Review of SIM 21-0304

This manuscript considers a method for deriving a personalized schedule for an invasive surveillance test based on dynamically predicted risks and an approach for determining a decision threshold that accounts for the benefit and burden of the decision. The questions addressed in the manuscript are interesting and of potential relative importance. The application to the prostate cancer study is thoughtful.

Main comments:

1. A key point of the paper is to determine subsequent schedules of multiple tests over the entire follow-up period at a specific clinical visit time *v*, whereas the existing literature mostly focused on the timing of the subsequent test following the time of visit. It is not clear why it is useful in this clinical setting to figure out all subsequent test schedules. The clinical decision at the visit is if a test should be scheduled earlier, on time, or late compared with the standard fixed schedule, please provide some justification why knowing a hypothetical schedule (also see #2) would help with that clinical decision.

2. Related to 1, the long-term schedule developed based on the information up to the visit day *v*, Y(*v*) without updating based on newer information is limited. As the decision on later tests after the subsequent one would be dependent on the additional finding from that test to be scheduled and newer clinical information collected at the next visit prior to the future scheduled visit. That is, predicting later tests based on information collected at the current clinical visit time may not be meaningful, as it is not updated by new information Y cumulated up to a more recent visit Y(v+s) for s>0).

3. Please provide details of how the cumulate risk $R(T>u|T>t,Y(v))$ can be calculated from a time-varying covariate model of the form $P(T>t|Y(t))$ from a joint modeling framework, given Y(v) with v>t.

4. The schedule of planned future tests S_j^k is determined by $R(u|t,v)$ with u and t vary but *v* is constant in determining the set of S_j^k, therefore S_j^k should be dependent on v. Should the calculation of $E\{N_j(S_j^k)\}$ and $D_j(S_j^k)$ be additionally conditioning on $T_j^*>v$?

5. In PRIAS study, is the dataset split into a training set for joint modeling, and a testing set for determining the personalized schedules and calculating the validation summaries? Also if k= 10% was decided as in Figure 4, should one consider cross-validation?

Page 14, last line, t_5 = 4.5 year, should it be t_5 = 5.5 year?