**Supplement to**
**"Bayesian Model Assessment in Joint Modeling of Longitudinal and Survival Data with Applications to Cancer Clinical Trials"**

Danjie Zhang, Ming-Hui Chen, Joseph G. Ibrahim, Mark E. Boye, and Wei Shen

# Appendix A: Prior Specification and Posterior Computation

For clarity of exposition, we consider both the trajectory model (TM) and the shared parameter model (SPM) and a piecewise constant hazard function for $\lambda_0(t)$ defined in (4.1) in this Appendix. Write $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)'$. We assume independent normal priors for $\boldsymbol{\gamma}$, $\boldsymbol{\theta}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$, with $\boldsymbol{\gamma} \sim N(\boldsymbol{\mu}_1, V_1)$, $\boldsymbol{\theta} \sim N(\boldsymbol{\mu}_2, V_2)$, $\boldsymbol{\alpha} \sim N(\boldsymbol{\mu}_3, V_3)$, $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_4, V_4)$, where $\boldsymbol{\mu}_j$ and $V_j$ (a positive definite matrix) for $j = 1, 2, 3, 4$ are prespecified hyperparameters. An inverse gamma (IG) prior is specified for $\sigma^2$, i.e., $\sigma^2 \sim \mathrm{IG}(a_0, b_0)$, where $a_0 \geq 0$ and $b_0 \geq 0$ are prespecified. For $\Omega$, we take an inverse Wishart (IW) prior,

$$\Omega \sim \mathrm{IW}_{q+1}(\nu_0, \Omega_0),$$

where $\nu_0$ is the degrees of freedom and $\Omega_0$ is a $(q+1) \times (q+1)$ positive definite matrix. We further assume independent gamma priors for $\boldsymbol{\lambda}$, $\lambda_k \sim \mathrm{Gamma}(a_k, b_k)$, where $a_k \geq 0$ and $b_k \geq 0$ for $k = 1, \ldots, K$. In Section 3 (a simulation study) and Section 4 (the real data analysis), the prespecified hyperparameters are given by $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \boldsymbol{\mu}_4 = \mathbf{0}$, $V_1 = 1000 I_p$, $V_2 = 1000 I_2$, $V_3 = 10000 I_2$, $V_4 = 10000 I_p$, $a_0 = 0.001$, $b_0 = 0.001$, $v_0 = 2$, $\Omega_0 = 0.001 I_2$, $a_1 = 0.001$, and $b_1 = 0.001$.

Finally, the joint prior is written as

$$\begin{aligned}
\pi(\boldsymbol{\varphi}) &= \pi(\boldsymbol{\gamma})\pi(\boldsymbol{\theta})\pi(\sigma^2)\pi(\Omega)\pi(\boldsymbol{\lambda})\pi(\boldsymbol{\alpha})\pi(\boldsymbol{\beta}) \\
&\propto \exp\Big[-\frac{1}{2}(\boldsymbol{\gamma} - \boldsymbol{\mu}_1)' V_1^{-1} (\boldsymbol{\gamma} - \boldsymbol{\mu}_1)\Big] \times \exp\Big[-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu}_2)' V_2^{-1} (\boldsymbol{\theta} - \boldsymbol{\mu}_2)\Big] \\
&\times (\sigma^2)^{-(a_0+1)} \exp\Big(-\frac{b_0}{\sigma^2}\Big) |\Omega|^{-\frac{\nu_0 + (q+1) + 1}{2}} \exp\Big[-\frac{1}{2}\mathrm{tr}(\Omega_0 \Omega^{-1})\Big] \times \Big[\prod_{k=1}^{K} \lambda_k^{a_k - 1} \exp(-b_k \lambda_k)\Big] \\
&\times \exp\Big[-\frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\mu}_3)' V_3^{-1} (\boldsymbol{\alpha}' - \boldsymbol{\mu}_3)\Big] \times \exp\Big[-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_4)' V_4^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_4)\Big].
\end{aligned} \tag{A.1}$$

Due to the complexity of the joint model, an analytical evaluation of the posterior distribution of $(\boldsymbol{\varphi}, \boldsymbol{\theta}^R)$ in (2.8) with the prior specified in (A.1) is not possible. Therefore, we use the Gibbs sampling algorithm to sample $(\boldsymbol{\varphi}, \boldsymbol{\theta}^R)$ from (2.8). To run the Gibbs sampling algorithm, we draw from the following full conditional distributions in turn: (i) $[\boldsymbol{\gamma} \mid \sigma^2, \boldsymbol{\theta}^R, D_{\mathrm{obs}}]$; (ii) $[\boldsymbol{\theta} \mid \Omega, \boldsymbol{\theta}^R, D_{\mathrm{obs}}]$; (iii) $[\Omega \mid \boldsymbol{\theta}, \boldsymbol{\theta}^R, D_{\mathrm{obs}}]$;

(iv) $[\sigma^2 \mid \boldsymbol{\gamma}, \boldsymbol{\theta}^R, D_{\text{obs}}]$; (v) $[\boldsymbol{\alpha} \mid \boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\theta}^R, D_{\text{obs}}]$; (vi) $[\boldsymbol{\beta} \mid \boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\theta}^R, D_{\text{obs}}]$; (vii) $[\boldsymbol{\lambda} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}^R, D_{\text{obs}}]$; and (viii) $[\boldsymbol{\theta}^R \mid \boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma^2, \Omega, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\beta}, D_{\text{obs}}]$.

We briefly discuss how to sample from the above conditional distributions. For (i) and (ii), the conditional distributions of $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ are multivariate normal; for (iii), the conditional distribution for $\Omega$ is inverse Wishart; and for (iv), the conditional distribution is inverse gamma. For (vii), given $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}^R, D_{\text{obs}})$, the $\lambda_k$'s are conditionally independent and each $\lambda_k$ follows a gamma distribution. Thus, sampling from these four conditional distributions is straightforward. For (vi), it can be shown that the conditional density is log-concave. Thus, we can use the adaptive rejection algorithm of Gilks and Wild (1992) to sample these parameters. For (v) and (viii), we use the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970) to sample $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}^R$.

Next, we derive the expressions of these full conditional distributions. For (i), we have

$$\boldsymbol{\gamma}|\sigma^2, \boldsymbol{\theta}^R,\ D_{\text{obs}} \ \sim\ N\Big(\Sigma_{\boldsymbol{\gamma}}\Big[\frac{1}{\sigma^2}\sum_{i=1}^{n}\begin{pmatrix}\mathbf{0} & I_p\end{pmatrix}\boldsymbol{X}_i'\Big(\boldsymbol{y}_i - \boldsymbol{X}_i\begin{pmatrix}I_{q+1}\\ \mathbf{0}\end{pmatrix}\boldsymbol{\theta}_i\Big) + V_1^{-1}\boldsymbol{\mu}_1\Big], \Sigma_{\boldsymbol{\gamma}}\Big),$$

where $\Sigma_{\boldsymbol{\gamma}} = \Big[V_1^{-1} + \frac{1}{\sigma^2}\sum_{i=1}^{n}\begin{pmatrix}\mathbf{0} & I_p\end{pmatrix}\boldsymbol{X}_i'\boldsymbol{X}_i\begin{pmatrix}\mathbf{0}\\ I_p\end{pmatrix}\Big]^{-1}$. For (ii), given $(\Omega, \boldsymbol{\theta}^R, D_{\text{obs}})$, $\boldsymbol{\theta}$ follows a multivariate normal distribution, given by

$$\boldsymbol{\theta}|\Omega, \boldsymbol{\theta}^R, D_{\text{obs}} \ \sim\ N\Big((n\Omega^{-1} + V_2^{-1})^{-1}\Big(\Omega^{-1}\sum_{i=1}^{n}\boldsymbol{\theta}_i + V_2^{-1}\boldsymbol{\mu}_2\Big), (n\Omega^{-1} + V_2^{-1})^{-1}\Big).$$

