Department of Biostatistics
Erasmus University Medical Center
PO Box 2040, 3000 CA Rotterdam
the Netherlands

February 19, 2018

Professor Michael J. Daniels
Department of Statistics
University of Florida
Gainesville, FL, 32611-8545
USA

Dear Professor Daniels,

We are writing to you with respect to the manuscript #BIOM2017609M, titled "Personalized schedules for surveillance of low risk prostate cancer patients" submitted to *Biometrics* and the reports we received after its review. We would like to thank you for giving us the opportunity to submit a revised version of our paper that tackles the weaknesses of the previous version.

Following the recommendations from the Reviewers, we have made several changes in the revised version of the manuscript. In particular, we updated the joint model fitted to the PRIAS dataset to account for non normal errors. To this end we assumed t-distribution for errors. We also added more diagnostic graphs to show the fit of the model to the dataset. We added a section in the supplementary material to show that joint model parameter estimates are not affected by the schedule of biopsies and PSA measurements, given that the latter two depend only upon the observed PSA values. In order to aid in medical decision making, we added discussion on suitable values of number of biopsies and offset. Lastly, we updated our graphs and their captions as per the suggestions of Reviewers. The previous version was 24 pages long and you have asked us to reduce that to 20 pages (including body of manuscript, acknowledgments and references). While according to the suggestions of the reviewers we have included new pieces of information, we have managed to reduce the length of the paper to 20.25 pages. We hope that this is acceptable.

Please find enclosed a detailed point-by-point response to the Reviewers' comments.

Yours sincerely,

the Authors

Response to Associate Editor's Comments

We would like to thank the Associate Editor (AE) for his/her constructive comments, which have allowed us to considerably improve our paper. The main differences of the new version of the manuscript compared to the previous one can be found in Sections 5 and 6, Web Appendix A.2, C and D. In addition, changes regarding the specific comments have been made throughout the text.

You may find below our ~~response~~ responses to the specific issues raised.

1. How would delaying in activating treatments translate into survival outcomes?

The AE raises a very important point.

~~The biopsies are already conducted for the patients according to the PRIAS schedule. Thus, we cannot test the effect of delay in detection of Gleason reclassification (GR) due to personalized schedules on the prostate cancer (PCa) specific mortality, unless they are applied on the patients. It may however be possible to compare the effect of delay in detection of GR on PCa mortality, due to PRIAS (less frequent biopsies) versus annual schedules (more frequent biopsies). However, given that we do not have PCa mortality results from cohorts which use annual schedules, this comparison cannot be done at the moment.~~

As far as the effect of active surveillance (AS) on PCa mortality is concerned, multiple studies ~~have reported small PCa specific mortality in low risk patients enrolled in AS (Klotz et al., 2009; Loeb et al., 2016; Tosoian et al., 2011). That is, less frequent schedules may be useful in such scenarios. For example, for slowly progressing patients (subgroup G_3) in our simulation study, we observed that even a personalized schedule which conducts on average two biopsies leads to an average delay of 10 months in detecting GR. This is only four months more delay than that of the annual schedule. Given, the low PCa mortality a relative difference of four months may not be that bad an alternative.~~

Response to 1st Referee's Comments

We would like to thank the Referee for his/her constructive comments, which have allowed us to considerably improve our paper. The main differences of the new version of the manuscript compared to the previous one can be found in Sections 5 and 6, Web Appendix A.2, C and D. In addition, changes regarding the specific comments have been made throughout the text.

You may find below our ~~response~~ respones to the specific issues raised.

1. Assumption of normality of random effects and error term.

   We thank the Referee for motivating us to check these assumptions in detail. We found that our model did not satisfy the assumptions of normality of error terms. To this end, we discuss our solution for this issue in the following paragraph. The issue of assumption of normality of random effects is discussed in the last paragraph.

   With regards to the assumption of normality of error term, we ~~conducted~~ used residual diagnostics to check this assumption. ~~The~~ The left panel of Figure 1 shows the quantile-quantile (q-q) plot of subject specific residuals ~~in Figure 1~~ under our original joint model. This figure ~~suggests~~ suggests that a symmetric long tailed distribution for errors is more plausible than the normal distribution. Based on this result, we further fitted two joint models (JMs) with t-distributed errors, with 4 and 3 degrees of freedom (df), respectively. We found that the model with t-distributed (df=3) errors satisfied the distributional assumptions the best (see Figure 1).

   We then compared the model with the assumption that errors are normally distributed and the model with assumption that errors are t-distributed. To this end, the fitted marginal $\log_2$ PSA profile for a hypothetical patient with age 70 years using the two models is shown in Figure 2. We also compared the subject specific fitted $\log_2$ PSA profiles for 9 randomly selected patients (each with more than 3 observations). ~~A seed of 2017 (year of submission of the article) was used to sample these patients from the PRIAS dataset sorted by patient ID.~~ Lastly, for the two models, Table 1 shows the association parameters. We can see that the association between the hazard of GR and slope of $\log_2$ PSA is stronger in the model with t-distributed (df=3) errors. We have updated the parameters estimates for the new JM in **??** of the updated supplementary material.

   Since the slope association between $\log_2$ PSA levels and hazard of Gleason reclassification (GR) in the model with t-distributed (df=3) has become stronger, we expect our schedules

