A Ready To Use Web-Application Providing a Personalized Biopsy Schedule for Men With Low-Risk PCa Under Active Surveillance*

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 Foundations 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 Foundations Global Action Plan Prostate Cancer Active Surveillance (GAP3) consortium members presented in Appendix A*

Abstract

Background: Prostate cancer active surveillance (AS) patients undergo repeat biopsies. Active treatment is advised when biopsy Gleason grade group ≥ 2 (*upgrading*). Many patients never experience upgrading, yet undergo biopsies frequently. Personalized biopsy decisions based on upgrading-risk may

*Word count abstract (headings excluded): 300; Word count text: 2484

*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)

reduce patient burden.

`Objective`: Develop a risk prediction model and web-application to assist patients/doctors in personalized biopsy decisions.

`Design, Setting, and Participants`: Model development: world's largest AS study PRIAS, 7813 patients, 1134 experienced upgrading; External validation: largest six cohorts of Movember Foundation's GAP3 database ($> 20,000$ patients, 27 centers worldwide); Data: repeat prostate-specific antigen (PSA) and biopsy Gleason grade.

`Outcome Measurements, and Statistical Analysis`: A Bayesian joint model fitted to the PRIAS dataset. This model was validated in GAP3 cohorts using risk prediction error, calibration, area under ROC (AUC). Model and personalized biopsy schedules based on predicted risks were implemented in a web-application.

`Results and Limitations`: Cause-specific cumulative upgrading-risk at year five of follow-up: 35% in PRIAS, at most 50% in GAP3 cohorts. 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). Validation: Moderate AUC (0.55–0.75) in PRIAS and GAP3 cohorts. Moderate prediction error (0.1–0.3) in GAP3 cohorts where impact of PSA value and velocity on upgrading-risk was similar to PRIAS, but large (0.3–0.45) otherwise. Recalibration advised for cohorts not vali-

dated in this work.

`Conclusions`: We successfully developed and validated a model for predicting upgrading-risk, and providing risk-based personalized biopsy decisions, in prostate cancer AS. The model made available via a web-application enables shared decision making of biopsy schedules by comparing fixed and personalized schedules on total biopsies and expected time delay in detecting upgrading.

`Patient Summary`: Personalized prostate biopsies are a novel alternative to fixed one-size-fits-all schedules. The underlying statistical models are made available through a user-friendly web-application and may help to reduce unnecessary prostate biopsies while maintaining cancer control.
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 usually recommended active surveillance (AS) instead of immediate radical treatment [1]. In AS, cancer progression is routinely monitored via prostate-specific antigen (PSA), digital rectal examination, and repeat biopsies. Among these, the strongest indicator of cancer-related outcomes is the biopsy Gleason grade group [2]. When the Gleason grade group increases from group 1 (Gleason 3+3) to 2 (Gleason 3+4) or higher, called *upgrading* [3], patients are commonly advised curative treatment [4].

