PERSONALIZED SCHEDULES FOR SHARED DECISION MAKING OF BURDENSOME SURVEILLANCE TESTS

Joint models been around for some time and one aspect that has had attention is prediction. This is a contribution in that direction. The idea is to derive, from a joint model for event times and longitudinal biomarkers, the predictive distribution function for time to event, and then schedule future tests at points where the predictive probability reaches a threshold, with a reset when a test is negative. The paper is built around scheduling biopsies in prostate cancer surveillance, with PSA and other markers intermittently observed, and the event of interest being progression to a more severe stage. The main novelty is an argument that a schedule of biopsies specified in advance is more useful than simply the time of the next biopsy. The success of a strategy is defined in terms of reduced number of biopsies and reduced delay between progression and detection. These quantities are investigated.

The paper is well motivated, nicely written and grounded in a real application. There is the right amount of detail on the underlying models. There is an example using a rich set of data on 7813 patients, with some case studies, and a large simulation study sensibly based on the real data. There is comparison between fixed and data-dependent threshold choices and with standard strategies, with proper attention to implementation in practice.

1. I have no concerns at all with the paper as it is. Everything makes good sense. The issue is that the paper is the latest in a succession of papers from the same group and although there is novelty in the focus on schedules, the impression is that it is an afterthought or appendix to the earlier work. The same basic ingredients are present in Tomer at al's 2019a Biometrics paper, and in two more applied papers from the same group: the cited 2019b Medical Decision Making paper and a more recent uncited Tomer et al 2021 article "Personalised biopsy schedules based on risk of Gleason upgrading for patients with low-risk prostate cancer on active surveillance" (BJU International, doi:10.1111/bju.15136).

   The papers are different in detail and there is certainly novelty in the current submission. But it does seem to fall into a gap between original methodology and application. The methodological contribution is limited and the application is covered, and for a more suitable audience, in the 2019b and 2021 papers. The latter states that "We successfully developed and validated a model for predicting upgrading risk, and providing risk-based personalised biopsy decisions in AS of prostate cancer". This raises the question of what extra the current paper achieves that would be of interest to Annals of Statistics readers?

2. The argument for the approach in this paper is that it allows a full testing schedule to be specified ahead of time, rather than just the next one. Is this really important in practice? As further longitudinal measurements are obtained then the predictive risk will change and so too presumably the proposed schedule. Won't this outweigh any advantage of telling a patient in advance what the full schedule is expected to be?

3. Something that might be interesting is to explore the robustness of the outcomes to misspecification of the joint model. Have the authors looked at that?

4. A minor aside: PRIAS looks competitive to me from the results in Figure 6.