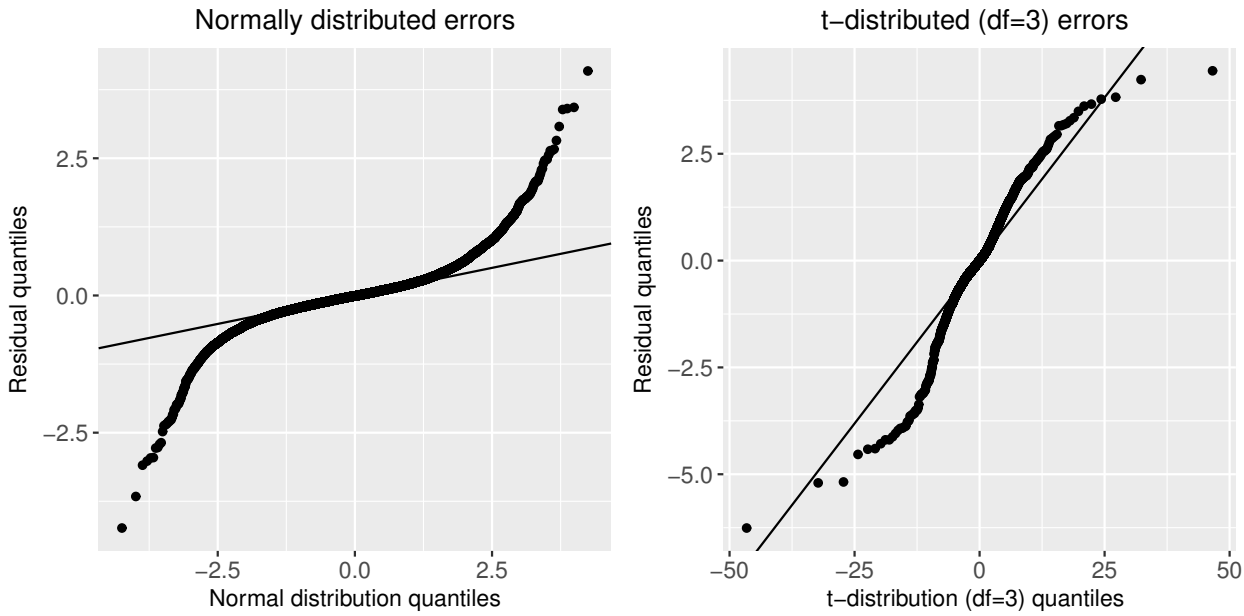

Figure 1: Quantile-quantile plots of subject specific residuals obtained from joint models with assumption of normally distributed errors, and t-distributed (df=3) errors, fitted to the PRIAS data set.

Table 1: Relative risk sub-model estimates for association parameters between hazard of GR and slope of $\log_2$ PSA levels. Mean and 95% credible interval (CI) are presented for fits obtained from joint model with assumption of normal distributed errors, and t-distributed (df=3) errors.

Error distribution	$\log_2$ PSA association [95% CI]	Slope($\log_2$ PSA) association [95% CI]
t-distribution (df=3)	-0.004 [-0.119, 0.117]	2.888 [2.318, 3.452]
Normal distribution	-0.049 [-0.172, 0.078]	2.407 [1.791, 3.069]

to become slightly more sensitive towards increase in $\log_2$ PSA velocity. This however, also depends on the type of personalized schedule. For example, we compared the personalized schedule based on dynamic risk of GR using the two different models for the three demonstration patients, and observed trivial differences. This is due to the fact that average risk (averaged over all time points) taken by dynamic risk of GR is not very high (5.3%). However quantiles corresponding to 50% risk (median time of GR) may differ by a bigger margin depending upon the profile of the patient (same for expected failure time). We next discuss the impact of the new association parameter on the schedules for the three patients.

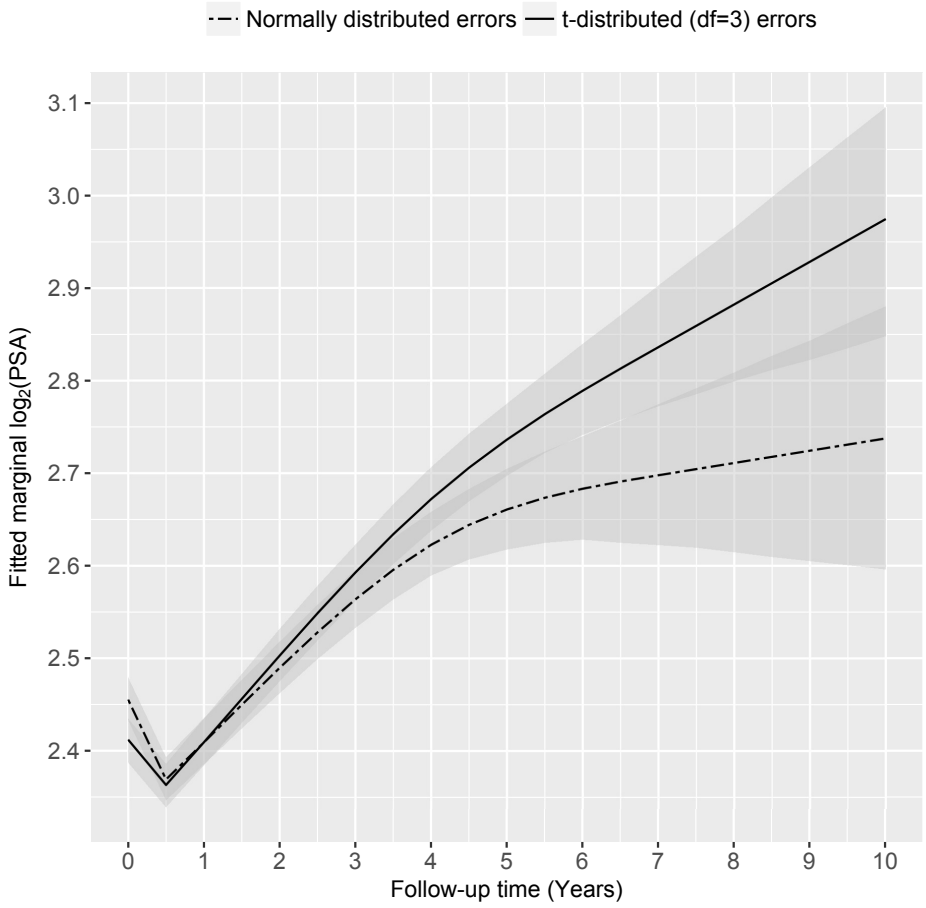

Figure 2: Fitted marginal 10 year $\log_2$ PSA profile with 95% credible interval (CI), for a hypothetical patient who was included in AS at the age of 70 years. Fits were obtained from joint models with assumption of normal distributed errors, and t-distributed (df=3) errors. The darker shaded region indicates the overlap in the two CI intervals, as well as demarcates the two sets of CIs.

