Supplementary materials for

"Joint longitudinal and time-to-event models for multilevel hierarchical data"

Samuel L Brilleman[1,2] *

Michael J Crowther[3]

Margarita Moreno-Betancur[2,4,5]

Jacqueline Buros Novik[6]

James Dunyak[7]

Nidal Al-Huniti[7]

Robert Fox[7]

Jeff Hammerbacher[6,8]

Rory Wolfe[1,2]

Author Affiliations: [1] Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia; [2] Victorian Centre for Biostatistics (ViCBiostat), Melbourne, Australia; [3] Biostatistics Research Group, Department of Health Sciences, University of Leicester, Leicester, UK; [4] Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Melbourne, Australia; [5] Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia; [6] Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA; [7] Quantitative Clinical Pharmacology, AstraZeneca, Waltham, MA, USA; [8] Department of Microbiology and Immunology, Medical University of South Carolina, Charleston, SC, USA;

*** Corresponding author:**

Postal: School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne, VIC 3004, Australia

Phone: +61 (4) 9903 0802

Email: sam.brilleman@monash.edu

Contents

1. Further details on the model estimation

For estimation of the model parameters we ran four MCMC chains in parallel, each with 1000 sample iterations preceded by a warm up period of 1000 iterations (i.e. 2000 iterations in total, of which 50% were warm up). Although this number of iterations would seem small for a complex model estimated using a Gibbs sampler, the estimation in Stan is based on a Hamiltonian Monte Carlo (HMC) algorithm, not Gibbs sampling. The HMC results in much lower autocorrelation between subsequent MCMC draws compared with Gibbs sampling, and therefore is much more efficient in terms of the effective sample size per iteration. For example, for each of the models with an association structure based on the expected value, the effective sample size for the estimated association parameter was 4000.

A potential limitation of the proposed approach is the additional computational complexity. Additional clustering factors mean that there are an increasing number of cluster-specific parameters (i.e. random effects) to be estimated and therefore computation time increases. In our application with 430 patients having a total of 1209 lesions, there were 430 patient-specific parameters (intercept only) and 3627 lesion-specific parameters (intercept and two polynomial terms) to be estimated. Computation time for the models with an association structure based on the expected value ranged between 1.5 and 3 hours. The differences in computation time were partly related to the random nature of the different MCMC chains, and partly related to the type of summary function used in the association structure (i.e. the sum, average, maximum, or minimum of the level 2 clusters). The type of association structure is of course part of the model definition and therefore the choice of association structure will have an influence on the shape of the target posterior distribution, with some resulting posteriors easier for the MCMC sampler to explore (i.e. less extreme curvature in

the posterior). When the association structure was based on both the expected value and the slope, the computation times were slightly longer; ranging between 2 and 5.5 hours. These times are based on 1000 warm up iterations, followed by 1000 sample iterations, on a standard quad-core desktop with a 3.30GHz processor and 8GB RAM.

2. Prior distributions used in our application

Table S1 shows the prior distributions used in our application, including the values of any relevant hyperparameters.

3. Sensitivity analysis using alternative priors for the baseline hazard parameters

At the suggestion of a reviewer, we conducted a sensitivity analysis to assess whether the choice of prior distribution for the B-spline coefficients (i.e. the log baseline hazard parameters) affected the results. Table S2 shows the estimated parameters for the joint model under three possible prior distributions for the B-spline coefficients:

1) Cauchy distribution with location 0 and scale 20 (this is the prior used for the main results presented in the manuscript)
2) Normal distribution with location 0 and scale (i.e. standard deviation) 50
3) Normal distribution with location 0 and scale (i.e. standard deviation) 200

There is very little difference in the estimated parameters across the three models, suggesting that the choice of prior distribution for the B-spline coefficients has little influence on the results.

