Joint models: when are treatment estimates improved?

Laura A. Hatfield[*], James S. Hodges, and Bradley P. Carlin

Treatments affect many aspects of disease; for example, a drug may improve symptoms, prolong survival, and cause serious side effects. A broader perspective on clinical effectiveness, considering multiple outcomes, requires analyses that account for relationships among outcomes. So-called joint modeling induces such relationships via shared parameters. Practical questions arise, including "When do we require a joint model?" and "How much do we gain by its use?" Motivated by these questions, we compare Gaussian joint models with shared latent parameters to separate models for each outcome individually. When we assume a single longitudinal measurement, known error variances, and no censoring, joint and separate treatment effect posteriors converge as the priors become improper. This result still holds when we add multiple longitudinal measurements and unknown error variance, but not when we make the prior informative for at least one treatment effect (longitudinal or survival). Joint models also improve inference under some censoring scenarios. Our results suggest that joint models are most useful when an information imbalance allows abundant information in one outcome to compensate for a paucity of information in another.

AMS 2000 subject classifications: Primary 62F15; secondary 62H20.

Keywords and phrases: Joint longitudinal-survival modeling, Bayesian learning, Multiple outcomes, Bimodal posteriors.

1. INTRODUCTION

Medical treatments rarely affect a single aspect of disease; rather, they manifest a spectrum of measurable features that may be elevated to "primary outcomes" or relegated to "side effects". Imagine a drug that simultaneously reduces some symptoms, prolongs survival, and causes serious side effects. In considering this therapy, patients and clinicians must trade off outcomes. Quality of life dominates in terminal patients seeking palliation, while survival trumps side effects in patients seeking a cure. Multiple outcomes also complicate statistical analysis. For example, after a patient dies, she cannot report her symptom severity, and more symptom-prone patients experience worse survival. Relationships among outcomes can bias estimates or

*Corresponding author.

sacrifice efficiency in single-outcome analyses. Ignoring missing symptom reports from patients who die early biases our estimate of the treatment effect on symptoms. Considering only survival underestimates the treatment's total effect on disease.

Let Z_1 and Z_2 represent two health outcomes of interest. If they share a measurement scale, a standard multivariate distribution may serve as the joint distribution, $[Z_1, Z_2]$, enabling standard models. However, variables on different scales preclude "brand name" joint distributions. Copula models handle this by constructing a joint using cumulative distribution functions. So-called joint models instead *induce* a joint distribution using conditional specification.

Three approaches to joint modeling may be distinguished [26]. First are models for survival, Z_2, conditional on longitudinal predictors, Z_1, written $[Z_2|Z_1]$ [14, 21]. Endogenous or error-prone predictors may themselves require a model, $[Z_1]$. Some authors model the predictor and plug it into a survival model [22, 11], while others simultaneously fit a latent trajectory and a survival model that depends on this [24].

A second class comprises survival events that censor longitudinal data collection [20]. Censoring matters when the missing outcomes differ systematically from the observed outcomes. For example, if patients with the worst symptoms drop out, the remaining data paints a too-rosy picture. Hu and Sale [12] showed that longitudinal predictions improve when the missing data mechanism, $[Z_2]$, is modeled. Examples include pattern-mixture models and selection models [15].

A final approach reflects the health outcomes perspective, emphasizing multiple dimensions of health [17]. In our clinical trial example, patients value treatment effects on both symptoms and survival. Shared parameter models assume independence between outcomes given a shared latent variable, u, specified via $[Z_1|u]$ and $[Z_2|u]$. Marginalizing these over the distribution $[u]$ yields a joint distribution. These approaches trade the benefit of symmetrical outcome models against the added complication of latent variable models [13].

After more than a decade of development, joint models now accommodate myriad data quirks, as recent reviews attest [18, 23, 3]. However, the properties and performance of joint models tend to receive less attention in the literature [10]. Exceptions include papers on residuals [19], sam-

ple size and power [2], robustness [28], and less parametric approaches [1, 27]. The original work on bias [5] and efficiency [4] prioritized the survival outcome. More recent comparisons show less benefit of joint modeling in other settings [7, 6].

Our work springs from two practical questions: "When do we require a joint model?" and "How much do we gain by its use?" To answer these, we compare posteriors from joint models to those of corresponding longitudinal- or survival-only models. Strong assumptions permit analytical methods, while simulations crack more realistic models. We conclude that inference improves when a joint model allows abundant information in one outcome type to compensate for an information deficit in another. Section 2 highlights such an asymmetry in longitudinal symptom and survival data from a real-world oncology trial. Section 3 specifies a Gaussian joint model for balanced data: one symptom observation and one survival time. With known error variance and improper priors, the treatment effect posteriors are the same in joint and separate models. This also holds in Section 4 when we use simulation to model multiple symptom observations. However, informative priors on one submodel can produce an asymmetry that facilitates joint modeling benefits. When censoring depends on the latent variables (Section 5), the joint model can sometimes improve posterior measures. We conclude in Section 6 with discussion and suggestions for future work. Algebraic details are relegated to a technical appendix.

2. CLINICAL TRIAL EXAMPLE

Previously-analyzed cancer clinical trials inspire our approach [8]. Our models are simpler, to permit focus on the relative merits of joint and separate models, but the inferential targets remain treatment effects on symptoms and survival. In the trial, patients were randomized to treatment or control, then self-reported their symptoms on a continuous scale. Progression-free survival (PFS) extended from enrollment to disease progression or death, whichever came first. Censored survival times were rare ($< 10\%$), and patients provided a median of 14 symptom reports.

We construct a shared random effects model, which assumes symptoms and survival are independent given a latent variable [23]. Person i reports her symptoms n_i times, forming the vector $\mathbf{z}_{1i} = (z_{1i1}, \ldots, z_{1in_i})'$, and contributes a single survival time, z_{2i}. Collecting all participants' symptom, survival, and latent variables into $\mathbf{z}_1$, $\mathbf{z}_2$, and $\mathbf{u}$, conditional independence leads to a simple joint likelihood, $f(\mathbf{z}_1, \mathbf{z}_2|\mathbf{u}, \boldsymbol{\theta}) = f(\mathbf{z}_1|\mathbf{u}, \boldsymbol{\theta})f(\mathbf{z}_2|\mathbf{u}, \boldsymbol{\theta})$. The parameter vector $\boldsymbol{\theta}$ contains the treatment effects of interest, as well as error variances, etc. We require a prior for these parameters, $\pi(\boldsymbol{\theta})$, and a distribution for the latent variables, $p(\mathbf{u})$. Bayes rule yields the full posterior,

$$(1) \quad f(\mathbf{u}, \boldsymbol{\theta}|\mathbf{z}_1, \mathbf{z}_2) = \frac{f(\mathbf{z}_1, \mathbf{z}_2|\boldsymbol{\theta}, \mathbf{u})\pi(\boldsymbol{\theta})p(\mathbf{u})}{\int f(\mathbf{z}_1, \mathbf{z}_2|\boldsymbol{\theta}, \mathbf{u})\pi(\boldsymbol{\theta})p(\mathbf{u})d\boldsymbol{\theta}d\mathbf{u}} .$$

In the clinical trial example, we model the symptoms, $\mathbf{z}_1$, and the log survival time, $\mathbf{z}_2$, as Gaussian, i.e.,

$$(2) \quad \begin{aligned} z_{1ij} &\sim \text{Normal}(\mathbf{X}_{1ij}\boldsymbol{\beta}_1 + u_i, \sigma_1^2) \\ z_{2i} &\sim \text{Normal}(\mathbf{X}_{2i}\boldsymbol{\beta}_2 + \alpha u_i, \sigma_2^2) \\ u_i &\sim \text{Normal}(0, \sigma_u^2) , \end{aligned}$$

where $i = 1, \ldots, N$ indexes individuals and $j = 1, \ldots, n_i$ indexes time. The two predictor vectors, $\mathbf{X}_1$ and $\mathbf{X}_2$, include treatment assignment, and the treatment effects are elements of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$. A single latent variable u_i links the two models; it appears as an intercept in the longitudinal mean and as a *scaled* intercept αu_i in the log survival mean. To complete the Bayesian model, we place vague $Normal(0, 100)$ priors on $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, and α and mildly informative $Gamma(1, 1)$ priors on the precisions, σ_1^{-2}, σ_2^{-2}, and σ_u^{-2}.

In the following, we ask, "Does a joint model change our conclusions?" The scaling parameter α mediates the latent variable's contribution to the two submodels. For example, a negative α implies that patients with higher symptoms have shorter survival. After fitting model (2) to clinical trial data, the resultant posterior credible interval for α lies entirely below zero, which supports a clinically sensible association between the outcomes. However, non-zero α does not mean treatment effect estimates benefit.

Does survival information alter symptom conclusions and vice versa? We compare treatment effect posteriors from the joint model (2) to those from a longitudinal-only model that omits survival and a survival-only model that omits symptoms. The left panel of Figure 1 displays posterior median symptom trajectories in the treatment (black) and control (gray) groups, from the joint (solid) and longitudinal-only (dashed) models. The right panel displays the posterior median survival curve in each treatment group, from the joint (solid) and survival-only (dashed) models. Vertical lines are point-wise, equal-tail 95% posterior credible intervals. Adding survival data via this joint model does not change the symptom medians or intervals (left panel). However, adding symptom data shifts the survival medians *and* narrows the intervals, especially at days 50 and 100 (right panel). Apparently, symptom reports carry the bulk of information in this data set and model.

However, this balance need not hold for all data sets, likelihoods, or priors. As noted in Section 1, joint models can reduce bias and increase efficiency, especially when focusing on one outcome. We consider outcomes of joint clinical interest, and we expect that their relative contributions will depend on the number of observations, censoring, and error variances.

