Personalized Biopsy Schedules Based on Risk of Gleason Upgrading for Low-Risk Prostate Cancer Active Surveillance Patients[⋆]

Anirudh Tomer, MSc[a,*], Daan Nieboer, MSc[b,c], Monique J. Roobol, PhD[c], Anders Bjartell, MD, PhD[d], Ewout W. Steyerberg, PhD[b,e], Dimitris Rizopoulos, PhD[a], Movember Foundation's Global Action Plan Prostate Cancer Active Surveillance (GAP3) consortium[f]

[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 Urology, Skåne University Hospital, Malmö, Sweden*
[e] *Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands*
[f] *The Movember Foundation's Global Action Plan Prostate Cancer Active Surveillance (GAP3) consortium members presented in Appendix A*

Abstract

`Objective`: To develop a model and methodology for predicting the risk of Gleason *upgrading* in prostate cancer active surveillance (AS) patients, and using the predicted risks to create risk-based *personalized* biopsy schedules as an alternative to one-size-fits-all schedules (e.g., annually). Furthermore,

[⋆] Word count: 3306

[*] 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.nieboer@erasmusmc.nl (Daan Nieboer, MSc), m.roobol@erasmusmc.nl (Monique J. Roobol, PhD), anders.bjartell@med.lu.se (Anders Bjartell, MD, PhD), e.w.steyerberg@lumc.nl (Ewout W. Steyerberg, PhD), d.rizopoulos@erasmusmc.nl (Dimitris Rizopoulos, PhD)

to assist patients and doctors in making shared decisions of biopsy schedules, by providing them quantitative estimates of the *burden* and *benefit* of opting for personalized versus any other schedule in AS. Last, to externally validate our model and implement it along with personalized schedules in a ready to use web-application.

`Materials and Methods`: Repeat prostate-specific antigen (PSA) measurements, timing and results of previous biopsies, and age at baseline from the world's largest AS study, Prostate Cancer Research International Active Surveillance or PRIAS (7813 patients, 1134 experienced upgrading). We fitted a Bayesian joint model for time-to-event and longitudinal data to this dataset. We then validated our model externally in the largest six AS cohorts of the Movember Foundation's Global Action Plan (GAP3) database ($> 20,000$ patients, 27 centers worldwide). Using the model predicted upgrading-risks, we scheduled biopsies whenever a patient's upgrading-risk was above a certain threshold. To assist patients/doctors in choice of this threshold, and to compare the resulting personalized schedule with currently practiced schedules, along with the timing and the total number of biopsies (burden) planned, for each schedule we provided them the time delay expected in detecting upgrading (shorter is better).

`Results`: The cause-specific cumulative upgrading-risk at year five of follow-up was 35% in PRIAS, and at most 50% in GAP3 cohorts. In the PRIAS based model, PSA velocity was a stronger predictor of upgrading (Hazard Ratio: 2.47, 95%CI: 1.93–2.99) than PSA value (Hazard Ratio: 0.99, 95%CI: 0.89–1.11). Our model had a moderate area under the receiver operating characteristic curve (0.6–0.7) in validation cohorts. The prediction error was moderate (0.1–0.2) in validation cohorts where the impact of PSA value and velocity on upgrading-risk was similar to PRIAS, but large (0.2–0.3) otherwise. Our model required recalibration of baseline upgrading-risk in validation cohorts. We implemented the validated models and the methodology for personalized schedules in a web-application (`http://tiny.cc/biopsy`).

`Conclusions`: We successfully developed and validated a model for predicting upgrading-risk, and providing risk-based personalized biopsy decisions, in prostate cancer AS. Personalized prostate biopsies are a novel alternative to fixed one-size-fits-all schedules that may help to reduce unnecessary prostate biopsies while maintaining cancer control. The model and schedules made available via a web-application enable shared decision making of biopsy schedules by comparing fixed and personalized schedules on total biopsies and expected time delay in detecting upgrading.

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 are recommended active surveillance (AS) usually, instead of immediate radical treatment [1]. In AS, cancer progression is monitored routinely via prostate-specific antigen (PSA), digital rectal examination (DRE), repeat biopsies, and recently, magnetic resonance imaging (MRI). Among these, the strongest indicator of cancer-related outcomes is the biopsy Gleason grade group [2]. When it increases from group 1 (Gleason 3+3) to 2 (Gleason 3+4) or higher, it is called *upgrading* [3]. Upgrading is an important endpoint in AS upon which patients are commonly advised curative treatment [4].

Biopsies in AS are always conducted with a time gap between them. Consequently, upgrading is always detected with a time delay (Figure 1) that cannot be measured directly. In this regard, to detect upgrading timely, many patients are prescribed fixed and frequent biopsies, most often annually [5]. However, such one-size-fits-all schedules lead to unnecessary biopsies in slow/non-progressing patients. Biopsies are invasive, may be painful, and are prone to medical complications such as bleeding and septicemia[6]. Thus, biopsy burden and patient non-compliance to frequent biopsies [7] have raised concerns regarding the optimal biopsy schedule [8, 9] in AS.

Except for the confirmatory biopsy at year one of AS [7], opinions and practice regarding the timing of remaining biopsies lack agreement [10]. Some AS programs utilize patients' observed PSA, DRE, previous biopsy Gleason grade, and lately, MRI results to decide biopsies [11, 4, 10]. In contrast, others discourage schedules based on clinical data and MRI results [12, 5], and instead support periodical one-size-fits-all biopsy schedules. Furthermore, some suggest replacing frequent periodical schedules with infrequent ones (e.g., biennially) [8, 13]. Each of these approaches has limitations. For example, one-size-fits-all schedules can lead to many unnecessary biopsies because of differences in baseline *upgrading-risk* across cohorts [8]. Whereas, since observed clinical data has measurement error (e.g., PSA fluctuations),

