Personalized Biopsies in Prostate Cancer Active Surveillance[☆]

Anirudh Tomer, MSc[a,*], Dimitris Rizopoulos, PhD[a], Daan Nieboer, MSc[b], Monique J. Roobol, PhD[c], Au Thor[d], Aut Hor[d], Auth Or[d]

[a]*Department of Biostatistics, Erasmus University Medical Center, Rotterdam, the Netherlands*
[b]*Department of Public Health, Erasmus University Medical Center, Rotterdam, the Netherlands*
[c]*Department of Urology, Erasmus University Medical Center, Rotterdam, the Netherlands*
[d]*Department of xxxx, xxxx University Medical Center, City, Country*

Abstract

`Background`: Low-risk prostate cancer patients enrolled in active surveillance (AS) undergo repeat biopsies. Treatment often advised when biopsy Gleason ≥ 7 (GS7). Many patients never obtain GS7, yet undergo biopsies frequently.

`Objective`: Better balance the number of biopsies and time delay in detection of GS7.

`Design, Setting, and Participants`: World's largest AS study PRIAS, 7813 patients, 1134 obtained GS7, 100 medical centers in 17 countries, common study protocol and online tool for data collection.

`Outcome Measurements, and Statistical Analysis`: Prostate-specific anti-

[☆]Word Count Abstract: 300; Word Count Text: 2800

[*]Corresponding author (Anirudh Tomer): Erasmus MC, kamer flex Na-2823, PO Box 2040, 3000 CA Rotterdam, the Netherlands. Tel: +31 10 70 43393

Email addresses: `a.tomer@erasmusmc.nl` (Anirudh Tomer, MSc), `d.rizopoulos@erasmusmc.nl` (Dimitris Rizopoulos, PhD), `d.nieboer@erasmusmc.nl` (Daan Nieboer, MSc), `m.roobol@erasmusmc.nl` (Monique J. Roobol, PhD)

gen (PSA) measured every six months, repeat biopsies every one to three years. Bayesian joint model fitted to the observed time of GS7 and longitudinal PSA measurements. Predictions for GS7 externally validated in five largest AS cohorts (GAP3 database). Predictions utilized to develop risk based biopsy schedules, and then compared with fixed schedules on the basis of total biopsies and expected delay in detection of GS7.

`Results and Limitations`: Roughly 50% patients do not obtain GS7 in first 10 years in PRIAS. Rate of change of (log-transformed) PSA was a stronger predictor of GS7 (Hazard Ratio: 2.45, 95%CI: 1.83–2.95) than PSA value (Hazard Ratio: 1.00, 95%CI: 0.98–1.02). Internal validation: Time varying area under ROC curve for GS7 prediction ranged between 0.xx and 0.xx, and prediction error between 0.xx and 0.xx. External validation: Results similar to internal validation only for Toronto and Johns Hopkins cohorts.

`Conclusions`: We developed risk based biopsy schedules as alternative to fixed schedules. For both fixed and risk based schedules we provide total biopsies, time of biopsies, and expected time delay in detection of GS7. Risk based schedules update over follow-ups with more patient data.

`Patient Summary`: Risk based biopsy schedules are a novel alternative to fixed biopsy schedules (e.g., yearly biopsies). They utilize a patient's PSA history and biopsy history to decide best time of biopsies in future. Such personalized schedules can offer a better balance between number of biopsies and delay in detection of Gleason upgrade than yearly biopsies.

Keywords: Active Surveillance, Biopsies, Personalized Medicine, Prostate Cancer, Shared Decision Making

1. Introduction

Patients with low- and very low-risk screening-detected localized prostate cancer (PCa) are usually advised active surveillance (AS) instead of immediate radical treatment [1]. In AS, PCa progression is routinely monitored via prostate-specific antigen (PSA), digital rectal examination, and the Gleason score from repeat prostate biopsies. Since Gleason score is the strongest indicator of cancer-related outcomes, patients are commonly advised curative treatment when Gleason ≥ 7 (GS7) is detected [2].