3. JOINT MODEL FOR ONE VALUE OF EACH TYPE

Although we now leave behind real data, we retain the language of symptoms and survival outcomes, treatment and
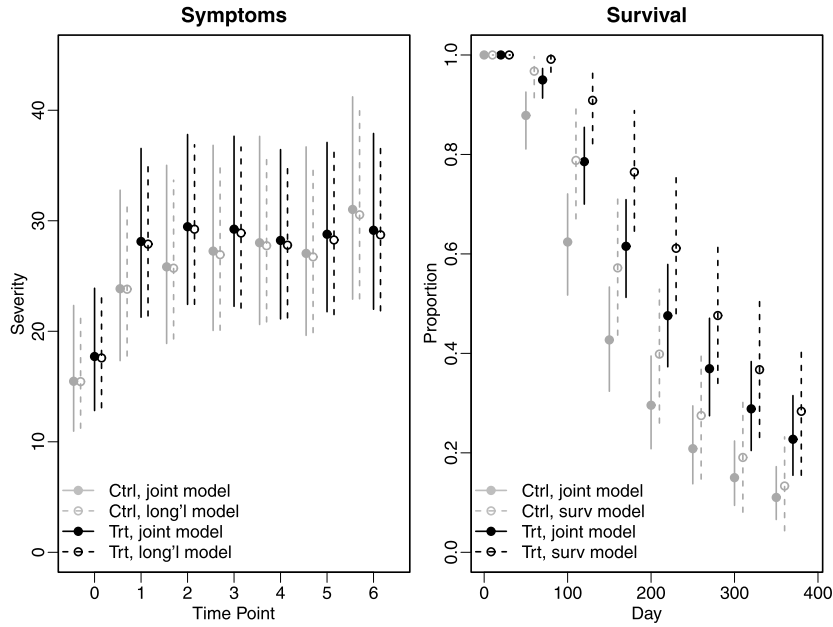
Figure 1. Fitted trajectories for symptom severity (left) and survival (right) in a Gaussian joint model fit to the clinical trial data. The treatment group fits are in black, control group in grey. Joint model results are plotted with solid lines; separate (longitudinal- or survival-only) model results are plotted as dashed lines.

control groups. We begin by assuming: 1) each person reports symptoms once; 2) symptoms z_{1i}, log survival z_{2i}, and the latent u_i are all Gaussian; 3) treatment is the only predictor; and 4) no data are missing. We also assume, without loss of generality, that treatment and control groups each have $N/2$ patients. Collect the observations into a single $2N$-vector of symptoms then survival, $\mathbf{Z} = (\mathbf{z}_1', \mathbf{z}_2')'$, each sorted so treatment values precede control. The covariate matrix $\mathbf{X}$ contains only intercepts and treatment indicators (coded as 1 for treatment and -1 for control). Coefficients for symptoms, $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12})'$, and survival, $\boldsymbol{\beta}_2 = (\beta_{21}, \beta_{22})'$, concatenate into $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$. The model can then be written

$$(3) \qquad \mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{u} + \boldsymbol{\epsilon} \,,$$

where

$$\mathbf{X} = \left(\begin{array}{cc} \left(\begin{array}{cc} \mathbf{1}_{\frac{N}{2}} & \mathbf{1}_{\frac{N}{2}} \\ \mathbf{1}_{\frac{N}{2}} & -\mathbf{1}_{\frac{N}{2}} \end{array} \right) & \begin{array}{cc} \mathbf{0}_N & \mathbf{0}_N \end{array} \\ \begin{array}{cc} \mathbf{0}_N & \mathbf{0}_N \end{array} & \left(\begin{array}{cc} \mathbf{1}_{\frac{N}{2}} & \mathbf{1}_{\frac{N}{2}} \\ \mathbf{1}_{\frac{N}{2}} & -\mathbf{1}_{\frac{N}{2}} \end{array} \right) \end{array} \right),$$

$$\mathbf{W} = \left(\begin{array}{c} \mathbf{I}_N \\ \alpha \mathbf{I}_N \end{array} \right),$$

$$\mathbf{u} \sim \text{Normal}(\mathbf{0}, \sigma_u^2 \mathbf{I}_N) \,,$$

and

$$\boldsymbol{\epsilon} \sim \text{Normal}\left(\left(\begin{array}{c} \mathbf{0} \\ \mathbf{0} \end{array} \right), \left(\begin{array}{cc} \sigma_1^2 \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 \mathbf{I}_N \end{array} \right) \right) \,.$$

We use $\mathbf{1}_K$ and $\mathbf{0}_K$ to denote K-vectors of ones and zeros, respectively. This expression resembles a factor analysis with factors $\mathbf{u}$, loading matrix $\mathbf{W}$, and factor loadings fixed at 1 in the symptom submodel and α in the survival submodel.

Standard calculations yield the marginal distribution

$$(4)$$
$$\mathbf{Z} \sim \text{Normal}\left(\mathbf{X}\boldsymbol{\beta}, \left(\begin{array}{cc} (\sigma_1^2 + \sigma_u^2)\mathbf{I}_N & \alpha\sigma_u^2\mathbf{I}_N \\ \alpha\sigma_u^2\mathbf{I}_N & (\sigma_2^2 + \alpha^2\sigma_u^2)\mathbf{I}_N \end{array} \right) \right) \,.$$

This model is over-parameterized; the variance-covariance matrix has 3 degrees of freedom but 4 parameters, σ_1^2, σ_2^2, σ_u^2, and α. We first solve this by treating *all* the variance parameters as known, though fixing one would suffice. Below, we consider data sufficient to ensure all variance parameters are identified.

The linking parameter α governs the correlation between a person's two outcomes,

$$(5) \qquad \text{Corr}(z_{1i}, z_{2i}) = \frac{\alpha\sigma_u^2}{\sqrt{(\sigma_1^2 + \sigma_u^2)(\sigma_2^2 + \alpha^2\sigma_u^2)}} \,.$$

Although shared random effect models treat outcomes more symmetrically than other approaches, the symmetry breaks because only one submodel contains α. That is, although in every other respect, the formulation of the models for the two outcomes may be identical (if desired), an arbitrary choice must be made with regard to the scaling of u_i. The model in which u_i appears without the α coefficient determines the scale of u_i, while α is included to re-scale the
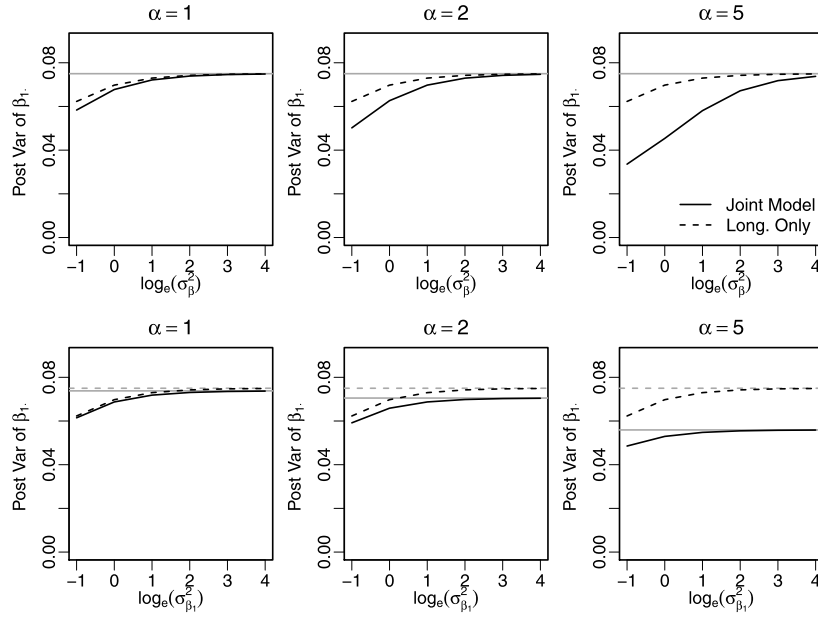
Figure 2. Posterior variance of $\beta_{1\cdot}$ as it depends on the prior variances when they are constrained to be the same (top row) or when $\sigma_{\beta_2}^2 = 2$ is fixed (bottom row). The joint model is plotted as a solid line, symptom-only model as a dashed line, and limits are shown as grey horizontal lines.

latent parameters for their contribution to the other outcome model. Known α obviates the problem, but the choice of a submodel in which to include α matters when its value is unknown, see Section 3.2.

3.1 Impact on regression coefficients

Consider the symptom treatment effect, β_{12}. With all variance parameters fixed, the data and prior on $\boldsymbol{\beta}$ determine its posterior. We place independent normal priors on the coefficients, $\pi(\boldsymbol{\beta}_1) = \text{Normal}((\mu_{11}, \mu_{12})', \sigma_{\beta_1}^2 \mathbf{I}_2)$ and $\pi(\boldsymbol{\beta}_2) = \text{Normal}((\mu_{21}, \mu_{22})', \sigma_{\beta_2}^2 \mathbf{I}_2)$. The intercept and treatment effect variances prove equal in this model, $Var(\beta_{11}|\mathbf{z}_1, \mathbf{z}_2) = Var(\beta_{12}|\mathbf{z}_1, \mathbf{z}_2)$, so we write $Var(\beta_{1\cdot}|\mathbf{z}_1, \mathbf{z}_2)$. Section A.1 of the appendix gives details of the joint model posterior variance,

$$(6) \qquad Var(\beta_{1\cdot}|\mathbf{z}_1, \mathbf{z}_2) = c_\beta \left(P_2 - N P_u^{-1} \left(\frac{\alpha}{\sigma_2^2} \right)^2 \right),$$

where

$$c_\beta = \left[P_1 P_2 - N P_u^{-1} \left(\left(\frac{\alpha}{\sigma_2^2} \right)^2 P_1 + \left(\frac{1}{\sigma_1^2} \right)^2 P_2 \right) \right]^{-1},$$

$$P_1 = \left(\frac{N}{\sigma_1^2} + \frac{1}{\sigma_{\beta_1}^2} \right), \quad P_2 = \left(\frac{N}{\sigma_2^2} + \frac{1}{\sigma_{\beta_2}^2} \right),$$

and

$$P_u = \left(\frac{1}{\sigma_1^2} + \frac{\alpha^2}{\sigma_2^2} + \frac{1}{\sigma_u^2} \right).$$

Section A.2 of the appendix derives the symptom-only model posterior variance,

$$(7) \quad Var(\beta_{1\cdot}|\mathbf{z}_1) = \left[P_1 - N \left(\frac{1}{\sigma_1^2} \right)^2 \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_u^2} \right)^{-1} \right]^{-1}.$$

These complex expressions do not immediately illuminate the differences, so we study their limits as the priors become improper. As the prior variances, $\sigma_{\beta_2}^2$ and $\sigma_{\beta_1}^2$, approach infinity, the joint and symptom-only model posterior variances approach the same limit, $\frac{\sigma_1^2 + \sigma_u^2}{N}$. This only holds when *both* prior variances increase together. If we fix $\sigma_{\beta_2}^2$ and let $\sigma_{\beta_1}^2$ approach ∞, the joint model posterior variance reaches a smaller limit than that of the symptom-only model. Results for the survival treatment effect are symmetrical.

Figure 2 illustrates the above results. In the top row, the prior variances are equal, $\sigma_{\beta_1}^2 = \sigma_{\beta_2}^2 = \sigma_\beta^2$. As they increase, the posterior variances in the joint (solid) and symptom-only model (dashed) approach the same limit. In the bottom row, we fix the survival prior variance, $\sigma_{\beta_2}^2 = 2$, and increase the symptom prior variance, $\sigma_{\beta_1}^2$. The joint model posterior variance approaches a lower limit, indicating a benefit of joint modeling. Larger α signals a stronger relationship between the outcomes and more benefit of joint modeling.