We can see in Figure 4 (bottom row) and Figure 5 that the third demonstration patient has a consistent profile, with quite slow rise in PSA. Consequently the effect of the increased $\log_2$ PSA slope association parameter does not affect the schedule much for this patient. Similar results are observed for the first demonstration patient when the PSA consistently remains low over nearly three years, starting at year two (Figure 4, top row, rightmost panel). Lastly, this can also be seen for the second demonstration patient wherein the schedules differ by a large margin initially when PSA rises very quickly. However the gap becomes slightly smaller after a negative biopsy indicating that GR is unlikely in near future. Thus we

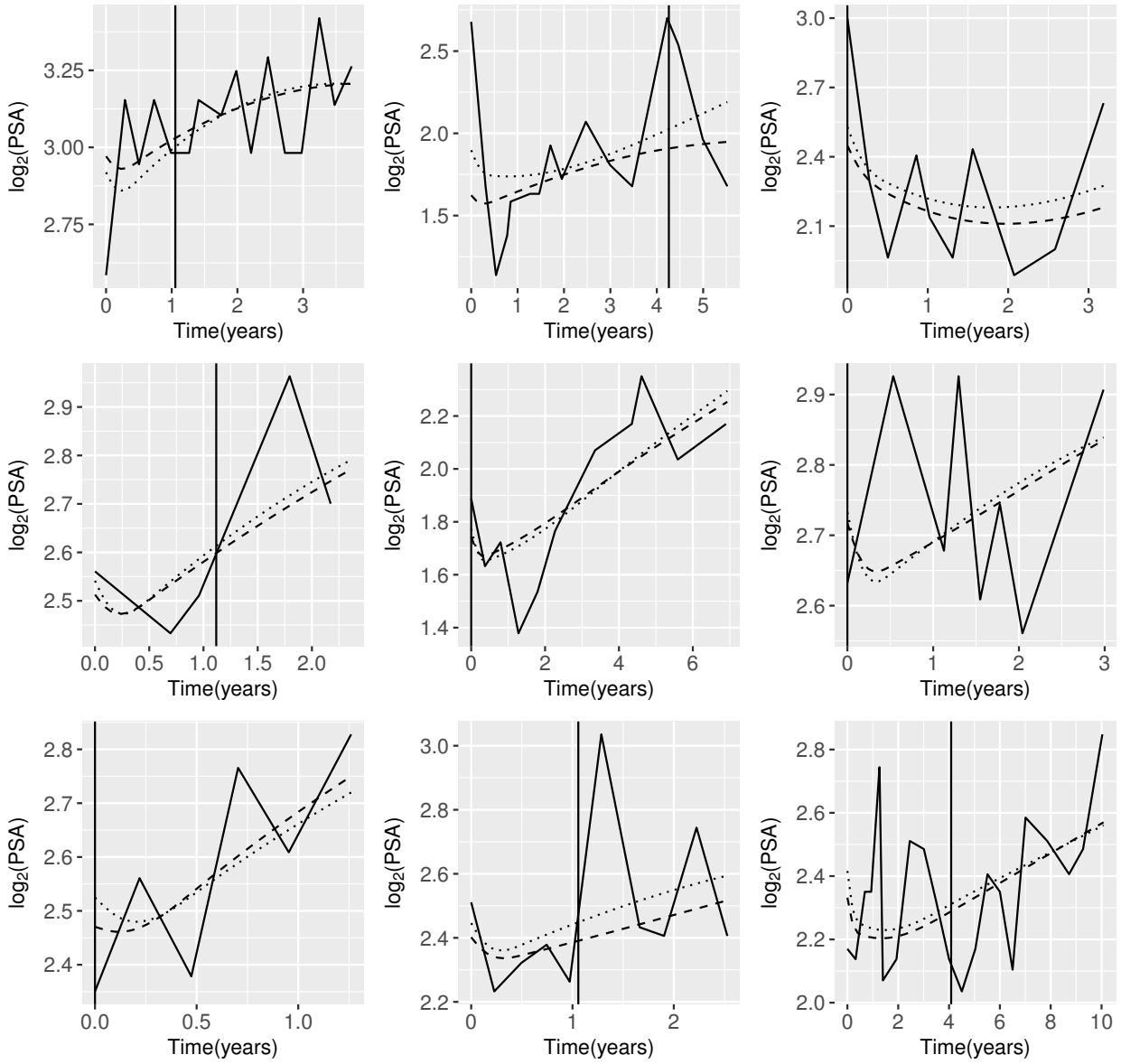

Figure 3: Fitted versus observed $\log_2$ PSA profiles for 9 randomly selected patients. Fits were obtained from joint models with assumption of normal distributed errors, and t-distributed (df=3) errors. The vertical line with a dot-dash pattern shows the time of the latest biopsy. The fitted profiles utilize information from both the observed PSA levels and time of latest biopsy.

expect that the sensitivity of the schedule based on expected failure time is within acceptable boundaries for slowly progressing (low risk) patients with consistent profiles.