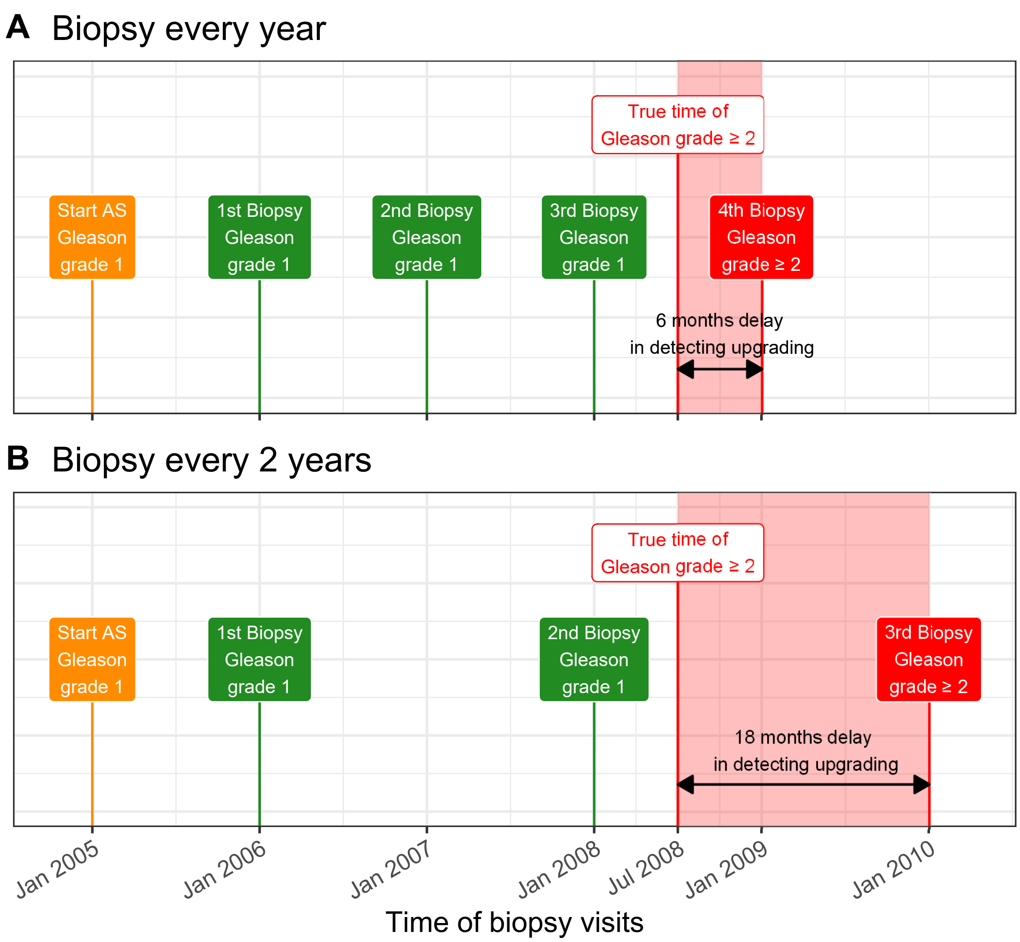

Figure 1: **Trade-off between the timing and number of biopsies (burden) and time delay in detecting Gleason upgrading (shorter is better):** The true time of Gleason upgrading (increase in Gleason grade group from group 1 to 2 or higher) for the patient in this figure is July 2008. When biopsies are scheduled annually (**Panel A**), upgrading is detected in January 2009 with a time delay of six months, and a total of four biopsies are scheduled. When biopsies are scheduled biennially (**Panel B**), upgrading is detected in January 2010 with a time delay of 18 months, and a total of three biopsies are scheduled. Since biopsies are conducted periodically, the time of upgrading is observed as an interval. For example, between Jan 2008–Jan 2009 in **Panel A** and between Jan 2008–Jan 2010 in **Panel B**. The phrase 'Gleason grade group' is shortened to 'Gleason grade' for brevity.

a flaw of using it directly is that it may lead to poor decisions. Also, decisions based on clinical data typically rely only on the latest data point and ignore previous repeated measurements. A novel alternative that counters these drawbacks is first processing patient data via a statistical model, and subsequently using model predicted upgrading-risks to create *personalized* biopsy schedules [10] (Figure 2). While, upgrading-risk calculators are not new [14, 15, 16, 17], not all are personalized either. Besides, they do not specify how risk predictions can be exploited to create a schedule.

This work is motivated by the problem of scheduling biopsies in AS. We have two goals. First, we want to assist practitioners in using clinical data in biopsy decisions in a statistically sound manner. To this end, we plan to develop a robust, generalizable statistical model that provides reliable individual upgrading-risk in AS. Subsequently, we will employ these predictions to derive risk-based personalized biopsy schedules. Our second goal is to enable shared decision making of biopsy schedules. We intend to achieve this by allowing patients and doctors to compare the *burden* and *benefit* (Figure 1) of opting for personalized schedules versus periodical schedules versus schedules based on clinical data. Specifically, we propose timing and number of planned biopsies (more/frequent are burdensome), and the expected time delay in detecting upgrading (shorter is beneficial) for any given schedule. While fulfilling our goals, we want to capture the maximum possible information from the available data. Hence, we will use all repeated measurements of patients, previous biopsy results, baseline characteristics, and keep our model flexible to accommodate novel biomarkers in the future. To fit this model, we will utilize data of the world's largest AS study, Prostate Cancer

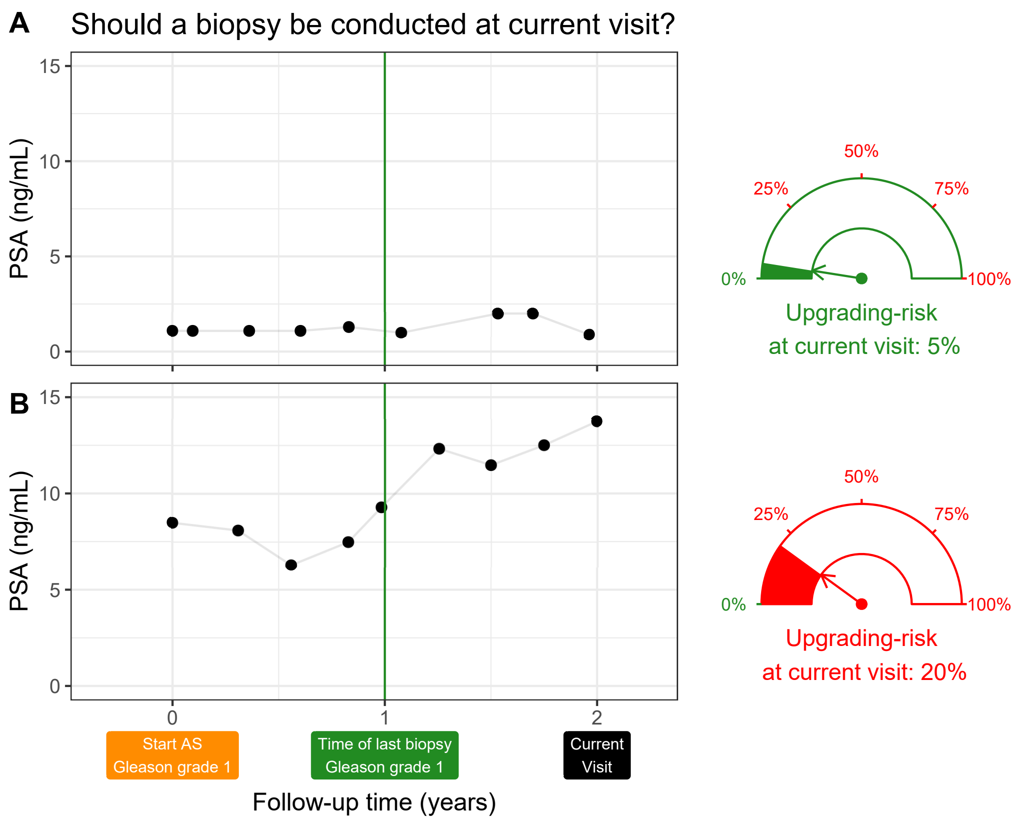

Figure 2: **Motivation for upgrading-risk based personalized biopsy decisions**: To utilize patients' complete longitudinal data and results from previous biopsies in making biopsy decisions. For this purpose, we first process data using a statistical model and then utilize the patient-specific predictions for risk of Gleason upgrading to schedule biopsies. For example, Patient A (**Panel A**) and B (**Panel B**) had their latest biopsy at year one of follow-up (green vertical line). Patient A's prostate-specific antigen (PSA) profile remained stable until his current visit at year two, whereas patient B's profile has shown a rise. Consequently, patient B's upgrading-risk at the current visit (year two) is higher than that of patient A. This makes patient B a more suitable candidate for biopsy than Patient A. Risk estimates in this figure are only illustrative.

