On Estimating the Relationship between Longitudinal Measurements and Time-to-Event Data Using a Simple Two-Stage Procedure

Paul S. Albert[1,*] **and Joanna H. Shih**[2]

[1]Biostatistics and Bioinformatics Branch, Division of Epidemiology, Statistics, and Prevention Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, Maryland 20892, U.S.A.
[2]Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, Maryland 20892, U.S.A.
email: albertp@mail.nih.gov

SUMMARY. Ye, Lin, and Taylor (2008, *Biometrics* **64**, 1238–1246) proposed a joint model for longitudinal measurements and time-to-event data in which the longitudinal measurements are modeled with a semiparametric mixed model to allow for the complex patterns in longitudinal biomarker data. They proposed a two-stage regression calibration approach that is simpler to implement than a joint modeling approach. In the first stage of their approach, the mixed model is fit without regard to the time-to-event data. In the second stage, the posterior expectation of an individual's random effects from the mixed-model are included as covariates in a Cox model. Although Ye et al. (2008) acknowledged that their regression calibration approach may cause a bias due to the problem of informative dropout and measurement error, they argued that the bias is small relative to alternative methods. In this article, we show that this bias may be substantial. We show how to alleviate much of this bias with an alternative regression calibration approach that can be applied for both discrete and continuous time-to-event data. Through simulations, the proposed approach is shown to have substantially less bias than the regression calibration approach proposed by Ye et al. (2008). In agreement with the methodology proposed by Ye et al. (2008), an advantage of our proposed approach over joint modeling is that it can be implemented with standard statistical software and does not require complex estimation techniques.

KEY WORDS: Informative dropout; Joint model; Regression calibration; Two-stage models.

1. Introduction

Ye et al. (2008) proposed a two-stage regression calibration approach for estimating the relationship between longitudinal measurements and time-to-event data. Their approach was motivated by trying to establish such a relationship when the longitudinal measurements follow a complex semi-parametric mixed model with subject-specific random stochastic processes and the time-to-event data follow a proportional hazards model. Specifically, they proposed a semi-parametric model with additive errors for the longitudinal measurements X_{ij} of the form

$$
\begin{aligned}
X_{ij} &= \mathbf{Z}_i'\boldsymbol{\beta} + \varphi(t_{ij}) + \mathbf{U}_i(t_{ij})\mathbf{b}_i + W_i(t_{ij}) + \epsilon_{ij} \\
&= X_i^*(t_{ij}) + \epsilon_{ij},
\end{aligned} \tag{1}
$$

where $\boldsymbol{\beta}$ is a vector of regression coefficients associated with fixed effect covariates $\mathbf{Z}_i$, $\varphi(t)$ is an unknown smooth function over time, $\mathbf{b}_i$ is a vector of subject-specific random effects corresponding to covariates $\mathbf{U}_i(t)$, which is assumed normally distributed with mean $\mathbf{0}$ and variance $\Sigma_{\mathbf{b}}$. Further, $W_i(t_{ij})$ is a zero mean integrated Wiener stochastic process. We denote $\mathbf{X}_i$ as all longitudinal measurements on the ith individual.

In Ye et al. (2008)'s approach, the relationship between the slope of the longitudinal process and a time-to-event outcome T_i is characterized by a Cox proportional hazards model with the slope at time t, denoted as $X_i^{*\prime}(t)$, being treated as a time-dependent covariate. The authors proposed a two-stage estimation procedure in which in the first stage, the mean of the posterior distribution of the slope at time t, $\mathrm{E}[X_i^{*\prime}(t)\,|\,\mathbf{X}_i, \mathbf{Z}_i]$, is estimated using model (1) without regard to the time-to-event process T_i. In the second stage, $E[X_i^{*\prime}(t)\,|\,\mathbf{X}_i, \mathbf{Z}_i]$ replaces $X_i^{*\prime}(t)$ in the Cox model. Ye et al. (2008) proposed two approaches: (i) the ordinary regression calibration (ORC) approach in which $\mathrm{E}[X_i^{*\prime}(t)\,|\,\mathbf{X}_i, \mathbf{Z}_i]$ is estimated using (1) with all available longitudinal measurements and (ii) the risk set regression calibration (RRC) approach in which these expectations are obtained by estimating model (1) after each event using only longitudinal measurements for subjects at risk at time t (i.e., subjects who have an event before time t are removed from the estimation).

The advantage of these regression calibration approaches are that they do not require the complex joint modeling of the longitudinal and time-to-event processes. In the discussion of their paper, Ye et al. (2008) acknowledge that these approaches may result in a biased estimation due to informative dropout and measurement error, and that improved performance will require incorporating informative dropout and the uncertainty of measurement error into the estimation. In this article, we show that an alternative two-stage procedure can be formulated, which reduces the bias

considerably without requiring complex joint modeling of both processes. For simplicity, we develop the approach for a longitudinal model without the smooth function $\varphi(t)$ and the stochastic component $W_i(t)$ in (1), but the proposed approach applies more generally. In this approach, we approximate the conditional distribution of the longitudinal process given the event time, simulate complete follow-up data based on the approximate conditional model, and then fit the longitudinal model with complete follow-up on each patient (hence avoiding the problem of informative dropout in Ye et al. [2008]'s approach). Section 2 develops the approach for a discrete event time distribution followed by an approximation for the continuous event time distribution. The results of simulations that show the advantages of the proposed approach over ORC and RRC are provided in Section 3. A discussion follows in Section 4.

2. Modeling Framework

We begin by considering a discrete event time distribution. Define T_i to be a discrete event time that can take on discrete values t_j, $j = 1, 2, \ldots, J$, and Y_{ij} to be a binary indicator of whether the ith patient is dead at time t_j. Then $J_i = \sum_{j=1}^{J}(1 - Y_{ij}) = J - Y_{i\cdot}$ where $Y_{i\cdot} = \sum_{j=1}^{J} Y_{ij}$ indicates the number of follow-up measurements before death or administrative censoring for the ith patient. Every patient will be followed until death or the end of follow-up at time t_j.

