

RESEARCH ARTICLE

Open Access



Comparison of a time-varying covariate model and a joint model of time-to-event outcomes in the presence of measurement error and interval censoring: application to kidney transplantation

Kristen R. Campbell¹, Elizabeth Juarez-Colunga^{2,3*} , Gary K. Grunwald², James Cooper⁴, Scott Davis⁴ and Jane Gralla^{1,2}

Abstract

Background: Tacrolimus (TAC) is an immunosuppressant drug given to kidney transplant recipients post-transplant to prevent antibody formation and kidney rejection. The optimal therapeutic dose for TAC is poorly defined and therapy requires frequent monitoring of drug trough levels. Analyzing the association between TAC levels over time and the development of potentially harmful de novo donor specific antibodies (dnDSA) is complex because TAC levels are subject to measurement error and dnDSA is assessed at discrete times, so it is an interval censored time-to-event outcome.

Methods: Using data from the University of Colorado Transplant Center, we investigated the association between TAC and dnDSA using a shared random effects (intercept and slope) model with longitudinal and interval censored survival sub-models (JM) and compared it with the more traditional interval censored survival model with a time-varying covariate (TVC). We carried out simulations to compare bias, level and power for the association parameter in the TVC and JM under varying conditions of measurement error and interval censoring. In addition, using Markov Chain Monte Carlo (MCMC) methods allowed us to calculate clinically relevant quantities along with credible intervals (CrI).

Results: The shared random effects model was a better fit and showed both the average TAC and the slope of TAC were associated with risk of dnDSA. The simulation studies demonstrated that, in the presence of heavy interval censoring and high measurement error, the TVC survival model underestimates the association between the survival and longitudinal measurement and has inflated type I error and considerably less power to detect associations.

Conclusions: To avoid underestimating associations, shared random effects models should be used in analyses of data with interval censoring and measurement error.

Keywords: Bayesian analysis, Interval censoring, Measurement error, Shared random effects

*Correspondence: elizabeth.juarez-colunga@cuanschutz.edu

²Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus, 80045 Aurora, Colorado, USA

³Adult and Child Consortium for Health Outcomes and Delivery Science, University of Colorado Anschutz Medical Campus, 80045 Aurora, Colorado, USA

Full list of author information is available at the end of the article



Background

There are nearly 100,000 patients on the U.S. kidney transplant waiting list, with 13 people dying every day while awaiting this life-saving therapy [1]. This national shortage of organs has made it imperative to ensure the longest possible kidney allograft survival for the approximately 20,000 patients that receive a transplant each year, yet over half of kidney grafts fail by 10 years after transplantation [1]. There has been an evolving recognition that antibody-mediated rejection represents one of the prominent barriers to improving long-term graft outcomes [2]. The antibodies that mediate this rejection process, de novo donor-specific antibodies (dnDSA), have been established as an early biomarker for post-transplant adverse kidney events and patients are now screened for dnDSA at regular intervals at most centers around the country [3, 4]. Immunosuppression therapy is likely the most directly modifiable factor in preventing dnDSA. The majority of centers in the U.S. (93%) use Tacrolimus (TAC) as the backbone of their immunosuppression protocols [5], but the drug has a narrow and poorly defined therapeutic range that varies by patient.

The relationship between TAC and dnDSA is still largely unexplored due to numerous complexities. TAC trough levels, or the lowest concentration of drug in the blood prior to the next dose of medication, are measured frequently post-transplant. The measurement error is high because of biological variability in how the drug is absorbed. In addition, TAC is considered an internal covariate in relation to dnDSA since both processes are occurring in the same individual and there may exist a feedback loop in which low TAC levels lead to dnDSA but the suspected presence of dnDSA causes a TAC dose increase. To add to the complexities, screening for dnDSA is conducted only periodically, e.g. every six months to one year, which makes dnDSA an interval censored outcome. It is known that dnDSA have developed during the interval, but the exact timing of dnDSA is unknown. In previous studies, one-dimensional functions of observed TAC values were found to be associated with development of dnDSA; these functions included the coefficient of variation of TAC, the percentage of TAC levels below 5 ng/ml [6], the mean TAC troughs of less than 8 ng/ml and the percentage of time TAC was in therapeutic range [7]. These studies did not address the longitudinal aspect of TAC, the measurement error or possible feedback from dnDSA to TAC, or the interval-censored nature of dnDSA.

We compare two methods to model a heavily interval censored (dnDSA) and a highly variable longitudinal outcome (TAC): (i) a joint model (JM) and (ii) a survival model with a time varying covariate (TVC). It has been shown that when exact event times are known, the presence of measurement error causes the association

parameter estimated by a TVC model to be biased towards null [8, 9]. We sought to further this exploration by studying how the amount of measurement error and the width of the censoring intervals around the time-to-event outcome affect the association estimate in JM and TVC. Interval censored time to event outcomes occur frequently in clinical situations, in particular for conditions that are subclinical and detectable only through a specialized assessment. However, to our knowledge, a thorough comparison of TVC versus JM in the presence of differing amounts of measurement error and interval censoring has yet to be performed.

First, we fit a joint model to analyze the association between longitudinal TAC levels and interval censored dnDSA. The literature of joint models of longitudinal and survival outcomes is extensive including textbooks [10] and [11] and comprehensive review articles [12–15]. Although methods have been developed for single survival outcomes [16–19], much less attention has been given to interval-censored outcomes in joint models [20–22]. Gueorguieva et al. developed a joint model for longitudinal and interval censored competing risk dropout outcomes using a family of parametric baseline hazards including Weibull and log-logistic and used maximum likelihood estimation [22]. We build our joint model for TAC and dnDSA based on this work, but with a single time to event interval censored outcome, a correlated error structure within the longitudinal sub-model, and estimation using Markov Chain Monte Carlo (MCMC).

Second, we fit an interval censored survival model with a time-varying covariate (TVC), following the work by Sparling et al. [23]. We compare the estimates of association between TAC and dnDSA, with the hypothesis that the association parameter in the TVC model will be underestimated. Finally, we perform a series of simulation studies to further explore how the two models perform in terms of bias, level and power based on (1) varying strengths of association, (2) varying amounts of measurement error in TAC, and (3) varying degrees of interval censoring in dnDSA.

In the first section, we introduce the kidney transplant study data. Next, the models are formulated and applied to the motivating kidney transplantation dataset. A simulation study comparing the models under different levels of interval censoring and different amounts of measurement error in the longitudinal process is presented and a discussion on the findings and comparison of the models follows.

Kidney Transplant Study

This retrospective study included patients who were at least 18 years old at time of kidney transplant between September 2007 and December 2013 at the University of Colorado Hospital. Patients were excluded if they

had primary non-function, simultaneous liver and kidney transplant, islet cell transplant, had pretransplant DSA, failed to undergo dnDSA screening, or had dnDSA within the first week post-transplant. Informed consent was provided by all patients, and this study was conducted in accordance with the Declaration of Helsinki and was approved by the ethics committee at the University of Colorado (COMIRB 13-3137). No organs or tissues were obtained from prisoners and all kidney transplant recipients received organs in accordance with the United Network of Organ Sharing, the Organ Procurement and Transplantation Network and a local organ procurement organization, the Donor Alliance, as directed by the United States National Organ Transplant Act.

