Rejoinder for "Predictive comparison of joint longitudinal-survival modeling: a case study illustrating competing approaches"

**Wesley Johnson · Adam J. Branscum ·
Timothy E. Hanson**

We were pleased that our paper elicited stimulating responses by Professors Craiu (RC) and Taylor (JT) that both complement our work and provide us the opportunity to further elaborate upon it. We confess that our original intention was to simply develop novel methodology for the joint modeling problem and to compare semi-parametric models using a predictive criterion. The choice of data was more for convenience, since it had been recently employed in published reports, than by design. Because of the obvious difficulties associated with these data, many of which are highlighted by JT and RC, the final form of the paper evolved into much more than it would have been under our original intentions.

In what follows, we address three connected themes that were described by the discussants: scientific purpose of the analysis, model selection and model fit, and alternative avenues to data analysis and inference. We emphasize how pleased we are to respond to such thoughtful comments.

W. Johnson
Department of statistics, University of california Irvine,
Irvine, CA, USA

A. J. Branscum
Departments of Biostatistics, Statistics and Epidemiology University of Kentucky,
Lexingtons, KY, USA

T. E. Hanson (✉)
Division of Biostatistics, University of Minnesota,
Minneapolis, MN, USA
e-mail: hanson@biostat.umn.edu

1 Scientific purpose of the analysis

JT wondered about the scientific goals of the analysis, and his hypothesized description was correct. As stated in our paper, the goal of the analysis was to provide a description of the joint distribution of lifetimes and reproduction, with particular emphasis on examining the connection between patterns of egg laying, or more generally fecundity, and longevity. In our analysis, we presented a variety of predictive densities for time to death among flies with egg laying profiles that were the same as those for flies in the data (with zero counts carried forward after death). While we used these densities to investigate "how well" they predicted the actual death times for flies in the study, a more scientifically oriented presentation would have attempted to characterize how different egg laying profiles might be associated with longevity. For example, once a model has been selected and its fit deemed reasonable, we can compare the predictive densities for flies with different types of reproductive histories. We stress here that our predictive densities are meant to characterize the life expectancy not for the flies in the study, but for flies that will have egg laying patterns that coincide with those for the flies in the study. This is of scientific interest in that it supports the scientist's agenda to learn how life expectancy changes for various types of reproduction histories.

To illustrate this idea, consider Figs. 3, 4 and 5, which present egg laying trajectories and corresponding predictive densities for flies 38, 45 and 6, respectively, under the parametric CO model. We focus on the predictive densities based on LOCF. Assuming that the particular model used is reasonable, fly 38 would be considered anomalous because it lived much longer than expected. The three predictive densities can be used to compare survival under the corresponding three (quite different) reproduction profiles. The model-based prediction is that medflies with egg production like fly 38 (with a long stretch of no eggs at the beginning of life) will die early relative to medflies with profiles like fly 45 (with fairly consistent production throughout life) and fly 6 (with a stretch of no production late in life). In addition, medflies like fly 6 are expected to die earlier than flies like 45.

Perhaps more interesting than comparing these predictive densities (from a scientific point of view) would be to look at estimated hazard functions, which condition on the history up to the present time, as they evolve through time. One could also construct different types of egg laying profiles and compare the corresponding hazard functions. This would be more interesting in other problems since no one is going to be too concerned about early death of a fruit fly. But if we were monitoring blood pressure for dialysis patients who are at considerable risk of early death, physicians might be particularly interested in seeing the pattern in a patient's hazard as a function of blood pressure, especially if it suddenly rose quickly.

A potential problem with all of our models, and given our stated goals, is the way the longitudinal process affects survival. All of the models for survival depend on the longitudinal profile in rather simplistic ways e.g., the hazard at time t depends on the covariate process only through a single number (in the case of the CO model, through $\int_0^t e^{x(s)\beta} ds/t$). Since we are thinking about profile shape being associated with time of event, it might make more sense to go beyond this single number summary of the covariate process for each time and instead condition on variables that describe the shape of time course trends. A simple possibility in the medfly study, assuming

the subject-specific log gamma shape is sensible, would be to condition the time of event on the corresponding two random effects, namely to model $S(t \mid b_{i1}, b_{i2})$. We could posit $h(t \mid \{x(s) : \forall s\}) = h(t \mid b_i' \beta)$, which would be $h_0(t)e^{b_i' \beta}$ in the case of the Cox model. This is a special case of the model proposed by Henderson et al. (2000) and subsequently used by Guo and Carlin (2004). This is no longer a traditional time dependent covariate survival model since it amounts to conditioning on objects that describe the entire trajectory for each fly. This type of model would extend to regressing survival on the random coefficients of the spline representations for profiles, or with the use of other basis functions (for example, the approximate principle components approach of Ding and Wang 2008). While the PO model could be analogously defined, the CO model could not. A natural extension of this would be to model these random effects through a Dirichlet process mixture (e.g., Li et al. 2010), which would allow for clustering of different types of profiles. Of course, with such a model we would lose the ability to make predictions or look at hazards conditional on history up to present, including knowledge of survival to present.

