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

May 22, 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 a 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 the 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 discussed suitable values of the number of biopsies and offset in the revised version of the manuscript. Lastly, we updated our graphs and their captions as per the suggestions of the Reviewers. The previous version was 24 pages long and you have asked us to reduce that to 20 pages (including the body of manuscript, acknowledgments, and references). While we have included new pieces of information according to the suggestions of the reviewers, we have managed to reduce the length of the paper to 20.4

pages. We hope that this is acceptable. Lastly, we have updated the title of our manuscript to "Personalized Schedules for Surveillance of Low-Risk Prostate Cancer Patients".

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 responses to the specific issues raised.

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

   The AE raises a very important point. A long delay of the biopsy time may have repercussions with respect to the optimal timing of initiating active treatment. In the specific context of prostate cancer, and given the characteristics of the disease, delaying active treatment for 10-12 months is considered acceptable by the urologists. The symmetric loss functions give similar delay for slowly-progressing patients (subgroup G_3 in Table 1 of the original manuscript). The acceptable delay of 10-12 months has also driven the specific choices we have made for the optimal cut-off value for individualized risk predictions in Section 3.3 of the original manuscript. However, in other contexts such a delay may not be acceptable, leading to another choice of the optimal cut-point. From a methodological perspective, accounting for the effect of delaying biopsy to survival would require modeling the survival outcome, and extending the relative risk sub-model in the specification of the joint model into a multi-state process. We are currently working on these extensions, but it falls outside the scope of our current paper.

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 responses 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 under the current heading. The abbreviation PSA is used instead of prostate-specific antigen.

   With regards to the assumption of normality of error term, we used residual diagnostics to check this assumption. The left panel of Figure 1 shows the quantile-quantile (q-q) plot of subject-specific residuals under our original joint model. This figure suggests that a symmetric long-tailed distribution for errors is more plausible than the normal distribution. Based on this result, we fitted two more joint models, both with t-distributed errors. However the t-distribution in one of them had 4 degrees of freedom and in another one, 3 df. 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 the 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). 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 parameter estimates for the new joint model in Web Appendix C of the revised supplementary material.

   Since the slope association between $\log_2$ PSA levels and hazard of Gleason reclassification

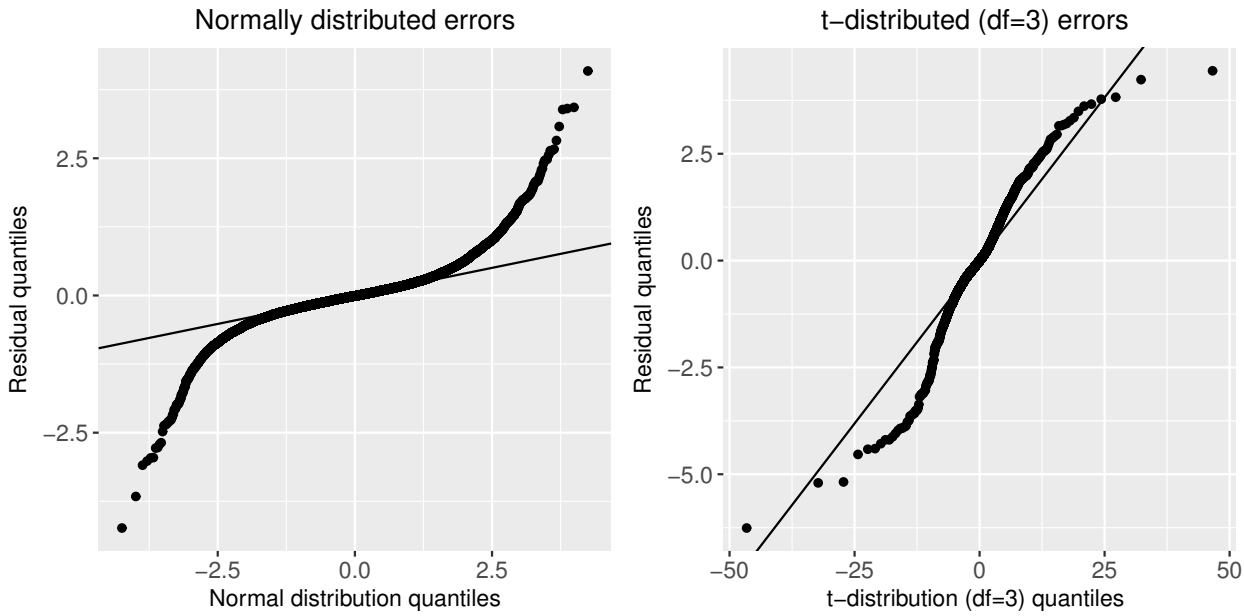

Figure 1: Quantile-quantile plots of subject specific residuals obtained from the joint models with assumption of normally distributed errors, and t-distributed (df=3) errors, fitted to the PRIAS data set.

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 the joint models with assumption of normally 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]

(GR) in the model with t-distributed (df=3) errors has become stronger, we expect our schedules to become slightly more sensitive towards an increase in $\log_2$ PSA velocity. However, this also depends on the type of personalized schedule. For example, we compared the personalized schedule based on the 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 the 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).

