally, most of these authors did not focus on models with high-dimensional random effects.

While several authors have considered joint modeling of a binary outcome, possibly with multiple longitudinal variables or covariates subject to DLs, no author has considered all three of these data aspects simultaneously. The censoring on the covariates makes it technically challenging to extend existing methods for fitting joint models with binary outcomes, such as those based on estimating equations, while also making an EM algorithm approach more computationally intensive since quadrature integration methods are more difficult to apply with censored covariate data. Incorporating multiple longitudinal covariates subject to DLs exacerbates the computational challenges.

In this paper, we propose a likelihood-based approach for fitting the joint model. We suggest using a censored linear mixed model with normal errors to describe the longitudinal covariates. We then include the associated censored-data distributions for the longitudinal covariates in the overall joint likelihood also containing the binary response and random effects distributions. In order to overcome the computational difficulties for fitting this joint likelihood, we propose a new EM algorithm approach that uses a normal density to approximate the distribution of the random effects given the data in the E-step of the algorithm. This approximation reduces the dimension of necessary integration to one, regardless of the number of random effects, and thus improves computational efficiency. We show through simulations that our approach for fitting the joint model of a binary response

and multiple longitudinal covariates subject to DLs leads to approximately consistent parameter and variance estimates while greatly reducing computational times compared to an EM algorithm based on standard Monte Carlo integration methods.

The remainder of this paper is organized as follows. In Section 2, we lay out the joint modeling framework when there is no censoring on the covariates, and present the approximate EM algorithm to obtain maximum likelihood estimates. We also describe a method for variance estimation. In Section 3, we extend this framework to include censoring on the covariates due to DLs. In Section 4, we evaluate our proposed normal approximation through simulations, both when censoring is and is not present in the covariates. In Section 5, we apply our proposed method to the GenIMS data set. Section 6 concludes the paper with discussions of limitations and avenues for further research. Technical details for the proposed approximate EM algorithm and some additional simulation results are provided in the online Supplementary Material.

2. Joint Modeling of a Binary Response and Multiple Longitudinal Covariates

In the following we present the longitudinal and logistic submodels for the covariates and binary response, as well as the joint likelihood that we wish to maximize to obtain parameter estimates. We then describe an approximate EM algorithm for maximizing the joint likelihood. In this section, we assume that the longitudinal covariates are not subject to DLs.

2.1. Models and Notation

Suppose we observe independent samples $\{Y_i, X_{i1}(t_{i1}), \dots, X_{iq}(t_{iq}), \mathbf{Z}_i\}$, $i = 1, \dots, n$, where Y_i is a binary response, $\mathbf{X}_{ik}(t_{ik}) = (X_{i1k}(t_{i1k}), \dots, X_{in_{ik}k}(t_{in_{ik}k}))^T$, $k = 1, \dots, q$, are longitudinal covariate vectors with n_{ik} repeated observations at times $t_{ik} = (t_{i1k}, \dots, t_{in_{ik}k})^T$, and $\mathbf{Z}_i$ is a $(p - q)$ -dimensional vector of baseline covariates. For convenience, we generally refer to $\mathbf{X}_{ik}(t_{ik})$ simply as $\mathbf{X}_{ik}$, though we note that the model for $\mathbf{X}_{ik}$ is actually conditional on t_{ik} . We adopt the following linear mixed model for the k^{th} longitudinal covariate:

$$X_{ijk} = \mathbf{r}_{ijk}^T \boldsymbol{\gamma}_k + \mathbf{s}_{ijk}^T \mathbf{b}_{ik} + \varepsilon_{ijk}, \quad i=1, \dots, n, \quad j=1, \dots, n_{ik}, \quad (1)$$

where $\varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \tau_k^2)$, $\mathbf{r}_{ijk}$ is a possibly time-dependent, fixed-effects design vector, potentially including the baseline covariates $\mathbf{Z}_i$, and $\mathbf{s}_{ijk}$ is a possibly time-dependent random-effects design vector. We assume that the matrix comprised of the row vectors $(\mathbf{r}_{ijk}^T, \mathbf{s}_{ijk}^T) \forall i, j, k$ is full rank, and that $n_{ik} > \text{length}\{\mathbf{b}_{ik}\}$ for some individuals. We also assume that the longitudinal covariates $\mathbf{X}_{ik}$, $k = 1, \dots, q$, are conditionally independent given the vector of baseline covariates $\mathbf{Z}_i$ and the vector of random effects $\mathbf{b}_i = (\mathbf{b}_{i1}^T, \mathbf{b}_{i2}^T, \dots, \mathbf{b}_{iq}^T)^T$, which we assume has the multivariate normal distribution $N(\mathbf{B}, \mathbf{D})$, where $\mathbf{B}$ is the mean vector for the random effects, $\mathbf{D}$ is an unstructured, positive-definite covariance matrix, and the dimensions of $\mathbf{B}$ and $\mathbf{D}$ correspond directly to the length of $\mathbf{b}_i$.

We assume that the binary response Y_i is related to the longitudinal covariates $\mathbf{X}_{ik}$, $k = 1, \dots, q$, only through the random effects $\mathbf{b}_i$ and possibly $\mathbf{Z}_i$. Specifically, we adopt the following logistic model for $\pi(\mathbf{Z}_i, \mathbf{b}_i) = \Pr(Y_i = 1 | \mathbf{Z}_i, \mathbf{b}_i)$:

$$\text{logit}\{\pi(\mathbf{Z}_i, \mathbf{b}_i)\} = \beta_0 + \beta_1^T \mathbf{Z}_i + \beta_2^T \mathbf{b}_i, \quad (2)$$

where $\text{logit}(p) = \log\{p(1-p)^{-1}\}$ and $\beta = (\beta_0, \beta_1^T, \beta_2^T)^T$ is the $(p+1)$ -dimensional vector of regression coefficients. We assume that the $\mathbf{Z}$ matrix whose i th row is $(1, \mathbf{Z}_i^T)$ is full rank.

The joint likelihood for all the data is given by

$$\prod_{i=1}^n \int_{-\infty}^{\infty} f(y_i | \mathbf{z}_i, \mathbf{b}_i; \beta) \left\{ \prod_{k=1}^q f(\mathbf{x}_{ik} | \mathbf{z}_i, \mathbf{b}_{ik}; \gamma_k, \tau_k^2) \right\} f(\mathbf{b}_i; B, D) d\mathbf{b}_i, \quad (3)$$

where $f(y_i | \mathbf{z}_i, \mathbf{b}_i; \beta)$ is the probability mass function of the

Bernoulli $[\{1 + \exp(-\beta_0 - \beta_1^T \mathbf{z}_i - \beta_2^T \mathbf{b}_i)\}^{-1}]$ distribution and

$f(\mathbf{x}_{ik} | \mathbf{z}_i, \mathbf{b}_{ik}; \gamma_k, \tau_k^2) = \prod_{j=1}^{n_{ik}} f(x_{ijk} | \mathbf{z}_i, \mathbf{b}_{ik}; \gamma_k, \tau_k^2)$ is the product of normal pdfs from model (1).

We henceforth denote all the parameters by $\theta = (\beta^T, \gamma^T, \tau^2, B^T, \text{vech}(D)^T)^T$, where

$\gamma = (\gamma_1^T, \dots, \gamma_q^T)^T$, $\tau^2 = (\tau_1^2, \dots, \tau_q^2)^T$, and $\text{vech}(D)$ is the column vector comprised of the elements in the upper right triangle in D . Also, we denote both the density and probability mass functions of discrete and continuous distributions by f . Note that because we assume mutual conditional independence of Y_i and $\mathbf{X}_{ik}$, $k = 1, \dots, q$, given $\mathbf{b}_i$ and $\mathbf{Z}_i$, each parameter is only involved in a single piece of the complete-data log-likelihood. In order to maximize the likelihood (3), we suggest using the EM algorithm, originally developed by Dempster et al. (1977).

2.2. Maximization via the EM Algorithm

The use of the EM algorithm for obtaining parameter estimates in joint models was originally proposed by Wulfsohn and Tsiatis (1997) in the context of a Cox proportional hazards submodel for a survival outcome. The details for the EM algorithm with a logistic submodel for the response variable were given by Hwang et al. (2011). While we propose a similar set-up as that given by Hwang et al. (2011), our longitudinal submodel (1) is more general by allowing for $\mathbf{r}_{ijk}$ to be non-empty and also permitting multiple longitudinal covariates subject to DLs. In this section, we describe a more traditional EM algorithm approach for obtaining parameter estimates, and, in Section 2.3, we propose a new approximate version of the EM algorithm that significantly increases computational efficiency by reducing the dimension of integration to one in the E-step of the algorithm.

The EM algorithm is most useful for maximizing the observed-data log-likelihood in the presence of missing data, and it does this by iterating between two steps, the E-step and M-step. In the E-step, the expected value of the log-likelihood of the complete data is

calculated with respect to the missing data conditional on all of the observed data at a set of current parameter estimates. In the M-step, this expected log-likelihood is maximized to obtain new parameter estimates. In the context of the joint model, the observed data for each individual are $\{Y_i, \mathbf{X}_{i1}(t_{i1}), \dots, \mathbf{X}_{iq}(t_{iq}), \mathbf{Z}_i\}$, and the random effects in $\mathbf{b}_i$ are treated as missing data since these values are not observed. The expected value of the complete data log-likelihood in the E-step is

