

Personalized Biopsies in Prostate Cancer Active Surveillance[☆]

Anirudh Tomer, MSc^{a,*}, Daan Nieboer, MSc^b, Monique J. Roobol, PhD^c, Anders Bjartell, 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 patients enrolled in active surveillance (AS) undergo repeat biopsies. When biopsy Gleason grade ≥ 2 (reclassification), treatment is commonly advised. Many patients never experience reclassification, yet undergo biopsies frequently.

[☆]Word Count Abstract: 300; Word Count Text: 2509

*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, PhD), e.w.steyerberg@lumc.nl (Ewout W. Steyerberg, PhD), d.rizopoulos@erasmusmc.nl (Dimitris Rizopoulos, PhD)

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

Design, Setting, and Participants: World’s largest AS study, PRIAS; 7813 patients, 1134 experienced reclassification; prostate-specific antigen (PSA) and repeat biopsy data available.

Outcome Measurements, and Statistical Analysis: Bayesian joint model based on accumulated clinical data used to predict patient-specific risk of reclassification. Predicted risks were utilized to schedule personalized biopsies. Personalized and fixed schedules were compared by number of biopsies, and model estimated expected time delay in detection of reclassification for each schedule. Model was validated externally in largest five AS cohorts from GAP3 database. Methodology implemented in a web-application.

Results and Limitations: Rate of reclassification in PRIAS was 50% in 10 years of follow-up. PSA velocity was a stronger predictor of reclassification (Hazard Ratio: 2.47, 95%CI: 1.93–2.99), than PSA value (Hazard Ratio: 0.99, 95%CI: 0.89–1.11). Internal validation: Time varying area under ROC curve for reclassification prediction between 0.62 and 0.69, and prediction error between 0.23 and 0.37. External validation: Results similar to internal validation only for Toronto, Memorial Sloan Kettering, and Johns Hopkins AS cohorts.

Conclusions: We developed personalized biopsy schedules as alternative

to fixed schedules for AS patients. To assist patients/doctors in biopsy decisions, we provided them expected time delay in detection of reclassification, for both personalized and fixed schedules. We implemented our methodology in a web-application.

Patient Summary: Personalized biopsy schedules are a novel alternative to fixed biopsy schedules (e.g., yearly biopsies). They rely on patient-specific risk of reclassification. 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

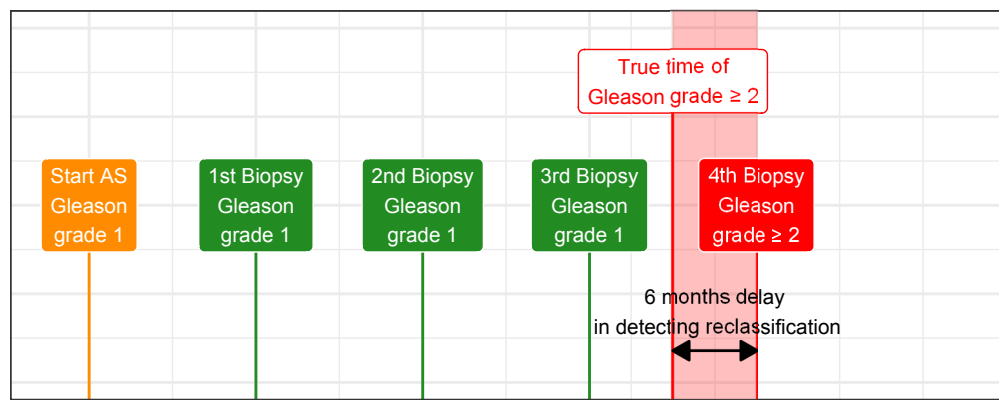
1. Introduction

Patients with low- and very low-risk screening-detected localized prostate cancer are usually advised 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 [2]. When it increases from grade 1 (Gleason 3+3) to grade 2 (Gleason 3+4) or higher, called *reclassification*, patients are commonly advised curative treatment [3].

Biopsies are conducted periodically. Consequently, reclassification is always detected with a time delay (Figure 1). For detecting reclassification timely, many AS programs schedule fixed and frequent biopsies (e.g., annually) for all patients [4, 5]. However, this also leads to many unnecessary biopsies in slow/non-progressing patients. Biopsies are invasive, painful and prone to medical complications. Thus, biopsy burden and patient non-compliance to frequent biopsies [6] has raised concerns regarding the optimal biopsy schedule [7, 8]. To this end, infrequent schedules such as biennial biopsies have been proposed as an alternative [7, 9]. Although, biennial biopsies may still lead to five unnecessary biopsies over ten years (current study period of large AS programs) for slow/non-progressing patients. A promising alternative to fixed and frequent biopsies is personalized biopsy schedules based on the patient-specific risk of reclassification (Figure 2).