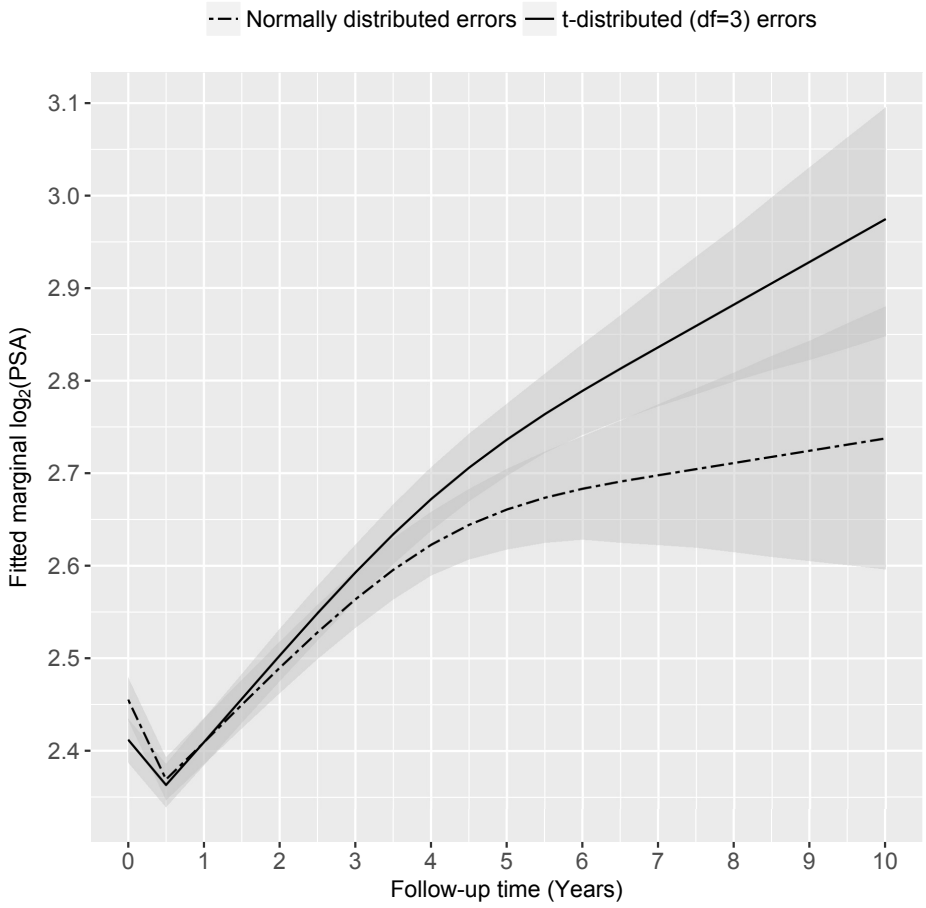

Figure 2: Fitted marginal $\log_2$ PSA profile with 95% credible interval (CI) over a 10 year follow-up period, for a hypothetical patient who was included in AS at the age of 70 years. Fits were obtained from joint models with the assumption of normally distributed errors, and t-distributed (df=3) errors.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

We next discuss the impact of the new association parameter on the schedules for the three demonstration patients.

We can see in Figure 4 (bottom row) and Figure 5 that the third demonstration patient has a consistent profile, with a 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).

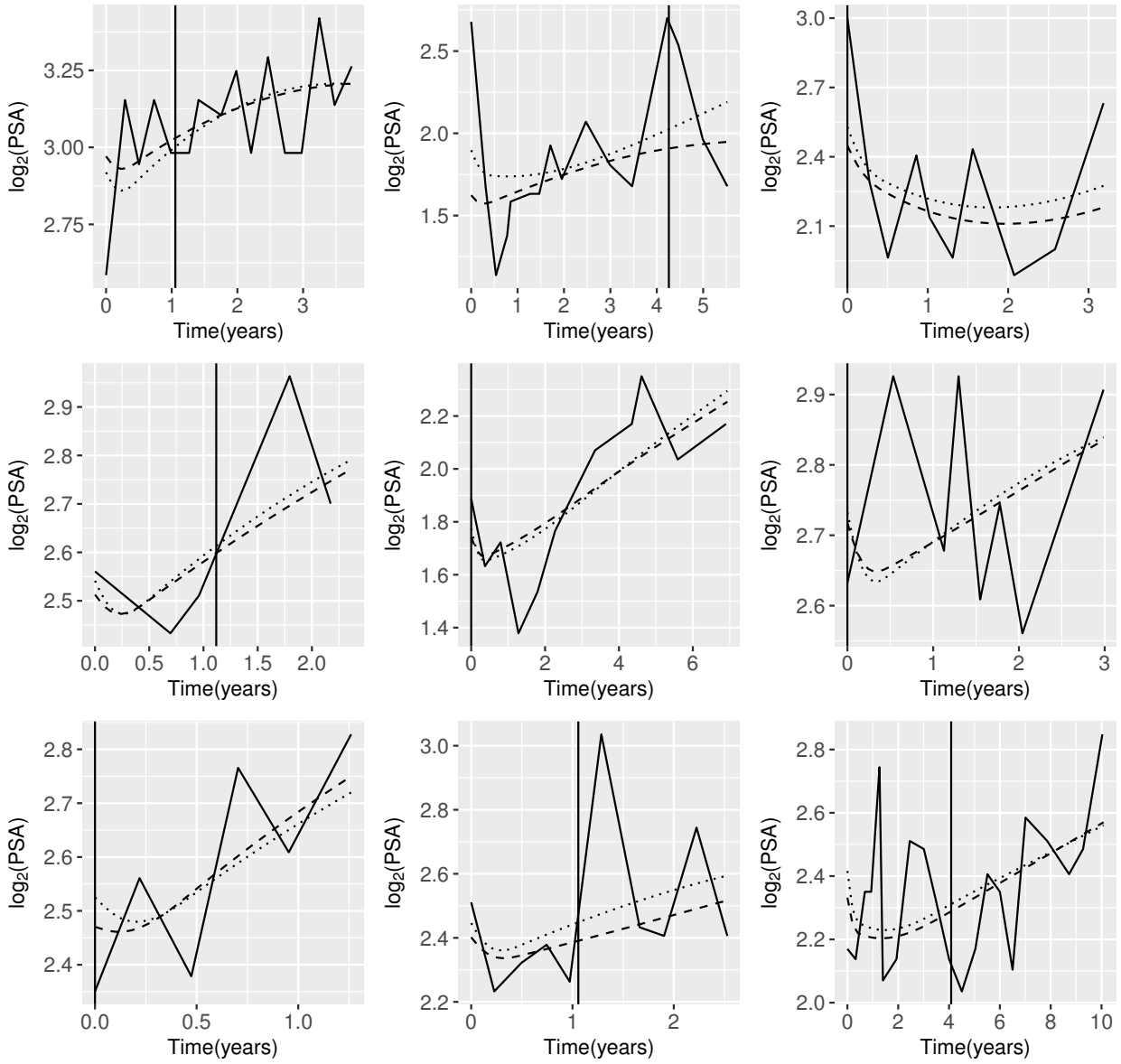

Figure 3: Fitted versus observed $\log_2$ PSA profiles for 9 randomly selected patients. Fits were obtained from joint models with assumption of normally distributed errors, and t-distributed (df=3) errors. The fitted profiles utilize information from both the observed PSA levels and time of latest biopsy.