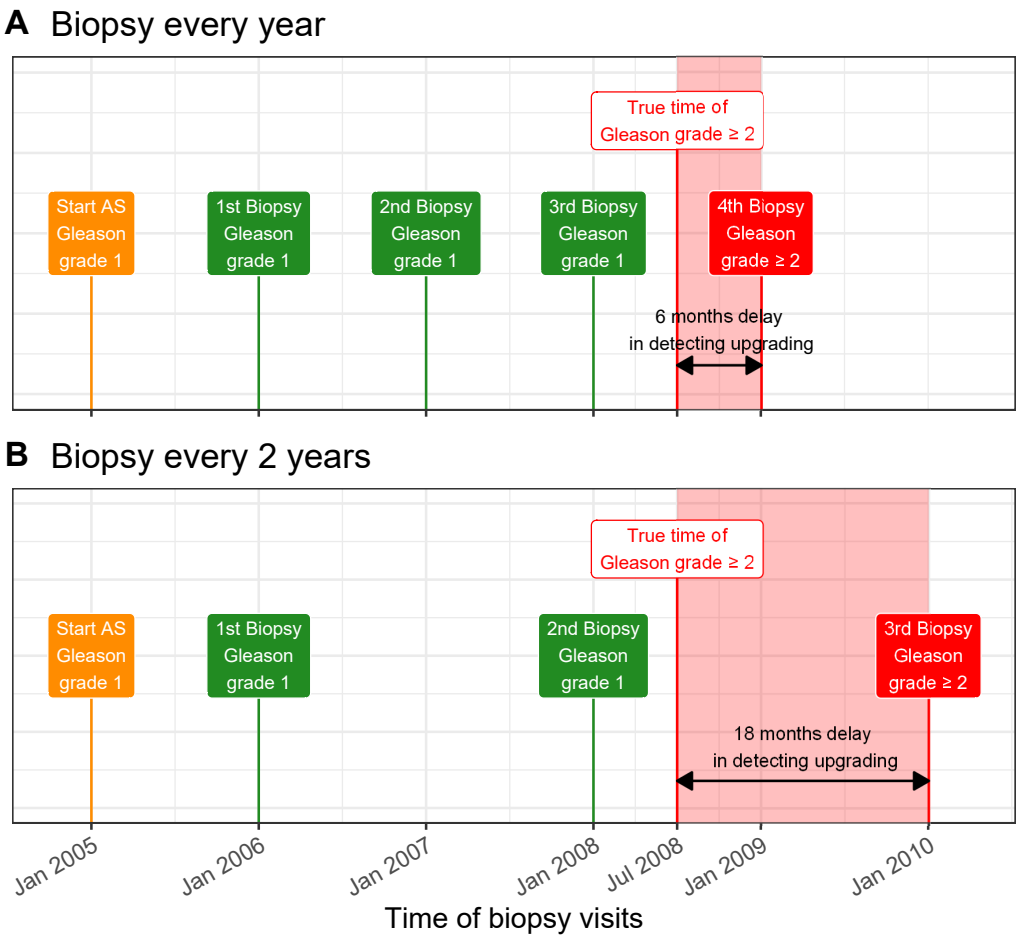

Figure 1: **Trade-off between the number of biopsies and time delay in detecting upgrading (Increase in Gleason grade group from 1 to 2 or higher):** The true time of upgrading for the patient in this figure is July 2008. When biopsies are scheduled annually (**Panel A**), upgrading is detected in January 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.

In most AS protocols, biopsies are scheduled periodically. Consequently, upgrading is always detected with a time delay (Figure 1). For detecting upgrading timely, many AS programs schedule fixed and frequent biopsies (e.g., annually) for all patients [5, 6]. However, this leads to many unnecessary biopsies in slow/non-progressing patients. Biopsies are invasive, may be painful, and are prone to medical complications such as bleeding and septicemia[7]. Thus, biopsy burden and patient non-compliance to frequent biopsies [8] has raised concerns regarding the optimal biopsy schedule [9, 10]. To this end, in some cohorts, MRI is used to explore the possibility of targeting visible tumor by biopsy, and to study the value for tumor monitoring. Although, due to currently limited AS data, MRI's value is not clear. Others have proposed infrequent schedules such as biennial biopsies as an alternative [9, 11]. Due to the differences in baseline upgrading-risk across cohorts [9], the fixed biopsy scheme may lead to many unnecessary biopsies per cohort, as well as across cohorts. A promising alternative to fixed and frequent biopsies is personalized biopsy schedules based on the patient-specific upgrading-risk (Figure 2).

The first challenge in creating personalized biopsy schedules is developing a statistical model to consolidate accumulated patient data (e.g., PSA, previous biopsy results) into estimates for upgrading-risk. Existing calculators for upgrading-risk [12, 13] use only the latest PSA measurement of a patient. In contrast, more information is captured by considering all repeated measurements of PSA, previous biopsy results, and baseline characteristics of a patient. To this end, a suitable model is the joint model for time-to-event and longitudinal data [14, 15, 16]. A joint model predicts the upgrading-risk in a