For illustrative purposes, we will consider a joint model for longitudinal and discrete time-to-event data in which the discrete event time distribution is modeled as a linear function of the slope of an individual's longitudinal process on the probit scale. Specifically,

$$P(Y_{ij} = 1 | Y_{i(j-1)} = 0) = \Phi(\alpha_{0j} + \alpha_1 b_{i1}), \qquad (2)$$

where $j = 1, 2, \ldots, J$, Y_{i0} is taken as 0, α_{0j} governs the baseline discrete event time distribution, and b_{i1} is the individual slope from the linear mixed model,

$$X_{ij} = X_i^*(t_j) + \epsilon_{ij}, \qquad (3)$$

$$X_i^*(t_j) = \beta_0 + \beta_1 t_j + b_{i0} + b_{i1} t_j, \qquad (4)$$

where $i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, J_i$. In (4), the parameters β_0 and β_1 are fixed-effect parameters characterizing the mean intercept and slope of the longitudinal process, respectively; $(b_{i0}, b_{i1})'$ is a vector of random effects, which are assumed multivariate normal with mean $\mathbf{0}$ and variance $\Sigma_{\boldsymbol{b}} = \begin{pmatrix} \sigma_{b_0}^2 & \sigma_{b_0, b_1} \\ \sigma_{b_0, b_1} & \sigma_{b_1}^2 \end{pmatrix}$; and ϵ_{ij} is a residual error term that is assumed normal with mean zero and variance σ_ϵ^2. In (2)–(4), the event time and the longitudinal process are linked through b_{i1}, and the parameter α_1 governs the relationship between the slope of the longitudinal process and the event time distribution. Denote $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{iJ_i})'$, $\boldsymbol{b}_i = (b_{i0}, b_{i1})'$, and $\boldsymbol{\beta} = (\beta_0, \beta_1)'$. As in Ye et al. (2008), the normality assumption for $\boldsymbol{b}_i$ is made for these joint models. Although not the focus of this article, various articles have proposed methods with flexible semi-parametric random effects distributions and have demonstrated that inferences are robust to departures from normality (Song, Davidian, and Tsiatis, 2002; Tsiatis and Davidian, 2004; Hsieh, Tseng, and Wang, 2006).

For estimating the relationship between the slope of the longitudinal process and the time-to-event process, the calibration approach of Ye et al. (2008) reduces to first, estimating $E[b_{i1} \mid \boldsymbol{X}_i, \boldsymbol{\beta}]$ using (3) and (4), and second, replacing b_{i1} by $E[b_{i1} | \boldsymbol{X}_i, \widehat{\boldsymbol{\beta}}]$ in estimating (2). As recognized by Ye et al. (2008), this methodology introduces bias in two ways. First, there is the problem of informative dropout, whereby b_{i0} and b_{i1} can depend on the event time T_i (which will occur if $\alpha_1 \neq 0$ in (2)). Ignoring this informative dropout may result in substantial bias. Second, not accounting for the measurement error in $E[b_{i1} | \boldsymbol{X}_i, \widehat{\boldsymbol{\beta}}]$ relative to true values of b_{i1} will result in attenuated estimation of α_1.

We propose a simple approach that reduces these two sources of bias. We first focus on the problem of informative dropout. The bias from informative dropout is a result of differential follow-up whereby the response process is related to the length of follow-up (i.e., in (2)–(4), when α_1 is positive, patients who die early are more likely to have large positive slopes). There would be no bias if all J follow-up measurements were observed on all patients. Thus, we recapture these missing measurements by generating data from the conditional distribution of $\boldsymbol{X}_i$ given T_i, denoted as $\boldsymbol{X}_i | T_i$. Since $\boldsymbol{X}_i | T_i$ under (2)–(4) does not have a tractable form, we propose a simple approximation for this conditional distribution.

Under model (2)–(4), the distribution of $\boldsymbol{X}_i | T_i$ can be expressed as

$$P(\boldsymbol{X}_i | T_i) = \int h(\boldsymbol{X}_i | \boldsymbol{b}_i, T_i) g(\boldsymbol{b}_i | T_i) d\boldsymbol{b}_i. \qquad (5)$$

Since T_i and the values of $\boldsymbol{X}_i$ are conditionally independent given $\boldsymbol{b}_i$, $h(\boldsymbol{X}_i | \boldsymbol{b}_i, T_i) = h(\boldsymbol{X}_i | \boldsymbol{b}_i)$, where $h(\boldsymbol{X}_i | \boldsymbol{b}_i)$ is the product of J_i univariate normal density functions each with mean $X_i^*(t_j)$ $(j = 1, 2, \ldots, J_i)$ and variance σ_ϵ^2. The distribution of $\boldsymbol{X}_i | T_i$ can easily be obtained with standard statistical software if we approximate $g(\boldsymbol{b}_i | T_i)$ by a normal distribution. Under the assumption that $g(\boldsymbol{b}_i | T_i)$ is normally distributed with mean $\boldsymbol{\mu}_{T_i} = (\mu_{0T_i}, \mu_{1T_i})'$ and variance $\Sigma_{bT_i}^*$, and by rearranging mean structure parameters in the integrand of (5) so that the random effects have mean zero, $\boldsymbol{X}_i | T_i$ corresponds to the following mixed model:

$$X_{ij} | \left(T_i, b_{i0T_i}^*, b_{i1T_i}^* \right) = \beta_{0T_i}^* + \beta_{1T_i}^* t_j + b_{i0T_i i}^* + b_{i1T_i}^* t_j + \epsilon_{ij}^*, \qquad (6)$$

where $i = 1, 2, \ldots, I$; $j = 1, 2, \ldots, J_i$; and the residuals ϵ_{ij}^* are assumed to have independent normal distributions with mean zero and variance σ_ϵ^{*2}. Further, the fixed-effects parameters $\beta_{0T_i}^*$ and $\beta_{1T_i}^*$ are intercept and slope parameters for patients who have an event at time T_i or who are censored at time $T_i = t_J$. In addition, the associated random effects $\boldsymbol{b}_{iT_i}^* = (b_{i0T_i}^*, b_{i1T_i}^*)'$ are multivariate normal with mean $\mathbf{0}$ and variance $\Sigma_{bT_i}^*$ for each T_i. Thus, this flexible conditional model involves estimating separate fixed effect intercept and slope parameters for each potential event-time and for subjects who are censored at time t_J. Likewise, separate random effects distributions are estimated for each of these discrete time points. For example, the intercept and slope fixed-effect parameters for those patients who have an event at time $T_i = 3$ are β_{03}^* and β_{13}^*, respectively. In addition, the intercept and slope random effects for those patients who have an event at

$T_i = 3, \boldsymbol{b}^*_{iT_i} = (b^*_{i03}, b^*_{i13})'$, are multivariate normal with mean $\boldsymbol{0}$ and variance Σ^*_{b3}. Model (6) can be fit with standard R code that is available from the first author.

