Bayesian Joint Modeling of Patient-Reported Outcomes and Survival Information: Discussion

A. Lawrence Gould
Merck Research Laboratories
351 N. Sumneytown Pike, North Wales, PA 19454

Abstract

This is a discussion of four presentations made during the session on Joint Modeling at the Joint Statistics Meetings (Session 321, 31 July 2012):

Learning and Information in Bayesian Joint Models for Longitudinal and Survival Data by Laura A. Hatfield, Harvard Medical School; James Hodges, University of Minnesota; Bradley P Carlin, University of Minnesota

Comparative Assessment of Joint Versus Conventional Modeling of Longitudinal and Survival Endpoints: A Reanalysis of P3 Oncology Data by Mark Boye; Joseph Ibrahim, The University of North Carolina at Chapel Hill; Ming-Hui Chen, University of Connecticut; Ping Wang, Eli Lilly and Company; Wei Shen, Eli Lilly and Company; Danjie Zhang, University of Connecticut

Joint Modeling of Longitudinal and Survival Data with Missing and Left-Censored Time-Varying Covariates by Ryan May, The EMMES Corporation

A Bayesian Joint Model for DAS28 Scores and Time-to-Dropout Data by Violeta Hennessey, Amgen, Inc.

Key Words: Bayes, joint modeling

1. Introduction

Longitudinal trials observe subjects over a period of time until completion of a planned measurement schedule or withdrawal from the trial for reasons that may informative such as toxicity, cure, or lack of efficacy, or for reasons that may not be informative such as administrative withdrawals or termination of the trial before all subjects could complete their scheduled observations. Joint modeling methods combine the longitudinal measurements with information about withdrawals to draw valid conclusions about intervention effects. Simply ignoring the missing information is not a viable option. Bayesian methods usually form the basis for analyses because they provide a convenient way to incorporate random effects at various stages of the models.

Joint modeling methods address either or both of two different questions, (1)"What is the effect of the intervention on the trajectory of measurements over time when withdrawals are considered?", and (2) "What is the effect of a subject's trajectory of measurements on the likelihood of that subject withdrawing from the trial for some reason?"

The longitudinal observations to which joint modeling methods are applied typically are modeled as the sum of a fixed effect and a random error, for example, the j-th longitudinal observation on the i-th subject, made at time t_{ij}, can be written as

$$Y_{ij} = Y_i(t_{ij}) = f(t_{ij}; \vec{X}_{ij}, \vec{\theta}_i) + \varepsilon_{ij} \text{ where } \vec{\theta}_i \sim g(\vec{\theta}_i; \vec{\Theta}) \qquad (1)$$

For any subject (fixed i), the expected value of Y_{ij}, $f(T_i; \vec{X}_i, \vec{\theta}_i)$, depends on covariates $\vec{X}_{ij}$ observed for that subject until measurement occasion j, with subject-specific

parameter(s) $\vec{\theta}_i$ that are drawn from a distribution of potential values characterized by a parameter vector $\vec{\Theta}$. The quantity ε_{ij} represents measurement error. In addition to the longitudinal observations, there are (potential) event times T_i for each subject, with distribution

$$T_i \sim h(T_i; f(T_i; \vec{X}_i, \vec{\theta}_i), \vec{Z}_i, \vec{\xi}_i) \text{ where } \vec{\xi}_i \sim k(\vec{\xi}_i; \vec{\Xi}) \quad (2)$$

The event time distribution for subject i could depend on the expected trajectory for that subject up to that time, $f(T_i; \vec{X}_i, \vec{\theta}_i)$, or on the observed trajectory, $\{Y_i(t_{ij}), j = 1, \ldots, m_i, m_i = \max(j \mid t_{ij} \le T_i\}$. Either way, models (1) and (2) share parameters and covariates wholly or in part, which is why they are called joint models. Missing longitudinal values (values of Y_{ij} scheduled for observation after an event/withdrawal time T_i, but not actually observed) are assumed to be missing at random so that the actual or expected trajectory of observations is potentially predictive of withdrawal.

2. Comments on Individual Presentations

2.1 Hatfield *et al*

This presentation considers how the choices of prior distribution affect inferences about the effects of longitudinal outcomes and survival information when these two quantities are jointly and separately distributed, assuming no informative censoring.