Research International Active Surveillance (PRIAS). To evaluate our model, we will externally validate it in the largest six AS cohorts from the Movember Foundation's Global Action Plan (GAP3) database [18]. Last, we aim to implement the validated model and methodology in a web-application.

2. Patients and Methods

2.1. Study Cohort

For developing a statistical model to predict upgrading-risk, we used the world's largest AS dataset, Prostate Cancer International Active Surveillance or PRIAS [4], dated April 2019 (Table 1). In PRIAS, biopsies were scheduled at year one, four, seven, ten, and additional yearly biopsies were scheduled when PSA doubling time was between zero and ten years. We selected all 7813 patients who had Gleason grade group 1 at inclusion in AS. Our primary event of interest is an increase in this Gleason grade group observed upon repeat biopsy, called *upgrading* (1134 patients). Upgrading is a trigger for treatment advice in PRIAS. Also, 2250 patients were provided treatment based on their PSA, the number of biopsy cores with cancer, or anxiety/other reasons. However, our reasons for focusing solely on upgrading are that upgrading is strongly associated with cancer-related outcomes, and other treatment triggers vary between cohorts [10].

For externally validating our model's predictions, we selected the following largest (by the number of repeated measurements) six cohorts from Movember Foundation's GAP3 database [18] version 3.1, covering nearly 73% of the GAP3 patients: the University of Toronto AS (Toronto), Johns Hopkins AS (Hopkins), Memorial Sloan Kettering Cancer Center AS (MSKCC), King's College London AS (KCL), Michigan Urological Surgery Improvement Collaborative AS (MUSIC), and University of California San Francisco AS (UCSF, version 3.2). Only patients with a Gleason grade group 1 at the time of inclusion in these cohorts were selected. Summary statistics are presented in Supplementary A.2.

Choice of predictors:. In our model, we used all repeated PSA measurements, the timing of the previous biopsy and Gleason grade, and age at inclusion in AS. Other predictors such as prostate volume, MRI results can also be important. MRI is utilized already for targeting biopsies, but regarding its use in deciding the time of biopsies, there are arguments both for and against it [11, 12, 19]. MRI is still a recent addition in most AS protocols. Consequently, repeated MRI data is very sparsely available in both PRIAS and GAP3 databases to make a stable prediction model. Prostate volume data is also sparsely available, especially in validation cohorts. Based on these reasons, we did not include them in our model. However, the model we propose next is extendable to include MRI and other novel biomarkers in the future.

2.2. Statistical Model

Modeling an AS dataset such as PRIAS, posed certain challenges. First, PSA was measured longitudinally, and over follow-up time, it did not always increase linearly. Also, PSA was available only until a patient observed upgrading. Hence, we need to accommodate the within-patient correlation for PSA, the association between the Gleason grades and PSA profiles of a patient, and handle missing PSA measurements after a patient experienced upgrading. Second, since the PRIAS biopsy schedule uses PSA, a patient's

Table 1: **Summary of the PRIAS dataset as of April 2019**. The primary event of interest is upgrading, that is, increase in Gleason grade group from group 1 [2] to 2 or higher. IQR: interquartile range, PSA: prostate-specific antigen. Study protocol URL: `https://www.prias-project.org`

Characteristic	**Value**
Total patients	7813
Upgrading (primary event)	1134
Treatment	2250
Watchful waiting	334
Loss to follow-up	249
Death (unrelated to prostate cancer)	95
Death (related to prostate cancer)	2
Median age at diagnosis (years)	66 (IQR: 61–71)
Median maximum follow-up per patient (years)	1.8 (IQR: 0.9–4.0)
Total PSA measurements	67578
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	15686
Median number of biopsies per patient	2 (IQR: 1–2)

observed time of upgrading was also dependent on their PSA. Thus, the effect of PSA on the upgrading-risk need to be adjusted for the effect of PSA on the biopsy schedule. Third, many patients obtained treatment and watchful waiting before observing upgrading. Since we considered events other than upgrading as censoring, the model needs to account for patients' reasons for treatment or watchful waiting (e.g., age, treatment based on observed data). A model that handles these challenges in a statistically sound manner is the joint model for time-to-event and longitudinal data [20, 14, 21].

Our joint model consisted of two sub-models. Namely, a linear mixed-effects sub-model [22] for longitudinally measured PSA (log-transformed), and a relative-risk sub-model (similar to the Cox model) for the interval-censored time of upgrading. Patient age was used in both sub-models. Results and timing of the previous negative biopsies were used only in the risk sub-model. To account for PSA fluctuations [23], we assumed t-distributed PSA measurement errors. The correlation between PSA measurements of the same patient was established using patient-specific random-effects. We fitted a unique curve to the PSA measurements of each patient (Panel A, Figure 3). Subsequently, we calculated the mathematical derivative of the patient's fitted PSA profile (Equation 2, Supplementary A), to obtain his follow-up time specific instantaneous PSA velocity (Panel B, Figure 3). This instantaneous velocity is a stronger predictor of upgrading than the widely used average PSA velocity [24]. We modeled the impact of PSA on upgrading-risk by employing fitted PSA value and instantaneous velocity as predictors in the risk sub-model (Panel C, Figure 3). We adjusted the effect of PSA on upgrading-risk for the PSA dependent PRIAS biopsy schedule by estimating parameters using a full likelihood method (proof in Supplementary A). This approach also accommodates watchful waiting and treatment protocols that are also based on patient data. Specifically, the parameters of our two sub-models were estimated jointly under the Bayesian paradigm (Supplementary A) using the R package **JMbayes** [25].

2.3. Risk Prediction and Model Validation

Our model provides predictions for upgrading-risk over the entire future follow-up period of a patient (Panel C, Figure 3). However, we recommend using predictions only after year one. This is because most AS programs recommend a confirmatory biopsy at year one, especially to detect patients who may be misdiagnosed as low-grade at inclusion in AS. The model also automatically updates risk-predictions over follow-up as more patient data becomes available (Figure 5, Supplementary B). We validated our model internally in the PRIAS cohort, and externally in the largest six GAP3 database cohorts. We employed calibration plots [26, 27] and follow-up *time-dependent* mean absolute risk prediction error or MAPE [28] to graphically and quantitatively evaluate our model's risk prediction accuracy, respectively. We assessed our model's ability to discriminate between patients who experience/do not experience upgrading via the time-dependent area under the receiver operating characteristic curve or AUC [28].