Bayesian models combine information about parameters from data and priors. It is clear how the prior variances influence parameter posteriors, since they are directly specified in relation to a parameter. The simplest way to provide more information from one side of the model is thus
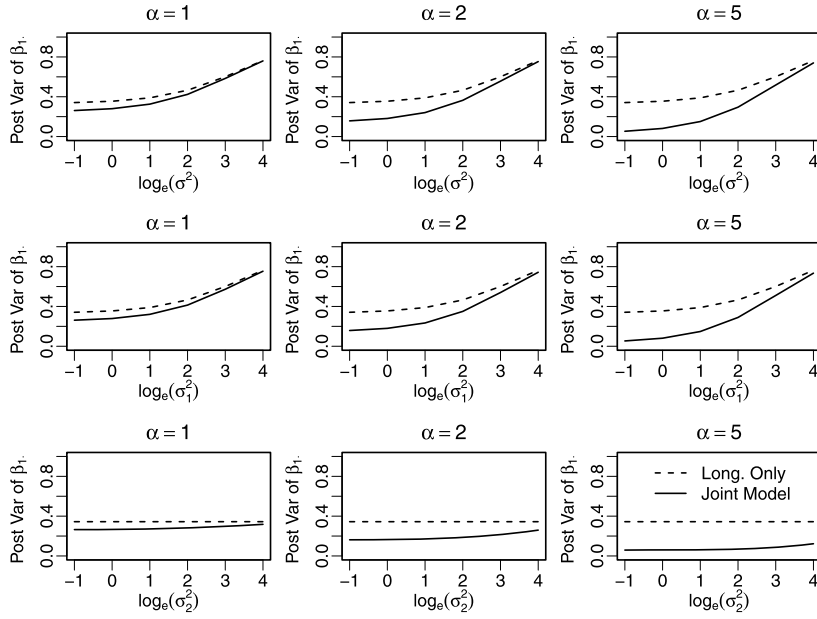
Figure 3. Posterior variance of $\beta_{1\cdot}$ when both data variances are equal (top), the survival data variance is fixed and σ_1^2 varies (middle), or the longitudinal data variance is fixed and σ_2^2 varies (bottom). The joint model is plotted as a solid line, symptom-only model as a dashed line.

to manipulate the priors. However, the data are related to the parameters in more complex ways. We can perform similar experiments by manipulating the data variances while keeping the priors fixed. Taking limits in the data variances, we find that the posterior parameter variances from both joint and separate models depend only on their respective prior variances. That is, as $\sigma_1^2 \to \infty$, $Var(\beta_{1\cdot}|\mathbf{z}_1, \mathbf{z}_2)$ and $Var(\beta_{1\cdot}|\mathbf{z}_1) \to \sigma_{\beta_1}^2$, regardless of σ_2^2. In Figure 3, we plot the posterior variance of the longitudinal parameters in the joint and separate models as a function of the data variances on both sides of the model. These plots represent the "information imbalance" of the *data*, which is different from that of the priors. The top row shows the results for both data variances equal, the middle row for fixed survival data variance σ_2^2, and the bottom row for fixed longitudinal data variance σ_1^2. In all cases, the joint model posterior variance is lower, though the benefit diminishes whenever the longitudinal data variance increases (top and middle rows). When only the survival data variance increases (bottom row), the posterior variance remains about constant. Together with the results in Figure 2, these demonstrate the complexity of quantifying the information content on the two sides of the model. While the prior variance is a fairly straightforward input of information on a parameter, the data inform parameter posteriors in a more complex way.

The expected benefits of joint modeling include both variance and bias reduction, so we also study posterior means. Consider again the symptom treatment effect; Section A.3 of the appendix contains details of the joint model posterior mean,

(8)
$$
\begin{aligned}
E(\beta_{12}|\mathbf{z}_1, \mathbf{z}_2) =& \left(z_{1+}^{trt} - z_{1+}^{ctrl}\right) \frac{Var(\beta_{1\cdot}) - Cov(\beta_{1\cdot}, u_\cdot)}{\sigma_1^2} \\
& + \mu_{12} \frac{Var(\beta_{1\cdot})}{\sigma_{\beta_1}^2} \\
& + \left(z_{2+}^{trt} - z_{2+}^{ctrl}\right) \frac{Cov(\beta_{1\cdot}, \beta_{2\cdot}) - \alpha Cov(\beta_{1\cdot}, u_\cdot)}{\sigma_2^2} \\
& + \mu_{22} \frac{Cov(\beta_{1\cdot}, \beta_{2\cdot})}{\sigma_{\beta_2}^2} \, ,
\end{aligned}
$$

where

$$
\begin{aligned}
Cov(\beta_{1\cdot}, \beta_{2\cdot}) &= c_\beta \left(NP_u^{-1} \frac{1}{\sigma_1^2} \frac{\alpha}{\sigma_2^2}\right) \, , \\
Cov(\beta_{1\cdot}, u_\cdot) &= c_{\beta u} \left(\frac{1}{\sigma_1^2} P_1^{-1}\right) \, ,
\end{aligned}
$$

and

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

Here, z_{1+}^{trt} and z_{1+}^{ctrl} are sums of symptom values from the treatment and control groups, and z_{2+}^{trt} and z_{2+}^{ctrl} are the same for survival observations. Notice that the mean's four terms are a weighted sum of the data mean and prior mean from the symptom submodel, followed by the same from the survival submodel.

The posterior mean (8) depends on the data, so to explore how it also depends on parameters, we must simulate data.
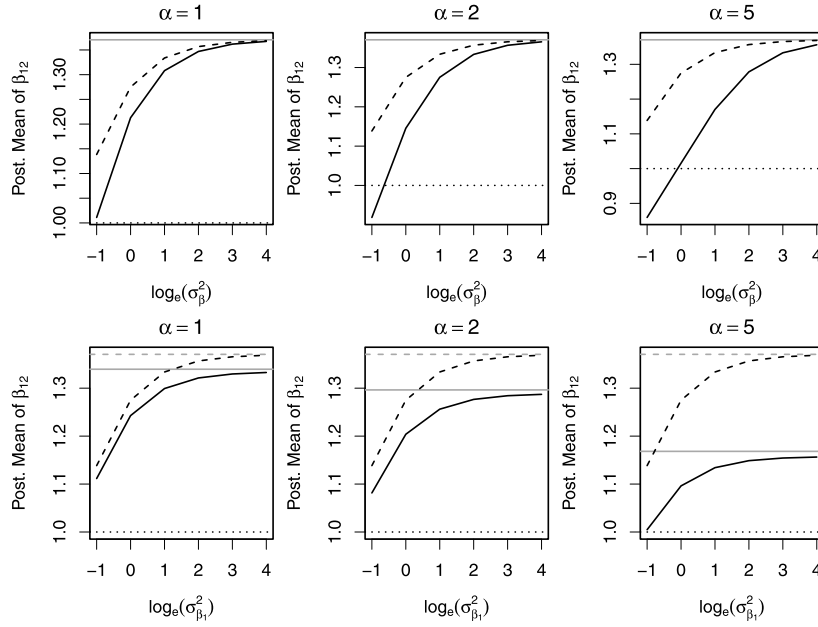
Figure 4. Posterior mean of β_{12} as it depends on the prior variances when they are constrained to be the same (top row) or when $\sigma_{\beta_2}^2 = 2$ is fixed (bottom row). The joint model is plotted as a solid line, symptom-only model as a dashed line. Limits and the true value (dotted line) are shown as grey horizontal lines.

Our simulation assumes true values $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = (1, 1)'$, $\sigma_1^2 = \sigma_2^2 = 1/2$, $\sigma_u^2 = 1$, and $N = 20$. We varied α and the prior variances to compare joint and symptom-only models.

Figure 4 shows β_{12}'s mean for a single simulated data set. In the top row, prior variances are equal, and the joint and symptom-only means approach the same limit. Specifically, $E(\beta_{12}|\mathbf{z}_1, \mathbf{z}_2)$ and $E(\beta_{12}|\mathbf{z}_1)$ both approach $(z_{1+}^{trt} - z_{1+}^{ctrl})/N$. In the bottom row, $\sigma_{\beta_2}^2 = 2$ is fixed, and the symptom-only mean also approaches this limit (dashed), but the joint model approaches a different limit (solid). As before, the larger α is, the further apart the joint and symptom-only models are. In this particular simulated data set, the symptom-only mean is closer to the true value, but this will not be true in general. Note that the joint model limit equals the MLE from the marginal (4), the usual least squares solution

(9)
$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z} = \frac{1}{N}\left(z_{1+},\ z_{1+}^{trt} - z_{1+}^{ctrl},\ z_{2+},\ z_{2+}^{trt} - z_{2+}^{ctrl}\right)'$$

This reproduces the well-known result that a flat-prior Bayesian posterior mode and a maximum likelihood estimate are equivalent.

3.2 Impact on linking parameter α

We next relax the assumption of known α. One observation of each type per person suffices to identify α when it is the only unknown variance parameter. Section A.4 of the appendix gives details of α's posterior. We illustrate the potential for learning about α by simulating data and highlighting a small sample issue.

The data-generating values were the same as in Section 3.1 with $\alpha = 2$. We varied the number of subjects $N \in \{10, 20\}$ and the longitudinal error variance $\sigma_1^2 \in \{1/2, 2\}$, simulating 20 data sets at each combination. To isolate information about α in the data, we specified an improper flat prior, $\pi(\alpha) \propto 1$, and vague but proper coefficient priors $\sigma_{\beta_1}^2 = \sigma_{\beta_2}^2 = 100$. With an improper prior, the propriety of the α posterior is not guaranteed. However, we can show that α's posterior is proper by demonstrating that the integral of the likelihood with respect to α is finite. The details are in Section A.4 of the Appendix.

Figure 5 displays the posterior of α. The average across replications, plotted in black, makes the basic shape clear; individual posteriors, plotted in grey, show the variation across simulations. The data provide more information about α when there are more subjects, N, (left vs right column) and less error variance, σ_1^2/n, (top vs bottom rows). With $N = 10$ patients (left column), α's posterior can be far from the truth, diffuse, and bi-modal. With $N = 20$ patients (right column), α's posterior is more symmetric and unimodal around the true value, $\alpha = 2$.

