Personalized Risk Based Shared Decision Making Framework for 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 is 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, com-

[☆]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)

mon study protocol as well as web-based tool for data collection.

`Outcome Measurements, and Statistical Analysis`: Prostate-specific antigen (PSA) measured every six months, repeat biopsies scheduled every one to three years. A Bayesian joint model was fitted to the observed time of GS7 and longitudinal PSA measurements. Predictions for GS7 externally validated in five largest AS cohorts (GAP3 database). Predictions used to develop risk based biopsy schedules. Comparison made with fixed schedules on the basis of total biopsies and expected delay in detection of GS7.

`Results and Limitations`: Median time of GS7 in PRIAS $>$ 10 years. Rate of change of (log-transformed) PSA found to be 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 AS cohorts.

`Conclusions`: Patients get objective estimates of their risk of GS7 over follow-up, and can choose among various risk based schedules and/or fixed schedules on the basis of

`Patient Summary`: Biopsies can be scheduled on the basis of patient-specific risk of Gleason upgrade. Such schedules can lead to fewer biopsies than fixed and frequent biopsies (e.g., annual biopsies) for all patients. In

many *slow progressing* patients, the window of opportunity (delay in detection of Gleason upgrade) for curative treatment is similar for risk based and fixed biopsy schedules.

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. Among these, the Gleason score is the strongest indicator of cancer related outcomes. Thus, patients are commonly advised curative treatment upon detecting a Gleason score ≥ 7 (GS7) [2].

Since biopsies are scheduled intermittently, reclassification is always detected with a delay. The smaller this delay is, the larger is the window of opportunity for curative treatment. To this end, majority of the AS programs worldwide, schedule biopsies every 12-24 months for all patients [3, 4]. Such fixed and frequent biopsies may benefit a small proportion of men with a high risk of reclassification. However, for many of the *slow progressing* patients (see Figure 1) frequent biopsies are redundant . Biopsies are also invasive, painful and prone to medical complications. The unnecessary burden of biopsies, and patient non-compliance [5] to frequent biopsies, has raised questions about the optimal time interval between subsequent biopsies [6, 7].

The simplest solution to frequent biopsies is reducing the frequency of biopsies for all patients. However, simulation studies have suggested that reducing the frequency beyond 24 months may not leave sufficient window of opportunity for curative treatment [6]. Although, even with a gap of 24 months, up to five unnecessary biopsies over ten years of follow-up may still be scheduled for *slow progressing* patients. A promising alternative to such fixed decision of biopsies are the risk based decisions of biopsies. Consider for

instance the two patients shown in Figure 2. Both patients had their latest biopsy at year one of follow-up and are now scheduled for a biopsy after a 24 month gap at year three. The PSA profile of patient A is stable and the PSA profile of patient B is rising. The cumulative risk of reclassification of patient B at year three is also higher than patient A's. Consequently, at year three he is a more suitable candidate for biopsy than patient A.

The first challenge in such a risk based approach is the consolidation of observed patient data (e.g., PSA, previous biopsy results) into estimates of the risk of reclassification (Figure 2). To this end, previous studies have employed joint models for time-to-event and longitudinal data [9, 10, 11]. A subsequent challenge however, is to translate these risk estimates into clinical decisions. For example, a 10% risk can be perceived as high/low depending upon the patient's age. Patients may also weigh the risk of reclassification with the potential consequences of another biopsy. Two such consequences are the delay in detection of reclassification (smaller is beneficial), and the total burden of biopsies. These consequences vary between the patients, and also over the follow-up period for the same patient.

The goal of this work is to assist patients and doctors in making better decisions of biopsies than fixed and frequent biopsies. We intend to achieve this by providing patient- and visit specific risks of reclassification. To further facilitate shared decision making, we also provide estimates of the delay in detection of reclassification and the total burden of biopsies. To this end, we fit a prediction (joint) model to the world's largest AS dataset, PRIAS [2]. We then validate the predictions in multiple external cohorts that are part of the GAP3 database. Lastly, we implement risk based schedules as a

web-application, and demonstrate them with real patient data.