Biopsies are conducted periodically, and hence GS7 is always detected with a time delay. The smaller the delay, the larger is the window of opportunity for curative treatment. For timely detection of GS7, many AS programs use fixed biopsy schedules (e.g., annual or biennial biopsies) for all patients [3, 4]. In Figure 1 we can see that in the PRIAS and JHAS cohorts, 50% patients do not need such frequent biopsies in first ten years of follow-up. Biopsies are also invasive, painful and prone to medical complications. Biopsy burden and patient non-compliance [5] to frequent biopsies, has raised questions about the optimal biopsy schedule [6, 7].

A simple alternative to frequent biopsies is infrequent biopsies. However, studies suggest not reducing biopsy frequency beyond 24 months, to have sufficient window of opportunity for curative treatment [6, 9]. Although, biopsy every 24 months, still leads to five unnecessary biopsies over ten years for 50% patients (Figure 1). A promising alternative to fixed biopsies is biopsies based on personalized risk of GS7. For example, among the two patients shown in Figure 2, patient B is a more suitable candidate for biopsy than patient A, because his risk of GS7 is much higher. Simulation studies

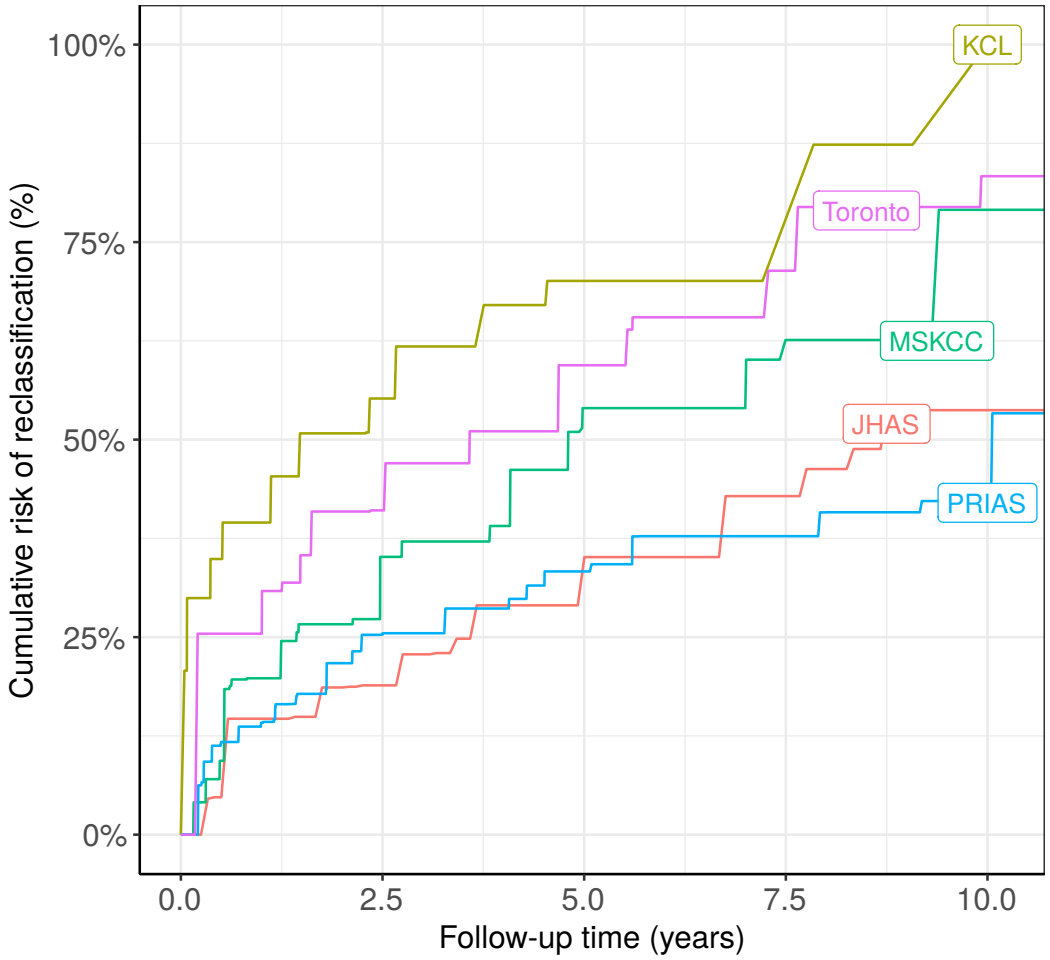

Figure 1: **Estimated cumulative risk of having Gleason $\geq$ 7 (GS7)** in five of the largest AS studies part of the GAP3 database [8], *JHAS*: Johns Hopkins Active Surveillance, *PRIAS*: Prostate Cancer International Active Surveillance, *Toronto*: University of Toronto Active Surveillance, *MSKCC*: Memorial Sloan Kettering Cancer Center Active Surveillance, *KCL*: King's College London Active Surveillance. In the world's largest AS cohort PRIAS and in JHAS, roughly 50% of patients do not obtain GS7 in the first ten years. That is, ideally no biopsies should be prescribed to 50% of the patients in first ten years of AS.

have shown that personalized schedules may better balance the number of biopsies per detected GS7 than fixed schedules [10].

The first challenge in developing risk-based schedules is consolidating observed patient data (e.g., PSA, previous biopsy results) into GS7 risk estimates. For this previous studies have employed joint models for time-to-event and longitudinal data [10, 11, 12]. However, translating risk estimates into clinical decisions is challenging. For example, a 10% risk can be perceived as high/low depending upon the patient's age. Patients may also weigh the risk of GS7 with the potential consequences of another biopsy. Two important consequences are the timing and total number of biopsies (burden), and the time delay in the detection of GS7 (smaller is better). These consequences vary between the patients, and also over the follow-up period for the same patient.

The goal of this work was to assist patients and doctors in making better decisions of biopsies than fixed and frequent biopsies. We intended to achieve this by providing the patients risk based personalized schedules of biopsies, and to allow them to compare the consequences of each schedule before making a decision. To this end, we took three steps. First we fitted a prediction (joint) model to the world's largest AS dataset, PRIAS [2]. We then externally validated the model predictions in five largest AS cohorts that are part of the GAP3 database. Lastly, we utilized the personalized GS7 risk predictions to calculate the timing and total number of biopsies, and the time delay in the detection of GS7 for risk-based and fixed biopsy schedules.