The first challenge in developing personalized biopsy schedules is consolidating observed patient data (e.g., PSA, previous biopsy results) into risk estimates for reclassification. Existing calculators for risk of reclassification

A Biopsy every year



B Biopsy every 2 years

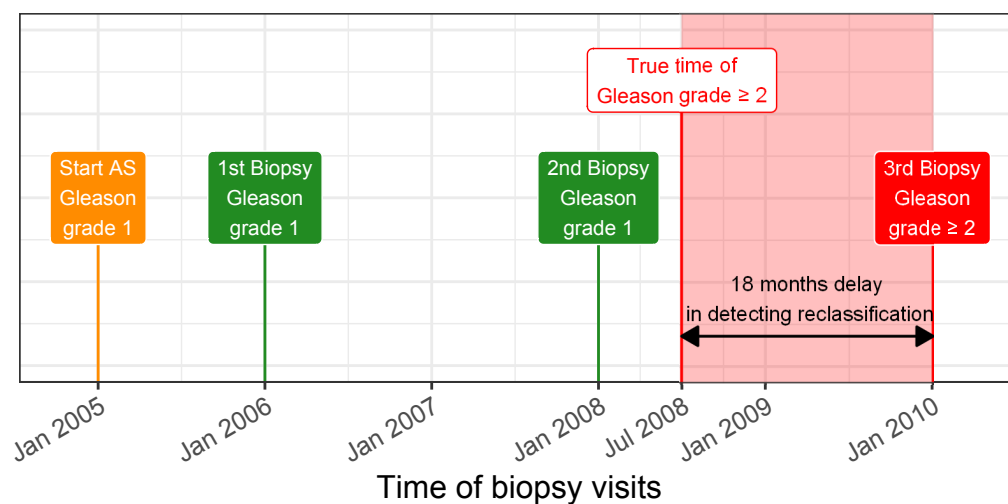


Figure 1: **Trade-off between the number of biopsies and time delay in detecting reclassification (Increase in Gleason grade from 1 to ≥ 2):** The true time of reclassification for the patient in this figure is July 2008. When biopsies are scheduled annually (**Panel A**), reclassification is detected 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**) reclassification 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 reclassification is observed as an interval. For example, Jan 2008–Jan 2009 in **Panel A** and Jan 2008–Jan 2010 in **Panel B**.

A Should a biopsy be conducted at current visit?

