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

June 11, 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.R1, 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 heavy tailed negative residuals evident in the quantile-quantile plot. To this end, we transformed the longitudinal outcome using a $\log_2(\text{PSA}+1)$ transformation instead of the original $\log_2 \text{PSA}$ transformation. We have updated the captions of figures and tables to make them more informative. We have also uploaded the R code with example data in a zip file, and moved the Supplementary Materials section to be the last numbered section before Acknowledgments. The previous version was 20.4 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 retain the length of the paper up to 20.4 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 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 C, D and E. In addition, changes regarding the specific comments have been made throughout the text.

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

### Major Concerns Shared by the 2nd Referee

1. Heavier tailed negative residuals in the quantile-quantile plot.

   We would like to thank the Referee for motivating us to check the model fit. As the Referee noted, the quantile-quantile plot for the residuals from the original model (top-left panel of Figure 1) shows heavy tailed negative residuals. Following the suggestion of the Referee we transformed the longitudinal outcome using a $\log_2(\text{PSA}+0.1)$ transformation instead of the original $\log_2(\text{PSA})$ transformation. In addition, we also tried $\log_2(\text{PSA}+1)$ transformation (Lin et al., 2000; Pearson et al., 1994). The resulting quantile-quantile plots of the residuals are in shown in Figure 1. Since the residuals from the model with $\log_2(\text{PSA}+1)$ transformation met the assumptions best, we use it in the revised version of the manuscript. The corresponding longitudinal sub-model of the joint model we fit is given by:

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

   where the error $\varepsilon_i(t)$ is assumed to be t-distributed with three degrees of freedom and scale $\sigma$.