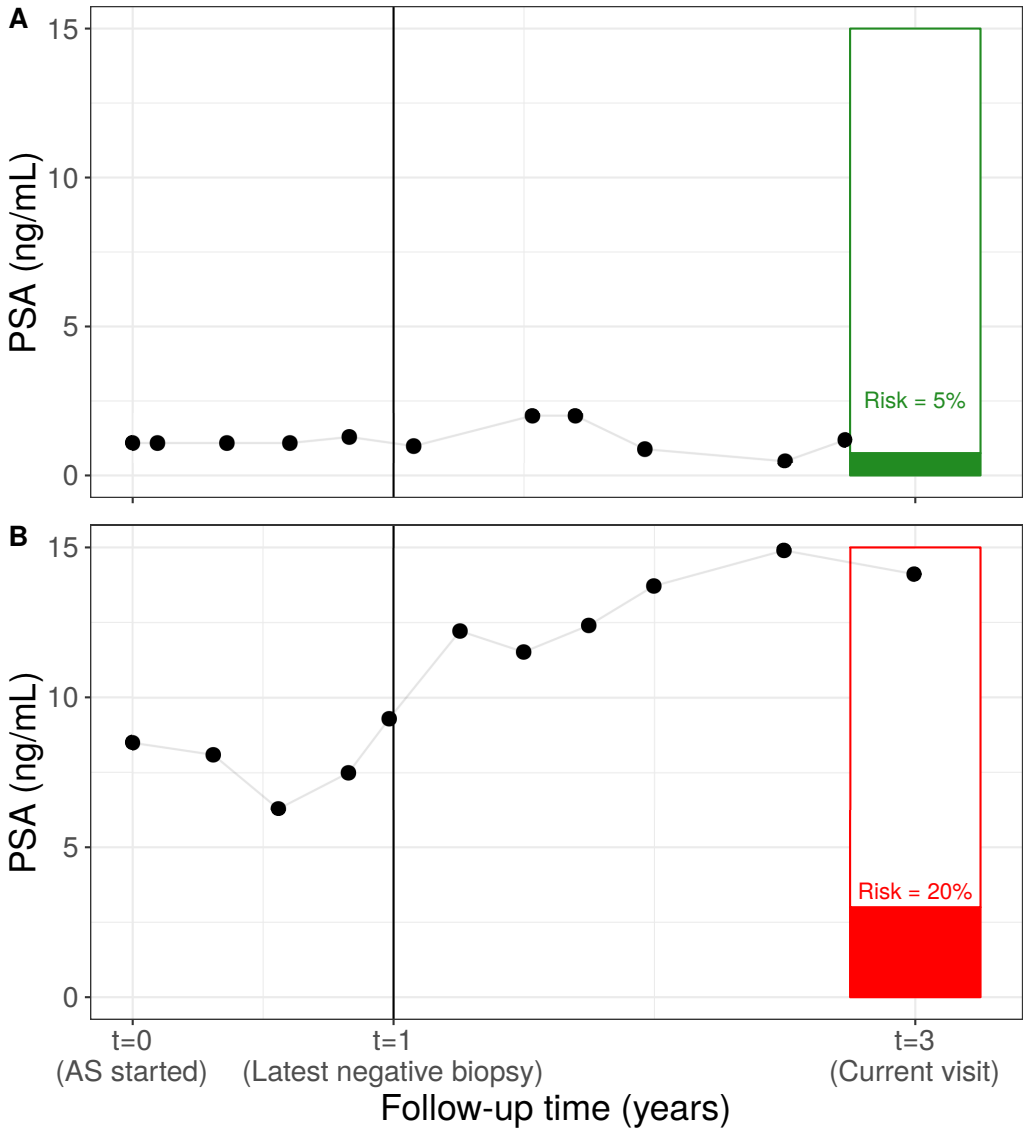

Figure 2: **Motivation for personalized risk-based decisions of biopsy**: Patient A and B had their latest biopsy at year one of follow-up. Patient A's prostate-specific antigen profile remained stable until the current visit time at year three, whereas patient B's profile has shown a rise. Consequently, patient B's estimated cumulative risk of GS7 at year three is higher than that of patient A, making patient B a more suitable candidate for biopsy than Patient A.

2. Patients and Methods

2.1. Study Cohort

Prostate Cancer International Active Surveillance (PRIAS) is an ongoing prospective cohort study of men with low- and very-low risk PCa diagnoses [2]. More than 100 medical centers from 17 countries contribute in PRIAS, using a common study protocol (`www.prias-project.org`). We used the data collected between December 2006 (beginning of PRIAS study) and May 2019. The PSA was measured every three months until year two of follow-up and every six months thereafter. Biopsy schedule was year one, four, seven, and ten, and additional yearly biopsies when PSA doubling time is between three and ten years. The primary event of this work is Gleason $\geq$ 7 (GS7). It was observed in 1134 patients, but 2250 were provided treatment (see Table 1). Treatment in absence of GS7 may have been advised on the basis of PSA, number of biopsy cores with cancer, anxiety, or other reasons. We focused only on GS7 because of its strong association with cancer-related outcomes. Due to the periodical nature of biopsies, the time of GS7 was only known as a time interval in which it occurred.