4. Example code for fitting the model

The model in the paper can be easily estimated after downloading the rstanarm R package from the Comprehensive R Archive Network (CRAN). To download and install rstanarm, type the following into your R console:

```
> install.packages("rstanarm")
```

And then an example of the code used to fit the model presented in Table 2 of the main manuscript would be:

```
> library(rstanarm)
```

```
> mod <- stan_jm(
    formulaLong = ldiam ~
      egfrcat * poly(months, degree = 2) +
      (poly(months, degree = 2) | lesion_id) +
      (1 | patient_id),
    dataLong = data$lesions,
    formulaEvent = Surv(eventtime, status) ~ physical,
    dataEvent = data$surv,
    seed = 9837355, time_var = "months", id_var = "patient_id",
    assoc = c("etavalue", "etaslope"), grp_assoc = "max")
```

Where `data$lesions` is a data frame containing the outcome and covariate data for the longitudinal submodel, and `data$surv` is a data frame containing the outcome and covariate data for the event submodel. The IPASS data used in the application in the main manuscript is not publicly available. However, to demonstrate the structure of the longitudinal and event submodel data we show some simulated data here.

The following shows the structure of the longitudinal data (i.e. `data$lesions`):

```
> tbl_df(data$lesions)
   patient_id lesion_id months ldiam egfrcat    physical
   <fct>      <fct>      <dbl> <dbl> <fct>      <fct>
 1 P00001     P00001_1     0      30 Positive   Restricted
 2 P00001     P00001_2     0       9 Positive   Restricted
 3 P00001     P00001_3     0      17 Positive   Restricted
 4 P00001     P00001_1     1.3    33 Positive   Restricted
```

```
 5 P00001     P00001_2     1.3     8 Positive   Restricted
 6 P00001     P00001_3     1.3    14 Positive   Restricted
 7 P00001     P00001_1     3.6    39 Positive   Restricted
 8 P00001     P00001_2     3.6     8 Positive   Restricted
 9 P00001     P00001_3     3.6     6 Positive   Restricted
10 P00002     P00002_1     0      42 Negative_A Normal
11 P00002     P00002_2     0      40 Negative_A Normal
12 P00002     P00002_1     2.2    43 Negative_A Normal
13 P00002     P00002_2     2.6    32 Negative_A Normal
14 P00002     P00002_1     4.5    41 Negative_A Normal
15 P00002     P00002_2     4.3    16 Negative_A Normal
```

These data are simulated to come from two patients, with the first patient having three lesions and the second patient having two lesions. Each lesion has a unique ID (the variable '`lesion_id`') and there is a separate row in the dataset for each time-specific lesion-specific biomarker measurement. We see that for the first patient the measurement times are common across lesions (t = 0, 1.3 and 3.6), whereas for the second patient the measurement times differ across lesions (t = 0, 2.2 and 2.5 for their first lesion, and t = 0, 2.6, and 4.3 for their second lesion). The factor variables '`egfrcat`' and '`physical`' are patient-specific and time-fixed so they are constant across rows within a patient.

The following shows the structure of the event data (i.e. `data$surv`):

```
> tbl_df(data$surv)
  patient_id egfrcat    physical   eventtime status
  <fct>      <fct>      <fct>          <dbl>  <dbl>
1 P00001     Positive   Restricted       4.3      1
2 P00002     Negative_A Normal           7.5      0
```

We see the same two patients and a single row of data for each patient. The patient-specific time-fixed variables '`egfrcat`' and '`physical`' are shown again. The event time (in months) for each patient is shown in the variable '`eventtime`' whilst the variable '`status`' is the event indicator (taking the value 1 if the patient experienced the event or value 0 if they were censored).

5. Simulation study

We performed a small simulation study to evaluate the performance of the **rstanarm** package with regard to estimating the proposed model. The objective of the simulation study was to assess whether **rstanarm** was able to recover the true value for each of the parameters used in the data generating model. The data generating model, the inferential quantities, and the results of the simulation study are described in the following sections.

5.1 Data generating model

We considered two data generating models (i.e. two scenarios). The two models differed only in the specification of the association structure. For both data generating models the longitudinal submodel was specified as

$$
\begin{aligned}
y_{ij}(t_{ijk}) &= \mu_{ij}(t_{ijk}) + \epsilon_{ij}(t_{ijk}) \\
\mu_{ij}(t_{ijk}) &= \beta_0 + b_{i0} + u_{ij0} + \beta_1 x_{1i} + \beta_2 x_{2i} + (\beta_3 + u_{ij1}) t_{ijk} + \beta_4 {t_{ijk}}^2 + \beta_5 {t_{ijk}}^3 \\
\epsilon_{ij}(t_{ijk}) &\sim N(0, \sigma_\epsilon^2) \\
b_{i0} &\sim N(0, \sigma_b^2) \\
\begin{pmatrix} u_{ij0} \\ u_{ij1} \end{pmatrix} &\sim N(0, \mathbf{\Sigma_u})
\end{aligned}
\tag{1}
$$