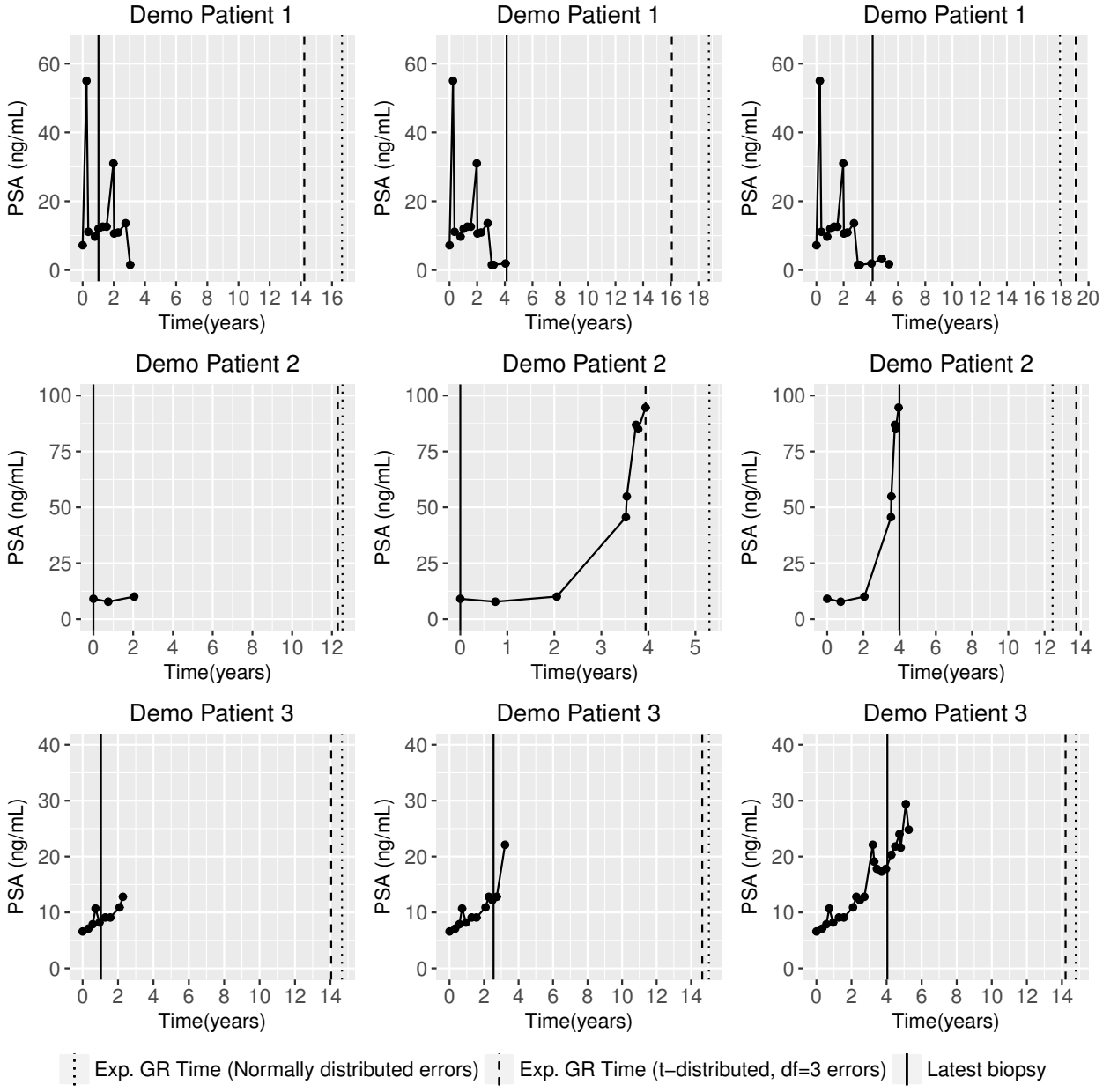

Figure 4: Dynamic expected failure time for the three demonstration patients at three different follow-up times, using joint models with assumption of normal distributed errors, and t-distributed (df=3) errors.

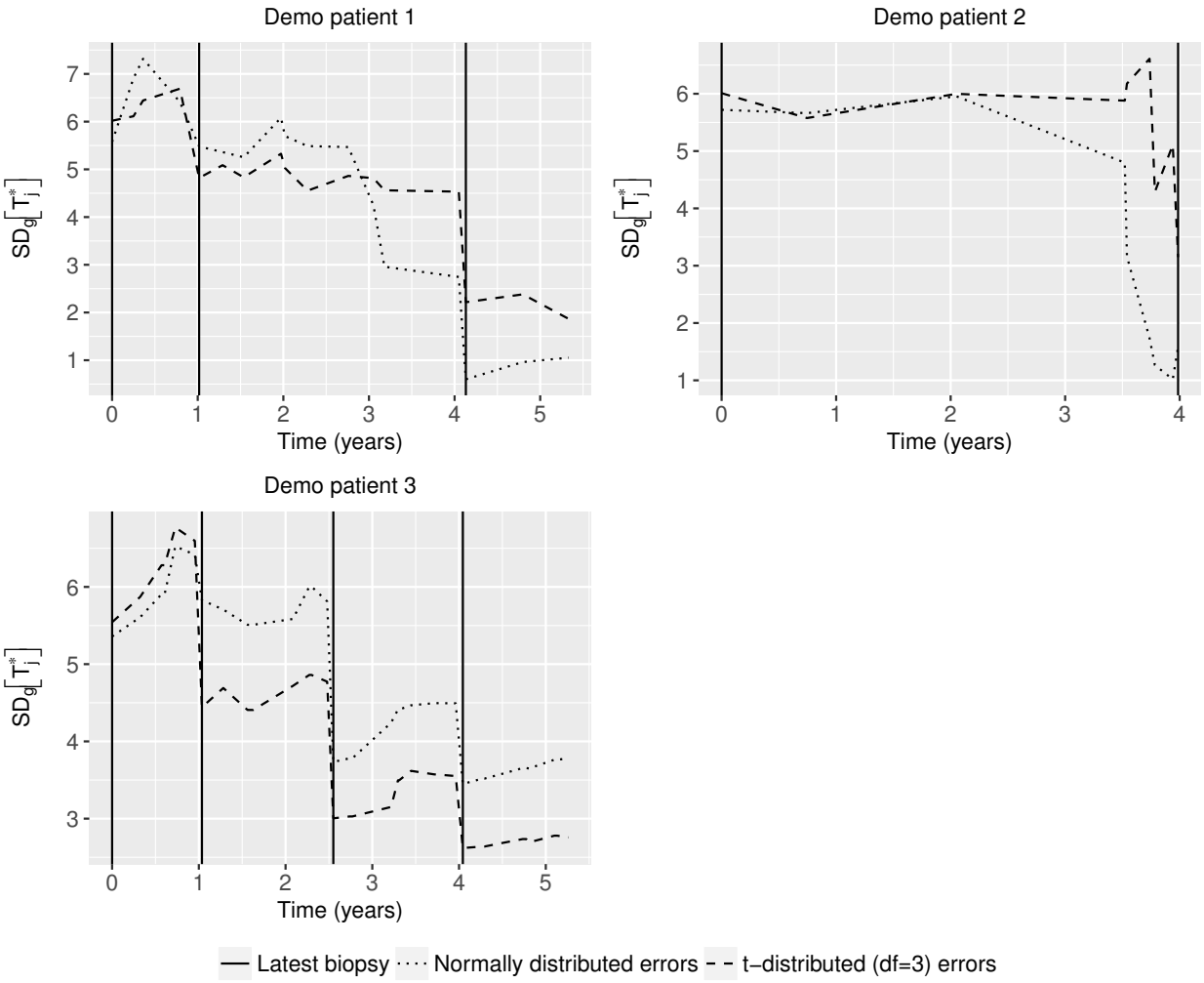

Figure 5: Dynamic variance of the posterior predictive distribution of event time for the three demonstration patients at three different follow-up times, using joint models with assumption of normal distributed errors, and t-distributed (df=3) errors.