Table 1: **Patient characteristics for the PRIAS dataset**. The primary event of interest is Gleason ≥ 7. IQR: interquartile range, PSA: prostate-specific antigen.

Characteristic	Value
Total patients	7813
Gleason ≥ 7 (primary event)	1134
Treatment	2250
Watchful waiting	334
Loss to follow-up	250
Death (unrelated to prostate cancer)	95
Death (related to prostate cancer)	2
Median age at diagnosis (years)	61 (IQR: 66–71)
Median follow-up period per patient (years)	1.8 (IQR: 0.9–3.99)
Total PSA measurements	65798
Median number of PSA measurements per patient	6 (IQR: 4–12)
Median PSA value (ng/mL)	5.7 (IQR: 4.1–7.7)
Total biopsies	15563
Median number of biopsies per patient	2 (IQR: 1–3)

2.2. Statistical Methods

Our aim was to develop a model for predicting the time of GS7. The available data for each patient were, age at the start of AS, all observed PSA measurements, and the history of biopsies. The PSA measurements of a patient were measured longitudinally and were likely correlated. They could also be higher when measured closer to the time of GS7. An additional complication was that such higher values were often missing once a patient obtained GS7. The vice versa, that is, GS7 could be indicated by rise in PSA was also plausible. Such complex correlations between a longitudinal PSA outcome and a time of GS7 outcome are commonly modeled via a joint model for time-to-event and longitudinal data [12, 10, 11].