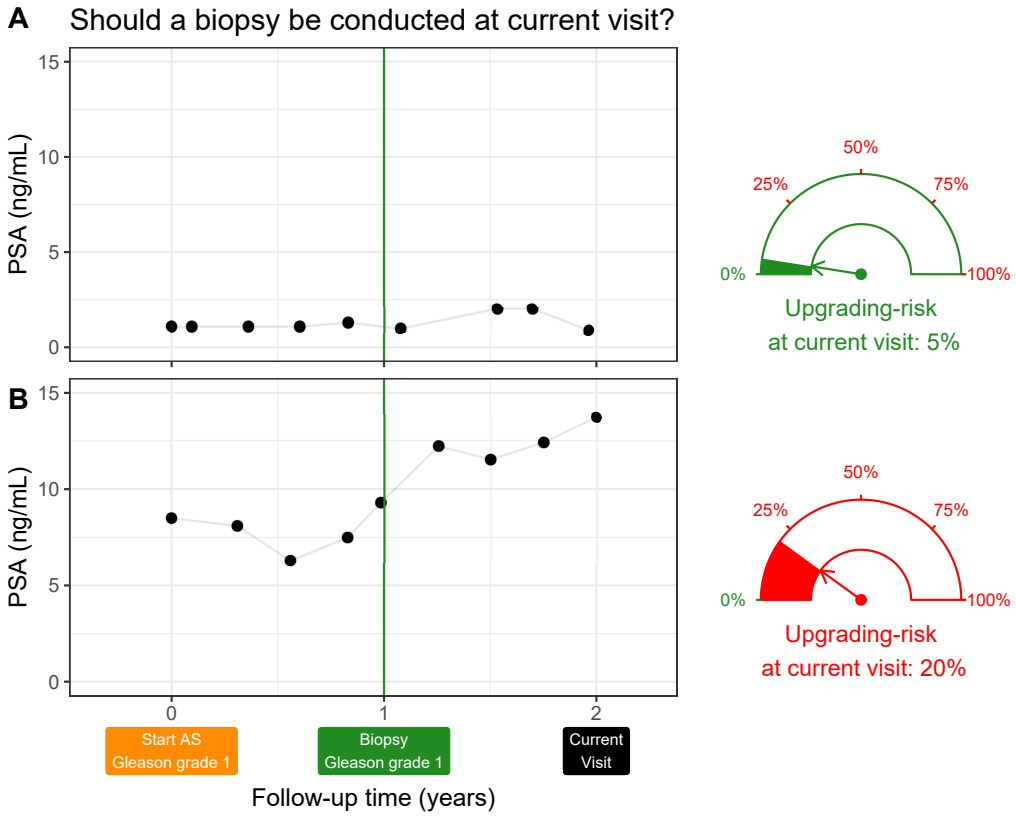

Figure 2: **Motivation for personalized upgrading-risk based decisions of biopsy**: Patient A (**Panel A**) and B (**Panel B**) had their latest biopsy at year one of follow-up (green vertical line). Patient A's prostate-specific antigen (PSA) profile remained stable until his current visit at year two, 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.

personalized manner. A subsequent challenge, however, is translating risks into clinical decisions. For example, a 10% upgrading-risk can be perceived high/low depending upon the patient's age. Patients may also weigh risks of upgrading with the potential *consequences* of another biopsy. Two relevant consequences of biopsies (Figure 1) are the timing and the total number of biopsies (burden), and the time delay in detecting upgrading (smaller is beneficial). The relative importance of these consequences can vary between the patients, and also over the follow-up period for the same patient.

The goal of this work is two fold. First, to develop a robust, generalizable model that gives reliable estimates for individual upgrading-risk in AS. Second, to utilize the predicted upgrading-risks to create personalized biopsy schedules. In order to facilitate shared decision making of biopsy schedules, we also intend to provide quantitative estimates of the *consequences* of opting for a personalized versus the standard fixed schedule. For developing our model, we will use the world's largest AS dataset PRIAS. Subsequently, we want to externally validate our model in the largest six AS cohorts from the Movember Foundation's GAP3 database [17]. Last, we intend to implement our 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] (Table 1). We used the database available as of May 2019. In PRIAS, PSA was measured quarterly for the first two years of follow-up and

semiannually thereafter. Biopsies were scheduled at year one, four, seven, and ten of follow-up. 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 the time of inclusion in PRIAS. Our primary event of interest is an increase in this Gleason grade group upon repeat biopsy, called *upgrading* (1134 patients). Upgrading is a trigger for treatment advice in PRIAS. Although, 2250 patients were provided treatment based on their PSA, or number of biopsy cores with cancer, or anxiety/other reasons. Our reasons for focusing solely on upgrading are, namely, upgrading is strongly associated with cancer-related outcomes, and other triggers for treatment vary between cohorts [5].

For model validation, we selected the largest (in terms of number of repeated measurements) six cohorts from Movember Foundation's GAP3 database version 3.1 [17]. These were, namely, the University of California San Francisco AS (UCSF, version 3.2), University of Toronto AS (Toronto), Johns Hopkins AS (Hopkins), Memorial Sloan Kettering Cancer Center AS (MSKCC), King's College London AS (KCL), and Michigan Urological Surgery Improvement Collaborative AS (MUSIC). Only patients with a Gleason grade group 1 at the time of inclusion in these cohorts were selected. Summary statistics for these cohorts are presented in Supplementary A.2.

2.2. Statistical Model

For developing an upgrading-risk prediction model, the available data in the PRIAS cohort was patient age at inclusion in AS, longitudinally measured PSA, timing of repeat biopsies and corresponding Gleason grades, and observed time of upgrading. Analysis of this data required modeling the within-

Table 1: **Summary of the PRIAS dataset as of May 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 centers	> 100
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)

patient correlation for PSA, the association between the Gleason grades and PSA profiles of a patient, and handling missing PSA measurements after a patient experienced upgrading. In such situations, a commonly used model is the joint model for time-to-event and longitudinal data [14, 15, 16].

Our joint model consisted of two sub-models. First, a linear mixed sub-model [18] for longitudinally measured PSA (log-transformed). Second, a relative-risk sub-model (similar to the Cox model) for obtaining the cause-specific upgrading-risk. Patient age was included as a predictor in both sub-models. In the PSA sub-model, 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 [19]. We modeled the impact of PSA on upgrading-risk by including fitted PSA value and instantaneous velocity as predictors in the relative-risk model. Also, the time of the latest negative biopsy was utilized in the relative-risk sub-model (Panel C, Figure 3). The parameters of the two sub-models were estimated jointly (Supplementary A) using the R package **JMbayes** [20].

