Published in final edited form as:

Stat Methods Med Res. 2016 August ; 25(4): 1512–1526. doi:10.1177/0962280213490342.

Joint Modeling of Longitudinal Data and Discrete-Time Survival Outcome

Feiyou Qiu^{a,b,*,+}, Catherine M. Stein^{a,c,*,+}, and Robert C. Elston^a for the Tuberculosis Research Unit (TBRU)

^aDepartment of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH 44106, U.S.A

^bCurrent affiliation: Medical Mutual of Ohio, Cleveland, OH 44115, U.S.A

^cUganda-CWRU Research Collaboration

Abstract

A predictive joint shared parameter model is proposed for discrete time-to-event and longitudinal data. A discrete survival model with frailty and a generalized linear mixed model for the longitudinal data are joined to predict the probability of events. This joint model focuses on predicting discrete time-to-event outcome, taking advantage of repeated measurements. We show that the probability of an event in a time window can be more precisely predicted by incorporating the longitudinal measurements. The model was investigated by comparison with a two-step model and a discrete time survival model. Results from both a study on the occurrence of tuberculosis and simulated data show that the joint model is superior to the other models in discrimination ability, especially as the latent variables related to both survival times and the longitudinal measurements depart from 0.

Keywords

joint modeling; discrete time-to-event; longitudinal; nonlinear; biomarker; Tuberculosis, immunology

1. Introduction

Time-to-event data are collected and analyzed for various scientific purposes. Event times can be continuous or discrete. If the exact occurrence time of the event can be recorded and the event can happen at any time, the event time is assumed to be continuous. It is inappropriate to assume event times are continuous when events can only take place at discrete time points [1]. When the event is only known to have occurred in a time window, the exact event time is continuous but unavailable. Such data are referred to as interval censored data [2]. Events either taking place at discrete times or interval censored are generally referred as discrete time events. The results from Discrete Time Survival Analyses

*Correspondence to: Feiyou Qiu or Catherine M. Stein, Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH 44106, U.S.A.

⁺feiyou.qiu@case.edu, or catherine.stein@case.edu

(DTSA)[3] have better comprehensibility than the results of analyses using continuous methods when event times are discrete. With interval-censored data, a special case of discrete time-to-event data, the likelihood constructed taking interval censoring into account should be used for analysis. A generalized survival model was specified by Cox for both continuous and discrete times [4]. In the discrete time case, the general model becomes a logistic model. After Cox, others [1, 3, 5-8] reparameterized the discrete time survival model using a logistic regression format. i.e., the likelihood of the discrete time survival model is equivalent to the likelihood of the Bernoulli distribution [1, 5, 7].

When both longitudinal measurement data and time-to-event data are observed, joint modeling efficiently uses both types of data. There are two perspectives when carrying out joint modeling. One takes the longitudinal data as the analysis of major interest but incorporates event data, such as historical events or non-ignorable dropout, to avoid bias. The other perspective of joint modeling focuses on estimating the probability of the event at a specific time or in a time window given certain conditions (covariate status), and the effects of the longitudinal observations on the event are modeled in an attempt to better predict the time to event outcome.

Longitudinal measures can be directly introduced into the Cox model as time-varying covariates [9]. Bias will be the primary problem for this approach when the observations are measured with error or there is within-subject variability over time. A two-step model was proposed by Tsiatis et al. [10]. Wulfsohn et al. developed a joint Cox proportional hazard model, which can be utilized to assess the impact of a longitudinal covariate on a failure time process. A linear mixed effects model is assumed for the longitudinal process [11].

Another joint modeling strategy employs a “shared parameter” [12]. The shared parameter model was proposed by Vonesh et al. to model longitudinal data and time to dropout [13]. They model random effects in both the longitudinal outcome and the survival model to allow for the dropout mechanism and call the random effects shared parameters. The primary longitudinal outcome is modeled conditional on both shared parameters and other covariates. The survival outcome (dropout) is also modeled conditional on the shared parameters. They assume independence between the primary outcome and missingness conditional on the shared latent effects.

In this paper, we propose a joint shared parameter model to model both discrete time survival data and continuous longitudinal data. A discrete survival model with frailty and a generalized linear mixed model for the longitudinal data are joined to predict the probability of events. This joint model focuses on the discrete time-to-event outcome, taking advantage of repeated measurements, i.e. the probability of an event in a time window is more precisely predicted by incorporating the longitudinal measurements. The paper is organized as follows. In section 2, we outline the joint shared parameter model. In section 3, estimation is discussed. Data on a study of tuberculosis (TB) are introduced in section 4. The joint model is then applied to the TB data and compared with two other models in Section 5. Finally we present some simulation studies and conclusions in Section 6.

2. Model specification

The Cox generalized regression model [4] has a general form for both continuous and discrete time. For the case of discrete time points, the likelihood function can be written as [14]:

$$L_i = \prod_{j=1}^{p_i} h_{ij}^{d_{ij}} (1 - h_{ij})^{(1-d_{ij})}, \quad (1)$$

where i is the index of a subject and j is the index of a discrete event time. The number of follow-up periods can vary from subject to subject. For the i can subject, the index j can be $1, 2, 3 \dots p_i$, where p_i is the time period in which subject i experiences the event, drops out of the study or reaches the subject's maximum follow-up period. Let d_{ij} represent the event status of subject i in time period j , i.e. if subject i experiences the event in time period j , then $d_{ij} = 1$; otherwise $d_{ij} = 0$. We assume that the probability (h_{ij}) of an event at time j for subject i can be expressed as a function of covariates and latent random effects as follows:

$$h_{ij} = \Pr(d_{ij} = 1 | d_{ik} \neq 1, k < j; X_{ij}) = \frac{1}{1 + e^{-\alpha_j - X'_{ij}\gamma - u'_i\eta}}, \quad (2)$$

where α_j is a parameter for period j , X_{ij} is the vector of possible covariates, γ is a parameter vector for X_{ij} , and η is a parameter vector for the random-effect vector u_i , which is shared by the longitudinal model.