Five hundred and thirty eight patients met the final inclusion criteria for this analysis. Baseline characteristics of interest are summarized in Table 1. It is hypothesized that age, race, and the degree of Human Leucocyte Antigen (HLA) mismatch between the donor and recipient are associated with development of dnDSA [7]. Post-transplant, each patient was closely monitored and data on TAC trough levels and formation of dnDSA were collected for up to 7 years. Each individual had up to 90 measures of TAC that ranged in value from 0 to 30. No obvious violation of normality assumption in the distribution of TAC was observed (histogram in Additional file 1:

Table 1 Kidney Transplant Study Descriptive Statistics

Summary Measure	No dnDSA (n=357)	dnDSA (n=181)	Overall (n=538)
Male	225 (63%)	111 (61%)	336 (62%)
Age			
Young (<30 years)	24 (7%)	23 (13%)	47 (9%)
Middle (30-49 years)	120 (34%)	80 (44%)	200 (37%)
Old (≥ 50 years)	213 (60%)	78 (43%)	291 (54%)
Ethnicity			
Caucasian	265 (74%)	107 (59%)	372 (69%)
African American	25 (7%)	26 (14%)	51 (9%)
Hispanic	53 (15%)	42 (23%)	95 (18%)
Other	14 (4%)	6 (3%)	20 (4%)
Number of HLA Mismatches median (IQR)	4 (2, 5)	4 (3, 5)	4 (3, 5)
Total TAC Measurements median (IQR)	26 (18, 37)	17 (11, 25)	22 (15, 32)
Months of Followup median (IQR)	48 (20, 60)	12 (6, 24)	36 (12, 60)

Figure S1). Due to the high variability and unreliability of TAC levels within the first week, only TAC levels after day 7 were included in the analysis. TAC levels of zero indicated that a patient was non-compliant and did not take their prescribed TAC dose or a patient was purposefully taken off the drug due to some other medical complication, such as an infection. The number and frequency of TAC measures varied by patient, with an overall median of 22 measurements per person (IQR: 15-32) (Table 1). dnDSA screening was performed at 1, 6, 12 months, annually, and when clinically indicated. Of the 538 patients, 181 developed dnDSA during the study period. For the purpose of this project, dnDSA was treated as a time to event variable, with the event being the first time an individual tested positive for dnDSA. Figure 1 shows the TAC trajectories and time to dnDSA or censoring of 4 random individuals who developed dnDSA and 4 random individuals who did not develop dnDSA.

Methods

To analyze the association between TAC and dnDSA, we propose a shared random effects model and compare it with a traditional interval censored survival model treating TAC as a time-varying covariate.

M1: Shared Random Intercept and Slope Model

Suppose there are N individuals each measured at n_i time points, $i = 1, \dots, n_i$. Let y_{ij} be the measurement for individual i at time t_{ij} , $j = 1, \dots, n_i$. We assume that given the vector of individual-specific random intercept and slope $(a_{0i}, a_{1i})'$, individual outcome measures y_{ij} are independent and normally distributed with mean μ_{ij} and residual variance σ_e^2 , where

$$\mu_{ij} = f(t_{ij}) + \beta_1' x_{1i} + a_{0i} + a_{1i} t_{ij}. \quad (1)$$

The time component t_{ij} can be modeled with any function f . Some examples include a linear trajectory, i.e. $f(t_{ij}) = b_0 + b_1 t_{ij}$; or b-spline basis functions, i.e. $f(t_{ij}) = \sum_{k=1}^K \beta_{0k}' \psi_k(t_{ij})$, where $\psi_k(t_{ij})$ is the k^{th} of K basis functions with coefficient β_{0k} and the number of basis functions K depends on the degree of the splines (e.g. $g = 3$) and number of inner knots ($h = 3$; $K = g + h + 1$). The vector x_{1i} represents baseline covariates with regression coefficients β_1 . The parameters a_{0i}, a_{1i} are the random intercepts and slopes that allow individuals to vary in their baseline TAC level and time trend; they are normally distributed with mean zero and a correlated variance/covariance matrix

$$\begin{bmatrix} a_{0i} \\ a_{1i} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \rho \sigma_0 \sigma_1 \\ \rho \sigma_0 \sigma_1 & \sigma_1^2 \end{bmatrix} \right). \quad (2)$$

Suppose the same N individuals are measured at discrete and varying timepoints for a time to event outcome. The time to event outcome is interval censored and only

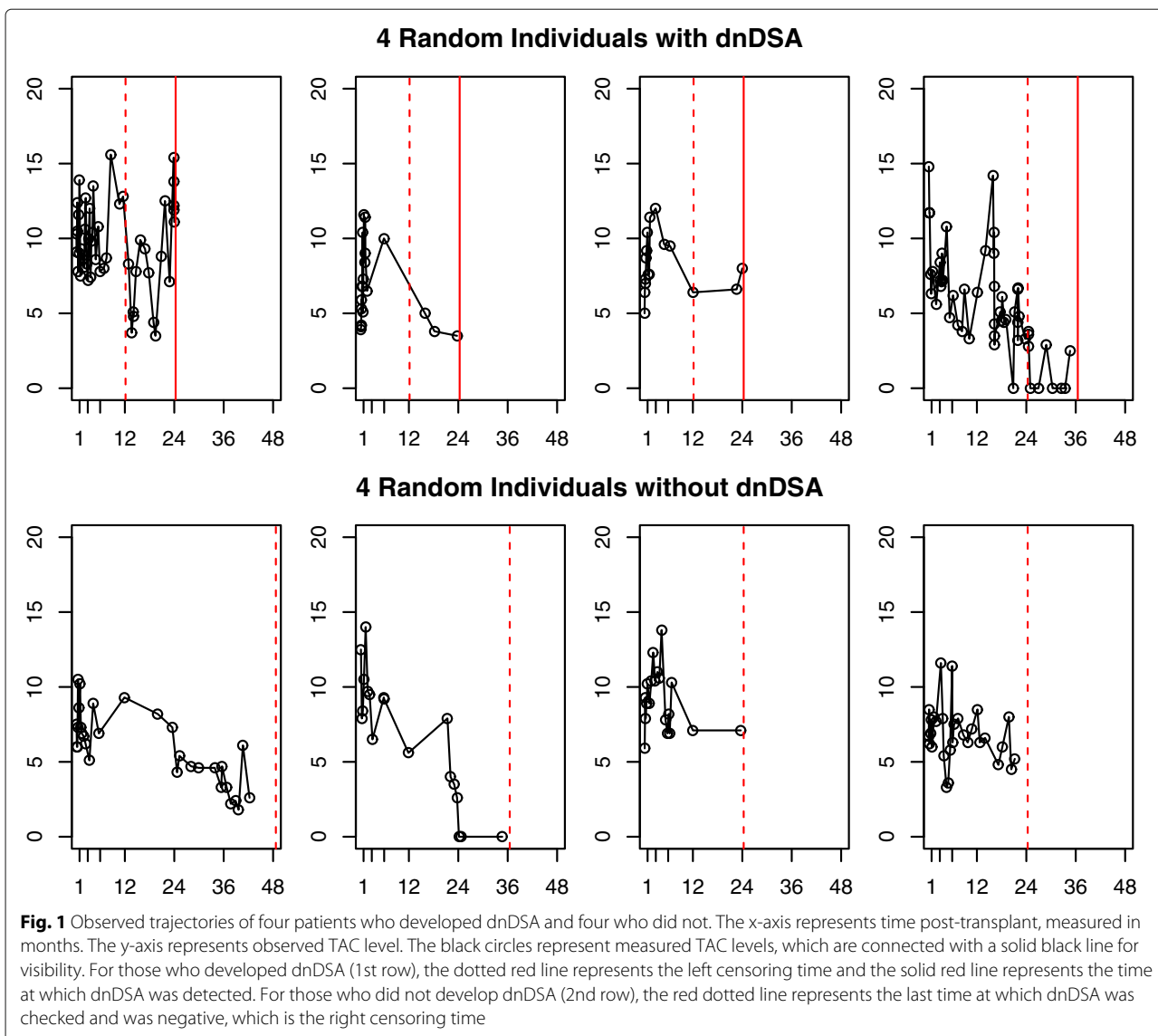


Fig. 1 Observed trajectories of four patients who developed dnDSA and four who did not. The x-axis represents time post-transplant, measured in months. The y-axis represents observed TAC level. The black circles represent measured TAC levels, which are connected with a solid black line for visibility. For those who developed dnDSA (1st row), the dotted red line represents the left censoring time and the solid red line represents the time at which dnDSA was detected. For those who did not develop dnDSA (2nd row), the red dotted line represents the last time at which dnDSA was checked and was negative, which is the right censoring time

