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 personalized risk based biopsy schedules as alternative to fixed schedules. For both fixed and personalized schedules we provide total biopsies, time of biopsies, and expected time delay in detection of GS7. Personalized schedules update with more patient data gathered over follow-up.

`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 concerns regarding 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

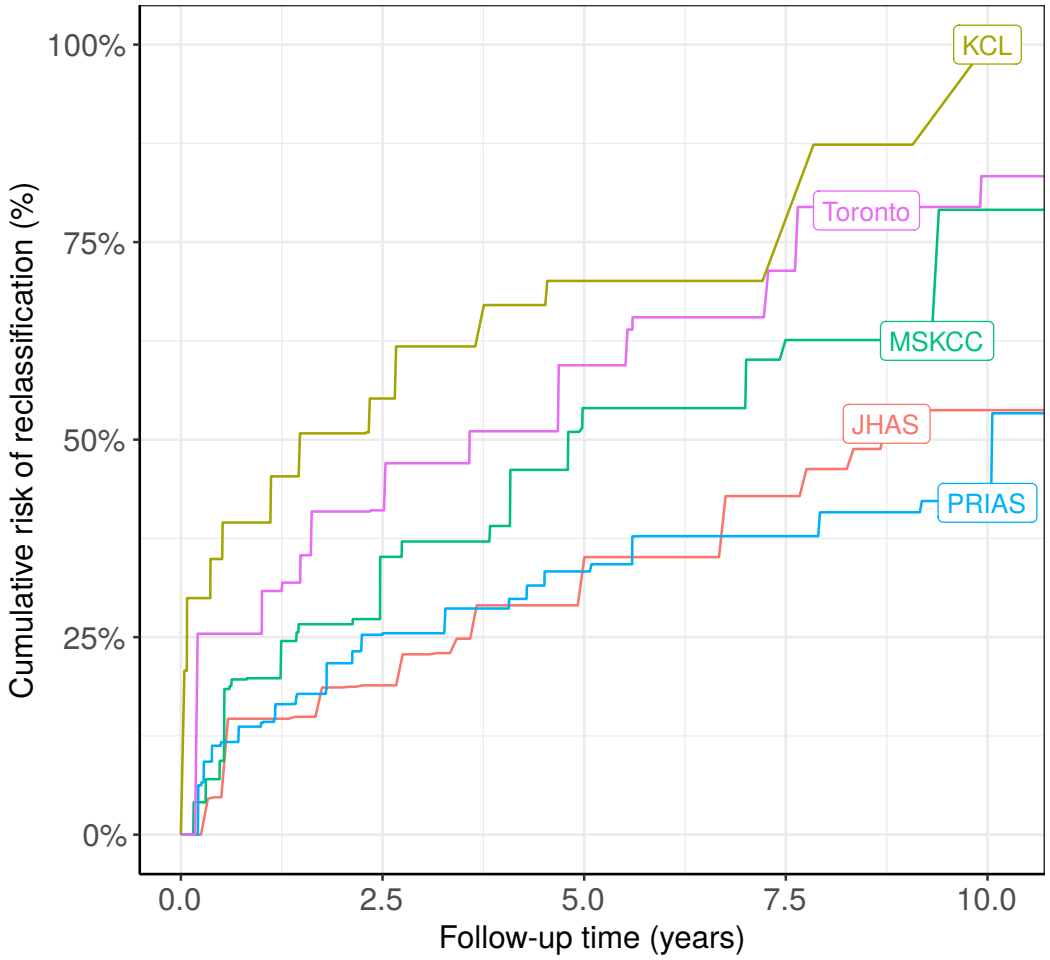

Figure 1: **Estimated cumulative risk of having Gleason $\geq$ 7 (GS7)** in five of the largest AS studies part of the GAP3 database [8], *JHAS*: Johns Hopkins Active Surveillance, *PRIAS*: Prostate Cancer International Active Surveillance, *Toronto*: University of Toronto Active Surveillance, *MSKCC*: 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.