Let $g_d(d_{ij}|u_i)$ denote the density function of d_{ij} , $j=1, 2, \dots, p_i$, conditional on u_i . The conditional distribution of $D'_i = (d_{i1}, d_{i2}, \dots, d_{ip_i})$ will then be:

$$g_D(D_i|u_i) = \prod_{j=1}^{p_i} g_d(d_{ij}|u_i). \quad (3)$$

Assuming the subjects are independent, the joint conditional distribution of all events for all subjects becomes

$$g_D(D|u) = \prod_{i=1}^n g_D(D_i|u_i) = \prod_{i=1}^n \prod_{j=1}^{p_i} g_d(d_{ij}|u_i). \quad (4)$$

The event status d_{ij} conditional on u_i at discrete time points can then be modeled through logistic regression model.

Let y_{ik} denote the measurement of a response variable at time t_{ik} for subject i , $i = 1, 2, \dots, n$ and $k = 1, 2, \dots, n_i$, where the t_{ik} are continuous. Theoretically, the repeated measures can be measured before or after the event or right censoring. The follow-up schedule for

longitudinal measurements can be different from the follow-up schedule for checking event, i.e. n_i can be less or greater than p_i . We use the same notation for time-to-event data as before. Both repeated measures and survival data are modeled conditional on an unobserved latent variable u_i . Longitudinal data y_{ik} conditional on u_i are modeled as a linear mixed model [15]. For example, when a response variable has a linear trajectory over time, the linear mixed model can be written as:

$$y_{ik} = \beta_0 + \beta_1 t_{ik} + Z_{ik}' u_i + \varepsilon_{ik}, \quad (5)$$

where β_0 and β_1 are the fixed effects parameters and u_i is a vector of random effects, i.e. $u_i = (u_{i0}, u_{i1})'$, which contains the unknown random intercept u_{i0} and slope u_{i1} for subject i . Here, Z_{ik}' is a known 1×2 design vector.

We assume both ε_{ik} and u_i have normal distributions centered at 0. Independence is also assumed between ε_{ik} and u_i . The repeated measurements of (y_{ik}) conditional on the random effects (u_i) are then also independent. Let $g_y(y_{ik}|u_i)$ represents the density function of y_{ik} conditional on u_i . Assuming subjects are independent, the conditional joint distribution of repeated measurements for all subjects will be

$$g_Y(Y|U) = \prod_{i=1}^n g_Y(Y_i|u_i) = \prod_{i=1}^n \prod_{k=1}^{n_i} g_y(y_{ik}|u_i), \quad (6)$$

where U is a matrix of random effects, i.e. $U = (u_1, u_2, \dots, u_n)'$. Under the assumption of conditional independence between $g_D(D_i|u_i)$ and $g_Y(Y_i|u_i)$ given u_i , the joint distribution for D_i and Y_i conditional on u_i can be written as:

$$f(D_i, Y_i|u_i) = g_D(D_i|u_i) g_Y(Y_i|u_i) = \prod_{j=1}^{p_i} g_d(d_{ij}|u_i) \prod_{k=1}^{n_i} g_y(y_{ik}|u_i). \quad (7)$$

Let y_{ij}^* be the joint response variable. When $\delta_{ij} = 0$, this variable indicates event status. When $\delta_{ij} = 1$, this variable represents the longitudinal measurement. Here index $j = 1, \dots, J_i$, where $J_i = p_i + n_i$. Equation (7) can be written as:

$$f(D_i, Y_i|u_i) = f(Y_i^*|u_i) = \prod_{j=1}^{J_i} g_y(y_{ij}^*|u_i)^{\delta_{ij}} g_d(y_{ij}^*|u_i)^{1-\delta_{ij}}, \quad (8)$$

where Y_i^* is the joint response vector. The conditional joint distribution for all subjects becomes:

$$f(D, Y|U) = \prod_{i=1}^n g_D(D_i|u_i) g_Y(Y_i|u_i) = \prod_{i=1}^n \left(\prod_{j=1}^{J_i} g_y(y_{ij}^*|u_i)^{\delta_{ij}} g_d(y_{ij}^*|u_i)^{1-\delta_{ij}} \right). \quad (9)$$

We can also obtain the unconditional joint density for all subjects, i.e.

$$f(D, Y) = \prod_{i=1}^n \int g_D(D_i|u_i) g_Y(Y_i|u_i) du_i. \quad (10)$$

Although the joint model and the two-step method have the same model forms for both longitudinal data and survival data, the random effects u_i are specified and estimated differently in the two methods. In the two-step approach, the u_i are estimated using only the longitudinal data, in a first step. The estimation in this step does not take account of possible bias due to non-ignorable dropout. Another drawback of the two-step method is that the $\hat{u}_i$ are only estimates of the u_i and have measurement error, which can lead to bias. Parameters related to the survival time, including the parameter η , are estimated using the conditional distribution of survival data given u_i , with $\hat{u}_i$ substituted for u_i . However, in the shared parameter joint model approach, all parameters are simultaneously estimated from the joint likelihood. The joint model takes advantage of more data in the estimation process than the two-step method and the two-step method uses more data than the pure survival model. Therefore, we intuitively expect that the joint model will provide better prediction compared to the two-step method, and the two-step method will be better than a pure survival model.

3. Estimation

Since the joint shared parameter model is non-linear in some parameters and includes random effects u_i , it is a nonlinear mixed effects model (NLMM). In general, the integral in (10) does not have a closed form. Maximum likelihood estimation for the NLMM becomes a challenge because there is no closed form for the log likelihood [16]. We choose the Newton-Raphson (NR) algorithm along with the Laplace approximation to estimate the parameters of our joint model.