If z_{1i} = mean of n longitudinal observations for subject i, i = 1, …, N and

z_{2i} = log survival time for subject i

then the joint likelihoods are

$$z_{1i} \sim N(\beta_{11} + \beta_{12}\, trt_i + u_i, \sigma_1^2/n), \; u_i \sim N(0, \sigma_u^2)$$

$$z_{2i} \sim N(\beta_{21} + \beta_{22}\, trt_i + \alpha u_i, \sigma_2^2/n$$

where u_i represents the shared random effect of subject i. The corresponding separate likelihoods are almost identical, except that $\alpha \equiv 1$. Normal prior distributions are assumed for $\vec{\beta}_1 = (\beta_{11}, \beta_{12})$ and $\vec{\beta}_2 = (\beta_{21}, \beta_{22})$: $\vec{\beta}_j \sim N(0, \eta_j \mathbf{I}_2)$, j = 1, 2. The key message of the presentation is that the choice of joint or separate likelihoods and the choice of priors (values of η_1 and η_2) can affect the inferences.

When applied to a set of data from a clinical trial, inferences about the longitudinal observations turned out not to depend on whether joint or separate or likelihoods were used. However, the inferences about the survival times did depend on the choice of model, as illustrated in Figure 1.

On the other hand, when joint modeling was employed, the *a priori* assumption about the variance of the treatment effect for the survival time also affected the posterior variance of the treatment effect for the longitudinal observations, depending on the value of the linking parameter α. Figure 2 illustrates the effect of different choices of prior distribution variance on the posterior variance of the treatment effect. The essential message here is that the benefit of joint modeling depends on the choice of prior distributions for the parameters.

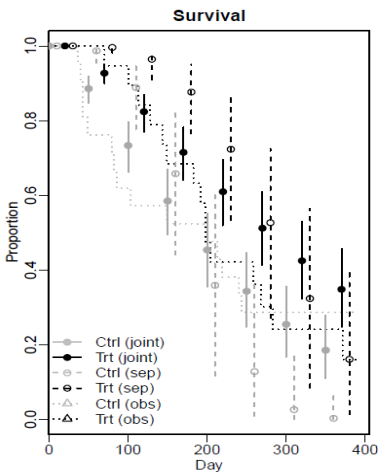

Figure 1. Prediction intervals and expected survival proportions on each treatment

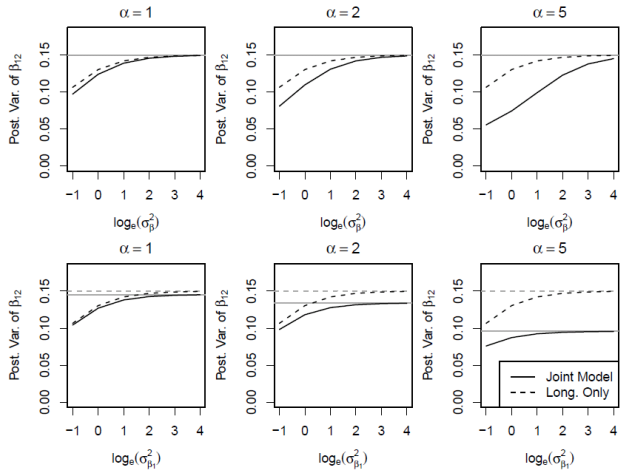

Figure 2. Variance of β_{12} as a function of the variance of β_{22}. Top row corresponds to $\eta_2 = \eta_1$, bottom row corresponds to $\eta_2 = 2$ (fixed)

2.2 Boye *et al*

This presentation addressed variations of longitudinal models like (1) with

$$f(T_i; \vec{X}_i, \vec{\theta}_i) = \theta_{0i} + \theta_{1i}\, t_{ij} + \gamma\, z_i, \quad \vec{\theta}_i \sim N(\boldsymbol{\Theta}, \boldsymbol{\Sigma})$$

and survival models with hazard functions of the form

$$\lambda_i(t_{ij}) = \exp\{\beta_1(\theta_{0i} + \gamma z_i) + \beta_2\theta_{1i}t_i + \alpha z_i\} \qquad \text{(Linear Random Effect = LRE)}$$

$$\lambda_i(t_{ij}) = \exp\{\beta(\theta_{0i} + \gamma z_i + \theta_{1i}t_i) + \alpha z_i\} \qquad \text{(Linear Trajectory = LT)}$$

and compared the gain in the fit of survival models from using longitudinal information by means of the change in Akaike Information Criterion (AIC),

$$\Delta AIC = AIC_{\text{survival}} - AIC_{\text{survival | longitudinal}}$$

The effect of misspecification of the longitudinal model on the findings of the survival analysis were evaluated by simulation:

1. A set of longitudinal data were generated using a LRE model and also a QRE (Quadratic Random Effect) model
2. A set of survival data were generated using a similar model (say method A)
3. The resulting data were analyzed using another method (say method B) that assumes some other model for the longitudinal data
4. The values of ΔAIC were compared using the various longitudinal models.

Figures 3 and 4 summarize the results of the comparisons of the models. Figure 3 summarizes the ΔAIC values when the survival data were constructed assuming that the longitudinal data were generated by a LRE model. The best results were obtained when the survival data were fit assuming the 'correct' longitudinal model. LRE2 is the same as the LRE model, except that a 2-piece exponential model was assumed for the baseline hazard function, and TS refers to a 2-stage model. Other reasonable longitudinal models gave reasonable results except for LOCF (Last Observation Carried Forward), which provided materially worse fits to the data. Figure 4 (note scale difference from Fig. 3) provides similar findings when the data were generated assuming a QRE longitudinal effects model. Assuming the longitudinal data were generated from an LRE model

produced results with the least AIC difference from the correct model. Assuming LOCF still gave the worst fits.

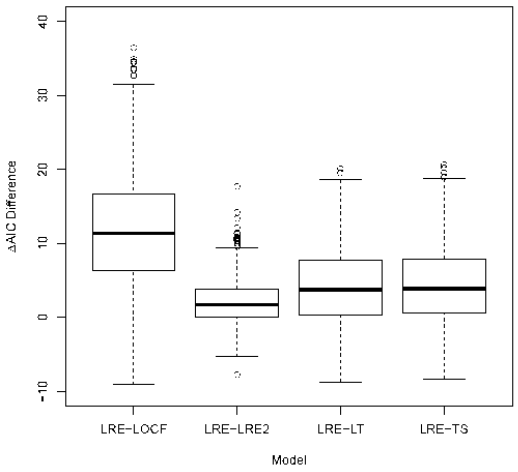

Figure 3. ΔAIC when LRE is correct longitudinal model

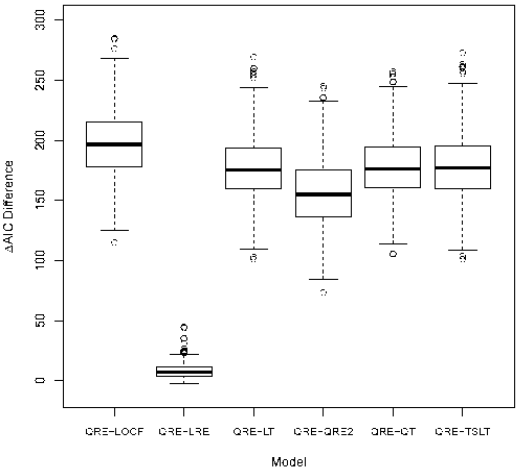

Figure 4. ΔAIC when QRE is correct longitudinal model.

2.3 May

This presentation was based on a simple mixed effect model linear in time for longitudinal data. A distinction was made between selection and shared parameter models. With selection models, which this presentation used, the survival model depends on fixed and random effects from the longitudinal model. With shared parameter models, the survival model depends on fixed effects or on random effects, but not both.

The analysis problem considered was the prediction of survival (progression to AIDS) in an HIV-infected population based on trajectories of CD4 counts and viral load. The complicating factor was that 27% of VL (viral load) data were missing, and 17% were left-censored (below the limit of detectability). The presentation described an approach for incorporating these potentially informative data.

For the censored VL findings, the analysis started with a normal prior distribution for VL, but truncated at the limit of detection (LD) so that the density was positive only if VL < LD. The missing data were assumed to be missing at random.