known to happen between two discrete time points. Let t_{Ri} denote the time at which the outcome was detected and t_{Li} indicate the visit time immediately preceding t_{Ri} . The true time to event, t_i , lies somewhere in between t_{Li} and t_{Ri} . The hazard for individual i at any time t is given by

$$h_i(t) = h_0(t) \exp(\beta_0 + \beta_2'x_{2i} + \lambda_0 a_{0i} + \lambda_1 a_{1i}) \quad (3)$$

where $h_0(t)$ is the baseline hazard, β_0 is the fixed intercept and scale parameter, x_{2i} is the vector of fixed covariates and β_2 is its parameter, $(\lambda_0, \lambda_1)'$ are the association parameters, and $(a_{0i}, a_{1i})'$ are the same random terms as above in the longitudinal sub-model (1). The baseline hazard can a number of flexible distributions, including parametric (typically exponential, Weibull, or gamma), piecewise constant or spline forms. The two sub-models (1) and

(3) are connected by the shared random intercepts from the survival portion with both the random intercepts and slopes from the longitudinal portion. The first association parameter, λ_0 , corresponds to the risk of event for a difference in average longitudinal measure of one unit. Parameter λ_1 corresponds to the risk of event for a unit increase in the slope of the longitudinal model.

Conditional on random effects, a_{0i}, a_{1i} , the log likelihood can be written for each individual:

$$LL_i = \sum_{j=1}^{n_i} \left\{ -\frac{1}{2} \ln(2\pi\sigma_e^2) - \frac{1}{2\sigma_e^2} (y_{ij} - f(t_{ij}) - \beta_1'x_{1i} - a_{0i} - a_{1i}t_{ij})^2 \right. \\ \left. - I_{Ri} \log\{1 - F_i(t_{Ri})\} + (1 - I_{Ri}) \log\{F_i(t_{Ri}) - F_i(t_{Li})\} \right\} \quad (4)$$

where I_{Ri} is an indicator variable that is 1 when the individual is right censored and 0 when interval censored, t_{Ri} is the right censoring time, and t_{Li} is the left censoring time. $F_i(t)$ corresponds to the cumulative density function at time t . Details on how this likelihood is constructed for the Weibull hazard that was used in the results can be found in the Additional file 1.

M2 and M3: Special Cases of M1

We also consider special cases of M1 where, instead of sharing both the random intercepts and slopes from the longitudinal model (1) with the random intercepts in the survival model (3), only the random intercepts from the longitudinal model are shared with the random intercepts from the survival model. This model is called M2, and the hazard function of the survival model is the special case when $\lambda_1 = 0$ in Equation (3). Model M3 is the special case when $\lambda_0 = 0$ in Equation (3). This model shares the random slope from the longitudinal model with the random intercept from the survival model.

M4: Time Varying Covariate Survival Model

We also consider an interval censored survival model with a time-varying covariate for this data. Assume that each individual, $i = 1, \dots, N$, has longitudinal measurements at update times t_{ij} , where $j = 1, \dots, n_i$. Let z_i denote a vector of baseline covariates for individual i and w_{ij} denote the value of the time-varying covariate for individual i at update time j . The hazard of event for individual i at time j can be written as:

$$h_{ij}(t) = h_{0tvc}(t_{ij}) \exp(\gamma_0 + \gamma' z_i + \eta w_{ij}) \tag{5}$$

where h_{0tvc} is the baseline hazard function (which can be specified as any flexible distribution, as in M1), γ_0 is the fixed intercept, γ' is the coefficient vector for baseline covariates, and η is the coefficient for the time-varying covariate, w_{ij} . In this case, w_{ij} is the same covariate as the longitudinal outcome in M1 (y_{ij}), since we are comparing the performance of M1 and M4 in modeling the same outcome, TAC from our motivating kidney dataset.

The likelihood for this model can be written as in Sparling et al. [23]:

$$LL_i = \prod_{i=1}^N -I_{Ri} [1 - F_i(t_i | \boldsymbol{\gamma}_i, \mathbf{w}_{i[(t_{Ri})]})] \times (1 - I_{Ri}) [F_i(t_{Ri} | \boldsymbol{\gamma}_i, \mathbf{w}_{i[(t_{Ri})]}) - F_i(t_{Li} | \boldsymbol{\gamma}_i, \mathbf{w}_{i[(t_{Li})]})] \tag{6}$$

where $F_i(t_i | \boldsymbol{\gamma}_i, \mathbf{w}_{i[(t_i)]})$ is the cumulative density, conditional on the fixed covariate values $\boldsymbol{\gamma}_i$ and the relevant time-dependent covariate values w_{ij} up to the right or interval censoring times. The cumulative density can be iteratively calculated based based on equation (5) for each subject, by using the cumulative hazard and survival functions.

Bayesian Model Fit

Model estimation and inference is based on the Bayesian framework where the hyper-parameters are assigned weakly informative priors. Priors for the parameters are as follows:

$$\begin{aligned} b_0, b_1, \beta_0, \beta'_1, \beta'_2, \lambda_0, \lambda_1, \gamma_0, \gamma', \eta &\sim N(0, 10000) \\ \sigma_e &\sim \text{Uniform}(0, 100) \\ \alpha, \alpha_{tvc} &\overset{ind}{\sim} \text{Gamma}(100, 100) \\ \begin{bmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{bmatrix}^{-1} &\sim \text{Wishart}\left(\begin{bmatrix} 0.00001 & 0 \\ 0 & 0.000001 \end{bmatrix}, 2\right) \end{aligned}$$

where $u \sim \text{Normal}(\mu, \sigma^2)$ with probability density $1/(\sigma\sqrt{2\pi}) \exp\{- (u - \mu)^2 / 2\sigma^2\}$; $u \sim \text{Uniform}(a, b)$ with probability density $1/(b - a)$ for $a < u < b$; $u \sim \text{Gamma}(r, \lambda)$ with probability density $\lambda^r u^{r-1} \exp(-\lambda u) / \Gamma(r)$; and $\Omega \sim \text{Wishart}(R, k)$ with probability density $(|\Omega|^{(k-p-1)/2} |R|^{k/2} \exp\{-Tr(R\Omega/2)\}) / (2^{pk/2} \Gamma_p(k/2))$ for $k \geq p$, where $p = 2$, Tr is the trace function and Γ_p is the multivariate gamma function.