where $y_{ij}(t_{ijk})$ denotes the longitudinal biomarker measurement at the k^{th} $(k = 1, \ldots, K_{ij})$ time point t_{ijk} for the j^{th} $(j = 1, \ldots, J_i)$ cluster within the i^{th} $(i = 1, \ldots, N)$ individual, $x_{1i} \sim \text{Bernoulli}(0.5)$ is an individual-level time-invariant binary covariate, $x_{2i} \sim N(0,1)$ is an individual-level time-invariant continuous covariate, b_{i0} denotes an individual-specific

parameter (i.e. random intercept) for individual i assumed to be drawn from a normal distribution with variance σ_b^2, u_{ij0} and u_{ij1} are cluster-specific parameters for cluster j assumed to be drawn from a multivariate normal distribution with variance-covariance matrix $\mathbf{\Sigma_u}$, β_0 through β_5 are population-level parameters, and $\epsilon_{ij}(t_{ijk})$ denotes the residual error term.

For the first data generating model (Model A) the event submodel was specified as

$$h_i(t) = \delta\ t^{\delta-1} \exp\left(\gamma_0 + \gamma_1\ x_{1i} + \gamma_2\ x_{2i} + \alpha^{(A)} \sum_{j=1}^{J_i} \mu_{ij}(t)\right) \tag{2}$$

and for the second data generating model (Model B) the event submodel was specified as

$$h_i(t) = \delta\ t^{\delta-1} \exp\left(\gamma_0 + \gamma_1\ x_{1i} + \gamma_2\ x_{2i} + \alpha^{(B)}\ \max(\mu_{ij}(t);\ j = 1, \ldots, J_i)\right) \tag{3}$$

where δ is the shape parameter for the Weibull baseline hazard, γ_0 is an intercept (i.e. log scale parameter for the Weibull baseline hazard), γ_1 and γ_2 are regression coefficients (i.e. log hazard ratios), and $\alpha^{(A)}$ and $\alpha^{(B)}$ are the association parameters in Models A and B, respectively. Therefore, the association structure in Model A is based on the summation of the cluster-specific expected values, whilst Model B is based on the maximum of the cluster-specific expected values (these association structures are discussed in greater detail in the main manuscript).

The true parameter values used in the simulation study are shown in Tables S3 and S4 (for data generating Models A and B, respectively).

5.2 Simulation scheme

We generated $D = 220$ datasets. For each dataset, we generated $N = 250$ individuals. For each individual, we generated J_i lower-level clusters where J_i was calculated as the integer component of a uniform random variable on the range 1 to 7. For each lower-level cluster, we generated 10 longitudinal measurements.

We simulated longitudinal measurements according to the longitudinal submodel defined in equation (1). The first longitudinal measurement was generated at baseline (i.e. time 0) with all remaining longitudinal measurements generated at times drawn from a uniform distribution on the range 0 to 20. We then generated an event time under the event submodel defined in equation (2) using the adapted cumulative hazard inversion method described by Crowther and Lambert [1] as implemented in the **simsurv** R package [2]. Any longitudinal measurement generated at a measurement time that occurred after the individual's event time was discarded.

5.3 Inferential quantities

We used the **simjm** R package to simulate the data [3]. An analysis model – intended to coincide with the data generating model – was fit to each simulated dataset using the **rstanarm** package [4,5]. We obtained a single chain of 2000 MCMC iterations, which included a warm-up phase of 1000 iterations that were not used for inference. Convergence for each simulated dataset was addressed by ensuring that each parameter had an estimated R-hat statistic less than 1.1 [6].