2.3. Risk Prediction and Model Validation

The predictions for upgrading-risk from our model are made for the entire future follow-up period of a patient. These predictions also automatically update over follow-up as more patient data becomes available (Figure 5, Supplementary B). We validated our PRIAS based risk prediction model internally in the PRIAS cohort, and externally in the largest six GAP3 database

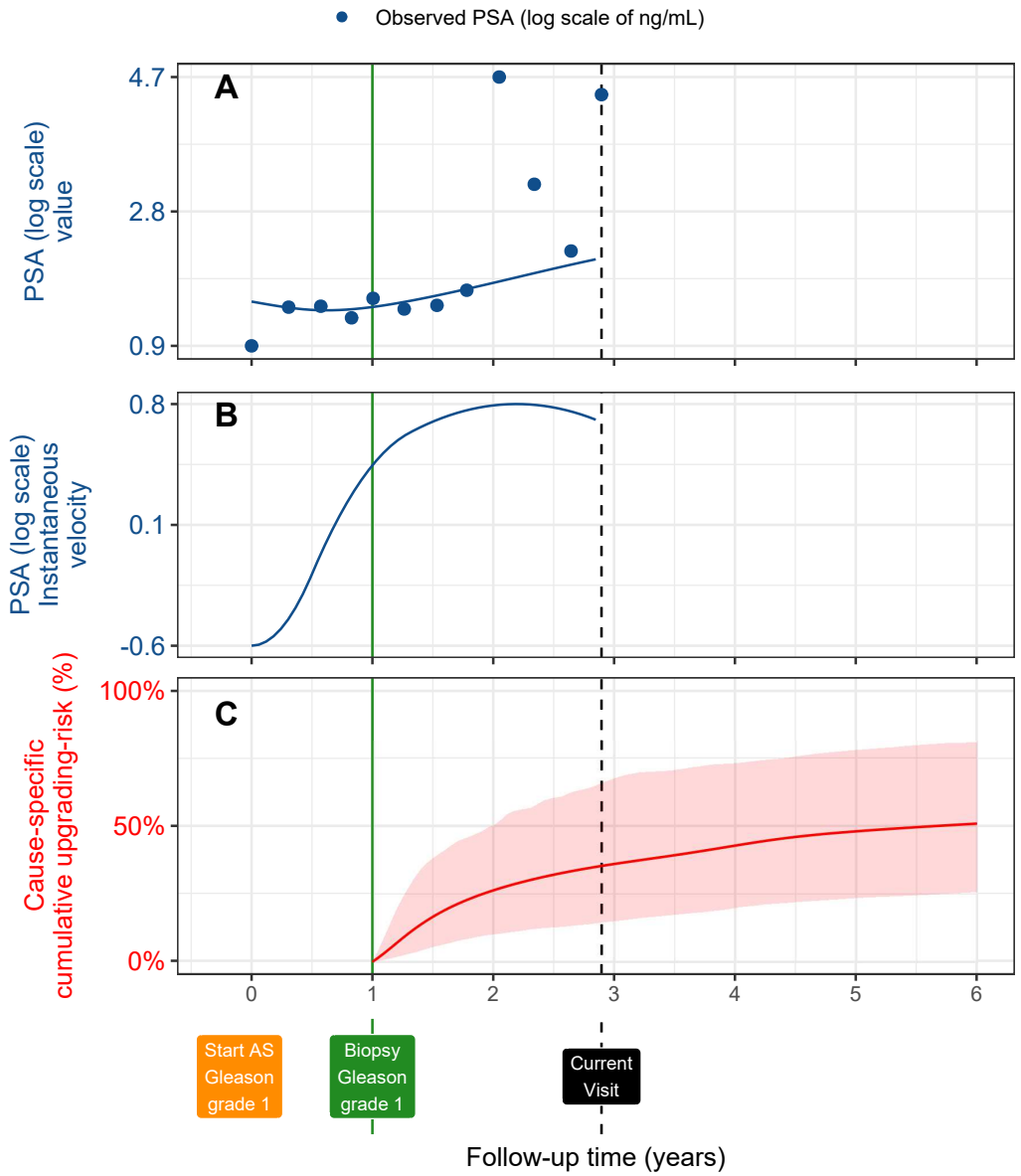

Figure 3: **Illustration of the joint model on a real PRIAS patient**. **Panel A:** Observed PSA (blue dots) and fitted PSA (solid blue line), log-transformed from ng/mL. **Panel B:** Estimated instantaneous velocity of PSA (log-transformed). **Panel C**: Predicted cause-specific cumulative upgrading-risk (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 1 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 3 denotes the time of current visit.

cohorts. We employed calibration plots [21, 22] and follow-up *time-dependent* mean absolute risk prediction error or MAPE [23] to graphically and quantitatively evaluate our model's risk prediction accuracy. 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 [23].