To understand this occasional bimodality, we select two draws from the least informative scenario ($N = 10$, $\sigma_1^2 = 2$). Figure 6 displays scatter plots of the simulated latent and longitudinal values (top row) and their posteriors (bottom row). Why are the posteriors so different? Recall the factor analysis interpretation of model (3): the loadings on u_i are assumed to equal 1 in the longitudinal submodel. That is, if we plot z_{1i} versus u_i in each treatment group, the slope of a least squares line fit to these values should be approx-
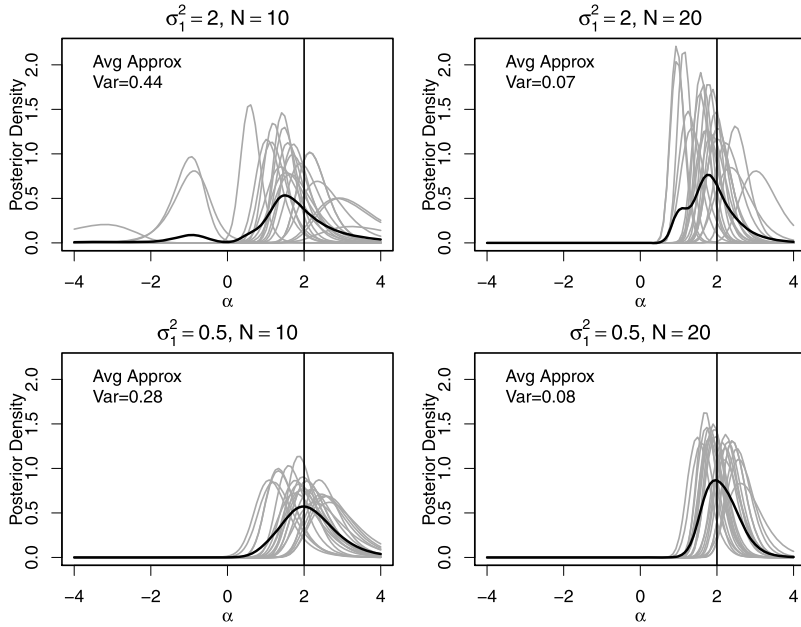
Figure 5. *The posterior of α from 20 simulated data sets with true $\alpha = 2$. The average is drawn in black and the rest in grey. Approximate variance is averaged across the inverse Hessian at the mode of each simulated posterior.*

imately 1. However, in the "bad" replicate (left panel), the actual relationship between z_{1i} and u_i is much weaker, a slope of just $= -0.03$ in the least squares fit line between these two variables shown on the figure. Compare this to the slope of the least squares fit to the two variables in the "good" replicate (right panel), which has a slope of 0.66, much closer to the assumed (modeled) value of 1.

Small sample size and large error variance conspire to produce "unlucky" data draws. As an extreme example, imagine that (by staggering coincidence) we happened to draw all identical values of z_{1i}. These longitudinal data would have no information about the latent variables (or any other parameters except the intercept β_{11}). The least squares line would have 0 slope. The survival data would have to do all the work of identifying the u_i. However, since each u_i is multiplied by the unknown α in the survival model, it is not well identified by only the survival data. Notice that the likelihood is identical for αu_i and $-\alpha(-u_i)$. Thus the posteriors of both the u_i and α will have symmetric modes around 0 corresponding to these two possibilities. A larger sample or smaller variance fixes this problem, as we see in the simulations.

4. JOINT MODEL FOR MULTIPLE OBSERVATIONS

Next, we increase the number of symptom observations, n, and create a centered n-vector of observation times, repeated N times, $\mathbf{t} = \mathbf{1}_N \otimes (t_1, \ldots, t_n)'$. We collect the individual vectors $\mathbf{z}_{1i} = (z_{11i}, \ldots, z_{1ni})'$ into $\mathbf{z}_1 = (\mathbf{z}_{11}', \ldots, \mathbf{z}_{1N}')'$ and expand model (3) to

$$(10) \qquad \mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{u} + \boldsymbol{\epsilon} \,,$$

where

$$\mathbf{X} = \begin{pmatrix} \begin{pmatrix} \mathbf{1}_{\frac{Nn}{2}} & \mathbf{t} & \mathbf{1}_{\frac{Nn}{2}} & \mathbf{t} \\ \mathbf{1}_{\frac{Nn}{2}} & \mathbf{t} & -\mathbf{1}_{\frac{Nn}{2}} & -\mathbf{t} \end{pmatrix} & \mathbf{0}_{Nn} \quad \mathbf{0}_{Nn} \\ \mathbf{0}_N \quad \mathbf{0}_N \quad \mathbf{0}_N \quad \mathbf{0}_N & \begin{pmatrix} \mathbf{1}_{\frac{N}{2}} & \mathbf{1}_{\frac{N}{2}} \\ \mathbf{1}_{\frac{N}{2}} & -\mathbf{1}_{\frac{N}{2}} \end{pmatrix} \end{pmatrix} \,,$$

$$\boldsymbol{\epsilon} \sim \text{Normal}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma_1^2 \mathbf{I}_{Nn} & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 \mathbf{I}_N \end{pmatrix} \right) \,,$$

and

$$\mathbf{u} \sim \text{Normal}\left(\mathbf{0}, \sigma_u^2 \mathbf{I}_N \right) \,.$$

The upper left submatrix of $\mathbf{X}$ contains the symptom predictors intercept, time, treatment, and time × treatment, with corresponding coefficient $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12}, \beta_{13}, \beta_{14})'$. The lower right submatrix contains the survival predictors intercept and treatment, with corresponding coefficient $\boldsymbol{\beta}_2 = (\beta_{21}, \beta_{22})'$.

Assuming again that random intercepts link the two submodels, the marginal covariance matrix of the $(n+1)N$ data vector $\mathbf{Z}$ is

$$(11)$$
$$Var(\mathbf{Z}) = \begin{pmatrix} \mathbf{I}_N \otimes \left(\sigma_1^2 \mathbf{I}_n + \sigma_u^2 \mathbf{J}_n \right) & \alpha \sigma_u^2 \mathbf{I}_N \\ \alpha \sigma_u^2 \mathbf{I}_N & (\sigma_2^2 + \alpha^2 \sigma_u^2) \mathbf{I}_N \end{pmatrix} \,,$$

where $\mathbf{J}_n$ is an $n \times n$ matrix of ones. In the joint model, the four variance parameters are identified for $n \geq 2$, because with at least three observations per person, the co-
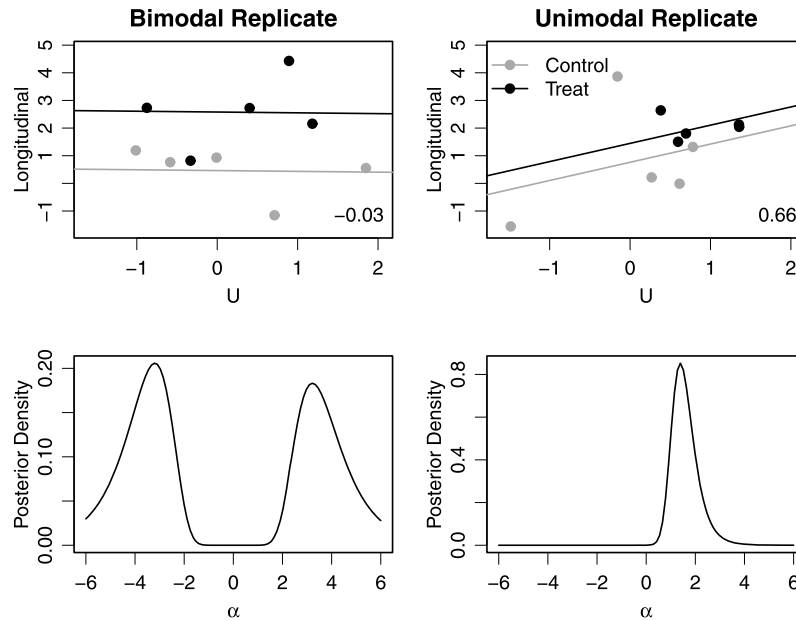
Figure 6. Two data draws from the $N = 10$, $\sigma_1^2 = 2$ scenario (top row) and their α posteriors (bottom row). On the left is an example of the worst bimodality and on the right, a unimodal and approximately correct posterior. Data from the treatment group are shown in black, control group in grey. Solid lines show ordinary least squares fits to latent and symptom data for each treatment group; their slopes are printed in the lower right.

variance matrix has at least six degrees of freedom. In the symptom-only model with $n = 2$, the covariance matrix is $Var(\mathbf{z}_1) = \mathbf{I}_N \otimes (\sigma_1^2 \mathbf{I}_n + \sigma_u^2 \mathbf{J}_n)$; its two parameters can also be identified because there are three degrees of freedom. In the survival-only model, the covariance matrix is $Var(\mathbf{z}_2) = (\sigma_2^2 + \alpha^2 \sigma_u^2) \mathbf{I}_N$; its three parameters cannot be identified with one survival observation per person (1 degree of freedom). For a fairer comparison, we assume α is known, so both the symptom- and survival-only models have two variance parameters.

4.1 Impact on regression coefficients

We fit the joint model in (10) using MCMC and compare the parameter posteriors to those from symptom- and survival-only models. The simulations all have $N = 20$ patients with $n = 3$ observation times and the same variance parameters as Section 3.1. We set $\boldsymbol{\beta}_2 = (1, 1)'$ in all models so that for $X_{2i} = (1, trt_i)'$, the expected means are 0 in the control group and 2 in the treatment group. The data-generating values of $\boldsymbol{\beta}_1$ vary across models, but all correspond to design vectors $X_{1ij} = (1, t_j, trt_i, trt_i * t_j)'$ where $t = \{-1, 0, 1\}$. In Model 1, the groups are different at baseline with no time trend, i.e., $\boldsymbol{\beta}_1 = (1, 0, 1, 0)'$. In Model 2, the groups start out equal and have different time trends, i.e., $\beta_{11} = (1, .5, 0, .5)$. Both of these include a random intercept in the longitudinal model, which also appears in the survival submodel multiplied by coefficient α. Next, Model 3 uses the same coefficients as Model 2, but random slopes appear in the longitudinal submodel and in the survival submodel

with coefficient α. We implement two versions of Model 4. Both have the same coefficients as Models 2 and 3 and both random slopes and intercepts in the longitudinal submodel. Model 4i uses the random intercepts in the survival submodel with linking parameter α, while Model 4s uses the random slopes, again with linking parameter α. The top row of Figure 7 illustrates these variations. Solid lines plot the fixed effect trajectories for treatment (black) and control (grey), and dashed lines show the influence of random effects. In all of these simulations, the priors on each element of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are normal with mean 0 and variance 100. The priors on the precision parameters are $Gamma(1, 1)$.

Figure 7 displays posterior MSE differences across 100 simulations, symptom-only versus joint in the second row and survival-only versus joint in the third row. Positive values indicate a benefit of joint modeling (smaller MSE). The regression parameters show no benefit in any scenario, while the variance parameters and random effects (not all shown) do benefit across all models. Symptom parameters unaffected by random effects show minimal simulation variability (i.e., Int and Trt effects in Model 3, Time and Time × Trt effects in Models 1 and 2), while all of the survival parameters have large variability across models.