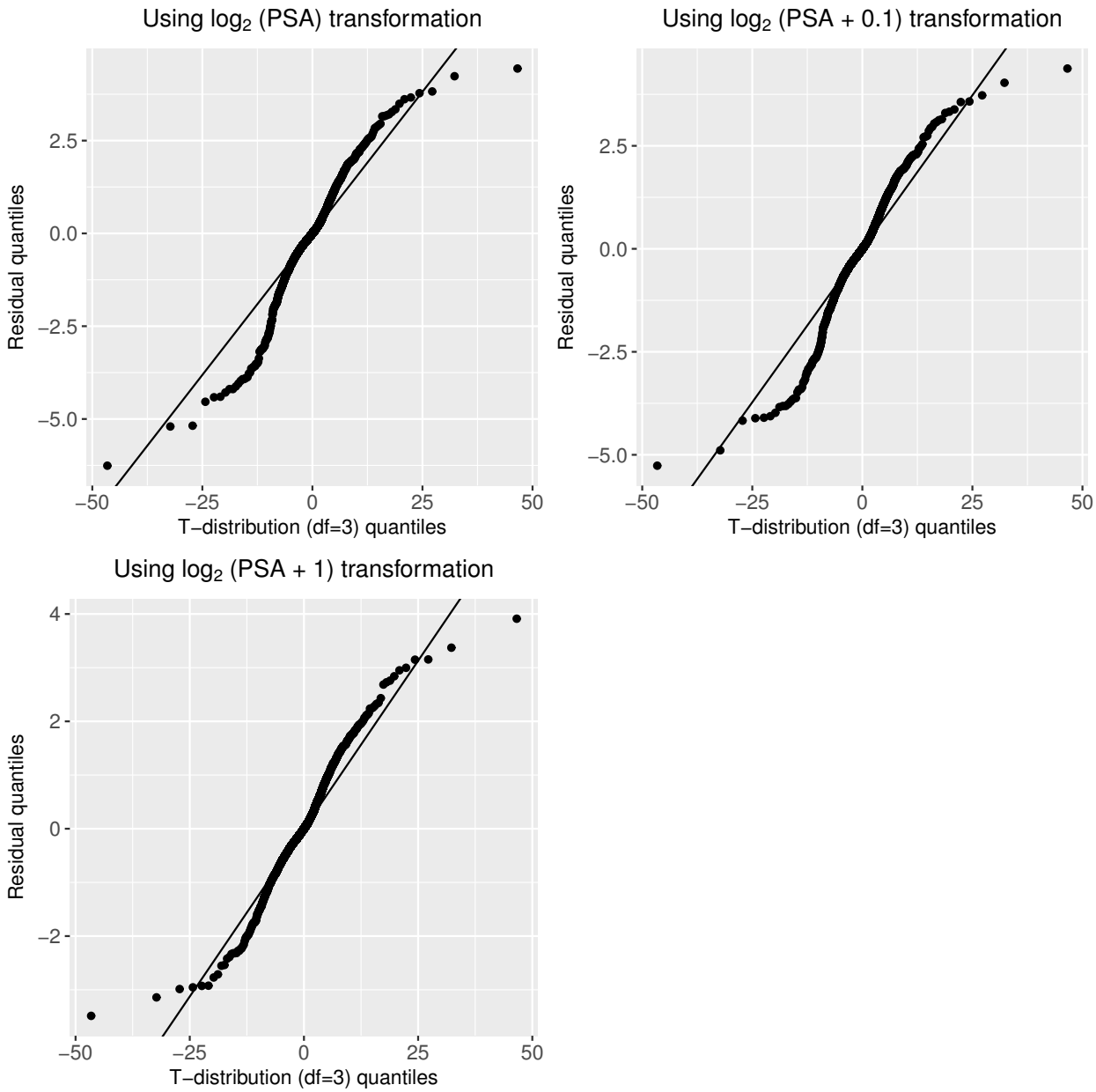


Figure 1: Quantile-quantile plots of subject-specific residuals obtained from joint models with $\log_2(\text{PSA})$, $\log_2(\text{PSA}+0.1)$, and $\log_2(\text{PSA}+1)$ transformed longitudinal outcome, and an assumption of t-distributed (df=3) errors, fitted to the PRIAS data set.

2. Explanation regarding fitted PSA profiles

We thank the Referee for motivating us to check the per patient fit of PSA profiles. The Referee expressed concern for the "fitted mean PSA values and velocities over the time" of the three demonstration patients. The original figure (Figure 10 of the previous response letter) mentioned by the Referee is shown as Figure 2 below. As we can see the mean fitted values are not as close to the observed values when the model with t-distributed errors is used in comparison to the model with normal distributed errors. The first reason for this behavior is that the t-distribution is more robust to sudden changes in the trend of the profiles, since the error distribution has longer tails (Lange, Little, and Taylor, 1989). Secondly, as in the case of Demo patient 1 (top panel of Figure 2) the actual trend may only become evident after enough observations show a consistent decrease in PSA levels. This is also the explanation for relatively better fits when number of measurements are more. To further mitigate concerns regarding the model fit, we present fitted profiles of eighteen randomly chosen PRIAS patients in Figure 3 and Figure 4. It can be seen that the model fit for all patients is reasonable.