The aforementioned *time-dependent* AUC and MAPE [28] are temporal extensions of their standard versions [27] in a longitudinal setting. Specifically, at every six months of follow-up, we calculated a unique AUC and MAPE for predicting upgrading-risk in the subsequent one year (Supplementary B.1). For emulating a realistic situation, we calculated the AUC and

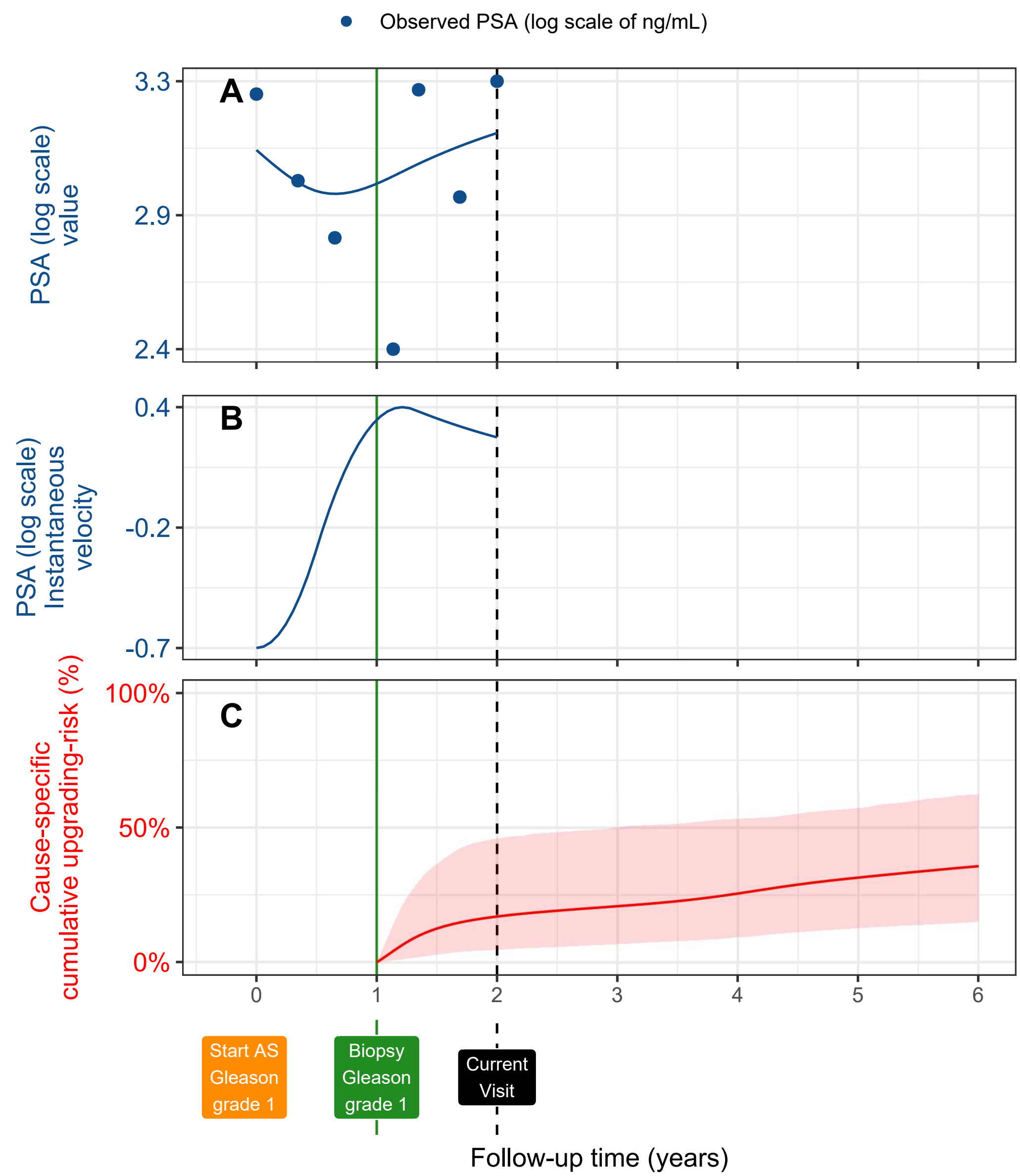

Figure 3: **Illustration of the joint model on a real PRIAS patient**. **Panel A:** Observed PSA (blue dots) and fitted PSA (solid blue line), log-transformed from ng/mL. **Panel B:** Estimated instantaneous velocity of PSA (log-transformed). **Panel C**: Predicted cause-specific cumulative upgrading-risk (95% credible interval shaded). Upgrading is defined as an increase in the Gleason grade group from group 1 [2] to 2 or higher. This upgrading-risk is calculated starting from the time of the latest negative biopsy (vertical green line at year one of follow-up). The joint model estimated it by combining the fitted PSA (log scale) value and instantaneous velocity, and time of the latest negative biopsy. Black dashed line at year two denotes the time of current visit.

MAPE at each follow-up using only the validation data available until that follow-up. Last, to resolve any potential model miscalibration in validation cohorts, we aimed to recalibrate our model's baseline hazard of upgrading (Supplementary B.1), individually for each cohort.

3. Results

The cause-specific cumulative upgrading-risk at year five of follow-up was 35% in PRIAS and at most 50% in validation cohorts (Panel B, Figure 4). In the fitted PRIAS model, the adjusted hazard ratio (aHR) of upgrading for an increase in patient age from 61 to 71 years (25-th to 75-th percentile) was 1.45 (95%CI: 1.30–1.63). For an increase in fitted PSA value from 2.36 to 3.07 (25-th to 75-th percentile, log scale), the aHR was 0.99 (95%CI: 0.89–1.11). The strongest predictor of upgrading-risk was instantaneous PSA velocity, with an increase from -0.09 to 0.31 (25-th to 75-th percentile), giving an aHR of 2.47 (95%CI: 1.93–2.99). The aHR for PSA value and velocity was different in each GAP3 cohort (Supplementary Table 8).

The time-dependent AUC, calibration plot, and time-dependent MAPE of our model are shown in Figure 4, and Supplementary Figure 8. In all cohorts, time-dependent AUC was moderate (0.6 to 0.7) over the whole follow-up period. Time-dependent MAPE was moderate (0.1 to 0.2) in those cohorts where the impact of PSA on upgrading-risk was similar to PRIAS (e.g., Hopkins cohort, Supplementary Table 8), and large (0.2 to 0.3) otherwise. Our model was miscalibrated for validation cohorts (Panel B, Figure 4). Recalibrating the baseline hazard of upgrading in validation cohorts resolved this issue (Supplementary Figure 6). We compared risk predictions from the recalibrated models, with predictions from separately fitted cohort-specific joint models (Supplementary Figure 7). The difference in predictions was lowest in the Johns Hopkins cohort (impact of PSA on upgrading-risk similar to PRIAS). Comprehensive results are in Supplementary A.3 and Supplementary B.