The ability to recover the true parameter values was assessed using the following approach. For the model that was fit to the d^{th} $(d = 1, \ldots, D)$ simulated dataset the following estimates were calculated:

- The mean of the posterior distribution for parameter k, denoted $\hat{\theta}_k^{(d)}$
- The bias for parameter k, defined as $\hat{B}_k^{(d)} = \hat{\theta}_k^{(d)} - \theta_k$ where θ_k denotes the true value of parameter k that was used to simulate the data
- The relative bias for parameter k, defined as $\hat{R}_k^{(d)} = 100 \times \theta_k^{-1}\left(\hat{\theta}_k^{(d)} - \theta_k\right)$ where θ_k denotes the true value of parameter k that was used to simulate the data
- The standard deviation of the posterior distribution (i.e. estimated standard error) for parameter k, denoted $\hat{S}_k^{(d)}$

For inference in the simulation study, the following quantities and plots were then calculated using the estimates obtained across the D datasets:

- The mean bias for parameter k, defined as $\bar{B}_k = \frac{1}{D}\sum_{\mathrm{d}=1}^{D} \hat{B}_k^{(d)}$ ("mean bias")
- The mean relative bias for parameter k, defined as $\bar{R}_k = \frac{1}{D}\sum_{\mathrm{d}=1}^{D} \hat{R}_k^{(d)}$ ("mean relative bias")
- The mean standard deviation of the posterior distribution for parameter k, defined as $\bar{S}_k = \frac{1}{D}\sum_{\mathrm{d}=1}^{D} \hat{S}_k^{(d)}$ ("mean estimated standard error")
- The standard error of the posterior mean for parameter k, defined as the standard deviation of the estimates $\left\{\hat{\theta}_k^{(d)}; d = 1, \ldots, D\right\}$ ("empirical standard error").

- Density plots of the posterior mean estimates $\left\{\hat{\theta}_k{}^{(d)}; d = 1, \ldots, D\right\}$ for parameter k; these were overlaid with a dashed line showing the true parameter value θ_k that was used to simulate the data ("empirical sampling distribution of the posterior mean").

5.4 Results

Tables S3 and S4 show the estimated mean bias, mean relative bias, mean estimated standard error, and empirical standard error for each of the parameters under Models A and B, respectively. Figures S1 and S2 show density plots of the posterior mean estimates for each parameter (i.e. the empirical distribution of $\hat{\theta}_k$) under Models A and B, respectively.

Overall, the results from the simulation study suggest that **rstanarm** was able to recover the true parameter values used in the data generating model. The true values for the parameters (the dashed lines in Figures S1 and S2) are located close to the centre of the sampling distribution for the posterior mean. Moreover, Tables S3 and S4 demonstrate that the mean estimated standard error (i.e. the mean standard deviation for the posterior distribution) was close to the empirical standard error for all parameters under both scenarios.

References

1. Crowther MJ, Lambert PC. Simulating biologically plausible complex survival data. Statistics in Medicine. 2013;32: 4118–4134. doi:10.1002/sim.5823

2. Brilleman SL. simsurv: Simulate Survival Data. R package version: 0.2.0. 2018; Available: https://CRAN.R-project.org/package=simsurv

3. Brilleman SL. simjm: Simulate Joint Longitudinal and Survival Data. R package version: 0.0.0. 2018; Available: https://github.com/sambrilleman/simjm

4. Stan Development Team. rstanarm: Bayesian applied regression modeling via Stan. R package version 2.17.1. 2017; Available: http://mc-stan.org/rstanarm

5. Brilleman SL, Crowther MJ, Moreno-Betancur M, Buros Novik J, Wolfe R. Joint longitudinal and time-to-event models via Stan. Proceedings of StanCon 2018. Pacific Grove, CA, USA; doi:10.5281/zenodo.1284334

6. Stan Development Team. Stan Modeling Language Users Guide and Reference Manual, Version 2.17.0. 2017. Available: http://mc-stan.org

Table S1. Prior distributions used in the non-small cell lung cancer (NSCLC) application presented in the main manuscript.