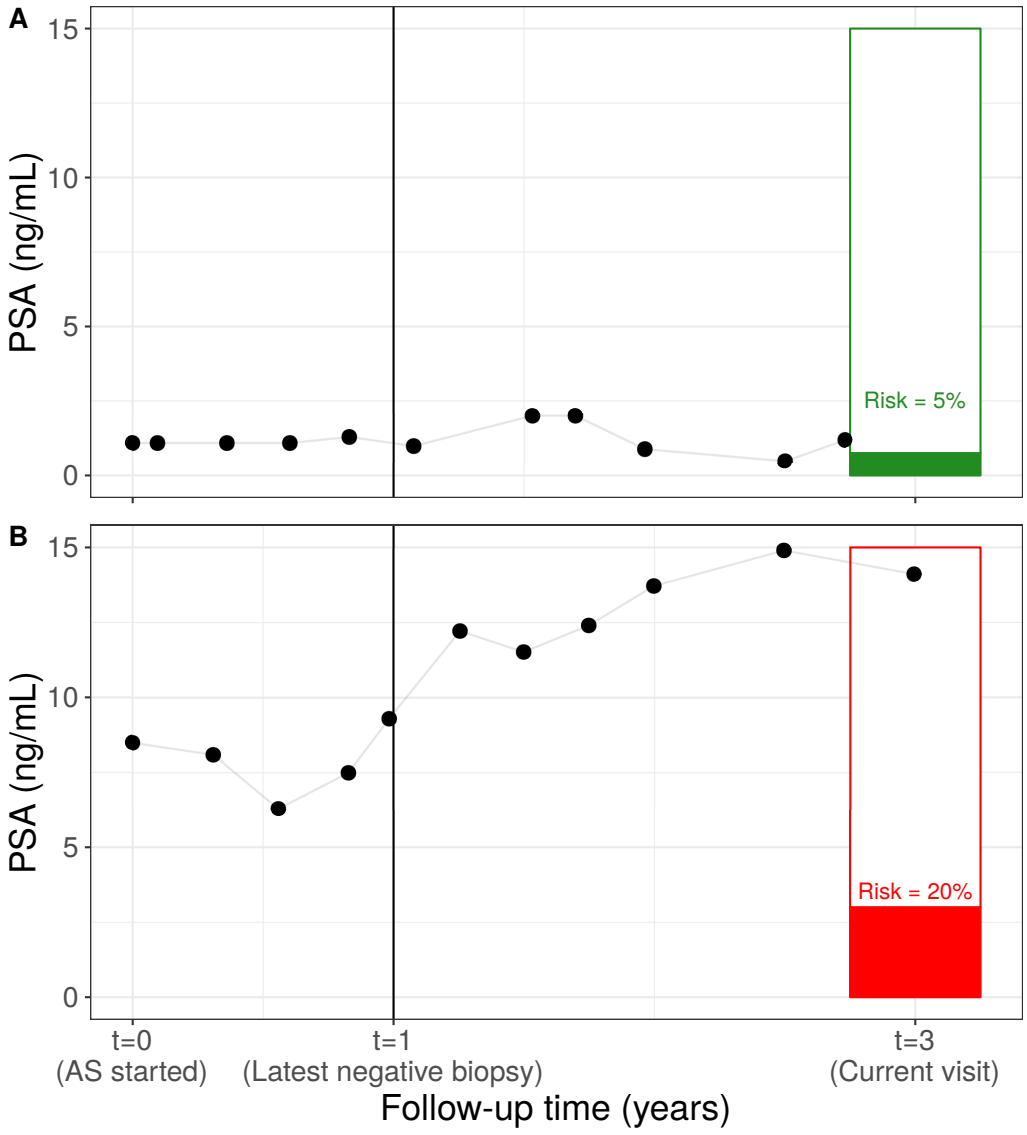

Figure 2: **Motivation for personalized risk-based decisions of biopsy**: Patient A and B had their latest biopsy at year one of follow-up. Patient A's prostate-specific antigen profile remained stable until the current visit time at year three, whereas patient B's profile has shown a rise. Consequently, 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 ≥ 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. We wanted to account for the correlation between the PSA measurements of the same patient, and also their correlation with the time of GS7. An additional complication was that PSA values were missing once a patient obtained GS7. A commonly used model to handle these issues is the joint model for time-to-event and longitudinal data [12, 10, 11].

The joint model we utilized, exploited patient-specific random effects [13] to act as a common source of correlation between the PSA and time of GS7 outcomes (see Figure 3). Random effects also represented 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 used indirectly by including 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. Hence, 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 of GS7

We validated the predictions of GS7 internally within the PRIAS dataset, as well as externally in five of the largest AS cohorts part of the GAP3 database [8]. To this end, we utilized the area under the receiver operating characteristic curve or AUC [15] as a measure of discrimination, and root mean squared prediction error or RMSPE [15] as a measure of calibration. Since AS studies are longitudinal in nature, we computed AUC and RMSPE 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 obtained his cumulative risk of GS7 over the entire follow-up period. A biopsy at his current visit may be suggested if the cumulative risk of GS7 at the current visit is above a certain threshold (e.g., 10% risk). By repeatedly applying the 10% threshold rule over the whole follow-up, we obtained his personalized schedule of biopsies. Similar schedules can be made with another risk threshold. These schedule are not fixed but are rather updated at each follow-up visit based on newly gathered patient data.