Lastly, this can also be seen in the second demonstration patient wherein the schedules differ by a large margin initially when the PSA rises very quickly. However, the gap becomes slightly smaller after a negative biopsy indicating that GR is unlikely in near future. Thus we 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.

With regards assumption of normality of random effects, joint models 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 patient is extremely small, such misspecification only and trivially affects the standard errors. In our dataset, we have a mean of 8.7 measurements per patient, which makes us feel confident with regards 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, the first demonstration patient had many repeat biopsies, and thus via his profile, we show how the variance of the 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 are not in concordance with each other.

   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, if our method proposes a time u of the biopsy, we cannot conduct it at time u because biopsies are already conducted for the patients as per PRIAS schedule. Secondly, 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. In this regard, the simulation study is our attempt to objectively evaluate our proposed method versus the fixed-schedule approach.

3. Stratified relative risk model for modeling baseline hazards in the simulation study.

   We would like to thank the Reviewer for raising this point. Indeed in our simulation study,

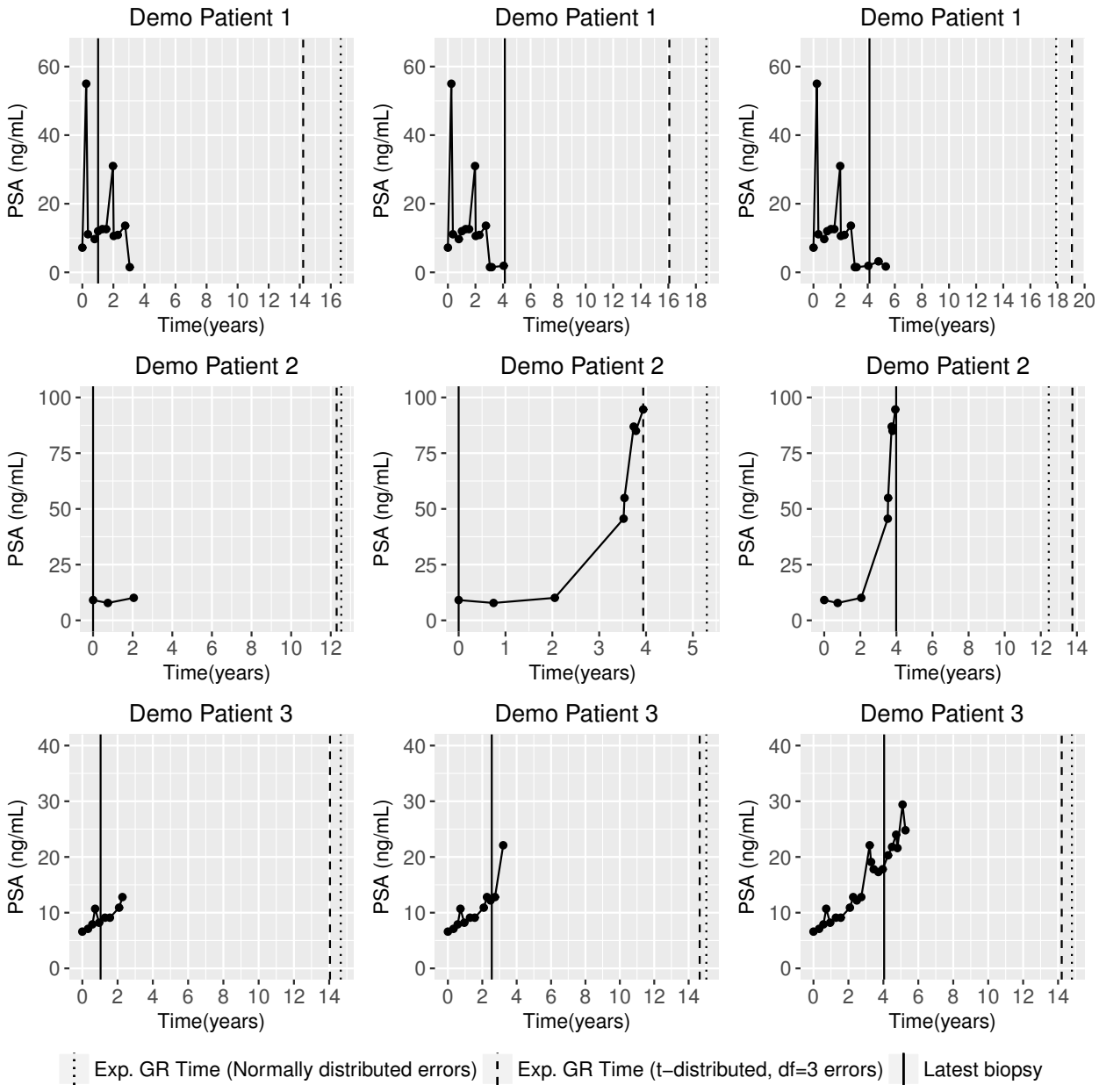

Figure 4: Dynamic expected failure time for the three demonstration patients at three different follow-up times, using joint models with assumption of normally distributed errors, and t-distributed (df=3) errors.