2 Model selection and model fit

Both discussants correctly point out that assessment of model selection and model fit should be guided by the goals of the analysis. A standard method of Bayesian model selection involves the Bayes Factor (BF). Our method parallels that approach through the use of the LPML since a Bayes Factor analogue is obtained by exponentiating the difference in LPMLs for two models. As with BFs, for the models under consideration, the best model selected according to this criterion may not fit the data very well. Alternative approaches like the deviance information criterion (DIC) or the Bayesian information criterion (BIC), which both involve a tradeoff between goodness of fit and complexity, are also problematic for determining goodness of fit. Because these measures lack standard reference values, it is difficult to determine from them whether particular models achieve sufficient predictive capability. Our method of model selection, LPML, is predictive in that its value will be large if the observed data are predicted well. This is in the sense that the n conditional predictive ordinates (CPO) are all large (each CPO is conditional on the actual egg laying history for a fly in the data).

JT reiterates that our CPOs are conditional on the history but not on the fact the fly was alive for all the time right up to the time of death. We could have computed alternative CPOs based on this further conditioning, which would result in a different model selection criterion. Our criterion is based on an inferential goal of predicting survival for flies *like the ones in the data*, for which we hypothesize egg trajectories that mimic the actual ones in the data. The CPO is a measure of the plausibility (as determined by the particular model) of the death time for the observed fly, which may be regarded as a sample of size one from the times for the class of flies with trajectories like the observed one. If we computed the CPOs conditional on all knowledge just prior to death, and of course deleting the current case from the data, then we would be seeing how plausible the observed death times were relative to the part of

the model that predicts death time for individual flies in the data. We agree that the utility of predictive discrepancy measures, as detailed for longitudinal data analysis by Crespi and Boscardin (2009), should be explored in future research on assessment of joint models. It would indeed be interesting to see how these criteria might differ in terms of model selection.

In addition, as pointed out by RC and alluded to above, all the CPOs for a selected model could be quite small, possibly indicating that none of the models would fit. Moreover, it could be the case that a few long-living flies are not predicted well by one or more models. Since we have non-parametric baseline hazards, we would not expect that part of the model to fail to account for a long tailed distribution. We expect the problem would be related to the modeling of the relationship between the covariate process and survival, or to a poorly modeled covariate process. Perhaps our suggested modeling at the end of the previous section could help in this regard.

Alas, we made no formal attempt in our paper to assess goodness of fit. However we did consider very flexible longitudinal models that provided better fitting trajectories compared to the simple two-term random effects model for egg counts. We also looked at predictive densities for particular flies and noted whether the observed death would have been reasonably predicted by the given models. For example, referring again to Figures 3, 4 and 5 just for illustration, we found that survival for fly 38 is poorly predicted under this model as the predictive density is anticipating death during the first 22 days when no eggs were laid. Fly 45 starts laying eggs much sooner followed by a longer period of productivity with a single zero count just before death, and the predictive density correspondingly anticipates a later death. Fly 6 is similar to fly 45 except that is has a sizeable window of zero counts towards the end of its life. The zero counts being carried forward after the actual death lead to spikes in the predictive densities for flies 6 and 45 after their deaths. We see that even though our model selection criterion favors LOCF to joint modeling or imputation, flies with many zero counts in the beginning of life may be poorly predicted.

In retrospect, we would have fit the longitudinal data first to assess quality of fit to the data (using standard diagnostics, e.g., normal plots, residual plots, outlier sensitivity analysis). However in our literature search, we found that methodology for assessing the goodness of fit of longitudinal models may be lacking. In particular, Sinharay and Stern (2003) found that identifying lack of fit through predictive model checking is challenging in a normal–normal hierarchical model (which has the same structure as the linear mixed model we considered for the medfly egg counts) unless the extent of the violation to be detected is large, or error variability small. For a recent application of predictive model checking in a somewhat related setting with spatially correlated sea surface temperature data, see Lemos and Sansó (2009).

