Report on manuscript AOAS2101-009 entitled "Personalized Schedules for Shared Decision Making of Burdensome Surveillance Tests", by Tomer A., Nieboer D., Roobol M.J., Steyerberg E.W. and Rizopoulos D.

March 8, 2021

Please find below a summary of the paper followed by some details of my assessment.

My summary of the paper

This paper is about generating personalized schedules for invasive diagnostic tests used to detect disease progression in early-stage chronic non-communicable disease surveillance. Based on a (standard) joint model for longitudinal and time-to-event data that is fitted to some training sample, a personalized schedule of future surveillance tests for a new patient with unknown progression time is derived. To this end, a cumulative-risk function, that was already proposed and used elsewhere, is employed and the patient's clinical data and previous diagnostic test results are combined with the training data to estimate the patient-specific cumulative-risk of progression over her/his current and future follow-up visits. Future invasive tests are planned whenever this cumulative-risk function is predicted to be above a certain threshold, where this threshold is selected in an automatic fashion by optimizing a loss function that incorporates the expected number of tests and the time delay in detecting disease progression of each schedule. The proposed models are fitted under the Bayesian paradigm using Markov chain Monte Carlo (MCMC) algorithms (with the R package JMbayes) and are applied to both simulated and real data in prostate cancer surveillance. The manuscript does genuinely concern applied statistics and addresses a topic that is timely and relevant. The paper is well-written and concise. However, I do have some concerns about this paper, as described below.

Details of my assessment of the paper

In the comments, the line numbers refer to the counted lines of text on the corresponding page of the manuscript. I hope these comments can be helpful in guiding the authors

towards improvements.

Comments

(a) Section 1 or Section 2: The authors are considering clustered data, where the repeated measurement data are clustered within individuals and where the time to disease progression is the primary endpoint. Such data could be modelled by a random-effect model for time-to-event data, in which the longitudinal biomarker is included as an endogenous time-varying covariate and the random-effect captures the association within clusters or individuals as well as the (unexplained) heterogeneity among clusters. Please add a short paragraph which states the advantages and disadvantages of choosing such an approach compared to a joint model for time-to-progression and longitudinal outcomes.

(b) Section 2: The random effects $\mathbf{b}_i \in \mathbb{R}^{K\times 1}$ for $i = 1, \ldots, n$ account for both the association between the longitudinal and time-to-event outcomes and the correlation between the repeated measurements in the longitudinal process. Could this assumption be relaxed such that one set of latent variables account for the former association structure and another set (correlated with the first set but not identical to it) accounts for the latter?

(c) Page 5 lines 5-6: A multivariate normal distribution is imposed on the random effects $\mathbf{b}_i \in \mathbb{R}^{K\times 1}$ for $i = 1, \ldots, n$. This implies certain dependence structures between the longitudinal outcomes and between the longitudinal and the time-to-event outcomes. But model dependence is not discussed at all in the manuscript. Please comment on this.

(d) In the context of the previous comment above, have other distributions for the random effects such as the gamma distribution been used and checked as well? I have noticed in the Supplementary B.4 that for the simulation study in Section 5 you have used different distributions for the error terms for the PSA measurements, that is, t-distributions with three and four degrees of freedom as well as a normal distribution. Similar things could be done for the random effects as well.

(e) Except for the last short paragraph in Supplementary A.2, Supplementary A.1 and A.2 are merely a repetition of Section 2. Moreover, from my point of view, Supplementary B.2 in which the joint model is explicitly formulated for the motivating example, would be better suited as a model section than the current Section 2, which is the general formulation of the joint model framework.

(f) Page 5 line 22 (here and elsewhere): replace "relative-risk" by "relative hazard" as risk ratios and hazard ratios are not the same and the model for the survival sub-process in Subsection 2.2 is a relative hazard model!

(g) Page 5: in the cumulative-risk function (1) the random effects for the new patient of interest are integrated out, which implies a population focus. Do the estimated random effects for the jth patient, $\widehat{\mathbf{b}}_j$, play a role in deriving personalized visiting schedules for the jth patient? If not, why not?

(h) Subsections 3.1, 3.4 and Figure 2: Wouldn't it be better to plan tests on future visits when the predicted cumulative-risk of progression is above the upper end of the 95% credible interval of the risk profile $R_j(u|t, v)$ in equation (1) rather than above the risk threshold κ?

(i) Page 11 line 2: enforce a line break.

(j) Section 4 and Section 5: in what respects does the dataset of the PRIAS study that is used as the motivating application for this manuscript differ from the data that have been used in Tomer et al. (2019a) and Tomer et al. (2019b)?

(k) Page 6, legend of Figure 2: "Panels A, B and C:" and not "Panels A, and B:".

(l) Section 6: Please discuss identifiability issues related to the estimation of model parameters in the joint model.

(m) The models are fitted using the R package JMbayes. In many regards the design of package JMbayes is similar to the one of package JM for fitting joint models under maximum likelihood. In particular, JMbayes has a basic model-fitting function called jointModelBayes(), which accepts as main arguments a linear mixed effects

object fit as returned by function lme() of package nlme and a survival object fit as returned by function coxph() of package survival. In other words, before one actually fits the joint model, two separate models are fitted beforehand just to extract information that are available without fitting these two models, e.g. response vectors, design matrices, the event indicator, et cetera (see also Rizopoulos (2012), p. 58). Could you please comment on that?