A similar approximation of the conditional distribution of the longitudinal process given dropout time has been proposed for estimating mean change over time in longitudinal measurements subject to informative dropout (Wu and Bailey, 1989; Wu and Follmann, 1999). In this article, we use the approximation to construct complete longitudinal datasets that in turn are used to estimate the mean of the posterior distribution of an individual's random effects given the data. Specifically, multiple complete longitudinal datasets can then be constructed by simulating X_{ij} values from (6) where the parameters are replaced by their estimated values. Since the simulated datasets have complete follow-up on each individual, the bias in estimating $E[b_{i1} \mid \boldsymbol{X}_i, \beta]$ caused by informative dropout is much reduced.

We provide a correction to account for the measurement error in using $E[b_{i1}|\boldsymbol{X}_i, \widehat{\beta}]$, denoted as $\widehat{b}_{i1}$, instead of using the actual random slope b_{i1}. As in Carroll et al. (1984) who adjust for measurement error in a covariate, we note that

$$P(Y_{ij} = 1|Y_{i(j-1)} = 0, \boldsymbol{X}_i) = \int \Phi(\alpha_{0j} + \alpha_1 b_{1i}) g(b_{1i}|\boldsymbol{X}_i) db_{1i}$$
$$= \Phi\left(\frac{\alpha_{0j} + \alpha_1 \widehat{b}_{i1}}{\sqrt{1 + \alpha_1^2 \text{Var}(\widehat{b}_{i1} - b_{i1})}}\right), \quad (7)$$

where $\text{Var}(\widehat{b}_{i1} - b_{i1})$ measures the error of estimation in $\widehat{b}_{i1}$ relative to b_{i1} that is the 1-1 element in the matrix $\text{Var}(\widehat{\boldsymbol{b}}_i - \boldsymbol{b}_i)$, given by

$$\text{Var}(\widehat{\boldsymbol{b}}_i - \boldsymbol{b}_i) = \Sigma_{\boldsymbol{b}} - \Sigma_{\boldsymbol{b}} \boldsymbol{R}'_i \{\boldsymbol{W}_i - \boldsymbol{W}_i \boldsymbol{F}_i \boldsymbol{Q} \boldsymbol{F}'_i \boldsymbol{W}_i\} \boldsymbol{R}_i \Sigma_{\boldsymbol{b}},$$

where $Q = \sum_{i=1}^{I}(\boldsymbol{F}'_i \boldsymbol{W}_i \boldsymbol{F}_i)^{-1}$, $\boldsymbol{W}_i = \boldsymbol{V}_i^{-1}$, and $\boldsymbol{V}_i$ is the variance of $\boldsymbol{X}_i$ (Laird and Ware, 1982; Verbeke and Molenberghs, 2000). Further, $\boldsymbol{F}_i$ and $\boldsymbol{R}_i$ are vectors of fixed and random effects for the ith subject. This variance formula incorporates the error in estimating the fixed effects in the longitudinal model. Expression (7) follows from that fact that $E[\Phi(a + V)] = \Phi[(a + \mu)/\sqrt{1 + \tau^2}]$, where $V \sim N(\mu, \tau^2)$.

Only individuals who have at least two longitudinal measurements provide useful information in assessing the relationship between an individual's slope and their time-to-event data, so we assume that all individuals in the analysis have at least two follow-up times. Thus, $\alpha_{01} = \alpha_{02} = -\infty$ and the regression parameters in the discrete-time model $\alpha_{0j}(j = 3, 4, \ldots, J)$ and α_1 can be estimated by maximizing the likelihood

$$L = \prod_{i=1}^{I}\left[\prod_{j=1}^{J_i}\{1 - P(Y_{ij} = 1|Y_{i(j-1)} = 0, \boldsymbol{X}_i)\}\right]$$
$$\times P(Y_{i(J_i+1)} = 1|Y_{iJ_i} = 0, \boldsymbol{X}_i)^{J_i < J}. \quad (8)$$

Thus, we propose the following algorithm for estimating $\alpha_{0j}(j = 3, 4, \ldots, J)$ and α_1 with a two-stage procedure

1. Estimate model (6) with all available longitudinal measurements using linear mixed-modeling software such as `lme` in R.

2. Simulate complete longitudinal pseudo measurements (i.e., X_{ij} for $i = 1, 2 \ldots, I$ and $j = 1, 2, \ldots, J$) using model (6) with model parameters estimated from step 1. Specifically, these measurements are simulated by first simulating values of $b^*_{iT_i}$ from a normal distribution with mean $\boldsymbol{0}$ and variance $\Sigma^*_{bT_i}$ and ϵ^*_{ij} from a normal distribution with mean 0 and variance $\sigma^*_{\epsilon^2}$, where the variance parameters are estimated in Step 1.

3. Estimate model (3) and (4) (without regard to the event time distribution (2)) with complete longitudinal measurements simulated in Step 2 using a linear mixed modeling software.

4. Estimate $\alpha_{0j}(j = 1, 2, \ldots, J)$ and α_1 (denoted as $\widehat{\alpha_{0j}}$ and $\widehat{\alpha}_1$, respectively) using (7) and (8) with $\widehat{b}_{i1}$ obtained from Step 3.