We use empirical Bayes (EB) estimation for the random effects. The observed data has conditional density function (8) and we assume that the random effect $u_i = (u_{i0}, u_{i1})'$ has a bivariate normal distribution, i.e. $u_i \sim N(0, G)$. Let $\pi(u_i|G)$ denote this distribution. The posterior distribution of u_i becomes

$$\begin{aligned} p(u_i|Y_i^*) &= \frac{f(Y_i^*|u_i)\pi(u_i|G)}{m(Y_i^*|G, M, \beta, \gamma)} \\ &= \frac{\left[\prod_{j=1}^{J_i} g_y(y_{ij}^*|u_i)^{\delta_{ij}} g_d(y_{ij}^*|u_i)^{1-\delta_{ij}} \right] \pi(u_i|G)}{\int_{-\infty}^{+\infty} \left\{ \left[\prod_{j=1}^{J_i} g_y(y_{ij}^*|u_i)^{\delta_{ij}} g_d(y_{ij}^*|u_i)^{1-\delta_{ij}} \right] \pi(u_i|G) \right\} du_i} \end{aligned} \quad (11)$$

In (11), G and M denote the variance of the random and residual vectors, respectively. Since there are unknown parameters (G, M, β, γ) in the posterior density function of u_j , estimators for these unknown parameters can be obtained through the marginal distribution $m(Y_i^* | G, M, \beta, \gamma)$ of all the observed data using MLE or some other estimation method. Once unknown parameters are estimated, the estimated posterior $\hat{p}(u_i | Y_i^*)$ can be obtained by plugging in these parameter estimates into (11). The estimated posterior mean of u_j , $\hat{p}(u_i | Y_i^*)$, is defined as the EB estimator of u_j for our shared parameter joint model. The expected value of u_j conditional on the observed data will be

$$E[u_i | Y_i^*] = \int_{-\infty}^{+\infty} u_i p(u_i | Y_i^*) du_i. \quad (12)$$

However, there are no known closed forms for (11) or (12). Therefore, we cannot directly estimate u_j using EB methods. A Monte Carlo Markov chain (MCMC) method can be used to integrate (12). The MCMC can generate a dependent sequence from a target distribution and then these samples can be used to estimate the quantities of interest using the Monte Carlo method.

The computational cost of estimation is sensitive to the starting values of parameters. In order to speed estimation, the estimates for the two-step model can be used as starting values for joint model estimation. Random effects are estimated using MCMC. Since the number of follow-up assessments for a subject is usually limited, the time of estimating random effect for one subject is relatively small. However, the aggregate time for all subjects will be considerable for large data, i.e. large number of subjects. Since our TB data has a small sample size, computation time is acceptable.

After obtaining the estimates of the fixed effect parameters, variance components and random effects, any quantity of interest, say $f(\beta, \Sigma, u_j)$, can be naturally predicted by

substituting these estimates into the function, i.e. as $f(\hat{\beta}, \hat{\Sigma}, \hat{u}_i)$. In our shared parameter model, the hazard of the event, i.e. the conditional probability of event h_{ij} given by equation (2), is of interest. The predicted hazard (h_{ij}) of an event can be estimated by substituting estimates of the fixed effect parameters, the variance parameters and the empirical Bayes estimates of random effects into (2), i.e.

$$\hat{h}_{ij}(\hat{\alpha}, \hat{\gamma}, \hat{\eta}, \hat{u}_i) = \frac{1}{1 + e^{-\hat{\alpha}_j - X_{ij}' \hat{\gamma} - \hat{u}_i' \hat{\eta}}} \quad (13)$$

The Delta method [17] is applied to compute the variance of the predicted hazard, i.e.

$$\text{var}(\hat{h}_{ij}) = \frac{\partial \hat{h}_{ij}(\theta)}{\partial \theta'} \text{Cov}(\theta) \frac{\partial \hat{h}_{ij}(\theta)}{\partial \theta}, \quad (14)$$

where θ is the vector of parameters, including the fixed effects parameters (α, γ, η) and random effects u_j , i.e. $\theta' = (\alpha', \gamma', \eta', u_j')$. In practice, the variance of prediction is usually

obtained through the unconditional mean squared error of prediction (UMSEP) [18], which is a naïve prediction variance. In this method, the parameters α , γ and η are assumed independent of u_i . The covariance matrix then becomes

$$Cov(\theta) = \begin{bmatrix} H^{-1} & 0 \\ 0 & Cov(u_i) \end{bmatrix}. \quad (15)$$

An estimate of the covariance matrix (H) for (α, γ, η) can be obtained from the inverse Hessian matrix, which is obtained during the estimation process, and $Cov(u_i)$ can be obtained using the MCMC process. However, in most situations the estimates of α , γ and η are not independent of u_i . The naïve method fails to account for the variability in estimating the parameters α , γ and η [19, 20]. Therefore, the naïve method leads to a biased estimate of the prediction variance. A correction term that takes the variability in the estimation of the hyperparameters into account must be added into the naïve estimate [19-23]. Booth and Hobert proposed conditional the mean squared error of prediction (CMSEP) as a general measure of prediction variance [18]. The CMSEP method is implemented in PROC NLMIXED for estimating the prediction variance. Using CMSEP, the variance matrix of θ can be constructed as [18],

$$Cov(\theta) = \begin{bmatrix} H^{-1} & H^{-1} \left(\frac{\partial \Omega}{\partial \theta} \right)^T \\ \left(\frac{\partial \Omega}{\partial \theta} \right) H^{-1} & \Lambda^{-1} + \left(\frac{\partial \Omega}{\partial \theta} \right) H^{-1} \left(\frac{\partial \Omega}{\partial \theta} \right)^T \end{bmatrix}, \quad (16)$$

where H is the estimated covariance matrix of (α, γ, η) and Ω is the EB estimator of the u_i , i.e. $\Omega = E[u_i | Y_i^*]$, and Λ is the covariance matrix of the u_i . Thus, the prediction variance $var(\hat{h}_{ij})$ is obtained by substituting the estimated and covariance matrix $Cov(\hat{\theta})$ into (14), i.e.

the covariance matrix $Cov(\theta)$ and the derivatives of $\frac{\partial \hat{h}_{ij}(\theta)}{\partial \hat{\theta}}$ and $\frac{\partial \hat{h}_{ij}(\theta)}{\partial \hat{\theta}}$ in (14) are all evaluated at the estimated $\hat{\theta}$.