With regards assumption of normality of random effects, JMs have been shown to be quite robust to random effects misspecification. More specifically, Huang, Stefanski, and Davidian (2009) and Rizopoulos, Verbeke, and Molenberghs (2008) have shown that unless the number of repeated measurements per person are extremely small, such misspecification only and trivially affects th andard errors. On the other hand in our dataset we have a mean of 8.7 measurements per person. which makes us feel confident with regard to this assumption.

2. Choice of demonstration patients, and cross validation using PRIAS dataset.

The three demonstration patients were chosen on the basis of specific characteristics of their data. This is because we wanted to demonstrate three main features of the personalized schedules. In particular, ~~For example,~~ the first demonstration patient had many repeat biopsies, and thus via his profile we show how the variance of posterior predictive distribution of GR time decreases with each biopsy. Via the second demonstration patient we show how the schedules change with changes in PSA alone (no repeat biopsies). Whereas, via the third demonstration patient we show how the schedules work when information from PSA and repeat biopsies is not in concordance with each other. ~~We would like to also mention that it is not the case that we conducted an exhaustive search to purposefully select only these three patients.~~

With regards to conducting cross-validation on real data, and to compare the true GR time of PRIAS patients who obtained GR, with the time proposed by personalized schedules, this is not possible for the following reason. For patients in PRIAS we only know the interval $l_i < T_i^* \leq r_i$ in which GR occurred and not the true GR time T_i^*. On top of that this is known only for 707 out of 5267 patients, and the rest are right censored. That is, in either case we cannot calculate the offset $T_i^S - T_i^*$ of our schedule, where $T_i^S > T_i^*$ ~~is~~ the time of the last biopsy at which GR is detected. The simulation study is our attempt to ~~circumvent~~ ~~this problem, because there we know the true event time T_i^* for each of the patients.~~

~~S~~tratified relative risk model for modeling baseline hazards in simulation study.

~~We assume~~ that there are three equal sized subgroups G_1, G_2 and G_3 of patients in the population, differing in the baseline hazard of GR. This was done because we wanted to test the performance of different schedules for a population with a mixture of patients, namely those with faster progressing PCa, as well as those with slowly progressing PCa. As advised by the Referee these can be modeled correctly using a stratified modeling approach. In our current case, this corresponds to the use of latent class JMs (Proust-Lima et al., 2014). ~~However this approach requires either knowing the number of subgroups in advance, which is unrealistic, or fitting multiple models to detect the correct number of subgroups. The latter would have been out of the scope of our paper. The alternative that we used was based on the fact that the baseline hazard in the simulated population corresponds to a mixture Weibull density~~ (Razali and Al-Wakeel, 2013). We model the log baseline hazard of the mixture Weibull density, flexibly using P-splines (see **??**). The mean of the fitted log baseline hazard between 0.1 years and 8 years (mean third quartile in the simulated progression times is 6.15 years), and 95% confidence interval obtained from the 500 simulations is shown in Figure 6 below. It can be seen that the fit is quite close to the theoretical baseline hazard, and hence we do

not expect our fitted model to give invalid results.

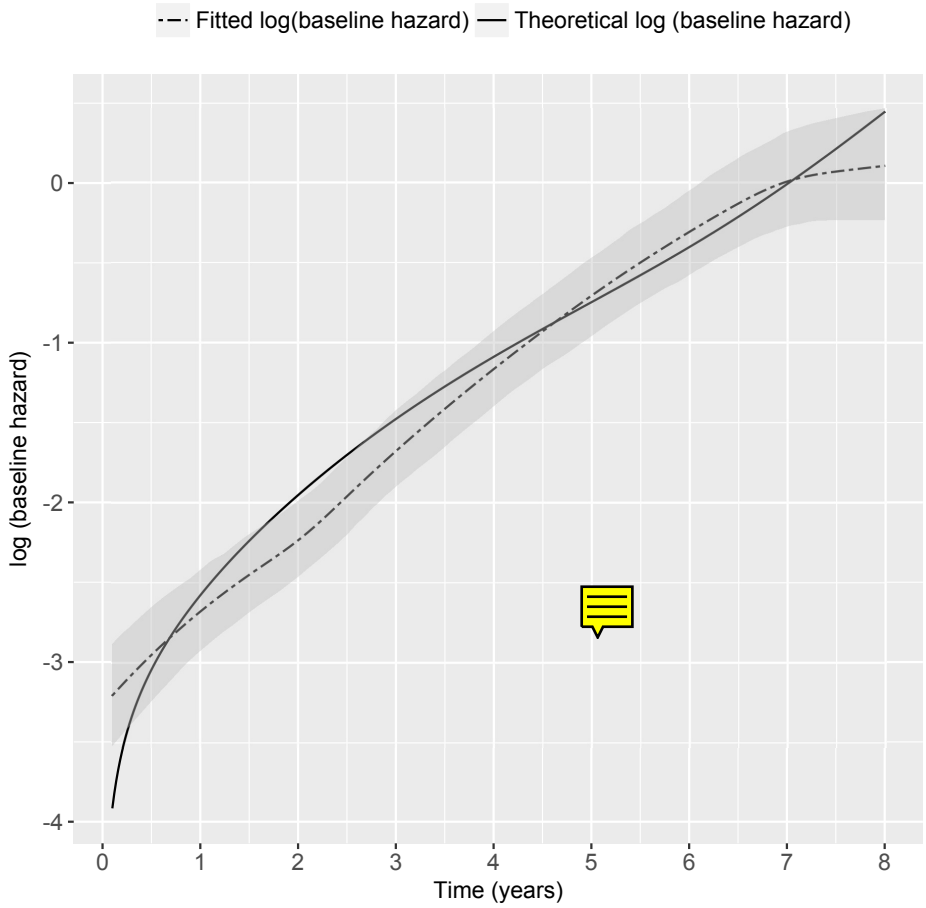

Figure 6: Theoretical log baseline hazard of the simulated population versus mean of the fitted log baseline hazard. The 95% confidence interval for the fitted log baseline hazard is obtained from the 500 simulations.

4 ias due to biopsy schedule depending upon PSA-DT.