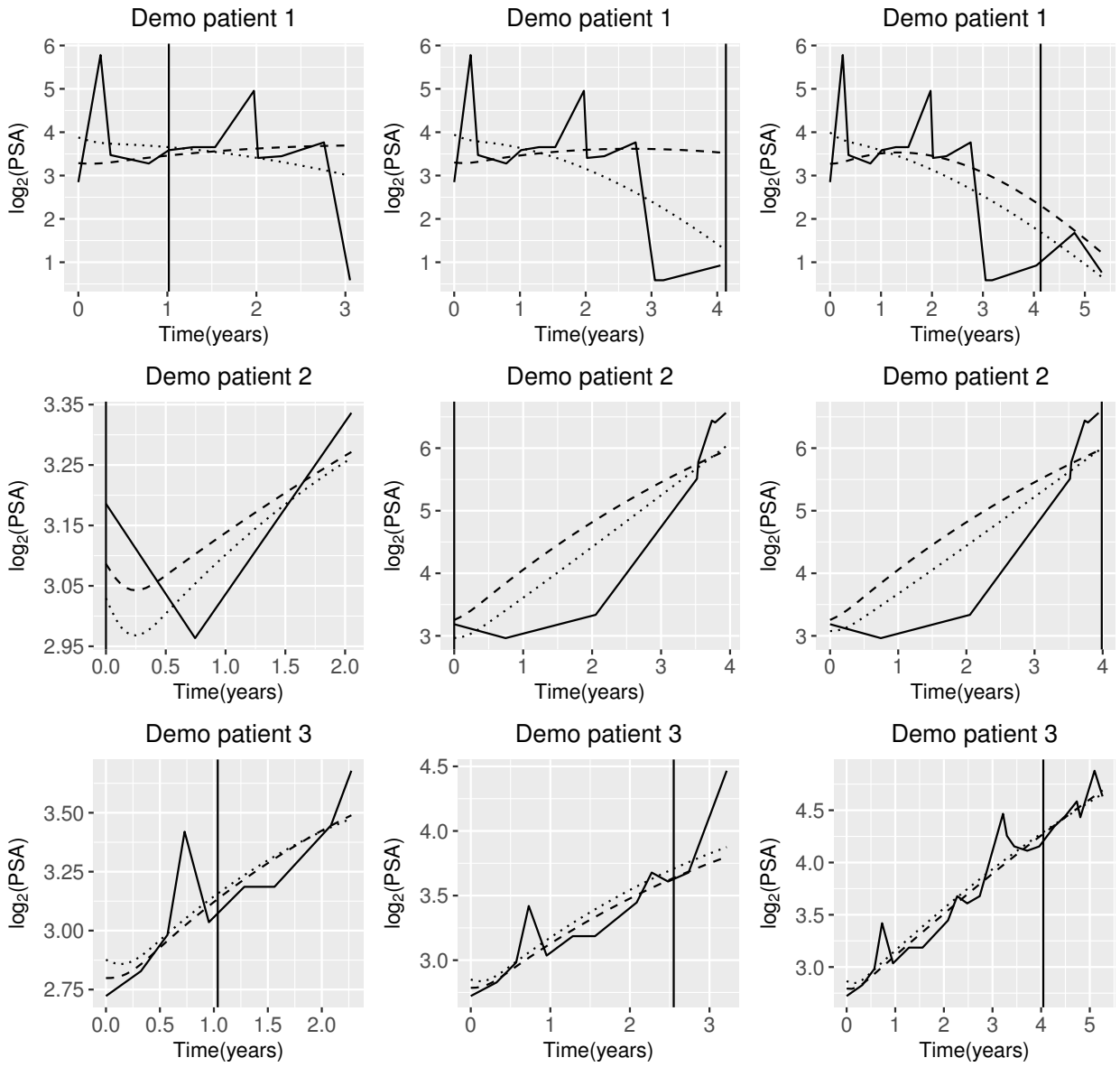


Figure 2: This is Figure 10 from the previous response letter. It shows fitted versus observed $\log_2$ PSA profiles for the three demonstration patients, at three different time points. The fitted profiles are dynamic in nature, and utilize information from both the observed PSA levels and time of latest biopsy.

