Combining longitudinal and survival information in Bayesian joint models: When are treatment estimates improved?

Supplementary Materials

LAURA A. HATFIELD

Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA, 02115, USA

hatfield@hcp.med.harvard.edu

JAMES S. HODGES, BRADLEY P. CARLIN

Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455, USA

1. Joint model fixed effect posterior covariance

We begin by recalling the core Bayesian hierarchical modeling result of Lindley and Smith (1972), henceforth abbreviated L&S. For \( n \times p_1 \)-dimensional response vector \( z \), \( p_1 \)-vector of parameters \( \theta \), known \( n \times p_1 \) design matrix \( A_1 \), and known \( n \times n \) covariance matrix \( C_1 \), let the likelihood be \( z \sim N(A_1\theta, C_1) \). Then for second-level \( p_2 \)-vector of parameters \( \mu \), known design and covariance matrices \( A_2 \) and \( C_2 \), let the prior be \( \theta \sim N(A_2\mu, C_2) \). L & S showed that the marginal distribution is \( z \sim N(A_1A_2\mu, C_1 + A_1C_2A_1') \) and the posterior is \( \theta|z \sim N(Dd, D) \) where \( D^{-1} = A_1'C_1^{-1}A_1 + C_2^{-1} \) and \( d = A_1'C_1^{-1}z + C_2^{-1}A_2\mu \). If we assume all covariance parameters (and \( \alpha \)) are known, we

*To whom correspondence should be addressed.
can directly apply these results in our setting.

In what follows, we make the following assumptions without loss of generality: there are equal numbers of subjects in the treatment and control groups (i.e., \( N/2 \) in each); and \( trt_i = 1 \) indicates observations from the treatment group while \( trt_i = -1 \) indicates the control group. Then we collect the data into a single \( 2N \)-vector, where longitudinal data come first, sorted into treatment group then control group outcomes, followed by survival data, similarly sorted: \( z = (z_{11}, \ldots, z_{1N}, z_{21}, \ldots, z_{2N})' \). The complete \( (4+N) \)-vector of parameters \( \theta = (\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}, u)' \) contains both fixed and latent effects. The \( 2N \times (4+N) \) regression design matrix is

\[
A_1 = \begin{pmatrix}
\begin{pmatrix}
1_N & 1_N \\
1_N & -1_N
\end{pmatrix} & 0_N & 0_N & I_N \\
0_N & 0_N & \begin{pmatrix}
1_N & 1_N \\
1_N & -1_N
\end{pmatrix} & \alpha I_N
\end{pmatrix},
\]

(1.1)

where \( 1_K \) and \( 0_K \) are \( K \)-vectors of ones and zeros, respectively. The covariance matrix \( C_1 \) is block diagonal since conditional on \( u_i \), all the responses are independent, thus \( C_1 = \text{Diag} \left( \frac{\sigma^2_{\beta_1}}{\pi} 1_N', \frac{\sigma^2_{\beta_2}}{\pi} 1_N' \right) \).

We place independent normal priors on \( \beta_1 \) and \( \beta_2 \) with means \( \mu_1 \) and \( \mu_2 \) and variance matrices \( \sigma^2_{\beta_1} I_2 \) and \( \sigma^2_{\beta_2} I_2 \), respectively. That is, we use the same prior variance, \( \sigma^2_{\beta_1} \), for both the intercept and treatment effect in the longitudinal model, and a separate prior variance, \( \sigma^2_{\beta_2} \), for both parameters of the survival model. We use an independent normal prior distribution on \( u \), centered at \( 0_N \) with variance matrix \( \sigma^2_u I_N \). Then the joint prior on \( \theta \) is \( N \left( (\mu_1, \mu_2, 0_N)', \text{Diag}(\sigma^2_{\beta_1} I_2, \sigma^2_{\beta_2} I_2, \sigma^2_u I_N) \right) \).