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(v)}) = \sum_{i=1}^n E \{ \log f(y_i|z_i, \mathbf{b}_i; \boldsymbol{\beta}) + \log f(\mathbf{x}_{i1}|z_i, \mathbf{b}_{i1}; \gamma_1, \tau_1^2) + \dots + \log f(\mathbf{x}_{iq}|z_i, \mathbf{b}_{iq}; \gamma_q, \tau_q^2) + \log f(\mathbf{b}_i; B, D) \}, \quad (4)$$

where $\boldsymbol{\theta}^{(v)} = (\boldsymbol{\beta}^{(v)T}, \boldsymbol{\gamma}^{(v)T}, \boldsymbol{\tau}^{2(v)T}, \mathbf{B}^{(v)T}, \text{vech}(D^{(v)T})^T)^T$ is the current set of parameter estimates and the expectation in (4), henceforth denoted as $E_i^{(v)}$ since it is different for each individual i , is with respect to the distribution $f(\mathbf{b}_i|y_i, \mathbf{x}_{i1}, \dots, \mathbf{x}_{iq}, z_i; \boldsymbol{\theta}^{(v)})$. To maximize (4), most of the parameters have closed form updates at each step in the EM algorithm. Specifically, we have that

$$\begin{aligned} \hat{B} &= \sum_{i=1}^n E_i^{(v)}(\mathbf{b}_i)/n, \\ \hat{D} &= \sum_{i=1}^n E_i^{(v)}(\mathbf{b}_i - \hat{B}^{(v)})(\mathbf{b}_i - \hat{B}^{(v)})^T / n, \\ \hat{\gamma}_k &= \sum_{i=1}^n (R_{ik}^T R_{ik})^{-1} \sum_{i=1}^n R_{ik}^T E_i^{(v)}(\mathbf{x}_{ik} - S_{ik} \mathbf{b}_{ik}), \quad k=1, \dots, q, \\ \hat{\tau}_k^2 &= \sum_{i=1}^n \sum_{j=1}^{n_{ik}} E_i^{(v)}(x_{ijk} - \mathbf{r}_{ijk}^T \hat{\gamma}_k - \mathbf{s}_{ijk}^T \mathbf{b}_{ik})^2 / \sum_{i=1}^n n_{ik}, \quad k=1, \dots, q, \end{aligned} \quad (5)$$

where $R_{ik} = [\mathbf{r}_{i1k}, \dots, \mathbf{r}_{in_{ik}k}]^T$ and $S_{ik} = [\mathbf{s}_{i1k}, \dots, \mathbf{s}_{in_{ik}k}]^T$ are the fixed and random effects design matrices for the i^{th} individual and k^{th} longitudinal covariate. There is no closed form update for the parameters in $\boldsymbol{\beta}$, which are of primary interest. Instead, at each iteration in the maximization procedure an update for $\boldsymbol{\beta}$ can be obtained by a one-step Newton-Raphson algorithm,

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^{(v)} - \left\{ Q''_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}^{(v)}) \right\}^{-1} Q'_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}^{(v)}), \quad (6)$$

where $Q'_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}^{(v)})$ is the vector of first partial derivatives of (4) with respect to $\boldsymbol{\beta}$ and $Q''_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}^{(v)})$ is the matrix of second partial derivatives of (4) with respect to $\boldsymbol{\beta}$, both evaluated at $\hat{\boldsymbol{\beta}}^{(v)}$.

Each estimate in (5) and (6) involves n expectations, which are taken with respect to the distributions $f(\mathbf{b}_i|y_i, \mathbf{x}_{i1}, \dots, \mathbf{x}_{iq}, z_i; \boldsymbol{\theta}^{(v)})$, $i = 1, \dots, n$. Wulfsohn and Tsiatis (1997) showed that these expectations can be calculated using Gauss-Hermite quadrature methods. However, the computational requirements using Gaussian quadrature grow exponentially as the dimension of $\mathbf{b}_i$ increases. When the dimension of $\mathbf{b}_i \geq 6$, it is generally recommended to compute the expectation using Monte Carlo methods (for example, James, 1980). However, with Monte Carlo methods, a large number of draws must be made from the distribution

$$f(\mathbf{b}_i|y_i, \mathbf{x}_{i1}, \dots, \mathbf{x}_{iq}, \mathbf{z}_i; \boldsymbol{\theta}^{(v)}) \propto f(y_i|\mathbf{z}_i, \mathbf{b}_i; \boldsymbol{\beta}^{(v)}) \left\{ \prod_{k=1}^q f(\mathbf{x}_{ik}|\mathbf{z}_i, \mathbf{b}_i; \boldsymbol{\gamma}_k^{(v)}, \tau_k^{2(v)}) \right\} f(\mathbf{b}_i; B^{(v)}, D^{(v)}).$$

2.3. Approximate EM Algorithm

When the number of longitudinal covariates or the number of random effects in (3) is high, both Gaussian quadrature methods and Monte Carlo methods can be extremely slow. For this reason, we propose approximating $f(\mathbf{b}_i|y_i, \mathbf{x}_{i1}, \dots, \mathbf{x}_{iq}, \mathbf{z}_i; \boldsymbol{\theta}^{(v)})$ by a multivariate normal distribution. Then, by taking advantage of the fact that linear combinations of $\mathbf{b}_i$ would also be normal, we show that the dimension of integration in our set-up can be reduced to one regardless of the number of random effects. Specifically, we consider that

$$\mathbf{b}_i|y_i, \mathbf{x}_{i1}, \dots, \mathbf{x}_{iq}, \mathbf{z}_i; \boldsymbol{\theta}^{(v)} \stackrel{a}{\sim} N(\hat{\mathbf{b}}_i, \hat{\Sigma}_i), \quad (7)$$

where $\stackrel{a}{\sim}$ indicates “is approximately distributed as,” $\hat{\mathbf{b}}_i = \operatorname{argmax}_{\mathbf{b}_i} \{\log f(y_i, \mathbf{x}_i, \mathbf{z}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(v)})\}$ and

$$\hat{\Sigma}_i = \left\{ - \frac{\partial^2 \log f(y_i, \mathbf{x}_i, \mathbf{z}_i, \mathbf{b}_i; \boldsymbol{\theta}^{(v)})}{\partial \mathbf{b}_i \partial \mathbf{b}_i^T} \bigg|_{\mathbf{b}_i = \hat{\mathbf{b}}_i} \right\}^{-1}.$$

That is, we propose approximating the conditional distribution for $\mathbf{b}_i$ using a multivariate normal centered at the posterior mode of $f(\mathbf{b}_i|y_i, \mathbf{x}_{i1}, \dots, \mathbf{x}_{iq}, \mathbf{z}_i; \boldsymbol{\theta}^{(v)})$ and scaled by the estimated variance of this mode estimate.

Heuristically, for a single longitudinal covariate, the normal approximation (7) follows because as $n_i \rightarrow \infty$, the density for $f(\mathbf{b}_i|y_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta})$, given by

$$f(\mathbf{b}_i|y_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta}) \propto f(y_i|\mathbf{z}_i, \mathbf{b}_i; \boldsymbol{\beta}) \left\{ \prod_{j=1}^{n_i} f(x_{ij}|\mathbf{z}_i, \mathbf{b}_i; \boldsymbol{\gamma}, \tau^2) \right\} f(\mathbf{b}_i; B, D), \quad (8)$$

is dominated by the term $\prod_{j=1}^{n_i} f(x_{ij}|\mathbf{z}_i, \mathbf{b}_i; \boldsymbol{\gamma}, \tau^2)$. More formally, Rizopoulos (2012a) noted that it can be shown that $f(\mathbf{b}_i|y_i, \mathbf{x}_i, \mathbf{z}_i; \boldsymbol{\theta})$ converges to a normal distribution as $n_i \rightarrow \infty$ using a variation of the Bayesian Central Limit Theorem. When there are multiple longitudinal covariates, asymptotic multivariate normality of $f(\mathbf{b}_i|y_i, \mathbf{x}_{i1}, \dots, \mathbf{x}_{iq}, \mathbf{z}_i; \boldsymbol{\theta})$ follows as $n_{ik} \rightarrow \infty \forall k = 1, \dots, q$. In the Supplementary Material, we provide some justification details for (7), though we note here that the approximation is reasonable for large n_i even when both $f(\mathbf{x}_i|\mathbf{z}_i, \mathbf{b}_i; \boldsymbol{\gamma}, \tau^2)$ and $f(\mathbf{b}_i; B, D)$ are not normal. However, when these distributions are normal, we have found in simulations that the approximation (7) works reasonably well even when n_{ik} is small (say $n_{ik} = 4$) $\forall k = 1, \dots, q$.

Normal approximations for the conditional posterior of $\mathbf{b}_i$ have been proposed previously in a variety of contexts, including linear and generalized linear mixed models (Baghishani and Mohammadzadeh, 2012) and joint models with a survival outcome (Rizopoulos, 2012a). In the latter paper, Rizopoulos (2012a) proposed approximating the conditional distribution of $\mathbf{b}_i$ as normal in order to eliminate the need to update quadrature points at each step in an EM algorithm for maximization when using adaptive Gauss-Hermite quadrature methods. In contrast, we propose a normal approximation for the conditional distribution of $\mathbf{b}_i$ in order to reduce the computational requirements when calculating expectations at each step in an EM algorithm. This approach is similar to the Laplace approximation method suggested by Rizopoulos et al. (2009), where the integrand of (4) is approximated using a second order Taylor expansion and then integrated with respect to a normal density, but with our approach the integrating distribution is approximated rather than the integrand itself. This strategy is somewhat more direct and does not require finding Taylor expansions.

With the normal approximation (7), the expectations in (5) and (6) are calculated with respect to one-dimensional normal distributions. In the update for γ_k , the expectation

$E_i(S_{ik}\mathbf{b}_{ik}) = E_i\{(s_{i1k}^T\mathbf{b}_{ik}, \dots, s_{in_{ik}k}^T\mathbf{b}_{ik})^T\}$ is computed component-wise with respect to the

one-dimensional normal distributions $N(s_{ijk}^T\hat{\mathbf{b}}_{ik}, s_{ijk}^T\sum_{ik}\hat{\mathbf{s}}_{ijk}, j = 1, \dots, n_{ik})$, where $\hat{\Sigma}_{ik}$ is the sub-matrix of $\hat{\Sigma}_i$ associated with the random effects vector $\mathbf{b}_{ik}$. Similarly, in the update for

τ_k^2 , the expectations $E_i(r_{ijk}^T\hat{\gamma}_k + s_{ijk}^T\mathbf{b}_{ik}), j = 1, \dots, n_{ik}$, are calculated with respect to the one-

dimensional normal distributions $N(r_{ijk}^T\hat{\gamma}_k + s_{ijk}^T\hat{\mathbf{b}}_{ik}, s_{ijk}^T\sum_{ik}\hat{\mathbf{s}}_{ijk}), j = 1, \dots, n_{ik}$. We can also show that in the update for β (refer to the online Supplementary Material for details), we

need to obtain the expectation $E_i[\log\{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1^T\mathbf{Z}_i + \hat{\beta}_2^T\mathbf{b}_i)\}]$, which using (7) is with

respect to the one-dimensional normal distribution $N(\hat{\beta}_0 + \hat{\beta}_1^T\mathbf{Z}_i + \hat{\beta}_2^T\hat{\mathbf{b}}_i, \hat{\beta}_2^T\sum_i\hat{\beta}_2)$. Most simply, with the normal approximation, updates for B and D do not require any integration since $E_i(\mathbf{b}_i) = \hat{\mathbf{b}}_i$ and $E_i(\mathbf{b}_i - B)(\mathbf{b}_i - B)^T = \hat{\Sigma}_i + (\hat{\mathbf{b}}_i - B)(\hat{\mathbf{b}}_i - B)^T$. Gauss-Hermite quadrature or Monte Carlo methods can be applied to quickly calculate each of the one-dimensional expectations.