4. TB immunology data

Data were taken from an ongoing longitudinal study of tuberculosis transmission within households, the Kawempe Community Health study (KCHS) [24, 25]. The KCHS systematically studies the processes of TB for both stages, i.e. latent tuberculosis infection (LTBI) and active TB disease, using a household contact design. People who live together in one residence were defined as a household. The first TB case identified in a household was defined as the index case. Household contacts were individuals who resided in a household with an index case at least 7 consecutive days during the 3-month period prior to the index case receiving a diagnosis of TB. As part of this study, household contacts were clinically evaluated, including tuberculin skin tests (TSTs) at baseline, and at 3, 6, 12, and 24 months after baseline [26, 27]. Blood draws for contacts who were TST negative at baseline were

scheduled at baseline, 1, 3, 6, 12, 24 months, as well as at sick visits. Details of the enzyme-linked immunosorbent assay (ELISA) are given in [28].

A total of 348 index cases and 1185 household contacts were enrolled into the KCHS between 2002 and 2006. Among 1185 household contacts, 85 were classified as coprevalent, 362 as TST negative, and 729 as TST positive with 9 of unknown status at baseline. A total of 139 of 362 TST negative contacts were found to have conversion to positive at follow-up visits. The exact time of TST conversion was unavailable, i.e. the event is interval censored. We define four discrete time periods for conversion, i.e. baseline to month 3 as period 1 (T_1), 3-6 months as period 2 (T_2), 6-12 months as period 3 (T_3), and 12-24 months as period 4 (T_4). None of the 362 contacts with negative TST at baseline were lost to follow-up prior to 24 months. Thus, censoring can only take place at 24 months because all TST negative contacts were followed at 3, 6, 12 and 24 months or until TST conversion was observed, and there were no dropouts.

The discrete time of conversion and longitudinal measurements are critical for answering the scientific question of whether these data, along with baseline information, can be used to predict the time of TST conversion. Therefore, subjects with missing or invalid times of conversions, or missing longitudinal measurements are excluded, and we only include the TST negative contacts who have at least one ELISA record with a valid time in this study. Among the cohort of 362 TST negative contacts, there are 264 contacts who have at least one ELISA record with a valid measurement time. Ninety eight of these (37.1%) have conversion event at T_1 , T_2 , T_3 or T_4 .

5. Predictive modeling using the TB data

To compare with our joint model, the TB data were also fitted using a two-step model and a pure survival model. The survival model should incorporate the longitudinal measurements as a regular covariate or a time-varying covariate. However, this strategy is not suited to highly unbalanced longitudinal data like the TB data. In the TB data, TST and biomarker measuring have different schedules and a high percentage of subjects did not follow the schedule of blood drawing. There are both discrete times for TST conversion and continuous times for repeated measure. A Cox regression model can be applied to discrete time survival data by treating the discrete time as continuous. Repeated measures can then be incorporated into survival model as time dependent covariates. However, the method of assuming the discrete time period as continuous time would lead to biased parameter estimation [2] and this approach will generally give inconsistent estimators of the relative risk function [29]. Furthermore, the asymptotic bias increases as the discrete times become coarser. The TB data only have four unequal time periods which are very coarse.

A pure survival model was fitted only using TST conversion data. Our intent to model this pure survival model is to show that the longitudinal measurements are useful in predicting TST conversion, by comparing the two-step model and the joint model to the pure survival model. The results of all the three models with the covariate (age) in the survival model are shown in Table 1. All parameters, including the coefficient (γ) for age, were significant at $\alpha = 0.05$. In particular, p-values for the coefficient (η) of the random effects and for the

intercept (β_0) and slope (β_1) in the longitudinal model are very small (<0.05), for both the two-step and joint models, whether including baseline age or not. The biomarker MTB6D7 was a statistically significant predictor of TST conversion in both the two-step and joint model, in contrast to the pure survival model. The estimated odds ratios for the effect of baseline age on survival are 1.042(se=0.013), 1.036(se=0.010) and 1.027(se=0.009) for the joint, the two-step, and the survival model, respectively. These odds ratios indicate increased odds of conversion when the baseline age increases and other covariates are fixed. Increasing age is associated with an increased risk of TST conversion. All p-values (0.0023, 0.0006 and 0.0005 for the survival model, the two-step model and joint model, respectively) from nested likelihood ratio tests indicate that the covariate age should be included in all three models. Other covariates might also be investigated using the same process. HIV status and gender were also separately modeled as covariates along with age in the survival model portion of each of the three models. The p-values of approximate t tests are 0.6730 (joint), 0.5811 (two-step) and 0.0282 (survival) for HIV status, and 0.0872 (joint), 0.0802 (two-step) and 0.0962 (survival) for gender. Large p-values (>0.05), except 0.0282 for HIV status in the discrete survival model, suggest that HIV status and gender need not be included in the model.

Owing to the small sample size (264 subjects), split-sample validation is inappropriate. The .632 bootstrap validation method [30] was used for validation and evaluating the performances of the three models. With the .632 bootstrap method, the apparent performance was first obtained by fitting and validating the three models using the original sample. The estimates of fixed effects parameters and variance components were obtained by fitting the three models using the bootstrap sample. These estimates were used to evaluate the prediction performance for subjects not in the bootstrap sample, which is called the test sample. The bootstrap procedure was repeated 150 times. After the average test performance was subtracted from the apparent performance, the remainder was weighted by 0.632. This weighted remainder is the estimated optimism. The overall performance was then obtained by subtracting the optimism from the apparent performance, i.e. estimated performance=apparent performance $- .632 \times$ [apparent performance - average (performance on the test samples)] = $0.368 \times$ apparent performance + $0.632 \times$ average (performance on the test sample).