5. SIMULATING INFORMATIVE CENSORING

Finally, we introduce informative censoring, that is, a missing data mechanism that induces dependence between the outcome types. We allow missingness to depend on the
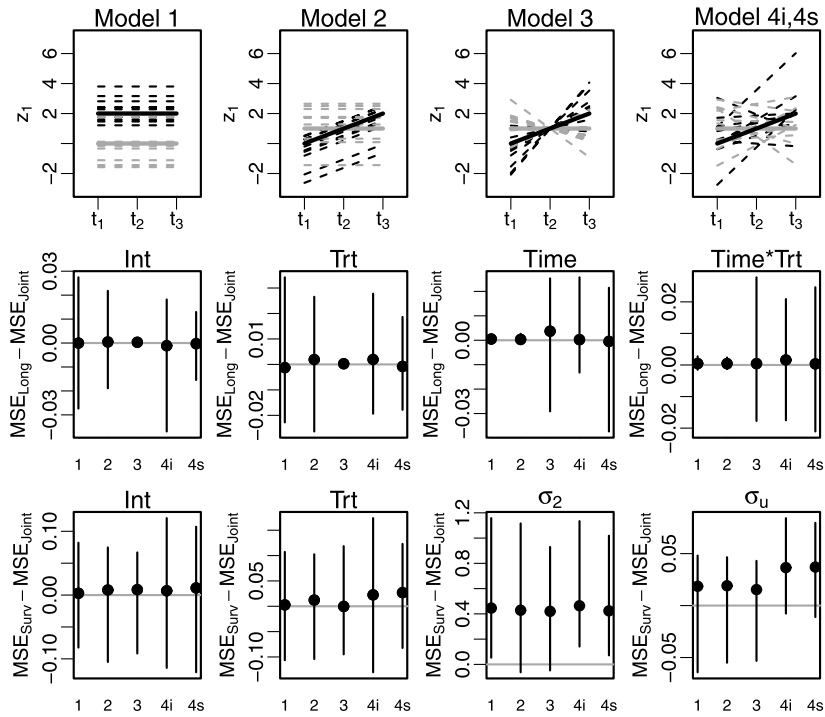
Figure 7. Simulated data and difference in posterior mean squared error (MSE) in joint and separate models fit to simulated data. The top row displays simulated symptom data from each model (fixed effects as solid lines, fixed plus random effects as dashed lines). The middle row displays the posterior MSE difference between joint and symptom-only models, while the bottom row shows the difference between joint and survival-only models. Each vertical segment extends from the lower .025 to the upper .975 quantile of the collection of differences across 100 simulations of each model.

other outcome or latent variables, but not on the missing value itself (i.e., missing at random). Our censoring scenarios emphasize opportunities for benefit from joint modeling. Model 3 from Section 4 generates the data, then various rules censor it. We begin with a reference case of complete symptom and survival data (Scenario 1). That is, each person has the same number of longitudinal observations regardless of survival time. In Scenario 2, we censor symptoms according to survival, so patients with shorter survival contribute fewer symptom observations. In Scenario 3, we censor survival according to the latent variable, so smaller u_i imply greater probability of survival censoring. Finally, in Scenario 4, we do the same for symptoms, so smaller u_i imply more symptom values are missing.

In these simulations, we assume $N = 20$ individuals are observed up to $n = 3$ times (longitudinal). The regression coefficients are $\boldsymbol{\beta}_1 = (1, .5, 0, .5)'$ and $\boldsymbol{\beta}_2 = (1, 1)$ as in Model 3 above, corresponding to no difference between the groups at baseline and time trends of .5 in the treatment arm and $-.5$ in the control arm. To each element of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, we assigned normal priors having mean 0 and variance 100. We fixed $\alpha = 2$ and put $Gamma(1, 1)$ priors on the precision parameters.

Figure 8 displays the difference in MSE between joint and separate models, with plotting symbols that indicate

the censoring scenario. The symptom regression coefficients do not benefit from joint modeling (top row), while the survival intercept (middle left panel) improvement increases with censoring in Scenario 3. The survival error variance σ_2^2 benefits across scenarios, the symptom error variance σ_1^2 improvement increases with censoring under Scenario 4, and the latent parameter variance σ_u^2 benefits modestly across scenarios.

6. DISCUSSION

We began with the questions, "When should we use joint models?" and "How much do we gain by their use?" In answer, we found that information imbalance is the key to unlocking the benefit of joint modeling. Symptom reports dominated survival in the real clinical trial data described in Section 2, with large changes to survival estimates by adding symptom information via joint modeling. In Section 3, the data were balanced, with one observation each of symptoms and survival. In models with improper priors on both submodels, joint modeling did not pay dividends. However, an informative prior on one submodel's coefficients allowed joint modeling to improve the other submodel's coefficient posteriors. Multiple symptom observations in Section 4 did not have the same "unbalancing" effect. We modeled only $n = 3$ symptom observations per person, compared to a me-
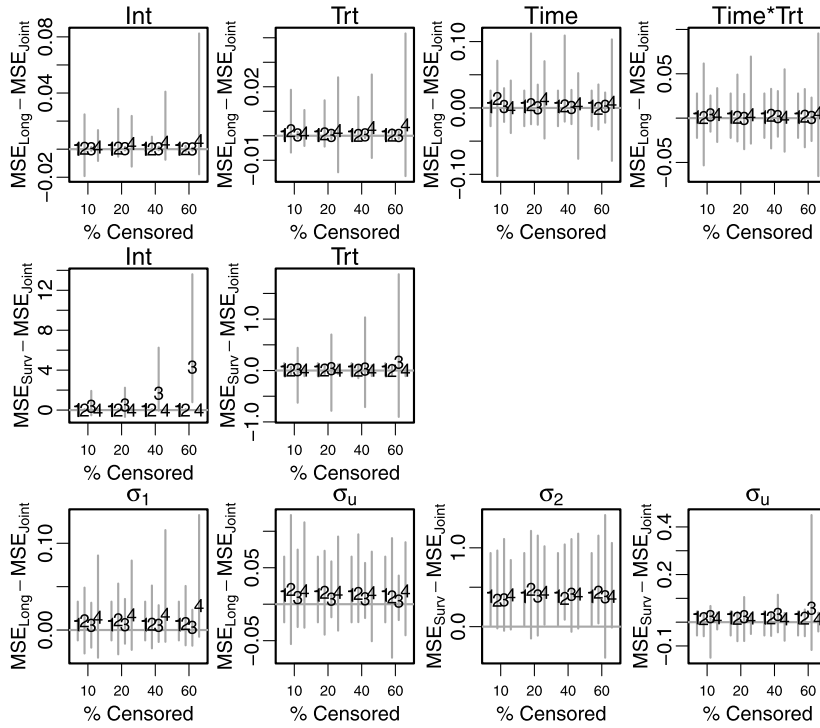
Figure 8. Difference in posterior mean squared error (MSE) from joint and separate models fit to simulated data. Each vertical segment extends from the lower $.025$ to the upper $.975$ quantile of the collection of differences across 100 simulations of each censoring scenario. The plotting symbols (at the median) indicate the censoring scenario, $1 = $ no censoring, $2 = $ symptom censored by survival, $3 = $ survival censored depending on U_i, $4 = $ symptom censored depending on U_i. The top row displays the difference between joint and symptom-only for regression coefficients, the middle row shows the difference between joint and survival-only models for regression coefficients, and the bottom row shows the differences for error and latent effect variance parameters.

dian of 14 symptom observations in the trial data, and much simpler time and treatment structures.

The question of how much we gain from joint modeling also depends on the parameters of interest. Even when joint modeling fails to pay dividends for the treatment effects, we still obtain more precise posteriors for the latent variables and variance parameters. Section A.5 of the appendix contains details of the posterior of u_i; it has smaller variance in the joint model, even with improper priors. Joint modeling provides more information about the random effects and thus improves allocation of variance among σ_u, σ_1, and σ_2. The linking parameter α is roughly the strength of connection between the two data types and determines the magnitude of the difference between joint and separate models. When it is treated as unknown, we found that sufficient sample size (or small error variance) is necessary to identify the signs of the latent variables and linking parameter. Therefore, the choice of submodel to contain α matters.

Joint modeling can improve treatment effect posteriors under certain forms of censoring. In Scenarios 3 and 4 of Section 5, when missingness depended on the latent variable, this was most apparent. This supports the idea that joint

modeling allows one data type to compensate for lack of information in the other via the latent variables.

Some readers may object that these stylized models do not resemble the complex joint models fit in practice. We nonetheless believe it is valuable to strip a model down to a minimal working example and explore its behavior both theoretically and empirically. Obvious next steps include extensions to more realistic models, which will yield *only* to simulation. For example, a generalized linear mixed model for the longitudinal data can be linked with a proportional hazard model for survival. Assuming a parametric baseline hazard, the joint model likelihood will still have a simple closed form. However, the integrals required for marginal posteriors are difficult due to non-linear links and non-Gaussian distributions. Approximation is a possible fix; for example, Gaussian approaches retain the nice features of normal distributions [e.g., 25]. Piecewise exponential models offer a computationally tractable way to generalize our parametric survival models.

Another appealing line of future inquiry is the relationship between latent and fixed effects. Hodges and Reich [9] found that latent parameters can compete with fixed treat-

ment effects in spatial models. It would be useful to investigate this connection in joint models, where latent variables may compete with either set of fixed effects and change the relationship between them.