Markov Chain Monte Carlo (MCMC) simulations were performed to estimate the posterior distribution. A Gibbs sampler was used to construct two Markov chains using JAGS software [24] and the runjags package in R [25]. Convergence of the samples was assessed by trace plot inspection, and Gelman and Geweke tests which test for equality of the means of the first and last part of a Markov chain [26]. After a burn-in period of 20,000 and thinning of 40 of 80,000 sampling iterations, 2000 samples per chain were used for inference.

Model Selection and Comparison

Model selection was carried out using (a) the Deviance Information Criterion (DIC) [27] and (b) the Watanabe-Akaike information criterion (WAIC) introduced by Watanabe [28]. The DIC is a measure calculated based on the deviance, $D(\boldsymbol{\theta}) = -2 \log L(\mathbf{y} | \boldsymbol{\theta})$, that penalizes for the number of effective degrees of freedom in the model and is defined as $DIC = \bar{D}(\boldsymbol{\theta}) + \rho_D = 2\bar{D}(\boldsymbol{\theta}) - D(\hat{\boldsymbol{\theta}})$, where \bar{D} and $\hat{\boldsymbol{\theta}}$ are obtained from MCMC output as the mean of the deviance and the posterior mean, respectively. WAIC approximates leave-one-out cross-validation [28] and works well for hierarchical models where the number of degrees of freedom is not obvious [29]. The best model should have the lowest values of DIC and WAIC; differences of at least 5 points will be considered important. We followed the traditional recommendations by Spiegelhalter and Ntzoufras [26, 27] to use the conditional likelihood rather than the marginal to compare model fits. More discussion on this topic can be found in the work by Kapur [30] and in the supplement (Part G) of Juarez-Colunga [31].

Results

Models and Model Selection

Models M1-M4 were fitted to the kidney transplant data. The three joint models (M1, M2, M3) had a longitudinal sub-model with TAC as the outcome, a fixed intercept and slope for time ($f(t_{ij}) = b_0 + b_1 t_{ij}$) and a random intercept and/or slope for time (depending on the model) for variations across individuals. Spline terms for time were tested, but DIC and WAIC indicated that a linear trend was sufficient. For the survival sub-models, various baseline hazard functions were tested (exponential, Weibull, gamma, piecewise constant) and the Weibull yielded the best model fit ($h_0(t) = \alpha t^{\alpha-1}$, where α is the Weibull shape parameter). The Weibull distribution makes intuitive sense for this data, as the hazard can flexibly be increasing or decreasing over time and the risk of dnDSA is thought to decrease as time from transplant progresses. The parameters from the Weibull model can easily be interpreted as the change in log-relative risk. All survival sub-models had time to dnDSA as the interval censored outcome, random intercept(s) associated with the longitudinal sub-model, and a vector of baseline covariates: age, ethnicity, and number of HLA mismatches. See Table 1 for the description of these covariates. Baseline covariates were tested in both the longitudinal and survival sub-models, but none were significant predictors in the longitudinal sub-model, so they were removed.

The interval censored survival model with a time-varying covariate (M4) had dnDSA as the outcome, TAC as the time-varying covariate, and the same baseline covariates as listed above. After considering other baseline hazard functions for M4, we chose the Weibull due to model fit and to compare to M1, so $h_{0TVC}(t) = \alpha_{TVC} t^{\alpha_{TVC}-1}$ where α_{TVC} is the Weibull shape parameter.

Table 2 displays the results of model comparisons based on the DIC and the WAIC. The survival sub-model fits in M1-M3 can be directly compared with the model fit for M4. The best fitting survival model according to both DIC and WAIC is Model 1, which shares both the random

intercept and slope from the longitudinal model with the random intercept from the survival model.

Estimation of Associations

Table 3 presents the results from the final model, M1, and the time-varying covariate model, M4. The longitudinal portion of M1 shows that on average, individuals start at a TAC level of about 7 ng/ml. For every one month post-transplant, the average individual's TAC levels decrease by 0.05 ng/ml. There is a significant negative correlation between the random intercept and slope, $\rho = -0.38$, indicating that individuals with higher starting TAC have a lower slope. The survival portion of the model has a decreasing hazard over time, parameterized by $\alpha = 0.57$. The degree of HLA mismatching is significantly associated with dnDSA; for every one unit increase in mismatching, the risk of dnDSA increases 1.26-fold (95% CrI: 1.13, 1.40). Race is also associated with risk of dnDSA; compared to Caucasian individuals, African Americans and Hispanics have a higher risk of dnDSA (2.09 and 1.60, respectively). Age is also a factor, as middle-aged (30-49 years) and older-aged (50+ years) individuals both have a reduced risk of dnDSA compared to individuals under the age of 30 (Hazard Ratios: 0.57 and 0.30, respectively). These covariate effects are all conditional on the same initial level and slope of TAC, so these results are not explained by different TAC levels across HLA mismatches, race or age groups. The first association parameter, λ_0 , explains the risk of dnDSA associated with a difference in overall average level of TAC. For every one ng/ml higher average TAC level, the risk of dnDSA lowers 0.64-fold (95% CrI: 0.52, 0.75). The second association parameter, λ_1 , explains the risk of dnDSA associated with a difference in the slope of TAC levels. For every 0.05 ng/ml/month higher TAC slope, the risk of dnDSA lowers 0.43-fold (95% CrI: 0.32, 0.58).

The interval censored survival model M4 does not have a linear sub-model, but instead uses a time-varying covariate for TAC. The scale parameter of the survival model is $\alpha = 0.46$, which also results in a decreasing hazard of dnDSA over time. The effects of baseline covariates, γ' , are similar in magnitude to β' in M1. The hazard ratio for the time-varying TAC is $\eta = 0.80$. Within an individual, following a one unit increase in TAC level, the risk of dnDSA decreases 0.80-fold (95% CrI: 0.75, 0.85). Though this is still a significant finding, the effect size of η in M4 is smaller than λ_0 in M1. We explore this further in the "Simulation Study" section.

Goodness of Fit of Kidney Transplant Study

To assess the adequacy of model fit for M1-M3, we simulated 1000 datasets from the posterior predictive distribution and compared them to the observed data. We performed these posterior predictive assessments

Table 2 Model Selection Criteria: DIC, and WAIC as defined in the "Model Selection and Comparison" section

Model	DIC	WAIC
M1: Shared Random Intercept/Slope with Intercept	68943 (67672+1271)	68917 (67608+1309)
M2: Shared Random Intercepts	69052 (67716+1336)	69029 (67675+1354)
M3: Shared Random Intercept with Slope	69104 (67739+1365)	69057 (67691+1366)
M4: TVC Model	1327	1329

The parentheses contain the (linear sub-model fit + survival sub-model fit). Lowest value indicating best model fit is in bold

Table 3 Results from M1 and M4. M1 is the joint model that shares both the random intercept and slope from the longitudinal sub-model with the random intercept from the survival model. M4 is the interval censored survival model with the longitudinal measurements as a time-varying covariate

