Joint Models for Time-to-Event Data and Longitudinal Biomarkers of High Dimension – Supplementary Materials

Molei Liu · Jiehuan Sun · Jose D. Herazo-Maya · Naftali Kaminski · Hongyu Zhao

the date of receipt and acceptance should be inserted later

Appendix A

The model and prior specifications are described in the paper. Here we provide detailed MCMC procedure to estimate each factor in the model iteratively following its conditional distribution. Denote that

$$\begin{aligned} \phi(t) &= \begin{pmatrix} 1 \\ t \end{pmatrix} \\ \Phi(t) &= \phi(t) \cdot \phi(t)^T = \begin{pmatrix} 1 & t \\ t & t^2 \end{pmatrix} \end{aligned} \tag{S1}$$

For β and $b^{(i)}$, we denote that

$$\beta = \begin{pmatrix} \beta_1^T \\ \beta_2^T \\ \vdots \\ \beta_K^T \end{pmatrix} = (\widehat{\beta}_0, \widehat{\beta}_1), b^{(i)} = \begin{pmatrix} (b_1^{(i)})^T \\ (b_2^{(i)})^T \\ \vdots \\ (b_K^{(i)})^T \end{pmatrix} = (\widehat{b}_0^{(i)}, \widehat{b}_1^{(i)}) \tag{S2}$$

Molei Liu
Department of Biostatistics, Harvard School of Public Health, Harvard University, Boston, MA 02115, U.S.A.

Jiehuan Sun
Department of Biostatistics, Yale University, New Haven, CT 06510, U.S.A.

Jose D. Herazo-Maya, Naftali Kaminski
Pulmonary, Critical Care and Sleep Medicine, Yale School of Medicine, New Haven, CT 06519, U.S.A

Naftali Kaminski
Pulmonary, Critical Care and Sleep Medicine, Yale School of Medicine, New Haven, CT 06519, U.S.A

Hongyu Zhao
Department of Biostatistics, Yale University, New Haven, CT 06510, U.S.A.
E-mail: hongyu.zhao@yale.edu

Prior distributions of the parameters are as follows:

- $\lambda_g \sim \mathscr{N}(\mathbf{0}, \Sigma_g)$;
- $\sigma_g^2 \sim \text{InverseGamma}(\nu_1, \nu_2)$;
- $(\alpha^T, \gamma^T)^T \sim \mathscr{N}(\mu_h, \Sigma_h)$;
- $\rho_q \sim \text{Gamma}(\nu_3, \nu_4)$;
- $D_k \sim \text{InverseWishart}(\Sigma_b, n_b)$;
- $\beta_k = (\beta_{0k}, \beta_{1k})^T \sim N(\mathbf{0}, \text{diag}\{\tau_0^2, \tau_1^2\})$

where the specific values of these priors are provided in the paper. We use "$-$" to represent complementary set of the parameter. Conditional distributions used in the algorithm are as follows.

- $\lambda_g | \lambda_g^- \sim \mathscr{N}(\widetilde{\mu}_g, \widetilde{\Psi}_g)$, where

$$\begin{aligned}\widetilde{\mu}_g =& \widetilde{\Psi}_g \frac{1}{\sigma_g^2} \sum_{i=1}^{n} \sum_{t=t_1}^{t_{n_i}} y_g^{(i)}(t)(\beta + b^{(i)})\phi(t) \\ \widetilde{\Psi}_g =& \left[\Sigma_g^{-1} + \frac{1}{\sigma_g^2} \sum_{i=1}^{n} \sum_{t=t_1}^{t_{n_i}} (\beta + b^{(i)})\Phi(t)(\beta + b^{(i)})^T\right]^{-1}\end{aligned} \tag{S3}$$

- $\sigma_g^2 | (\sigma_g^2)^- \sim$

$$\text{InverseGamma}\left(\nu_1 + \frac{1}{2}\sum_{i=1}^{n} n_i, \nu_2 + \frac{1}{2}\sum_{i=1}^{n}\sum_{t=t_1}^{t_{n_i}} [y_g^{(i)}(t) - \lambda_g^T(\beta + b^{(i)})\phi(t)]^2\right) \tag{S4}$$