A 2-stage strategy was used to fit the models: the longitudinal model was fit first, then the results were used to fit the survival model. The key findings were that observations with missing or left-censored covariates can strongly influence posterior estimates of the parameters of the survival distribution when joint models are used. As it turned out, the point estimates of the effect of the HAART (Highly Active Anti Retroviral Treatment) calendar period were not greatly affected, but the precision of the estimates was better when all of the data were included.

2.4 Hennessey

This presentation considered the presence of competing withdrawal risks (toxicity and lack of efficacy), a change point model for the longitudinal observations, and the use of latent variables to link the longitudinal and survival information.

The longitudinal model assumed that the slope changed at Week 12,

$$E(Y_i(t_{ij})) = \beta_{0i} + \beta_{1i}t_{ij} \qquad t_{ij} < 12$$

$$\beta_{0i} + \beta_{1i}\times 12 + \beta_{2i}(t_{ij} - 12) \qquad t_{ij} > 12$$

$$\beta_{mi} \sim N(\mu_{mi}, \sigma_m^2) \qquad \mu_{mi} = \gamma_{3m}\text{-}2 + \gamma_{3m-1}z_{1i} + \gamma_{3m}z_{2i} \qquad z_1 = E \qquad z_2 = E+M$$

Vague priors were assumed for γ and σ_m^2. The event times (times to withdrawal for toxicity or lack of efficacy) T_{AE}, T_{Eff} were assumed to have lognormal distributions with

$$E(T_{xi}) = \mu_{xi} = \phi_{x1} + \phi_{x2}z_{1i} + \phi_{x3}z_{2i} + W_{xi} \quad x = AE \text{ or } Eff$$

W = linear function of longitudinal model parameters

Figures 5 (before week 12) and 6 (after week 12) summarize the results from using joint and separate models; MTX, ETAN, and MTX+ETAN were the interventions used in the trial. There clearly was a substantial difference between the findings before and after week 12. However, the analysis results were essentially unaffected by the choice of model (joint or separate).

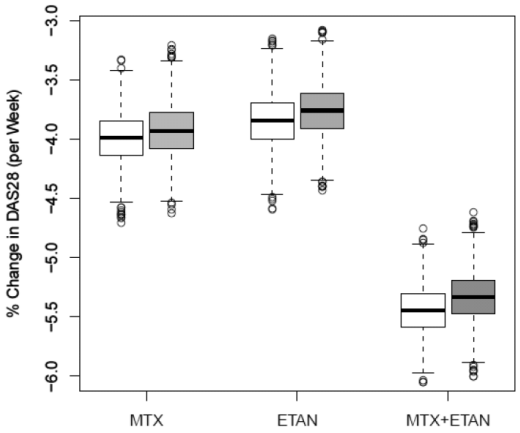

Figure 5. Pre-Week 12 findings (open = joint, shaded = separate)

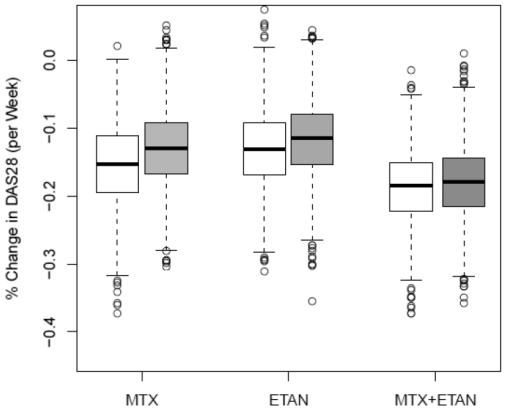

Figure 6. Post-Week 12 findings (open = joint, shaded = separate)

3. Summing Up

The presentations addressed different aspects of the application of joint models. Generally, simple linear time trajectories were used, which the results of Boye *et a*l and Hennessey suggest may be unrealistic. It is possible, but computationally expensive, to account for left censoring and missing data. It is not, however, clear, how much benefit actually would be realized in practice; this may depend on the circumstances and it would be useful to have some guidelines as to when the censored and missing data actually would be important when joint models are used. h

With careful attention to *a priori* assumptions in the context of Bayesian analyses, joint modeling can provide increased insight, although how much may be very dependent on the actual situation. The precision of the posterior inferences depends on the models used for the longitudinal data

As often is the case with evolving areas of development, there is considerable scope for future work.