For (iii), $[\Omega|\boldsymbol{\theta}, \boldsymbol{\theta}^R, D_{\text{obs}}]$ is an inverse Wishart (IW) distribution given by

$$\Omega|\boldsymbol{\theta}, \boldsymbol{\theta}^R, D_{\text{obs}} \ \sim\ \text{IW}_{q+1}\Big(\nu_0 + n, \Omega_0 + \sum_{i=1}^{n}(\boldsymbol{\theta}_i - \boldsymbol{\theta})(\boldsymbol{\theta}_i - \boldsymbol{\theta})'\Big).$$

For (iv), the conditional distribution of $\sigma^2$ is given by

$$\sigma^2|\boldsymbol{\gamma}, \boldsymbol{\theta}^R, D_{\text{obs}} \ \sim\ \text{IG}\Big(a_0 + \sum_{i=1}^{n} m_i/2, b_0 + \sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{X}_i(\boldsymbol{\theta}_i', \boldsymbol{\gamma}')')'(\boldsymbol{y}_i - \boldsymbol{X}_i(\boldsymbol{\theta}_i', \boldsymbol{\gamma}')')/2\Big).$$

For (v), the full conditional distribution of $\boldsymbol{\alpha}$ takes the form

$$\pi(\boldsymbol{\alpha}|\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\theta}^R, D_{\text{obs}}) \propto \exp\Big[\sum_{i=1}^{n}\delta_i\{h(\boldsymbol{\alpha}, \boldsymbol{\theta}_i, \boldsymbol{g}(t_i)) + \boldsymbol{z}_i'\boldsymbol{\beta}\} - \sum_{i=1}^{n}\exp(\boldsymbol{z}_i'\boldsymbol{\beta})\sum_{k=1}^{K}\lambda_k H_{ik} - \frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\mu}_3)'V_3^{-1}(\boldsymbol{\alpha} - \boldsymbol{\mu}_3)\Big],$$

where

$$H_{ik} = 1\{s_{k-1} < t_i\}\int_{s_{k-1}}^{\min\{s_k, t_i\}}\exp\{h(\boldsymbol{\alpha}, \boldsymbol{\theta}_i, \boldsymbol{g}(u))\}du,$$

and the indicator function $1\{s_{k-1} < t_i\} = 1$ if $s_{k-1} < t_i$ and 0 otherwise. For TM, the Riemann integral is used to compute $H_{ik}$ as the integral involved is intractable. For (vi), the full conditional distribution of $\boldsymbol{\beta}$ is given by

$$\pi(\boldsymbol{\beta}|\boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\theta}^R, D_{\text{obs}}) \propto \exp\Big[\sum_{i=1}^n \delta_i\{h(\boldsymbol{\alpha}, \boldsymbol{\theta}_i, \boldsymbol{g}(t_i)) + \boldsymbol{z}_i'\boldsymbol{\beta}\} - \sum_{i=1}^n \exp(\boldsymbol{z}_i'\boldsymbol{\beta}) \sum_{k=1}^K \lambda_k H_{ik} - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\mu}_4)' V_4^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}_4)\Big].$$

It is easy to show that $\log(\pi(\boldsymbol{\beta}|\boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\theta}^R, D_{\text{obs}}))$ is concave. For (vii), given $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}^R, D_{\text{obs}})$, the $\lambda_k$'s are conditionally independent and for each $\lambda_k$, we have

$$\lambda_k|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}^R, D_{\text{obs}} \sim \text{Gamma}(a_k + d_k, b_k + \tau_k),$$

where $d_k = \sum_{i=1}^n \delta_i I(s_{k-1} < t_i \leq s_k)$, and $\tau_k = \sum_{i=1}^n \exp\{\boldsymbol{z}_i'\boldsymbol{\beta}\} H_{ik}$. Finally, for (viii), the $\boldsymbol{\theta}_i$'s are conditionally independent and the full conditional distribution of $\boldsymbol{\theta}_i$ takes the form

$$\pi(\boldsymbol{\theta}_i|\boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma^2, \Omega, \boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\beta}, D_{\text{obs}}) \propto \exp\Big[\delta_i\{h(\boldsymbol{\alpha}, \boldsymbol{\theta}_i, \boldsymbol{g}(t_i)) + \boldsymbol{z}_i'\boldsymbol{\beta}\} - \exp(\boldsymbol{z}_i'\boldsymbol{\beta}) \sum_{k=1}^K \lambda_k H_{ik} - \frac{1}{2\sigma^2}\Big[(\boldsymbol{y}_i - \boldsymbol{X}_i(\boldsymbol{\theta}_i', \boldsymbol{\gamma}')')'(\boldsymbol{y}_i - \boldsymbol{X}_i(\boldsymbol{\theta}_i', \boldsymbol{\gamma}')')\Big] - \frac{1}{2}(\boldsymbol{\theta}_i - \boldsymbol{\theta})'\Omega^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\theta})\Big].$$

# Appendix B: Decomposition II

## Appendix B.1: DIC Decomposition II and $\Delta\text{DIC}_{\text{Long}}$

Write $\boldsymbol{\psi}_1 = (\boldsymbol{\gamma}, \sigma^2)$ and $\boldsymbol{\psi}_2 = (\boldsymbol{\lambda}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \Omega)$. In this Appendix, we extend AIC decomposition II of Zhang et al. (2015a) to develop the second decomposition of DIC. The DIC decomposition II quantifies the contribution of the survival data to the fit of the longitudinal data.

Let $f(\boldsymbol{\theta}_i|\boldsymbol{\psi}_2, t_i, \delta_i, \boldsymbol{z}_i)$ be the conditional density of the random effects $\boldsymbol{\theta}_i$ given $(t_i, \delta_i, \boldsymbol{z}_i)$, and also let $f(t_i|\boldsymbol{\psi}_2, \delta_i, \boldsymbol{z}_i) = \int f(t_i|\boldsymbol{\varphi}_2, \boldsymbol{\theta}_i, \delta_i, \boldsymbol{z}_i) f(\boldsymbol{\theta}_i|\boldsymbol{\theta}, \Omega) d\boldsymbol{\theta}_i$, which is the marginal function of $t_i$. Let $\bar{\boldsymbol{\psi}}_1$ and $\bar{\boldsymbol{\psi}}_2$ denote the posterior means of $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$. Define $\text{Dev}_{\text{Surv}}(\bar{\boldsymbol{\varphi}}) = -2\sum_{i=1}^n \log f(t_i|\bar{\boldsymbol{\psi}}_2, \delta_i, \boldsymbol{z}_i)$, $p_{D[\text{Surv}]} = E\Big[-2\sum_{i=1}^n f(t_i|\boldsymbol{\psi}_2, \delta_i, \boldsymbol{z}_i)\Big|D_{\text{obs}}\Big] + 2\sum_{i=1}^n \log f(t_i|\bar{\boldsymbol{\psi}}_2, \delta_i, \boldsymbol{z}_i)$, $\text{Dev}_{\text{Long|Surv}}(\bar{\boldsymbol{\varphi}}) = -2\sum_{i=1}^n \log( \int f(\boldsymbol{y}_i|\bar{\boldsymbol{\psi}}_1, \boldsymbol{\theta}_i, \boldsymbol{x}_i) f(\boldsymbol{\theta}_i|\bar{\boldsymbol{\psi}}_2, t_i, \delta_i, \boldsymbol{z}_i) d\boldsymbol{\theta}_i)$, and $p_{D[\text{Long|Surv}]} = E\Big[-2\sum_{i=1}^n \log \int f(\boldsymbol{y}_i|\boldsymbol{\psi}_1, \boldsymbol{\theta}_i, \boldsymbol{x}_i) f(\boldsymbol{\theta}_i|\boldsymbol{\psi}_2, t_i, \delta_i, \boldsymbol{z}_i) d\boldsymbol{\theta}_i\Big|D_{\text{obs}}\Big] + 2\sum_{i=1}^n \log \int f(\boldsymbol{y}_i|\bar{\boldsymbol{\psi}}_1, \boldsymbol{\theta}_i, \boldsymbol{x}_i) f(\boldsymbol{\theta}_i|\bar{\boldsymbol{\psi}}_2, t_i, \delta_i, \boldsymbol{z}_i) d\boldsymbol{\theta}_i$. We are led to the following result.