3.1. Personalized Biopsy Schedules

We employed the PRIAS based fitted model to create personalized biopsy schedules for real PRIAS patients. Particularly, first using the model and patient's observed data, we predicted his cumulative upgrading-risk (Figure 5) on all of his future follow-up visits (biannually in PRIAS). Subsequently, we planned biopsies on those future visits where his conditional cumulative upgrading-risk was more than a certain threshold (see Supplementary C for mathematical details). The choice of this threshold dictates the timing of biopsies in a risk-based personalized schedule. For example, personalized schedules based on 5% and 10% risk thresholds are shown in Figure 5, and in Supplementary Figure 10–12.

To facilitate the choice of a risk-threshold, and for comparing the consequences of opting for a risk-based schedule versus any other schedule (e.g., annual, PRIAS), we predict expected time delay in detecting upgrading for following a schedule. We are able to predict this delay for any schedule. For example, in Panel C of Figure 5, the annual schedule has the least expected delay. In contrast, a personalized schedule based on a 10% risk threshold has a slightly larger expected delay, but it also schedules much fewer biopsies. An important aspect of this delay is that it is personalized as well. That is, even if two different patients are prescribed the same biopsy schedule, their

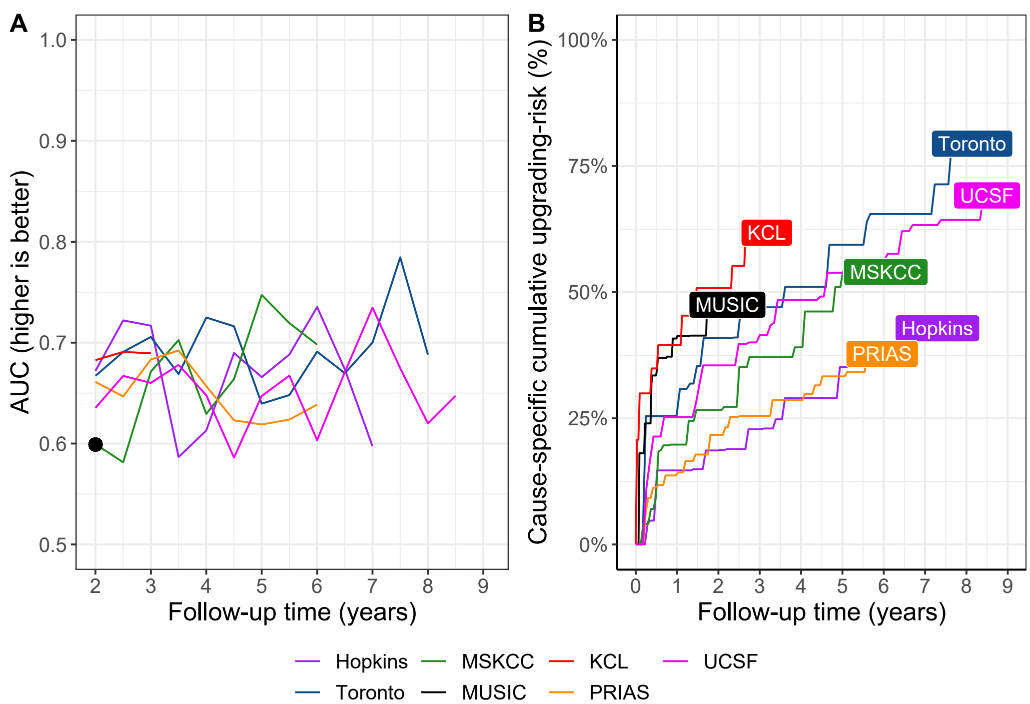

Figure 4: **Model Validation Results**. **Panel A**: time-dependent area under the receiver operating characteristic curve or AUC (measure of discrimination). AUC at year one is not shown because we do not intend to replace the confirmatory biopsy at year one. **Panel B**: calibration-at-large indicates model miscalibration. This is because solid lines depicting the non-parameteric estimate of the cause-specific cumulative upgrading-risk [29], and dashed lines showing the average cause-specific cumulative upgrading-risk obtained using the joint model fitted to the PRIAS dataset, are not overlapping. Recalibrating the baseline hazard of upgrading resolved this issue (Supplementary Figure 6). Full names of Cohorts are *PRIAS*: Prostate Cancer International Active Surveillance, *Toronto*: University of Toronto Active Surveillance, *Hopkins*: Johns Hopkins Active Surveillance, *MSKCC*: Memorial Sloan Kettering Cancer Center Active Surveillance, *KCL*: King's College London Active Surveillance, *MUSIC*: Michigan Urological Surgery Improvement Collaborative Active Surveillance, *UCSF*: University of California San Francisco AS.

expected delays will be different. This is because delay is estimated using all available clinical data of the patient (see Supplementary C). While the timing and the total number of planned biopsies denote the burden of a schedule, a shorter expected time delay in detecting upgrading can be a benefit. These two, along with other measures such as a patient's comorbidities, anxiety, etc., can help to make an informed biopsy decision.

3.2. Web-Application

We implemented the PRIAS based model, recalibrated models for GAP3 cohorts, and personalized schedules in a user-friendly web-application `https://emcbiostatistics.shinyapps.io/prias_biopsy_recommender/`. This application works on both desktop and mobile devices. Patient data can be entered in Microsoft Excel format. The maximum follow-up time up to which predictions can be obtained depends on each cohort (Supplementary Table 9). The web-application supports personalized, annual, and PRIAS schedules. For personalized schedules, users can control the choice of risk-threshold. The web-application also compares the resulting risk-based schedule's timing of biopsies, and expected time delay in detecting upgrading, with annual and PRIAS schedules, to enable sharing biopsy decision making.

4. Discussion

We successfully developed and externally validated a statistical model for predicting upgrading-risk [3] in prostate cancer AS, and providing risk-based personalized biopsy decisions. Our work has four novel features over earlier risk calculators [14, 15]. First, our model was fitted to the world's largest AS dataset PRIAS and externally validated in the largest six cohorts of the