Parameter	M1: Shared Random Effects				M4: Time Varying Covariate				
	Estimate	SD	Lower 95% CrI	Upper 95% CrI	Estimate	SD	Lower 95% CrI	Upper 95% CrI	
Linear Sub-Model									
b_0	7.24	0.06	7.12	7.36	-	-	-	-	-
b_1	-0.05	0.005	-0.06	-0.04	-	-	-	-	-
σ_ϵ^2	7.44	0.09	7.26	7.63	-	-	-	-	-
σ_0^2	1.66	0.14	1.40	1.94	-	-	-	-	-
σ_1^2	0.004	0.0006	0.003	0.005	-	-	-	-	-
ρ	-0.38	0.07	-0.51	-0.24	-	-	-	-	-
Survival Sub-Model									
α	0.57	0.06	0.47	0.69	α_{TVC}	0.46	0.03	0.40	0.53
	Hazard					Hazard			
	Ratio					Ratio			
β_0	0.03	0.01	0.01	0.06	γ_0	0.17	0.06	0.08	0.30
β_1 (HLA Mismatch Number)	1.26	0.07	1.13	1.40	γ_1	1.29	0.06	1.17	1.43
β_2 (African American)	2.09	0.56	1.23	3.39	γ_2	1.79	0.40	1.11	2.70
β_3 (Hispanic)	1.60	0.33	1.05	2.31	γ_3	1.67	0.31	1.14	2.33
β_4 (Other Race)	1.14	0.56	0.34	2.41	γ_4	1.12	0.47	0.40	2.20
β_5 (30-49 years)	0.57	0.15	0.32	0.93	γ_5	0.67	0.17	0.42	1.05
β_6 (≥ 50 years)	0.30	0.08	0.17	0.49	γ_6	0.41	0.10	0.25	0.65
λ_0	0.64	0.06	0.52	0.75	η	0.80	0.03	0.75	0.85
λ_1 (per 0.05 change)	0.43	0.07	0.32	0.58	-	-	-	-	-

For M1, the posterior mean and 95% credible intervals are presented for the linear portion of the model, the survival portion of the model, and the association parameters. The longitudinal portion is comprised of a fixed intercept (b_0), a fixed slope (b_1), and a random error term ($\sim N(0, \sigma_\epsilon^2)$). There is also a random intercept ($a_{0i} \sim N(0, \sigma_0^2)$) and a random slope for each individual ($a_{1i} \sim N(0, \sigma_1^2)$), which have a correlation parameter ρ . The survival model is comprised of fixed covariates (β), the Weibull association parameter α , and a random intercept for each individual that is related to the random intercept and slope of the longitudinal sub-model through two association parameters. The first association parameter (λ_0) links the two sub-models through their shared random intercepts. The second association parameter (λ_1) links the two sub-models through the longitudinal random slope and the survival random intercept. M4 contains the Weibull scale parameter, α_{TVC} , baseline covariates with parameter γ , and the coefficient for the time-varying covariate, η .

separately for the longitudinal and survival sub-models. This is usually done through a discrepancy function represented by a scalar or a vector, often represented graphically [32, 33]. For comparing the observed longitudinal TAC data with the posterior distribution of longitudinal measures on given individuals, we have conditioned on the posterior mean of the random effects at the individual level; e.g. for M1: $\hat{a}_{0i}, \hat{a}_{1i}$ in (1) are considered fixed [34]. We compared 1000 posterior predictive TAC trajectories against the observed trajectory for 3 random individuals using M1, M2 and M3. Due to the large variability in TAC levels within and between individuals, the plots are all nearly identical, which is indicative that no one model fits the longitudinal trajectory better than another model. The plots are shown in Additional file 1: Figure S2.

For the survival sub-model predictive check, we created artificial intervals (that resembled the kidney study data intervals) around the survival times taken from the

1000 posterior predictive datasets. Next, using these intervals, we plotted the predicted Weibull cumulative density functions from the 1000 posterior datasets and then overlaid the estimated cumulative density curve from the raw data. We chose to plot the cumulative density function instead of the survival function because of the low event rate in these sub-groups [35]. We had to choose certain values of baseline covariates for these checks, so we presented three different types of individuals that represent the most prevalent combinations of covariates. The results of the survival predictive checks for three particular combinations of baseline covariates, for all three models M1-M3, are found in the Additional file 1: Figure S3. The coverage of the observed survival curve by the posterior predictive survival curves is best for M1. This is indicative that M1 fits the survival data better than the other two joint models.

Sensitivity Analysis

To determine if the choice of priors affected the estimates of M1 parameters, an alternative prior distribution of Uniform ($U(-100, 100)$) was compared against a Normal ($N(0, 1000)$) prior for b_0 , b_1 , β , λ_0 and λ_1 , and the half-Cauchy ($HC(0, 25)$) distribution for σ_e was compared against the uniform prior ($U(0, 100)$) for this variance component. The results, tabulated in Additional file 1: Table S1, indicate the choice of priors has very little effect on the parameter estimates.

Simulation Study

We conducted a simulation study to compare how the joint M1 and the time varying covariate model perform under (1) varying degrees of association, (2) varying amounts of interval censoring (IC), and (3) varying amounts of measurement error (ME). We simulated data using the joint model with shared random intercepts (M2) under nine scenarios, each with a different combination of measurement error (none [$\sigma_e^2 = 0$], low [$\sigma_e^2 = 1$], and high [$\sigma_e^2 = 8$]) and degree of interval censoring (lighter [0-1 month, 1-6 month, 6-12 month, 12-24 month, yearly after], visits missed at random, and heavier [3 year intervals]). For the random interval censoring, we assumed patients had a fixed follow-up schedule matching the lighter IC scenario, but missed 50% of visits at random (we also set the missing-ness to 25% and the results were almost identical). All other variables were set to be similar to the kidney transplant data, with Weibull parameters $\alpha = 0.50$ and $\beta_0 = -2$. Two hundred datasets were simulated under each of these six conditions, and the JM and TVC models were fitted to each dataset using MCMC. Each dataset contained 300 individuals, and approximately half developed dnDSA. The results from all simulations were obtained using two parallel chains; the TVC model had a burn-in of 1000 samples and 1000 samples for inference, while the JM had a burn-in of 2000 samples and 2000 samples for inference. The average model estimates from the light and heavy IC scenarios are reported in Table 4. The light and random IC scenarios did not differ significantly, so the results from the random IC scenarios are in the Additional file 1: Table S2.

To compare power for detecting an association between the longitudinal and the time to event outcomes, we varied the association between TAC and dnDSA (λ_0 in the JM and η in the TVC) between -0.50 and 0 within each scenario. We calculated power based on each model at each true association value using the credible intervals of λ_0 or η . For example, for simulation run m , $m = 1, \dots, M$ ($M=200$ for our simulations), let $\hat{\lambda}_{0L}$ and $\hat{\lambda}_{0U}$ denote the 2.5% and 97.5% percentiles of the distribution of the MCMC samples of $\hat{\lambda}_0$. The power is calculated as: Power

$= 1 - 1/M \sum_{m=1}^M I(\hat{\lambda}_{0L} < 0 < \hat{\lambda}_{0U})$, where $I(E)$ is the indicator function of event E . The power curves for all nine scenarios are presented in Fig. 2.

As expected, the association parameter was consistently attenuated toward zero in the TVC compared to the JM. As more measurement error was introduced, the TVC underestimated the association even more (true value: $\lambda_0 = -0.50$, low ME: $\hat{\eta} = -0.32$ and high ME: $\hat{\eta} = -0.18$), while there was no evidence of substantial bias for the JM (low ME: $\hat{\lambda}_0 = -0.53$ and high ME: $\hat{\lambda}_0 = -0.54$). As heavier interval censoring was introduced, the association estimates were further from the true value in both modeling approaches, but the standard deviations also increased. The parameter that was most affected by heavier interval censoring was the Weibull shape parameter, α . Although the bias of α was lower in the TVC context, this model should not be preferred when there is high measurement error and heavy censoring, due to the underestimation of the association between the two outcomes, which is typically of more interest.