Using the L\&S result, the joint posterior precision matrix for \( (\beta_1, \beta_2, u)' \) is

\[
\begin{pmatrix}
\begin{pmatrix}
\frac{N}{\sigma^2} + \frac{1}{\sigma^2_{\beta_1}} \\
\frac{N}{\sigma^2} + \frac{1}{\sigma^2_{\beta_2}}
\end{pmatrix} I_2 & 0 \\
0 & \begin{pmatrix}
\frac{N}{\sigma^2} + \frac{1}{\sigma^2_{\beta_1}} \\
\frac{N}{\sigma^2} + \frac{1}{\sigma^2_{\beta_2}}
\end{pmatrix} I_2
\end{pmatrix}
\begin{pmatrix}
P'_{12u} & \phantom{0}
\end{pmatrix}
\]

(1.2)

where

\[
P'_{12u} = \begin{pmatrix}
\begin{pmatrix}
\frac{N}{\sigma^2} & 1_N \\
1_N & -1_N
\end{pmatrix} & \frac{\alpha}{\sigma^2} \begin{pmatrix}
1_N & 1_N \\
1_N & -1_N
\end{pmatrix}
\end{pmatrix}
\]

(1.3)
we write this joint posterior precision matrix as

\[ D^{-1} = \begin{pmatrix} P_1 I_2 & 0 & P_{12u} \\ 0 & P_2 I_2 \\ P_{12u} & P_{12u} & P_{12u} \end{pmatrix} \]  \tag{1.4} \]

that is, \( P_1 = \left( \frac{N}{\sigma_1^2/n} + \frac{1}{\sigma_{\beta_1}} \right) \), \( P_2 = \left( \frac{N}{\sigma_2^2} + \frac{1}{\sigma_{\beta_2}} \right) \), and \( P_u = \left( \frac{n}{\sigma_1^2} + \frac{\alpha^2}{\sigma_2^2} + \frac{1}{\sigma_u^2} \right) \). Inverting this, we obtain the posterior variance-covariance matrix, the submatrices of which are described below.

The 4 × 4 posterior variance-covariance matrix of the fixed effects \( \beta \) is

\[
\begin{pmatrix}
\text{Var}(\beta_1) I_2 & \text{Cov}(\beta_1, \beta_2) I_2 \\
\text{Cov}(\beta_1, \beta_2) I_2 & \text{Var}(\beta_2) I_2
\end{pmatrix}
\]  \tag{1.5} \]

where the scalar variances and covariances are given by

\[
\text{Var}(\beta_1) = c_{\beta_1} \left( P_2 - NP_u^{-1} \left( \frac{\alpha}{\sigma_2^2} \right)^2 \right),
\]

\[
\text{Var}(\beta_2) = c_{\beta_2} \left( P_1 - NP_u^{-1} \left( \frac{n}{\sigma_1^2} \right)^2 \right),
\]

\[
\text{Cov}(\beta_1, \beta_2) = c_{\beta_1, \beta_2} \left( NP_u^{-1} \frac{n}{\sigma_1^2} \frac{\alpha}{\sigma_2^2} \right),
\]

with

\[
c_{\beta_1, \beta_2} = \left[ P_1 P_2 - NP_u^{-1} \left( \left( \frac{\alpha}{\sigma_2^2} \right)^2 P_1 + \left( \frac{n}{\sigma_1^2} \right)^2 P_2 \right) \right]^{-1}.
\]

The \( N \times 4 \) covariance matrix between \( u \) and \( \beta \) is given by

\[
\begin{pmatrix}
-Cov(\beta_1, u) I_N & -Cov(\beta_1, u) I_N & -Cov(\beta_2, u) I_N & -Cov(\beta_2, u) I_N \\
-Cov(\beta_1, u) I_N & Cov(\beta_1, u) I_N & -Cov(\beta_2, u) I_N & Cov(\beta_2, u) I_N
\end{pmatrix}
\]  \tag{1.7} \]

where the scalar covariances are given by

\[
\text{Cov}(\beta_1, u) = c_{\beta_1, u} \left( \frac{n}{\sigma_1^2} P_1^{-1} \right),
\]

\[
\text{Cov}(\beta_2, u) = c_{\beta_2, u} \left( \frac{\alpha}{\sigma_2^2} P_2^{-1} \right),
\]

with

\[
c_{\beta_1, u} = \left[ \frac{1}{N \sigma_{\beta_1}^2 + \sigma_1^2/n} + \frac{\alpha^2}{N \sigma_{\beta_2}^2 + \sigma_2^2} + \frac{1}{\sigma_u^2} \right]^{-1}.
\]