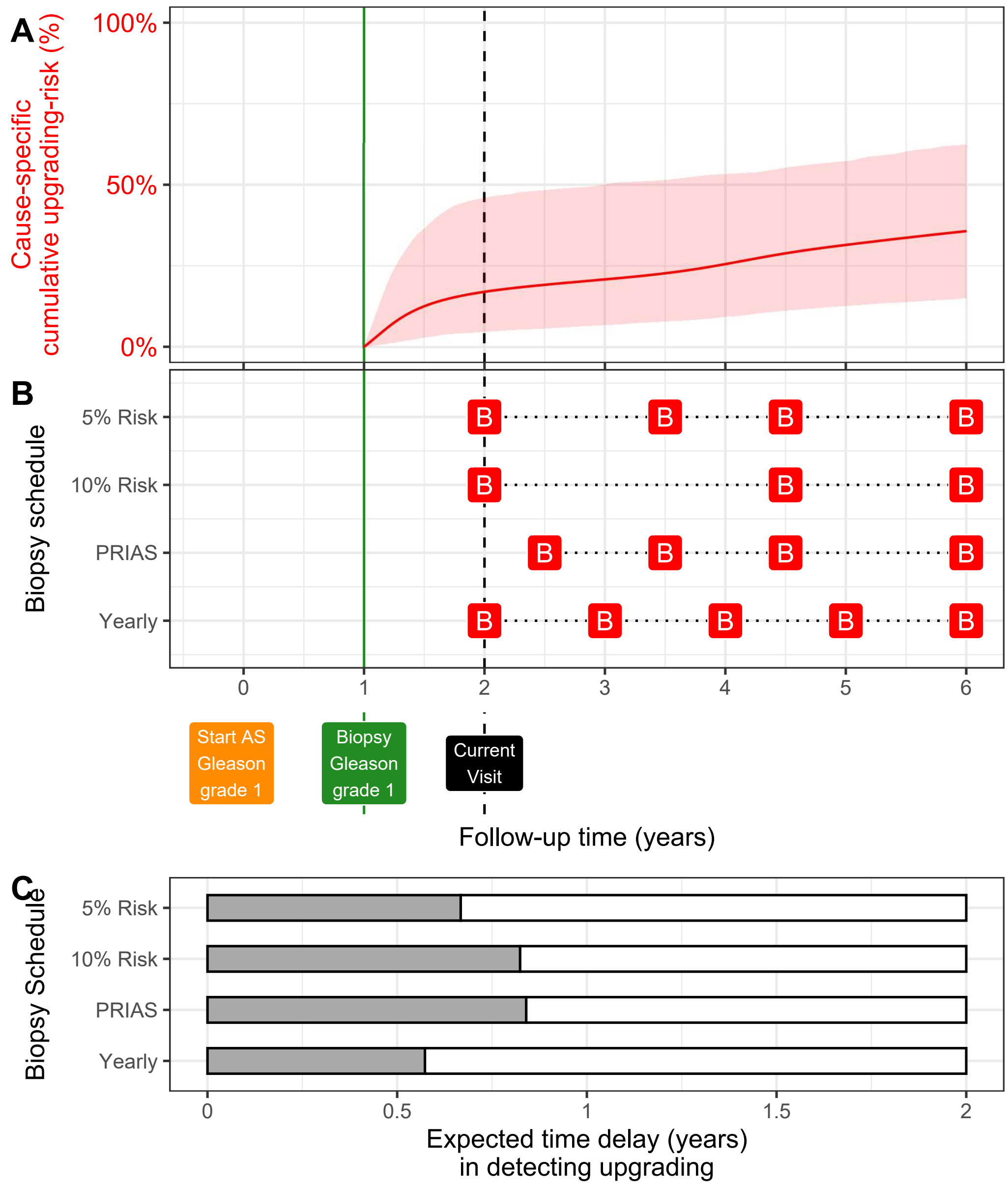

Figure 5: **Illustration of personalized and fixed schedules of biopsies for patient from Figure 3**. **Panel A:** Predicted cumulative upgrading-risk (95% credible interval shaded). **Panel B:** Different biopsy schedules with a red 'B' indicating a future biopsy. Risk: 5% and Risk: 10% are personalized schedules in which a biopsy is planned whenever the conditional cause-specific cumulative upgrading-risk is above 5% or 10% risk, respectively. Green vertical line at year one is the time of the latest negative biopsy. Black dashed line at year two denotes the time of the current visit. **Panel C:** Expected time delay in detecting upgrading (years) if patient progresses before year six. A compulsory biopsy was scheduled at year six (maximum biopsy scheduling time in PRIAS, Supplementary C) in all schedules for a meaningful comparison between them.

Movember Foundation's GAP3 database [18]. Second, the model predicts a patient's current and future upgrading-risk in a personalized manner. Third, using the predicted risks, we created personalized biopsy schedules. We also calculated the expected time delay in detecting upgrading (less is beneficial) for following any schedule. Thus, patients/doctors can compare schedules before making a choice. Fourth, we implemented our methodology in a user-friendly web-application (`https://emcbiostatistics.shinyapps.io/prias_biopsy_recommender/`) for both PRIAS and validated cohorts.

Our model and methods can be useful for numerous patients from PRIAS and the validated GAP3 cohorts (nearly 73% of all GAP3 patients). The model utilizes all repeated PSA measurements, results of previous biopsies, and baseline characteristics of a patient. We could not include MRI and PSA volume because of sparsely available data in both PRIAS and GAP3 databases. But, our model is extendable to include them in the near future. The current discrimination ability of our model, exhibited by the *time-dependent* AUC, was between 0.6 and 0.7 over-follow. While this is moderate, it is also so because unlike the standard AUC [27] the time-dependent AUC is more conservative as it utilizes only the validation data available until the time at which it is calculated. The same holds for the time-dependent MAPE (mean absolute prediction error). Although, MAPE varied much more between cohorts than AUC. In cohorts where the effect size for the impact of PSA value and velocity on upgrading-risk was similar to that for PRIAS (e.g., Hopkins cohort), MAPE was moderate. Otherwise, MAPE was large (e.g., KCL and MUSIC cohorts). We required recalibration of our model's baseline hazard of upgrading for all validation cohorts.

The clinical implications of our work are as follows. First, the cause-specific cumulative upgrading-risk at year five of follow-up was at most 50% in all cohorts (Panel B, Figure 4). That is, many patients may not require some of the biopsies planned in the first five years of AS. Given the non-compliance and burden of frequent biopsies [7], the availability of our methodology as a web-application may encourage patients/doctors to consider upgrading-risk based personalized schedules instead. An additional advantage of personalized schedules is that they update as more patient data becomes available over follow-up. We have shown via a simulation study [30] that personalized schedules plan, on average, six fewer biopsies compared to annual schedule and two fewer biopsies than the PRIAS schedule in slow/non-progressing AS patients, while maintaining almost the same time delay in detecting upgrading as PRIAS schedule. Personalized schedules with different risk thresholds indeed have different performance. In this regard, to assist patients/doctors in choosing between fixed schedules and personalized schedules based on different risk thresholds, the web-application provides a patient-specific estimate of the expected time delay in detecting upgrading, for both personalized and fixed schedules. We hope that this will objectively address patient apprehensions regarding adverse outcomes in AS. Last, we note that our web-application should only be used to decide biopsies after the compulsory confirmatory biopsy at year one of follow-up.

This work has certain limitations. Predictions for upgrading-risk and personalized schedules are available only for a currently limited, cohort-specific, follow-up period (Supplementary Table 9). This problem can be mitigated by refitting the model with new follow-up data in the future. Recently, some cohorts started utilizing MRI to explore the possibility of targeting visible lesions by biopsy. Presently, the GAP3 database has limited MRI follow-up data available. As more such data becomes available, the current model can be extended to include MRI based predictors. We scheduled biopsies using cause-specific cumulative upgrading-risk, which ignores competing events such as treatment based on the number of positive biopsy cores. Employing a competing-risk model may lead to improved personalized schedules. Upgrading is susceptible to inter-observer variation too. Models which account for this variation [14, 31] will be interesting to investigate further. Even with an enhanced risk prediction model, the methodology for personalized scheduling and calculation of expected time delay (Supplementary C) need not change. Last, our web-application only allows uploading patient data in Microsoft Excel format. Connecting it with patient databases can increase usability.