The power of the TVC model was consistently lower than the power of the JM (Fig. 2). When more measurement error was introduced, the power of the TVC model significantly decreased while the power of the JM stayed approximately the same. When visits were missed at random, the decrease in power on both approaches was minimal. The wider interval censoring around dnDSA resulted in a high type I error in the case of the TVC model (type I error=0.23 in the high ME and heavy IC simulation from Fig. 2). The average credible intervals and coverage probabilities for all simulations are reported in the Additional file 1: Figure S3.

Discussion

In this paper, we compare two modeling techniques for two outcomes: a longitudinal outcome with high measurement error and a survival outcome with heavy interval censoring. For our motivating dataset, both DIC and WAIC suggested that a shared random effects model with random effects for both the intercept and slope of the longitudinal process (TAC) was the preferred model. Results from this model and from the TVC model (M4) support the hypothesis that the association between TAC and dnDSA is underestimated when a TVC model is fit. The simulation study further confirmed this hypothesis and showed that the TVC has lower power and higher type I error compared to the JM when the outcomes have high measurement error and heavy interval censoring.

TAC is the most commonly used immunosuppressant drug in kidney transplantation and yet optimal therapeutic dose to prevent dnDSA, a common reason for graft loss, is unknown. Most published studies on the influence of immunosuppressant drugs on development of dnDSA

Table 4 Results from the simulation study

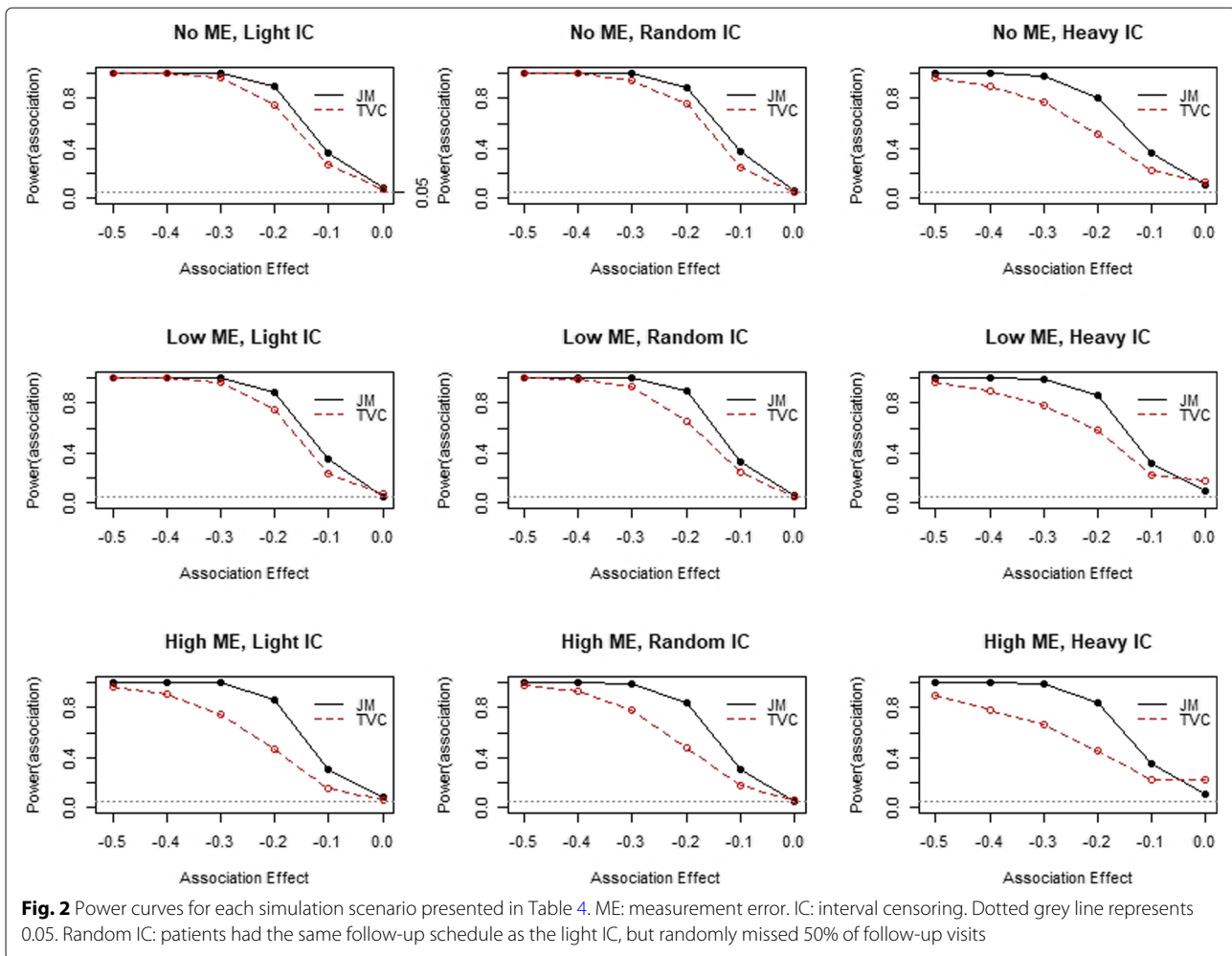
	Variable	No Measurement Error			Low Measurement Error			High Measurement Error		
		True Value	M2	M4	True Value	M2	M4	True Value	M2	M4
Lighter interval censoring	b_1	-0.03	-0.02 (0.00)		-0.03	-0.03 (0.005)		-0.03	-0.03 (0.005)	
	b_0	7.00	6.92 (0.001)		7.00	7.02 (0.05)		7.00	6.99 (0.07)	
	σ_e^2	0	0.00 (0.00)		1.00	1.00 (0.01)		8.00	8.01 (0.09)	
	ρ	-0.005	-0.01 (0.001)		-0.005	-0.005 (0.06)		-0.005	-0.001 (0.08)	
	σ_0^2	1.75	1.76 (0.003)		1.75	1.76 (0.15)		1.75	1.75 (0.16)	
	σ_1^2	0.004	0.004 (0.00)		0.004	0.004 (0.000)		0.004	0.004 (0.001)	
	β_0, γ_0	-2.00	-2.10 (0.006)	0.63 (0.009)	-2.00	-2.14 (0.26)	0.21 (0.39)	-2.00	-2.14 (0.27)	-0.82 (0.31)
	β_1 (HLA)	0.25	0.26 (0.002)	0.24 (0.002)	0.25	0.26 (0.07)	0.24 (0.07)	0.25	0.26 (0.07)	0.23 (0.07)
	α	0.50	0.54 (0.00)	0.41 (0.00)	0.50	0.53 (0.04)	0.47 (0.04)	0.50	0.54 (0.04)	0.47 (0.04)
	λ_0, η	-0.50	-0.52 (0.002)	-0.33 (0.001)	-0.50	-0.53 (0.07)	-0.32 (0.05)	-0.50	-0.54 (0.08)	-0.18 (0.04)
Heavier interval censoring	b_1	-0.03	-0.02 (0.00)		-0.03	-0.03 (0.004)		-0.03	-0.03 (0.005)	
	b_0	7.00	6.93 (0.00)		7.00	7.00 (0.05)		7.00	7.00 (0.08)	
	σ_e^2	0	0.00 (0.00)		1.00	1.00 (0.01)		8.00	8.01 (0.09)	
	ρ	-0.005	-0.01 (0.001)		-0.005	-0.005 (0.06)		-0.005	-0.001 (0.08)	
	σ_0^2	1.75	1.76 (0.003)		1.75	1.76 (0.15)		1.75	1.75 (0.16)	
	σ_1^2	0.004	0.004 (0.00)		0.004	0.004 (0.000)		0.004	0.004 (0.001)	
	β_0, γ_0	-2.00	-3.99 (0.01)	-0.80 (0.01)	-2.00	-4.00 (0.49)	-1.61 (0.56)	-2.00	-4.20 (0.55)	-2.24 (0.49)
	β_1 (HLA)	0.25	0.28 (0.002)	0.24 (0.002)	0.25	0.27 (0.09)	0.23 (0.08)	0.25	0.28 (0.09)	0.23 (0.08)
	α	0.50	1.08 (0.002)	0.69 (0.002)	0.50	1.09 (0.11)	0.87 (0.10)	0.50	1.15 (0.13)	0.89 (0.09)
	λ_0, η	-0.50	-0.56 (0.002)	-0.26 (0.001)	-0.50	-0.59 (0.10)	-0.25 (0.06)	-0.50	-0.64 (0.12)	-0.18 (0.05)