APPENDIX A. COMPUTATIONAL DETAILS

A.1 Balanced joint model posterior covariance

We begin by recalling the core Bayesian hierarchical modeling result of Lindley and Smith [16], henceforth abbreviated L&S. For n-vector of responses $\mathbf{z}$, p_1-vector of parameters $\boldsymbol{\theta}$, known $n \times p_1$ design matrix $\mathbf{A}_1$, and known $n \times n$ covariance matrix $\mathbf{C}_1$, let the likelihood be $\mathbf{z} \sim N(\mathbf{A}_1\boldsymbol{\theta}, \mathbf{C}_1)$. Then for second-level p_2-vector of parameters $\boldsymbol{\mu}$, known design and covariance matrices $\mathbf{A}_2$ and $\mathbf{C}_2$, respectively, let the prior be $\boldsymbol{\theta} \sim N(\mathbf{A}_2\boldsymbol{\mu}, \mathbf{C}_2)$. L & S showed that the marginal distribution is $\mathbf{z} \sim N(\mathbf{A}_1\mathbf{A}_2\boldsymbol{\mu}, \mathbf{C}_1 + \mathbf{A}_1\mathbf{C}_2\mathbf{A}_1')$ and the posterior is $\boldsymbol{\theta}|\mathbf{z} \sim N(\mathbf{Dd}, \mathbf{D})$ where $\mathbf{D}^{-1} = \mathbf{A}_1'\mathbf{C}_1^{-1}\mathbf{A}_1 + \mathbf{C}_2^{-1}$ and $\mathbf{d} = \mathbf{A}_1'\mathbf{C}_1^{-1}\mathbf{z} + \mathbf{C}_2^{-1}\mathbf{A}_2\boldsymbol{\mu}$. If we assume all covariance parameters (and α) are known, we 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: $\mathbf{z} = (z_{11}, \ldots, z_{1N}, z_{21}, \ldots, z_{2N})'$. The complete $(4 + N)$-vector of parameters $\boldsymbol{\theta} = (\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}, \mathbf{u})'$ contains both fixed and latent effects. The $2N \times (4 + N)$ regression design matrix is

(12)
$$
\mathbf{A}_1 = \begin{pmatrix} \begin{pmatrix} \mathbf{1}_{\frac{N}{2}} & \mathbf{1}_{\frac{N}{2}} \\ \mathbf{1}_{\frac{N}{2}} & -\mathbf{1}_{\frac{N}{2}} \end{pmatrix} & \mathbf{0}_N & \mathbf{0}_N & \mathbf{I}_N \\ \mathbf{0}_N & \mathbf{0}_N & \begin{pmatrix} \mathbf{1}_{\frac{N}{2}} & \mathbf{1}_{\frac{N}{2}} \\ \mathbf{1}_{\frac{N}{2}} & -\mathbf{1}_{\frac{N}{2}} \end{pmatrix} & \alpha\mathbf{I}_N \end{pmatrix},
$$

where $\mathbf{1}_K$ and $\mathbf{0}_K$ are K-vectors of ones and zeros, respectively. The covariance matrix $\mathbf{C}_1$ is block diagonal since conditional on u_i, all the responses are independent, thus $\mathbf{C}_1 = \text{Diag}(\sigma_1^2\mathbf{1}_N', \sigma_2^2\mathbf{1}_N')$. We place independent normal priors on $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ with mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and variance matrices $\sigma_{\beta_1}^2\mathbf{I}_2$ and $\sigma_{\beta_2}^2\mathbf{I}_2$, respectively. That is, we use the same prior variance, $\sigma_{\beta_1}^2$, for both the intercept and treatment effect in the longitudinal model, and a separate prior variance, $\sigma_{\beta_2}^2$, for both parameters of the survival model. We use an independent normal distribution on $\mathbf{u}$, centered at $\mathbf{0}_N$ with variance matrix $\sigma_u^2\mathbf{I}_N$. Then the joint prior on $\boldsymbol{\theta}$ is $N((\boldsymbol{\mu}_1', \boldsymbol{\mu}_2', \mathbf{0}_N')', \text{Diag}(\sigma_{\beta_1}^2\mathbf{1}_2', \sigma_{\beta_2}^2\mathbf{1}_2', \sigma_u^2\mathbf{1}_N'))$.

Using the L& S result, the joint posterior precision matrix for $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{u})'$ is

(13)
$$
\begin{pmatrix} \left(\frac{N}{\sigma_1^2} + \frac{1}{\sigma_{\beta_1}^2}\right)\mathbf{I}_2 & \mathbf{0} & \mathbf{P}'_{12u} \\ \mathbf{0} & \left(\frac{N}{\sigma_2^2} + \frac{1}{\sigma_{\beta_2}^2}\right)\mathbf{I}_2 & \\ \hline \mathbf{P}_{12u} & \left(\frac{1}{\sigma_1^2} + \frac{\alpha^2}{\sigma_2^2} + \frac{1}{\sigma_u^2}\right)\mathbf{I}_N \end{pmatrix}
$$

where

(14)
$$
\mathbf{P}_{12u} = \left(\frac{1}{\sigma_1^2}\begin{pmatrix} \mathbf{1}_{\frac{N}{2}} & \mathbf{1}_{\frac{N}{2}} \\ \mathbf{1}_{\frac{N}{2}} & -\mathbf{1}_{\frac{N}{2}} \end{pmatrix} \;\middle|\; \frac{\alpha}{\sigma_2^2}\begin{pmatrix} \mathbf{1}_{\frac{N}{2}} & \mathbf{1}_{\frac{N}{2}} \\ \mathbf{1}_{\frac{N}{2}} & -\mathbf{1}_{\frac{N}{2}} \end{pmatrix} \right)
$$

we write this joint posterior precision matrix as

(15)
$$
\mathbf{D}^{-1} = \begin{pmatrix} P_1\mathbf{I}_2 & \mathbf{0} & \mathbf{P}'_{12u} \\ \mathbf{0} & P_2\mathbf{I}_2 & \\ \hline \mathbf{P}_{12u} & P_u\mathbf{I}_N \end{pmatrix}
$$

that is, $P_1 = (\frac{N}{\sigma_1^2} + \frac{1}{\sigma_{\beta_1}^2})$, $P_2 = (\frac{N}{\sigma_2^2} + \frac{1}{\sigma_{\beta_2}^2})$, and $P_u = (\frac{1}{\sigma_1^2} + \frac{\alpha^2}{\sigma_2^2} + \frac{1}{\sigma_u^2})$. 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 regression coefficients $\boldsymbol{\beta}$ is

(16)
$$
\begin{pmatrix} Var(\beta_{1\cdot})\mathbf{I}_2 & Cov(\beta_{1\cdot}, \beta_{2\cdot})\mathbf{I}_2 \\ Cov(\beta_{1\cdot}, \beta_{2\cdot})\mathbf{I}_2 & Var(\beta_{2\cdot})\mathbf{I}_2 \end{pmatrix},
$$

where the scalar variances and covariances are given by

(17)
$$
Var(\beta_{1\cdot}|\mathbf{z}_1, \mathbf{z}_2) = c_\beta\left(P_2 - NP_u^{-1}\left(\frac{\alpha}{\sigma_2^2}\right)^2\right),
$$

(18)
$$
Var(\beta_{2\cdot}|\mathbf{z}_1, \mathbf{z}_2) = c_\beta\left(P_1 - NP_u^{-1}\left(\frac{1}{\sigma_1^2}\right)^2\right),
$$

(19)
$$
Cov(\beta_{1\cdot}, \beta_{2\cdot}|\mathbf{z}_1, \mathbf{z}_2) = c_\beta\left(NP_u^{-1}\frac{1}{\sigma_1^2}\frac{\alpha}{\sigma_2^2}\right),
$$

with

$$
c_\beta = \left[P_1P_2 - NP_u^{-1}\left(\left(\frac{\alpha}{\sigma_2^2}\right)^2 P_1 + \left(\frac{1}{\sigma_1^2}\right)^2 P_2\right)\right]^{-1}.
$$

The $N \times 4$ covariance matrix between $\mathbf{u}$ and $\boldsymbol{\beta}$ is given by

(20)
$$
\begin{pmatrix} -(21)\mathbf{1}_{\frac{N}{2}} & -(21)\mathbf{1}_{\frac{N}{2}} & -(22)\mathbf{1}_{\frac{N}{2}} & -(22)\mathbf{1}_{\frac{N}{2}} \\ -(21)\mathbf{1}_{\frac{N}{2}} & (21)\mathbf{1}_{\frac{N}{2}} & -(22)\mathbf{1}_{\frac{N}{2}} & (22)\mathbf{1}_{\frac{N}{2}} \end{pmatrix}
$$

where the referenced scalar covariances are given by

(21)
$$
Cov(\beta_{1\cdot}, u_\cdot) = c_{\beta u}\left(\frac{1}{\sigma_1^2}P_1^{-1}\right)
$$

and

(22)
$$Cov(\beta_{2\cdot}, u_\cdot) = c_{\beta u} \left(\frac{\alpha}{\sigma_2^2} P_2^{-1} \right)$$

with

$$c_{\beta u} = \left[\left(\frac{1}{N\sigma_{\beta_1}^2 + \sigma_1^2} + \frac{\alpha^2}{N\sigma_{\beta_2}^2 + \sigma_2^2} + \frac{1}{\sigma_u^2} \right) \right]^{-1}$$

Notice that when $\alpha > 0$, the treatment group u_i have negative covariance with both β_{11} and β_{12}, while the control group u_i have negative covariance with β_{11} and positive covariance with β_{12}. Again, this is a consequence of the treatment assignment parameterization and balance assumptions in this model.

A.2 Separate model posterior covariances

Turning to the posterior variance of the fixed effects obtained from a longitudinal-only model, we use elements $\mathbf{A}_1$, $\mathbf{C}_1$, $\boldsymbol{\theta}$, and $\mathbf{C}_2$ that are simply the reduced forms obtained by deleting the survival data and parameters. The joint posterior precision matrix of $(\boldsymbol{\beta}_1', \mathbf{u}')'$ is given by

(23)
$$\left(\begin{array}{c|c} P_1 \mathbf{I}_2 & \frac{n}{\sigma_1^2} \left(\begin{array}{cc} \mathbf{1}_{\frac{N}{2}} & \mathbf{1}_{\frac{N}{2}} \\ \mathbf{1}_{\frac{N}{2}} & -\mathbf{1}_{\frac{N}{2}} \end{array} \right)' \\ \hline \frac{1}{\sigma_1^2} \left(\begin{array}{cc} \mathbf{1}_{\frac{N}{2}} & \mathbf{1}_{\frac{N}{2}} \\ \mathbf{1}_{\frac{N}{2}} & -\mathbf{1}_{\frac{N}{2}} \end{array} \right) & \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_u^2} \right) \mathbf{I}_N \end{array} \right).$$

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

(24)
$$Var(\beta_{1\cdot} | \mathbf{z}_1) = \left[P_1 - N \left(\frac{1}{\sigma_1^2} \right)^2 \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_u^2} \right)^{-1} \right]^{-1}.$$

We can obtain the same result by setting $\alpha = 0$ in (17) above, since P_u becomes $(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_u^2})$ and P_2 cancels out of the remaining terms after a bit of algebra. To obtain the analogous result for the posterior variance of $\boldsymbol{\beta}_2$, we re-compute the L&S posterior using a model that involves only the survival submodel,

(25)
$$Var(\beta_{2\cdot} | \mathbf{z}_2) = \left[P_2 - N \left(\frac{\alpha}{\sigma_2^2} \right)^2 \left(\frac{\alpha^2}{\sigma_2^2} + \frac{1}{\sigma_u^2} \right)^{-1} \right]^{-1}.$$