To summarize, we propose the following algorithm for obtaining approximate maximum likelihood estimates, with calculation tips based on R software (R Core Team, 2012):

1. Using all the data, $\{y_i, \mathbf{x}_{i1}(t_{i1}), \dots, \mathbf{x}_{iq}(t_{iq}), \mathbf{z}_i\}$, $i = 1, \dots, n$, obtain an initial estimate for $\theta, \theta^{\hat{1}}$. This can be done by fitting separate linear mixed models for each longitudinal covariate to get estimates for γ_k and τ_k^2 , $k = 1, \dots, q$. The matrix D can be estimated as the diagonal matrix of the individually estimated D_k , $k = 1, \dots, q$. Initial estimates for β can be obtained by fitting a logistic model for Y_i based on $\mathbf{Z}_i$ and the posterior estimates for the random effects from the fitted linear mixed models. With R, the “lmer” function in the “lme4” package (Bates et al., 2012) can be used to obtain the linear mixed model parameter estimates and the “glm” function can be used to obtain the logistic model parameter estimates.

2. Maximize $f(\mathbf{b}_i|y_i, \mathbf{x}_{i1}, \dots, \mathbf{x}_{iq}, z_i; \hat{\boldsymbol{\theta}}^1)$ with respect to $\mathbf{b}_i, i = 1, \dots, n$, in order to obtain $\hat{\mathbf{b}}_i$ and $\hat{\Sigma}_i$. With R, the “nlm” or “optim” function can be used to obtain these estimates.
3. Update $\hat{\boldsymbol{\theta}}^1$ as $\hat{\boldsymbol{\theta}}^2$ by using (5) and (6), obtaining the expected values by assuming that $\mathbf{b}_i|y_i, \mathbf{x}_{i1}, \dots, \mathbf{x}_{iq}, z_i; \boldsymbol{\theta}^{(v)} \stackrel{a}{\sim} N(\hat{\mathbf{b}}_i, \hat{\Sigma}_i), i = 1, \dots, n$.
4. Repeat Steps 2–3, each time obtaining $\hat{\boldsymbol{\theta}}^{(v+1)}$ based on $\hat{\boldsymbol{\theta}}^{(v)}$.
5. When $\hat{\boldsymbol{\theta}}^{(v+1)}$ is sufficiently close to $\hat{\boldsymbol{\theta}}^{(v)}$, define $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{(v+1)}$ as the vector of approximate maximum likelihood estimates.

In Step 5, we suggest defining consecutive estimates for $\boldsymbol{\theta}$ as close when

$$\max \left\{ \frac{|\theta_1^{(v+1)} - \theta_1^{(v)}|}{|\theta_1^{(v)}|}, \dots, \frac{|\theta_s^{(v+1)} - \theta_s^{(v)}|}{|\theta_s^{(v)}|} \right\} < \varepsilon, \quad (9)$$

where s is the number of elements in $\boldsymbol{\theta}$ and ε is a predefined distance. In practice, we have found that this approximate EM algorithm works well. However, this algorithm could also be used to quickly find an initial estimate for $\boldsymbol{\theta}$, which could then be used as the starting value in an EM algorithm where the true distribution $f(\mathbf{b}_i|y_i, \mathbf{x}_{i1}, \dots, \mathbf{x}_{iq}, z_i; \boldsymbol{\theta})$ is used in each E-step.

While several methods are available for estimating the variances of the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$, we suggest using the approach proposed by Rizopoulos (2012b) based on directly calculating derivatives of the observed-data likelihood functions. To obtain the variance estimate, first note that the observed-data score can be represented by

$$S(\boldsymbol{\theta}) = \sum_{i=1}^n \int_{-\infty}^{\infty} A(\boldsymbol{\theta}, \mathbf{b}_i) f(\mathbf{b}_i|y_i, \mathbf{x}_{i1}, \dots, \mathbf{x}_{iq}, z_i; \boldsymbol{\theta}) d\mathbf{b}_i, \quad (10)$$

where

$$A(\boldsymbol{\theta}, \mathbf{b}_i) = \frac{\partial}{\partial \boldsymbol{\theta}^T} \{ \log f(y_i|z_i, \mathbf{b}_i; \boldsymbol{\beta}) + \log f(\mathbf{x}_{i1}|z_i, \mathbf{b}_{i1}; \gamma_1, \tau_1^2) + \dots + \log f(\mathbf{x}_{iq}|z_i, \mathbf{b}_{iq}; \gamma_q, \tau_q^2) + \log f(\mathbf{b}_i; B, D) \}.$$

Then, the contribution to the Hessian matrix by the i^{th} individual can be written as

$$\frac{\partial S_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \int_{-\infty}^{\infty} \frac{\partial A(\boldsymbol{\theta}, \mathbf{b}_i)}{\partial \boldsymbol{\theta}^T} f(\mathbf{b}_i|y_i, \mathbf{x}_{i1}, \dots, \mathbf{x}_{iq}, z_i; \boldsymbol{\theta}) d\mathbf{b}_i + \int_{-\infty}^{\infty} A(\boldsymbol{\theta}, \mathbf{b}_i) \{A(\boldsymbol{\theta}, \mathbf{b}_i) - S_i(\boldsymbol{\theta})\}^T f(\mathbf{b}_i|y_i, \mathbf{x}_{i1}, \dots, \mathbf{x}_{iq}, z_i; \boldsymbol{\theta}) d\mathbf{b}_i. \quad (11)$$

Finally, we can calculate the variance of the parameter estimates as

$$\hat{\text{Var}}(\hat{\theta}) = \left\{ - \sum_{i=1}^n \frac{\partial S_i(\theta)}{\partial \theta} \bigg|_{\theta=\hat{\theta}} \right\}^{-1}. \quad (12)$$

We suggest using the normal approximation (7) to approximate the integrals in (10) and (11), though since these calculations only need to be made once, Monte Carlo integration based on the true distribution $f(\mathbf{b}_i|y_i, \mathbf{x}_{i1}, \dots, \mathbf{x}_{iq}, \mathbf{z}_i; \theta)$ is also computationally reasonable. Since the estimates obtained via the proposed approximate EM algorithm are not technically maximum likelihood estimates, a sandwich variance estimator may be obtained as

$$\hat{\text{Var}}(\theta) = \left\{ - \sum_{i=1}^n \frac{\partial S_i(\theta)}{\partial \theta} \bigg|_{\theta=\hat{\theta}} \right\}^{-1} \left\{ \sum_{i=1}^n S_i(\hat{\theta}) S_i(\hat{\theta})^T \right\} \left[\left\{ - \sum_{i=1}^n \frac{\partial S_i(\theta)}{\partial \theta} \bigg|_{\theta=\hat{\theta}} \right\}^{-1} \right]^T. \quad (13)$$

2.4. Convergence of the Approximate EM Algorithm

Since the proposed method involves approximating the E-step of the EM algorithm, unless $n_i \rightarrow \infty$, the resulting estimator will not be exactly the same as the maximum likelihood estimate. However, using the results of Wu (1983), it can still be shown that the proposed approximate EM algorithm converges to a stationary point. Wu (1983) showed that under some typical regularity conditions, the algorithm will converge as long as $Q(\theta|\theta)$ is continuous in both θ' and θ . This condition is straightforward to show for (4) where the expected value is taken with respect to $N(\mathbf{b}_i, \hat{\Sigma})$ instead of $f(\mathbf{b}_i|y_i, \mathbf{x}_{i1}, \dots, \mathbf{x}_{iq}, \mathbf{z}_i; \theta)$.

3. Joint Modeling of a Binary Response and Multiple Covariates Subject to Detection Limits

When the longitudinal covariates are subject to DLs, the joint model described in Section 2 no longer applies. While it is common in practice, in a variety of contexts, to simply replace covariate values which are censored by a function of the DL, such as DL/2, Helsel (2012) and many others have shown that this leads to biased parameter and variance estimates in many modeling situations.

In the following, we recommend an extension of the joint likelihood approach described in Section 2 to incorporate the censoring on the longitudinal covariates in the linear mixed submodels. We also propose an adaptation of the approximate EM algorithm described in Section 2.3 that is still much faster than a traditional EM approach.

3.1. Longitudinal Model with Censoring

Suppose we observe independent samples $\{Y_i, \mathbf{X}_{i1}^*(t_{i1}), \dots, \mathbf{X}_{iq}^*(t_{iq}), \rho_{i1}, \dots, \rho_{iq}, \mathbf{Z}_i\}$, $i = 1, \dots, n$, where Y_i is a binary response,

$\mathbf{X}_{ik}^*(t_{ik}) = (\max\{X_{i1k}(t_{i1k}), d_k\}, \dots, \max\{X_{in_{ikk}}(t_{in_{ikk}}), d_k\})^T$, $k = 1, \dots, q$, are longitudinal covariate vectors with n_{ik} repeated observations at times $\mathbf{t}_{ik} = (t_{i1k}, \dots, t_{in_{ikk}})^T$, $\rho_{ik} = (\rho_{i1k}, \dots, \rho_{in_{ikk}})^T$, $k = 1, \dots, q$, are vectors of censoring indicators, $\mathbf{Z}_i$ is a $(p - q)$ -dimensional vector of baseline covariates, and $\mathbf{d} = (d_1, \dots, d_q)^T$ is a vector of lower DLs

corresponding to the q longitudinal covariates. For the same reason as stated in Section 2.1, we generally refer to $\mathbf{X}_{ik}^*(t_{ik})$ simply as $\mathbf{X}_{ik}^*$. Additionally, while in this paper we focus on lower DLs, the methods we propose can easily be extended to handle upper DLs.