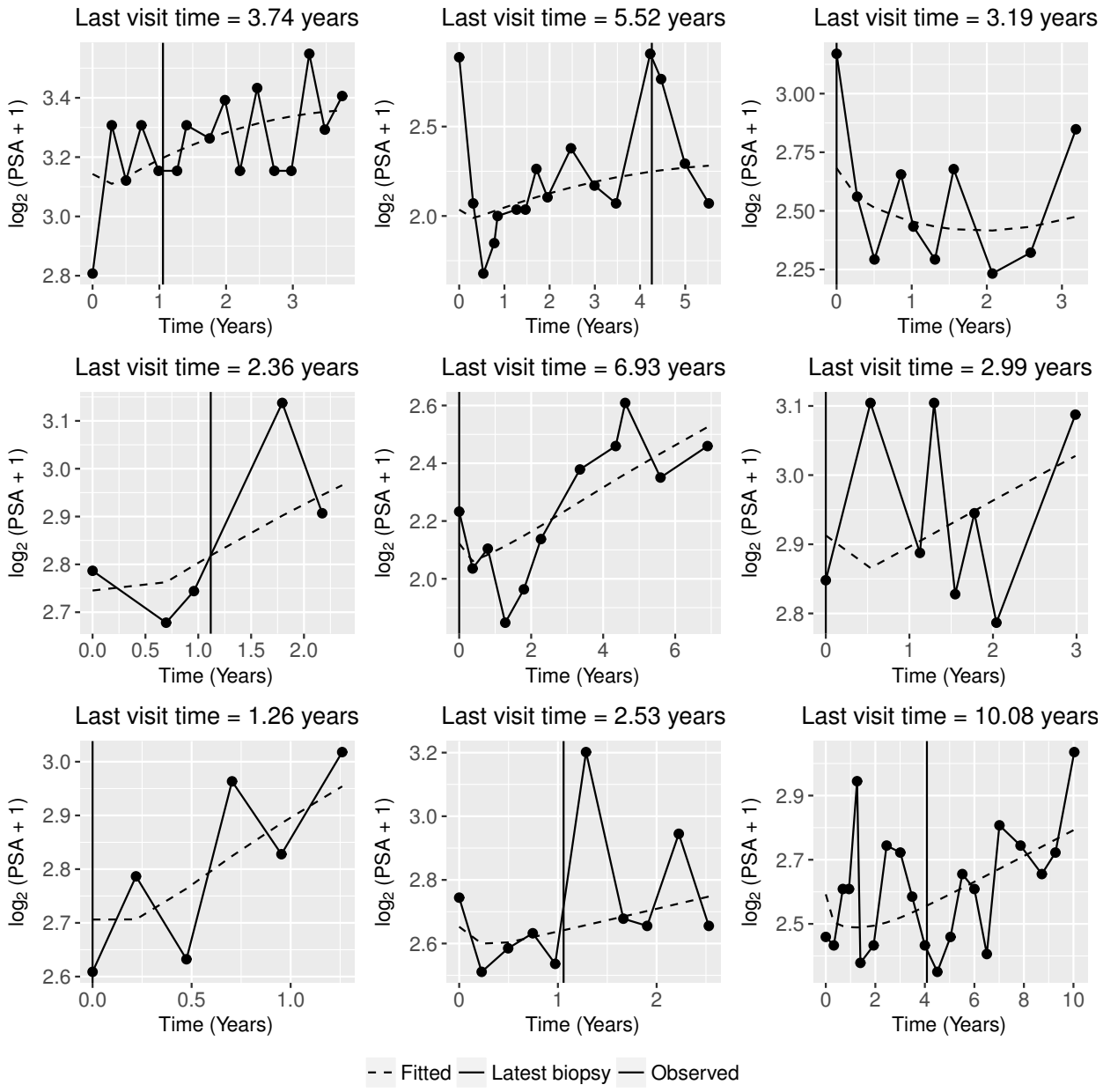


Figure 3: Fitted versus observed $\log_2(\text{PSA}+1)$ profiles of nine randomly selected PRIAS patients. The fitted profiles utilize information from both the observed PSA levels and time of latest biopsy.