**Result B.1:** *DIC and $p_D$ in (2.9) have the decomposition*

$$\text{DIC} = \text{DIC}_{\text{Surv}} + \text{DIC}_{\text{Long|Surv}},$$

$$p_D = p_{D[\text{Surv}]} + p_{D[\text{Long|Surv}]},$$

*where* $\text{DIC}_{\text{Surv}} = \text{Dev}_{\text{Surv}}(\bar{\boldsymbol{\varphi}}) + 2p_{D[\text{Surv}]}$, *and* $\text{DIC}_{\text{Long|Surv}} = \text{Dev}_{\text{Long|Surv}}(\bar{\boldsymbol{\varphi}}) + 2p_{D[\text{Long|Surv}]}$.

The proof of Result B1 is based on the key identity

$$f(t_i|\boldsymbol{\varphi}_2, \boldsymbol{\theta}_i, \delta_i, \boldsymbol{z}_i) f(\boldsymbol{\theta}_i|\boldsymbol{\theta}, \Omega) = f(t_i|\boldsymbol{\varphi}_2, \boldsymbol{\theta}, \Omega, \delta_i, \boldsymbol{z}_i) f(\boldsymbol{\theta}_i|\boldsymbol{\varphi}_2, \boldsymbol{\theta}, \Omega, t_i, \delta_i, \boldsymbol{z}_i), \tag{A.2}$$

where

$$f(\boldsymbol{\theta}_i|\boldsymbol{\varphi}_2, \boldsymbol{\theta}, \Omega, t_i, \delta_i, \boldsymbol{z}_i) = \frac{f(t_i|\boldsymbol{\varphi}_2, \boldsymbol{\theta}_i, \delta_i, \boldsymbol{z}_i) f(\boldsymbol{\theta}_i|\boldsymbol{\theta}, \Omega)}{\int f(t_i|\boldsymbol{\varphi}_2, \boldsymbol{\theta}_i, \delta_i, \boldsymbol{z}_i) f(\boldsymbol{\theta}_i|\boldsymbol{\theta}, \Omega) d\boldsymbol{\theta}_i}.$$

We are also interested in how much the survival data can contribute to the fit of the longitudinal component in the joint model. In the same fashion as $\Delta\text{DIC}_{\text{Surv}}$, $\Delta\text{DIC}_{\text{Long}}$ is given by

$$\Delta\text{DIC}_{\text{Long}} = \text{DIC}_{\text{Long,alone}} - \text{DIC}_{\text{Long|Surv}}. \tag{A.3}$$

where $\text{DIC}_{\text{Long,alone}} = \text{Dev}_{\text{Long,alone}}(\bar{\boldsymbol{\varphi}}_1) + 2p_{D[\text{Long,alone}]}$, $\text{Dev}_{\text{Long,alone}}(\bar{\boldsymbol{\varphi}}_1) = -2\sum_{i=1}^n \log \int f(\boldsymbol{y}_i|\bar{\boldsymbol{\psi}}_1, \boldsymbol{\theta}_i, \boldsymbol{x}_i) f(\boldsymbol{\theta}_i|\bar{\boldsymbol{\theta}}, \bar{\Omega}) d\boldsymbol{\theta}_i$, and $p_{D[\text{Long,alone}]} = E\big[-2\sum_{i=1}^n \log \int f(\boldsymbol{y}_i|\boldsymbol{\psi}_1, \boldsymbol{\theta}_i, \boldsymbol{x}_i) f(\boldsymbol{\theta}_i|\boldsymbol{\theta}, \Omega) d\boldsymbol{\theta}_i \Big| D_{\text{Long,obs}}\big] + 2\sum_{i=1}^n \log \int f(\boldsymbol{y}_i|\bar{\boldsymbol{\psi}}_1, \boldsymbol{\theta}_i, \boldsymbol{x}_i) f(\boldsymbol{\theta}_i|\bar{\boldsymbol{\theta}}, \bar{\Omega}) d\boldsymbol{\theta}_i$.

$\Delta\text{DIC}_{\text{Long}}$ in (A.3) quantifies the trade-off between the fitting improvement in the longitudinal component due to the survival data and the dimension penalty for the additional parameters in the longitudinal component of the joint model. A model with a large value of $\Delta\text{DIC}_{\text{Long}}$ is more preferred.

## Appendix B.2: CPO Decomposition II, LPML Decomposition II, and $\Delta\text{LPML}_{\text{Long}}$

We have already shown that (2.18) holds for any fixed value $\boldsymbol{\varphi}^*$. Let $\boldsymbol{\psi}_1^*$ and $\boldsymbol{\psi}_2^*$ denote the posterior means of $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$. Given (A.2), we have

$$f(\boldsymbol{y}_i, t_i|\boldsymbol{\varphi}^*, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i) = f(t_i|\boldsymbol{\psi}_2^*, \delta_i, \boldsymbol{z}_i) f(\boldsymbol{y}_i|\boldsymbol{\psi}_1^*, \boldsymbol{\psi}_2^*, t_i, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i), \tag{A.4}$$

where $f(y_i|\boldsymbol{\psi}_1^*, \boldsymbol{\psi}_2^*, t_i, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i) = \int f(\boldsymbol{y}_i|\boldsymbol{\psi}_1^*, \boldsymbol{\theta}_i, \boldsymbol{x}_i) f(\boldsymbol{\theta}_i|\boldsymbol{\psi}_2^*, t_i, \delta_i, \boldsymbol{z}_i) d\boldsymbol{\theta}_i$. Similar to (2.19),

$$\begin{aligned} \pi(\boldsymbol{\varphi}^*|D_{\text{obs}}^{(-i)}) &= \pi(\boldsymbol{\psi}_2^*|D_{\text{obs}}^{(-i)}) \pi(\boldsymbol{\psi}_1^*|\boldsymbol{\psi}_2^*, D_{\text{obs}}^{(-i)}), \\ \pi(\boldsymbol{\varphi}^*|D_{\text{obs}}) &= \pi(\boldsymbol{\psi}_2^*|D_{\text{obs}}) \pi(\boldsymbol{\psi}_1^*|\boldsymbol{\psi}_2^*, D_{\text{obs}}). \end{aligned} \tag{A.5}$$