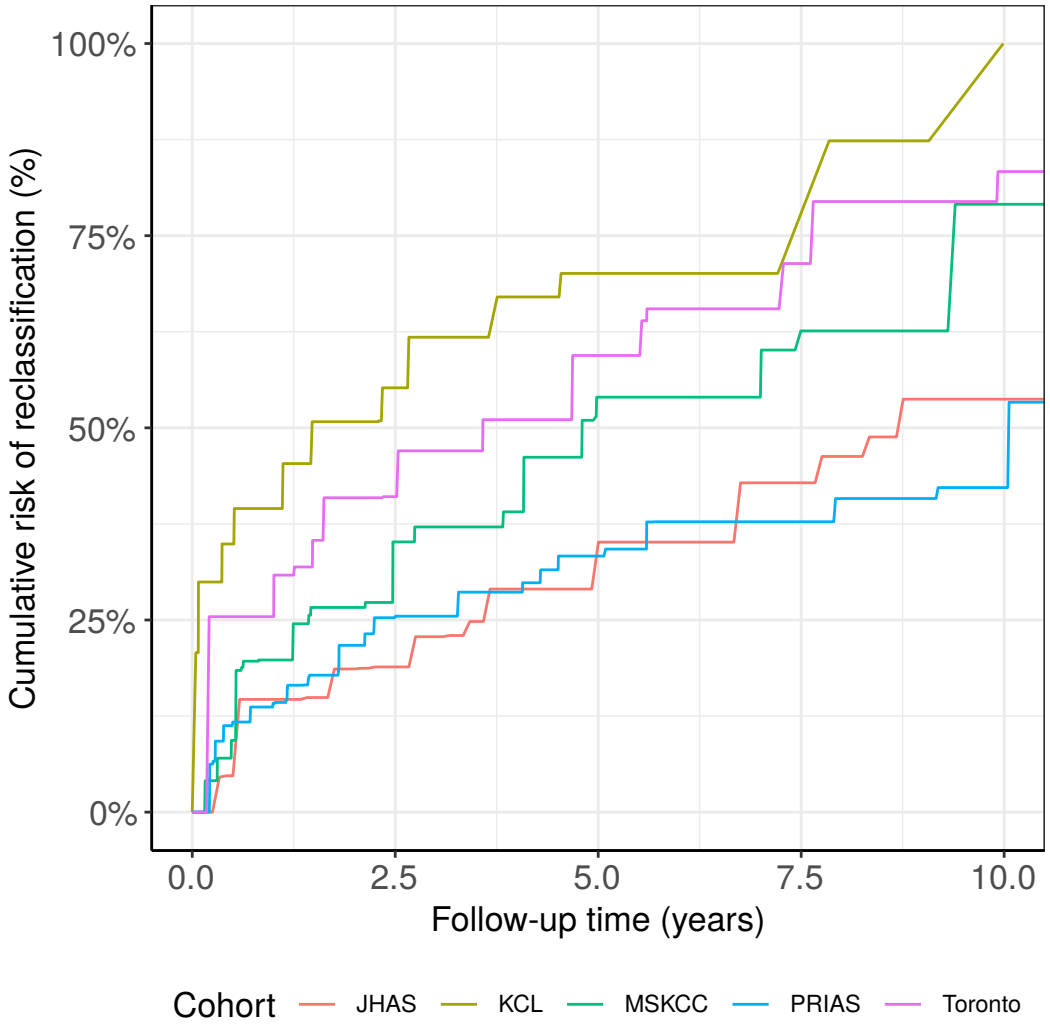

Figure 1: **Active surveillance cancer patients are often *slow progressing***. Graph shows estimated cumulative risk of having a Gleason score > 6 in five of the largest AS studies that are part of the GAP3 database [8]. In all cohorts except KCL, roughly 50% patients may not require any biopsy in first five years. In the world's largest AS cohort PRIAS and in JHAS, roughly 50% patients may not require any biopsy in the first ten years. **Legend**: *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