The performance of prediction for the three models for the TB data were evaluated and compared using the Brier score, AUC, Hosmer-Lemeshow (H-L) statistic and Kolmogorov-Smirnov (K-S) statistic, which were estimated using bootstrapping. The Brier score can be partitioned into a calibration part and a refinement part [31]. The model with the smaller Brier score is preferred. The AUC and K-S are statistics for measuring the discrimination ability of a model. Larger values for AUC and K-S are desirable for a model. The K-S test was also conducted to test the discrimination significance of the model. The calibration for all models was assessed using the H-L statistic and H-L test. The results of these measures for each period using the bootstrap validation method are tabulated in Table 2.

For periods 1 and 2, the joint model has the smallest Brier Score and the two-step model has a smaller Brier Score than the pure survival model. However, the two-step model has a smaller Brier score than the other two models for periods 3 and 4. The pure survival model

has the largest Brier score across all periods. When ranked from highest to lowest according to the AUCs and K-S statistics, the rankings are the joint model (highest), the two-step model, and the survival model (lowest), across all periods. The average p-values for the K-S tests are all small with the exception of the survival model for period 1. All three models have a high percentage of p-values for the K-S test that are less than 0.05 in all periods, the highest being the survival model for period 1 (60.7%).

Paired t-tests show that the AUCs differed significantly between all three models at all periods ($p\text{-value} < 0.0001$). Therefore, the shared parameter joint model is significantly better than the other two models in terms of discrimination ability as measured by the AUC, and the two-step model had significantly better discrimination ability than the discrete survival model. We conclude that the shared parameter joint model has the best discrimination performance among the three models, and that the two-step model has better discrimination performance than the discrete time survival model.

The calibration of the three models was investigated using H-L statistics and H-L tests. Bootstrap samples with prediction for all bootstrapping steps were pooled together. For the H-L method, the subjects were partitioned into 5 groups according to their ranked predicted probabilities for each model at each period, for both the bootstrap sample and test sample. The expected number of events and the observed number of events for these 5 groups at each time period were used to calculate the H-L statistic for the bootstrap samples. The H-L statistic test samples (original TB data) were also obtained using the same procedure. The overall H-L statistic was also obtained using the .632 method and the p-value for the H-L test was calculated by assuming a chi-square distribution with 3 degree of freedom for the H-L statistic. The results of H-L statistics and H-L tests are listed in Table 2. The shared parameter joint model had the highest H-L statistic across all periods. The H-L statistic for the two-step model is higher than the survival model for all periods except period 3. The p-values of the H-L tests for the joint model are low (< 0.05) for periods 1 and 4. The two-step and survival models have large p-values of H-L tests across all periods. Therefore, the two-step and the survival models for all periods are at an acceptable level for calibration. The joint model's calibration for periods 2 and 3 is acceptable. However, based on calibration results, we should be cautious in applying the joint model to periods 1 and 4 due to low p-values for the H-L test. A trade-off exists between discrimination and calibration since a model usually cannot be perfect in both [32]. If the ability to discriminate subjects with versus without TST conversion is of more interest than the calibration of the prediction, the joint model is suitable for the prediction of periods 1 and 4. In summary, for these data the shared parameter joint model had better overall performance and discrimination ability than the two-step model and the discrete survival model when bootstrap validation was performed.

6. Simulation

We conducted various simulation studies to investigate our model's estimation accuracy and prediction performance compared to the two-step model and the pure survival model. Factors of prediction performance were also studied in simulations. All simulation studies were based on the TB data, i.e. the true values for all parameters were from the joint model

of the TB data or one parameter was varied and others remained unchanged. The sample size of the simulated data was 300, which was similar to the TB data ($n=264$). The missing rate of longitudinal measurements was set to be 0.8 which was the rate of missingness in the TB data. We simulated four measurement periods which were the same as for the TB data, but we did not consider non-scheduled measurements, such as sick visits. We also did not consider the situation of dropout in the simulations. A total of 1000 replicates were generated using estimates from the joint model under our TB data as the true values of parameters and were used to fit the discrete time survival model, the two-step model and the shared parameter joint model. From the results of fitting the three models on simulated data sets, estimates and the 95% confidence intervals of parameters were obtained. For each we also calculated the percentage of CIs that covered the true value.

The estimated bias was found by subtracting the true value from the average of the estimates and the mean square error (MSE) was defined by

$$MSE = Bias^2 + Variance \quad (17)$$

The percentage of bias for an estimator was also estimated as

$$\%bias = \frac{\text{Average of estimates} - \text{True}}{|\text{True}|} \times 100 \quad (18)$$

In Table 3, the estimates from the joint model had larger variance and MSE than the other two models for all parameters except the variance components. The joint model also had smaller absolute bias and % bias. We also noticed that the joint model had the best coverage of 95% confidence intervals for almost every parameter as compared to the two-step model and the pure survival model. The average estimates from the survival model had larger %bias and lower coverage probability due to the misspecification of the discrete survival model for the simulated data. That is, the survival model did not account for the random effect that was included when simulating the discrete survival data. Also, the occurrence of a conversion event for a subject led to an informative dropout (i.e. missing values for the biomarker after the time of conversion) for this subject when the coefficient η was not zero. While the joint model correctly accounted for this informative (non-ignorable) dropout, the two-step model did not. This was one of the reasons that explain the large %bias and the low coverage probability for η and β_I when estimated by the two-step model. Another explanation for the larger negative bias in η for the two-step model is that it uses an estimate of the random effect u_i for the true value, and the presence of measurement error causes attenuation in the estimate of η . All properties indicated that the joint model and the two-step model have better estimation as compared to the survival model, and the joint model has either better or similar properties as compared to the two-step model.

Another similar simulation was conducted setting the true value of η to zero. In this simulation, the joint model, the two-step model and the survival model had low absolute bias, low absolute % bias, low MSE and high coverage of 95% confidence intervals. Since

the magnitude of the coefficient (η) was zero, the occurrence of the conversion event for a subject leads to a non-informative (ignorable) dropout. The two-step model did not have an informative dropout problem in this case. There was also no misspecification for the survival model when the coefficient (η) was zero since the event is independent of the longitudinal measurements.