Using (A.4) and (A.5), we propose the CPO Decomposition II as

$$\mathrm{CPO}_i = \mathrm{CPO}_{i,\mathrm{Surv}} \cdot \mathrm{CPO}_{i,\mathrm{Long|Surv}}, \tag{A.6}$$

where

$$\mathrm{CPO}_{i,\mathrm{Surv}} = \frac{f(t_i|\boldsymbol{\psi}_2^*, \delta_i, \boldsymbol{z}_i)\pi(\boldsymbol{\psi}_2^*|D_{\mathrm{obs}}^{(-i)})}{\pi(\boldsymbol{\psi}_2^*|D_{\mathrm{obs}})}, \tag{A.7}$$

and

$$\mathrm{CPO}_{i,\mathrm{Long|Surv}} = \frac{f(\boldsymbol{y}_i|\boldsymbol{\psi}_1^*, \boldsymbol{\psi}_2^*, t_i, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)\pi(\boldsymbol{\psi}_1^*|\boldsymbol{\psi}_2^*, D_{\mathrm{obs}}^{(-i)})}{\pi(\boldsymbol{\psi}_1^*|\boldsymbol{\psi}_2^*, D_{\mathrm{obs}})}. \tag{A.8}$$

**Result B.2:** *For CPO$_i$, CPO$_{i,\mathrm{Surv}}$, and CPO$_{i,\mathrm{Long|Surv}}$, we have the identities*

$$\frac{\pi(\boldsymbol{\psi}_2^*|D_{\mathrm{obs}}^{(-i)})}{\pi(\boldsymbol{\psi}_2^*|D_{\mathrm{obs}})} = \mathrm{CPO}_i \int \frac{1}{f(\boldsymbol{y}_i, t_i|\boldsymbol{\psi}_1, \boldsymbol{\psi}_2^*, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \pi(\boldsymbol{\psi}_1|\boldsymbol{\psi}_2^*, D_{\mathrm{obs}})d\boldsymbol{\psi}_1,$$

$$\mathrm{CPO}_{i,\mathrm{Surv}} = \mathrm{CPO}_i \int \frac{1}{f(\boldsymbol{y}_i|\boldsymbol{\psi}_1, \boldsymbol{\psi}_2^*, t_i, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \pi(\boldsymbol{\psi}_1|\boldsymbol{\psi}_2^*, D_{\mathrm{obs}})d\boldsymbol{\psi}_1,$$

*and*

$$\mathrm{CPO}_{i,\mathrm{Long|Surv}} = \frac{1}{\int \frac{1}{f(\boldsymbol{y}_i|\boldsymbol{\psi}_1, \boldsymbol{\psi}_2^*, t_i, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \pi(\boldsymbol{\psi}_1|\boldsymbol{\psi}_2^*, D_{\mathrm{obs}})d\boldsymbol{\psi}_1}. \tag{A.9}$$

Using (A.4), (A.9) can be rewritten as

$$\mathrm{CPO}_{i,\mathrm{Long|Surv}} = \frac{1}{f(t_i|\boldsymbol{\psi}_2^*, \delta_i, \boldsymbol{z}_i) \int \frac{1}{f(\boldsymbol{y}_i, t_i|\boldsymbol{\psi}_1, \boldsymbol{\psi}_2^*, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \pi(\boldsymbol{\psi}_1|\boldsymbol{\psi}_2^*, D_{\mathrm{obs}})d\boldsymbol{\psi}_1}.$$

To avoid the calculation of $f(\boldsymbol{y}_i, t_i|\boldsymbol{\psi}_1, \boldsymbol{\psi}_2^*, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)$, we use the same idea as in (2.16) and obtain

$$\frac{\pi(\boldsymbol{\psi}_2^*|D_{\mathrm{obs}}^{(-i)})}{\pi(\boldsymbol{\psi}_2^*|D_{\mathrm{obs}})} = \mathrm{CPO}_i \int \frac{w_i(\boldsymbol{\theta}_i)}{f(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i|\boldsymbol{\psi}_1, \boldsymbol{\psi}_2^*, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \pi(\boldsymbol{\psi}_1, \boldsymbol{\theta}^R|\boldsymbol{\psi}_2^*, D_{\mathrm{obs}})d\boldsymbol{\theta}^R d\boldsymbol{\psi}_1,$$

where the optimal choice of $w_i(\boldsymbol{\theta}_i)$ is $\frac{f(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i|\boldsymbol{\psi}_1, \boldsymbol{\psi}_2^*, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)}{f(\boldsymbol{y}_i, t_i|\boldsymbol{\psi}_1, \boldsymbol{\psi}_2^*, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)}$. Similarly,

$$\mathrm{CPO}_{i,\mathrm{Long|Surv}} = \frac{1}{f(t_i|\boldsymbol{\psi}_2^*, \delta_i, \boldsymbol{z}_i) \int \frac{w_i(\boldsymbol{\theta}_i)}{f(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i|\boldsymbol{\psi}_1, \boldsymbol{\psi}_2^*, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \pi(\boldsymbol{\psi}_1, \boldsymbol{\theta}^R|\boldsymbol{\psi}_2^*, D_{\mathrm{obs}})d\boldsymbol{\theta}^R d\boldsymbol{\psi}_1},$$

where the optimal choice of $w_i(\boldsymbol{\theta}_i)$ is $\frac{f(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i|\boldsymbol{\psi}_1, \boldsymbol{\psi}_2^*, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)}{f(\boldsymbol{y}_i, t_i|\boldsymbol{\psi}_1, \boldsymbol{\psi}_2^*, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)}$.

**Result B.3:** *LPML can be decomposed as*

$$\mathrm{LPML} = \mathrm{LPML}_{\mathrm{Surv}} + \mathrm{LPML}_{\mathrm{Long|Surv}}, \tag{A.10}$$

*where* $\text{LPML}_{\text{Surv}} = \sum_{i=1}^n \log \text{CPO}_{i,\text{Surv}}$, $\text{LPML}_{\text{Long}|\text{Surv}} = \sum_{i=1}^n \log \text{CPO}_{i,\text{Long}|\text{Surv}}$, *with* $CPO_{i,\text{Surv}}$ *defined in* (A.7) *and* $CPO_{i,\text{Long}|\text{Surv}}$ *defined in* (A.8), *respectively.*

Define $\text{LPML}_{\text{Long,alone}} = \sum_{i=1}^n \log(\text{CPO}_{i,\text{Long alone}})$ where $\text{CPO}_{i,\text{Long alone}}$ is given by (2.23). We propose the following model assessment criterion:

$$\Delta\text{LPML}_{\text{Long}} = \text{LPML}_{\text{Long}|\text{Surv}} - \text{LPML}_{\text{Long,alone}}.$$

$\Delta\text{LPML}_{\text{Long}}$ quantifies the gain of the fit in the longitudinal component due to the survival data with a penalty for the additional parameters in the longitudinal component of the joint model. A model with a large value of $\Delta\text{LPML}_{\text{Long}}$ is more preferred.