Notice that when \( \alpha > 0 \), the treatment group \( u_i \) have negative covariance with both \( \beta_{11} \) and \( \beta_{12} \), while the control group \( u_i \) have negative covariance with \( \beta_{11} \) and positive covariance with \( \beta_{12} \). Again, this is a consequence of the treatment assignment parameterization and balance assumptions in this simplified model.
Turning to the posterior variance of the fixed effects obtained from a longitudinal-only model, we use elements $A_1$, $C_1$, $\theta$, and $C_2$ that are simply the reduced forms obtained by deleting the survival data and parameters. The joint posterior precision matrix of $(\beta_1, \mathbf{u})'$ is given by

$$
\begin{pmatrix}
P_1 \mathbf{I}_2 & \frac{n}{\sigma_1^2} \left( \frac{1}{N} \begin{pmatrix} 1 & 1 \end{pmatrix} \right)' \\
\frac{n}{\sigma_1^2} \left( \frac{1}{N} \begin{pmatrix} 1 & 1 \end{pmatrix} \right) & \left( \frac{n}{\sigma_1^2} + \frac{1}{\sigma_u^2} \right) \mathbf{I}_N
\end{pmatrix}.
$$

(2.1)

Inverting this yields the posterior variance of the longitudinal fixed effects,

$$
\text{Var}(\beta_1 | \mathbf{z}_1) = \left[ P_1 - N \left( \frac{n}{\sigma_1^2} \right)^2 \left( \frac{n}{\sigma_1^2} + \frac{1}{\sigma_u^2} \right)^{-1} \right]^{-1}.
$$

(2.2)

We can obtain the same result by setting $\alpha = 0$ in (1.6) above, since $P_u$ becomes $\left( \frac{n}{\sigma_1^2} + \frac{1}{\sigma_u^2} \right)$ and $P_2$ cancels out of the remaining terms after a bit of algebra. To obtain the analogous result for the posterior variance of $\beta_2$, we note that we obtain the same result either by re-computing the L&S posterior using a model that involves only the survival submodel, or by setting $n = 0$ in (1.6). Either method produces

$$
\text{Var}(\beta_2 | \mathbf{z}_2) = \left[ P_2 - N \left( \frac{\alpha}{\sigma_2^2} \right)^2 \left( \frac{\alpha^2}{\sigma_2^2} + \frac{1}{\sigma_u^2} \right)^{-1} \right]^{-1}.
$$

(2.3)

### 3. Joint model posterior mean

Turning to the mean of the complete parameter vector $\theta$, by L&S this is given by $D\mathbf{d}$, where

$$
\mathbf{d} = \begin{pmatrix}
\frac{n}{\sigma_1^2} z_{1+} + \frac{\mu_1}{\sigma_1^2} \\
\frac{n}{\sigma_1^2} (z_{1+}^{\text{trt}} - z_{1+}^{\text{ctrl}}) + \frac{\mu_{12}}{\sigma_1^2} \\
\frac{1}{\sigma_2^2} (z_{2+}^{\text{trt}} - z_{2+}^{\text{ctrl}}) + \frac{\mu_{22}}{\sigma_2^2} \\
\frac{1}{\sigma_1^2} z_1 + \frac{1}{\sigma_2^2} z_2
\end{pmatrix}.
$$

(3.1)