To assist patients in making an informed choice for a schedule, be it personalized or fixed, we provided them patient-specific consequences of following each schedule. To this end, we first calculated the probability of

occurrence of GS7 between successive biopsies of each schedule. Using these probabilities we then obtained the expected delay in detection of GS7 for following that schedule. Thus, patients have a method to compare across various schedules in terms of the personalized burden (time and total biopsies), and personalized benefit (less delay in detection of GS7 is beneficial). Lastly, we implemented this approach in a web-application.

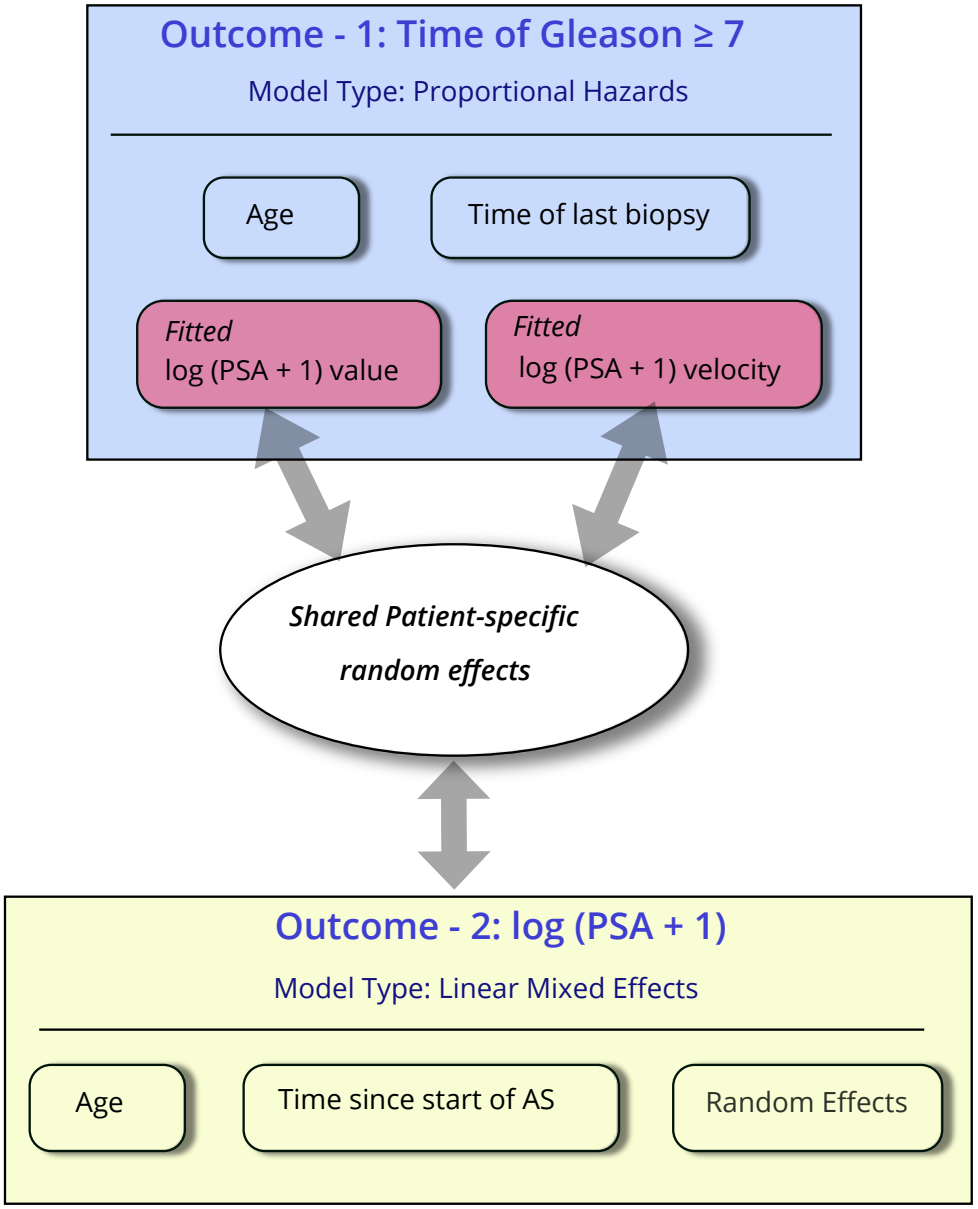

Figure 3: **Diagram of the joint model**: Patient-specific random effects (ellipse in center) are shared between sub-models for all outcomes pertaining to prostate cancer progression. The random effects model the correlation between the outcomes. In the linear mixed effects sub-model for $\log_2\{\text{PSA} + 1\}$ 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.