We continue to assume the longitudinal and logistic submodels (1) and (2) so that the joint likelihood for the longitudinal covariates and binary response can still be represented by (3). However, we propose that the contribution to the likelihood for the k^{th} longitudinal covariate should be represented by

$$f(\mathbf{x}_{ik}^* | \boldsymbol{\rho}_{ik}, \mathbf{z}_i, \mathbf{b}_{ik}; \boldsymbol{\gamma}_k, \tau_k^2) = \prod_{j=1}^{n_{ik}} \left\{ \frac{1}{\tau_k} \phi \left(\frac{\pi_{ijk} - \mathbf{r}_{ijk}^T \boldsymbol{\gamma}_k - \mathbf{s}_{ijk}^T \mathbf{b}_{ik}}{\tau_k} \right) \right\}^{\rho_{ijk}} \times \left\{ \Phi \left(\frac{d_k - \mathbf{r}_{ijk}^T \boldsymbol{\gamma}_k - \mathbf{s}_{ijk}^T \mathbf{b}_{ik}}{\tau_k} \right) \right\}^{1-\rho_{ijk}}, \quad (14)$$

where ϕ is the standard normal density and Φ is the cdf of the standard normal distribution. That is, we propose using the cdf of the standard normal to represent the contribution to the overall likelihood for all observations that are observed to be below the DLs. This approach has been taken by numerous authors in the context of linear mixed models with censored responses (Pettitt, 1986; Carriquiry et al., 1987; Lyles et al., 2000; Vock et al., 2012, among others).

3.2. Approximate EM Algorithm with Covariates Subject to Detection Limits

The likelihood given by (3) together with (14) can be maximized using an EM algorithm very similar to that described in Section 2.2. The main difference is that there are no closed form representations for $\boldsymbol{\gamma}$ and $\boldsymbol{\tau}^2$ at each step in the algorithm since the part of the likelihood involving these parameters includes the standard normal cdf. However, an update for $(\boldsymbol{\gamma}, \boldsymbol{\tau}^2)$ may still be obtained by using the one-step Newton-Raphson algorithm,

$$\begin{pmatrix} \hat{\boldsymbol{\gamma}} \\ \hat{\boldsymbol{\tau}}^2 \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{\gamma}}^{(v)} \\ \hat{\boldsymbol{\tau}}^{2(v)} \end{pmatrix} - \left\{ Q''_{\boldsymbol{\gamma}, \boldsymbol{\tau}^2}(\hat{\boldsymbol{\gamma}}^{(v)}, \hat{\boldsymbol{\tau}}^{2(v)}) \right\}^{-1} \begin{pmatrix} Q'_{\boldsymbol{\gamma}}(\hat{\boldsymbol{\gamma}}^{(v)}, \hat{\boldsymbol{\tau}}^{2(v)}) \\ Q'_{\boldsymbol{\tau}^2}(\hat{\boldsymbol{\gamma}}^{(v)}, \hat{\boldsymbol{\tau}}^{2(v)}) \end{pmatrix}, \quad (15)$$

where $Q'_{\boldsymbol{\gamma}}(\hat{\boldsymbol{\gamma}}^{(v)}, \hat{\boldsymbol{\tau}}^{2(v)})$ and $Q'_{\boldsymbol{\tau}^2}(\hat{\boldsymbol{\gamma}}^{(v)}, \hat{\boldsymbol{\tau}}^{2(v)})$ are the vectors of first partial derivatives of (4) with respect to $\boldsymbol{\gamma}$ and $\boldsymbol{\tau}^2$, respectively, and $Q''_{\boldsymbol{\gamma}, \boldsymbol{\tau}^2}(\hat{\boldsymbol{\gamma}}^{(v)}, \hat{\boldsymbol{\tau}}^{2(v)})$ is the matrix of second partial derivatives of (4) with respect to $\boldsymbol{\gamma}$ and $\boldsymbol{\tau}^2$, all of which are evaluated at $(\hat{\boldsymbol{\gamma}}^{(v)}, \hat{\boldsymbol{\tau}}^{2(v)})$. Refer to the online Supplementary Material for the derivation of this updating formula and additional details.

As before, in order to compute the expectations in the E-step of the EM algorithm, Monte Carlo integration may be used. While Gaussian quadrature methods may also be applied, the conditional distribution $f(\mathbf{b}_i | \mathbf{x}_{i1}^*, \dots, \mathbf{x}_{iq}^*, \boldsymbol{\rho}_{i1}, \dots, \boldsymbol{\rho}_{iq}, \mathbf{z}_i)$ is no longer normal, and thus is less convenient and may require more evaluation points for similar accuracy.

With either Monte Carlo or Gaussian quadrature integration methods, calculating the expectations is very slow for high-dimensional random effects. For this reason, we propose using the normal approximation suggested in Section 2.3. With a single longitudinal

covariate subject to censoring, the normal approximation follows when the number of uncensored values grows large compared to the number of censored values, since then the density for $f(\mathbf{b}_i|y_i, \mathbf{x}_i^*, \boldsymbol{\rho}_i, \mathbf{z}_i; \boldsymbol{\theta})$ is dominated by the term $\prod_{j=1}^{n_i} f(x_{ij}|\mathbf{z}_i, \mathbf{b}_i)$. With multiple longitudinal covariates, the normal approximation follows as the number of uncensored values grows large for each of the covariates. Unfortunately, for those individuals with very few or no observations above the DLs, $f(\mathbf{b}_i|y_i, \mathbf{x}_i^*, \boldsymbol{\rho}_i, \mathbf{z}_i; \boldsymbol{\theta})$ will not necessarily be approximately normal. Thus, we must modify the approximate EM algorithm described in Section 2.3. Specifically, we propose using the normal approximation when reasonable and Monte Carlo integration for those individuals with very few covariates observed above the DLs.

To empirically get a better idea at how the normal approximation performs at various levels of censoring, we plotted the marginal posterior distributions of the random effects in a few scenarios. Figure 1 displays the marginal distributions of $f(\mathbf{b}_i|y_i, \mathbf{x}_i^*, \boldsymbol{\rho}_i, \mathbf{z}_i; \boldsymbol{\theta})$ for a case with a single covariate and a random intercept and a random slope, b_0 and b_1 , for varying degrees of censoring. To generate the distributions in Figure 1, we considered data from individuals in a data set simulated as described in Section 4 with the true parameter values plugged into $f(\mathbf{b}_i|y_i, \mathbf{x}_i^*, \boldsymbol{\rho}_i, \mathbf{z}_i; \boldsymbol{\theta})$. We see that even for high degrees of censoring the normal approximation appears quite reasonable. However, when all longitudinal observations are censored at the DLs, the marginal distribution of the random slope conditional on the data becomes very skewed. In the online Supplementary Material, we supply additional examples when three covariates are in the model, each having a random intercept and a random slope for a total of six random effects. The normal approximation again seems reasonable when there are at least a few covariate values observed above the DLs, even if no values for one or two of the covariates are observed to be above the DLs.

Based on our experience and empirical examples such as those displayed in Figure 1, we recommend using the normal approximation if at least 20% of the total number of longitudinal observations for an individual are observed above the DLs, regardless of the number of longitudinal covariates q . Then, to summarize, we propose the following algorithm for obtaining approximate maximum likelihood estimates in the presence of censoring on the longitudinal covariates, with calculation tips based on R software (R Core Team, 2012):

1. Using all the data, $\{y_i, \mathbf{x}_{i1}^*(t_{i1}), \dots, \mathbf{x}_{iq}^*(t_{iq}), \boldsymbol{\rho}_{i1}, \dots, \boldsymbol{\rho}_{iq}, \mathbf{z}_i\}$, $i = 1, \dots, n$, obtain an initial estimate for $\boldsymbol{\theta}$, $\boldsymbol{\theta}^{(1)}$. This can be done by fitting separate censored linear mixed models for each longitudinal covariate to get estimates for γ_k and τ_k^2 , $k = 1, \dots, q$. The matrix D can be estimated as the diagonal matrix of the individually estimated D_k , $k = 1, \dots, q$. Initial estimates for $\boldsymbol{\beta}$ can be obtained by fitting a logistic model for Y_i based on $\mathbf{Z}_i$ and the posterior estimates for the random effects from the fitted censored linear mixed models. With R, the “lme4” package (Vaida and Liu, 2012) can be used to obtain the censored linear mixed model parameter estimates and the “glm” function can be used to obtain the logistic model parameter estimates.

2. For those individuals with at least 20% of the total number of longitudinal observations observed above the DLs, maximize

$f(\mathbf{b}_i|y_i, \mathbf{x}_{i1}^*, \dots, \mathbf{x}_{iq}^*, \boldsymbol{\rho}_{i1}, \dots, \boldsymbol{\rho}_{iq}, z_i; \hat{\boldsymbol{\theta}}^{(1)})$ with respect to $\mathbf{b}_i$, $i = 1, \dots, n$, in order to obtain $\hat{\mathbf{b}}_i$ and $\hat{\Sigma}_i$. With R, the “nlm” or “optim” function can be used to obtain these estimates. For those individuals with less than 20% of the total number of longitudinal observations above the DLs, obtain a sample from

$f(\mathbf{b}_i|y_i, \mathbf{x}_{i1}^*, \dots, \mathbf{x}_{iq}^*, \boldsymbol{\rho}_{i1}, \dots, \boldsymbol{\rho}_{iq}, z_i; \hat{\boldsymbol{\theta}}^{(1)})$ using Monte Carlo sampling techniques. With R, the function “Metrop1R” in the package “MCMCpack” (Martin et al., 2011) can be used to obtain these samples.

3. Update $\hat{\boldsymbol{\theta}}^{(1)}$ as $\hat{\boldsymbol{\theta}}^{(2)}$ by using (5) for B and D , (6) for $\boldsymbol{\beta}$, and (15) for γ and τ^2 , obtaining the expected values by assuming $\mathbf{b}_i|y_i, \mathbf{x}_{i1}^*, \dots, \mathbf{x}_{iq}^*, \boldsymbol{\rho}_{i1}, \dots, \boldsymbol{\rho}_{iq}, z_i$; $\hat{\boldsymbol{\theta}}^{(1)} \stackrel{a}{\sim} N(\hat{\mathbf{b}}_i, \hat{\Sigma}_i)$ for each individual i with at least 20% of the total number of longitudinal observations observed above the DLs and using Monte Carlo integration methods based on the true posterior distribution otherwise.
4. Repeat Steps 2–3, each time obtaining $\hat{\boldsymbol{\theta}}^{(v+1)}$ based on $\hat{\boldsymbol{\theta}}^{(v)}$.
5. When $\hat{\boldsymbol{\theta}}^{(v+1)}$ is sufficiently close to $\hat{\boldsymbol{\theta}}^{(v)}$, define $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{(v+1)}$ as the vector of approximate maximum likelihood estimates.