Our joint model exploited patient-specific random effects [13] to act as a common source of correlation between the sub-models for the PSA and time of GS7 outcomes (see Figure 3). Random effects manifested the underlying state of PCa, and were included in both the linear mixed effects sub-model for $\log_2\{\text{PSA} + 1\}$ transformed measurements, and the relative risk sub-model (similar to cox model) for time of GS7. In the PSA sub-model, random effects non-linearly modeled the evolution of PSA over time. Simultaneously, in the relative risk model they were included indirectly by using fitted $\log_2\{\text{PSA} + 1\}$ value and velocity as time dependent covariates. This established the correlation between PSA and time of GS7. Unlike observed $\log_2\{\text{PSA} + 1\}$ values, the fitted values were free of measurement errors. The $\log_2\{\text{PSA} + 1\}$ velocity was mathematically derived from fitted $\log_2\{\text{PSA} + 1\}$ values. The $\log_2\{\text{PSA} + 1\}$ velocity was also allowed to change non-linearly over follow-up.

The parameters of the two sub-models were estimated jointly using the R **JMbayes** [14]. This package utilizes the Bayesian methodology to estimate model parameters. The parameters and 95% credible intervals are presented in Table.. of Appendix.

2.3. Assessment of Predictions

We assessed the goodness of fit of our model using both in-sample and out-of-sample predictions of GS7. For out-of-sample predictions we utilized the five largest AS cohorts that constitute the GAP3 database [8]. We measured the accuracy of these predictions via the root mean squared prediction error or RMSPE [15] and the area under the receiver operating characteristic curve or AUC [15]. Both of these measures take a value between zero and one. The RMSPE is a measure of calibration representing the difference between the true GS7 status of a patient, and the predicted risk of GS7. Ideally the RMSPE should be zero. The AUC indicates if the model is able to discriminate between patients who obtain GS7 and those do not obtain it. Ideally it should be equal to one. In practice it should not be less than 0.5 (AUC of random discrimination). Since PRIAS is a longitudinal study, we compute these measures in a time dependent manner, at a gap of every one year until xx years of follow-up (95% quantile of observed GS7 times).

2.4. Estimate Risk of GS7 and Consequences of Biopsies

Consider a new patient P shown in Figure .... Using the joint model fitted to the PRIAS dataset, we first obtained his profile of the cumulative risk of GS7 over the follow-up period. We then suggest a biopsy at a follow-up visit if the cumulative risk at that visit is above a certain threshold (e.g., 10%

risk). The cumulative risk is updated at each new visit, by accounting for latest PSA measurements and decisions of biopsies. One can then repeatedly apply the threshold based decision rule for biopsies at each new visit.

The choice of a threshold is not easy. To this end, we exploit the entire cumulative risk profile of a patient to estimate the consequences of following a particular threshold based schedule (Figure ...). The consequences we use in this paper are the expected delay in detection of GS7, the corresponding number of biopsies required, at the estimated visit times at which they are scheduled. These estimates are patient specific and also updated with new data at each visit. Since we calculate the consequences for various fixed biopsy schedules as well, patients can make a more informed decision of biopsy. Lastly, we implemented this approach in a web-based application for use in medical centers.