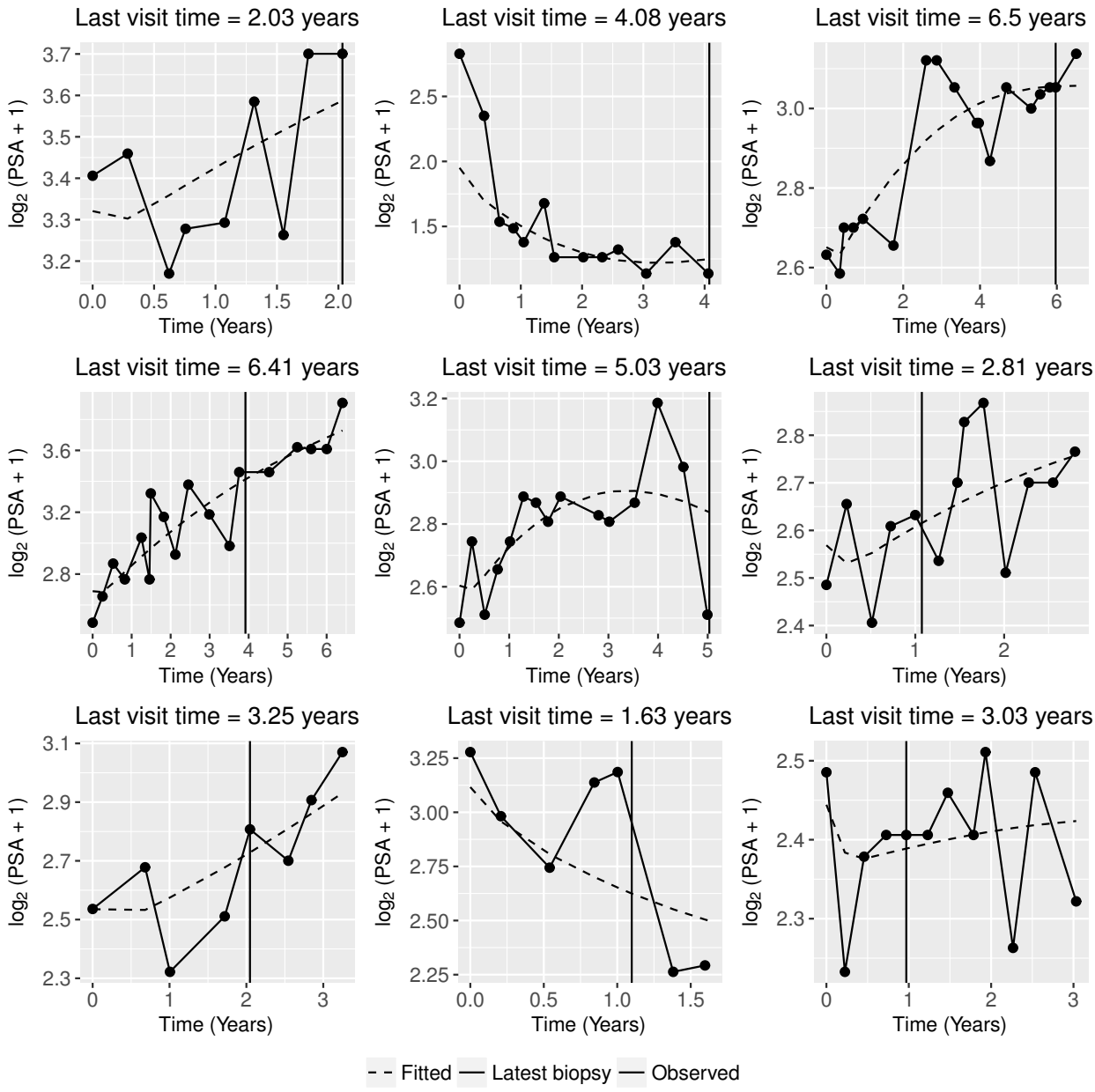


Figure 4: Fitted versus observed $\log_2(\text{PSA}+1)$ profiles of nine randomly selected PRIAS patients different than the ones shown in Figure 3. The fitted profiles utilize information from both the observed PSA levels and time of latest biopsy.

### Minor Concerns Shared by the 2nd Referee

1. More informative captions for Table 1 and Figure 3-5.

   We have updated the captions of Table 1 and Figure 3-5 in the revised manuscript. The revised captions along with Table/Figures are shown in Table 1 and Figure 5-7 in this reply letter.

2. Missing subscript $i$ in Equation 7 of the original manuscript.

   We thank the Referee for noticing this error. Equation (1) in this reply letter shows the new equation that we use in the revised version of the manuscript.