While in our experience this algorithm has worked well in practice, if desired, the normal approximation can be suspended after a certain number of iterations and the EM algorithm can be completed using the true conditional distribution of the random effects. This should generally be faster than using the EM algorithm without the approximation, and the final estimate for $\boldsymbol{\theta}$ will be the maximum likelihood estimate. Variance estimation can be conducted as described in Section 2.3, but for those individuals with greater than 80%

censoring, we suggest using the true distribution $f(\mathbf{b}_i|y_i, \mathbf{x}_{i1}^*, \dots, \mathbf{x}_{iq}^*, \boldsymbol{\rho}_{i1}, \dots, \boldsymbol{\rho}_{iq}, z_i; \hat{\boldsymbol{\theta}}^{(v)})$ to calculate the integrals in (10) and (11). Since the variance only needs to be calculated once, rather than at each step in the EM algorithm, the additional computation requirements are minimal.

4. Simulation Study

We conducted a series of simulations to assess the performance of the proposed approximate EM algorithm with and without censoring on the covariates. In Section 4.1, we use numerical studies to assess the effectiveness of the proposed approximate EM algorithm when there is a single longitudinal covariate while in Section 4.2 we use numerical studies to investigate the approximate EM algorithm in the case when there are three longitudinal covariates. In Section 4.3, we discuss the computational advantages and limitations of the proposed method as observed in the simulations. Finally, in Section 4.4, we assess the effect of reducing the number of repeated observations and changing the true distribution of the longitudinal covariates.

4.1. Joint Model with One Longitudinal Covariate

We generated 500 data sets of 500 individuals with a binary response, baseline covariate, and up to eight repeated observations on a single longitudinal covariate. We generated data to mimic the GenIMS data set that we analyze in Section 5. For each individual, we first generated a baseline “age” covariate Z_i as $\text{beta}(3, 2) \cdot 84 + 18$. We then generated a random number from one to eight, each with equal probability, corresponding to the number of observations for an individual. We also generated random intercepts and slopes as

$$\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N \left\{ \begin{pmatrix} 1 \\ 0.2 \end{pmatrix}, \begin{pmatrix} 1 & 0.2 \\ 0.2 & 0.3 \end{pmatrix} \right\}.$$

We then let $X_{ij} \sim N(Z_i\gamma + b_{0i} + b_{1i}t_{ij}, \tau^2)$, where $\gamma = 1/70$, $\tau^2 = 1$, and $0 \leq t_{ij} \leq 7$ is the j^{th} observed time for the i^{th} individual. To mimic the GenIMS data set, which has unbalanced repeated measurements, we randomly deleted 10% of the longitudinal observations. Finally, we generated a binary response by $Y_i|Z_i, b_{0i}, b_{1i} \sim \text{Bernoulli}[\{1 + \exp(-(1, Z_i)\beta_1 - (b_{0i}, b_{1i})\beta_2)\}]$, where $\beta_1 = (-2, 0.035)^T$ and $\beta_2 = (0.3, -0.45)^T$.

Table 1 gives the bias, empirical standard deviation, average standard error estimate, and empirical 95% coverage probability for each of the parameters in β and the bias and empirical standard deviation for γ , τ^2 , B , and D . We only report standard errors and coverage probabilities for the parameters in β since they are of primary interest. The standard errors are estimated using (12) rather than (13) because we found that the former performed slightly better. We emphasize that there is no censoring on the covariates for the results in Table 1.

None of the parameter estimates have statistically significant biases at the $\alpha = 0.01$ level, where the standard error of $\hat{\beta}$ is calculated as $\text{SD}/\sqrt{500}$. Standard error estimation for the parameters in β seems good and empirical coverage probabilities are maintained near the 95% level.

We repeated the simulation as described above, but we introduced a DL d so that 20%, 40%, and 60% of the longitudinal observations were censored below d . Table 2 gives the bias, empirical standard deviation, average standard error estimate, and empirical 95% coverage probability for each of the parameters in β as well as the bias and empirical standard deviation for the remaining parameters for simulations with 20%, 40%, and 60% censoring due to DLs.

The estimates for all the parameters in β are unbiased at the $\alpha = 0.01$ level, though, with the exception of τ^2 and D_{12} for 20% censoring and D_{12} for 60% censoring, each of the remaining parameter estimates is slightly biased. Standard error estimates and empirical coverage probabilities for the parameters in β are good for all the levels of censoring. Thus, while several of the parameter estimates are slightly biased, estimation and inference for the parameters of main interest seems satisfactory.

For comparison, Table 3 gives the bias in parameter estimates when censored observations are replaced by the DL or by DL/2 and a full-data analysis is conducted, as is often done in practice. Statistically significant biases exist for all the parameters with the exception of the β_{12} parameter associated with the uncensored covariate Z_i . The biases generally are more severe for higher levels of censoring and when using DL versus DL/2.

We also conducted simulations to assess the effect of increasing the random effect variances in relation to the variance of the repeated observations. As before, we defined $\tau = 1$, but we considered two additional cases for D : $\text{vech}(D) = (3, 1, 3)^T$ and $\text{vech}(D) = (10, 3, 5)^T$. We provide the results of these simulations in the online Supplementary Material, though we note here that the findings regarding bias, standard error estimation, and coverage probability are largely the same as observed above.

4.2. Joint Model with Three Longitudinal Covariates

We conducted a second set of simulations to assess the performance of the approximate EM algorithm in the presence of multiple longitudinal covariates, since, in practice, the proposed normal approximation described in Section 2.3 is most useful when the dimension of the random effects is high. We generated 300 data sets of 500 individuals with a binary response, baseline covariate, and up to eight repeated observations on three longitudinal covariates. We again modeled this simulation after the GenIMS data set, first generating a baseline “age” covariate and number of observations for each longitudinal covariate as in Section 4.1. We then generated an intercept and slope random effect for each of the three covariates as $\mathbf{b}_i \sim N(\mathbf{B}, D)$, with $\mathbf{B} = (1, 0.2, 3, -0.1, -2, 0.3)^T$ and $D_{11} = D_{33} = D_{55} = 1$, $D_{22} = D_{44} = D_{66} = 0.3$, $D_{12} = D_{21} = D_{34} = D_{43} = D_{56} = D_{65} = 0.05$, and all other elements of D set equal to 0.1, where D_{ij} is the element in the i^{th} row and j^{th} column of D . The first and second elements of $\mathbf{B}$ correspond to the first longitudinal covariate, the third and fourth to the second longitudinal covariate, and the fifth and sixth to the third longitudinal covariate.

We let $X_{ijk} \sim N((1, t_{ijk})\mathbf{b}_{ik}, \tau_k^2)$, $i = 1, \dots, n$, $j = 1, \dots, n_{ik}$, $k = 1, 2, 3$, where $\tau^2 = (1, 1.5, 1.25)^T$ and $0 \leq t_{ijk} \leq 7$ is the j^{th} observed time for the i^{th} individual and k^{th} covariate. To mimic the GenIMS data, we randomly deleted 10% of all the observations at a given time for an individual and an additional 5% of the remaining observed longitudinal observations. Finally, we generated a binary response by

$Y_i | Z_i, \mathbf{b}_i \sim \text{Bernoulli}[\{1 + \exp(-(1, Z_i)\beta_1 - \mathbf{b}_i^T \beta_2)\}]$, where $\beta_1 = (-1, 0.01)^T$ and $\beta_2 = (0.2, -0.1, 0.4, -0.3, 0.25, -0.3)^T$.

Table 4 gives the bias, empirical standard deviation, average standard error estimate, and empirical 95% coverage probability for each of the parameters in β and the bias and empirical standard deviation for each of the remaining parameters except D , which we do not report due to the large number of parameters in D , though we note the estimator generally performs well. As before, we only report standard error estimates and coverage probabilities for the parameters in β since they are of primary interest. We note that for the results in Table 4 none of the covariates are subject to censoring.

All of the parameter estimates in Table 4 are unbiased at the $\alpha = 0.01$ level except β_{23} , which has a slightly significant bias. The standard error estimates and coverage probabilities are good for all of the parameters in β , including β_{23} .

We repeated the simulation as described above, but we introduced censoring due to DLs on each of the longitudinal covariates. We defined the DLs d_1 , d_2 , and d_3 so that censoring on the three covariates corresponded to approximately 25%, 12.5%, and 37.5%, respectively, for a total censoring rate of 25%, as well as approximately 50%, 25%, and 75%, respectively, for a total censoring rate of 50%.

Table 5 gives the bias, empirical standard deviation, average standard error estimate, and empirical 95% coverage probability for each of the parameters in β and the bias and empirical standard deviation for the remaining parameters except D for simulations with 25% and 50% total censoring due to DLs. Table 5 also contains estimates for a full-data analysis where censored observations are replaced by the DL or DL/2. Note that obtaining convergence was challenging when replacing censored covariates with the DL for 50% censoring, and thus these results are not included in the table.

For our proposed algorithm, all of the parameters in β are unbiased at the 0.01 α -level, with the exception of the intercept in the model with 25% censoring. While most of the parameters in τ^2 and B have statistically significant biases, they are generally small relative to the size of the true parameter values. Standard error estimates and coverage probabilities are fairly good for both levels of censoring. All the parameters are biased when replacing censored values with the DL or DL/2.

4.3. Computational Advantages and Limitations

All of simulations above were conducted using the statistical software program R (R Core Team, 2012) on a computer with an Intel Core i7 870 @ 2.93 GHz processor and 8 GB of RAM. Based on this software, we compared the computational times required for fitting the joint models in the simulations using our proposed approximation versus Monte Carlo integration in the E-step of the EM algorithm. For the case when there is a single longitudinal covariate with no censoring, the proposed approximate EM algorithm ran for 2.0 minutes per data set on average, not including estimation of the initial values or variances, which is about 101.7 times faster than an EM algorithm based on Monte Carlo integration using 300 independent random draws. When 20%, 40%, or 60% of the longitudinal covariate values are censored, the proposed EM algorithm ran for 21.3, 55.8, and 99.8 minutes per data set, respectively, which is about 10.1, 3.9, and 2.1 times faster than using Monte Carlo integration. For the case when there are three longitudinal covariates and 0%, 25%, or 50% total censoring, the proposed EM algorithm ran for 11.0, 23.1, and 78.8 minutes per data set, respectively, which is about 54.1, 26.3, and 8.0 times faster than using Monte Carlo integration.