In this expression, $z_{1+}$ is the sum of all the longitudinal observations; $z_{1+}^{\text{trt}}$ and $z_{1+}^{\text{ctrl}}$ are sums of longitudinal observations from the treatment and control groups, respectively; and $z_{2+}$, $z_{2+}^{\text{trt}}$, and $z_{2+}^{\text{ctrl}}$.
and $z_{2+}^{ctrl}$ are defined analogously for the survival observations. Multiplying (3.1) by the inverse of (1.4) yields the vector of posterior means of $(\beta_1, \beta_2, u)'$.

\[
E(\beta_{11}|z_1, z_2) = \frac{z_{1+}}{\sigma_1^2/n} \left( Var(\beta_1) - \frac{nN_{\alpha, u}}{\sigma_1^2P_1} \right) + \frac{Var(\beta_1)\mu_{11}}{\sigma_1^2} \\
+ \frac{z_{2+}}{\sigma_2^2} \left( Cov(\beta_1, \beta_2) - \frac{nN_{\alpha, u}}{\sigma_2^2P_2} \right) + \frac{Cov(\beta_1, \beta_2)\mu_{21}}{\sigma_2^2} \\
E(\beta_{12}|z_1, z_2) = \left(z_{1+}^{ctrl} - z_{1+}^{ctrl} \right) \frac{Var(\beta_1) - Cov(\beta_1, u)}{\sigma_1^2/n} + \mu_{12} \left( \frac{Var(\beta_1)}{\sigma_1^2} \right) \\
+ \left(z_{2+}^{ctrl} - z_{2+}^{ctrl} \right) \left( Cov(\beta_1, \beta_2) - \alpha Cov(\beta_1, u) \right) \frac{Cov(\beta_1, \beta_2)}{\sigma_2^2} + \mu_{22} \left( \frac{Cov(\beta_1, \beta_2)}{\sigma_2^2} \right) \\
E(\beta_{21}|z_1, z_2) = \frac{z_{2+}}{\sigma_2^2} \left( Var(\beta_2) - \frac{\alpha^2 N_{\alpha, u}}{\sigma_2^2P_2} \right) + \frac{Var(\beta_2)\mu_{21}}{\sigma_2^2} \\
+ \frac{z_{1+}}{\sigma_1^2/n} \left( Cov(\beta_1, \beta_2) - \frac{\alpha^2 N_{\alpha, u}}{\sigma_1^2P_1} \right) + \frac{Cov(\beta_1, \beta_2)\mu_{11}}{\sigma_1^2} \\
E(\beta_{22}|z_1, z_2) = \left(z_{2+}^{ctrl} - z_{2+}^{ctrl} \right) \frac{Var(\beta_2) - \frac{\alpha^2 N_{\alpha, u}}{\sigma_2^2P_2} \right) + \frac{Var(\beta_2)\mu_{22}}{\sigma_2^2} \\
+ \left(z_{1+}^{ctrl} - z_{1+}^{ctrl} \right) \left( Cov(\beta_1, \beta_2) - \frac{\alpha^2 N_{\alpha, u}}{\sigma_1^2P_1} \right) + \frac{Cov(\beta_1, \beta_2)\mu_{12}}{\sigma_1^2}.
\]

4. Posterior for the linking parameter $\alpha$

To put a prior on $\alpha$ and derive its posterior, we begin by writing $A_1(\alpha)$ to emphasize the dependence of the design matrix on $\alpha$. Then the joint posterior of $(\theta, \alpha)'$ is

\[
p(\theta, \alpha|z) = \frac{f(z|A_1(\alpha), \theta)f(\theta|\mu)\pi(\alpha)}{\int \int f(z|A_1(\alpha), \theta)f(\theta|\mu)\pi(\alpha)d\theta d\alpha} \\
\propto \exp \left\{ -\frac{1}{2} \left[ (\theta - Dd)'D^{-1}(\theta - Dd) - d'Dd + z'C_1^{-1}z + (A_2\mu)'C_2^{-1}(A_2\mu) \right] \right\} \pi(\alpha).
\]