Parameter	Description of parameter	Prior distribution
Longitudinal submodel		
β_0	Population-level intercept	Normal(0, 179)
β_1	Population-level linear effect for time ("time")	Normal(0, 3475)
β_2	Population-level quadratic effect for time ("time^2")	Normal(0, 3480)
β_{31}	Population-level group effects, group 2	Normal(0 45)
β_{32}	Population-level group effects, group 3	Normal(0, 45)
β_{41}	Population-level group*time interaction effects, group 2	Normal(0, 1059)
β_{42}	Population-level group*time interaction effects, group 3	Normal(0, 11187)
β_{51}	Population-level group*time^2 interaction effects, group 2	Normal(0, 10812)
β_{52}	Population-level group*time^2 interaction effects, group 3	Normal(0, 11440)
Event submodel		
γ_{1g}	Population-level physical activity group effects	Normal(0, 2.50)
$\alpha_1^{(1)}$	Association parameter for the *sum* of the lesion-specific *expected values*	Normal(0, 0.05)
$\alpha_1^{(2)}$	Association parameter for the *average* of the lesion-specific *expected values*	Normal(0, 0.14)
$\alpha_1^{(3)}$	Association parameter for the *maximum* of the lesion-specific *expected values*	Normal(0, 0.12)
$\alpha_1^{(4)}$	Association parameter for the *minimum* of the lesion-specific *expected values*	Normal(0, 0.14)

$\alpha_2^{(1)}$	Association parameter for the *sum* of the lesion-specific *slopes*	Normal(0, 0.41)
$\alpha_2^{(2)}$	Association parameter for the *average* of the lesion-specific *slopes*	Normal(0, 1.75)
$\alpha_2^{(3)}$	Association parameter for the *maximum* of the lesion-specific *slopes*	Normal(0, 1.76)
$\alpha_2^{(4)}$	Association parameter for the *minimum* of the lesion-specific *slopes*	Normal(0, 1.67)
$\omega_1, \ldots., \omega_6$	Coefficients for the B-spline basis terms in the log baseline hazard	Cauchy(0, 20)

Random effects terms[a]

σ_ϵ	SD of residual error	Half-Cauchy(0, 90)
σ_b, σ_{u1}, σ_{u2}, σ_{u3},	Standard deviations for the lesion-specific and patient-specific parameters	Half-Cauchy(0, 10)
R_u	Correlation matrix for the lesion-specific parameters	LKJ (regularization = 1)[b]

[a] The variance-covariance matrix for the lesion-specific parameters, Σ_u is decomposed into a correlation matrix R_u and a vector of standard deviations $\sigma_u = (\sigma_{u1}, \sigma_{u2}, \sigma_{u3})$. Since there is only one patient-specific parameter there is no correlation matrix R_b, rather the variance-covariance matrix for the patient-specific parameters Σ_b is simply $\Sigma_b = \sigma_b^2$ where σ_b is the standard deviation of the patient-specific intercept.

[b] With a regularization parameter equal to 1, the LKJ correlation matrix prior distribution corresponds to a joint uniform prior over all possible correlation matrices. The technical details of the distribution are described in 'Lewandowski et al. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*. 2009; 100: 1989–2001'.

Table S2. Fixed effect parameter estimates (posterior means and 95% credible interval limits) from the joint model under three possible prior distributions for the B-spline coefficients (i.e. the log baseline hazard parameters). The estimates for the event submodel are on the log hazard scale.

	Prior distribution for the B-spline coefficients		
Parameter	Cauchy (0, 20)	Normal (0, 50)	Normal (0, 200)
Longitudinal submodel			
Intercept	22.98 (21.33 to 24.73)	23.03 (21.35 to 24.72)	22.94 (21.21 to 24.64)
Group (ref: EGFR+)			
EGFR-, carboplatin plus paclitaxel	4.03 (0.82 to 7.06)	3.93 (0.81 to 6.97)	4.00 (0.96 to 7.16)
EGFR-, gefitinib	16.92 (13.22 to 20.38)	16.94 (13.39 to 20.48)	16.95 (13.39 to 20.65)
Time effects			
Linear term (orthogonalised)	-0.07 (-73.26 to 76.72)	0.66 (-70.61 to 77.72)	-4.81 (-79.78 to 73.27)
Quadratic term (orthogonalised)	450.31 (391.57 to 512.48)	449.89 (391.12 to 511.41)	446.49 (385.13 to 510.54)
Group * Linear interaction			
EGFR-, carboplatin plus paclitaxel * Linear	315.21 (195.05 to 438.40)	313.88 (194.48 to 431.53)	313.24 (195.95 to 434.32)
EGFR-, gefitinib * Linear	389.96 (127.54 to 660.39)	391.70 (129.33 to 655.02)	390.24 (127.06 to 660.47)
Group * Quadratic interaction			
EGFR-, carboplatin plus paclitaxel * Quadratic	23.74 (-74.32 to 123.44)	23.65 (-76.39 to 124.86)	23.56 (-77.07 to 122.15)