A.3 Balanced joint model posterior mean

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

(26)
$$\mathbf{d} = \left(\begin{array}{c} \frac{1}{\sigma_1^2} z_{1+} + \frac{\mu_{11}}{\sigma_{\beta_1}^2} \\ \frac{1}{\sigma_1^2} \left(z_{1+}^{trt} - z_{1+}^{ctrl} \right) + \frac{\mu_{12}}{\sigma_{\beta_1}^2} \\ \frac{1}{\sigma_2^2} z_{2+} + \frac{\mu_{21}}{\sigma_{\beta_2}^2} \\ \frac{1}{\sigma_2^2} \left(z_{2+}^{trt} - z_{2+}^{ctrl} \right) + \frac{\mu_{22}}{\sigma_{\beta_2}^2} \\ \frac{1}{\sigma_1^2} \mathbf{z}_1 + \frac{\alpha}{\sigma_2^2} \mathbf{z}_2 \end{array} \right).$$

In this expression, z_{1+} is the sum of all the longitudinal observations; z_{1+}^{trt} and z_{1+}^{ctrl} are sums of longitudinal observations from the treatment and control groups, respectively; and z_{2+}, z_{2+}^{trt}, and z_{2+}^{ctrl} are defined analogously for the survival observations. Multiplying (26) by the inverse of (15) yields the vector of posterior mean. We display the latent effect means in Section A.5; the regression coefficient means are as follows:

(27)
$$E(\beta_{11} | \mathbf{z}_1, \mathbf{z}_2) = z_{1+} \frac{Var(\beta_{1\cdot}) - Cov(\beta_{1\cdot}, u_\cdot)}{\sigma_1^2}$$
$$+ \mu_{11} \frac{Var(\beta_{1\cdot})}{\sigma_{\beta_1}^2}$$
$$+ z_{2+} \frac{Cov(\beta_{1\cdot}, \beta_{2\cdot}) - \alpha Cov(\beta_{1\cdot}, u_\cdot)}{\sigma_2^2}$$
$$+ \mu_{21} \frac{Cov(\beta_{1\cdot}, \beta_{2\cdot})}{\sigma_{\beta_2}^2},$$

(28)
$$E(\beta_{12} | \mathbf{z}_1, \mathbf{z}_2) = \left(z_{1+}^{trt} - z_{1+}^{ctrl} \right) \frac{Var(\beta_{1\cdot}) - Cov(\beta_{1\cdot}, u_\cdot)}{\sigma_1^2}$$
$$+ \mu_{12} \frac{Var(\beta_{1\cdot})}{\sigma_{\beta_1}^2}$$
$$+ \left(z_{2+}^{trt} - z_{2+}^{ctrl} \right) \frac{Cov(\beta_{1\cdot}, \beta_{2\cdot}) - \alpha Cov(\beta_{1\cdot}, u_\cdot)}{\sigma_2^2}$$
$$+ \mu_{22} \frac{Cov(\beta_{1\cdot}, \beta_{2\cdot})}{\sigma_{\beta_2}^2},$$

(29)
$$E(\beta_{21} | \mathbf{z}_1, \mathbf{z}_2) = z_{2+} \frac{Var(\beta_{2\cdot}) - \alpha Cov(\beta_{2\cdot}, u_\cdot)}{\sigma_2^2}$$
$$+ \mu_{21} \frac{Var(\beta_{2\cdot})}{\sigma_{\beta_2}^2}$$
$$+ z_{1+} \frac{Cov(\beta_{1\cdot}, \beta_{2\cdot}) - Cov(\beta_{2\cdot}, u_\cdot)}{\sigma_1^2}$$
$$+ \mu_{11} \frac{Cov(\beta_{1\cdot}, \beta_{2\cdot})}{\sigma_{\beta_1}^2},$$

and

(30)
$$E(\beta_{22} | \mathbf{z}_1, \mathbf{z}_2) = \left(z_{2+}^{trt} - z_{2+}^{ctrl} \right) \frac{Var(\beta_{2\cdot}) - \alpha Cov(\beta_{2\cdot}, u_\cdot)}{\sigma_2^2}$$
$$+ \mu_{22} \frac{Var(\beta_{2\cdot})}{\sigma_{\beta_2}^2}$$
$$+ \left(z_{1+}^{trt} - z_{1+}^{ctrl} \right) \frac{Cov(\beta_{1\cdot}, \beta_{2\cdot}) - Cov(\beta_{2\cdot}, u_\cdot)}{\sigma_1^2}$$
$$+ \mu_{12} \frac{Cov(\beta_{1\cdot}, \beta_{2\cdot})}{\sigma_{\beta_1}^2}.$$

A.4 Balanced joint model posterior for the linking parameter α

To put a prior on α and derive its posterior, we begin by writing $\mathbf{A}_1(\alpha)$ to emphasize the dependence of the design matrix on α. Then the joint posterior is

(31)
$$p(\boldsymbol{\theta}, \alpha|\mathbf{z}) = \frac{f(\mathbf{z}|\mathbf{A}_1(\alpha), \boldsymbol{\theta})f(\boldsymbol{\theta}|\boldsymbol{\mu})\pi(\alpha)}{\int\int f(\mathbf{z}|\mathbf{A}_1(\alpha), \boldsymbol{\theta})f(\boldsymbol{\theta}|\boldsymbol{\mu})\pi(\alpha)d\boldsymbol{\theta}d\alpha}$$
$$\propto \exp\left\{-\frac{1}{2}\left[(\boldsymbol{\theta} - \mathbf{Dd})'\mathbf{D}^{-1}(\boldsymbol{\theta} - \mathbf{Dd}) - \mathbf{d}'\mathbf{Dd}\right.\right.$$
$$\left.\left. + \mathbf{z}'\mathbf{C}_1^{-1}\mathbf{z} + (\mathbf{A}_2\boldsymbol{\mu})'\mathbf{C}_2^{-1}(\mathbf{A}_2\boldsymbol{\mu})\right]\right\}\pi(\alpha) .$$

Recall that α appears in elements of $\mathbf{D}$ and $\mathbf{d}$. Obtaining an expression proportional to the *marginal* posterior of α requires integration of this expression with respect to $\boldsymbol{\theta}$. The first part of the exponential is simply a normal kernel in $\boldsymbol{\theta}$, leading to $\int \exp\{-\frac{1}{2}(\boldsymbol{\theta} - \mathbf{Dd})'\mathbf{D}^{-1}(\boldsymbol{\theta} - \mathbf{Dd})\}d\boldsymbol{\theta} \propto |\mathbf{D}|^{1/2}$. Then recall that $\mathbf{D}^{-1}$ in (15) and $\mathbf{d}$ in (26) depend on α, but that the other two terms in the exponential do not contain α. Thus the expression for the posterior of α is straightforward

(32)
$$p(\alpha|\mathbf{z}) = \frac{|\mathbf{D}|^{1/2}\exp\left\{\frac{1}{2}\mathbf{d}'\mathbf{Dd}\right\}\pi(\alpha)}{m(\mathbf{z})}$$

where

$$m(\mathbf{z}) \propto \int |\mathbf{D}|^{1/2}\exp\left\{-\frac{1}{2}\left[\mathbf{z}'\mathbf{C}_1^{-1}\mathbf{z}\right.\right.$$
$$\left.\left. + (\mathbf{A}_2\boldsymbol{\mu})'\mathbf{C}_2^{-1}(\mathbf{A}_2\boldsymbol{\mu}) - \mathbf{d}'\mathbf{Dd}\right]\right\}\pi(\alpha)d\alpha .$$

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

To show that the posterior of α is finite even when $\pi(\alpha)$ is improper, we can show that the integral of the likelihood with respect to α is finite. We first write $f(\mathbf{z}|\alpha, \boldsymbol{\theta})$ to emphasize the elements that depend on α:

(33)
$$f(\mathbf{z}|\alpha, \boldsymbol{\theta}) \propto \exp\left\{-\frac{1}{2}\left((\boldsymbol{\theta} - \mathbf{Dd})'\mathbf{D}^{-1}(\boldsymbol{\theta} - \mathbf{Dd}) - \mathbf{d}'\mathbf{Dd}\right)\right\}$$
$$= \exp\left\{-\frac{1}{2}\left(\boldsymbol{\theta}'\mathbf{D}^{-1}\boldsymbol{\theta} - 2\mathbf{d}'\boldsymbol{\theta}\right)\right\}$$
$$\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\theta}'\mathbf{A}_1'\mathbf{C}_1^{-1}\mathbf{A}_1\boldsymbol{\theta} - 2\mathbf{d}'\boldsymbol{\theta}\right)\right\} .$$

Recall that $\mathbf{C}_1$ and $\boldsymbol{\theta}$ are free of α, and both $\mathbf{d}$ and $\mathbf{A}_1$ are linear in α, making this a quadratic form which is integrable.

A.5 Balanced joint model posterior for latent variables

We next turn our attention to the posterior for an individual's random effect u_i; here we return to assuming the variance parameters σ_1^2, σ_2^2, and σ_u^2 and α are all known and there is no censoring. The conditional posterior distribution of the latent parameter, $p(u_i|z_{1i}, z_{2i}, \boldsymbol{\theta})$, is proportional to

$$\propto \phi(z_{1i}|u_i, \boldsymbol{\theta})\phi(z_{2i}|u_i, \boldsymbol{\theta})\phi(u_i|\boldsymbol{\theta})$$
$$\propto \exp\left\{-\frac{1}{2}\left[\frac{(z_{1i} - \beta_{11} - \beta_{12}trt_i - u_i)^2}{\sigma_1^2}\right.\right.$$
$$\left.\left. + \frac{(z_{2i} - \beta_{21} - \beta_{22}trt_i - \alpha u_i)^2}{\sigma_2^2} + \frac{u_i^2}{\sigma_u^2}\right]\right\}$$
$$\propto \exp\left\{u_i\left(\frac{(z_{1i} - \beta_{11} - \beta_{12}trt_i)}{\sigma_1^2} + \frac{\alpha(z_{2i} - \beta_{21} - \beta_{22}trt_i)}{\sigma_2^2}\right)\right.$$
$$\left. - \frac{u_i^2}{2}\left(\frac{1}{\sigma_1^2} + \frac{\alpha^2}{\sigma_2^2} + \frac{1}{\sigma_u^2}\right)\right\} .$$

To find the posterior mode of u_i, we differentiate the log of this, set the derivative to 0, and solve to obtain the conditional posterior mode of u_i,