we have assumed 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

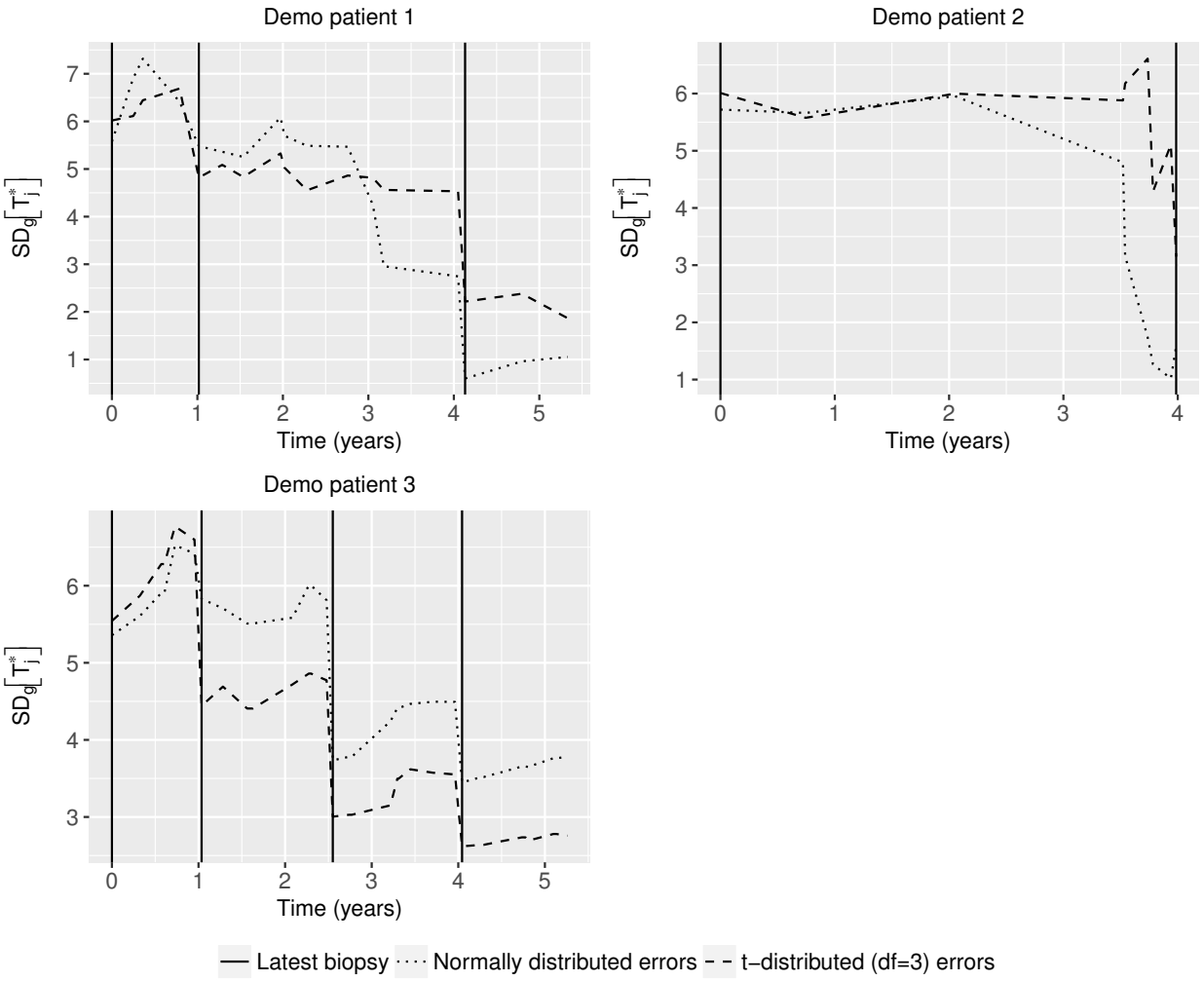

Figure 5: Dynamic variance of the posterior predictive distribution of event time for the three demonstration patients at three different follow-up times, using joint models with assumption of normally distributed errors, and t-distributed (df=3) errors.

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 correctly advised by the Referee these can be modeled using a stratified modeling approach. In the current case, this corresponds to the use of latent class joint models (Proust-Lima et al., 2014). Even though our approach could also be formulated under the latent class model, this extension falls outside the scope of our paper that primarily aims to introduce our proposed procedure for personalized scheduling of biopsies. Moreover, we expect that our postulated

penalized B-spline approximation of the log baseline hazard (see Web Appendix A) captures the assumed mixture of Weibull hazards adequately. Figure 6 shows a comparison between the fitted and theoretical hazard. We observe that the fit of the B-spline is close to the theoretical baseline hazard.