5. Conclusions

We successfully developed a statistical model and methodology for predicting upgrading-risk, and providing risk-based personalized biopsy decisions, in prostate cancer AS. We externally validated our model, covering nearly 73% patients from the Movember Foundations' GAP3 database. The model made available via a user-friendly web-application (`https://emcbiostatistics.shinyapps.io/prias_biopsy_recommender/`) enables shared decision making of biopsy schedules by comparing fixed and personalized schedules on total biopsies and expected time delay in detecting upgrading. Novel biomarkers and MRI data can be added as predictors in the model to improve predictions in the future. Recalibration of baseline upgrading-risk is advised for cohorts not validated in this work.

Author Contributions

Anirudh Tomer had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Tomer, Nieboer, Roobol, Bjartell, and Rizopoulos
Acquisition of data: Tomer, Nieboer, and Roobol
Analysis and interpretation of data: Tomer, Nieboer, and Rizopoulos
Drafting of the manuscript: Tomer, and Rizopoulos
Critical revision of the manuscript for important intellectual content: Tomer, Nieboer, Roobol, Bjartell, Steyerberg, and Rizopoulos
Statistical analyses: Tomer, Nieboer, Steyerberg, and Rizopoulos
Obtaining funding: Roobol, Steyerberg, and Rizopoulos
Administrative, technical or material support: Nieboer
Supervision: Roobol, and Rizopoulos
Other: none

Acknowledgments

We thank Jozien Helleman from the Department of Urology, Erasmus University Medical Center, for coordinating the project. The first and last authors would like to acknowledge support by Nederlandse Organisatie voor Wetenschappelijk Onderzoek (the national research council of the Netherlands) VIDI grant nr. 016.146.301, and Erasmus University Medical Center funding. Part of this work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative. The authors also thank the Erasmus University Medical Center's Cancer Computational Biology Center for giving access to their IT-infrastructure and software that was used for the computations and data analysis in this study. The PRIAS website is funded by the Prostate Cancer Research Foundation, Rotterdam (SWOP). We would like to thank the PRIAS consortium for enabling this research project.

This work was supported by the Movember Foundation. The funder did not play any role in the study design, collection, analysis or interpretation of data, or in the drafting of this paper.

Conflicts of Interest

The authors do not report any conflict of interest, and have nothing to disclose.

Appendix A. Members of The Movember Foundation's Global Action Plan Prostate Cancer Active Surveillance (GAP3) consortium

Principle Investigators: Bruce Trock (Johns Hopkins University, The James Buchanan Brady Urological Institute, Baltimore, USA), Behfar Ehdaie (Memorial Sloan Kettering Cancer Center, New York, USA), Peter Carroll (University of California San Francisco, San Francisco, USA), Christopher Filson (Emory University School of Medicine, Winship Cancer Institute, Atlanta, USA), Jeri Kim / Christopher Logothetis (MD Anderson Cancer Centre, Houston, USA), Todd Morgan (University of Michigan and

Michigan Urological Surgery Improvement Collaborative (MUSIC), Michigan, USA), Laurence Klotz (University of Toronto, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada), Tom Pickles (University of British Columbia, BC Cancer Agency, Vancouver, Canada), Eric Hyndman (University of Calgary, Southern Alberta Institute of Urology, Calgary, Canada), Caroline Moore (University College London & University College London Hospital Trust, London, UK), Vincent Gnanapragasam (University of Cambridge & Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK), Mieke Van Hemelrijck (King's College London, London, UK & Guy's and St Thomas' NHS Foundation Trust, London, UK), Prokar Dasgupta (Guy's and St Thomas' NHS Foundation Trust, London, UK), Chris Bangma (Erasmus Medical Center, Rotterdam, The Netherlands/ representative of Prostate cancer Research International Active Surveillance (PRIAS) consortium), Monique Roobol (Erasmus Medical Center, Rotterdam, The Netherlands/ representative of Prostate cancer Research International Active Surveillance (PRIAS) consortium), The PRIAS study group, Arnauld Villers (Lille University Hospital Center, Lille, France), Antti Rannikko (Helsinki University and Helsinki University Hospital, Helsinki, Finland), Riccardo Valdagni (Department of Oncology and Hemato-oncology, Università degli Studi di Milano, Radiation Oncology 1 and Prostate Cancer Program, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy), Antoinette Perry (University College Dublin, Dublin, Ireland), Jonas Hugosson (Sahlgrenska University Hospital, Göteborg, Sweden), Jose Rubio-Briones (Instituto Valenciano de Oncología, Valencia, Spain), Anders Bjartell (Skåne University Hospital, Malmö, Sweden), Lukas Hefermehl (Kantonsspital Baden,

Baden, Switzerland), Lee Lui Shiong (Singapore General Hospital, Singapore, Singapore), Mark Frydenberg (Monash Health; Monash University, Melbourne, Australia), Yoshiyuki Kakehi / Mikio Sugimoto (Kagawa University Faculty of Medicine, Kagawa, Japan), Byung Ha Chung (Gangnam Severance Hospital, Yonsei University Health System, Seoul, Republic of Korea)

Pathologist: Theo van der Kwast (Princess Margaret Cancer Centre, Toronto, Canada). Technology Research Partners: Henk Obbink (Royal Philips, Eindhoven, the Netherlands), Wim van der Linden (Royal Philips, Eindhoven, the Netherlands), Tim Hulsen (Royal Philips, Eindhoven, the Netherlands), Cees de Jonge (Royal Philips, Eindhoven, the Netherlands).

Advisory Regional statisticians: Mike Kattan (Cleveland Clinic, Cleveland, Ohio, USA), Ji Xinge (Cleveland Clinic, Cleveland, Ohio, USA), Kenneth Muir (University of Manchester, Manchester, UK), Artitaya Lophatananon (University of Manchester, Manchester, UK), Michael Fahey (Epworth HealthCare, Melbourne, Australia), Ewout Steyerberg (Erasmus Medical Center, Rotterdam, The Netherlands), Daan Nieboer (Erasmus Medical Center, Rotterdam, The Netherlands); Liying Zhang (University of Toronto, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada)

Executive Regional statisticians: Ewout Steyerberg (Erasmus Medical Center, Rotterdam, The Netherlands), Daan Nieboer (Erasmus Medical Center, Rotterdam, The Netherlands); Kerri Beckmann (King's College London, London, UK & Guy's and St Thomas' NHS Foundation Trust, London, UK), Brian Denton (University of Michigan, Michigan, USA), Andrew Hayen (University of Technology Sydney, Australia), Paul Boutros (Ontario Institute of

Cancer Research, Toronto, Ontario, Canada).