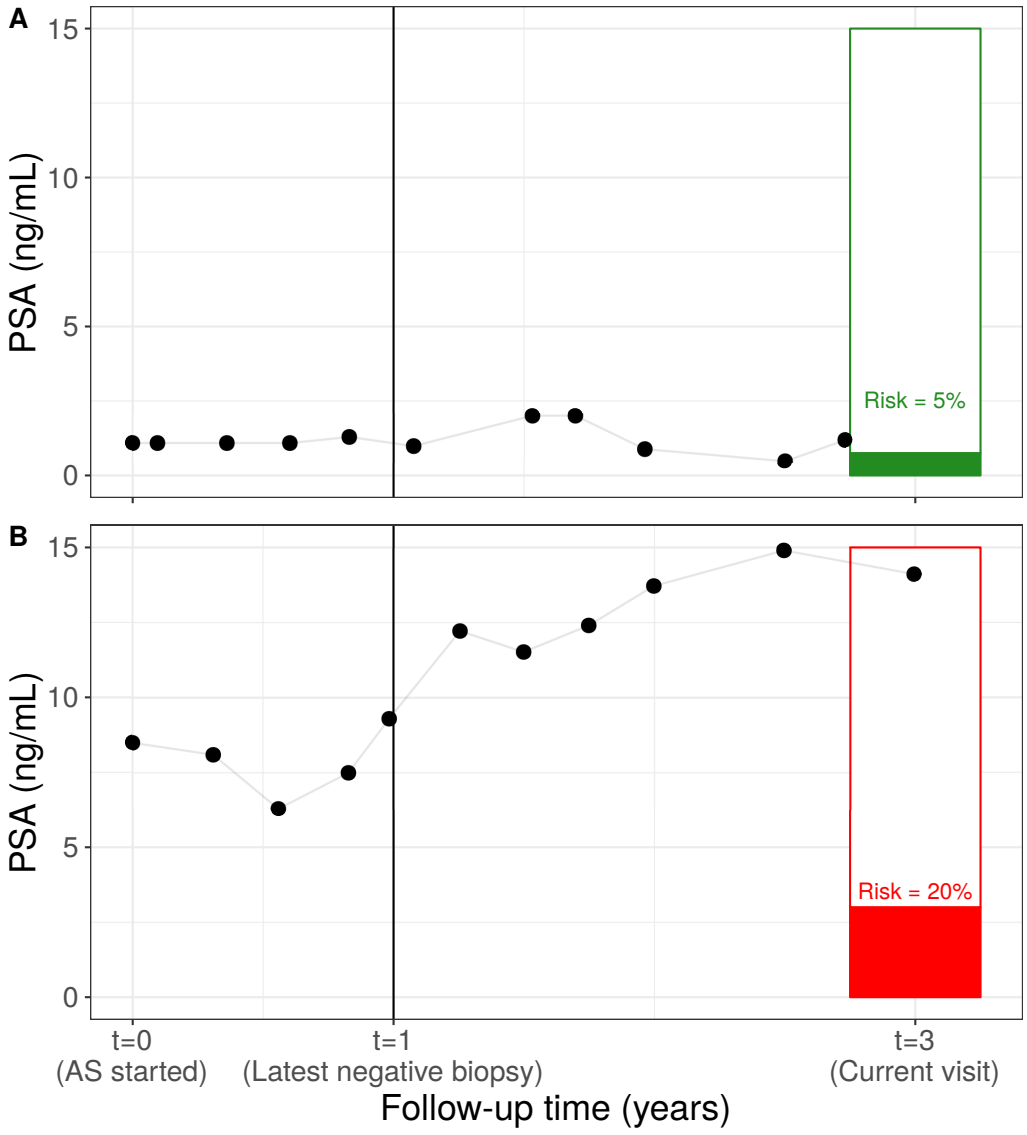

Figure 2: **Motivation for risk based decisions of biopsy**: Patient A and B had their latest biopsy at year one of follow-up. Patient A's prostate-specific antigen (PSA) profile remained stable until the current visit time at year three. Consequently, his cumulative risk of reclassification at year three is 5%. On the other hand patient B's PSA profile has shown a rise since the latest biopsy, and his cumulative risk of reclassification is also 20%. Patient B is a better 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 worldwide contribute in PRIAS `www.prias-project.org`. We use the data collected between December 2006 (beginning of PRIAS study) and May 2019. The follow-up protocol scheduled PSA measurements (ng/mL) every three months for the first two years and every six months thereafter. Repeat biopsies were scheduled after one, four, seven, and ten years. Additional yearly biopsies were scheduled for patients having PSA doubling time between three and ten years. Reclassification (Gleason > 6) was observed in 1134 patients, and 2250 were provided treatment (see Table 1). Treatment in absence of reclassification may have been advised on the basis of PSA, number of biopsy cores with cancer, anxiety, or other undocumented reasons. However, we focus only on Gleason reclassification because of its strong association with cancer related outcomes. Due to the periodical nature of biopsies, the time of reclassification 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 disease reclassification. IQR: interquartile range, PSA: prostate-specific antigen.