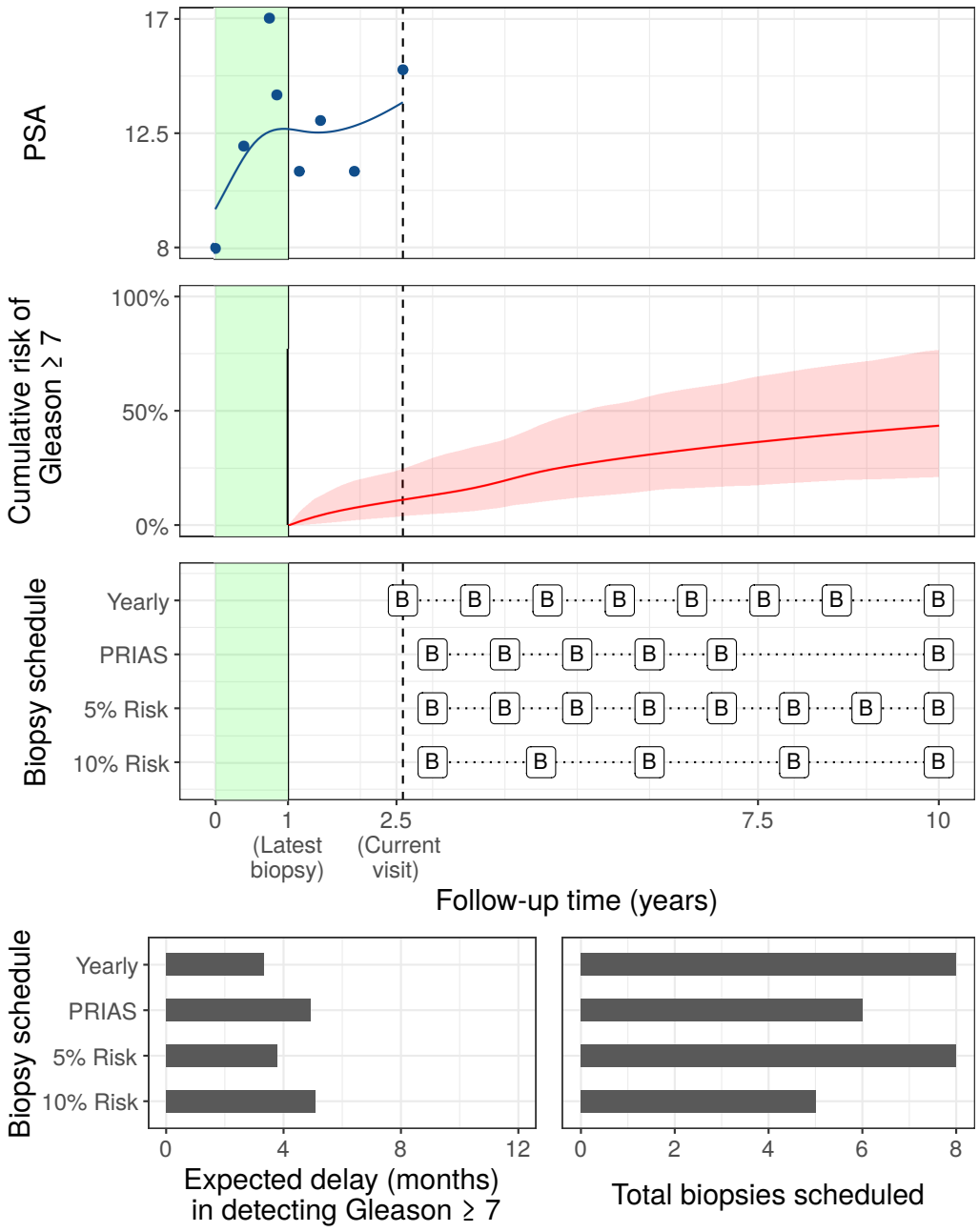
PSA
17
12.5
8
Cumulative risk of Gleason ≥ 7
100%
50%
0%
Biopsy schedule
Yearly
PRIAS
5% Risk
10% Risk
B
0
1
(Latest biopsy)
2.5
(Current visit)
7.5
10
Follow-up time (years)
Biopsy schedule
Yearly
PRIAS
5% Risk
10% Risk
0
4
8
12
Expected delay (months) in detecting Gleason ≥ 7
0
2
4
6
8
Total biopsies scheduled

Figure 4

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

Table 2: **Assessment of predictions of Gleason $\geq$ 7 (GS7) from our model using**, area under the receiver operating characteristic curve (AUC), and root mean squared prediction error (RMSPE).