In addition to these computational advantages, the proposed method does have some computational challenges. First, while finding starting values for the parameters in the model is fairly automatic when using the R functions outlined in Sections 2.3 and 3.2, issues do rarely arise when finding starting values for maximizing

$f(\mathbf{b}_i | y_i, \mathbf{x}_{i1}^*, \dots, \mathbf{x}_{iq}^*, \boldsymbol{\rho}_{i1}, \dots, \boldsymbol{\rho}_{iq}, z_i; \hat{\boldsymbol{\theta}})$ with respect to $\mathbf{b}_i$. However, we have found these issues usually only occur when at least one element of $\text{diag}(\mathbf{D})$ is fairly large and can be easily solved by manually testing reasonable starting values or by attempting alternative optimization algorithms such as those provided by the R function “optim”. Another computational issue we faced was determining convergence of the algorithm. We assumed estimates had converged when (9) was satisfied with $\varepsilon = 0.005$. However, with censored covariates, the number of iterations of the EM algorithm until convergence was occasionally large (>30). We employed two strategies to reduce the necessary computation as much as possible. First, for those individuals with $>80\%$ censoring, we used the technique recommended by Wei and Tanner (1990) for implementing a Monte Carlo EM algorithm where the number of sample points used to estimate (4) starts small and increases with each iteration. Second, we redefined $\varepsilon = 0.01$ for the non- $\boldsymbol{\beta}$ parameters. Under these settings, for the simulations with one covariate, an average of 13.3, 21.2, 24.3, and 25.3 iterations of the approximate EM algorithm were needed at the 0%, 20%, 40%, and 60% censoring levels, respectively, while for the simulations with three covariates, an average of 19.1, 21.0, and 22.1 iterations were needed at the 0%, 25%, and 50% censoring levels, respectively.

4.4. Robustness to Failure of Assumptions

The proposed approximate EM algorithm method for maximizing the joint likelihood of a binary response and longitudinal covariates relies heavily on two assumptions regarding the data. First, the number of repeated observations for each individual is assumed to be relatively large, and second, the distribution of the repeated covariate measurements is assumed to be normal about the individual’s longitudinal trajectory.

To assess the effect of this first assumption failing, we repeated the simulations with a single longitudinal covariate described in Section 4.1, but restricted the number of repeated observations for each individual to be at most four. While the earlier simulation only averaged 4.5 observations per individual, the number of observations per individual here only averaged 2.55, and some of these observations may be unknown due to DLs.

Table 6 gives the bias, empirical standard deviation, average standard error estimate, and empirical 95% coverage probability for each of the parameters in $\boldsymbol{\beta}$ and the bias and empirical standard deviation for each of the remaining parameters in the case with no censoring. Additional results when censoring occurs due to DLs are provided in the online Supplementary Material. All of the parameters in $\boldsymbol{\beta}$ except β_{12} have small statistically significant biases when no censoring is present on the covariates. Additionally, τ^2 , B_2 , and D_{12} have small observed biases. Interestingly, when censoring is present, the size of the biases remains essentially the same for the $\boldsymbol{\beta}$ of main interest, but is larger for the remaining parameters. Standard error estimates are still reasonable and coverage probabilities are close to 95%.

While biases in the parameter estimates do arise in this simulation scenario, we note that this example is a fairly extreme violation of the assumption of a large number of observations per individual. If these biases are not considered tolerable, a more traditional Monte Carlo EM algorithm could be employed. While more traditional methods may be computationally

intensive, with so few observations per individual it may be reasonable to simply model an intercept random effect, thus reducing the number of random effect terms.

To assess the robustness to the second major assumption regarding normality of repeated covariate measurements, we repeated the simulations in Section 4.1 but instead of assuming $X_{ij} \sim N(Z_i\gamma + b_{0i} + b_{1i}t_{ij}, \tau^2)$ we let $X_{ij} \sim \text{Gamma}(1, \tau^2) + Z_i\gamma + b_{0i} + b_{1i}t_{ij}$ where τ^2 is a scale parameter so that the mean and the variance of the X_{ij} 's remain the same but the distribution is very skewed right about the longitudinal trajectory.

Table 7 gives the bias, empirical standard deviation, average standard error estimate, and empirical 95% coverage probability for each of the parameters in β and the bias and empirical standard deviation for each of the remaining parameters in the case of no censoring. Additional results when censoring occurs due to DLs are provided in the online Supplementary Material. All of the parameters in β except β_{12} have small statistically significant biases when no censoring is present. However, none of the additional parameters have observed biases. When censoring is present, the size of the biases remains similar but increases gradually with the degree of censoring due to DLs. Standard error estimates are still reasonable and coverage probabilities are close to 95%.

This simulation indicates that a failure of the assumption of normally distributed covariate measurements is not very costly when there are a moderate number of longitudinal observations per individual. If this assumption was a concern, it could be checked by plotting residuals for the longitudinal observations about their estimated trajectory, and if it fails, either a transformation could be applied or a different distribution could be assumed for the repeated covariate measurements. Since the normal approximation described in Section 2.3 does not rely on normally distributed covariate measurements when the number of observation points grows large, any reasonable distribution could be used for the repeated measurements, though the update formulas provided in the Supplementary Material would need to be amended.

5. Application to the GenIMS Study

We illustrate the proposed approximate EM algorithm for maximizing the joint model of a binary variable and several longitudinal covariates subject to DLs by applying it to the Genetic and Inflammatory Markers of Sepsis (GenIMS) data set. One of the main purposes of the GenIMS study was to identify relationships between cytokine levels in the body and the event of survival after 90-days for patients with community acquired pneumonia (CAP).

Three cytokines were measured in this study: tumor necrosis factor (TNF), interleukin-6 (IL-6), and interleukin-10 (IL-10). Cytokines are cell-signaling protein molecules that are sent out by the immune system for the purpose of communicating with the rest of the body. The TNF and IL-6 cytokines serve as biomarkers of pro-inflammatory responses to CAP while IL-10 serves as a biomarker of anti-inflammatory responses to CAP. The three biomarkers, TNF, IL-6, and IL-10, were all subject to censoring below the detection thresholds 4, 2 or 5, and 5 pg/ml, respectively, and measured repeatedly for up to eight days for the majority of patients but for as many as 30 days for a few. We only considered the 1875 patients that truly had CAP, required a hospital stay, and had at least one, possibly

censored, measurement for each cytokine. The censoring proportions across all time-repeated measurements for the three biomarkers are 39.74%, 28.20%, and 70.59%, respectively, and 46.87% overall. A total of 98.53% of the individuals had at least one censored measurement.

We used the joint model described in Sections 2 and 3 to explain the relationship between the binary event of 90-day survival (1 indicating patient survived 90 days) and six covariates: the levels of the three cytokines, TNF, IL-6, and IL-10, sex (1 representing males, 0 representing females), race (1 representing Caucasians, 0 representing all other races), and age. We took a log-transformation of the three cytokine biomarkers before analysis, though we continue refer to the log-cytokines simply as TNF, IL-6, and IL-10. While longer-term survival times were available for those surviving beyond 90 days, we only considered the binary event of survival at 90 days since this was of main interest in the original study, based on recommendations for sepsis trials from international expert panels, presumably since any observed deaths after 90 days were likely due to causes unrelated to CAP.

We assumed the longitudinal model (1) with no elements in $\mathbf{r}_{ijk}$ and $\mathbf{s}_{ijk} = (1, t_{ijk})^T$, $i = 1, \dots, 1875$, $j = 1, \dots, n_{ik}$, $k = 1, \dots, 3$. Using the longitudinal plots shown in Figure 2 as a guide, we decided to include an intercept and slope random effect term for each of the cytokine biomarkers. We assumed that the joint distribution of the six random effects is multivariate normal. We allowed for unique variances τ_{TNF}^2 , τ_{IL-6}^2 , and τ_{IL-10}^2 , but we assumed that these variances are constant over time for each covariate. Finally, we assumed the logistic model (2) to relate 90-day survival to the random effects and baseline covariates sex, race, and age. With this modeling set-up, the logistic regression parameters associated with the intercept random effects can be interpreted as the average effect, conditional on sex, race, and age, that the patient's cytokine levels upon entering the hospital have on 90-day survival probability (in the logit scale). Similarly, the logistic regression parameters associated with the slope random effects can be interpreted as the conditional average effect that the linear rate of change of cytokine levels during the first several days in the hospital have on 90-day survival probability.

In the GenIMS data set, the longitudinal covariates were subject to missingness in addition to the censoring at DLs. Specifically, 48.5%, 15.3%, and 15.4% of the TNF, IL-6, and IL-10 observations, respectively, were missing completely. The high level of missingness for the TNF biomarker was by design, as TNF measurements were only collected after day one on a random subset of the individuals. Most of the remaining missingness resulted from the fact that no measurements were immediately taken on individuals arriving at the hospital during certain times on weekends and also due to reduced measurement taking during holidays (Kellum et al., 2007). Based on these explanations for missingness, we felt that it was reasonable to conclude that the missing values are missing at random. That is, we have no reason to believe that the missingness is in any way related to unobserved factors.

Table 8 summarizes the parameter and standard error estimates for the intercept, age, race, and sex covariates, as well as the intercept and random effects associated with the TNF,

IL-6, and IL-10 covariates. We only include the parameters in the logistic submodel of the joint model since they are of primary interest for the original GenIMS study.

The sex and race covariates are both significantly related to 90-day survival at the $\alpha = 0.05$ level, with p-values of 0.012 and 0.027, respectively, and the age covariate is very significant, with a p-value of less than 0.001. Thus, being male, white, and older is associated with a decreased probability of survival after admission to the hospital due to CAP. Two of the three cytokine biomarkers, IL-6 and IL-10, are significantly related to 90-day survival. Specifically, the baseline levels of IL-6 and IL-10, as well as the rates of change over time for IL-6 and IL-10 levels, are very strongly negatively related to 90-day survival with p-values all less than 0.001. The p-values for the baseline TNF levels and rate of change over time are 0.077 and 0.418.