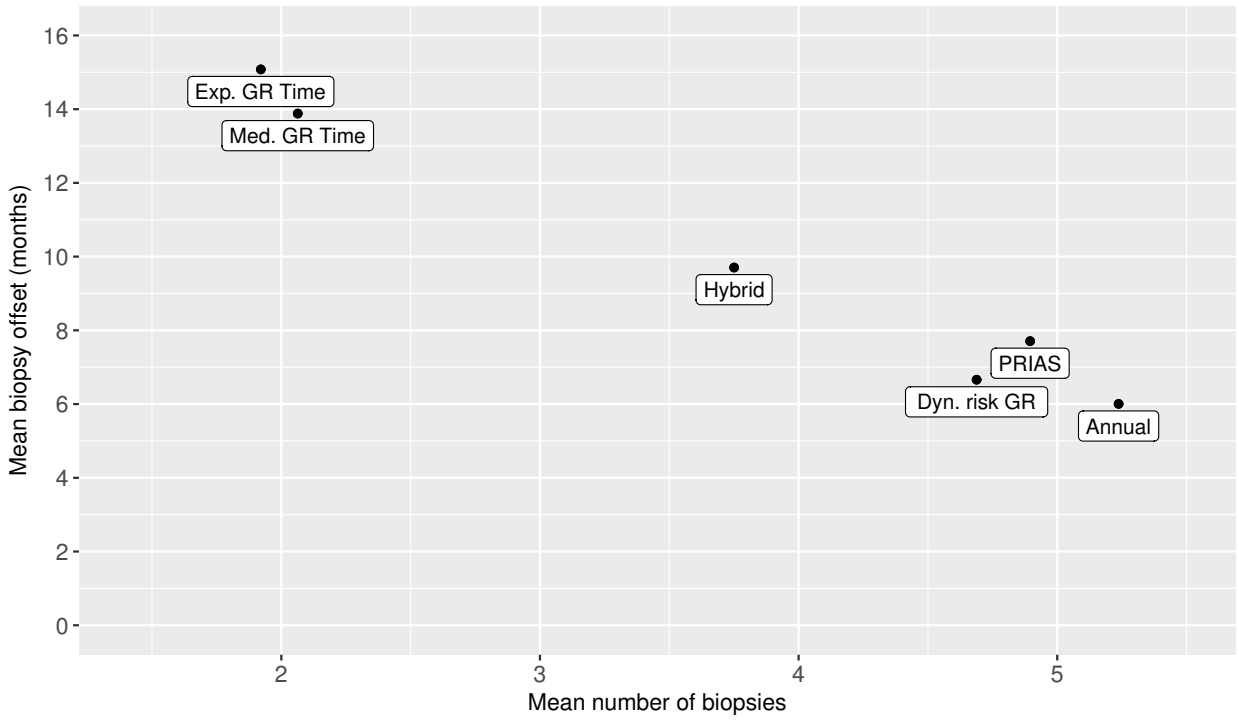


Figure 5: Estimated mean number of biopsies conducted until Gleason reclassification (GR) is detected, and mean offset (difference in time at which GR is detected and the true time of GR, in months) for the simulated (500 datasets) test patients, across different schedules. Types of personalized schedules (full names in brackets): Exp. GR Time (expected time of GR), Med. GR Time (median time of GR), Dyn. risk GR (schedules based on dynamic risk of GR), Hybrid (a hybrid approach between median time of GR and dynamic risk of GR). Annual corresponds to a schedule of yearly biopsies and PRIAS corresponds to biopsies as per PRIAS protocol.

Table 1: Estimated mean and standard deviation (SD), of the number of biopsies $N_j^S$ conducted until Gleason reclassification (GR) is detected, and of the offset $O_j^S$ (difference in time at which GR is detected and the true time of GR, in months), for the simulated (500 datasets) test patients, across different schedules and subgroups. Patients in subgroup $G_1$ have the fastest prostate cancer progression rate, whereas patients in subgroup $G_3$ have the slowest progression rate. Types of personalized schedules (full names in brackets): Exp. GR Time (expected time of GR), Med. GR Time (median time of GR), Dyn. risk GR (schedules based on dynamic risk of GR), Hybrid (a hybrid approach between median time of GR and dynamic risk of GR). Annual corresponds to a schedule of yearly biopsies and PRIAS corresponds to biopsies as per PRIAS protocol.

| a) All hypothetical subgroups | | | | |
|---|---|---|---|---|
| Schedule | $E(N_j^S)$ | $E(O_j^S)$ | SD($N_j^S$) | SD($O_j^S$) |
| Annual | 5.24 | 6.01 | 2.53 | 3.46 |
| PRIAS | 4.90 | 7.71 | 2.36 | 6.31 |
| Dyn. risk GR | 4.69 | 6.66 | 2.19 | 4.38 |
| Hybrid | 3.75 | 9.70 | 1.71 | 7.25 |
| Med. GR time | 2.06 | 13.88 | 1.41 | 11.80 |
| Exp. GR time | 1.92 | 15.08 | 1.19 | 12.11 |
| b) Hypothetical subgroup $G_1$ | | | | |
| Schedule | $E(N_j^S)$ | $E(O_j^S)$ | SD($N_j^S$) | SD($O_j^S$) |
| Annual | 4.32 | 6.02 | 3.13 | 3.44 |
| PRIAS | 4.07 | 7.44 | 2.88 | 6.11 |
| Dyn. risk GR | 3.85 | 6.75 | 2.69 | 4.44 |
| Hybrid | 3.25 | 10.25 | 2.16 | 8.07 |
| Med. GR time | 1.84 | 20.66 | 1.76 | 14.62 |
| Exp. GR time | 1.72 | 21.65 | 1.47 | 14.75 |
| c) Hypothetical subgroup $G_2$ | | | | |
| Schedule | $E(N_j^S)$ | $E(O_j^S)$ | SD($N_j^S$) | SD($O_j^S$) |
| Annual | 5.18 | 5.98 | 2.13 | 3.47 |
| PRIAS | 4.85 | 7.70 | 2.00 | 6.29 |
| Dyn. risk GR | 4.63 | 6.66 | 1.82 | 4.37 |
| Hybrid | 3.68 | 10.32 | 1.37 | 7.45 |
| Med. GR time | 1.89 | 12.33 | 1.16 | 9.44 |
| Exp. GR time | 1.77 | 13.54 | 0.98 | 9.83 |
| d) Hypothetical subgroup $G_3$ | | | | |
| Schedule | $E(N_j^S)$ | $E(O_j^S)$ | SD($N_j^S$) | SD($O_j^S$) |
| Annual | 6.20 | 6.02 | 1.76 | 3.46 |
| PRIAS | 5.76 | 7.98 | 1.71 | 6.51 |
| Dyn. risk GR | 5.58 | 6.58 | 1.56 | 4.33 |
| Hybrid | 4.32 | 8.55 | 1.26 | 5.91 |
| Med. GR time | 2.45 | 8.70 | 1.15 | 6.32 |
| Exp. GR time | 2.27 | 10.09 | 0.99 | 7.47 |