With regard to the survival part of the model, we have attempted to evaluate model fit in a semi-parametric survival analysis with time dependent covariates under LOCF (Hanson et al. 2009). This involved assessments using standard residuals and integrated Cox–Snell residual plots, in addition to calculating LPML for selecting among different families for the baseline survivor function as we have done here (these plots were not included in the published version of that paper). We believe that model assessment should be done in two stages, with consideration of the longitudinal part first,

and then conditional on the selected longitudinal model, assessment of the survival part. It is not clear to us how to proceed with this second stage assessment, except in the case of actual conditioning on either the raw or imputed longitudinal processes. Even in these instances, we believe that additional attention is required. We expect the methods found in Gelman et al. (1996), notwithstanding the findings in Sinharay and Stern (2003), will ultimately be found to be useful in applications of joint modeling. Recently, Rizopoulos et al. (2009) have developed residual plots for assessing fit in the joint modeling context. Furthermore, Rizopoulis has contributed the free R package JM to the Comprehensive R Archive Network that fits certain classes of joint models relatively painlessly.

3 Alternative modeling

RC provides a useful alternative model and analysis based on the observation that the longitudinal count data, when looked at by day, are skewed. This observation establishes once again that hindsight is 20–20. We were too busy developing our methodology, and subsequently attempting to fix the obvious problem with some fly trajectories failing to fit the observed data, to notice that the daily transformed fly counts were skewed. We thus very much appreciate this interesting observation and subsequent approach to analyzing the medfly data based on a median zero error distribution that allows for skewness. RC has shown that, using a piecewise exponential model for survival, and with mean (normal) and median zero (Laplace) error distributions for the transformed egg counts (and with varying distributions on the random effects), that there is improvement based on LPML using the median zero models versus the mean zero models.

With complex data sets, the possibilities for additional modeling of the data are limited only by our imagination. A discrete longitudinal model such as a mixed Poisson or negative binomial regression could be used for the egg counts. Various other flexible regression structures such as wavelets or neural networks could be used for the temporal trend of the covariate process, and other survival distributions or nonparametric Bayesian priors could be used. Alternative analyses of these data have used simple random effects models for the longitudinal component.

JT remarked about the potential advantages of using an additional time dependent covariate, like the rate of change of the original one, egg counts in our case. Of course we agree, and we actually tried this in our original analysis, both with rate of change of egg production and current total egg production, but there was no improvement in our analysis as neither covariate brought about gains in the LPML compared to the single covariate model.

JT also suggested the possibility of marginally modeling the survival data and specifying the longitudinal data conditional on the failure times. Such an approach is taken by Zhang et al. (2010) to model survival of prostate cancer patients. This might actually be necessary with the fruit fly data if all of the data, including flies with shorter lives, were to be analyzed.

We intend to pursue the alternative models implied by our additional thoughts, stimulated by RC and JT, about scientific goals discussed in Sect. 1.

References

Crespi CM, Boscardin WJ (2009) Bayesian model checking for multivariate outcome data. Comput Stat Data Anal 53:3765–3772

Ding J, Wang J-L (2008) Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. Biometrics 64:546–556

Gelman A, Meng X-L, Stern H (1996) Posterior predictive assessment of model fitness via realized discrepancies. Stat Sin 66:733–807

Guo X., Carlin BP (2004) Separate and joint modeling of longitudinal and event time data using standard computer packages. Am Statist 58:16–24

Hanson T, Johnson WO, Laud P (2009) Semiparametric inference for survival models with step process covariates. Can J Stat 37:60–79

Henderson R, Diggle P, Dobson A (2000) Joint modelling of longitudinal measurements and event time data. Biostatistics 1:465–480

Lemos RT, Sansó B (2009) A spatio-temporal model for mean, anomaly, and trend fields of north Atlantic sea surface temperature. J Am Stat Assoc 104:5–18

Li Y, Lin X, Müller P (2010) Bayesian inference in semi-parametric mixed models for longitudinal data. Biometrics 66:70–78

Rizopoulos D, Verbeke G, Molenberghs G (2009) Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. Biometrics 66:20–29

Sinharay S, Stern HS (2003) Posterior predictive model checking in hierarchical models. J Stat Plan Inference 111:209–221

Zhang S, Müller P, Do K-A (2010) A Bayesian semiparametric survival model with longitudinal markers. Biometrics 66:435–443