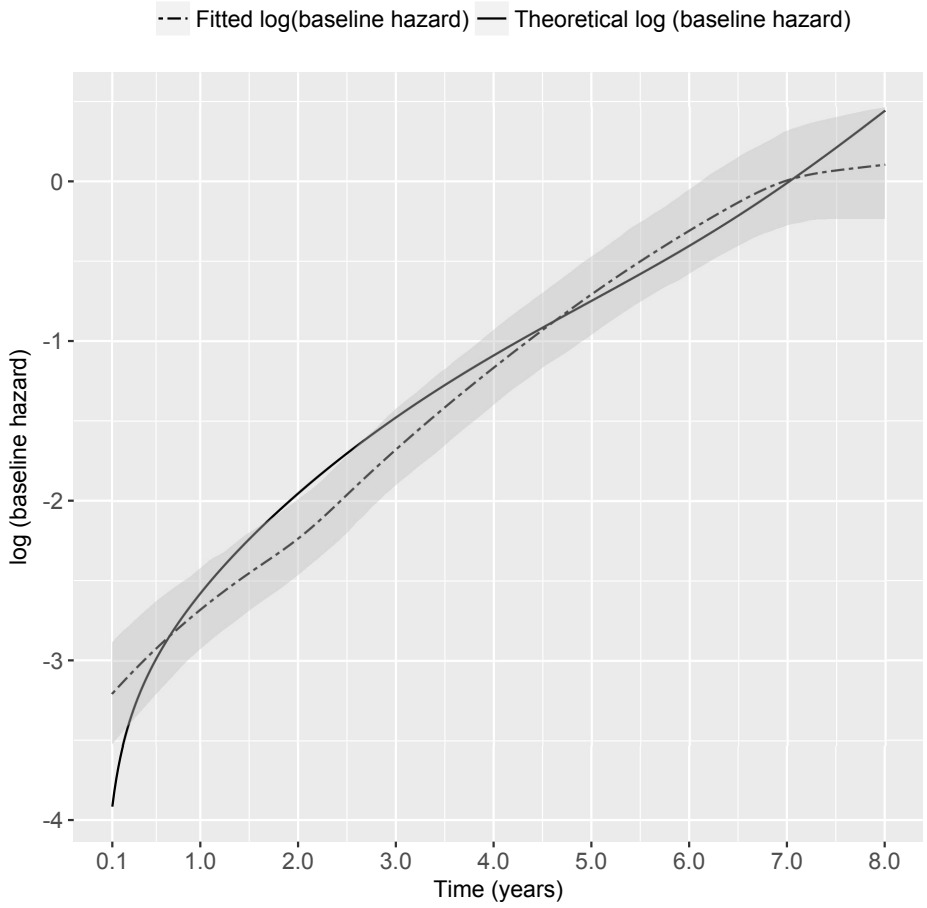

Figure 6: Theoretical log baseline hazard of the simulated population versus mean of the fitted log baseline hazard. The 95% confidence interval for the fitted log baseline hazard is obtained from the 500 simulations.

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

   The Reviewer raises an important point. Indeed PRIAS switches to the more frequent annual schedule if a patient's PSA doubling time (PSA-DT), measured as the inverse of the slope of the regression line through the base two logarithm of PSA values, is less than 10 years. This raises the concern of (ascertainment) bias caused by the fact that the schedule of biopsies

depends on past PSA levels. Nevertheless, working under the framework of joint models has the advantageous feature that we can ignore the PSA-DT process and obtain valid parameter estimates, under the condition that the model is correctly specified. This is due to the fact that the joint model uses a full likelihood specification for the longitudinal and event time processes (Tsiatis and Davidian, 2004). To show this, consider the following full general specification of the joint model 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 denote 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:

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

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, the second term only contributes a constant in the likelihood and hence can be ignored. To check if we correctly specified the joint model, we performed several sensitivity analysis in our model (e.g., changing the position of the knots, etc.) to investigate the fit of the model and also the robustness of the results. In all of our attempts, the same conclusions were reached, namely that the $\log_2$ PSA velocity is more strongly associated with the hazard of GR compared to the $\log_2$ PSA levels.

The Referee has also asked an interesting question about the use of right censored data in the simulation study instead of interval censored data, given that the PRIAS data is also interval censored. However, in the simulation study even if we would have generated interval censored data it would not have led 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), under the condition that the model is correctly specified. In the current context, we correctly use an interval censoring specification of the joint likelihood when the data is interval censored, and right censoring specification when the (simulated) data is right censored.

5. AUC comparison: model with value and velocity association versus only value association.

We thank the Referee for noticing the erroneous result that was reported. The two sets of

AUC's were mistakenly swapped while creating Web Table 3 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. The resulting estimates are presented in Table 2 below, as well as added in Supplementary material (Web Table 3).

Table 2: Area under the receiver operating characteristic curves (AUC), and 95% confidence interval in brackets. AUC's are calculated for two joint models: first one having association between hazard of GR and $\log_2$ PSA value as well as velocity, and second one having association with only $\log_2$ PSA value.

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 corrected these in the revised version of the 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 responses 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 assumptions 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. However, this is the general form of the equation for the longitudinal sub-model and is only used to introduce the joint model notation. The actual equation, showing the log-transformed PSA levels, baseline covariates and B-spline for the effect of time is Equation (2) below (it is Equation 7 in the revised 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.

$$\begin{aligned}\log_2 \text{PSA}(t) = \beta_0 + \beta_1(\text{Age} - 70) + \beta_2(\text{Age} - 70)^2 + \sum_{k=1}^{4} \beta_{k+2} B_k(t, \mathcal{K}) \\ + b_{i0} + b_{i1} B_7(t, 0.1) + b_{i2} B_8(t, 0.1) + \varepsilon_i(t),\end{aligned} \tag{2}$$

Since concerns regarding the 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 7. In addition, the fitted marginal $\log_2$ PSA profiles, and subject-specific fitted versus observed $\log_2$ PSA profiles of 9 randomly selected patients (each with more than 3 observations), using the two different models are presented in Figure 8 and Figure 9, respectively.

With regards to the fitted profiles for the three demonstration patients, we show their fitted profiles in Figure 10. 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

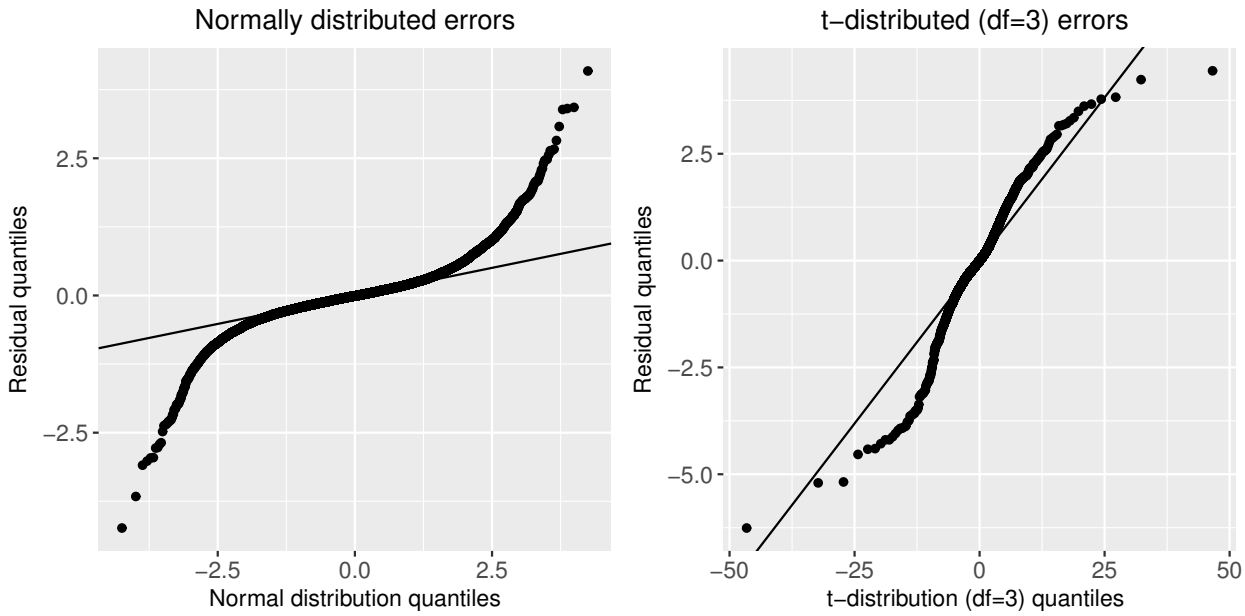

Figure 7: Quantile-quantile plots of subject specific residuals obtained from joint models with assumption of normally distributed errors, and t-distributed (df=3) errors, fitted to the PRIAS data set.

these patients in the main 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 revised supplementary material as well.