5. Repeat Steps 2–4 M times and average $\widehat{\alpha_{0j}}$ and $\widehat{\alpha}_1$ to get final estimates.

The approach can be generalized for continuous event-time distributions where T_i is the continuous event time for the i individual, all individuals are followed up to time T_E, and where patients are administratively censored at the end of the study when $T_i > T_E$. In addition, the Cox model, $\lambda(t, b_{i1}) = \lambda_0(t) \exp(\alpha b_{i1})$, is used to relate the longitudinal measurements to time-to-event data. We can approximate this conditional distribution by first discretizing the follow-up interval into K equally spaced intervals. We define d_i as a discretized version of the continuous event time distribution, whereby $d_i = k$ when $T_i \in ((k - 1)T_E/K, kT_E/K]$, $k = 1, 2, \ldots, K$, and where $d_i = K + 1$ when patient i's event time is administratively censored at time T_E. The conditional distribution of the longitudinal measurements given the continuous event time, $\boldsymbol{X}_i \mid T_i$, can be approximated by the distribution of the longitudinal measurements given the discretized version d_i, $\boldsymbol{X}_i \mid d_i$, where, as for the discrete event time model, this conditional distribution can be approximated by a linear mixed model

$$X_{ij} \mid (d_i, b^*_{i0d_i}, b^*_{i1d_i}) = \beta^*_{0d_i} + \beta^*_{1d_i} t_j + b^*_{i0d_i} + b^*_{i1d_i} t_j + \epsilon^*_{ij}, \quad (9)$$

where $i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots, J_i$, and where J_i is the number of follow-up measurements before death or administrative censoring for the ith patient. Similar to (6), $\beta^*_{0d_i}$ and $\beta^*_{1d_i}$ are intercept and slope parameters for patients with a discretized event time of d_i. Also, $\boldsymbol{b}^*_{id_i} = (b^*_{i0d_i}, b^*_{i1d_i})'$ are assumed to be normally distributed with mean $\boldsymbol{0}$ and variance $\Sigma^*_{bd_i}$. For continuous event times, we apply the previous algorithm for discrete-time data except that in Step 1 we fit model (9) for a reasonably large K, and in Step 3, we fit a Cox model without a measurement error correction instead of the discrete-time model.

Asymptotic standard errors from the discrete or continuous event time models cannot be used for inference since they fail to account for the missing data uncertainty in our procedure. The bootstrap (Efron and Tibshirani, 1993) can be used for valid standard error estimation.

3. Simulations

We evaluated the procedure for both discrete and continuous time-to-event data with a simulation study. For

Table 1
Estimates of α_1 from model (2)–(4) when
$\beta_0 = 1, \beta_1 = 3, \sigma_{b_0} = 1, \sigma_{b_1} = 1$, and $\sigma_{b_0,b_1} = 0$. We assume
that $\sigma_\epsilon = 0.75$, $\alpha_0 = -1.5$, $\alpha_1 = 0.50$, $J = 5$, and $I = 300$.
Further, we assume that $t_j = j$ and all individuals who are
alive at $t_5 = 5$ are administratively censored at that time
point. The means (standard deviations) from 1,000
simulations are presented.

Estimator	$\widehat{\alpha}_1$	SD
Known[1]	0.50	0.076
Complete[2]	0.50	0.077
ORC	0.45	0.078
RRC	0.45	0.075
Prop M=3 w/o MC	0.49	0.097
Prop M=3 w/MC	0.49	0.098
Prop M=10 w/o MC	0.49	0.090
Prop M=10 w/MC	0.49	0.092
Prop M=20 w/o MC	0.49	0.089
Prop M=20 w/MC	0.49	0.090
Prop M=50 w/o MC	0.49	0.088
Prop M=50 w/MC	0.49	0.089
Prop M=100 w/o MC	0.49	0.087
Prop M=100 w/MC	0.49	0.089

[1]Model (2) fit with b_{i1} assumed known.

[2]Model (2) fit with $\widehat{b}_{i1}$ replacing b_{i1}. The empirical Bayes estimates $\widehat{b}_{i1}$ are obtained by fitting (3) and (4) with complete longitudinal measurements.

discrete event-time data, we assume that there are potentially five follow-up times $J = 5$ at discrete times $t_j = j$ ($j = 1, \ldots, 5$) and $I = 300$ subjects having at least two or more longitudinal measurements (i.e., $\alpha_{01} = \alpha_{02} = -\infty$) with $\alpha_{0j} = -1.5$, $j = 3, 4$, and 5, and $\alpha_1 = 0.50$. Table 1 shows the mean and standard deviation for various estimators of α_1. These values are provided for estimators in which b_{i1} is assumed known, estimators that use complete simulated data, ORC and RRC, and our proposed approach with and without measurement error correction for different numbers of simulated datasets (M). The results show that ORC and RRC have an approximately 10% bias, while the proposed approach is unbiased. We also found that choosing $M = 10$ provided a good balance between efficiency and computational efficiency. Further, not incorporating the measurement error correction in the proposed approach had little effect on the results. We found this to be the case even when we increased the measurement error above 1, suggesting that this adjustment is not

particularly important for the simple model in which longitudinal measurements and survival are linked through a random slope parameter. The measurement error correction may be more important for a more complex model such as that presented by Ye et al. (2008).

For continuous time-to-event data, the simulation was conducted with an exponential survival distribution with a mean of 5 years when $b_i = 0$, administrative censoring after 5 years, and $\alpha = 0.5$. We also assume that longitudinal measurements are taken at $t_1 = 0$, $t_2 = 0.125$, $t_3 = 0.25$, $t_4 = 0.75$, $t_5 = 1$, $t_6 = 2$, $t_7 = 3$, $t_8 = 4$, and $t_9 = 5$ ($J = 9$), with survival times being categorized into one-year intervals with $K = 5$. Table 2 shows the results of these simulations with $I = 300$, $\alpha = 0.5$, $M = 10$, and different values of the measurement error σ_ϵ. We present the mean (standard deviation) of parameter estimates with complete longitudinal data, ORC, RRC, and the proposed approach. Although the proposed approach has increasing bias as σ_ϵ becomes large, this approach has less bias than both ORC and RRC for all values of σ_ϵ. Further, we conducted an additional simulation in which measurements t_3 to t_9 were missing with probability 0.5, creating datasets with fewer observations on each subject. Results were essentially the same as reported in Table 2, suggesting that our approach does well even with shorter sequences of longitudinal data (data not shown).