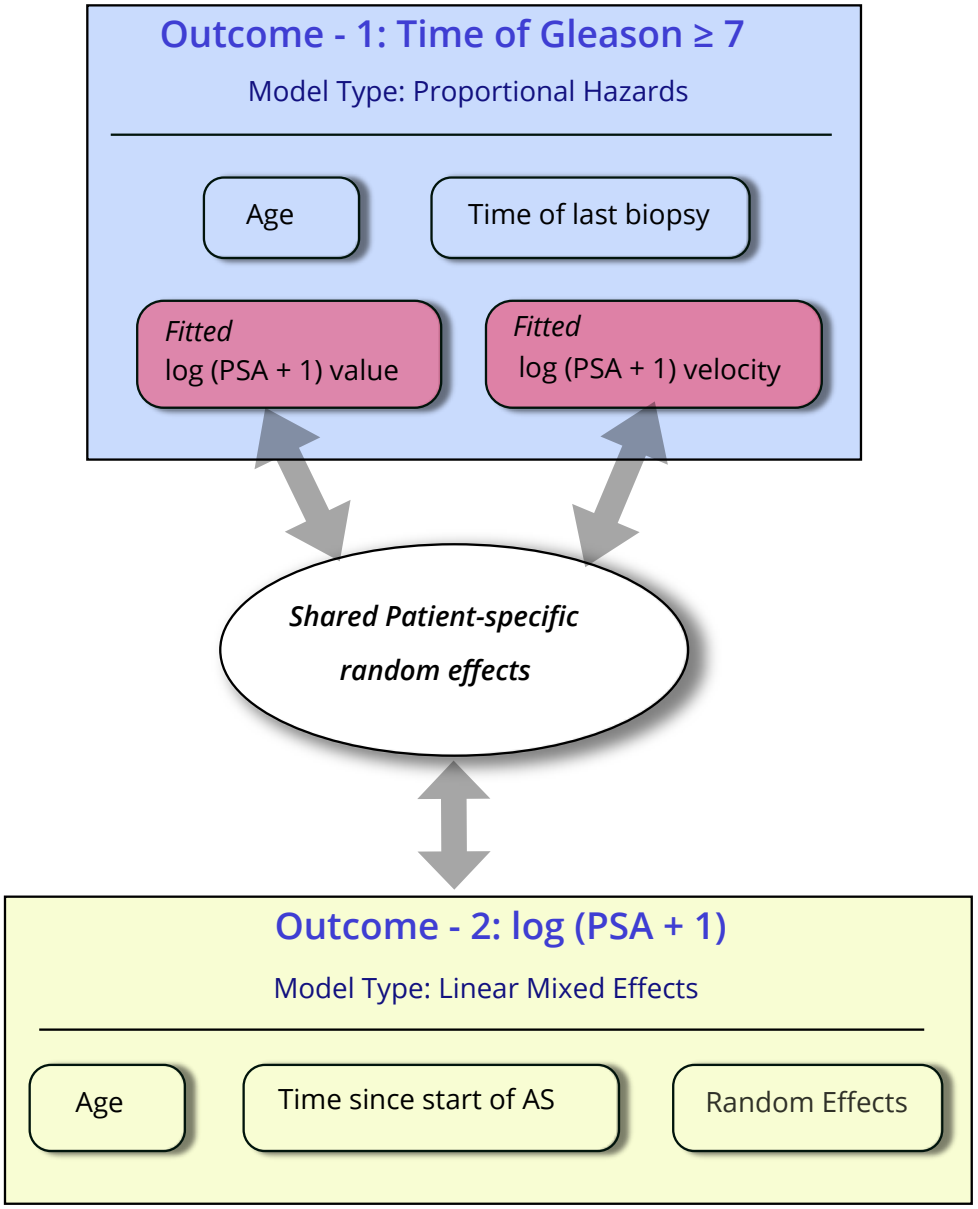

Figure 3: **Diagram of the joint model**: Patient-specific random effects (ellipse in center) are shared between sub-models for all outcomes pertaining to prostate cancer progression, to model the correlation between them. In the linear mixed effects sub-model for $\log_2\{\text{PSA} + 1\}$ transformed PSA outcome (bottom rectangle), random effects are used as covariates. In the relative-risk sub-model (similar to Cox model) for time of Gleason ≥ 7 (GS7), random effects are utilized indirectly, by including fitted $\log_2\{\text{PSA} + 1\}$ value and velocities as covariates. Age of patient at baseline is included in both sub-models. Parameters of both sub-models are estimated jointly.