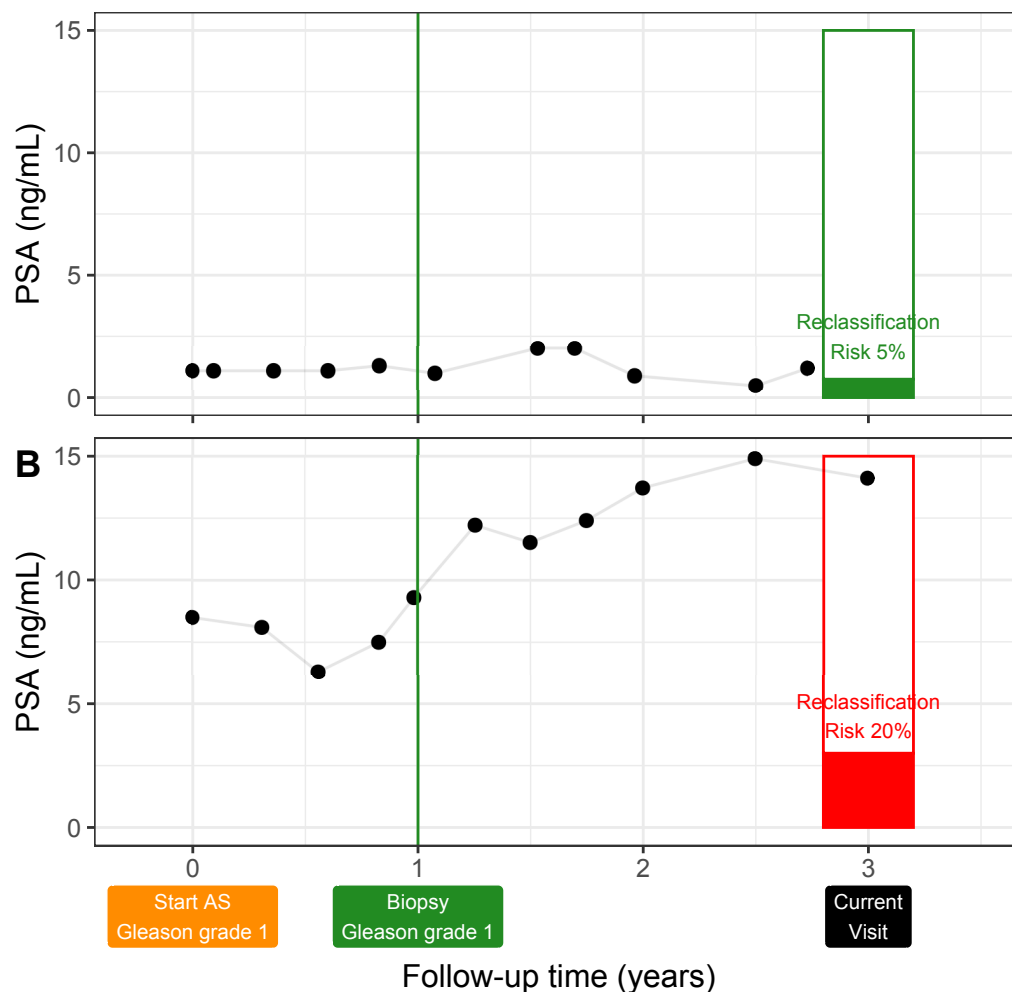


Figure 2: **Motivation for personalized risk-based decisions of biopsy:** Patients A (**Panel A**) and B (**Panel B**) had their latest biopsy at year one of follow-up (green vertical line). Patient A's prostate-specific antigen (PSA) profile remained stable until his current visit at year three, whereas patient B's profile has shown a rise. Consequently, patient B's estimated cumulative risk of reclassification at the current visit (year three) 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.*

[10, 11] ignore longitudinally measured PSA in AS, and utilize only the latest PSA measurement of a patient. In contrast, we intend to utilize complete longitudinal PSA data, 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 [12, 13, 14]. While a joint model predicts risk of reclassification in a personalized manner, a subsequent challenge is translating risks into clinical decisions. For example, a 10% risk of reclassification can be perceived high/low depending upon patient age. Patients may also weigh these risks with the potential *consequences* of another biopsy. Two relevant *consequences* of biopsies (Figure 1) are the timing and total number of biopsies (burden), and the time delay in detecting reclassification (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 was to assist patients and doctors in making better decisions of biopsies than fixed and frequent biopsies. For this purpose, we developed a web-application that provides patients their current and future risk of reclassification. It also suggests them risk-based personalized schedules of biopsies. For each biopsy schedule, be it fixed or personalized, the web-application provides expected *consequences* of following it. Thus, patients can compare schedules before making a decision. The web-application uses a prediction joint model fitted to the world’s largest AS dataset, PRIAS [3]. We externally validated this model in five largest AS cohorts from the GAP3 database [15], so that the tool can be used by a large number of patients worldwide.

50 2. Patients and Methods

51 2.1. Study Cohort

52 Prostate Cancer International Active Surveillance (PRIAS) is an ongoing
 53 (December 2006 – to date) prospective cohort study of men with low-
 54 and very-low risk prostate cancer diagnoses [3]. More than 100 medical
 55 centers from 17 countries contributed to PRIAS, using a common protocol
 56 (www.prias-project.org). PSA was measured every three months for
 57 the first two years of follow-up and every six months thereafter. Biopsies
 58 were scheduled at year one, four, seven, and ten. Additional yearly biopsies
 59 were scheduled when PSA doubling time fell between zero and ten years.

60 We selected all 7813 patients who had Gleason grade 1 [2] at the time of
 61 inclusion in PRIAS (Table 1). Our primary event of interest is increase in this
 62 Gleason grade to grade 2 or higher upon repeat biopsy, called *reclassification*
 63 (1134 patients). Reclassification is a trigger for treatment advice in PRIAS.
 64 Although, 2250 patients were also advised treatment on the basis of their
 65 PSA, or number of biopsy cores with cancer, or anxiety/other reasons. We
 66 focused solely on reclassification because of its strong association with cancer-
 67 related outcomes.

Table 1: **Summary of the PRIAS dataset.** The primary event of interest is reclassification, that is, increase in Gleason grade from grade 1 to 2 or higher. IQR: interquartile range, PSA: prostate-specific antigen.