- $\rho_q | \rho_q^- \sim \text{Gamma}(\widetilde{\nu}_3, \widetilde{\nu}_4)$, where

$$\begin{aligned}\widetilde{\nu}_3 =& \nu_3 + \sum_{i=1}^{n} \delta_i I(B_q < T_i \le B_{q+1}) \\ \widetilde{\nu}_4 =& \nu_4 + \sum_{i=1}^{n} I(T_i > B_q) \int_{B_q}^{\min(B_{q+1}, T_i)} \exp\{\gamma^T \omega_i + \alpha^T(\beta + b^{(i)})\phi(t)\} ds \\ =& \nu_4 + \sum_{i=1}^{n} \frac{I(T_i > B_q)}{\alpha^T(\widehat{\beta}_1 + \widehat{b}_1^{(i)})} \left[\exp\{\gamma^T \omega_i + \alpha^T(\beta + b^{(i)})\phi(t)\}\Big|_{B_q}^{\min(B_{q+1}, T_i)}\right]\end{aligned} \tag{S5}$$

- $D_k | D_k^- \sim$

$$\text{InverseWishart}\left(\left[\sum_{i=1}^{n} b_k^{(i)}(b_k^{(i)})^T + \Sigma_b^{-1}\right], n_b + n\right) \tag{S6}$$

- Implement Metropolis-Hastings Algorithm to sample α and γ from their conditional distributions. Denote the jump covariance matrix as Σ_h^{jump}. In the tth iteration, we draw $(\alpha;\gamma)^{new}$ from $\mathscr{N}((\alpha;\gamma)^{t-1}, \Sigma_h^{jump})$ and let

$$
\begin{aligned}
r =&\exp\left\{-\frac{1}{2}[(\alpha;\gamma)-\mu_h]^T\Sigma_h^{-1}[(\alpha;\gamma)-\mu_h]\Big|_{(\alpha;\gamma)^{t-1}}^{(\alpha;\gamma)^{new}}\right\}\\
&\times\exp\left\{\sum_{i=1}^{n}\left[\delta_i(\gamma^T\omega_i+\alpha^T\eta^{(i)}(T_i))-\int_0^{T_i}h_0(s)\exp\{\gamma^T\omega_i+\alpha^T\eta^{(i)}(s)\}ds\right]\Bigg|_{(\alpha;\gamma)^{t-1}}^{(\alpha;\gamma)^{new}}\right\}
\end{aligned}
\tag{S7}
$$

where $(\cdot;\cdot)$ is for the column juxtaposition of two vectors. If $r\geq 1$, we set $(\alpha;\gamma)^t=(\alpha;\gamma)^{new}$. Otherwise, we set $(\alpha;\gamma)^t=(\alpha;\gamma)^{new}$ with probability r.

To sample β and $b^{(i)}$ more efficiently, we adopt the strategy that samples $\widehat{\beta}_0$, $\widehat{\beta}_1$ from their joint distribution. Similarly, we sample factors in each $b^{(i)}$ jointly.

- For $\widehat{\beta}=(\widehat{\beta}_0;\widehat{\beta}_1^T)$, take the jump covariance matrix as Σ_β^{jump}. In the tth iteration, we draw $\widehat{\beta}^{new}$ from $N(\widehat{\beta}^{t-1},\Sigma_\beta^{jump})$, denote that

$$
\begin{aligned}
\widehat{\mu} =&\widehat{V}\sum_{g=1}^{G}\sum_{i=1}^{n}\sum_{t=t_1}^{t_{n_i}}\frac{1}{\sigma_g^2}\left[y_g^{(i)}(t)\phi(t)\otimes\lambda_g-\phi(t)\otimes\lambda_g\lambda_g^Tb^{(i)}\phi(t)\right]\\
\widehat{V} =&\left[\tau\otimes I_{k\times k}+\sum_{g=1}^{G}\sum_{i=1}^{n}\sum_{t=t_1}^{t_{n_i}}\frac{1}{\sigma_g^2}\Phi(t)\otimes\lambda_g\lambda_g^T\right]^{-1}
\end{aligned}
\tag{S8}
$$

Operator $\otimes$ in Equation (S8) represents Kronecker Product. And

$$
\tau=\begin{pmatrix}\frac{1}{\tau_0^2} & 0\\ 0 & \frac{1}{\tau_1^2}\end{pmatrix}
\tag{S9}
$$

Let

$$
\begin{aligned}
r =&\exp\left\{-\frac{1}{2}(\widehat{\beta}-\widehat{\mu})^T\widehat{V}^{-1}(\widehat{\beta}-\widehat{\mu})\Big|_{\widehat{\beta}^{t-1}}^{\widehat{\beta}^{new}}\right\}\\
&\times\exp\left\{\sum_{i=1}^{n}\left(\delta_i\alpha^T\beta\phi(T_i)-\int_0^{T_i}h_0(s)\exp\{\gamma^T\omega_i+\alpha^T\eta^{(i)}(s)\}ds\right)\Bigg|_{\widehat{\beta}^{t-1}}^{\widehat{\beta}^{new}}\right\}
\end{aligned}
\tag{S10}
$$

If $r\geq 1$, we set $\widehat{\beta}^t=\widehat{\beta}^{new}$. Otherwise, we set $\widehat{\beta}^t=\widehat{\beta}^{new}$ with probability r.