# Appendix C: Proofs of Identities, Results, and Theorems

**Proof of Result 1.** We first observe that

$$f(\boldsymbol{y}_i|\gamma, \sigma^2, \boldsymbol{\theta}_i, \boldsymbol{x}_i) f(\boldsymbol{\theta}_i|\boldsymbol{\theta}, \Omega) = f(\boldsymbol{\theta}_i|\boldsymbol{\varphi}_1, \boldsymbol{y}_i, \boldsymbol{x}_i) f(\boldsymbol{y}_i|\boldsymbol{\varphi}_1, \boldsymbol{x}_i). \tag{C.1}$$

Using (2.5) and (C.1), we have

$$\begin{aligned} L(\boldsymbol{\varphi}|D_{\text{obs}}) &= \prod_{i=1}^n \int f(t_i|\boldsymbol{\varphi}_2, \boldsymbol{\theta}_i, \delta_i, \boldsymbol{z}_i) f(\boldsymbol{y}_i|\gamma, \sigma^2, \boldsymbol{\theta}_i, \boldsymbol{x}_i) f(\boldsymbol{\theta}_i|\boldsymbol{\theta}, \Omega) d\boldsymbol{\theta}_i \\ &= \prod_{i=1}^n \int f(t_i|\boldsymbol{\varphi}_2, \boldsymbol{\theta}_i, \delta_i, \boldsymbol{z}_i) f(\boldsymbol{\theta}_i|\boldsymbol{\varphi}_1, \boldsymbol{y}_i, \boldsymbol{x}_i) f(\boldsymbol{y}_i|\boldsymbol{\varphi}_1, \boldsymbol{x}_i) d\boldsymbol{\theta}_i \\ &= \prod_{i=1}^n f(\boldsymbol{y}_i|\boldsymbol{\varphi}_1, \boldsymbol{x}_i) \prod_{i=1}^n \int f(t_i|\boldsymbol{\varphi}_2, \boldsymbol{\theta}_i, \delta_i, \boldsymbol{z}_i) f(\boldsymbol{\theta}_i|\boldsymbol{\varphi}_1, \boldsymbol{y}_i, \boldsymbol{x}_i) d\boldsymbol{\theta}_i. \end{aligned} \tag{C.2}$$

Taking -2log of (C.2) yields

$$-2\log L(\boldsymbol{\varphi}|D_{\text{obs}}) = -2\sum_{i=1}^n \log f(\boldsymbol{y}_i|\boldsymbol{\varphi}_1, \boldsymbol{x}_i) - 2\sum_{i=1}^n \log \int f(t_i|\boldsymbol{\varphi}_2, \boldsymbol{\theta}_i, \delta_i, \boldsymbol{z}_i) f(\boldsymbol{\theta}_i|\boldsymbol{\varphi}_1, \boldsymbol{y}_i, \boldsymbol{x}_i) d\boldsymbol{\theta}_i. \tag{C.3}$$

Taking the expectation of (C.3) with respect to the posterior distribution of $\boldsymbol{\varphi}$, we obtain

$$\begin{aligned} & E\left[-2\log L(\boldsymbol{\varphi}|D_{\text{obs}})\middle|D_{\text{obs}}\right] \\ &= E\left[-2\sum_{i=1}^n \log f(\boldsymbol{y}_i|\boldsymbol{\varphi}_1, \boldsymbol{x}_i) - 2\sum_{i=1}^n \log \int f(t_i|\boldsymbol{\varphi}_2, \boldsymbol{\theta}_i, \delta_i, \boldsymbol{z}_i) f(\boldsymbol{\theta}_i|\boldsymbol{\varphi}_1, \boldsymbol{y}_i, \boldsymbol{x}_i) d\boldsymbol{\theta}_i \middle| D_{\text{obs}}\right] \\ &= E\left[-2\sum_{i=1}^n \log f(\boldsymbol{y}_i|\boldsymbol{\varphi}_1, \boldsymbol{x}_i)\middle|D_{\text{obs}}\right] + E\left[-2\sum_{i=1}^n \log \int f(t_i|\boldsymbol{\varphi}_2, \boldsymbol{\theta}_i, \delta_i, \boldsymbol{z}_i) f(\boldsymbol{\theta}_i|\boldsymbol{\varphi}_1, \boldsymbol{y}_i, \boldsymbol{x}_i) d\boldsymbol{\theta}_i \middle| D_{\text{obs}}\right]. \end{aligned} \tag{C.4}$$

The DIC decomposition in (2.10) directly follows from (2.9), (C.3), and (C.4), which completes the proof.

**Proof of CPO Identity II.** Using (2.4) and (2.8), we have

$$\begin{aligned}
\text{CPO}_i &= \frac{\int f(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i \mid \boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i) \prod_{j\neq i} f(\boldsymbol{y}_j, t_j, \boldsymbol{\theta}_j \mid \boldsymbol{\varphi}, \delta_j, \boldsymbol{x}_j, \boldsymbol{z}_j)\pi(\boldsymbol{\varphi}) d\boldsymbol{\theta}^R d\boldsymbol{\varphi}}{\int \prod_{j\neq i} f(\boldsymbol{y}_j, t_j, \boldsymbol{\theta}_j \mid \boldsymbol{\varphi}, \delta_j, \boldsymbol{x}_j, \boldsymbol{z}_j)\pi(\boldsymbol{\varphi}) d\boldsymbol{\theta}^R_{(-i)} d\boldsymbol{\varphi}} \\
&= \frac{c(D_{\text{obs}})}{\int w_i(\boldsymbol{\theta}_i) \prod_{j\neq i} f(\boldsymbol{y}_j, t_j, \boldsymbol{\theta}_j \mid \boldsymbol{\varphi}, \delta_j, \boldsymbol{x}_j, \boldsymbol{z}_j)\pi(\boldsymbol{\varphi}) d\boldsymbol{\theta}_i d\boldsymbol{\theta}^R_{(-i)} d\boldsymbol{\varphi}} \\
&= \frac{1}{\int \frac{w_i(\boldsymbol{\theta}_i)}{f(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i | \boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \frac{\prod_{j=1}^n f(\boldsymbol{y}_j, t_j, \boldsymbol{\theta}_j | \boldsymbol{\varphi}, \delta_j, \boldsymbol{x}_j, \boldsymbol{z}_j)\pi(\boldsymbol{\varphi})}{c(D_{\text{obs}})} d\boldsymbol{\theta}_i d\boldsymbol{\theta}^R_{(-i)} d\boldsymbol{\varphi}} \\
&= \frac{1}{\int \frac{w_i(\boldsymbol{\theta}_i)}{f(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i | \boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_j)} \pi(\boldsymbol{\varphi}, \boldsymbol{\theta}^R \mid D_{\text{obs}}) d\boldsymbol{\theta}^R d\boldsymbol{\varphi}}.
\end{aligned}$$

**Proof of Theorem 1.** The proof of this theorem is similar to that of Theorem 2.1 in Chen (1994). The detail of the proof is given as follows. If $\text{Var}\Big(\frac{w_i(\boldsymbol{\theta}_i)}{f(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i | \boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)}\Big| D_{\text{obs}}\Big) = \infty$, then the inequality holds automatically. Now, we assume that $\text{Var}\Big(\frac{w_i(\boldsymbol{\theta}_i)}{f(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i | \boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)}\Big| D_{\text{obs}}\Big) < \infty$. We first see that from (2.14),

$$E\Big[\frac{1}{f(\boldsymbol{y}_i, t_i | \boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)}\Big| D_{\text{obs}}\Big] = \text{CPO}_i^{-1},$$

and from (2.16),

$$E\Big[\frac{w_i(\boldsymbol{\theta}_i)}{f(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i | \boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)}\Big| D_{\text{obs}}\Big] = \text{CPO}_i^{-1}.$$