The aforementioned *time-dependent* AUC and MAPE [23] are temporal extensions of their standard versions [22] in a longitudinal setting. More 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 MAPE at each follow-up using only the validation data available until that follow-up. Lastly, to resolve any potential model miscalibration in validation cohorts, we aimed to recalibrate our model's baseline hazard of upgrading (Supplementary A), 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 the six validation cohorts (Panel B, Figure 4). That is, many patients may not require any biopsy in the first five years of AS.

In the joint model fitted to the PRIAS dataset, the adjusted hazard ratio 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 (log scale) from 2.36 to 3.07 (25-th to 75-th percentile), the adjusted

hazard ratio was 0.99 (95%CI: 0.89–1.11). In contrast to PSA value, instantaneous PSA velocity was a stronger predictor of upgrading-risk, because an increase in velocity from -0.09 to 0.31 (25-th to 75-th percentile) had a hazard ratio of 2.47 (95%CI: 1.93–2.99). The impact of PSA value and velocity on upgrading-risk varied between cohorts (Table 6, Supplementary A.2). Detailed results are in Supplementary A.2.

The time-dependent MAPE; time-dependent AUC; and calibration plot of our model in different validation cohorts are shown in Panel B, Figure 8, Supplementary B; Panel A, Figure 4; and Panel B, Figure 4, respectively. In all cohorts, time-dependent AUC was moderate (0.55 to 0.75) over the whole follow-up period. Time-dependent MAPE was large (0.3 to 0.45) in those cohorts where the impact of PSA value and velocity on upgrading-risk was different from PRIAS (e.g., MUSIC cohort, Table 6, Supplementary A.2), and moderate (0.1 to 0.3) otherwise. In all cohorts the MAPE decreased rapidly after year one of follow-up. To resolve issues in calibration-at-large (Panel B, Figure 4), we recalibrated the baseline hazard of upgrading in all cohorts (Figure 6, Supplementary B). We compared risk predictions from the recalibrated models with predictions from separately fitted joint models to each cohort (Figure 7, Supplementary B). The difference in predictions was lowest in Johns Hopkins cohort (impact of PSA similar to PRIAS, Table 5, Supplementary A.2). Comprehensive validation results are in Supplementary B.

3.1. Personalized Biopsy Schedules

We utilized the fitted joint model to create upgrading-risk based personalized biopsy schedules. To this end, given a new patient's accumulated PSA