Data from the simulations carried out under the TB data scenario were also used to evaluate the prediction performance for the discrete time survival model, the two-step model and the shared parameter joint model. Apparent validation was applied to every simulated data set. The averages of AUC, Brier score, K-S and H-L for the three models across all simulated data sets are tabulated in Table 4. The shared parameter model has the smallest average Brier scores for all periods. The two-step model has smaller average Brier scores than the survival model. The averages of the AUCs and K-S statistics ranked from high to low for the joint model, the two-step model and the survival model across all periods. The average p-values for the K-S tests were all small (<0.05) with the exception of the survival model for period 1. We also noticed that the joint model and the two-step model had the highest percentage (100%) of K-S test p-values that were less than 0.05 in all periods. With the exception of the shared parameter model and the two-step model in period 1, all average p-values of H-L tests were large (>0.05). Therefore, simulation studies led to the conclusion that the joint model had the best discrimination ability, and the two-step model had better discrimination ability than the pure survival model. However, the shared parameter model had a noticeably high percentage of the H-L test p-values less than 0.05 in all periods. Thus, in all periods the shared parameter model was not as well calibrated as the two-step model and the survival model. A model usually cannot be perfect in both discrimination and calibration [32]. In practice, discrimination is of more interest than calibration. Therefore, simulation studies show that joint model is preferable according to discrimination ability.

We also studied the models under different scenarios in order to investigate potential factors, in particular the coefficient of longitudinal measurement (η) and the intrasubject correlation coefficient (ICC) that might have an effect on the prediction performance. A base scenario was defined slightly differently from the TB data. In this base scenario, η was set to 1, the missing rate of longitudinal measurements was 0.4, sample size increased to 600 and ICC $\rho = 0.5$. We simulated 500 data sets for each scenario, varying ρ and η . The average AUCs are listed in Table 5. The results of these simulation studies under the 5 scenarios provided evidence that the ICC (ρ) and the coefficient of random intercepts (η) have a large effect on the prediction performance. The effect of the coefficient of longitudinal measurement (η) on prediction performance in the two-step model and the shared parameter joint model were then investigated through a more intensive simulation study. The coefficient (η) was varied from -3.2 to 3.2 by steps of 0.015 , i.e. a total of 426 data sets were simulated. Apparent validation was then conducted for each of the four periods of each data set. Then the AUCs of the three models for all periods for each data set were calculated. The AUCs of period 4 are plotted against the coefficient (η) in Figure 1 as a demonstration. In Figure 1, the solid line and the dashed line are two smoothed lines using LOESS for the joint shared parameter model and the two-step model, respectively. These show that, except when η is near zero, the joint model has higher AUCs than the two-step model. The other measures (Brier score and K-S) were also computed and plotted against the coefficient (η). The plot in Figure 1

confirms that the shared parameter joint model has better prediction performance (calibration) than the two-step model when $\eta = 0$, with the difference increasing the farther η is from zero.

In conclusion, the results of applying the three models to TB data and simulation studies based on these data show that the shared parameter model has the best discrimination ability in predicting the probability of TST conversion at a specific time period by incorporating the longitudinal biomarker measurements.

Acknowledgments

The authors are thankful to Dr. Mark D. Schluchter at Case Western Reserve University for useful discussion, insightful comments and key direction. This work was funded in part by National Institutes of Health (NIH) grants R01-HL096811 and R01-HL106798 from NHLBI, and Tuberculosis Research Unit (grant HHSN266200700022C/N01-AI70022 from NIAID). We would also like to thank Dr. W. Henry Boom (PI of TBRU), LaShaunda Malone, Bonnie Thiel, and Keith Chervenak for their roles in data collection and management. This study could not have been conducted without the efforts of medical officers, home visitors, immunology lab technicians, study coordinators, and data management staff in Kampala, Uganda.

BIBLIOGRAPHY

1. Allison PD. Discrete-Time Methods for the Analysis of Event Histories. *Sociological methodology*. 1982; 13:61–98. (ArticleType: research-article / Full publication date: 1982 / Copyright © 1982 John Wiley & Sons).
2. Lindsey JC, Ryan LM. Methods for interval-censored data. *Statistics in Medicine*. 1998; 17(2):219–238. [PubMed: 9483730]
3. Singer JD, Willett JB. It's about Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events. *Journal of Educational Statistics*. 1993; 18(2):155–195.
4. Cox DR. Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1972; 34(2):187–220.
5. Brown CC. On the use of indicator variables for studying the time-dependence of parameters in a response-time model. *Biometrics*. 1975; 31(4):863–72. [PubMed: 1203428]
6. Efron B. Logistic Regression, Survival Analysis, and the Kaplan-Meier Curve. *Journal of the American Statistical Association*. 1988; 83(402):414–425.
7. Laird N, Olivier D. Covariance Analysis of Censored Survival Data Using Log-Linear Analysis Techniques. *Journal of the American Statistical Association*. 1981; 76(374):231–240.
8. Thompson WA, J. On the Treatment of Grouped Observations in Life Studies. *Biometrics*. 1977; 33(3):463–470. [PubMed: 911970]
9. Ibrahim JG, Chu H, Chen LM. Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*. 2010; 28(16):2796–801. [PubMed: 20439643]
10. Tsiatis AA, DeGruttola V, Wulfsohn MS. Modeling the Relationship of Survival to Longitudinal Data Measured with Error. Applications to Survival and CD4 Counts in Patients with AIDS. *Journal of the American Statistical Association*. 1995; 90(429):27–37.
11. Wulfsohn MS, Tsiatis AA. A Joint Model for Survival and Longitudinal Data Measured with Error. *Biometrics*. 1997; 53(1):330–339. [PubMed: 9147598]
12. Follmann D, Wu M. An approximate generalized linear model with random effects for informative missing data. *Biometrics*. 1995; 51(1):151–68. [PubMed: 7766771]
13. Vonesh EF, Greene T, Schluchter MD. Shared parameter models for the joint analysis of longitudinal data and event times. *Stat Med*. 2006; 25(1):143–63. [PubMed: 16025541]
14. Wu L, et al. Analysis of longitudinal and survival data: Joint modeling, inference methods, and issues. *Journal of Probability and Statistics*. 2011; 2012
15. McCullagh, P.; Nelder, J. *Generalized Linear Models*. Second Edition. Taylor and Francis; 1989.