4. Discussion

This article proposes a simple regression calibration approach for estimating the relationship between longitudinal measurements and time-to-event data that account for informative dropout in the longitudinal process. The approach is not completely unbiased since the conditional distribution of the longitudinal process given the event time is approximated by a multivariate normal distribution. Particularly when the longitudinal and time-to-event processes are strongly linked, there may be small amounts of departure from normality. The effect of this lack of normality on a bias appears to increase as the measurement error increases. However, in most situations, the bias is substantially smaller than the ORC and RRC approaches proposed in Ye et al. (2008). The simulation results demonstrate that, in general, the proposed approach results in estimates with increased variance relative to ORC and RRC. More precise estimation is possible under a more parsimonious parameterization. For example, $\beta^*_{1T_i}$ in (6) may be modeled as linear in T_i.

Table 2
Estimates of α from model (3) and (4) and $\lambda(t, b_{i1}) = \lambda_0(t) \exp(\alpha \, b_{i1})$ where $I = 300$ and $M = 10$. We also assume that
$\beta_0 = 1, \beta_1 = 3, \sigma_{b_0} = 1, \sigma_{b_1} = 1$, and $\sigma_{b_{i0},b_{i1}} = 0$. The means (standard deviations) from 1,000 simulations are presented.

Parameters			Estimators of α		
σ_ϵ	α	Complete	ORC	RRC	Proposed
0.2	0.25	0.25(0.072)	0.22(0.069)	0.21 (0.071)	0.24 (0.084)
	0.50	0.50(0.077)	0.43(0.072)	0.42(0.072)	0.48 (0.088)
0.5	0.25	0.25(0.072)	0.19(0.066)	0.19 (0.067)	0.24(0.104)
	0.50	0.50(0.077)	0.37(0.069)	0.36(0.068)	0.48(0.109)
1.0	0.25	0.25(0.072)	0.14(0.067)	0.16 (0.063)	0.23(0.149)
	0.50	0.50(0.078)	0.32(0.067)	0.30(0.064)	0.46(0.155)

The proposed approach could be applied to a setting in which the two processes are linked through the true value of the longitudinal processes and time-to-event distribution. Further, the approach could be extended to allow for a more complex stochastic processes mean structure for the longitudinal process and for a semi-parametric fixed-effect structure as proposed by Ye et al. (2008). This would involve fitting model (6) or (9) with a different smooth curve $\varphi(t)$ and stochastic process $W_i(t)$ for each discretized dropout time. Such a model could be fit within the framework proposed by Zhang et al. (1998).

Our setup assumes that event times are only administratively censored after a fixed follow-up at the end of the study. For the case in which patients are censored prematurely, dropout times can be imputed based on a model fit using only patients who had the potential to be followed over the entire study duration.

REFERENCES

Carroll, R. J., Spiegelman, C. H., Lan, K. K. G., Bailey, K. T., and Abbott, R. D. (1984). On errors-in-variables for binary regression models. *Biometrika* **71,** 19–25.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Boostrap.* New York: Chapman and Hall.

Hsieh, F., Tseng, Y. K., and Wang, J. L. (2006). Joint modeling of survival and longitudinal data: Likelihood approach revisited. *Biometrics* **62,** 1037–1043.

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

Song, X., Davidian, M., and Tsiatis A. A. (2002). A semiparametric likelihood approach for joint modeling of longitudinal data and time-to-event data. *Biometrics* **58,** 742–753.

Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica* **14,** 809–834.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data.* New York: Springer Verlag.

Wu, M. C. and Bailey, K. R. (1989). Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. *Biometrics* **45,** 939–955.

Wu, M. C. and Follmann, D. A. (1999). Use of summary measurements to adjust for informative missingness repeated measures data with random effects. *Biometrics* **55,** 75–84.

Zhang, D., Lin, X., Raz, J., and Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association* **93,** 710–719.

Ye, W., Lin, X., and Taylor, J. M. G. (2008). Semiparametric modeling of longitudinal measurements and time-to-event data: A two stage regression calibration approach. *Biometrics* **64,** 1238–1246.

The authors replied as follows:

1. Introduction

The focus of our article, Ye, Lin, and Taylor (2008), hereafter referred to as YL&T, is on fitting a joint longitudinal survival model in the situation where we can make minimal assumptions about the population and individual longitudinal trajectories other than that they are smooth functions of time. We proposed a two-stage approach to avoid the complexities of joint estimation of longitudinal and survival models when the longitudinal model was specified as a semiparametric mixed model.

Albert and Shih (2009), hereafter referred to as A&S, note the two sources of potential bias with the two-stage method, ordinary regression calibration (ORC). As described in YL&T, in the first step of ORC a longitudinal model is fit to all the data. The first source of bias is from informative dropout in this longitudinal model. One way to remove this source of bias is the riskset regression calibration (RRC) method in YL&T, in which separate longitudinal models are fit to the historical data of subjects in the riskset at each event time. While this more computationally intensive approach should reduce the bias, the simulation results from both us and A&S found very little difference between the ORC and RRC results. The second source of bias is due to uncertainty in the estimate of the best linear unbiased prediction (BLUP) estimate, which is directly plugged into the partial likelihood. With the goal of reducing this bias, A&S proposed a useful modified two-stage approach. They nicely demonstrated that the proposed method has a good performance and outperforms the existing two-stage methods in simple settings.

The idea of A&S of imputing pseudo observations has the potential to be extended to deal with more complicated settings. However, we would like to point out that some challenges remain. We will concentrate on the two major assumptions made by A&S: (1) the longitudinal variable has a linear trajectory (Section 2); (2) event times are only administratively censored after a fixed follow-up at the end of the study (Section 3). Furthermore, we discuss different possible strategies for imputing pseudo observations and conduct simulation studies to compare two of them to the method of A&S, and the ORC method in two scenarios (Section 4). Conclusions and discussion are provided in Section 5.

2. The Linear Trajectory Assumption

In the preliminary stage of their proposed method, A&S used a pattern mixture model to estimate $\mathbf{X}_i \,|\, T_i$. In this model, each subpopulation (subjects with same event-time) has a unique set of parameters for its population mean trajectory. The only parameter shared between these subpopulations is the measurement error term.

In A&S's setting, since the true trajectory of $\mathbf{X}_i$ is linear and only subjects with more than two longitudinal observations are included in the analysis, the pattern mixture model of $\mathbf{X}_i \,|\, T_i$ can capture the feature of the complete trajectory with a good linear approximation. However, for very low T_i for which subjects had just one or two observations, the reliability of the estimates from the pattern mixture model would seem questionable. Consequently, there would be concerns about whether use of the future imputed values of X_i was actually reducing the bias due to informative dropout or increasing the bias due to unreliability of the values. This concern could be reduced if the separate longitudinal models given T_i are linked to each other, for example by forcing the parameters from each model to be a smooth function of T_i (Hogan, Lin, and Herman, 2004). The price to pay for this approach would be more computational complexity.