Characteristic	Value
Total patients	7813
Disease reclassification (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

The goal of the statistical analysis of the PRIAS data was to develop a model for predicting the time of reclassification. To this end, for each patient we have the information about his age at the start of AS, all observed PSA measurements, and the history of biopsies. Since PRIAS data is longitudinal in nature, the PSA measurements of a patient are correlated. PSA can be higher when measured closer to the time of reclassification. An additional complication is that such higher values are also often missing once a patient obtains reclassification. The vice versa, that is, reclassification is more likely when PSA increases is also plausible. A commonly used statistical method to model such complex correlation between a longitudinal outcome (PSA) and a time-to-event (reclassification) outcome is the joint model for time-to-event and longitudinal data [11, 9, 10].

A joint model exploits patient-specific random effects (similar to random effects of a linear mixed effects model) to act as a common source of correlation between various outcomes (see Figure 3). These random effects manifest the unobservable patient-specific state of PCa. The joint model has separate sub-models for PSA and time of reclassification. However, both models utilize these random effects as covariates in the model. We used a linear mixed effects model for $\log_2\{\text{PSA} + 1\}$ transformed measurements, and a relative risk model (similar to cox model) for time of reclassification. The mixed effects model for PSA uses random effects to non-linearly model the evolution of PSA over time in a patient-specific manner. Simultaneously, in the relative risk model we establish the correlation between time of reclassification and PSA. This is achieved by using fitted $\log_2\{\text{PSA} + 1\}$ value and velocity as

time dependent covariates, that is, random effects are used indirectly. Unlike observed $\log_2\{\text{PSA} + 1\}$ values, the fitted values are free of measurement errors. The $\log_2\{\text{PSA} + 1\}$ velocity is not modeled separately, but is rather mathematically derived as the rate of change of fitted $\log_2\{\text{PSA} + 1\}$ value over time. Since fitted $\log_2\{\text{PSA} + 1\}$ profiles are modeled non-linearly, the corresponding velocity is also allowed to change over follow-up.

The various parameters of the two sub-models estimated jointly using the R **JMbayes** [12]. 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 reclassification. 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 two commonly used measures, namely the root mean squared prediction error (RMSPE) and the area under the receiver operating characteristic curve (AUC). Both of these measures take a value between zero and one. The RMSPE represents the difference between the true reclassification status of a patient, and the predicted risk of reclassification. Ideally the RMSPE should be zero. The AUC indicates if the model is able to discriminate between patients who obtain reclassification 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 reclassification times).

2.4. Estimate Risk of Reclassification and Consequences of Biopsies

Consider a new patient with a certain history of biopsies, and PSA measurements. Using the joint model fitted to the PRIAS dataset, we first obtain his profile of the cumulative risk of reclassification over the follow-up period (Figure ...). 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 reclassification, 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.