16. Wang J. EM algorithms for nonlinear mixed effects models. *Computational Statistics & Data Analysis*. 2007; 51(6):3244–3256.
17. Oehlert GW. A Note on the Delta Method. *The American Statistician*. 1992; 46(1):27–29.
18. Booth JG, Hobert JP. Standard Errors of Prediction in Generalized Linear Mixed Models. *Journal of the American Statistical Association*. 1998; 93(441):262–272.
19. Carlin BP, Gelfand AE. Approaches for Empirical Bayes Confidence Intervals. *Journal of the American Statistical Association*. 1990; 85(409):105–114.
20. Carlin BP, Gelfand AE. A Sample Reuse Method for Accurate Parametric Empirical Bayes Confidence Intervals. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1991; 53(1):189–200.
21. Laird NM, Louis TA. Empirical Bayes Confidence Intervals Based on Bootstrap Samples. *Journal of the American Statistical Association*. 1987; 82(399):739–750.
22. Morris CN. Parametric Empirical Bayes Inference: Theory and Applications. *Journal of the American Statistical Association*. 1983; 78(381):47–55.
23. Piegorsch WW, Casella G. Empirical Bayes Estimation for Logistic Regression and Extended Parametric Regression Models. *Journal of Agricultural, Biological, and Environmental Statistics*. 1996; 1(2):231–249.
24. Guwatudde D, et al. Tuberculosis in Household Contacts of Infectious Cases in Kampala, Uganda. *American Journal of Epidemiology*. 2003; 158(9):887–898. [PubMed: 14585767]
25. Stein CM, et al. Heritability Analysis of Cytokines as Intermediate Phenotypes of Tuberculosis. *The Journal of Infectious Diseases*. 2003; 187(11):1679–1685. [PubMed: 12751024]
26. Mahan CS, et al. Innate and Adaptive Immune Responses during Acute M. tuberculosis Infection in Adult Household Contacts in Kampala, Uganda. *The American Journal of Tropical Medicine and Hygiene*. 2012; 86(4):690–697. [PubMed: 22492155]
27. Stein CM, et al. Genome Scan of M. tuberculosis Infection and Disease in Ugandans. *PLoS One*. 2008; 3(12):e4094. [PubMed: 19116662]
28. Mahan Cs Fau - Zalwango S, et al. Innate and adaptive immune responses during acute M. tuberculosis infection in adult household contacts in Kampala, Uganda. 2012 (1476-1645 (Electronic)).
29. Prentice RL, Gloeckler LA. Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data. *Biometrics*. 1978; 34(1):57–67. [PubMed: 630037]
30. Efron, B.; Tibshirani, R. Monographs on statistics and applied probability. 1st CRC Press reprint. Vol. xvi. Chapman & Hall/CRC; Boca Raton, Fla.: 1998. An introduction to the bootstrap; p. 436
31. Abramowitz, M.; Stegun, IA. Handbook of mathematical functions with formulas, graphs, and mathematical tables. U.S. Govt. Print. Off; 1964.
32. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 2007; 115(7):928–935. [PubMed: 17309939]

Figure 1.
AUCs vs. the coefficient (η) at period 4 for joint and two-step models

Table 1

Results for three models with age as a covariate

	Model	alpha1 (Period 1)	alpha2 Period 2)	alpha3 (Period 3)	alpha4 Period 4)	gamma (Age)	eta (Coefficient of Random effect)	beta0 (Intercept)	beta1 (Slope)	Psi_Sq (random)	Sigma_Sq (residual)
Estimate	Joint	-1.9965	-3.3536	-3.9501	-3.5721	0.0411	1.0508	5.7510	0.0013	1.4594	1.8513
	Two- step	-1.7082	-3.1514	-3.8010	-3.4677	0.0350	0.8733	5.7563	0.0008	1.4473	1.8588
	Survival	-1.4533	-2.9384	-3.6005	-3.2870	0.0271	-	-	-	-	-
StdErr	Joint	0.3122	0.3799	0.4604	0.4165	0.0120	0.2525	0.1026	0.0005	0.2212	0.1596
	Two- step	0.2221	0.3277	0.4261	0.3854	0.0099	0.1568	0.0991	0.0003	0.2237	0.1620
	Survival	0.1934	0.3085	0.4107	0.3689	0.0086	-	-	-	-	-
P-Value	Joint	<0.0001	<0.0001	<0.0001	<0.0001	0.0007	<0.0001	<0.0001	0.0059	<0.0001	<0.0001
	Two- step	<0.0001	<0.0001	<0.0001	<0.0001	0.0004	<0.0001	<0.0001	0.0240	<0.0001	<0.0001
	Survival	<0.0001	<0.0001	<0.0001	<0.0001	0.0017	-	-	-	-	-

Table 2

Assessment measures of models using .632 bootstrap validation

Period	Model	AUC	Brier Score	H-L	H-L Test p-value	K-S	Average K-S Test p-value	% K-S P-Value <0.05
1	Joint	0.7408	0.1634	24.7185	<0.0001	0.3829	0.0272	86.7%
	Twostep	0.6671	0.1668	7.4472	0.0589	0.2979	0.0167	94.0%
	Survival	0.5982	0.1896	3.3271	0.3439	0.2336	0.0924	60.7%
2	Joint	0.7379	0.0631	6.7444	0.0805	0.5055	0.0058	100.0%
	Twostep	0.6690	0.0634	1.7860	0.6180	0.4237	0.0063	100.0%
	Survival	0.5207	0.0679	0.7043	0.8722	0.2614	0.0465	82.7%
3	Joint	0.8239	0.0348	6.6741	0.0830	0.6950	<0.0001	96.7%
	Twostep	0.7624	0.0345	2.5890	0.4594	0.5980	0.0001	96.7%
	Survival	0.7158	0.0360	3.1196	0.3736	0.5049	0.0001	96.7%
4	Joint	0.8285	0.0460	10.0474	0.0182	0.7073	<0.0001	98.7%
	Twostep	0.7792	0.0455	4.5833	0.2050	0.6154	<0.0001	98.7%
	Survival	0.5923	0.0492	2.2600	0.5202	0.3774	0.0150	91.3%