Hence, it is sufficient to show that

$$\int \frac{\pi(\boldsymbol{\varphi} | D_{\text{obs}})}{f^2(\boldsymbol{y}_i, t_i | \boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} d\boldsymbol{\varphi} \leq \int \frac{w_i^2(\boldsymbol{\theta}_i)}{f^2(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i | \boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \pi(\boldsymbol{\varphi}, \boldsymbol{\theta}^R | D_{\text{obs}}) d\boldsymbol{\theta}^R d\boldsymbol{\varphi}. \tag{C.5}$$

Observe that

$$\begin{aligned}
1 &= \Big(\int w_i(\boldsymbol{\theta}_i) d\theta_i\Big)^2 \\
&= \Bigg(\int \frac{w_i(\boldsymbol{\theta}_i)}{\sqrt{f(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i | \boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)}} \sqrt{f(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i | \boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} d\boldsymbol{\theta}_i\Bigg)^2.
\end{aligned}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\Bigg(\int \frac{w_i(\boldsymbol{\theta}_i)}{\sqrt{f(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i | \boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)}} \sqrt{f(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i | \boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} d\theta_i\Bigg)^2 \\
&\leq \int \frac{w_i^2(\boldsymbol{\theta}_i)}{f(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i | \boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} d\boldsymbol{\theta}_i \int f(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i | \boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i) d\boldsymbol{\theta}_i \\
&= \int \frac{w_i^2(\boldsymbol{\theta}_i)}{f(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i | \boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} d\boldsymbol{\theta}_i f(\boldsymbol{y}_i, t_i | \boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i). \qquad \text{(C.6)}
\end{aligned}$$

Moving $f(\boldsymbol{y}_i, t_i|\boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)$ to the left side of (C.6) yields

$$\frac{1}{f(\boldsymbol{y}_i, t_i|\boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \le \int \frac{w_i^2(\boldsymbol{\theta}_i)}{f(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i|\boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} d\boldsymbol{\theta}_i. \tag{C.7}$$

Multiplying $\pi(\boldsymbol{\varphi}, \boldsymbol{\theta}^R_{(-i)}|D^{(-i)}_{\text{obs}})$ to both sides of (C.7) and integrating over $\boldsymbol{\theta}^R_{(-i)}$ and $\boldsymbol{\varphi}$ gives

$$\int \frac{\pi(\boldsymbol{\varphi}, \boldsymbol{\theta}^R_{(-i)}|D^{(-i)}_{\text{obs}})}{f(\boldsymbol{y}_i, t_i|\boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} d\boldsymbol{\theta}^R_{(-i)} d\boldsymbol{\varphi} \le \int \frac{w_i^2(\boldsymbol{\theta}_i)}{f(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i|\boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \pi(\boldsymbol{\varphi}, \boldsymbol{\theta}^R_{(-i)}|D^{(-i)}_{\text{obs}}) d\boldsymbol{\theta}_i d\boldsymbol{\theta}^R_{(-i)} d\boldsymbol{\varphi}. \tag{C.8}$$

After some algebra, we have

$$\begin{aligned} \text{The left side of (C.8)} &= \int \frac{f(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i|\boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)\pi(\boldsymbol{\varphi}, \boldsymbol{\theta}^R_{(-i)}|D^{(-i)}_{\text{obs}})}{f^2(\boldsymbol{y}_i, t_i|\boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} d\boldsymbol{\theta}_i d\boldsymbol{\theta}^R_{(-i)} d\boldsymbol{\varphi} \\ &= \int \frac{\pi(\boldsymbol{\varphi}, \boldsymbol{\theta}^R|D_{\text{obs}})}{f^2(\boldsymbol{y}_i, t_i|\boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \frac{c(D_{\text{obs}})}{c(D^{(-i)}_{\text{obs}})} d\boldsymbol{\theta}^R d\boldsymbol{\varphi} \\ &= \int \frac{\pi(\boldsymbol{\varphi}|D_{\text{obs}})}{f^2(\boldsymbol{y}_i, t_i|\boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \frac{c(D_{\text{obs}})}{c(D^{(-i)}_{\text{obs}})} d\boldsymbol{\varphi} \end{aligned} \tag{C.9}$$

and

$$\text{The right side of (C.8)} = \int \frac{w_i^2(\boldsymbol{\theta}_i)}{f^2(\boldsymbol{y}_i, t_i, \boldsymbol{\theta}_i|\boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \pi(\boldsymbol{\varphi}, \boldsymbol{\theta}^R|D_{\text{obs}}) \frac{c(D_{\text{obs}})}{c(D^{(-i)}_{\text{obs}})} d\boldsymbol{\theta}^R d\boldsymbol{\varphi}. \tag{C.10}$$

Canceling the common constant $\frac{c(D_{\text{obs}})}{c(D^{(-i)}_{\text{obs}})}$ from (C.9) and (C.10), we obtain (C.5), which completes the proof.

**Proof of CPO Identity III.** Plugging (2.13) into (2.12) yields

$$\text{CPO}_i = \frac{c(D_{\text{obs}})}{c(D^{(-i)}_{\text{obs}})}. \tag{C.11}$$

We define

$$\pi(\boldsymbol{\varphi} \mid D^{(i)}_{\text{obs}}, \pi(\boldsymbol{\varphi} \mid D^{(-i)}_{\text{obs}})) = \frac{f(\boldsymbol{y}_i, t_i \mid \boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)\pi(\boldsymbol{\varphi} \mid D^{(-i)}_{\text{obs}})}{\text{CPO}_i}. \tag{C.12}$$

where $D^{(i)}_{\text{obs}} = \{(\boldsymbol{y}_i, t_i, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)\}$ denotes the observed data for the $i^{th}$ subject. Then, $\pi(\boldsymbol{\varphi} \mid D^{(i)}_{\text{obs}}, \pi(\boldsymbol{\varphi} \mid D^{(-i)}_{\text{obs}}))$ can be viewed as the posterior distribution of $\boldsymbol{\varphi}$ given the data $D^{(i)}_{\text{obs}}$ with $\pi(\boldsymbol{\varphi} \mid D^{(-i)}_{\text{obs}})$ as the prior. Clearly, $\text{CPO}_i$ is the normalizing constant of $\pi(\boldsymbol{\varphi} \mid D^{(i)}_{\text{obs}}, \pi(\boldsymbol{\varphi} \mid D^{(-i)}_{\text{obs}}))$ in (C.12). Using (2.6), (2.13), (C.11), and (C.12), we obtain

$$\begin{aligned} \pi(\boldsymbol{\varphi} \mid D^{(i)}_{\text{obs}}, \pi(\boldsymbol{\varphi} \mid D^{(-i)}_{\text{obs}})) &= \frac{c(D^{(-i)}_{\text{obs}})}{c(D_{\text{obs}})} \times f(\boldsymbol{y}_i, t_i \mid \boldsymbol{\varphi}, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i) \times \frac{\prod_{j \neq i} f(\boldsymbol{y}_j, t_j \mid \boldsymbol{\varphi}, \delta_j, \boldsymbol{x}_j, \boldsymbol{z}_j)\pi(\boldsymbol{\varphi})}{c(D^{(-i)}_{\text{obs}})} \\ &= \frac{\prod_{j=1}^n f(\boldsymbol{y}_j, t_j \mid \boldsymbol{\varphi}, \delta_j, \boldsymbol{x}_j, \boldsymbol{z}_j)\pi(\boldsymbol{\varphi})}{c(D_{\text{obs}})} = \pi(\boldsymbol{\varphi} \mid D_{\text{obs}}). \end{aligned} \tag{C.13}$$

Combining (C.12) and (C.13) gives (2.17), which completes the proof.