~~As noted by the referee,~~ indeed PRIAS switches to the more frequent annual schedule if a patient's PSA doubling time (PSA-DT), measured as the inverse of the sl of the regression line through the base two logarithm of PSA values, is less than 10 years. ~~In this regard, the JM allows the schedule to depend upon the observed PSA values (e.g., via PSA-DT). This is because the parameters are estimated using~~ a full likelihood approach (Tsiatis and Davidian, 2004). To show this, consider the following full general specification of the JM that we use. Let $\boldsymbol{y}_i$ denote the $n_i \times 1$ vector of PSA measurements for the i-th patient, and l_i, r_i te

the two time points of the interval in which GR occurs for the i-th patient. In addition let T_i^S and $\mathcal{V}_i$ denote the schedule of biopsies and schedule of PSA measurements, respectively. Under the assumption that both of these schedules may depend upon only the observed $\boldsymbol{y}_i$, the joint likelihood of all four processes is given by:

$$p(\boldsymbol{y}_i, l_i, r_i, T_i^S, \mathcal{V}_i \mid \boldsymbol{\theta}, \boldsymbol{\psi}) = p(\boldsymbol{y}_i, l_i, r_i \mid \theta) \times p(T_i^S, \mathcal{V}_i \mid \boldsymbol{y}_i, \boldsymbol{\psi}) \quad (1)$$

From this decomposition we can see that even if the processes T_i^S and $\mathcal{V}_i$ may be determined from $\boldsymbol{y}_i$, if we are interested in the parameters $\boldsymbol{\theta}$ of the joint distribution of longitudinal and event outcome, we can maximize the likelihood based on the first term and ignore the second term. In other words, the second term will not carry information for $\boldsymbol{\theta}$.

~~In order to demonstrate this we simulated a dataset with 750 patients. The true event times T_i^* for these patients were generated using parameters from a JM fitted to the PRIAS dataset. However this JM did not include association between velocity of log PSA values and hazard of GR. That is, the hazard of GR $h_i(t)$ at any time t depends only on the underlying log PSA value $m_i(t)$ at that time. Furthermore, for the simulated patients, we used the schedule of PRIAS to generate the interval $l_i \leq T_i^* \leq r_i$ in which GR is detected. Thus the observed data for i th patient is $\{\boldsymbol{y}_i, l_i, r_i\}$. Our aim is to show that if there is no association between $h_i(t)$ and velocity of log PSA value $m_i'(t)$, then even if the biopsy schedule depends on PSA DT (which is a crude measure of PSA velocity), a JM fitted with both value and velocity associations will have an insignificant velocity association. In the fitted JM we found the value association (95% credible interval in brackets) to be 0.182 [0.090, 0.274], and the velocity association to be -0.001 [0.295, 0.254]. That is even though the schedule of biopsies depended upon observed PSA values it did not lead to a spurious association.~~

With regards to the second question about the use of right censoring in the simulation study instead of interval censored data, this should not lead to any nontrivial differences in results. The reason is that we use a full likelihood approach as described in Section 2 of the original manuscript. Parameter estimation using full likelihood approaches always gives consistent and asymptotically unbiased results (Gentleman and Geyer, 1994).

5. AUC comparison for JM with value and velocity association, and only value association. We thank the Referee for noticing the erroneous result that was reported ~~earlier.~~ The two sets of AUC's were mistakenly swapped while creating **??** in the original supplementary material, and hence the counterintuitive results. We have corrected this mistake, and in addition as

advised by the referee, we have also reported the confidence interval. ~~To obtain the 95% confidence interval, we used 15 bootstrapped datasets of patients from the training dataset utilized in this paper. The choice of the number of datasets was based on the fact that for our large dataset estimating each of these models takes time. We calculate the AUC at year one, year two and year three of follow up in AS. The time window for which the AUC is calculated is one year.~~ The resulting estimates are presented in Table 2 below, as well as added in Supplementary material (??).

Table 2: Area under the receiver operating characteristic curves, and 95% confidence interval in brackets, obtained using 15 bootstrapped datasets. Bootstrap datasets were obtained from the PRIAS dataset.

Year	$\log_2$ PSA value and velocity association	$\log_2$ PSA value association
1	0.613 [0.582, 0.632]	0.595 [0.565, 0.618]
2	0.648 [0.608, 0.685]	0.609 [0.568, 0.654]
3	0.593 [0.560, 0.638]	0.590 [0.536, 0.628]

6. Typographic errors in equations.

   We thank the referee for pointing out these errors. We have fixed these in the ~~updated~~ manuscript.

Response to 2nd Referee's Comments

We would like to thank the Referee for his/her constructive comments, which have allowed us to considerably improve our paper. The main differences of the new version of the manuscript compared to the previous one can be found in Sections 5 and 6, Web Appendix A.2, C and D. In addition, changes regarding the specific comments have been made throughout the text.

You may find below our response to the specific issues raised.

1,4. Validity of the model for PSA.

We would like to thank the Referee for motivating us to check the model assumption and fit. As the Referee noted, the equation for the longitudinal sub-model on page 4 of the original manuscript does not indicate that we used a log transform for PSA levels. This is however the general form of the equation for the longitudinal sub-model, and is only used to introduce the joint model (JM) notation. The actual equation, showing the log transformed PSA levels, baseline covariates and B-spline for the effect of time is Equation (7) in the original manuscript. That is, it is not the case that the log transformation is used only in simulation study as noted by the referee, but also used for fitting the PRIAS data.

Since concerns regarding assumption of normality on errors were also raised by the first Referee, we refitted our model with an assumption that the errors are t-distributed (df=3). The residual quantile-quantile plots for the model with normally distributed errors as well as the T-distributed (df=3) errors are shown in Figure 1. In addition, the fitted marginal $\log_2$ PSA profiles, and subject specific fitted versus observed $\log_2$ PSA profiles of 9 randomly selected patients, using the two different models are presented in Figure 2 and Figure 3, respectively.

With regards to the fitted profiles for the three demonstration patients, we show their fitted profiles in Figure 7. The fitted profiles are dynamic in nature, and utilize information from both the observed PSA levels and time of latest biopsy. The first two panels for each of the patients are corresponding to the time points at which we made personalized schedules for these patients in the original manuscript. The third panel for each patient shows the fitted profile for the entire follow up period.