Data were simulated from the joint model with shared random intercepts, M2, and used to fit M2 and M4 by MCMC. Six simulations were performed with varying amounts of measurement error (none ($\sigma_e^2 = 0$), low ($\sigma_e^2 = 1$), and high ($\sigma_e^2 = 8$)) and varying amounts of interval censoring (lighter (as in our data) and heavier (3 year intervals)). Unlike in Table 3, none of these estimates are converted into hazard ratios, because the interest here is comparing the results from the simulation to the true values. The numbers presented are mean (standard deviation) of the estimates from the 200 datasets for each simulation condition

compare drugs rather than evaluate dosing of a specific drug [36, 37]. A study in the liver transplant setting by Kaneku et al. analyzed risk factors for dnDSA in 749 adult liver transplant recipients [38]. They found 8.1% of individuals developed dnDSA at one year and at least one low TAC (<3 ng/ml) or cyclosporine (<75 ng/ml) level were associated with dnDSA (OR 2.66, $p = 0.015$). In our dataset, 21.7% of individuals had dnDSA by 1 year and we also found that for every one ng/ml lower average TAC the risk of dnDSA was higher (HR [95% CrI]: 1.56 [1.33, 1.92]) and the hazard of dnDSA also increased 2.33-fold (95% CrI: 1.72, 3.13) for every 0.05 ng/ml/month decrease in the slope of TAC. The higher levels of dnDSA detected in our study may be due to several reasons, including center differences in detection levels of antibody counted as positive and inclusion criteria. Advancing previous models, we modeled the time to dnDSA while accounting for changes in TAC, where TAC is modeled as a variable with measurement error. Although minimization of TAC exposure with lower drug trough levels may decrease its associated toxicities, our results suggest that this strategy may place

individuals at increased risk for dnDSA, which may have important implications for intermediate and long-term graft survival.

Similar to the findings presented by Kolagmunnage and Prentice [8, 9], our simulation found severe underestimation of the association between the longitudinal and survival outcomes when using the time-varying covariate model for longitudinal data with a considerable amount of measurement error. We also found a similarly severe underestimation of the association using the TVC model when the survival outcome was interval censored, even in the case of no measurement error. Importantly, the TVC model had a lower power to detect an association compared to the JM, and it had a higher type I error rate when heavy interval censoring was introduced. Interestingly, when half of the follow-up visits were missed at random for the scenarios studied, there was a negligible overall effect on power for both the TVC and JM. These results are important in understanding the behavior of the JM and TVC since in many clinical settings time to event outcomes can often be observed only within intervals. In



future work, we would like to study the effect of informative missing-ness (e.g. probability of missing depends on a covariate, such as age, or the probability of missing increases as time from transplant progresses) on the power of detecting associations within these models.

We recognize the inherent differences in the two modeling approaches discussed here. The joint models (M1-M3) all result in between-subject association interpretations. For example, λ_0 is interpreted as the increased hazard of dnDSA for an individual with an average TAC level of one ng/ml lower than another individual. In contrast, the time-varying covariate model (M4) yields a within-subject interpretation. In this model, η relates to the increased hazard of dnDSA for a given individual, immediately following a one ng/ml decrease in TAC.

By employing MCMC to fit all models, we were able to easily obtain the posterior distribution of all estimated parameters. Since the main focus of this analysis was to assess the association between TAC and dnDSA, we were particularly interested in the association parameters: λ_0 and λ_1 in M1 and η in M4. We calculated credible intervals

(CrI) for these parameters, which give the straightforward interpretation as an interval that contains the association parameter with 95% certainty. The MCMC framework also allows for easy calculations of dynamic predictions and this will be explored in future work. The code to fit the final joint model (M1) and time-varying covariate model (M4) on simulated data in JAGS is located in the Additional file 2. The code for fitting M1 in SAS using PROC NLMIXED is also provided.

Conclusion

The joint model examined in this paper allows for flexible modeling of dnDSA development as it relates to TAC levels over time. As our simulation study showed, this approach accommodates modeling heavily interval censored survival data jointly with highly variable longitudinal data, and if a time varying covariate (TVC) approach is used with heavy interval censoring and high measurement error, the associations may be severely underestimated, missed completely, or detected spuriously.

Additional files

Additional file 1: Supplemental materials. Contains details on likelihood construction, additional figures and tables, and selected output from the reproducible examples. (PDF 393 kb)

Additional file 2: Reproducible Example Files:

- example-1-JM.R: Code to fit M1
- longitudinal-data.csv: simulated TAC data for M1
- survival-data.csv: simulated dnDSA data for M1
- model-1-JM.txt: JAGS model, called by example-1-JM.R
- example-1-NLMIXED.SAS: Code to fit M1 with PROC NLMIXED, uses same simulated data
- example-4-TVC.R: Code to fit M4
- longitudinal data tvc.csv: simulated TAC data for M4 (carried forward values of TAC)
- model-4-TVC.txt: JAGS model, called by example-4-TVC.R (ZIP 199 kb)

Abbreviations

CrI: Credible interval; DIC: Deviance information criterion; dnDSA: de novo Donor specific antibody; HLA: Human leucocyte antigen; IC: Interval censoring; JAGS: Just another Gibbs sampler; JM: Joint model; MCMC: Markov chain Monte Carlo; ME: Measurement error; TAC: Tacrolimus; TVC: Time-varying covariate; WAIC: Watanabe-Akaike information criterion

Acknowledgments

This paper is based on Kristen Campbell's master's thesis supervised by Dr. Elizabeth Juarez-Colunga, Dr. Jane Gralla and Dr. Gary Grunwald.

Authors' contributions

KC, EJC, and GKG performed all statistical analysis mentioned in the paper. SD and JC provided clinical context and guidance. JG served as the liaison between the statistical and clinical co-authors. All authors read and approved the final manuscript.

Funding

Supported by NIH/NCATS Colorado CTSA Grant Number UL1 TR002535. The design of the study and the collection, analysis, and interpretation of data are the authors' sole responsibility and do not necessarily represent official NIH views.

Availability of data and materials

Simulated datasets and the code used to fit the joint model and time varying covariate model are publicly available in the Additional file 2.

Ethics approval and consent to participate

Study data collection was conducted in accordance with the Declaration of Helsinki and was approved by the ethics committee at the University of Colorado (COMIRB 13-3137). All data collected was retrospective in nature and de-identified, so individual consent forms were not needed and administrative data privileges were not needed.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Pediatrics, University of Colorado Anschutz Medical Campus, 80045 Aurora, Colorado, USA. ²Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus, 80045 Aurora, Colorado, USA. ³Adult and Child Consortium for Health Outcomes and Delivery Science, University of Colorado Anschutz Medical Campus, 80045 Aurora, Colorado, USA. ⁴Division of Renal Diseases and Hypertension, University of Colorado Anschutz Medical Campus, 80045 Aurora, Colorado, USA.