Table 3

Estimation results from simulation study based on the TB data

Parameter	TRUE	Model	Average of Estimate	Variance	Bias	%Bias	MSE	% Coverage of CI at $\alpha=0.05$
alpha1	-1.9965	Joint	-2.181	8.311	-0.184	-9.22%	8.345	95.08%
		Two Step	-1.744	0.104	0.253	12.66%	0.168	84.45%
		Survival	-1.542	0.091	0.454	22.74%	0.297	64.54%
alpha2	-3.3536	Joint	-3.504	5.576	-0.150	-4.49%	5.599	95.75%
		Two Step	-3.206	0.149	0.148	4.41%	0.171	93.29%
		Survival	-3.047	0.137	0.306	9.14%	0.230	85.12%
alpha3	-3.9501	Joint	-4.131	4.860	-0.181	-4.58%	4.893	94.97%
		Two Step	-3.903	0.257	0.047	1.20%	0.260	92.73%
		Survival	-3.767	0.233	0.183	4.64%	0.267	90.38%
alpha4	-3.5721	Joint	-3.717	3.241	-0.145	-4.06%	3.262	95.19%
		Two Step	-3.555	0.208	0.017	0.49%	0.208	94.52%
		Survival	-3.450	0.194	0.122	3.42%	0.208	91.95%
gamma	0.0411	Joint	0.045	0.004	0.004	9.35%	0.004	95.30%
		Two Step	0.036	<0.0001	-0.005	-12.53%	<0.0001	93.51%
		Survival	0.031	<0.0001	-0.010	-24.19%	<0.0001	88.93%
eta	1.0508	Joint	1.190	6.950	0.140	13.29%	6.970	95.08%
		Two Step	0.911	0.036	-0.140	-13.28%	0.055	72.71%
beta0	5.751	Joint	5.748	0.011	-0.003	-0.06%	0.011	93.18%
		Two Step	5.741	0.011	-0.010	-0.17%	0.011	93.18%
beta1	0.0013	Joint	0.001	<0.0001	-2.12E-06	-0.16%	<0.0001	95.75%
		Two Step	0.001	<0.0001	-0.001	-50.82%	<0.0001	79.75%
Psi_Sq	1.459441	Joint	1.470	0.059	0.010	0.70%	0.059	94.74%
		Two Step	1.385	0.065	-0.075	-5.12%	0.070	97.32%
Sigma_Sq	1.8513	Joint	1.833	0.033	-0.018	-0.99%	0.033	93.29%
		Two Step	1.883	0.037	0.032	1.73%	0.038	94.63%

Table 4

Average assessment measures of models using apparent validation under TB scenario*

Period	Model	Average AUC	Average Brier Score	Average H-L	Average H-L Test P-Value	% H-L P-Value <0.05	Average K-S	Average K-S Test P-Value	% K-S P-Value <0.05
1	Joint	0.8780	0.1302	28.3484	0.0122	94	0.5156	<0.0001	100
	Twostep	0.7089	0.1723	7.9086	0.5008	4	0.3141	<0.0001	100
	Survival	0.5642	0.1921	8.3711	0.4700	5	0.1408	0.0789	71
2	Joint	0.8673	0.0595	14.5311	0.1698	37	0.6114	<0.0001	100
	Twostep	0.7020	0.0669	8.9718	0.4339	8	0.3775	0.0001	100
	Survival	0.5591	0.0698	8.9115	0.4260	6	0.2286	0.0146	92
3	Joint	0.8603	0.0351	10.6052	0.3326	13	0.6570	<0.0001	100
	Twostep	0.7004	0.0377	8.8672	0.4600	9	0.4339	0.0003	100
	Survival	0.5597	0.0386	9.1527	0.4151	8	0.3101	0.0029	98
4	Joint	0.8521	0.0446	11.0120	0.3015	16	0.6221	<0.0001	100
	Twostep	0.6967	0.0483	8.9641	0.4460	8	0.4077	0.0005	100
	Survival	0.5494	0.0497	8.9948	0.4227	7	0.2761	0.0059	97

Note: The TB scenario was simulated based on the TB data, i.e. the true values for all parameters are from the joint model of the TB data, the sample size is 300 and the missing rate of longitudinal measurements is set to be 0.8, which is the rate of missingness in the TB data.

Table 5

Average AUCs for 5 simulation scenarios

Scenario	$(\eta, \% \text{missing}, N, \text{ICC})$	Period 1		Period 2		Period 3		Period 4	
		Joint	Two step	Joint	Two step	Joint	Two step	Joint	Two step
Base	(1.0, 0.4, 600, 0.5)	0.8703	0.7074	0.8512	0.6908	0.8492	0.6885	0.8380	0.6920
LowICC	(1.0, 0.4, 600, <u>0.2</u>)	0.7937	0.6115	0.7644	0.6073	0.7498	0.6053	0.7453	0.6208
HighICC	(1.0, 0.4, 600, <u>0.8</u>)	0.9327	0.8604	0.9025	0.8325	0.8794	0.8162	0.8740	0.8194
LowEta	(<u>0.0</u> , 0.4, 600, 0.5)	0.5899	0.5856	0.5907	0.5882	0.5681	0.5640	0.5691	0.5672
HighEta	(<u>2.0</u> , 0.4, 600, 0.5)	0.9755	0.7943	0.9526	0.7720	0.9382	0.7804	0.9229	0.7747