We have added the aforementioned figures in Web Appendix C of the updates supplementary material as well.

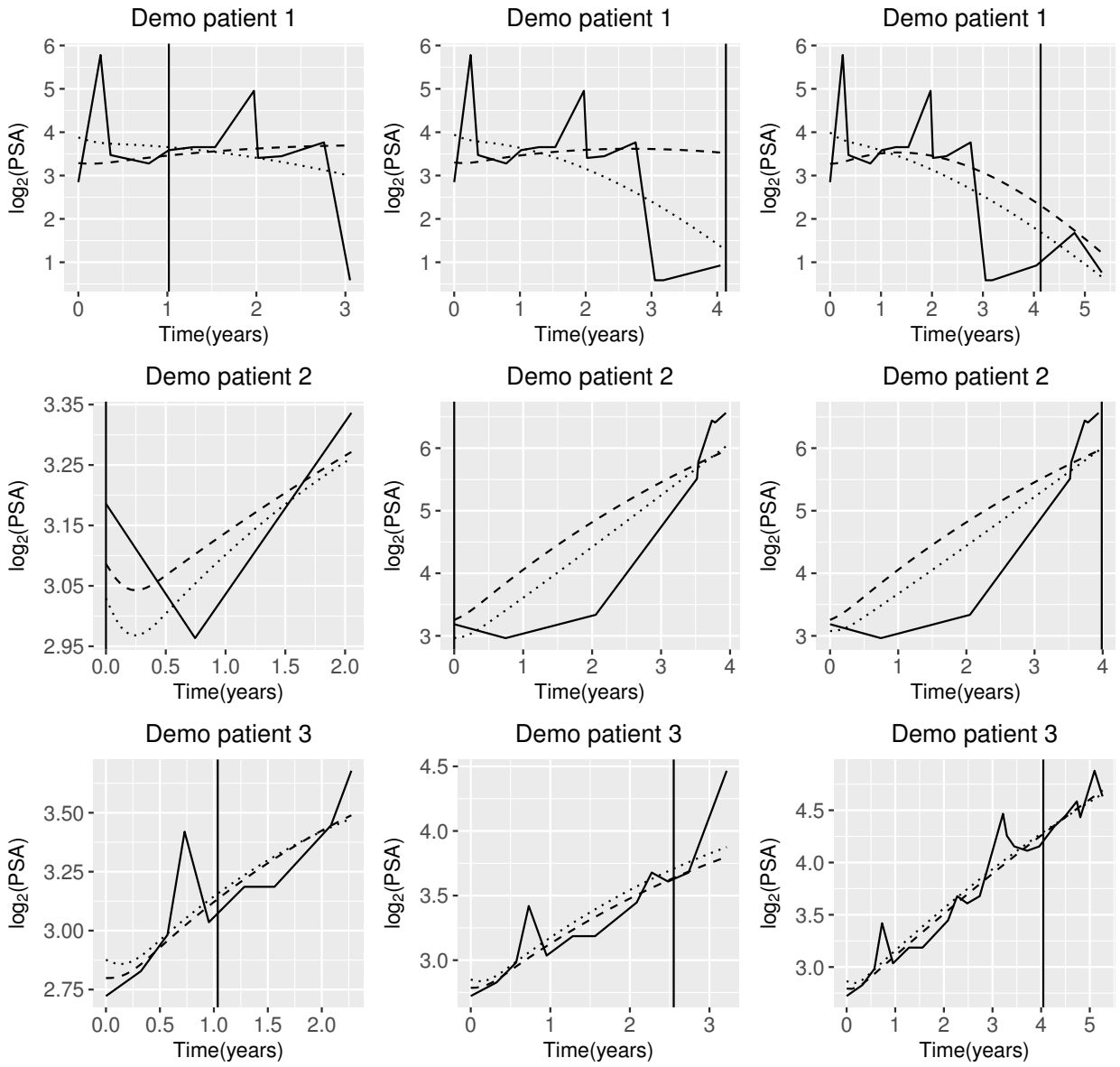

Figure 7: Fitted versus observed $\log_2$ PSA profiles for the three demonstration patients, at two different time points. The fitted profiles are dynamic in nature, and utilize information from both the observed PSA levels and time of latest biopsy.

2. If expected T_j^* is in past, we should be able to suggest to take biopsy right now.

   We agree to the Reviewer that our approach should be dynamic, that is, it should be able

to take into account entire PSA history and repeat biopsies, and also give a decision on immediate/delayed biopsy. We indeed provide a method to "evaluate biopsy time from current time, particularly when there is new information, such as new PSA measure after last biopsy". To illustrate this, suppose for the j-th patient, the last biopsy was conducted at time t, and the current visit time at which PSA is measured is $s > t$, then we are interested in finding the time $u > s$ of the next biopsy which utilizes all the available information up to s. To this end, all of our approaches are based on the posterior predictive distribution of GR time, given by $p\{T_j^* \mid T_j^* > t, \mathcal{Y}_j(s), \mathcal{D}_n\}$. Here $\mathcal{Y}_j(s)$ is the history of PSA up to s and the information that no GR was found at last biopsy is included via the condition $T_j^* > t$. Indeed as the referee noted it is possible that $t < T_j^* \leq s$, and then biopsy should be conducted immediately. However, it is often the case that difference between consecutive biopsies is required to be at least an year. Thus even if the schedule also suggests a time $t < u \leq s$, the biopsy should not be conducted immediately, but rather with a delay of $1-t$. ~~We have discussed~~ We now explain this scenario in Section 3.4 of the original manuscript. In addition, we have shown the entire decision making process related to conducting a biopsy in the flowchart in Figure 1 of the original manuscript.

Take the opportunity and provide extra clarification

We would also like to ~~provide clarification~~ for the definition of $\mathcal{M}_i(t)$ on page 5 of the original manuscript. We define $\mathcal{M}_i(t) = \{m_i(v), 0 \leq v \leq t\}$ as the history of the underlying PSA levels up to time t, or as noted by the Referee, PSA level up to last biopsy (Rizopoulos, 2012; Tsiatis and Davidian, 2004). The reason for such a definition is that the association between hazard of GR and PSA may depend on the entire history of PSA levels. For example, if hazard of GR at time t depends on the cumulative PSA levels up to t, then it is manifested by the following functional form:

$$f\{\mathcal{M}_i(t), \boldsymbol{b}_i, \boldsymbol{\alpha}\} = \alpha \int_0^t m_i(t)\mathrm{d}t \tag{2}$$

3. Robustness of the schedules based on dynamic risk of GR

We agree with the Referee that the term robust was used inappropriately. The meaning we wanted to imply was that schedules based on dynamic risk of GR are robust to large overshooting margins (offset). We have been more specific in this matter in the updated manuscript. Furthermore, we are providing a detailed explanation in the following paragraph.