Received: 11 February 2019 Accepted: 9 June 2019

Published online: 26 June 2019

References

1. Hart A, Smith J, Skeans M, Gustafson S, Stewart D, Cherikh W, Wainright J, Kucheryavaya A, Woodbury M, Snyder J, et al. Optn/srtr 2015 annual data report: kidney. *Am J Transplant.* 2017;17(S1):21–116.
2. Sellares J, de Freitas D, Mengel M, Reeve J, Einecke G, Sis B. Understanding the causes of kidney transplant failure: The dominant role of antibody-mediated rejection and nonadherence. *Am J Transplant.* 2011;2(12):388–99.
3. Parajuli S, Reville PK, Ellis TM, Djamali A, Mandelbrot DA. Utility of protocol kidney biopsies for de novo donor-specific antibodies. *Am J Transplant.* 2017;17(12):3210–8.
4. Schinstock C, Cosio F, Cheungpasitporn W, Dadhania D, Everly M, Samaniego-Picota M, Cornell L, Stegall M. The value of protocol biopsies to identify patients with de novo donor-specific antibody at high risk for allograft loss. *Am J Transplant.* 2017;17(6):1574–84.
5. Matas A, Smith J, Skeans M, Thompson B, Gustafson S, Schnitzler M, Stewart D, Cherikh W, Wainright J, Snyder J, et al. Optn/srtr 2012 annual data report: kidney. *Am J Transplant.* 2014;14(S1):11–44.
6. Wiebe C, Rush DN, Nevens TE, Birk PE, Blydt-Hansen T, Gibson IW, Goldberg A, Ho J, Karpinski M, Pochinco D, et al. Class II eplet mismatch modulates tacrolimus trough levels required to prevent donor-specific antibody development. *J Am Soc Nephrol: JASN.* 2017;28(11):3353–62.
7. Davis S, Gralla J, Klem P, Tong S, Wedermyer G, Freed B, Wiseman A, Cooper JE. Lower tacrolimus exposure and time in therapeutic range increase the risk of de novo donor-specific antibodies in the first year of kidney transplantation. *Am J Transplant.* 2017;18(4):907–15.
8. Kolamunnage-Dona R, Williamson PR. Time-dependent efficacy of longitudinal biomarker for clinical endpoint. *Stat Methods Med Res.* 2018;27(6):1909–24.
9. Prentice R. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika.* 1982;69(2):331–42.
10. Elashoff R, Li N, et al. *Joint Modeling of Longitudinal and Time-to-event Data.* Boca Raton, FL: CRC Press; 2016.
11. Rizopoulos D. *Joint Models for Longitudinal and Time-to-event Data: With Applications in R.* Boca Raton, FL: CRC Press; 2012.
12. Hogan JW, Laird NM. Model-based approaches to analysing incomplete longitudinal and failure time data. *Stat Med.* 1997;16(3):259–72.
13. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Stat Sin.* 2004;14:809–34.
14. Lawrence Gould A, Boye ME, Crowther MJ, Ibrahim JG, Quartey G, Micallef S, Bois FY. Joint modeling of survival and longitudinal non-survival data: current methods and issues. report of the dia bayesian joint modeling working group. *Stat Med.* 2015;34(14):2181–95.
15. Papageorgiou G, Mauff K, Tomer A, Rizopoulos D. An overview of joint modeling of time-to-event and longitudinal outcomes. *Ann Rev Stat Appl.* 2019;6:223–40.
16. Li J, Ma S. *Survival Analysis in Medicine and Genetics.* Boca Raton, FL: Chapman and Hall/CRC; 2013.
17. Finkelstein DM, Wolfe RA. A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics.* 1985;4:933–45.
18. Goetghebuer E, Ryan L. Semiparametric regression analysis of interval-censored data. *Biometrics.* 2000;56(4):1139–44.
19. Odell PM, Anderson KM, D'Agostino RB. Maximum likelihood estimation for interval-censored data using a weibull-based accelerated failure time model. *Biometrics.* 1992;48:951–9.
20. Rouanet A, Joly P, Dartigues J-F, Proust-Lima C, Jacqmin-Gadda H. Joint latent class model for longitudinal data and interval-censored semi-competing events: Application to dementia. *Biometrics.* 2016;72(4):1123–35.
21. Chen C-M, Shen P-s, Tseng Y-K. Semiparametric transformation joint models for longitudinal covariates and interval-censored failure time. *Comput Stat Data Anal.* 2018;128:116–27.
22. Gueorguieva R, Rosenheck R, Lin H. Joint modelling of longitudinal outcome and interval-censored competing risk dropout in a schizophrenia clinical trial. *J R Stat Soc: Ser A (Stat Soc).* 2012;175(2):417–33.
23. Sparling YvonneH, Younes Naji, Lachin John. Parametric survival models for interval-censored data with time-dependent covariates. *Biostatistics.* 2006;7(4):599–614. <https://doi.org/10.1093/biostatistics/kxj028>.
24. Plummer M. JAGS Version 3.3.0 User Manual. 2012. <http://mcmc-jags.sourceforge.net>. Accessed 1 Aug 2018.

25. Denwood MJ. runjags: An r package providing interface utilities, model templates, parallel computing methods and additional distributions for mcmc models in jags. *J Stat Softw.* 2016;71(9):1–25.
26. Ntzoufras I. *Bayesian Modeling Using WinBUGS* vol. 698. Hoboken, NJ: Wiley; 2011.
27. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc: Ser B (Stat Methodol).* 2002;64(4):583–639.
28. Watanabe S. Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *J Mach Learn Res.* 2010;11:3571–94.
29. Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. *Stat Comput.* 2014;24:997–1016.
30. Kapur K, Li X, Blood EA, Hedeker D. Bayesian mixed-effects location and scale models for multivariate longitudinal outcomes: an application to ecological momentary assessment data. *Stat Med.* 2015;34(4):630–51.
31. Juarez-Colunga E, Silva G, Dean C. Joint modeling of zero-inflated panel count and severity outcomes. *Biometrics.* 2017;73(4):1413–23.
32. Gelman A, Meng X-L, Stern H. Posterior predictive assessment of model fitness via realized discrepancies. *Stat Sin.* 1996;6(4):733–60.
33. Lynch SM, Western B. Bayesian posterior predictive checks for complex models. *Sociol Methods Res.* 2004;32(3):301–35. <http://smr.sagepub.com/content/32/3/301.full.pdf+html>.
34. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*, 3rd edn. Boca Raton, FL: CRC press; 2013.
35. Pocock SJ, Clayton TC, Altman DG. Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. *Lancet.* 2002;359(9318):1686–9.
36. O’Leary JG SM. The influence of immunosuppressive agents on the risk of de novo donor-specific hla antibody production in solid organ transplant recipients. *Transplantation.* 2016;1(100):39–53.
37. Thaanat O, Koenig A, Leibler C, Grimbert P. Effect of immunosuppressive drugs on humoral allosensitization after kidney transplant. *J Am Soc Nephrol.* 2016;27(7):1890–1990.
38. Kaneku H, O’leary J, Banuelos N, Jennings L, Susskind B, Klintmalm G, Terasaki P. De novo donor-specific hla antibodies decrease patient and graft survival in liver transplant recipients. *Am J Transplant.* 2013;13(6):1541–8.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