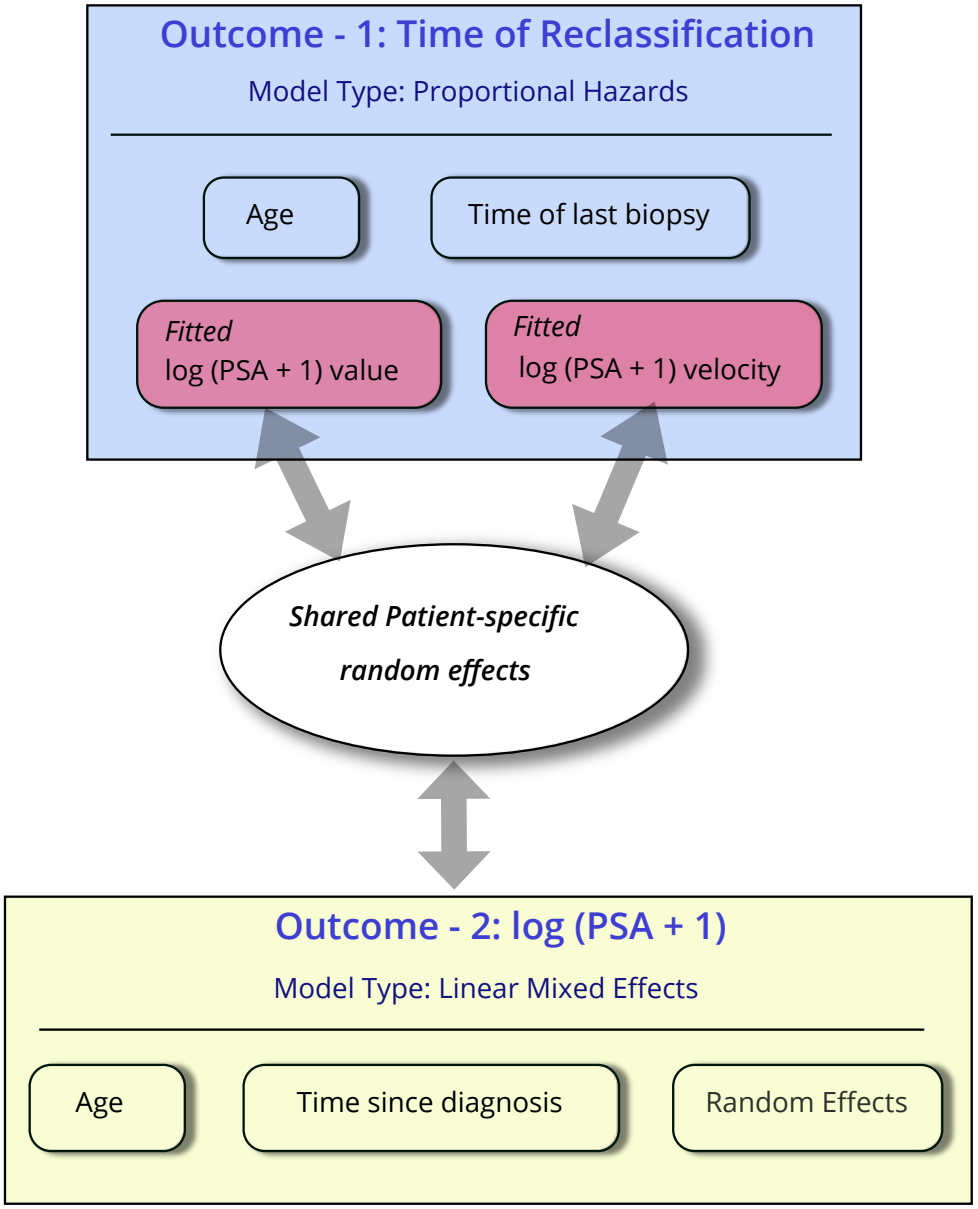

Figure 3: **Diagram of the joint model**: Per patient we observe the $\log_2\{\text{PSA} + 1\}$ transformed PSA, and the results of biopsies. We combine information from these observations to estimate the time of disease reclassification. To this end, we use a linear mixed effects model for $\log_2\{\text{PSA} + 1\}$ measurements, and proportional hazards model for time of disease reclassification. The time of disease reclassification depends on patient age, time of latest negative biopsy and underlying trend of PSA. To account for the correlation between PSA measurements and time of reclassification, the two models share patient-specific random effects in their model equations.

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 decision making of biopsies in low-risk PCa patients enrolled in AS programs. In this methodology patients are first provided risk of reclassification based personalized biopsy schedules, as an alternative to fixed and frequent biopsy schedules (e.g., annual biopsies). Subsequently, to assist patients in making an appropriate choice among the personalized and fixed schedules, we also provide them the consequences of following each schedule. More specifically, for each schedule we give the total number of biopsies (burden), the time at which they will be conducted, and the expected delay in detection of reclassification. This delay is estimated after accounting for the probability of not having any reclassification 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 reclassification 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 (reclassification). In this regard, accounting for competing risks (see Table 1), and for inter-observer variation [10] 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. 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. doi:10.1111/biom.12940.

10. 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.

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

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