EGFR-, gefitinib * Quadratic	-524.81 (-696.96 to -351.09)	-522.12 (-689.88 to -356.56)	-523.40 (-695.13 to -356.48)
Event submodel			
Intercept (adjusts for centering of predictors)	-1.22 (-1.77 to -0.67)	-1.21 (-1.74 to -0.69)	-1.21 (-1.74 to -0.64)
Physical functioning (ref: in bed >50% of the time)			
Normal activity	-0.46 (-0.93 to 0.02)	-0.47 (-0.91 to 0.01)	-0.47 (-0.94 to 0.01)
Restricted activity	-0.43 (-0.87 to 0.01)	-0.44 (-0.85 to 0.00)	-0.44 (-0.86 to -0.00)
Association parameters			
Value (diameter of largest lesion)	0.011 (0.004 to 0.017)	0.011 (0.004 to 0.017)	0.011 (0.004 to 0.017)
Slope (rate of change in fastest growing lesion)	0.447 (0.352 to 0.559)	0.449 (0.354 to 0.565)	0.448 (0.353 to 0.559)
Log baseline hazard parameters			
B-spline coefficient 1	-5.22 (-6.82 to -3.88)	-5.25 (-6.79 to -3.92)	-5.21 (-6.75 to -3.90)
B-spline coefficient 2	-0.28 (-0.99 to 0.42)	-0.27 (-0.97 to 0.42)	-0.28 (-0.96 to 0.40)
B-spline coefficient 3	-1.28 (-1.85 to -0.76)	-1.29 (-1.86 to -0.76)	-1.28 (-1.83 to -0.74)
B-spline coefficient 4	-0.61 (-1.88 to 0.62)	-0.60 (-1.94 to 0.63)	-0.58 (-1.91 to 0.65)
B-spline coefficient 5	-2.25 (-4.07 to -0.50)	-2.31 (-4.15 to -0.54)	-2.32 (-4.25 to -0.54)
B-spline coefficient 6	0.88 (-1.35 to 2.69)	0.90 (-1.25 to 2.67)	0.91 (-1.32 to 2.71)

Abbreviations. ref: reference category; EGFR: epidermal growth factor receptor (mutation status).

Table S3. Simulation study results for Model A (i.e. association structure based on the **summation** of the cluster-specific expected values); estimated mean bias, mean relative bias, mean estimated standard error, and empirical standard error for each of the parameters.

Parameter		True value	$\bar{B}_k$	$\bar{R}_k$	$\bar{S}_k$	$sd(\hat{\theta}_k)$
Longitudinal submodel						
β_0	Population-level intercept	10	0.0044	0.0441	0.0701	0.0781
β_1	Population-level coefficient for binary covariate	-1	0.0018	-0.1819	0.0981	0.1084
β_2	Population-level coefficient for continuous covariate	1	0.0016	0.1559	0.0497	0.0529
β_3	Population-level linear effect for time	-0.25	-0.0001	0.0417	0.0139	0.0136
β_4	Population-level quadratic effect for time	0.03	<0.0001	0.0523	0.0020	0.0018
β_5	Population-level cubic effect for time	-0.0015	<0.0001	0.0733	0.0001	0.0001
σ_ϵ	Standard deviation of the residual errors	0.5	0.0001	0.0248	0.0080	0.0081
Event submodel						
γ_0	Intercept (i.e. log scale parameter for Weibull baseline hazard)	-5	-0.1121	2.2419	0.3451	0.3340
γ_1	Coefficient (log hazard ratio) for binary covariate	-0.5	0.0028	-0.5655	0.1533	0.1738
γ_2	Coefficient (log hazard ratio) for continuous covariate	0.5	0.0013	0.2652	0.0838	0.0854