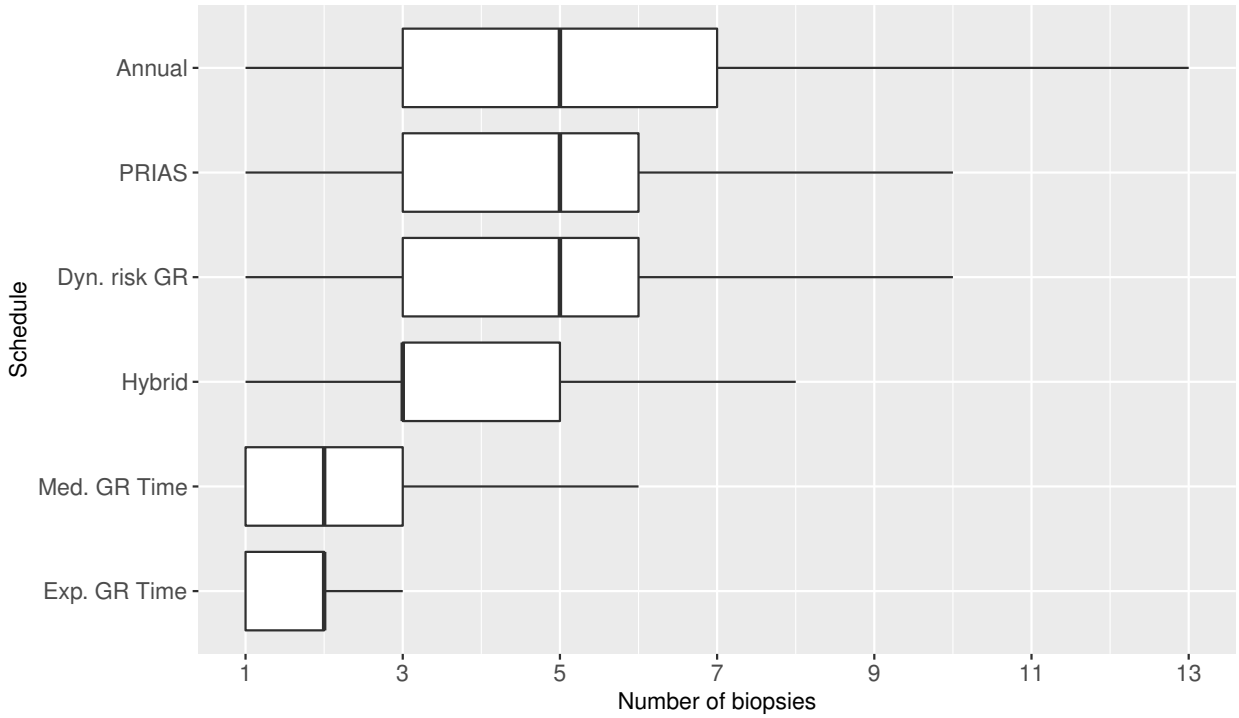


Figure 6: Boxplot showing variation in number of biopsies conducted by various biopsy schedules for the simulated (500 datasets) test patients. Biopsies are conducted until Gleason reclassification (GR) is detected. Types of personalized schedules (full names in brackets): Exp. GR Time (expected time of GR), Med. GR Time (median time of GR), Dyn. risk GR (schedules based on dynamic risk of GR), Hybrid (a hybrid approach between median time of GR and dynamic risk of GR). Annual corresponds to a schedule of yearly biopsies and PRIAS corresponds to biopsies as per PRIAS protocol.