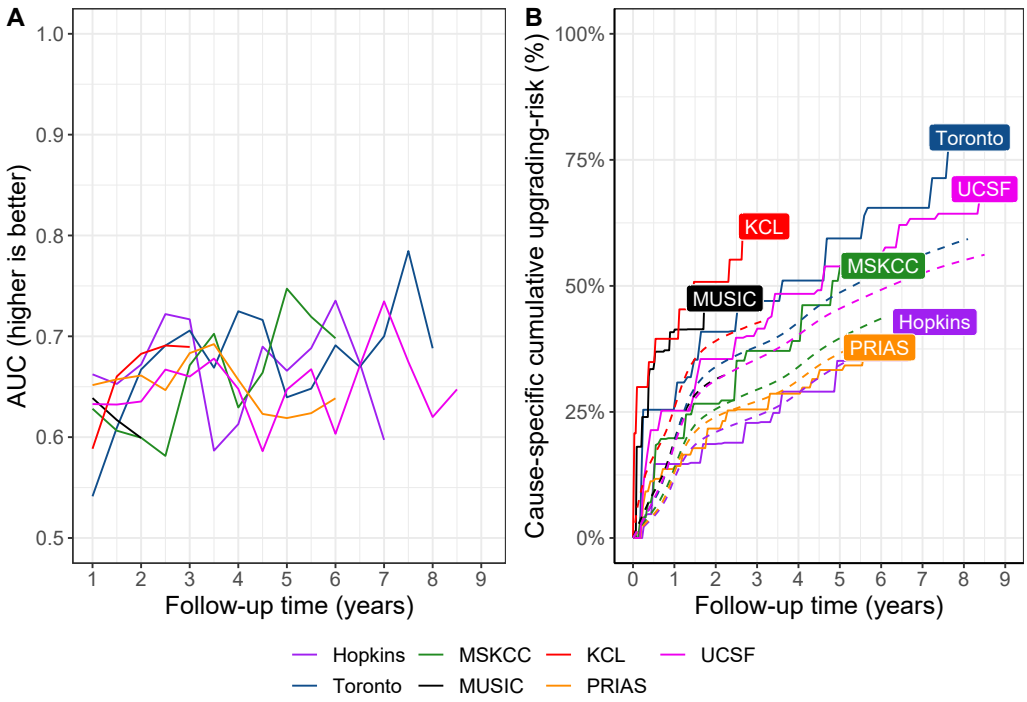

Figure 4: **Model Validation Results**. **Panel A**: time dependent area under the receiver operating characteristic curve or AUC (measure of discrimination). **Panel B**: calibration-at-large indicates model miscalibration. This is because solid lines depicting the non-parameteric estimate of the cause-specific cumulative upgrading-risk [24], and dashed lines showing the average cause-specific cumulative upgrading-risk obtained using the joint model fitted to the PRIAS dataset, are not overlapping. Same plot after recalibration is shown in Figure 6, Supplementary B. 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.

158 measurements (Panel A, Figure 3) and biopsy results, we first predicted his
159 cause-specific cumulative upgrading-risk at his current as well as future PSA
160 follow-up visits (Panel A, Figure 5). These PSA visits occur every six months
161 in PRIAS. Subsequently, we scheduled personalized biopsies on those future
162 follow-up visits of a patient, where his conditional cumulative upgrading-risk
163 was more than a certain threshold (Supplementary C), for example, 10%
164 risk. We maintained a minimum gap of one year between consecutive biop-
165 sies (PRIAS recommendation). Example personalized schedules based on 5%
166 and 10% risk thresholds are shown in Panel B, Figure 5, and in Figure 9–
167 11, Supplementary C. Both the risk predictions and resulting personalized
168 schedules were dynamic because they were updated as more follow-up data
169 became available over follow-up (Figure 5, Supplementary B).

170 The choice of the risk threshold in the personalized schedule dictates
171 the timing and the total number of biopsies, and the expected time delay
172 (Figure 1) in detecting upgrading. We estimated the time delay for both
173 personalized and fixed schedules (Panel C in Figure 5 and Figure 9–11, Sup-
174 plementary C). Since we estimated the time delay in a personalized manner as
175 well, patients/doctors can compare personalized schedules based on different
176 risk thresholds, with fixed schedules, before making a choice.

177 *3.2. Web-Application*

178 We implemented our model and personalized schedules in a user-friendly
179 web-application `https://emcbiostatistics.shinyapps.io/prias_biopsy_`
180 `recommender/`. Currently, the web-application supports PRIAS and the six
181 validation cohorts. Patient data can be entered manually and in Microsoft
182 Excel format. Predictions for upgrading-risk are available for a currently