**Proof of Theorem 2.** Using augmented posteriors,

$$\begin{aligned}\frac{\pi(\boldsymbol{\varphi}_1^*|D_{\text{obs}}^{(-i)})}{\pi(\boldsymbol{\varphi}_1^*|D_{\text{obs}})} &= \frac{\int \pi(\boldsymbol{\varphi}_1^*, \boldsymbol{\varphi}_2|D_{\text{obs}}^{(-i)})d\boldsymbol{\varphi}_2}{\int \pi(\boldsymbol{\varphi}_1^*, \boldsymbol{\varphi}_2|D_{\text{obs}})d\boldsymbol{\varphi}_2} \\ &= \frac{\int \frac{\prod_{j\neq i} f(\boldsymbol{y}_j, t_j|\boldsymbol{\varphi}_1^*, \boldsymbol{\varphi}_2, \delta_j, \boldsymbol{x}_j, \boldsymbol{z}_j)\pi(\boldsymbol{\varphi}_1^*, \boldsymbol{\varphi}_2)}{c(D_{\text{obs}}^{(-i)})} d\boldsymbol{\varphi}_2}{\int \frac{\prod_{j=1} f(\boldsymbol{y}_j, t_j|\boldsymbol{\varphi}_1^*, \boldsymbol{\varphi}_2, \delta_j, \boldsymbol{x}_j, \boldsymbol{z}_j)\pi(\boldsymbol{\varphi}_1^*, \boldsymbol{\varphi}_2)}{c(D_{\text{obs}})} d\boldsymbol{\varphi}_2} \\ &= \text{CPO}_i \frac{\int \frac{1}{f(\boldsymbol{y}_i, t_i|\boldsymbol{\varphi}_1^*, \boldsymbol{\varphi}_2, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \prod_{j=1} f(\boldsymbol{y}_j, t_j|\boldsymbol{\varphi}_1^*, \boldsymbol{\varphi}_2, \delta_j, \boldsymbol{x}_j, \boldsymbol{z}_j)\pi(\boldsymbol{\varphi}_1^*, \boldsymbol{\varphi}_2)d\boldsymbol{\varphi}_2}{\int \prod_{j=1} f(\boldsymbol{y}_j, t_j|\boldsymbol{\varphi}_1^*, \boldsymbol{\varphi}_2, \delta_j, \boldsymbol{x}_j, \boldsymbol{z}_j)\pi(\boldsymbol{\varphi}_1^*, \boldsymbol{\varphi}_2)d\boldsymbol{\varphi}_2} \\ &= \text{CPO}_i \int \frac{1}{f(\boldsymbol{y}_i, t_i|\boldsymbol{\varphi}_1^*, \boldsymbol{\varphi}_2, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \pi(\boldsymbol{\varphi}_2|\boldsymbol{\varphi}_1^*, D_{\text{obs}})d\boldsymbol{\varphi}_2. \end{aligned} \tag{C.14}$$

Using (2.21) and (C.14), we have

$$\begin{aligned}\text{CPO}_{i,\text{Long}} &= \text{CPO}_i f(\boldsymbol{y}_i|\boldsymbol{\varphi}_1^*, \boldsymbol{x}_i) \int \frac{1}{f(\boldsymbol{y}_i, t_i|\boldsymbol{\varphi}_1^*, \boldsymbol{\varphi}_2, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \pi(\boldsymbol{\varphi}_2|\boldsymbol{\varphi}_1^*, D_{\text{obs}})d\boldsymbol{\varphi}_2 \\ &= \text{CPO}_i \int \frac{f(\boldsymbol{y}_i|\boldsymbol{\varphi}_1^*, \boldsymbol{x}_i)}{f(\boldsymbol{y}_i|\boldsymbol{\varphi}_1^*, \boldsymbol{x}_i) f(t_i|\boldsymbol{\varphi}_2, \boldsymbol{\varphi}_1^*, \delta_i, \boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \pi(\boldsymbol{\varphi}_2|\boldsymbol{\varphi}_1^*, D_{\text{obs}})d\boldsymbol{\varphi}_2 \\ &= \text{CPO}_i \int \frac{1}{f(t_i|\boldsymbol{\varphi}_2, \boldsymbol{\varphi}_1^*, \delta_i, \boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \pi(\boldsymbol{\varphi}_2|\boldsymbol{\varphi}_1^*, D_{\text{obs}})d\boldsymbol{\varphi}_2. \end{aligned}$$

Thus, (2.26) can be immediately obtained based on (2.20) and (2.25). We can also prove (2.26) directly from (2.22). Using (2.22),

$$\begin{aligned}&\text{CPO}_{i,\text{Surv}|\text{Long}} \\ =&\frac{f(t_i|\boldsymbol{\varphi}_2^*, \boldsymbol{\varphi}_1^*, \delta_i, \boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i) \frac{\pi(\boldsymbol{\varphi}^*|D_{\text{obs}}^{(-i)})}{\pi(\boldsymbol{\varphi}_1^*|D_{\text{obs}}^{(-i)})}}{\frac{\pi(\boldsymbol{\varphi}^*|D_{\text{obs}})}{\pi(\boldsymbol{\varphi}_1^*|D_{\text{obs}})}} \\ =&f(t_i|\boldsymbol{\varphi}_2^*, \boldsymbol{\varphi}_1^*, \delta_i, \boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i)\Big[\frac{\pi(\boldsymbol{\varphi}^*|D_{\text{obs}}^{(-i)})}{\pi(\boldsymbol{\varphi}^*|D_{\text{obs}})}\Big]\Big[\frac{\pi(\boldsymbol{\varphi}_1^*|D_{\text{obs}})}{\pi(\boldsymbol{\varphi}_1^*|D_{\text{obs}})}\Big] \\ =&f(t_i|\boldsymbol{\varphi}^*, \delta_i, \boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i) \cdot \frac{\text{CPO}_i}{f(\boldsymbol{y}_i, t_i|\boldsymbol{\varphi}^*, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \cdot \frac{1}{\text{CPO}_i \int \frac{1}{f(\boldsymbol{y}_i, t_i|\boldsymbol{\varphi}_2, \boldsymbol{\varphi}_1^*, \delta_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \pi(\boldsymbol{\varphi}_2|\boldsymbol{\varphi}_1^*, D_{\text{obs}})d\boldsymbol{\varphi}_2} \\ =&\frac{f(t_i|\boldsymbol{\varphi}^*, \delta_i, \boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i)}{f(\boldsymbol{y}_i|\boldsymbol{\varphi}_1^*, \boldsymbol{x}_i) f(t_i|\boldsymbol{\varphi}^*, \delta_i, \boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \cdot \frac{1}{\int \frac{1}{f(\boldsymbol{y}_i|\boldsymbol{\varphi}_1^*, \boldsymbol{x}_i) f(t_i|\boldsymbol{\varphi}_2, \boldsymbol{\varphi}_1^*, \delta_i, \boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \pi(\boldsymbol{\varphi}_2|\boldsymbol{\varphi}_1^*, D_{\text{obs}})d\boldsymbol{\varphi}_2} \\ =&\frac{1}{\int \frac{1}{f(t_i|\boldsymbol{\varphi}_2, \boldsymbol{\varphi}_1^*, \delta_i, \boldsymbol{y}_i, \boldsymbol{x}_i, \boldsymbol{z}_i)} \pi(\boldsymbol{\varphi}_2|\boldsymbol{\varphi}_1^*, D_{\text{obs}})d\boldsymbol{\varphi}_2}. \end{aligned}$$