The results above compare strongly with previous studies. In the original analysis conducted by Kellum et al. (2007), each of the three cytokines were considered separately rather than jointly, and both IL-6 and IL-10 were found to be significantly related to 90-day survival. Bernhardt et al. (2013), which considered the cytokines jointly in a logistic model but only using first day observations, found that none of the cytokines were related to 90-day survival. Recognizing the multicollinearity involved, an additional analysis was then conducted on each cytokine separately and each was found to be at least marginally statistically significant. A few papers, such as D' Angelo and Weissfeld (2008), Sattar et al. (2012), and Bernhardt et al. (2014) modeled the actual survival times using one or more of the cytokines as measured on first day or last day of hospitalization. Sattar et al. (2012) and D' Angelo and Weissfeld (2008) both used techniques based on Cox proportional hazards models. Sattar et al. (2012) only considered IL-10 in the model and found it to be statistically significant in predicting survival time while D' Angelo and Weissfeld (2008) jointly modeled IL-6 and IL-10 and only found IL-6 to be strongly statistically significant. Bernhardt et al. (2014) used accelerated failure time models to jointly model TNF, IL-6, and IL-10 and only found IL-6 and IL-10 to be moderately statistically significant, though it was noted that a global test for the three biomarkers was strongly significant.

No previous study of the GenIMS data used all of the longitudinal data for all three cytokines of interest simultaneously in a model for 90-day survival. By taking advantage of this additional information, the analysis above shows a clearer relationship of IL-6 and IL-10 to survival when both are considered jointly with TNF in the model.

6. Discussion

We have proposed a joint model for a binary outcome and multiple longitudinal covariates subject to DLs. To overcome difficulties in fitting this model, we have developed a computationally-efficient approximate EM algorithm. We have shown that this algorithm performs very well, giving unbiased or only slightly biased estimates with good coverage probabilities, both in cases when there is no censoring on the covariates and when there is censoring due to DLs. The main advantage of this approximate algorithm is that the necessary computational time does not increase exponentially as the number of random effects increases since only one-dimensional integrations are required.

The joint model that we have proposed is more flexible than those presented by several authors in the context of a binary outcome, specifically by allowing for multiple longitudinal covariates, all of which may be subject to DLs, and also by allowing for the inclusion of baseline covariates in the linear mixed models for the longitudinal covariates. However, some of our assumptions may be restrictive in some scenarios. We presumed that the trajectories for the longitudinal covariates can be explained by a linear mixed model, though in some contexts, a non-linear mixed model may be more appropriate. Additionally, we have assumed independent and identically distributed errors in the longitudinal model where sometimes allowing for heterogeneous errors is more accurate. Lastly, we presupposed that the random effects are normally distributed. Several authors have shown that assuming normality may lead to biases in the parameter estimates when the random effects distribution is not normal, and thus it may be useful to consider more flexible distributions for the random effects. Fortunately, as long as the number of longitudinal observations grows large for each individual, the main idea of the approximate EM algorithm that we proposed is still valid since the conditional distribution $f(\mathbf{b}_i|y_i, \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iq}, z_i; \boldsymbol{\theta})$ converges to a normal.

Finally, we note that the proposed normal approximation could also be used for joint models with a survival response. A similar approximate EM algorithm procedure should be applicable as long as the number of longitudinal observations for each individual grows large. Future work could focus on studying the properties of the approximate EM algorithm in this context. Additionally, it may be interesting to compare the approximations used for obtaining estimates for joint models in this paper to previous proposals such as those in Rizopoulos et al. (2009) and Rizopoulos (2012a).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Dr. Lan Kong and the CRISMA (Clinical Research, Investigation, and Systems Modeling of Acute Illness) Center at the University of Pittsburgh for providing the GenIMS data set as well as the editor, associated editor, and reviewers for their helpful comments and suggestions. The research of Wang is supported by the National Science Foundation CAREER award [DMS-1149355], and the research of Zhang is supported by the National Institutes of Health [R01 CA85848-12] and the National Institutes of Health/National Institute of Allergy and Infectious Diseases [R37 AI031789-20].

References

- Baghishani H, Mohammadzadeh M. Asymptotic normality of posterior distribution for generalized linear mixed models. *Journal of Multivariate Analysis*. 2012; 111:66–77.
- Bates, D.; Maechler, M.; Bolker, B. lme4: Linear mixed-effects models using Eigen and R syntax. 2012. R package version 0.999999-0
- Bernhardt PW, Wang HJ, Zhang D. Statistical methods for generalized linear models with covariates subject to detection limits. *Statistics in Biosciences*. 2013;1–22.
- Bernhardt PW, Wang HJ, Zhang D. Flexible modeling of survival data with covariates subject to detection limits via multiple imputation. *Computational Statistics and Data Analysis*. 2014; 69:81–91.
- Carriquiry AL, Gianola D, Fernando RL. Mixed-model analysis of a censored normal distribution with reference to animal breeding. *Biometrics*. 1987; 4:929–939. [PubMed: 3509964]

- Cox, D.; Hinkley, D. Theoretical Statistics. Chapman and Hall; 1973.
- D'Angelo GD, Weissfeld L. An index approach for the Cox model with left censored covariates. *Statistics in Medicine*. 2008; 27:4502–4514. [PubMed: 18407573]
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*. 1977; 39:1–38.
- Have TRT, Kunselman AR, Pulkstenis ER. Mixed effects logistic regression models for longitudinal binary response data with informative dropout. *Biometrics*. 1998; 54:367–383. [PubMed: 9544529]
- Have TRT, Miller ME, Reboussin BA, James MK. Mixed effects logistic regression models for longitudinal ordinal functional response data with multiple-cause drop-out from the longitudinal study of aging. *Biometrics*. 2000; 56:279–287. [PubMed: 10783807]
- Helsel, DR. *Statistics for Censored Environmental Data Using Minitab and R*. 2. Wiley; 2012.
- Hwang Y-T, Tsai H-Y, Chang Y-J, Kuo H-C, Wang CC. The joint model of the logistic model and linear random effect model – an application to predict orthostatic hypertension for subacute stroke patients. *Computational Statistics and Data Analysis*. 2011; 55:914–923.
- James F. Monte Carlo theory and practice. *Reports on Progress in Physics*. 1980; 43:1145–1189.
- Kellum JA, Kong L, Fink MP, Weissfeld LA, Yealy DM, Pinsky MR, Fine J, Krichevsky A, Delude R, Angus D. Understanding the inflammatory cytokine response in pneumonia and sepsis. *Archives of Internal Medicine*. 2007; 167:1655–1663. [PubMed: 17698689]
- Li E, Wang N, Wang N-Y. Joint models for primary endpoint and multiple longitudinal covariate processes. *Biometrics*. 2007a; 63:1068–1078. [PubMed: 17501940]
- Li E, Zhang D, Davidian M. Conditional estimation for generalized linear models when covariates are subject-specific parameters in a mixed model for longitudinal measurements. *Biometrics*. 2004; 60:1–7. [PubMed: 15032767]
- Li E, Zhang D, Davidian M. Likelihood and pseudo-likelihood methods for semiparametric joint models for a primary endpoint and longitudinal data. *Computational Statistics and Data Analysis*. 2007b; 51:5776–5790. [PubMed: 18704154]
- Lin H, McCulloch CE, Mayne ST. Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. *Statistics in Medicine*. 2002; 21:2369–2382. [PubMed: 12210621]
- Lyles RH, Lyles CM, Taylor DJ. 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*. 2000; 49:485–497.
- Martin AD, Quinn KM, Park JH. MCMCpack: Markov Chain Monte Carlo in R. *Journal of Statistical Software*. 2011;42, 22.
- May, R. PhD thesis. University of North Carolina; 2011. Estimation methods for data subject to detection limits.
- Pettitt AN. Censored observations, repeated measures and mixed effects models – an approach using the EM algorithm and normal errors. *Biometrika*. 1986; 73:635–643.
- Pike F. Joint modeling of censored longitudinal and event time data. *Journal of Applied Statistics*. 2013; 40:17–27.
- Proust-Lima C, Joly P, Dartigues J, Jacqmin-Gadda H. Joint modelling of multivariate longitudinal outcomes and a time-to-event: a non-linear latent class approach. *Computational Statistics and Data Analysis*. 2009; 53:1142–1154.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; Vienna, Austria; 2012.
- Rizopoulos D. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Computational Statistics and Data Analysis*. 2012a; 56:491–501.
- Rizopoulos, D. *Joint Models for Longitudinal and Time-to-Event Data*. Chapman and Hall; 2012b.
- Rizopoulos D, Verbeke G, Lesaffre E. Fully exponential Laplace approximation for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society, Series B*. 2009; 71:637–653.
- Sattar A, Sinha SK, Morris NJ. A parametric survival model when a covariate is subject to left-censoring. *Journal of Biometrics and Biostatistics*. 2012;S3:002.10.4172/2155-6180.S3-002

- Sattar A, Weissfeld LA, Molenberghs G. Analysis of non-ignorable missing and left-censored longitudinal data using a weight random effects Tobit model. *Statistics in Medicine*. 2011; 30:3167–3180. [PubMed: 21898524]
- Su L, Tom BDM, Farewell VT. Bias in 2-part mixed models for longitudinal semicontinuous data. *Biostatistics*. 2009; 10:374–389. [PubMed: 19136448]
- Vaida F, Liu L. lme4: Linear mixed-effects models with censored responses. 2012 R package version 1.0.
- Vock DM, Davidian M, Tsiatis AA. Mixed model analysis of censored longitudinal data with flexible random-effects density. *Biostatistics*. 2012; 13:61–73. [PubMed: 21914727]
- Wang CY, Wang N, Wang S. Regression analysis when covariates are regression parameters of a random effects model for observed longitudinal measurements. *Biometrics*. 2000; 56:487–495. [PubMed: 10877308]
- Wei GCG, Tanner MA. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*. 1990; 85:699–704.
- Wu CFJ. On the convergence properties of the EM algorithm. *The Annals of Statistics*. 1983; 11:95–103.
- Wu L. A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with applications to AIDS studies. *Journal of the American Statistical Association*. 2002; 97:955–964.
- Wu L, Hu XJ, Wu H. Joint inference for nonlinear mixed-effects models and time to event at the presence of missing data. *Biostatistics*. 2008; 9:308–320. [PubMed: 17728318]
- Wu M, Carroll R. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*. 1988; 44:175–188.
- Wulfsohn M, Tsiatis A. A joint model for survival and longitudinal data measured with error. *Biometrics*. 1997; 53:330–339. [PubMed: 9147598]
- Ye W, Lin X, Taylor J. A penalized likelihood approach to joint modeling of longitudinal measurements and time-to-event data. *Statistics and Its Interface*. 2008; 1:33–45.
- Yuan Y, Little RJA. Mixed-effect hybrid models for longitudinal data with nonignorable dropout. *Biometrics*. 2009; 1:478–486.