- For each $\widehat{b}^{(i)}=(\widehat{b}_0^{(i)};\widehat{b}_1^{(i)})$, take the jump covariance matrix as Σ_b^{jump}. In the tth iteration, we draw $\widehat{\beta}^{new}$ from $N(\widehat{\beta}^{t-1},\Sigma_b^{jump})$, denote that

$$
\begin{aligned}
\widehat{\mu} =& \widehat{V} \sum_{g=1}^{G} \sum_{t=t_1}^{t_{n_i}} \frac{1}{\sigma_g^2} \left[y_g^{(i)}(t)\phi(t) \otimes \lambda_g - \phi(t) \otimes \lambda_g \lambda_g^T \beta \phi(t) \right] \\
\widehat{V} =& \left[\widehat{\Sigma}_b + \sum_{g=1}^{G} \sum_{t=t_1}^{t_{n_i}} \frac{1}{\sigma_g^2} \Phi(t) \otimes \lambda_g \lambda_g^T \right]^{-1}
\end{aligned}
\tag{S11}
$$

The covariance matrix $\widehat{\Sigma}_b$ in Equation (S11) is obtained from marginal distribution of $b_k^{(i)}$ by inserting each factor to corresponding entry.
Let

$$
\begin{aligned}
r =& \exp\left\{ -\frac{1}{2} (\widehat{b}^{(i)} - \widehat{\mu})^T \widehat{V}^{-1} (\widehat{b}^{(i)} - \widehat{\mu}) \Big|_{(\widehat{b}^{(i)})^{t-1}}^{(\widehat{b}^{(i)})^{new}} \right\} \\
& \times \exp\left\{ \left(\delta_i \alpha^T b^{(i)} \phi(T_i) - \int_0^{T_i} h_0(s) \exp\{\gamma^T \omega_i + \alpha^T \eta^{(i)}(s)\} ds \right) \Bigg|_{(\widehat{b}^{(i)})^{t-1}}^{(\widehat{b}^{(i)})^{new}} \right\}
\end{aligned}
\tag{S12}
$$

If $r \geq 1$, we set $(\widehat{b}^{(i)})^t = (\widehat{b}^{(i)})^{new}$. Otherwise, we set $(\widehat{b}^{(i)})^t = (\widehat{b}^{(i)})^{new}$ with probability r.

Appendix B

As the scheme for individualized prediction is given in the paper, we implement the following algorithm to estimate $\zeta_j(t+c|t) = \zeta_j(u|t)$ for subject j in the testing data and given $c > 0$:

1. Draw parameters $\theta = (\alpha, \gamma, \rho, \Lambda, \beta, \sigma_g^2, D_k, \widehat{\Sigma}_b)$ from their posterior distribution fitted by the training data.
2. Draw the initial value $(b^{(j)})^0$ that each $b_k^{(j)}$ following $N(0, D_k)$.
3. Implement Metropolis-Hasting algorithm to draw $b^{(j)}$ from its posterior distribution conditional on θ and $\mathscr{R}_j(t)$. More specifically, repeat the following steps for 100 to 200 times:
   - Draw $(\widehat{b}^{(j)})^{new}$ from $\mathscr{N}(\widetilde{\mu}, \widetilde{V})$, where

$$
\begin{aligned}
\widehat{\mu} =& \widehat{V} \sum_{g=1}^{G} \sum_{t=t_1}^{t_{n_j}} \frac{1}{\sigma_g^2} \left[y_g^{(j)}(t)\phi(t) \otimes \lambda_g - \phi(t) \otimes \lambda_g \lambda_g^T \beta \phi(t) \right] \\
\widehat{V} =& \left[\widehat{\Sigma}_b + \sum_{g=1}^{G} \sum_{t=t_1}^{t_{n_j}} \frac{1}{\sigma_g^2} \Phi(t) \otimes \lambda_g \lambda_g^T \right]^{-1}
\end{aligned}
\tag{S13}
$$

- Let

$$r = \exp\left\{ \left(-\int_0^t h_0(s) \exp\{\gamma^T \omega_j + \alpha^T \eta^{(j)}(s)\} ds \right) \Bigg|_{(\widehat{b}^{(j)})^{t-1}}^{(\widehat{b}^{(j)})^{new}} \right\} \tag{S14}$$

- If $r \geq 1$, we set $(\widehat{b}^{(j)})^t = (\widehat{b}^{(j)})^{new}$. Otherwise, we set $(\widehat{b}^{(j)})^t = (\widehat{b}^{(j)})^{new}$ with probability r.