**Proof of Result 2.** By the CPO decomposition in (2.20), we obtain

$$
\begin{aligned}
\text{LPML} &= \sum_{i=1}^{n} \log(\text{CPO}_i) \\
&= \sum_{i=1}^{n} \log(\text{CPO}_{i,\text{Long}} \cdot \text{CPO}_{i,\text{Surv}|\text{Long}}) \\
&= \sum_{i=1}^{n} \log(\text{CPO}_{i,\text{Long}}) + \sum_{i=1}^{n} \log(\text{CPO}_{i,\text{Surv}|\text{Long}}) \\
&= \text{LPML}_{\text{Long}} + \text{LPML}_{\text{Surv}|\text{Long}}.
\end{aligned}
$$

# Appendix D: Additional Tables

Table S1: DIC, $p_D$, and LPML for fitting survival alone with different $K$

| $K$ | DIC | $p_D$ | LPML |
|---|---|---|---|
| 2 | 2206.05 | 8.86 | -1103.10 |
| 10 | 2049.45 | 17.00 | -1024.68 |
| 20 | 2022.76 | 27.18 | -1012.13 |
| 25 | 2026.85 | 32.25 | -1014.27 |
| 29 | 2031.12 | 36.43 | -1016.32 |
| 30 | 2018.49 | 37.36 | -1010.07 |
| 31 | 2020.72 | 38.47 | -1011.34 |
| 35 | 2022.56 | 42.76 | -1012.62 |
| 50 | 2040.11 | 58.21 | -1022.65 |
| 75 | 2048.33 | 85.52 | -1037.82 |
| 100 | 2070.61 | 113.64 | -1070.94 |

Table S2: $p_D$'s and $p_{D[\text{Surv}|\text{Long}]}$'s for five PROs under SPML and TML with different $K$

| $K$ | Model | | Anorexia | Cough | Dyspnea | Fatigue | Pain |
|---|---|---|---|---|---|---|---|
| 25 | SPML | $p_D$ | 47.16 | 47.30 | 47.70 | 47.23 | 47.28 |
| | | $p_{D[\text{Surv}|\text{Long}]}$ | 34.34 | 34.26 | 34.42 | 34.41 | 34.51 |
| | TML | $p_D$ | 46.20 | 46.52 | 46.23 | 46.39 | 46.48 |
| | | $p_{D[\text{Surv}|\text{Long}]}$ | 33.23 | 33.40 | 33.32 | 33.25 | 33.37 |
| 30 | SPML | $p_D$ | 52.45 | 52.27 | 52.55 | 52.16 | 52.36 |
| | | $p_{D[\text{Surv}|\text{Long}]}$ | 39.49 | 39.37 | 39.51 | 39.51 | 39.60 |
| | TML | $p_D$ | 51.44 | 51.32 | 51.32 | 51.38 | 51.43 |
| | | $p_{D[\text{Surv}|\text{Long}]}$ | 38.48 | 38.35 | 38.28 | 38.41 | 38.52 |
| 35 | SPML | $p_D$ | 57.88 | 57.69 | 57.12 | 57.64 | 57.73 |
| | | $p_{D[\text{Surv}|\text{Long}]}$ | 44.75 | 44.62 | 44.57 | 44.69 | 44.80 |
| | TML | $p_D$ | 56.42 | 56.84 | 56.70 | 56.58 | 56.52 |
| | | $p_{D[\text{Surv}|\text{Long}]}$ | 43.52 | 43.66 | 43.64 | 43.68 | 43.75 |

Table S3: The Decomposition of LPML for five PROs under SPML and TML with different $K$ using Gaussian quadrature

| $K$ | Model | | Anorexia | Cough | Dyspnea | Fatigue | Pain |
|---|---|---|---|---|---|---|---|
| 25 | SPML | LPML | -7015.03 | -7145.63 | -5965.59 | -6504.33 | -6428.84 |
| | | $\text{LPML}_{\text{Surv}\mid\text{Long}}$ | -1003.19 | -1012.16 | -1004.83 | -998.64 | -988.22 |
| | | $\Delta\text{LPML}_{\text{Surv}}$ | 11.08 | 2.11 | 9.44 | 15.63 | 26.05 |
| | TML | LPML | -7016.15 | -7144.97 | -5968.85 | -6507.35 | -6436.11 |
| | | $\text{LPML}_{\text{Surv}\mid\text{Long}}$ | -1004.32 | -1011.10 | -1008.79 | -1001.93 | -996.01 |
| | | $\Delta\text{LPML}_{\text{Surv}}$ | 9.95 | 3.17 | 5.47 | 12.33 | 18.25 |
| 30 | SPML | LPML | -7011.19 | -7141.37 | -5960.95 | -6500.15 | -6424.75 |
| | | $\text{LPML}_{\text{Surv}\mid\text{Long}}$ | -999.00 | -1008.05 | -1000.59 | -994.66 | -984.14 |
| | | $\Delta\text{LPML}_{\text{Surv}}$ | 11.07 | 2.01 | 9.48 | 15.41 | 25.93 |
| | TML | LPML | -7012.15 | -7140.22 | -5964.64 | -6503.22 | -6431.93 |
| | | $\text{LPML}_{\text{Surv}\mid\text{Long}}$ | -1000.25 | -1006.80 | -1004.67 | -998.01 | -992.00 |
| | | $\Delta\text{LPML}_{\text{Surv}}$ | 9.82 | 3.27 | 5.40 | 12.06 | 18.07 |
| 35 | SPML | LPML | -7013.76 | -7143.94 | -5962.76 | -6502.81 | -6427.28 |
| | | $\text{LPML}_{\text{Surv}\mid\text{Long}}$ | -1001.52 | -1010.42 | -1002.91 | -997.13 | -986.48 |
| | | $\Delta\text{LPML}_{\text{Surv}}$ | 11.09 | 2.19 | 9.71 | 15.49 | 26.14 |
| | TML | LPML | -7014.32 | -7143.11 | -5967.16 | -6505.62 | -6434.14 |
| | | $\text{LPML}_{\text{Surv}\mid\text{Long}}$ | -1002.68 | -1009.38 | -1007.07 | -1000.39 | -994.41 |
| | | $\Delta\text{LPML}_{\text{Surv}}$ | 9.94 | 3.24 | 5.54 | 12.23 | 18.21 |

## References


Chen, M.-H. (1994). Importance-weighted marginal Bayesian posterior density estimation. *Journal of the American Statistical Association* **89**, 818-824.

Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **41**, 337-348.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97-109.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087-1092.

Zhang, D., Chen, M.-H., Ibrahim, J. G., Boye, M. E., and Shen, W. (2015a). Assessment of fit in longitudinal data for joint models with applications to cancer clinical trials. New York: Springer. In press. In: *Applied Statistics in Biomedicine and Clinical Trials Design - Selected Papers from 2013 ICSA/ISBS Joint Statistical Meetings.* Eds. Z. Chen, A. Liu, Y. Qu, L. Tang, N. Ting, and Y. Tsong. New York: Springer. pp. 347-365.