$\alpha^{(A)}$	Association parameter	0.1	0.0024	2.4182	0.0071	0.0070
δ	Shape parameter for Weibull baseline hazard	1.1	0.0335	3.0444	0.0691	0.0667
Random effects terms						
σ_b^2	Variance for individual-level random intercept term	0.25	-0.0003	-0.1156	0.0554	0.0562
$\mathbf{\Sigma}_{\mathrm{u}}[1,1]$	Variance for cluster-level random intercept term	1	0.0069	0.6872	0.0654	0.0669
$\mathbf{\Sigma}_{\mathrm{u}}[1,2]$	Covariance for cluster-level random intercept and slope terms	-0.014	<0.0001	0.0105	0.0058	0.0058
$\mathbf{\Sigma}_{\mathrm{u}}[2,2]$	Variance for cluster-level random (linear) slope term	0.0049	<0.0001	0.9981	0.0006	0.0006

Abbreviations. $\bar{B}_k$: mean bias for parameter k; $\bar{R}_k$: mean relative bias (%) for parameter k; $\bar{S}_k$: mean estimated standard error for parameter k; $sd(\hat{\theta}_k)$: empirical standard error of the posterior mean for parameter k; n/a: not applicable.

Table S4. Simulation study results for Model B (i.e. association structure based on the **maximum** of the cluster-specific expected values); estimated mean bias, mean relative bias, mean estimated standard error, and empirical standard error for each of the parameters.

Parameter		True value	$\bar{B}_k$	$\bar{R}_k$	$\bar{S}_k$	$sd(\hat{\theta}_k)$
Longitudinal submodel						
β_0	Population-level intercept	10	0.0035	0.0348	0.0711	0.0704
β_1	Population-level coefficient for binary covariate	-1	-0.0069	0.6862	0.0987	0.0996
β_2	Population-level coefficient for continuous covariate	1	-0.0009	-0.0853	0.0500	0.0502
β_3	Population-level linear effect for time	-0.25	-0.0002	0.0696	0.0117	0.0115
β_4	Population-level quadratic effect for time	0.03	-0.0001	-0.1720	0.0017	0.0016
β_5	Population-level cubic effect for time	-0.0015	<0.0001	-0.0667	0.0001	0.0001
σ_ϵ	Standard deviation of the residual errors	0.5	0.0012	0.2414	0.0069	0.0063
Event submodel						
γ_0	Intercept (i.e. log scale parameter for Weibull baseline hazard)	-5	-0.1026	2.0520	0.7795	0.7900
γ_1	Coefficient (log hazard ratio) for binary covariate	-0.5	-0.0006	0.1135	0.1598	0.1700
γ_2	Coefficient (log hazard ratio) for continuous covariate	0.5	0.0118	2.3506	0.1039	0.1016

$\alpha^{(B)}$	Association parameter	0.3	0.0075	2.4930	0.0691	0.0695
δ	Shape parameter for Weibull baseline hazard	1.1	0.0314	2.8577	0.0705	0.0730
Random effects terms						
σ_b^2	Variance for individual-level random intercept term	0.25	0.0042	1.6858	0.0548	0.0548
$\mathbf{\Sigma}_{\mathrm{u}}[1,1]$	Variance for cluster-level random intercept term	1	0.0070	0.7046	0.0638	0.0640
$\mathbf{\Sigma}_{\mathrm{u}}[1,2]$	Covariance for cluster-level random intercept and slope terms	-0.014	<0.0001	0.0489	0.0050	0.0051
$\mathbf{\Sigma}_{\mathrm{u}}[2,2]$	Variance for cluster-level random (linear) slope term	0.0049	<0.0001	0.3286	0.0005	0.0005

Abbreviations. $\bar{B}_k$: mean bias for parameter k; $\bar{R}_k$: mean relative bias (%) for parameter k; $\bar{S}_k$: mean estimated standard error for parameter k; $sd(\hat{\theta}_k)$: empirical standard error of the posterior mean for parameter k; n/a: not applicable.

Figure S1. Simulation study results for Model A (i.e. association structure based on the **summation** of the cluster-specific expected values); kernel density plots showing the distribution (across the 280 simulated datasets) of the estimated posterior mean for each parameter. The dashed line shows the true parameter value used in the data generating model.