Characteristic	Value
Total patients	7813
Gleason $\geq$ 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)

4. Discussion

We developed personalized schedules for repeat biopsies in PCa patients enrolled in AS programs. These schedules were based on a patient's risk for having a Gleason ≥ 7 (GS7). Patient- and visit-specific risks of GS7 were estimated using their entire history of PSA and repeat biopsies, and baseline characteristics. Consequently, the personalized schedules were updated as more data was gathered over follow-up. Risk calculators for GS7 are not new [11, 16]. However, the novelty of our work is that we developed a methodology for scheduling personalized biopsies using those risks, as well a methodology to compare schedules, be it personalized or fixed, in simple terms of burden and benefit. More specifically, for each schedule we provided patients the times of biopsy and total biopsies (burden), and the delay in detection of GS7 (less is beneficial) expected due to that schedule. We also implemented our methodology in a web-application.

The proposed joint model accounted for the complex correlation structure that exists between longitudinal PSA measurements and time of GS7 of a patient. It also accounted for PSA measurements that were missing in patients who obtained GS7. This model adjusts the risks of GS7 upon a negative repeat biopsy. Thus complete patient information in consolidated into a single patient risk profile. Our model is fitted to the world's largest PCa AS program, PRIAS. We also validated our model predictions for GS7 in external cohorts that are part of the GAP3 database [8]. We found that the discrimination and calibration measures of the predictions from our model (Table 2) were similar for PRIAS (internal validation), and Toronto and Johns Hopkins cohorts (external validation). Given the large size of these

cohorts, we expect that our model and the methodology will be useful to a large number of AS patients. Extending our model and methodology in other cohorts only requires fitting the model to their AS dataset.

The clinical implications of our work are as follows. The median survival time for GS7 is more than ten years in PRIAS, and in some other cohorts (Figure 1). That is, more than 50% of AS patients do not require any biopsy during the first ten years of follow-up. We hope that our work will address patient apprehensions regarding adverse outcomes in AS, in a more objective manner. Many AS programs still utilize a rigorous schedule of yearly biopsies [3]. However, with concerns about non-compliance and burden of biopsies [5], the availability of our web based tool may encourage patients and doctors to consider personalized schedules.

Our work has certain limitations. The proposed model is valid only for the first thirteen years of follow-up in PRIAS, whereas GS7 may occur much later in many patients. Due to this issue, the calibration and discrimination measures of predictions were also less accurate in later follow-up periods. These issue can be mitigated by refitting the model as more follow-up data is gathered in PRIAS. While we focused only on GS7, it is susceptible to inter-observer variation. Models which account for this variation [11, 17] will be interesting to investigate further. However, the methodology to schedule biopsies, and to estimate the consequences of following a schedule need not change. There is also a potential for including diagnostic information from novel biomarkers, quality of life measures, and magnetic resonance imaging (MRI). Currently, this data is very sparsely available in the PRIAS dataset. However, in future, adding this information in our model is trivial. This is

194 because modeling correlation for extra outcomes, mainly entails connecting
195 sub-models for the outcomes to shared random effects (see Figure 3). Our
196 model can also be used to schedule MRI scans, since they are expensive in
197 developing countries.

5. Conclusions

We developed a novel methodology for scheduling biopsies to detect Gleason ≥ 7 (GS7) in PCa patients enrolled in AS. Our methodology consolidates a patient's entire history of PSA and repeat biopsies, and baseline characteristics into risk profile of GS7 over his follow-up period. It then utilizes this risk profile to schedule biopsies in a personalized manner. The personalized schedule is updated as more patient data is gathered over follow-up. To assist patients in making the choice of the best biopsy schedule, we provided them personalized burden (time and total biopsies), and personalized benefit (less delay in detection of GS7 is beneficial), for both personalized and currently used schedules. Lastly, we implemented this approach in a web-application.

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.

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

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

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

262 16. Ankerst DP, Xia J, Thompson Jr IM, Hoefler J, Newcomb LF, Brooks
263 JD, Carroll PR, Ellis WJ, Gleave ME, Lance RS, et al. Precision
264 medicine in active surveillance for prostate cancer: development of the
265 canary–early detection research network active surveillance biopsy risk
266 calculator. *European urology* 2015;68(6):1083–8.

267 17. Balasubramanian R, Lagakos SW. Estimation of a failure time distribu-
268 tion based on imperfect diagnostic tests. *Biometrika* 2003;90(1):171–82.