When the true trajectory is not linear, e.g., the nonlinear prostate-specific antigen (PSA) trajectory presented by YL&T, using a pattern mixed model similar to the one proposed by A&S, one might not be able to capture the features of the complete trajectory for some subpopulation, especially those who have very short follow-up time. In this situation, the quality of the imputation of the PSA values would be heavily reliant on the assumptions of the nonlinear model. In YL&T, we used a semiparametric mixed model to capture the smooth nonlinear pattern of PSA. Because of the structure of this model the imputation of the missing longitudinal measurements after the time of the last observed PSA, measurements would be based on linear extrapolation. This could be considerably biased for some subjects with early events.

3. Censoring Assumption

A&S assume that event times are only administratively censored after a fixed follow-up at the end of the study. However, in many applications, censoring occurs at a broad range of times, either because subjects often drop out of the study before it ends, or because subjects enter the study at different times. Thus follow-up time differs greatly for subjects who never experience the event during the study. For these censored patients, longitudinal measurements are also missing after the censoring time. Subjects who are censored at a specific time could have very different longitudinal trajectories from each other and from those whose event-time is close to their censoring time. Consequently, it is not appropriate to fit the pattern mixture model grouping together subjects with similar censoring times and event times. In discussing this, A&S stated that "For the case in which patients are censored prematurely dropouts can be imputed based on a model fit using only patients who had the potential to be followed over the entire study duration."

To follow up with their idea on imputing event-time, one could potentially impute the event times using a survival model based on baseline covariates for censored subjects. One might also use ORC or RRC in YL&T to include the longitudinal measurements as a time-dependent covariate, or other parametric (Faucett, Schenker, and Taylor, 2002) or nonparametric approaches (Hsu et al., 2006). The prediction of exact times is not precise, so a multiple imputation strategy would be preferred. This would detract from the computational simplicity of A&S's approach.

4. Simulation Studies to Compare A&S's Method with Alternative Methods

An interesting feature of A&S's procedure is that they simulate complete observations for each person. Thus the simulated data are supposed to "look like" the original data before the dropout time, and to exist after the dropout time. An alternative strategy would be to preserve the actual observed data for each subject and only simulate pseudo observations after the dropout time. Intuitively there would seem to be some inherent robustness in not replacing the observed data with pseudo data. For example, if the observed data are far from being multivariate normally distributed, then the A&S simulated data before the dropout time may not "look like" the real data. In this case it would seem less damaging to

subsequent analysis only to simulate pseudo data after the dropout time, rather than at all times.

To develop this further, we consider the models described in equations (2)–(4) of A&S and an approximation of the Bayesian imputation procedure. Recall that β^* and b^* denote the fixed and random effect vectors in the pattern mixture model, where $b^* \sim N(0, \Sigma_b^*)$ and ϵ is the error term, and $\epsilon \sim N(0, \sigma^{*2})$. We consider the following strategy as a modified A&S method (MAS).

Let X, X_0, and X_m denote the complete, observed, and missing due to dropout observations, respectively, and note that

$$[X_m \mid X_o] = \int [X_m \mid X_o, \theta, b^*][\theta, b^* \mid X_o]d\theta\, db^*$$

$$= \int [X_m \mid \theta, b^*][\theta, b^* \mid X_o]d\theta\, db^*, \tag{1}$$

where $\theta = (\beta^*, \sigma_\epsilon^{*2}, \Sigma_b^*)$. One can impute X_m by first drawing (θ, b^*) from the posterior and then drawing pseudo observations from a normal distribution conditional on them. In our approach, we replace the draw of $[\theta, b^* \mid X_0]$ by a draw from $[\beta^*, b^* \mid \hat{\sigma}_\epsilon^*, \hat{\Sigma}_b^*, X_0]$ at fixed values of $\hat{\sigma}_\epsilon^*$, $\hat{\Sigma}_b^*$, and then draw X_m conditional on the drawn value of β^* and b^*. A detailed description of this procedure is given in the Web Appendix.

We also note that the analytic formula for the correction for uncertainty (equation [7] in A&S) can be directly applied to the ORC method. We repeated and extended the simulation study of A&S to compare the ORC, A&S method, MAS, and their corresponding versions with uncertainty correction (UC).

We fit the probit model (2) of A&S with true b_{1i} as a benchmark method, which we refer to as the TRUE method. We vary the measurement error and the value of a_1 to investigate how the two strategies ((1) imputing the missing longitudinal measurements to reduce the bias; and (2) UC) work under difference scenarios.

For each simulation study, we simulated 500 data sets of sample size N = 300 based on equations (2)–(4) in A&S, and $\beta_0 = 1$, $\beta_1 = 3$, $\Sigma_b = I_{2 \times 2}$, $a_0 = -1.5$, $J = 5$. All individuals who are at risk at $t_5 = 5$ are administratively censored at that time point. To save computation time, we simulated M = 3 data sets in the multiple simulation step for all of our simulation studies.

Table 1 shows the results. When $a_1 = 0.5$, both the A&S and MAS methods perform very well. Even when σ_ϵ is at the largest, the bias in these two methods are still quite small (approximately 2%). The efficiency of the A&S and MAS methods is similar. The bias of ORC increases as measurement error σ_ϵ increases. UC works very well for the ORC method: the relative bias in all the three means of the ORC estimates with UC is less than 0.5%, which is even better than the A&S and MAS methods without UC. In addition, the efficiency of the ORC method with UC is better than both the A&S and MAS methods, especially when measurement error is large. Under this set-up, since both A&S and MAS work quite well without UC, it is hard to judge how well UC works for these two methods.

In Table 1, when $a_1 = 1$, all methods yield more bias compared to when $a_1 = 0.5$. Bias in all the three methods

Table 1

Estimates of a_1 from models (2)–(4) of A&S. Entries are Monte Carlo average of estimates (MEAN) Monte Carlo standard deviation (SD) from 500 simulations. Relative Bias (RB) is calculated with respect to the mean estimate from the TRUE method.