Recall that $\alpha$ appears in elements of $D$ and $d$. Obtaining an expression proportional to the marginal posterior of $\alpha$ requires integration of this expression with respect to $\theta$. The first part of the exponential is simply a normal kernel in $\theta$, leading to $\int \exp \left\{ -\frac{1}{2} (\theta - Dd)'D^{-1}(\theta - Dd) \right\} d\theta \propto |D|^{1/2}$. Then recall that $D^{-1}$ in (1.4) and $d$ in (3.1) depend on $\alpha$, but that the other two terms
in the exponential do not contain $\alpha$. Thus the expression for the posterior of $\alpha$ is straightforward

$$p(\alpha|z) = \frac{|D|^{1/2} \exp \left\{ \frac{1}{2} d'Dd \right\} \pi(\alpha)}{m(z)}$$

where $m(z) \propto \int |D|^{1/2} \exp \left\{ -\frac{1}{2} \left[ z' (C_1^{-1} + C_2^{-1}) z + (A_2\mu)' C_2^{-1} (A_2\mu) - d'Dd \right] \right\} \pi(\alpha) d\alpha.$ \hspace{1cm} (4.2)

There is no tidy analytical expression for this, as both $d'Dd$ and $|D|^{1/2}$ are complicated functions of $\alpha$. However, the integration required to obtain the marginal distribution $m(z)$ is only one-dimensional, so for any data set $z$, we can readily evaluate the posterior numerically.

5. Latent effect posterior

The conditional posterior distribution of the latent parameter, $p(u_i|z_{1i}, z_{2i}, \theta)$, is proportional to

$$\phi(z_{1i}|u_i, \theta) \phi(z_{2i}|u_i, \theta) \phi(u_i|\theta)$$

$$\propto \exp \left\{ -\frac{1}{2} \left[ \frac{(z_{1i} - \beta_{11} - \beta_{12} trt_i - u_i)^2}{\sigma_{1i}^2} + \frac{(z_{2i} - \beta_{21} - \beta_{22} trt_i - \alpha u_i)^2}{\sigma_{2i}^2} + \frac{u_i^2}{\sigma_u^2} \right] \right\}$$

$$\propto \exp \left\{ u_i \left( \frac{(z_{1i} - \beta_{11} - \beta_{12} trt_i)}{\sigma_{1i}^2} + \frac{\alpha (z_{2i} - \beta_{21} - \beta_{22} trt_i)}{\sigma_{2i}^2} \right) - \frac{u_i^2}{2} \left( \frac{n_i}{\sigma_{1i}^2} + \frac{\alpha^2}{\sigma_{2i}^2} + \frac{1}{\sigma_u^2} \right) \right\}.$$ \hspace{1cm} (5.1)

To find the posterior mode of $u_i$, we differentiate the log of (5.1), set the derivative to 0, and solve to obtain the conditional posterior mode of $u_i$,

$$\left( \frac{z_{1i} - \beta_{11} - \beta_{12} trt_i}{\sigma_{1i}^2} + \frac{\alpha (z_{2i} - \beta_{21} - \beta_{22} trt_i)}{\sigma_{2i}^2} \right) \sigma_{u_i}^2,$$ \hspace{1cm} (5.2)

where $\sigma_{u_i}^2 = \left( \frac{n_i}{\sigma_{1i}^2} + \frac{\alpha^2}{\sigma_{2i}^2} + \frac{1}{\sigma_u^2} \right)^{-1}$. Notice that the posterior mode is increasing with the sum of scaled residuals of the linear predictors from the longitudinal and survival submodels. The second derivative of the log posterior, $\frac{\partial^2 \log f(u_i|\theta, z_{1i}, z_{2i})}{\partial u_i^2} = -\sigma_{u_i}^{-2}$, is negative everywhere, and thus the Fisher information is an inverse of summed precisions, also intuitively sensible.