Characteristic	Value
Total patients	7813
Reclassification (primary event)	1134
Treatment	2250
Watchful waiting	334
Loss to follow-up	250
Death (unrelated to prostate cancer)	95
Death (related to prostate cancer)	2
Median age at diagnosis (years)	66 (IQR: 61–71)
Median follow-up period 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)

68 2.2. Statistical Model

69 To create personalized biopsy schedules based on patient-specific risk of
 70 reclassification, we required a risk prediction model. Available data was pa-
 71 tient age at inclusion in AS, PSA measurements over follow-up, the timing
 72 of repeat biopsies and corresponding Gleason grades, and observed time of
 73 reclassification. Analysis of this data required modeling the correlation be-
 74 tween the PSA measurements of the same patient, the correlation between
 75 the Gleason grade and the PSA profile, and to account for PSA measure-
 76 ments missing after a patient experienced reclassification. To this end, a
 77 commonly used model is the joint model for time-to-event and longitudinal
 78 data [12, 13, 14].

79 Our joint model consisted of two sub-models. First, a linear mixed model
 80 [16] for longitudinally measured PSA (log-transformed). Second, a relative-
 81 risk model (similar to Cox model) for obtaining the risk of reclassification.
 82 In the model for PSA, we fitted a curve to PSA measurements (Panel A,
 83 Figure 3). From each patient’s fitted PSA profile we extracted the instan-
 84 taneous PSA velocity. This velocity varies over time (Panel B, Figure 3).
 85 Consequently, it is more precise than the currently used definition of PSA
 86 velocity [17]. We connected the two sub-models by using the fitted PSA and
 87 instantaneous velocity as predictors in the sub-model for risk of reclassifica-
 88 tion (Panel C, Figure 3). Patient age was included in both sub-models. The
 89 parameters of the two sub-models were estimated jointly (Appendix A) using
 90 the R package **JMbayes** [18].