	$\sigma_\varepsilon = 0.25$			$\sigma_\varepsilon = 0.75$			$\sigma_\varepsilon = 1$		
	MEAN	RB(%)	SD	MEAN	RB(%)	SD	MEAN	RB(%)	SD
$a_1 = 0.5$									
True	0.503	–	0.077	0.503	–	0.077	0.503	–	0.077
ORC	0.489	−2.8	0.077	0.451	−10.3	0.075	0.437	−13.1	0.075
A&S	0.493	−2.0	0.089	0.494	−1.8	0.101	0.492	−2.2	0.105
MAS	0.503	0.0	0.083	0.498	−1.0	0.098	0.494	−1.8	0.103
ORC with UC	0.505	0.4	0.080	0.505	0.4	0.087	0.504	0.2	0.090
A&S with UC	0.494	−1.8	0.086	0.496	−1.4	0.102	0.496	−1.4	0.108
MAS with UC	0.503	0.0	0.083	0.501	−0.4	0.099	0.498	−1.0	0.106
$a_1 = 1$									
True	1.012	–	0.098	1.012	–	0.098	1.012	–	0.098
ORC	0.971	−4.1	0.097	0.871	−13.9	0.097	0.837	−17.3	0.097
A&S	0.944	−6.7	0.110	0.927	−8.4	0.120	0.923	−8.8	0.138
MAS	1.007	−0.5	0.105	0.979	−3.3	0.123	0.966	−4.5	0.131
ORC with UC	1.015	0.3	0.105	1.021	0.9	0.126	1.024	1.2	0.136
A&S with UC	0.948	−6.3	0.112	0.945	−6.6	0.127	0.949	−6.2	0.150
MAS with UC	1.012	0.0	0.107	1.000	−1.2	0.131	0.994	−1.8	0.143

without UC increases as measurement error σ_ϵ increases. MAS has the least bias among all the methods without UC at all three σ_ϵ values. When $\sigma_\epsilon = 1$, bias in the MAS estimator is approximately half as in the A&S estimator. Efficiency of the A&S and MAS methods seems to be similar in general. When measurement error is small, $\sigma_\epsilon = 0.25$, the ORC method is slightly better than the A&S method, but when $\sigma_\epsilon = 0.75$, 1, the ORC method yields the largest bias among the three methods without UC. Similar to where $a_1 = 0.5$, UC works very well for the ORC method: the relative bias in all the three means of the ORC estimates with UC is less than 1.5%. UC also works very well on the MAS methods (relative bias < 2%), although MAS with UC is less efficient than the ORC method with UC. UC does not have much effect on the A&S method, although it seems to decrease the bias a little bit when measurement error is large.

In A&S, they assume $P(Y_{ij} = 1 \,|\, Y_{i(j-1)} = 0) = \Phi(a_0 + a_1 \times b_{i1})$, where b_{i1} is the difference between the subject-specific slope and the population mean slope. Here we will also study the above methods under a slightly different survival model where

$$P\big(Y_{ij} = 1 \,|\, Y_{i(j-1)} = 0\big) = \Phi(a_0 + a_1 \times (b_{i1} + \beta_1)). \quad (2)$$

In this model it is the individual's slope, rather than the deviation of that person's slope from the population average that determines the hazard. Under this scenario, another alternative method, which is computationally simpler than MAS, is to replace $(b_{i1} + \beta_1)$ in (2) using draws from $[\beta^*, b^* \,|\, \hat{\sigma}_\epsilon^*, \hat{\Sigma}_b^*, X_0]$ directly rather than going through imputing X_m and ORC as in MAS. Similar to the MAS and A&S methods, three sets (M = 3) of (β^*, b^*) are drawn and a multiple imputation strategy is used. We call this method DirectMAS and study it together with the other six methods under the above model (2) in this article, and (3) and (4) of A&S. In addition, to study the behavior of all these seven methods under a

nonnormal scenario, we simulated random effects using two independent exponential distributions: $b_{0i} = \xi_{0i} - 1$, and $b_{1i} = \xi_{1i} - 1$, where $\xi_{0i} \sim \exp(1)$, $\xi_{1i} \sim \exp(1)$, and $\text{cov}(\xi_{0i}, \xi_{1i}) = 0$.

Table 2 shows simulation results in the setting where the risk of event depends on the subject-specific slope $(\beta_1 + b_{1i})$. To make the setting in Tables 1 and 2 more comparable, we adjust $a_0 = -3$ so that the event rate is similar between the scenarios presented in Tables 1 and 2. When (b_{0i}, b_{1i}) are normally distributed, conclusions that can be drawn from Table 2 are in general similar to those in Table 1. The additional method DirectMAS also works very well and yields results that are similar to MAS and better than ORC and A&S.

When (b_{0i}, b_{1i}) are from nonnormal distribution, all methods yield more biased estimate than in the normal setting. The A&S method yields severely biased estimates (19–22%), even when measurement error is small. The ORC method yields small bias when measurement error is small (1%), but substantial bias when measurement error is large (8%, 18%). MAS also yields considerable bias but much better than A&S at all three σ_ϵ values. The DirectMAS method seems to work very well, and the bias is only 3% even when measurement error is at its largest value. As expected, DirectMAS has much larger SE compared to other methods, which is a trade-off for less bias. Unlike in the normal setting, UC does not work for the ORC method, but instead increases the bias largely. This is not surprising, because the uncertainty equation (7) in A&S was derived under normality assumption for random effects. Similar to the normal setting, UC does not improve the A&S method. Under this nonnormal setting, UC also has little effect on the MAS method.

5. Conclusion and Discussion

A&S suggested a simple approach, which shows good performance in simple settings. For more complicated settings,

Table 2
Estimates of a_1 from model (2) in this article, and (3) and (4) of A&S. Entries are Monte Carlo average of estimates (MEAN), Monte Carlo standard deviation (SD) from 500 simulations. Relative Bias (RB) is calculated with respect to the mean estimate from the TRUE method.