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 measurement after the 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

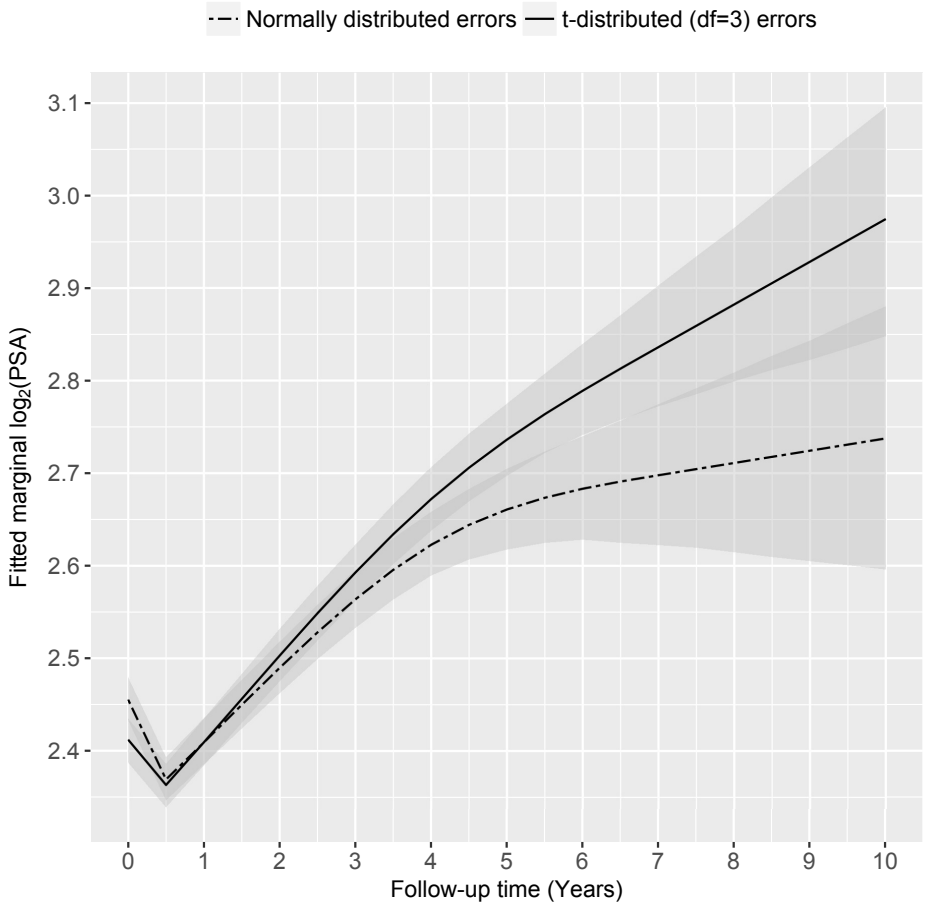

Figure 8: Fitted marginal 10 year $\log_2$ PSA profile with 95% credible interval (CI), for a hypothetical patient who was included in AS at the age of 70 years. Fits were obtained from joint models with assumption of normal distributed errors, and t-distributed (df=3) errors. The darker shaded region indicates the overlap in the two CI intervals, as well as demarcates the two sets of CIs.

as the referee noted it is possible that $t < T_j^* \leq s$, in which case a biopsy should be conducted immediately. However, it is often the case that difference between consecutive biopsies is required to be at least a 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 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 11 (it is Figure 1 in the revised manuscript).

We would also like to take the opportunity and provide extra clarification for the definition

