Discussion on "Joint modeling of survival and longitudinal non-survival data" by Gould *et al.*

Alessio Farcomeni* Bhuvanesh Pareek† Pulak Ghosh‡

First, we would like to congratulate the authors on a very accurate and clearly written work. The paper provides a thorough literature review on various aspects of the joint model and certainly provides a strong motivation to the use of joint modeling techniques in medical applications.

Jointly modelling two or more processes together has been an active area of research for quite sometime now. This is particularly useful when the processes have different distributions, as in mixed data frameworks. A particularly interesting feature is the possibility to predict one event conditional on the others.

We believe three extremely interesting experimental designs arise which are amenable to joint modeling techniques. First of all, joint models must be used when the focus is on the longitudinal outcome(s), but there is informative drop-out. Secondly, joint models must be used when the focus is on the survival outcome, and there are time-varying covariates which might be measured with error [1]. Finally, in many applications both aspects of the joint model might be of interest, where one might discover that a treatment is beneficial for survival but detrimental for a longitudinal outcome measuring quality of life or occurrence of a side effect.

While much research happened on this topic, it is yet to become popular in the applied field despite the fact that, as surveyed by the authors, there are now many software options for easy implementation of joint models. Reasons in our opinion range from lack of user-friendly software to lack of understanding of the the advantages of the joint model framework.

*Sapienza - University of Rome
†Indian Institute of Management Indore
‡Indian Institute of Management Bangalore

Most importantly, editorial (e.g., by referees) pressure to use the joint model framework when appropriate in applied and clinical journals seems to be very light. A quick and non-systematic search on PubMed reveals that these models are mentioned mostly only in journals in the areas of statistics and probability. On the other hand, clinical papers based on longitudinal (observational or randomized) studies are prevalent. A quick skim through some reveals that the most common approach is to focus on estimation of survival probabilities based on *baseline* measurements, ignoring any following measurements. We believe there is a good rationale behind this practice, that is, one can assess survival probabilities based on the information available at baseline. On the other hand, ignoring additional measurements is a terrible waste of information. We must report that we also have found a few papers ignoring the problem of informative drop-out and mistreating data as missing at random. These papers report possibly biased results which might have mislead the clinical conclusions and recommendations.

We would like to complement the very nice review of Gould *et al.* by mentioning few additional recent advances.

In [2] and [3] it is described how to obtain dynamic predictions of survival probabilities. A particularly intriguing feature is that the estimates can easily be updated after each longitudinal measurement, providing up to date prognosis assessments.

Additionally, many works is this field are based on Gaussian longitudinal outcomes, but we would like to mention few exceptions that allow to model semi-continuous outcomes [4], and quantiles of a continuous outcome [5]. Generalized linear models for the outcomes are proposed in [6] from the Bayesian perspective and in [7] in a classical framework.

Further, in many cases a right-censored survival outcome will be available, but other available options include competing risks as in [8], multiple events per subject [9], also considered in the seminal paper of [10], and interval-censored survival outcomes [11, 12].

Observation times are commonly not informative (e.g., they might be scheduled in advance). In other situations anyway observation times might instead be informative, e.g., patients in worse conditions may be visited more often. Joint models with informative observation times have been derived for instance in [13] and [14].

A technical issue in traditional joint models is the use of Gaussian quadrature. The use of Gaussian quadrature techniques (e.g., to obtain the expected complete likelihood) in our opinion has often limited the dimensionality of

random effects involved. In many cases only a subject-specific intercept and a random slope are included, which is fine. In our opinion anyway quadrature may become unstable or too computationally expensive when more random effects are specified. A first attempt to tackle this issue can be found in [5], where a general Monte Carlo Expectation Maximization strategy is developed.

Another open issue is a full exploration of the flexibility of copulas for developing joint models.

Contacting author: Pulak Ghosh
Bannerghatta Road-560076, India
email: pulak.ghosh@iimb.ernet.in

References

[1] Keogh RH, White IR. A toolkit for measurement error correction, with a focus on nutritional epidemiology. *Statistics in Medicine* 2014; :available online.

[2] Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* 2011; **67**:819–829.

[3] Mauguen A, Rachet B, Mathoulin-Pélissier S, MacGrogan G, Laurent A, Rondeau V. Dynamic prediction of risk of death using history of cancer recurrences in joint frailty models. *Statistics in Medicine* 2013; **32**:5366–5380.

[4] Liu L. Joint modeling longitudinal semi-continuous data and survival, with application to longitudinal medical cost data. *Statistics in Medicine* 2009; **28**:972–986.

[5] Farcomeni A, Viviani S. Longitudinal quantile regression in presence of informative drop-out through longitudinal-survival joint modeling 2014; :arXiv:1404.1175.

[6] Rizopoulos D, Ghosh P. A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine* 2011; **30**:1366–1380.

[7] Viviani S, Alfó M, Rizopoulos D. Generalized linear mixed joint model for longitudinal and survival outcomes. *Statistics and Computing* 2014; **24**:417–427.

[8] Huang X, Li G, Elashoff RM, Pan J. A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. *Lifetime data analysis* 2011; **17**:80–100.

[9] Elashoff RM, Li G, Li N. A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics* 2008; **64**:762–771.

[10] Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000; **1**:465–480.

[11] Gueorguieva R, Rosenheck R, Lin H. Joint modelling of longitudinal outcome and interval-censored competing risk dropout in a schizophrenia clinical trial. *Journal of the Royal Statistical Society (Series A)* 2012; **175**:417–433.

[12] Bartolucci F, Farcomeni A. A discrete time event-history approach to informative drop-out in mixed latent Markov models with covariates. *Biometrics* 2014; :in press.

[13] Liu L, Huang X, O'Quigley J. Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics* 2008; **64**:950–958.

[14] Sun L, Song X, Zhou J, Liu L. Joint analysis of longitudinal data with informative observation times and a dependent terminal event. *Journal of the American Statistical Association* 2012; **107**:688–700.