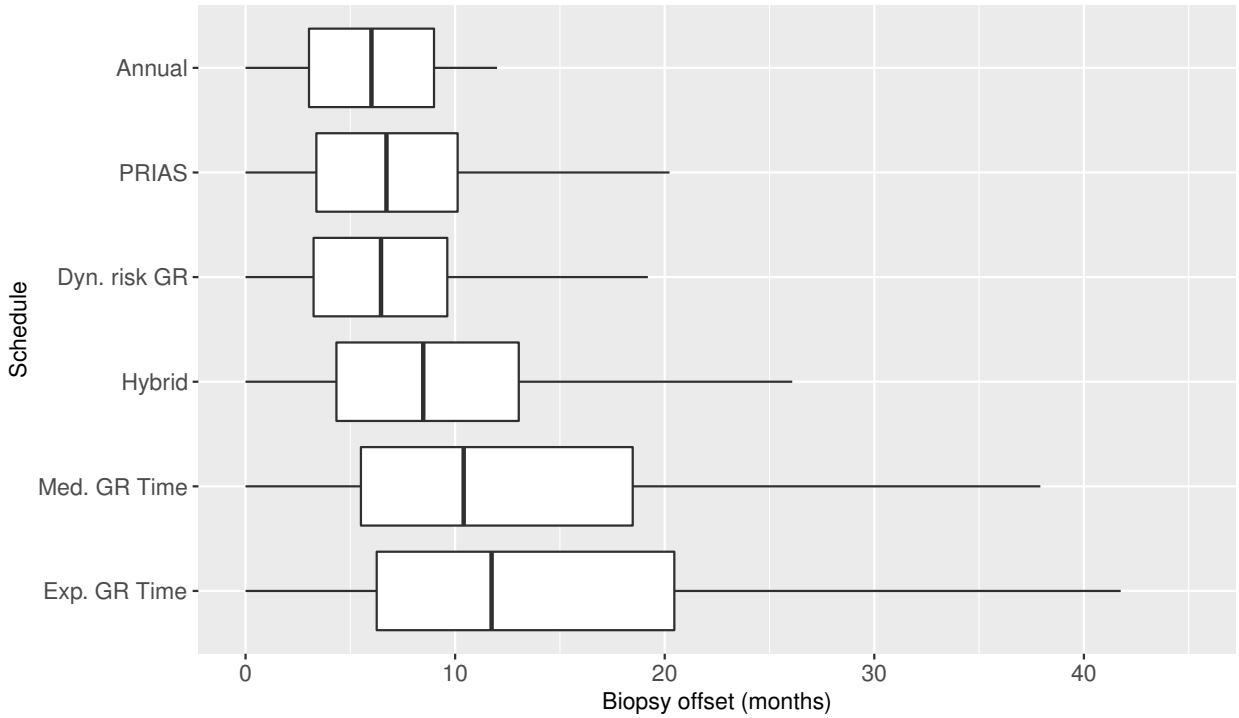


Figure 7: Boxplot showing variation in biopsy offset (difference in time at which Gleason reclassification, also known as GR, is detected and the true time of GR, in months) for the simulated (500 datasets) test patients, across different schedules. Types of personalized schedules (full names in brackets): Exp. GR Time (expected time of GR), Med. GR Time (median time of GR), Dyn. risk GR (schedules based on dynamic risk of GR), Hybrid (a hybrid approach between median time of GR and dynamic risk of GR). Annual corresponds to a schedule of yearly biopsies and PRIAS corresponds to biopsies as per PRIAS protocol.

## References

Lange, Kenneth L, Roderick JA Little, and Jeremy MG Taylor (1989). "Robust statistical modeling using the t distribution". In: *Journal of the American Statistical Association* 84.408, pp. 881–896.

Lin, Haiqun et al. (2000). "A latent class mixed model for analysing biomarker trajectories with irregularly scheduled observations". In: *Statistics in Medicine* 19.10, pp. 1303–1318.

Pearson, Jay D et al. (1994). "Mixed-effects regression models for studying the natural history of prostate disease". In: *Statistics in Medicine* 13.5-7, pp. 587–601.