We observed in Figure 2 of the original manuscript that the variance of posterior predictive distribution of event time decreases as more information is gathered over time. That is, a schedule based on expected/median time of Gleason reclassification (GR) is less accurate

(the consistency property) in predicting true event time when less information is available. In comparison, the schedule based on dynamic risk of GR is robust in the sense that it is more risk averse than the schedule based on median time of GR (50% risk), at all time points. For example, in PRIAS, on average it schedules biopsies whenever the risk increases more than 5.3%. Thus, it is less likely to overshoot the true GR time by a big margin even if less information is available for the patients. This is also demonstrated via the simulation study, wherein the schedule based on dynamic risk of GR leads to almost the same mean offset and variance of offset across the three subgroups of patients.

Minor Concerns Shared by the 2nd Referee

1. More informative captions for tables and graphs.

   We have now updated the captions of tables and graphs in the updated version of the manuscript. We specifically mention that these graphs pertain to the results from the simulation study.

2,4. Recommended values for parameters κ and η.

   For the two parameters κ and η, we do not use fixed set of values. We compute the parameter κ (dynamic risk of GR) from the data, as shown in Section 3.3 of the original manuscript. That is, we obviate choosing this value manually. We also provide an example for a comm recommended risk threshold of 5%. It is quite a risk averse threshold, however as shown in **??** in supplementary material the performance of this schedule is exactly same as that of annual schedule. That is, it gives a very small offset at the cost of too many biopsies.

   With regards to the choice of weights η_1, η_2, as discussed in Section 4.1 of the original manuscript, this choice can be obviated by reformulating the optimization of original weighted sum as a constrained optimization problem. For example, if η_1 is the weight corresponding to average number of biopsies $E(N^S)$ and η_2 is the weight corresponding to average offset $E(O^S)$, then we can instead put a constraint C on average offset, and then optimize for only the number of biopsies. Since multiple studies have reported small prostate cancer specific mortality in low risk patients enrolled in active surveillance programs (Klotz et al., 2009; Loeb et al., 2016; Tosoian et al., 2011),a recommended cutoff C on rage offset is 1 year. We have also added this information in the discussion section of the updated manuscript.

3. Age effect is not interpretable because of quadratic form of age.

We thank the Referee for noticing that due to the quadratic form of age in our model, the interpretation of relative difference of age that we did is incorrect. We have ~~fixed~~ this issue in our updated manuscript. We now illustrate the effect of age with an example, namely an increase in age at the time of inclusion in AS from 65 years to 75 years (first and third quartiles of age in PRIAS dataset) corresponds to a 1.419 fold increase in the hazard of GR.

Addressed

Good and bad O_j^S and N_j^S

The discussion of good and bad O_j^S and N_j^S entails discussion of patients tolerance for burden (N_j^S), and the amount of risk(O_j^S) that doctors consider manageable. Hence there are no fixed cutoffs for good and bad O_j^S and N_j^S. However, because PRIAS and annual schedules are already in practice, it can be argued that the maximum possible offsets due to these schedules (one and three years, respectively) are acceptable to doctors. In addition, multiple studies have reported small PCa specific mortality in low risk AS patients (Klotz et al., 2009; Loeb et al., 2016; Tosoian et al., 2011). Thus, less frequent schedules are an interesting alternative. For example, for slowly progressing patients in our simulation study, we observed that the schedule based on expected time of GR conducts on average two biopsies and has an average offset of 10 months. In comparison, annual schedule conducts six biopsies on average and gives an offset smaller by only four months, making the personalized schedule a suitable alternative.

For high risk patients however, early detection (annual or PRIAS schedule) may be necessary, given the rapidness of progression. When it is not known in advance if a patient will have a fast or slow progression of PCa, the hybrid approach may be used. It conducts one biopsy less than the annual schedule in faster progressing PCa patients and has an average offset of 10.25 months. For slowly progressing PCa patients it conducts two biopsies less than the annual schedule and has an average offset of 8.55 months.

References

Gentleman, Robert and Charles J Geyer (1994). "Maximum likelihood for interval censored data: Consistency and computation". In: *Biometrika* 81.3, pp. 618–623.

Huang, Xianzheng, Leonard A Stefanski, and Marie Davidian (2009). "Latent-model robustness in joint models for a primary endpoint and a longitudinal process". In: *Biometrics* 65.3, pp. 719–727.

Klotz, Laurence et al. (2009). "Clinical results of long-term follow-up of a large, active surveillance cohort with localized prostate cancer". In: *Journal of Clinical Oncology* 28.1, pp. 126–131.

Loeb, Stacy et al. (2016). "Immediate versus delayed prostatectomy: Nationwide population-based study". In: *Scandinavian journal of urology* 50.4, pp. 246–254.

Proust-Lima, Cécile et al. (2014). "Joint latent class models for longitudinal and time-to-event data: A review". In: *Statistical methods in medical research* 23.1, pp. 74–90.

Razali, Ahmad Mahir and Ali A Al-Wakeel (2013). "Mixture Weibull distributions for fitting failure times data". In: *Applied Mathematics and Computation* 219.24, pp. 11358–11364.

Rizopoulos, Dimitris (2012). *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. CRC Press.

Rizopoulos, Dimitris, Geert Verbeke, and Geert Molenberghs (2008). "Shared parameter models under random effects misspecification". In: *Biometrika* 95.1, pp. 63–74.

Tosoian, Jeffrey J et al. (2011). "Active surveillance program for prostate cancer: an update of the Johns Hopkins experience". In: *Journal of Clinical Oncology* 29.16, pp. 2185–2190.

Tsiatis, Anastasios A and Marie Davidian (2004). "Joint modeling of longitudinal and time-to-event data: an overview". In: *Statistica Sinica* 14.3, pp. 809–834.