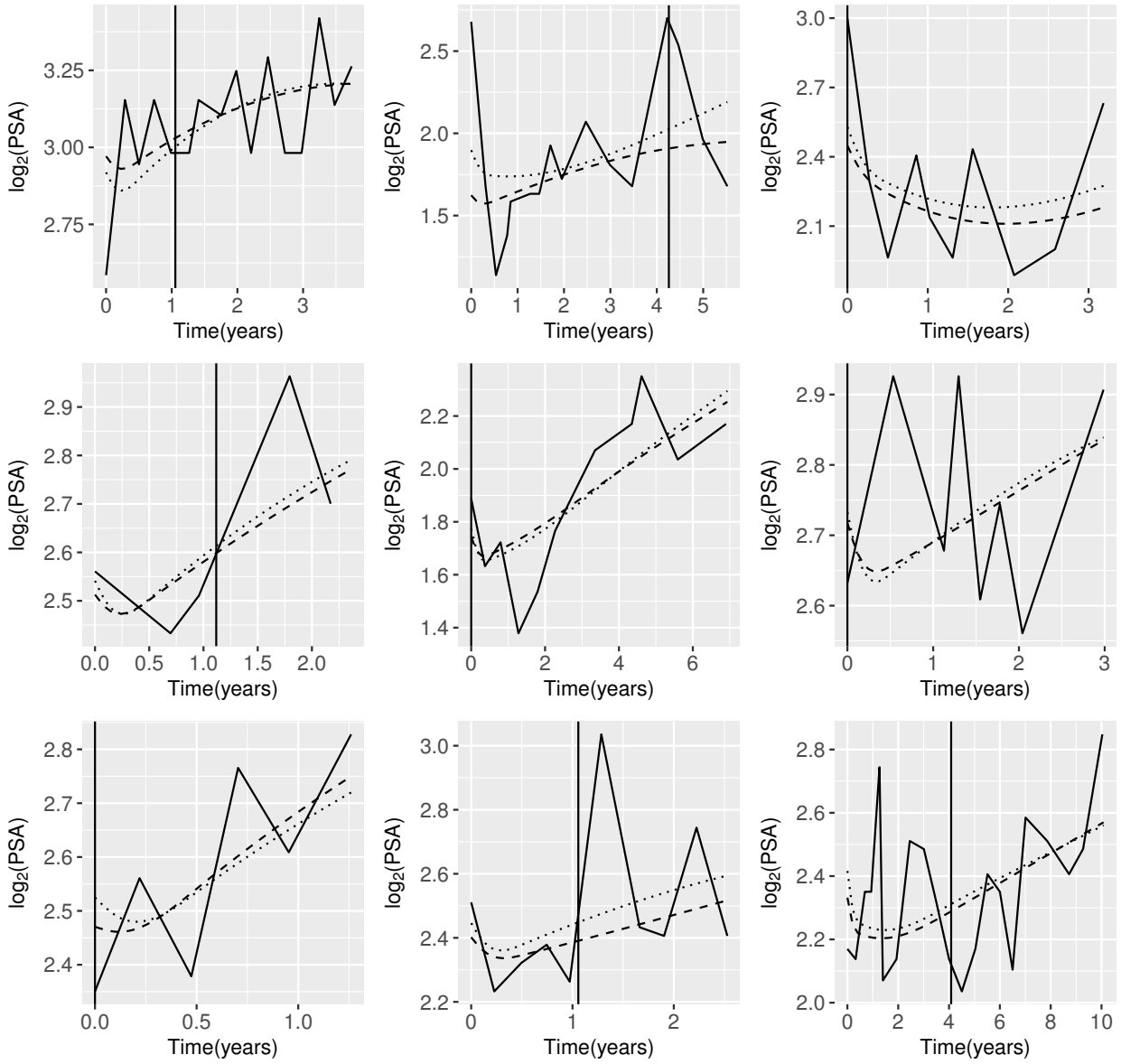

Figure 9: Fitted versus observed $\log_2$ PSA profiles for 9 randomly selected patients. Fits were obtained from joint models with assumption of normal distributed errors, and t-distributed (df=3) errors. The fitted profiles utilize information from both the observed PSA levels and time of latest biopsy.

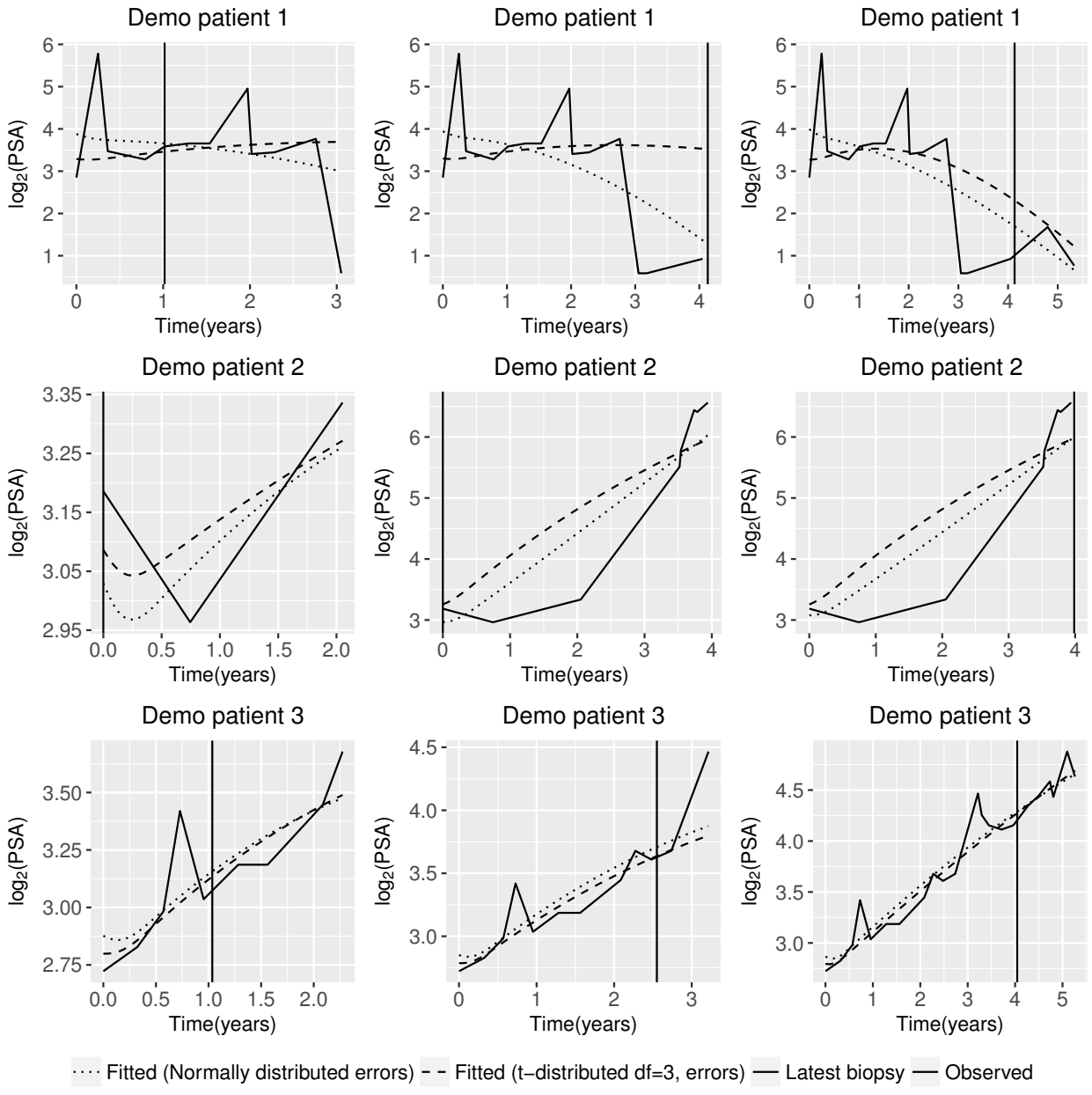

Figure 10: Fitted versus observed $\log_2$ PSA profiles for the three demonstration patients, at three different time points. The fitted profiles are dynamic in nature, and utilize information from both the observed PSA levels and time of latest biopsy.