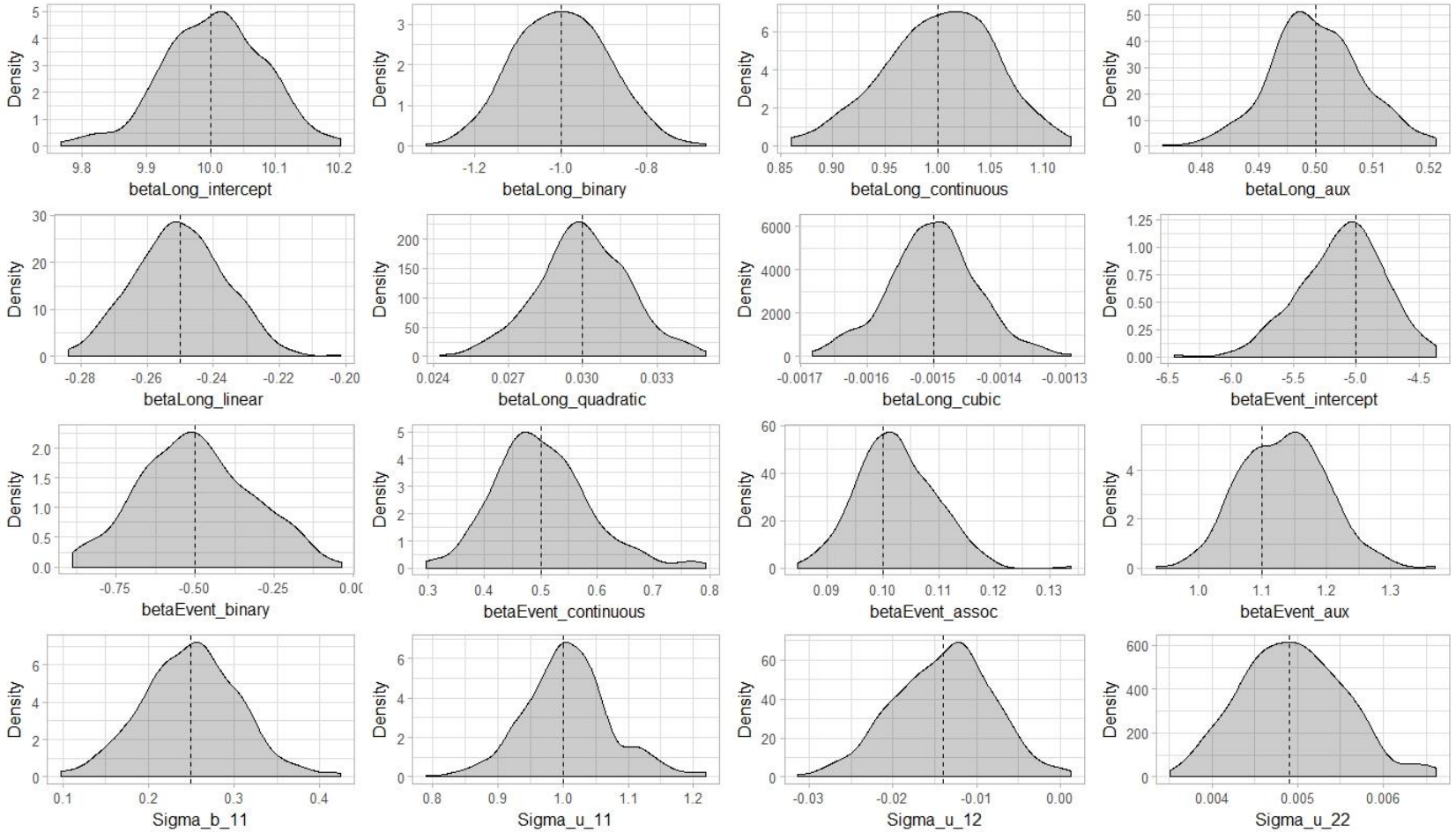

Figure S2. Simulation study results for Model B (i.e. association structure based on the **maximum** of the cluster-specific expected values); kernel density plots showing the distribution (across the 280 simulated datasets) of the estimated posterior mean for each parameter. The dashed line shows the true parameter value used in the data generating model.