3. Results

For patients in the PRIAS dataset, the probability of obtaining reclassification within the first five and ten years is 33% and 42%, respectively (see Figure 1). That is, more than 50% of the patients may not require any biopsy in the first ten years. We refer to them as *slow progressing* patients hereafter. For every ten years increase in a patient age the corresponding adjusted hazard ratio of reclassification is 1.45 (95%CI: 1.29–1.61). For an increase in fitted $\log_2\{\text{PSA} + 1\}$ value from the first quartile of fitted value (2.67) to the third quartile (2.82), the corresponding adjusted hazard ratio of reclassification is 1.00 (95%CI: 0.98–1.02). On the other hand an increase in fitted $\log_2\{\text{PSA} + 1\}$ velocity from the first quartile of fitted velocity (-0.04) to the third quartile (0.15), the corresponding adjusted hazard ratio of reclassification is 2.45 (95%CI: 1.83–2.95). These results indicate that the velocity of $\log_2\{\text{PSA} + 1\}$ measurements is a stronger predictor of hazard of reclassification than the $\log_2\{\text{PSA} + 1\}$ value.

The time dependent area under the receiver operating characteristic curves (AUC) and the root mean squared prediction error (RMSPE) with 95% CI are shown in Figure. The results are comparable for internal and external validation in cohorts that are similar to the PRIAS cohort. That is, the model may also be useful for risk prediction in other cohorts such as the Toronto AS cohort.

Using the fitted model we next predict the

4. Discussion

We developed a novel methodology for personalized biopsies in low-risk PCa patients enrolled in AS programs. These biopsies are based on a patient's risk profile for having a Gleason ≥ 7 (GS7). To assist patients in making a choice between the personalized and currently practiced fixed schedules, we give objective estimates of the consequences of following each schedule. More specifically, for a schedule we give the total number of biopsies (burden), the time at which they will be conducted, and the expected delay in detection of GS7. This delay is estimated after accounting for the probability of not having any GS7 at all over the follow-up period. Lastly, our approach dynamically updates the aforementioned schedules and consequences as more patient data becomes available over follow-up.

The aforementioned methodology is based on the world's largest PCa AS program, PRIAS. Consequently, a lot of patients may get benefited from this study. To this end, we have developed a web-application implementing our methodology. The web-application only requires patient data in well known file formats (e.g., SPSS, CSV etc.), but does not require any separate integration with the electronic health record of the PRIAS program. We hope that this will lead to improvement in the shared decision making of biopsies, with patients having objective estimates of the consequences of their decisions.

Clinical implications: The median survival time for GS7 is more than ten years in PRIAS. That is, more than 50% patients do not require any biopsy during the first ten years of follow-up. The situation is similar in many other cohorts. Hence frequent biopsies may not be recommended for

all patients.

Existing work on reducing the burden of biopsies in AS primarily advocates less frequent heuristic schedules of biopsies [6] (e.g., biopsies biennially instead of annually). To our knowledge, risk-based biopsy schedules have barely been explored yet in AS [3**?**]. The part of our results pertaining to the fixed/heuristic schedules is comparable with corresponding results obtained in existing work [6], even though the AS cohorts are not the same. Thus, we anticipate similar validity for the results pertaining to the personalized schedules.

Our work has certain limitations. The prediction model that we developed is valid only for the first thirteen years of follow-up in AS, whereas PCa in AS patients progresses slowly. This issue can be mitigated by refitting the model as more follow-up data is gathered in PRIAS. The results of external validation indicate that the use of our model may be restricted in cohorts with AUC, and RMSPE results similar to that of PRIAS. To this end, in other cohorts, refitting the model to their dataset will be required before making risk based schedules, and estimating the consequences of each schedule. There is also a potential for including diagnostic information from novel biomarkers, quality of life measures, and magnetic resonance imaging. Currently, this data is very sparsely available in the PRIAS dataset. However, in future, adding this information in our model is trivial. This is because modeling correlation for extra outcomes (see Figure 3), mainly entails sharing the random effects in the joint model structure. Since MRI scans are expensive in developing countries, our model can also be used to trigger MRI scans. Lastly, in this study focus only on biopsy Gleason upgrade (reclassi-

fication). In this regard, accounting for competing risks (see Table 1), and for inter-observer variation [11] in biopsy Gleason scores can be interesting to investigate further.