4. Calculate and record $\zeta_j(t+c|t)$ in Equation (S15) with $b^{(j)}$ given by step 3.

$$\zeta_j(t+c|t) = \frac{S(t+c|\theta, b^{(j)})}{S(t|\theta, b^{(j)})} = \exp\left\{ -\int_t^{t+c} h_0(s) \exp\{\gamma^T \omega_j + \alpha^T \eta^{(j)}(s)\} ds \right\} \tag{S15}$$

5. Repeat steps 1 to 4 for 1000 to 2000 times to estimate average value and confidence interval of $\zeta_j(t+c|t)$.

Appendix C

In this section, we include supplemental numerical results to our paper, including (i) the MCMC convergence diagnostic (trace plot and Deweke score) for simulation study and real data analysis; (ii) the results of sensitivity analysis for our simulation study in described in Section 4.4 and (iii) the results of the sensitivity analysis on the choice of the prior in the real data analysis.

Fig. S1 Diagnostic of MCMC sampling in simulation studies. Some important and representative parameters are picked (from left hand side to right hand side): $b_{2,1}^{(8)}$ (coefficient for random effect of subject 8 of factor 2), $\beta_{3,1}$ (coefficient for fixed effect of factor 3) and $\boldsymbol{\Lambda}_{10,2}$ (loading factor for the 10th gene corresponding to factor 2). Their Geweke score between the first 10% and the last 50% samples are respectively: -1.00, -1.04 and -0.67. When the MCMC converge well, the Geweke score can be regarded as a Z-score. So when it leaves far from standard normal distribution, there is indication for bad convergence of MCMC. In our cases, their is no indication from trace plots and the Geweke scores for convergence issue of MCMC.

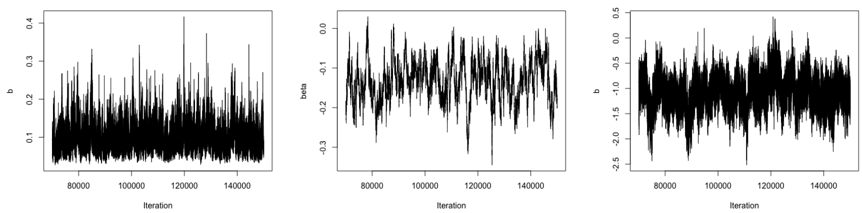

Fig. S2 Diagnostic of MCMC sampling in simulation studies. Some important and representative parameters are picked (from left hand side to right hand side): $b_{2,1}^{(1)}$ (coefficient for random effect of subject 1 of factor 2), $\beta_{3,1}$ (coefficient for fixed effect of factor 3) and $\boldsymbol{\Lambda}_{15,3}$ (loading factor for the 15th gene corresponding to factor 3). Their Geweke score between the first 10% and the last 50% samples are respectively: 0.00, -0.08 and -0.63. Again, their is no indication from trace plots and the Geweke scores for convergence issue of MCMC.

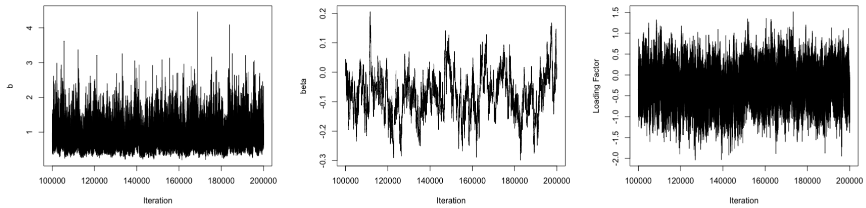

Fig. S3 Sensitivity analysis on the hyper-parameters in our prior. For the real data analysis, we double all the non-zero hyper-parameters except those of the loading factors $\boldsymbol{\Lambda}$ in the prior (the variance parameter for the prior of $\boldsymbol{\Lambda}$ will have non-essential effect on the scale of the loading factors and the latent variables). In the left panel, we show the trace plots of the same parameters as Figure S3 for our original choice on the prior. While in the right panel, we show the corresponding plots for the prior with doubled hyper-parameters. It can be seen that there is no essential difference between the results from the two choices on our prior.