	$\sigma_\varepsilon = 0.25$			$\sigma_\varepsilon = 0.75$			$\sigma_\varepsilon = 1$		
	MEAN	RB(%)	SD	MEAN	RB(%)	SD	MEAN	RB(%)	SD
$a_1 = 0.5$									
b_{0i} and b_{1i} are from normal distribution									
True	0.505	–	0.073	0.505	–	0.073	0.505	–	0.073
ORC	0.489	−3.2	0.071	0.451	−10.7	0.070	0.437	−13.5	0.071
A&S	0.495	−2.0	0.083	0.498	−1.4	0.095	0.493	−2.4	0.099
MAS	0.505	0.0	0.077	0.499	−1.4	0.086	0.495	−2.0	0.090
DirectMAS	0.506	0.2	0.080	0.499	−1.4	0.090	0.496	−1.8	0.083
ORC with UC	0.505	0.0	0.074	0.505	0.0	0.082	0.505	0.0	0.086
A&S with UC	0.496	−1.8	0.083	0.500	−1.0	0.097	0.497	−1.6	0.102
MAS with UC	0.505	0.0	0.077	0.502	−0.6	0.087	0.499	−1.2	0.092
$a_1 = 1$									
b_{0i} and b_{1i} are from normal distribution									
True	1.012	–	0.088	1.012	–	0.088	1.012	–	0.088
ORC	0.997	−1.5	0.088	0.954	−5.7	0.096	0.941	−7.0	0.102
A&S	0.984	−2.8	0.092	0.971	−4.1	0.114	0.962	−4.9	0.118
MAS	1.009	−0.3	0.094	1.004	−0.8	0.117	0.998	−1.4	0.127
DirectMAS	1.017	0.5	0.103	1.024	1.2	0.116	1.020	0.8	0.165
ORC with UC	1.018	0.6	0.094	1.034	2.2	0.118	1.039	2.7	0.130
A&S with UC	0.986	−2.6	0.093	0.990	−2.2	0.121	0.989	−2.3	0.129
MAS with UC	1.012	0.0	0.095	1.024	1.2	0.124	1.027	1.5	0.139
$a_1 = 1$									
b_{0i} and b_{1i} are from nonnormal distribution									
True	1.012	–	0.106	1.012	–	0.106	1.012	–	0.106
ORC	1.001	−1.1	0.111	1.089	7.6	0.150	1.186	17.2	0.195
A&S	0.824	−18.6	0.110	0.802	−20.8	0.122	0.791	−21.8	0.128
MAS	0.972	−4.0	0.110	0.933	−7.8	0.131	0.920	−9.1	0.146
DirectMAS	0.985	2.7	0.113	0.983	−2.9	0.155	0.981	−3.1	0.182
ORC with UC	1.043	3.1	0.124	1.216	20.2	0.191	1.324	30.8	0.250
A&S with UC	0.825	−18.5	0.111	0.813	−19.7	0.127	0.807	−20.3	0.136
MAS with UC	0.972	−4.0	0.110	0.948	−6.3	0.138	0.941	−7.0	0.157

such as the settings with nonlinear models or different censoring times, modifications of their simple imputation procedures might be feasible although not straightforward and can be computationally complex. If computational consideration is not an issue, then of course methods that use the full likelihood of the longitudinal and survival data are preferred rather than two-stage approximations.

The idea of improving the first stage of a two-stage approach to joint modeling of longitudinal and survival data can be implemented in a number of different ways. In this article, we suggested two other multiple imputation approaches, MAS and DirectMAS, which are slightly more computationally complex and do appear to have better properties. In addition, we found that the ORC method with UC also works very well when random effects are from multivariate normal distribution.

We have discussed a number of different multiple imputation strategies that are followed by the ORC method. A more computationally intensive alternative would be to follow the

imputation with the RRC method. In such a case the imputed values after the event-time would not be used at all and are irrelevant. If we had implemented the MAS procedure followed by RRC, it would have corresponded exactly to the original RRC procedure, which was biased. This leads us to believe that the major source of bias for the ORC and RRC methods in the situation that we and Albert and Shih (2009) have considered is the uncertainty in the BLUP estimates, rather than the bias due to informative drop-out. This suggests that the reason the MAS and DirectMAS methods we proposed did work and the A&S method is useful is because the multiple imputation method does introduce some uncertainty into the BLUP estimates, which helps reduce the bias due to this source.

The DirectMAS method, which directly uses imputed subject-specific slopes from the pattern mixture model in the second stage survival model, works very well not only under normal scenarios but also under nonnormal scenarios. This is probably due to the fact that the BLUP estimates in the

pattern mixture model are quite robust to nonnormal distribution. In contrast, in the MAS method, BLUP estimates from the linear mixed model fit to the pseudo observations are assumed to be normally distributed and therefore are less accurate than those from the pattern mixture model.

6. Supplementary Materials

The Web Appendix referenced in Section 4 is available under the Paper Information link at the *Biometrics* website `http://www.biometrics.tibs.org`.

References

Albert, P. S. and Shih, J. H. (2009). On estimating the relationship between longitudinal measurements and time-to-event data using a simple two-stage procedure. *Biometrics* DOI 10.1111/j.1541-0420.2009.01324.x.

Faucett, C. L., Schenker, N., and Taylor, J. M. G. (2002). Survival analysis using auxiliary variables via multiple imputation, with application to AIDS clinical trial data. *Biometrics* **58,** 37–47.

Hogan, J. W., Lin, X., and Herman, B. (2004). Mixtures of varying coefficient models for longitudinal data with discrete or continuous nonignorable dropout. *Biometrics* **60,** 854–864.

Hsu, C.-H., Taylor, J. M. G., Murray, S., and Commenges, D. (2006). Survival analysis using auxiliary variables via nonparametric multiple imputation. *Statistics in Medicine* **25,** 3503–3517.

Ye, W., Lin, X., and Taylor, J. M. G. (2008). Semiparametric modeling of longitudinal measurements and time-to-event data—A two-stage regression calibration approach. *Biometrics* **64,** 1238–1246.

Wen Ye[*] *and Jeremy M.G. Taylor*[**]
Department of Biostatistics,
University of Michigan,
Ann Arbor, Michigan 48109, U.S.A.
[*]email: *wye@umich.edu*
[**]email: *jmgt@umich.edu*

and

Xihong Lin
Department of Biostatistics,
Harvard School of Public Health,
Boston, Massachusetts 02115, U.S.A.
email: *xlin@hsph.harvard.edu*