(34)
$$\left(\frac{z_{1i} - \beta_{11} - \beta_{12}trt_i}{\sigma_1^2} + \frac{\alpha(z_{2i} - \beta_{21} - \beta_{22}trt_i)}{\sigma_2^2}\right)\sigma_{u_i}^2 ,$$

where $\sigma_{u_i}^2 = (\frac{1}{\sigma_1^2} + \frac{\alpha^2}{\sigma_2^2} + \frac{1}{\sigma_u^2})^{-1}$. 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, $\frac{\partial^2}{\partial u_i^2}\log f(u_i|\boldsymbol{\theta}, z_{1i}, z_{2i}) = -\sigma_{u_i}^{-2}$, 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 $\boldsymbol{\theta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \mathbf{u}')'$ in the joint model when σ_1^2, σ_2^2, σ_u^2, and α are assumed known. We obtain the posterior covariance matrix of $\mathbf{u}$ by inverting (15) and taking the lower left $N \times N$ submatrix of the result,

(35)
$$\mathbf{I}_2 \otimes \left[P_u^{-1}\mathbf{I}_{\frac{N}{2}} + Cov(u., u.)\mathbf{J}_{\frac{N}{2}}\right] ,$$

where

(36)
$$Cov(u., u.) = \frac{2P_u^{-1}\left(\left(\frac{1}{\sigma_1}\right)^2 P_1^{-1} + \left(\frac{\alpha}{\sigma_2}\right)^2 P_2^{-1}\right)}{P_u - N\left(\left(\frac{1}{\sigma_1}\right)^2 P_1^{-1} + \left(\frac{\alpha}{\sigma_2}\right)^2 P_2^{-1}\right)} ,$$

and $\mathbf{J}_K$ is a $K \times K$ matrix of ones. Notice that the u_i for subjects in the same treatment group have correlation $Cov(u., u.)/(P_u^{-1} + Cov(u., u.))$, 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 (35) to obtain

$$(37) \qquad \mathbf{I}_2 \otimes \left[\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_u^2} \right)^{-1} \mathbf{I}_{\frac{N}{2}} + Cov(u., u.|\mathbf{z}_1)\mathbf{J}_{\frac{N}{2}} \right],$$

where

$$Cov(u., u.|\mathbf{z}_1) = \frac{2 \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_u^2} \right)^{-1} \left(\left(\frac{1}{\sigma_1} \right)^2 P_1^{-1} \right)}{\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_u^2} \right) - N \left(\left(\frac{1}{\sigma_1} \right)^2 P_1^{-1} \right)}.$$

When we compare the results of taking limits in both of these expressions as $\sigma_{\beta_1}^2$ and $\sigma_{\beta_2}^2$ approach infinity,

$$(38)$$
$$Var(u_i|\mathbf{z}_1, \mathbf{z}_2) \to \frac{1}{\sigma_u^2} - \left(\frac{\sigma_2^2 + \alpha^2 \sigma_1^2}{\sigma_2^2 \sigma_u^2 + \alpha^2 \sigma_u^2 \sigma_1^2 + \sigma_2^2 \sigma_1^2} \right) \left(1 + \frac{2}{N} \right)$$

and

$$(39) \qquad Var(u_i|\mathbf{z}_1) \to \frac{1}{\sigma_u^2} - \left(\frac{1}{\sigma_u^2 + \sigma_1^2} \right) \left(1 + \frac{2}{N} \right).$$

The difference in limiting posterior variances (joint minus longitudinal-only) is negative

$$Var(u_i|\mathbf{z}_1, \mathbf{z}_2) - Var(u_i|\mathbf{z}_1)$$
$$\to \left(1 + \frac{2}{N} \right) \left[\frac{-\alpha^2(\sigma_1^2)^2}{(\sigma_u^2 + \sigma_1^2)(\sigma_2^2 \sigma_u^2 + \alpha^2 \sigma_u^2 \sigma_1^2 + \sigma_2^2 \sigma_1^2)} \right],$$

indicating a smaller posterior variance for u_i in the joint model, even with improper priors.

We can also compute the posterior mean of the random effects in the joint model with known variances. The posterior mean of the jth random effect in the treatment group $E(u_j|\mathbf{z}_1, \mathbf{z}_2)$ is

$$(40) \qquad \frac{1}{\sigma_1^2} \qquad \left(Var(u.)z_{1j} + Cov(u., u.)z_{1+}^{trt} \right.$$
$$- \frac{Cov(\beta., u.)}{P_1} \left[\frac{z_{1+}}{\sigma_1^2} + \frac{\mu_{11}}{\sigma_{\beta_1}^2} + \frac{(z_{1+}^{trt} - z_{1+}^{ctrl})}{\sigma_1^2} + \frac{\mu_{12}}{\sigma_{\beta_1}^2} \right] \right)$$
$$+ \frac{\alpha}{\sigma_2^2} \qquad \left(Var(u.)z_{2j} + Cov(u., u.)z_{2+}^{trt} \right.$$
$$- \frac{Cov(\beta., u.)}{P_2} \left[\frac{z_{2+}}{\sigma_2^2} + \frac{\mu_{12}}{\sigma_{\beta_2}^2} + \frac{(z_{2+}^{trt} - z_{2+}^{ctrl})}{\sigma_2^2} + \frac{\mu_{22}}{\sigma_{\beta_2}^2} \right] \right).$$

where $Var(u.) = (P_u^{-1} + Cov(u., u.))$. Notice that this has an intuitive interpretation similar to the regression coefficients in (27). 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, α ensures 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

[1] BROWN, E. R., J. G. IBRAHIM, and V. DEGRUTTOLA (2005). A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics 61*, 64–73. MR2129202

[2] CHEN, L. M., J. G. IBRAHIM, and H. CHU (2011). Sample size and power determination in joint modeling of longitudinal and survival data. *Statistics in Medicine 30*(18), 2295–309. MR2830009

[3] CHEN, M.-H. and P. GUSTAFSON (2011). Bayesian methods in survival analysis. *Lifetime Data Analysis 17*, 1–2. MR2764576

[4] FAUCETT, C., N. SCHENKER, and R. M. ELASHOFF (1998). Analysis of censored survival data with intermittently observed time-dependent binary covariates. *Journal of the American Statistical Association 93*(442), 427–437.

[5] FAUCETT, C. and D. C. THOMAS (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine 15*, 1663–1685.

[6] GESKUS, R. B. (2014). Which individuals make dropout informative? *Statistical Methods in Medical Research 23*, 91–106.

[7] HANSON, T. E., A. J. BRANSCUM, and W. O. JOHNSON (2011). Predictive comparison of joint longitudinal-survival modeling: a case study illustrating competing approaches. *Lifetime Data Analysis 17*(1), 3–28. MR2764577

[8] HATFIELD, L. A., M. E. BOYE, M. D. HACKSHAW, and B. P. CARLIN (2012). Multilevel Bayesian models for survival times and longitudinal patient-reported outcomes with many zeros. *Journal of the American Statistical Association 107*(499), 875–85. MR3010877

[9] HODGES, J. S. and B. J. REICH (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician 64*(4), 325–334. MR2758564

[10] HOGAN, J. (2009). Comments on: Missing data methods in longitudinal studies: a review. *Test 18*, 59–64. MR2495962

[11] HOLTE, S. E., T. W. RANDOLPH, J. DING, J. TIEN, R. S. MCCLELLAND, J. M. BAETEN, and J. OVERBAUGH (2012). Efficient use of longitudinal CD4 counts and viral load measures in survival analysis. *Statistics in Medicine 31*(19), 2086–2097. MR2956063

[12] HU, C. and M. E. SALE (2003). A joint model for nonlinear longitudinal data with informative dropout. *Journal of Pharmacokinetics and Pharmacodynamics 30*(1), 83–103.

[13] HUANG, X., G. LI, R. M. ELASHOFF, and J. PAN (2011). A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. *Lifetime Data Analysis 17*(1), 80–100. MR2764585

[14] IBRAHIM, J. G., M.-H. CHEN, and D. SINHA (2004). Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Statistica Sinica 14*, 863–883. MR2087976

[15] KENWARD, M. G., G. MOLENBERGHS, and H. THIJS (2003). Pattern-mixture models with proper time dependence. *Biometrika 90*(1), 53–71. MR1966550

[16] LINDLEY, D. V. and A. F. M. SMITH (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B (Methodological) 34*, 1–41. MR0415861

[17] PORTER, M. (2010). What is value in health care? *New England Journal of Medicine 363*(26), 2477–2481.

[18] PROUST-LIMA, C., M. SÉNE, J. M. TAYLOR, and H. JACQMIN-GADDA (2014). Joint latent class models for longitudinal and time-to-event data: A review. *Statistical Methods in Medical Research 23*, 74–90.

[19] RIZOPOULOS, D., G. VERBEKE, and G. MOLENBERGHS (2010). Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. *Biometrics 66*(1), 20–9. MR2756687

[20] SUN, L., X. SONG, and J. ZHOU (2011). Regression analysis of longitudinal data with time-dependent covariates in the presence of informative observation and censoring times. *Journal of Statistical Planning and Inference 141*(8), 2902–2919. MR2787754

[21] SWEETING, M. J. and S. G. THOMPSON (2011). Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal 53*(5), 750–63. MR2861484

[22] TSIATIS, A. A., V. DEGRUTTOLA, and M. S. WULFSOHN (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association 90*(429), 27–37.

[23] VERBEKE, G., S. FIEUWS, G. MOLENBERGHS, and M. DAVIDIAN (2014). The analysis of multivariate longitudinal data: A review. *Statistical Methods in Medical Research 23*, 42–59.

[24] WULFSOHN, M. S. and A. A. TSIATIS (1997). A joint model for survival and longitudinal data measured with error. *Biometrics 53*(1), 330–9. MR1450186

[25] XIE, Y. and B. P. CARLIN (2006). Measures of Bayesian learning and identifiability in hierarchical models. *Journal of Statistical Planning and Inference 136*, 3458–3477. MR2256284

[26] XU, J. and S. L. ZEGER (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics 50*(3), 375–387. MR1856332

[27] ZHANG, S., P. MULLER, and K. A. DO (2010). A Bayesian semiparametric survival model with longitudinal markers. *Biometrics 66*(2), 435–43. MR2758823

[28] ZHU, H., J. G. IBRAHIM, Y.-Y. CHI, and N. TANG (2012). Bayesian influence measures for joint models for longitudinal and survival data. *Biometrics 68*(3), 954–964.

Laura A. Hatfield
Department of Health Care Policy
Harvard Medical School
180 Longwood Avenue
Boston, MA 02115
USA
E-mail address: hatfield@hcp.med.harvard.edu

James S. Hodges
Division of Biostatistics
University of Minnesota School of Public Health
2221 University Ave SE, Suite 200
Minneapolis, MN 55414
USA
E-mail address: hodge003@umn.edu

Bradley P. Carlin
Division of Biostatistics
University of Minnesota School of Public Health
Mayo Mail Code 303
Minneapolis, MN 55455
USA
E-mail address: brad@biostat.umn.edu