Dropping the conditioning on the fixed effects, we consider the full posterior of $\theta = (\beta'_1, \beta'_2, u')'$ in the joint model when $\sigma_{1i}^2, \sigma_{2i}^2, \sigma_u^2$, and $\alpha$ are assumed known. We obtain the posterior covariance
Benefits of joint modeling

matrix of \( u \) by inverting (1.4) and taking the lower left \( N \times N \) submatrix of the result,

\[
I_2 \otimes \left[ P_u^{-1} I_N + \text{Cov}(u, u) J_N \right],
\]

where \( \text{Cov}(u, u) = \frac{2P_u^{-1} \left( \frac{n}{\sigma_1^2} P_1^{-1} + \frac{\alpha}{\sigma_2} \right)^2 P_2^{-1}}{P_u - N \left( \frac{n}{\sigma_1^2} P_1^{-1} + \frac{\alpha}{\sigma_2} \right)^2 P_2^{-1}} \),

(5.3)

and \( J_K \) is a \( K \times K \) matrix of ones. Notice that the \( u_i \) for subjects in the same treatment group have correlation \( \text{Cov}(u, u)/ (P_u^{-1} + \text{Cov}(u, u)) \), while those for subjects in different treatment groups are uncorrelated. Again, this is a consequence of the assumed balance and the treatment group coding.

To find the analogous result for the model that uses only the longitudinal data, we simply set \( \alpha = 0 \) in (5.3) to obtain

\[
I_2 \otimes \left[ \left( \frac{n}{\sigma_1^2} + \frac{1}{\sigma_u^2} \right)^{-1} I_N + \text{Cov}(u, u|z_1) J_N \right],
\]

where \( \text{Cov}(u, u|z_1) = \frac{2 \left( \frac{n}{\sigma_1^2} + \frac{1}{\sigma_u^2} \right)^{-1} \frac{n}{\sigma_1^2} P_1^{-1}}{\left( \frac{n}{\sigma_1^2} + \frac{1}{\sigma_u^2} \right)^{-1} P_1^{-1}} \).

(5.4)

We can also compute the posterior mean of the random effects in the joint model with known variances. The posterior mean of the \( j^{th} \) random effect in the treatment group \( E(u_j|z_1, z_2) \) is

\[
\frac{n}{\sigma_1^2} \left( \text{Var}(u) z_{1j} + \text{Cov}(u, u) z_{1jt} \right) = \frac{\text{Cov}(\beta, u)}{P_1} \left[ \frac{z_{1t}}{\sigma_1^2/n} + \frac{\mu_{11}}{\sigma_{\beta_1}^2} + \frac{(z_{1t}^* - z_{1t}^*)}{\sigma_{\beta_1}^2/n} + \frac{\mu_{12}}{\sigma_{\beta_2}^2} \right]
\]

\[
+ \frac{\alpha}{\sigma_2^2} \left( \text{Var}(u) z_{2j} + \text{Cov}(u, u) z_{2jt} \right) = \frac{\text{Cov}(\beta, u)}{P_2} \left[ \frac{z_{2t}}{\sigma_2^2} + \frac{\mu_{12}}{\sigma_{\beta_2}^2} + \frac{(z_{2t}^* - z_{2t}^*)}{\sigma_{\beta_2}^2} + \frac{\mu_{22}}{\sigma_{\beta_2}^2} \right].
\]

(5.5)

where \( \text{Var}(u) = (P_u^{-1} + \text{Cov}(u, u)) \). Notice that this has an intuitive interpretation similar to that of the longitudinal treatment effect. On both the longitudinal and survival sides, we see a weighted sum of contributions from the individual’s data and the data from individuals in the same treatment group, subtracting off a piece that resembles a scaled, naïve fit (in square brackets). As before, \( \alpha \) appears to ensure that the contribution from the survival data goes in the
right direction. A similar expression is obtained for individuals in the control group, substituting $z_{1*}^{ctrl}$ and $z_{2*}^{ctrl}$ for $z_{1*}^{trt}$ and $z_{2*}^{trt}$.

References