Clinical Research Partners' IT Experts: Wei Guo (Johns Hopkins University, The James Buchanan Brady Urological Institute, Baltimore, USA), Nicole Benfante (Memorial Sloan Kettering Cancer Center, New York, USA), Janet Cowan (University of California San Francisco, San Francisco, USA), Dattatraya Patil (Emory University School of Medicine, Winship Cancer Institute, Atlanta, USA), Emily Tolosa (MD Anderson Cancer Centre, Houston, Texas, USA), Tae-Kyung Kim (University of Michigan and Michigan Urological Surgery Improvement Collaborative, Ann Arbor, Michigan, USA), Alexandre Mamedov (University of Toronto, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada), Vincent LaPointe (University of British Columbia, BC Cancer Agency, Vancouver, Canada), Trafford Crump (University of Calgary, Southern Alberta Institute of Urology, Calgary, Canada), Vasilis Stavrinides (University College London & University College London Hospital Trust, London, UK), Jenna Kimberly-Duffell (University of Cambridge & Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK), Aida Santaolalla (King's College London, London, UK & Guy's and St Thomas' NHS Foundation Trust, London, UK), Daan Nieboer (Erasmus Medical Center, Rotterdam, The Netherlands), Jonathan Olivier (Lille University Hospital Center, Lille, France), Tiziana Rancati (Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy), Helén Ahlgren (Sahlgrenska University Hospital, Göteborg, Sweden), Juanma Mascarós (Instituto Valenciano de Oncología, Valencia, Spain), Annica Löfgren (Skåne University Hospital, Malmö, Sweden), Kurt Lehmann (Kantonsspital Baden, Baden, Switzerland), Catherine Han Lin (Monash University and Epworth

HealthCare, Melbourne, Australia), Hiromi Hirama (Kagawa University, Kagawa, Japan), Kwang Suk Lee (Yonsei University College of Medicine, Gangnam Severance Hospital, Seoul, Korea).

Research Advisory Committee: Guido Jenster (Erasmus MC, Rotterdam, the Netherlands), Anssi Auvinen (University of Tampere, Tampere, Finland), Anders Bjartell (Skåne University Hospital, Malmö, Sweden), Masoom Haider (University of Toronto, Toronto, Canada), Kees van Bochove (The Hyve B.V. Utrecht, Utrecht, the Netherlands), Ballentine Carter (Johns Hopkins University, Baltimore, USA – until 2018).

Management team: Sam Gledhill (Movember Foundation, Melbourne, Australia), Mark Buzza / Michelle Kouspou (Movember Foundation, Melbourne, Australia), Chris Bangma (Erasmus Medical Center, Rotterdam, The Netherlands), Monique Roobol (Erasmus Medical Center, Rotterdam, The Netherlands), Sophie Bruinsma / Jozien Helleman (Erasmus Medical Center, Rotterdam, The Netherlands).

References

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

[2] Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA. The 2014 international society of urological pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma. The American journal of surgical pathology 2016;40(2):244–52.

[3] Bruinsma SM, Roobol MJ, Carroll PR, Klotz L, Pickles T, Moore CM, et al. Expert consensus document: semantics in active surveillance for men with localized prostate cancer—results of a modified Delphi consensus procedure. Nature reviews urology 2017;14(5):312.

[4] Bul M, Zhu X, Valdagni R, Pickles T, Kakehi Y, Rannikko A, et al. Active surveillance for low-risk prostate cancer worldwide: the PRIAS study. European urology 2013;63(4):597–603.

[5] 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.

[6] Loeb S, Vellekoop A, Ahmed HU, Catto J, Emberton M, Nam R, et al. Systematic review of complications of prostate biopsy. European urology 2013;64(6):876–92.

[7] Bokhorst LP, Alberts AR, Rannikko A, Valdagni R, Pickles T, Kakehi Y, et al. Compliance rates with the Prostate Cancer Research International Active Surveillance (PRIAS) protocol and disease reclassification in noncompliers. European Urology 2015;68(5):814–21.

[8] Inoue LY, Lin DW, Newcomb LF, Leonardson AS, Ankerst D, Gulati R, et al. Comparative analysis of biopsy upgrading in four prostate cancer active surveillance cohorts. Annals of internal medicine 2018;168(1):1–9.

[9] Bratt O, Carlsson S, Holmberg E, Holmberg L, Johansson E, Josefsson A, 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.

[10] 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.

[11] Kasivisvanathan V, Giganti F, Emberton M, Moore CM. Magnetic resonance imaging should be used in the active surveillance of patients with localised prostate cancer. European urology 2020;77(3):318–9.

[12] Chesnut GT, Vertosick EA, Benfante N, Sjoberg DD, Fainberg J, Lee T, et al. Role of changes in magnetic resonance imaging or clinical stage in evaluation of disease progression for men with prostate cancer on active surveillance. European Urology 2020;77(4):501–7.

[13] 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.

[14] 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.

[15] Ankerst DP, Xia J, Thompson Jr IM, Hoefler J, Newcomb LF, Brooks JD, et al. Precision medicine in active surveillance for prostate cancer: development of the canary–early detection research network active surveillance biopsy risk calculator. European urology 2015;68(6):1083–8.

[16] Partin AW, Yoo J, Carter HB, Pearson JD, Chan DW, Epstein JI, et al. The use of prostate specific antigen, clinical stage and Gleason score to predict pathological stage in men with localized prostate cancer. The Journal of urology 1993;150(1):110–4.

[17] Makarov DV, Trock BJ, Humphreys EB, Mangold LA, Walsh PC, Epstein JI, et al. Updated nomogram to predict pathologic stage of prostate cancer given prostate-specific antigen level, clinical stage, and biopsy Gleason score (Partin tables) based on cases from 2000 to 2005. Urology 2007;69(6):1095–101.

[18] Bruinsma SM, Zhang L, Roobol MJ, Bangma CH, Steyerberg EW, Nieboer D, 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.

[19] Schoots IG, Petrides N, Giganti F, Bokhorst LP, Rannikko A, Klotz L, et al. Magnetic resonance imaging in active surveillance of prostate cancer: a systematic review. European urology 2015;67(4):627–36.

[20] 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.

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

[22] Laird NM, Ware JH, et al. Random-effects models for longitudinal data. Biometrics 1982;38(4):963–74.

[23] Nixon RG, Wener MH, Smith KM, Parson RE, Strobel SA, Brawer MK. Biological variation of prostate specific antigen levels in serum: an evaluation of day-to-day physiological fluctuations in a well-defined cohort of 24 patients. The Journal of urology 1997;157(6):2183–90.

[24] Cooperberg MR, Brooks JD, Faino AV, Newcomb LF, Kearns JT, Carroll PR, et al. Refined analysis of prostate-specific antigen kinetics to predict prostate cancer active surveillance outcomes. European urology 2018;74(2):211–7.

[25] 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.

[26] Royston P, Altman DG. External validation of a cox prognostic model: principles and methods. BMC medical research methodology 2013;13(1):33.

[27] Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology (Cambridge, Mass) 2010;21(1):128.

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

[29] Turnbull BW. The empirical distribution function with arbitrarily

grouped, censored and truncated data. Journal of the Royal Statistical Society Series B (Methodological) 1976;38(3):290–5.

[30] Tomer A, Rizopoulos D, Nieboer D, Drost FJ, Roobol MJ, Steyerberg EW. Personalized decision making for biopsies in prostate cancer active surveillance programs. Medical Decision Making 2019;39(5):499–508.

[31] Balasubramanian R, Lagakos SW. Estimation of a failure time distribution based on imperfect diagnostic tests. Biometrika 2003;90(1):171–82.