5. Conclusions

References

1. Briganti A, Fossati N, Catto JW, Cornford P, Montorsi F, Mottet N, Wirth M, Van Poppel H. Active surveillance for low-risk prostate cancer: the European Association of Urology position in 2018. *European urology* 2018;74(3):357–68.

2. Bul M, Zhu X, Valdagni R, Pickles T, Kakehi Y, Rannikko A, Bjartell A, Van Der Schoot DK, Cornel EB, Conti GN, et al. Active surveillance for low-risk prostate cancer worldwide: the prias study. *European urology* 2013;63(4):597–603.

3. Nieboer D, Tomer A, Rizopoulos D, Roobol MJ, Steyerberg EW. Active surveillance: a review of risk-based, dynamic monitoring. *Translational andrology and urology* 2018;7(1):106–15.

4. Loeb S, Carter HB, Schwartz M, Fagerlin A, Braithwaite RS, Lepor H. Heterogeneity in active surveillance protocols worldwide. *Reviews in urology* 2014;16(4):202–3.

5. Bokhorst LP, Alberts AR, Rannikko A, Valdagni R, Pickles T, Kakehi Y, Bangma CH, Roobol MJ, PRIAS study group . Compliance rates with the Prostate Cancer Research International Active Surveillance (PRIAS) protocol and disease reclassification in noncompliers. *European Urology* 2015;68(5):814–21.

6. Inoue LY, Lin DW, Newcomb LF, Leonardson AS, Ankerst D, Gulati R, Carter HB, Trock BJ, Carroll PR, Cooperberg MR, et al. Comparative

analysis of biopsy upgrading in four prostate cancer active surveillance cohorts. *Annals of internal medicine* 2018;168(1):1–9.

7. Bratt O, Carlsson S, Holmberg E, Holmberg L, Johansson E, Josefsson A, Nilsson A, Nyberg M, Robinsson D, Sandberg J, et al. The study of active monitoring in sweden (sams): a randomized study comparing two different follow-up schedules for active surveillance of low-risk prostate cancer. *Scandinavian journal of urology* 2013;47(5):347–55.

8. Bruinsma SM, Zhang L, Roobol MJ, Bangma CH, Steyerberg EW, Nieboer D, Van Hemelrijck M, consortium MFGAPPCASG, Trock B, Ehdaie B, et al. The movember foundation's gap3 cohort: a profile of the largest global prostate cancer active surveillance database to date. *BJU international* 2018;121(5):737–44.

9. de Carvalho TM, Heijnsdijk EA, de Koning HJ. Estimating the risks and benefits of active surveillance protocols for prostate cancer: a microsimulation study. *BJU international* 2017;119(4):560–6.

10. Tomer A, Nieboer D, Roobol MJ, Steyerberg EW, Rizopoulos D. Personalized schedules for surveillance of low-risk prostate cancer patients. *Biometrics* 2019;75(1):153–62.

11. Coley RY, Zeger SL, Mamawala M, Pienta KJ, Carter HB. Prediction of the pathologic gleason score to inform a personalized management program for prostate cancer. *European urology* 2017;72(1):135–41.

12. Rizopoulos D. Joint Models for Longitudinal and Time-to-Event Data: With Applications in R. CRC Press; 2012. ISBN 9781439872864.

250 13. Laird NM, Ware JH, et al. Random-effects models for longitudinal data.
251 *Biometrics* 1982;38(4):963–74.

252 14. Rizopoulos D. The R package JMbayes for fitting joint models for lon-
253 gitudinal and time-to-event data using MCMC. *Journal of Statistical*
254 *Software* 2016;72(7):1–46.

255 15. Rizopoulos D, Molenberghs G, Lesaffre EM. Dynamic predictions with
256 time-dependent covariates in survival analysis using joint modeling and
257 landmarking. *Biometrical Journal* 2017;59(6):1261–76.