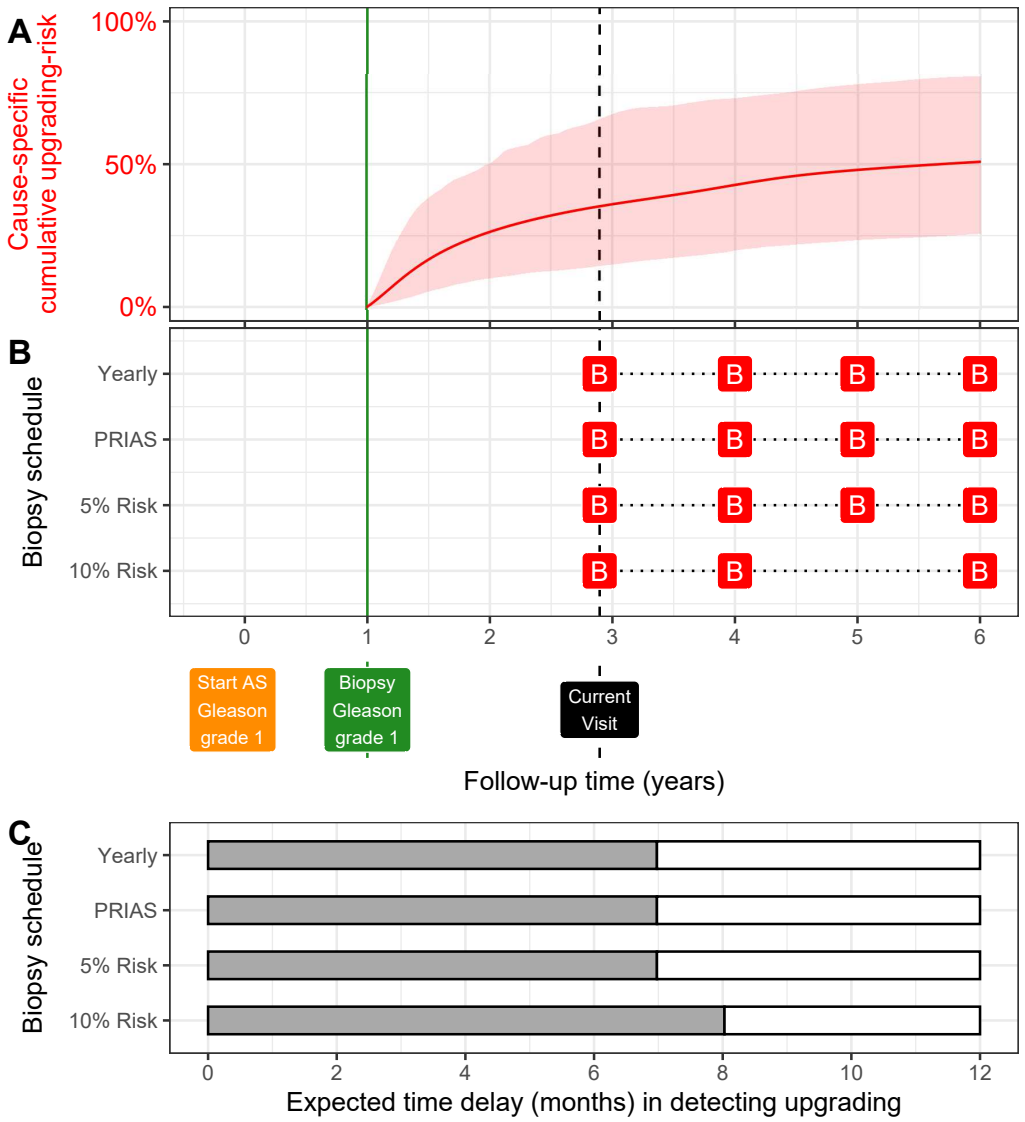

Figure 5: **Illustration of personalized and fixed schedules of biopsies**. Due to a lack of space, the PSA profile of this patient is shown in Figure 3. **Panel A:** Predicted cumulative upgrading-risk (95% cred ible interval shaded). **Panel B:** Personalized and fixed schedules of biopsies, with a red 'B' indicating a scheduled biopsy. The green vertical line at year 1 denotes the time of the latest negative biopsy. Black dashed line at year 3 denotes the time of the current visit. **Panel C:** Expected time delay in detecting upgrading (months) for different schedules. 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.

limited, cohort-specific, follow-up period (Table 7, Supplementary C). The web-application visualizes the timing of biopsies, and expected time delay in detecting upgrading, for personalized schedules based on 5%, 10%, and 15% risk threshold; annual biopsies; biennial biopsies; and PRIAS schedule.

4. Discussion

We successfully developed and externally validated a model for predicting upgrading-risk [3], and providing risk-based personalized biopsy decisions, in prostate cancer AS. Our work has four novel features over earlier risk calculators [15, 25]. First, our model was fitted to the world's largest AS dataset PRIAS and externally validated in the largest six cohorts of the Movember Foundation's GAP3 database [17]. Second, the model predicts a patient's current and future upgrading-risk in a dynamic and personalized manner. Third, we use the risks to make a personalized schedule, and also calculate expected time delay in detecting upgrading (less is beneficial) if that schedule is followed. 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 PRIAS and validated cohorts.

Our PRIAS based model is useful for a large number of patients from the PRIAS and the following validation cohorts: University of California San Francisco AS (UCSF), Johns Hopkins AS (Hopkins), Memorial Sloan Kettering Cancer Center AS, King's College London AS (KCL), and Michigan Urological Surgery Improvement Collaborative AS (MUSIC). The model had a moderate time-dependent AUC (0.55–0.75), a measure of discrimination,

in all validation cohorts. The moderate AUC can be explained by the fact that, unlike the standard AUC [22], the time-dependent AUC 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 it varied much more between cohorts than AUC. It was moderate in cohorts where the effect size for impact of PSA value and velocity on upgrading-risk was similar to that for PRIAS (e.g., Hopkins cohort). Otherwise, as in the case of KCL or MUSIC cohorts, the MAPE was large. In all cohorts, MAPE decreased rapidly after year one of follow-up. This may be explained by the fact at year one the validation data also consists of those patients who may have been misclassified incorrectly as Gleason grade group 1 at the time of inclusion in AS. Last, we required recalibration of our model's baseline hazard of upgrading, individually 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 every biopsy they receive in the first five years of AS. Given the non-compliance and burden of frequent biopsies [8], 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 [26] that personalized schedules may reduce up to a median of six biopsies compared to annual schedule, and a median of two biopsies compared to PRIAS schedule in slow/non-progressing AS patients, while maintaining almost the same time

delay in detection of progression as PRIAS schedule. Personalized schedules with different risk thresholds indeed have different performance. In this regard, to assist patients/doctors in choosing between various personalized, and fixed schedules, the web-application provides a patient-specific estimate of expected time delay in detecting upgrading, for following both personalized and fixed schedules. We hope that this will objectively address patient apprehensions regarding adverse outcomes in AS.