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

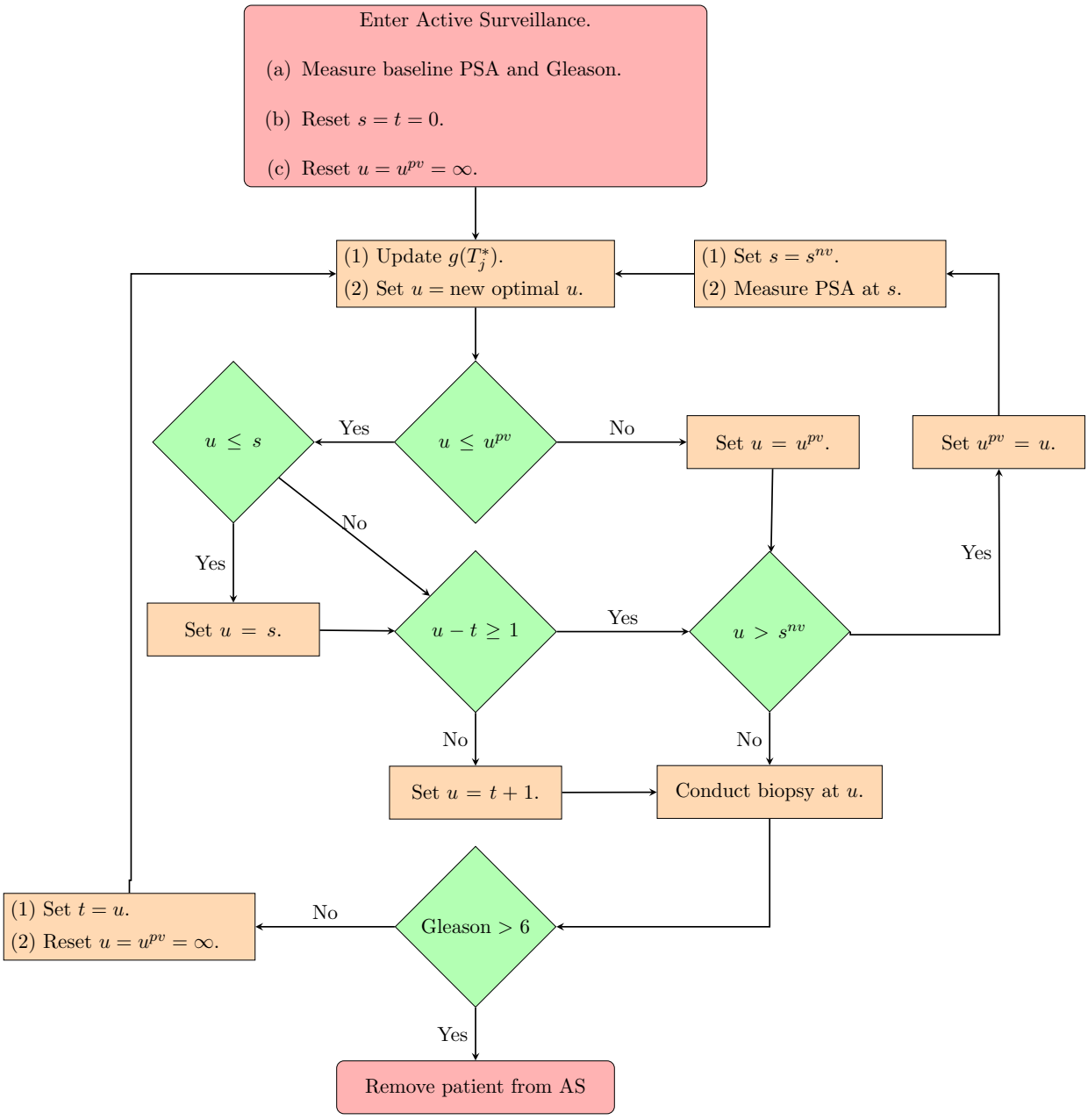

Figure 11: Algorithm for creating a personalized schedule for patient j. The time of the latest biopsy is denoted by t. The time of the latest available PSA measurement is denoted by s. The proposed personalized time of biopsy is denoted by u. The time at which a repeat biopsy was proposed on the last visit to the hospital is denoted by u^{pv}. The time of the next visit for the measurement of PSA is denoted by s^{nv}.

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:

$$f\{\mathcal{M}_i(t), \boldsymbol{b}_i, \boldsymbol{\alpha}\} = \alpha \int_0^t m_i(t)\mathrm{d}t \tag{3}$$

3. Robustness of the schedules based on the 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 the dynamic risk of GR are robust to large overshooting margins (offset). We observed in Figure 2 of the original manuscript that the variance of the 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 the 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 the dynamic risk of GR leads to almost the same mean offset and variance of offset across the three subgroups of patients.

   Due to space restrictions in the main manuscript, we have provided the following brief explanation. In practice, for some patients, we may not have sufficient information to accurately estimate their PSA profile. The resulting high variance of $g(T_j^*)$ could lead to a mean (or median) time of GR which overshoots the true T_j^* by a big margin. In such cases, the approach based on the dynamic risk of GR with smaller risk thresholds is more risk-averse and thus could be more robust to large overshooting margins.

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 revised manuscript. We specifically mention that these graphs pertain to the results of the simulation study.

2. 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 commonly recommended risk threshold of 5%. It is quite a risk-averse threshold, however as shown in Web Table 4 in the 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 the 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. The choice of offset cutoff C is discussed in point number 4 below.

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 addressed this issue in our revised 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.

4. Good and bad O_j^S and N_j^S

   The Reviewer raises a very important point with regards to the practicality of our approach. In this regard, 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 low-risk patients who obtain GR in the latter years of their follow-up. 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.

However, for high-risk patients, 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.

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.