Figure 1.

Distribution of the random effects conditional on the data in a joint model with a single longitudinal covariate described by a linear mixed model with a random intercept, b_0 , and a random slope, b_1 , using simulated data and various degrees of censoring.

Figure 2. Longitudinal plots of $\log(\text{TNF})$, $\log(\text{IL-6})$, and $\log(\text{IL-10})$ for a random sample of patients in the GenIMS study.

Table 1

Simulation results for the joint model of a binary response and a single longitudinal covariate not subject to censoring. SD: empirical standard deviation; SE: average estimated standard error; CP: empirical 95% coverage probability.

Parameter	Bias	SD	SE	CP
β_{11}	-0.024	0.475	0.469	0.966
β_{12}	0.000	0.006	0.006	0.966
β_{21}	0.017	0.215	0.209	0.942
β_{22}	-0.029	0.326	0.331	0.962
γ	0.000	0.004	-	-
τ^2	-0.001	0.038	-	-
B_1	-0.014	0.276	-	-
B_2	0.001	0.029	-	-
D_{11}	-0.007	0.134	-	-
D_{12}	-0.001	0.043	-	-
D_{22}	-0.002	0.027	-	-

Table 2

Simulation results for the joint model of a binary response and a single longitudinal covariate subject to 20%, 40%, and 60% censoring. Par.: Parameter; SD: empirical standard deviation; SE: average estimated standard error; CP: empirical 95% coverage probability.

Par.	20% Censoring				40% Censoring				60% Censoring			
	Bias	SD	SE	CP	Bias	SD	SE	CP	Bias	SD	SE	CP
β_{11}	-0.019	0.476	0.474	0.960	-0.040	0.492	0.491	0.958	-0.045	0.527	0.522	0.958
β_{12}	0.000	0.006	0.006	0.958	0.000	0.006	0.006	0.956	0.000	0.006	0.006	0.946
β_{21}	-0.002	0.213	0.210	0.950	0.009	0.234	0.236	0.958	-0.003	0.255	0.267	0.952
β_{22}	-0.039	0.357	0.364	0.958	-0.044	0.394	0.407	0.952	-0.030	0.421	0.436	0.950
γ	-0.001	0.004	-	-	-0.002	0.004	-	-	-0.002	0.005	-	-
ρ^2	-0.001	0.041	-	-	-0.004	0.045	-	-	-0.010	0.052	-	-
B_1	0.089	0.287	-	-	0.119	0.320	-	-	0.174	0.387	-	-
B_2	0.019	0.029	-	-	0.024	0.032	-	-	0.017	0.041	-	-
D_{11}	0.041	0.140	-	-	0.041	0.154	-	-	0.051	0.189	-	-
D_{12}	0.002	0.042	-	-	0.006	0.047	-	-	0.005	0.047	-	-
D_{22}	-0.026	0.027	-	-	-0.027	0.031	-	-	-0.015	0.038	-	-

Table 3

Bias for parameter estimates in the joint model of a binary response and a single longitudinal covariate subject to 20%, 40%, and 60% censoring when censored values are replaced by DL or DL/2.

Parameter	<u>20% Censoring</u>		<u>40% Censoring</u>		<u>60% Censoring</u>	
	DL	DL/2	DL	DL/2	DL	DL/2
β_{11}	0.027	0.112	-0.524	-0.085	-1.974	-1.024
β_{12}	0.000	0.000	0.001	0.000	0.000	0.000
β_{21}	-0.016	-0.056	0.176	0.028	0.476	0.314
β_{22}	-0.104	-0.088	-0.061	-0.098	-0.258	0.084
γ	-0.002	0.000	-0.006	-0.003	-0.010	-0.008
ρ^2	-0.236	-0.117	-0.447	-0.290	-0.655	-0.499
B_1	0.138	-0.061	0.801	0.339	1.798	1.148
B_2	0.087	0.074	0.094	0.103	0.059	0.091
D_{11}	-0.133	0.115	-0.643	-0.344	-0.881	-0.792
D_{12}	-0.065	-0.040	-0.136	-0.093	-0.233	-0.177
D_{22}	-0.120	-0.103	-0.145	-0.133	-0.160	-0.148

Table 4

Simulation results for the joint model of a binary response and three longitudinal covariates not subject to censoring. SD: empirical standard deviation; SE: average estimated standard error; CP: empirical 95% coverage probability.

Parameter	Bias	SD	SE	CP
β_{11}	-0.015	0.887	0.939	0.970
β_{12}	0.000	0.006	0.006	0.940
β_{21}	0.009	0.195	0.184	0.947
β_{22}	-0.009	0.282	0.286	0.967
β_{23}	0.035	0.231	0.234	0.963
β_{24}	-0.021	0.325	0.332	0.953
β_{25}	0.029	0.203	0.196	0.957
β_{26}	-0.033	0.310	0.310	0.963
τ_1^2	-0.006	0.041	-	-
τ_2^2	0.006	0.059	-	-
τ_3^2	-0.002	0.052	-	-
B_1	-0.007	0.064	-	-
B_2	-0.001	0.029	-	-
B_3	-0.007	0.073	-	-
B_4	0.002	0.032	-	-
B_5	-0.011	0.062	-	-
B_6	-0.004	0.029	-	-

Table 5

Simulation results for the joint model of a binary response and three longitudinal covariates subject to 25% and 50% total censoring. Par.: Parameter; SD: empirical standard deviation; SE: average estimated standard error; CP: empirical 95% coverage probability; DL: bias when replacing censored observations with DL; DL/2: bias when replaced censored observations with DL/2.

Par.	25% Censoring						50% Censoring					
	Bias	SD	SE	CP	DL	DL/2	Bias	SD	SE	CP	DL/2	DL/2
β_{11}	0.173	0.808	0.922	0.963	0.508	0.578	0.066	0.867	0.967	0.967	0.967	-0.090
β_{12}	0.000	0.006	0.006	0.933	0.002	-0.001	0.000	0.006	0.006	0.937	0.937	-0.001
β_{21}	0.000	0.197	0.198	0.953	-0.055	-0.019	0.000	0.226	0.218	0.957	0.957	0.091
β_{22}	-0.027	0.319	0.332	0.970	0.226	-0.010	0.016	0.357	0.368	0.967	0.967	0.047
β_{23}	-0.024	0.195	0.213	0.963	-0.058	-0.059	-0.005	0.204	0.215	0.957	0.957	-0.070
β_{24}	-0.001	0.331	0.345	0.943	-0.626	-0.169	0.001	0.369	0.360	0.940	0.940	-0.176
β_{25}	0.013	0.205	0.228	0.973	0.300	0.133	0.024	0.268	0.289	0.943	0.943	1.183
β_{26}	-0.042	0.352	0.377	0.953	0.034	-0.048	-0.038	0.429	0.436	0.967	0.967	1.125
τ_1^2	-0.006	0.046	-	-	-0.312	-0.160	-0.016	0.052	-	-	-	-0.386
τ_2^2	0.006	0.061	-	-	-0.204	-0.112	-0.019	0.071	-	-	-	-0.248
τ_3^2	-0.002	0.062	-	-	0.579	-0.3353	-0.133	0.101	-	-	-	-0.872
B_1	-0.007	0.066	-	-	0.107	-0.062	-0.007	0.073	-	-	-	0.255
B_2	-0.001	0.030	-	-	0.074	0.075	0.022	0.035	-	-	-	0.087
B_3	-0.007	0.074	-	-	-0.161	-0.134	-0.030	0.082	-	-	-	-0.163
B_4	0.002	0.032	-	-	0.116	0.080	0.020	0.033	-	-	-	0.127
B_5	-0.011	0.069	-	-	0.491	0.147	0.023	0.142	-	-	-	1.602
B_6	-0.004	0.030	-	-	0.034	0.061	0.060	0.037	-	-	-	-0.043

Table 6

Simulation results for the joint model of a binary response and a single longitudinal covariate not subject to censoring with no more than four longitudinal observations per individual. SD: empirical standard deviation; SE: average estimated standard error; CP: empirical 95% coverage probability.

Parameter	Bias	SD	SE	CP
β_{11}	-0.081	0.563	0.542	0.966
β_{12}	0.001	0.007	0.007	0.954
β_{21}	0.057	0.369	0.346	0.958
β_{22}	-0.093	0.634	0.587	0.956
γ	0.000	0.004	-	-
τ^2	-0.007	0.057	-	-
B_1	-0.018	0.285	-	-
B_2	0.002	0.040	-	-
D_{11}	0.008	0.221	-	-
D_{12}	-0.011	0.084	-	-
D_{22}	0.002	0.051	-	-

Table 7

Simulation results for the joint model of a binary response and a single longitudinal covariate not subject to censoring with repeated observations distributed Gamma. SD: empirical standard deviation; SE: average estimated standard error; CP: empirical 95% coverage probability.

Parameter	Bias	SD	SE	CP
β_{11}	-0.055	0.483	0.473	0.954
β_{12}	0.001	0.006	0.006	0.940
β_{21}	0.039	0.247	0.212	0.952
β_{22}	-0.059	0.367	0.332	0.940
γ	0.000	0.004	-	-
τ^2	0.004	0.069	-	-
B_1	-0.015	0.273	-	-
B_2	0.000	0.031	-	-
D_{11}	-0.009	0.153	-	-
D_{12}	0.001	0.046	-	-
D_{22}	-0.001	0.027	-	-

Table 8

Parameter and standard error estimates for the coefficients of log-cytokine and demographic covariates in the joint model for 90-day survival, based on the GenIMS data set. Each pair of estimates for the log-cytokines represents the baseline effect of the covariate as well as the effect corresponding to the longitudinal trajectory.

	Par. Est.	Std. Error
Intercept	6.446	0.794
Sex	−0.330	0.131
Race	−0.266	0.120
Age	−0.053	0.005
TNF, Intercept	0.267	0.151
TNF, Slope	−0.590	1.577
IL-6, Intercept	−0.382	0.105
IL-6, Slope	−1.943	0.448
IL-10, Intercept	−0.363	0.110
IL-10, Slope	−1.334	0.293