This work has certain limitations. Predictions for upgrading-risk and personalized schedules are available only for a currently limited, cohort-specific, follow-up period (Table 7, Supplementary C). This problem can be mitigated by refitting the model with new follow-up data in the future. Along with PSA, in some cohorts recently, MRI is also used to explore the possibility of targeting visible tumor by biopsy. However, the utility of MRI can only be determined with more follow-up data in the future. Subsequently, MRI data can also be added as a predictor in our model. Decisions based on information combined from both MRI and PSA can potentially improve the currently developed model. We scheduled biopsies using cause-specific cumulative upgrading-risk. Accounting for competing events, such as treatment based on the number of positive biopsy cores, may lead to improved personalized biopsy decisions. Although, in this work, we did not consider such additional triggers for treatment because, unlike upgrading, they differ between cohorts [5]. Upgrading is susceptible to inter-observer variation too. Models which account for this variation [15, 27] will be interesting to investigate further. However, the methodology for personalized scheduling, and for comparison of various schedules need not change.

5. Conclusions

We successfully developed and validated a model for predicting upgrading-risk, and providing risk-based personalized biopsy decisions, in prostate cancer AS. 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 the baseline hazard of upgrading-risk is advised before using the model in cohorts other than the PRIAS cohort.

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

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 Rotterdam Prostate Cancer Research Foundation Rotterdam (SWOP). 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.

Appendix A. Members of The Movember Foundations 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 Insti-

tute, 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 & Guys and St Thomas NHS Foundation Trust, London, UK), Prokar Dasgupta (Guys 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), 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, Universit 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, Gteborg, Sweden), Jose Rubio-Briones (Instituto Valenciano de Oncologa,

Valencia, Spain), Anders Bjartell (Skne University Hospital, Malm, 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 & Guys and St Thomas NHS Foundation Trust, London, UK), Brian Denton (University of Michigan, Michigan, USA), Andrew Hayen (Uni-

versity 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 & Guys 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), Heln Ahlgren (Sahlgrenska University Hospital, Gteborg, Sweden), Juanma Mascars (Instituto Valenciano de Oncologa, Valencia, Spain), Annica Lfgren (Skne University Hospital, Malm, Sweden), Kurt Lehmann (Kantonsspital Baden, Baden,

Switzerland), Catherine Han Lin (Monash University and Epworth Health-Care, 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 (Skne University Hospital, Malm, 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, 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. 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, Gnanapragasam VJ, Villers A, Rannikko A, Valdagni R, et al. Expert consensus document: semantics in active surveillance for men with localized prostate cancerresults 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, 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.

5. 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.

6. 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.

7. Loeb S, Vellekoop A, Ahmed HU, Catto J, Emberton M, Nam R, Rosario DJ, Scattoni V, Lotan Y. Systematic review of complications of prostate biopsy. *European urology* 2013;64(6):876–92.

8. 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)

425 protocol and disease reclassification in noncompliers. *European Urology*
426 2015;68(5):814–21.

427 9. Inoue LY, Lin DW, Newcomb LF, Leonardson AS, Ankerst D, Gulati R,
428 Carter HB, Trock BJ, Carroll PR, Cooperberg MR, et al. Comparative
429 analysis of biopsy upgrading in four prostate cancer active surveillance
430 cohorts. *Annals of internal medicine* 2018;168(1):1–9.

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

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

439 12. Partin AW, Yoo J, Carter HB, Pearson JD, Chan DW, Epstein JI, Walsh
440 PC. The use of prostate specific antigen, clinical stage and gleason score
441 to predict pathological stage in men with localized prostate cancer. *The*
442 *Journal of urology* 1993;150(1):110–4.

443 13. Makarov DV, Trock BJ, Humphreys EB, Mangold LA, Walsh PC, Ep-
444 stein JI, Partin AW. Updated nomogram to predict pathologic stage of
445 prostate cancer given prostate-specific antigen level, clinical stage, and
446 biopsy gleason score (partin tables) based on cases from 2000 to 2005.
447 *Urology* 2007;69(6):1095–101.

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

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

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

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

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

463 19. Cooperberg MR, Brooks JD, Faino AV, Newcomb LF, Kearns JT, Car-
464 roll PR, Dash A, Etzioni R, Fabrizio MD, Gleave ME, et al. Refined
465 analysis of prostate-specific antigen kinetics to predict prostate cancer
466 active surveillance outcomes. *European urology* 2018;74(2):211–7.

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

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

22. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology (Cambridge, Mass)* 2010;21(1):128.

23. 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.

24. 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.

25. Ankerst DP, Xia J, Thompson Jr IM, Hoefler J, Newcomb LF, Brooks JD, Carroll PR, Ellis WJ, Gleave ME, Lance RS, 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.

26. 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.

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