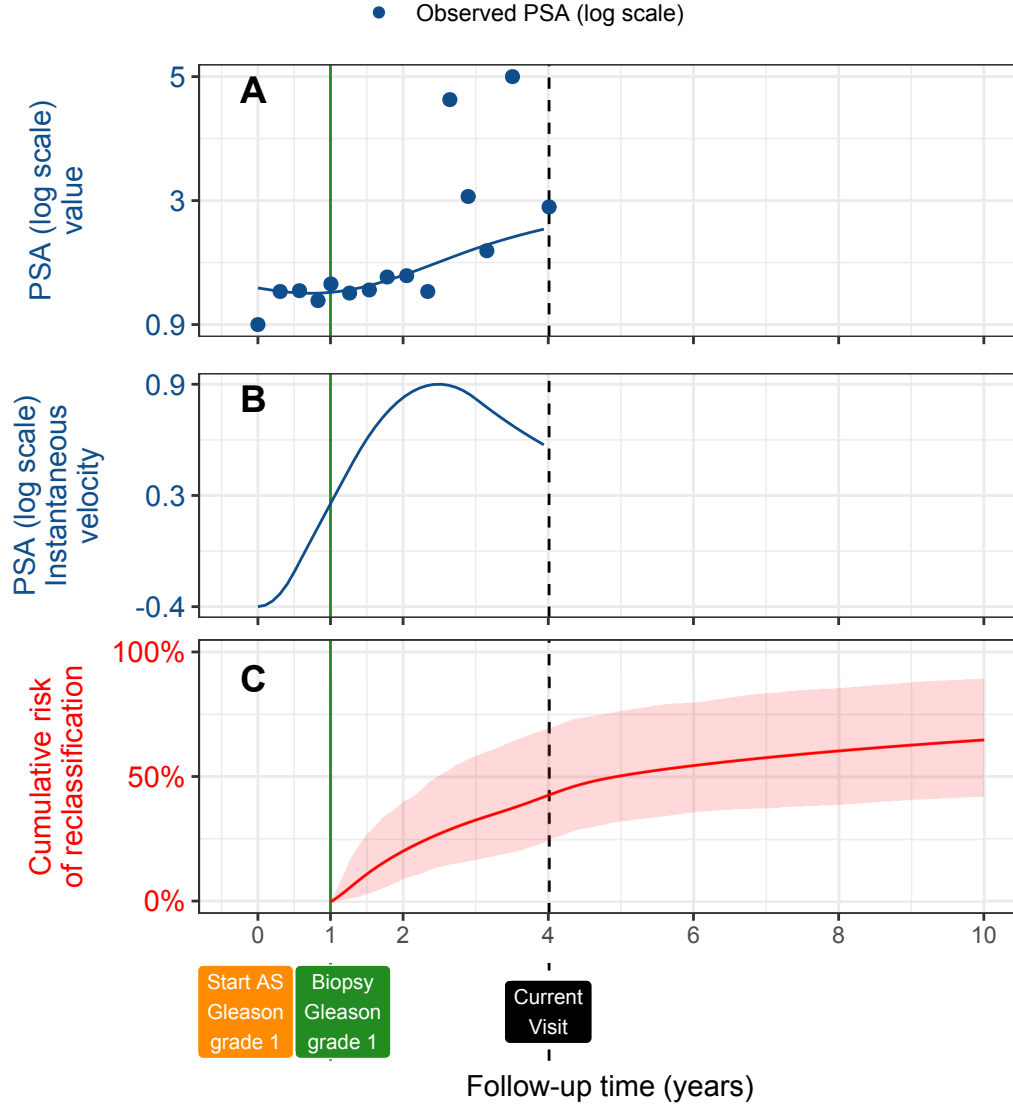


Figure 3: **Illustration of the joint model on a real PRIAS dataset patient.** **Panel A:** Observed (blue dots) and fitted PSA (solid blue line) measurements, log-transformed. **Panel B:** Estimated instantaneous velocity of PSA (log-transformed). **Panel C:** Predicted cumulative-risk of reclassification (95% credible interval shaded). Reclassification is defined as increase in Gleason grade [2] from grade 1 to 2 or higher. This risk of reclassification is available starting from the time of the latest negative biopsy (vertical green line at year 1 of follow-up). Joint model estimated it by combining the fitted PSA value and velocity (both on log scale of PSA) and time of latest negative biopsy. Black dashed line at year 4 denotes time of current visit.

91 2.3. *Predicting Risk of Reclassification*

92 The key component in personalized schedules is the cumulative-risk of
 93 reclassification. Given, a patient’s PSA measurements and previous biopsy
 94 results, our joint model predicts the cumulative-risk of reclassification at his
 95 current as well as future visit times (e.g., Panel C, Figure 3). The cumulative-
 96 risk profile is also updated as more patient data becomes available over follow-
 97 up (Appendix B).

98 We validated our model internally as well as externally. Internal valida-
 99 tion was done using PRIAS dataset. External validation was done in largest
 100 five AS cohorts from the GAP3 database [15], namely University of Toronto
 101 AS (Toronto-AS), Johns Hopkins AS (JH-AS), Memorial Sloan Kettering
 102 Cancer Center AS (MSKCC-AS), King’s College London AS (KCL-AS), and
 103 Michigan Urological Surgery Improvement Collaborative AS (MUSIC-AS).
 104 We assessed our model’s discrimination via the area under the receiver oper-
 105 ating characteristic curve or AUC [19]. For calibration we utilized calibration
 106 plots for visual assessment, and root mean squared prediction error [19] for
 107 quantitative assessment. We also recalibrated our model’s baseline hazard of
 108 reclassification for each of the five external cohorts (Appendix B).

109 2.4. *Personalized Biopsy Schedules*

110 We created personalized schedules of biopsies upon a patient’s visit for
 111 testing PSA (every 6 months in PRIAS). Specifically, if a patient’s cumulative-
 112 risk of reclassification at his current PSA visit was more than a certain thresh-
 113 old (e.g., 10%) we scheduled an immediate biopsy. Our model predicted his
 114 cumulative-risk of reclassification at his future PSA follow-up visits as well.
 115 Thus, by applying the same risk threshold rule at each future follow-up visit,

we scheduled future biopsies (Appendix C). We kept a minimum gap of one year between consecutive biopsies (PRIAS recommendation). Example personalized schedules based on different risk thresholds are shown in Figure 4.

The choice of the risk threshold in the personalized schedule dictates the *consequences* of following that schedule. *Consequences* are the timing and the total number of biopsies, and the expected delay in detecting reclassification. Larger risk thresholds will schedule infrequent personalized biopsies. However, the consequent delay in detecting reclassification may also be longer. Our model also estimated these *consequences* in a personalized manner given any schedule of biopsies. Thus patients can compare fixed schedules with various risk threshold based personalized schedules before making a choice.

2.5. Web-Application

We implemented our methodology in a web-application https://emcbiostatistics.shinyapps.io/prias_biopsy_recommender/. It utilizes our joint model fitted to the PRIAS dataset. Currently, the web-application supports PRIAS and the five external cohorts in which we validated our model. Patient data can be entered manually or can be uploaded in Microsoft Excel format. Predictions for risk of reclassification are shown for the first ten years of follow-up only (current study period of PRIAS). The web-application shows the *consequences* of following these schedules: personalized schedules based on 5%, 10%, and 15% risk threshold; annual biopsies; biennial biopsies; and PRIAS schedule.

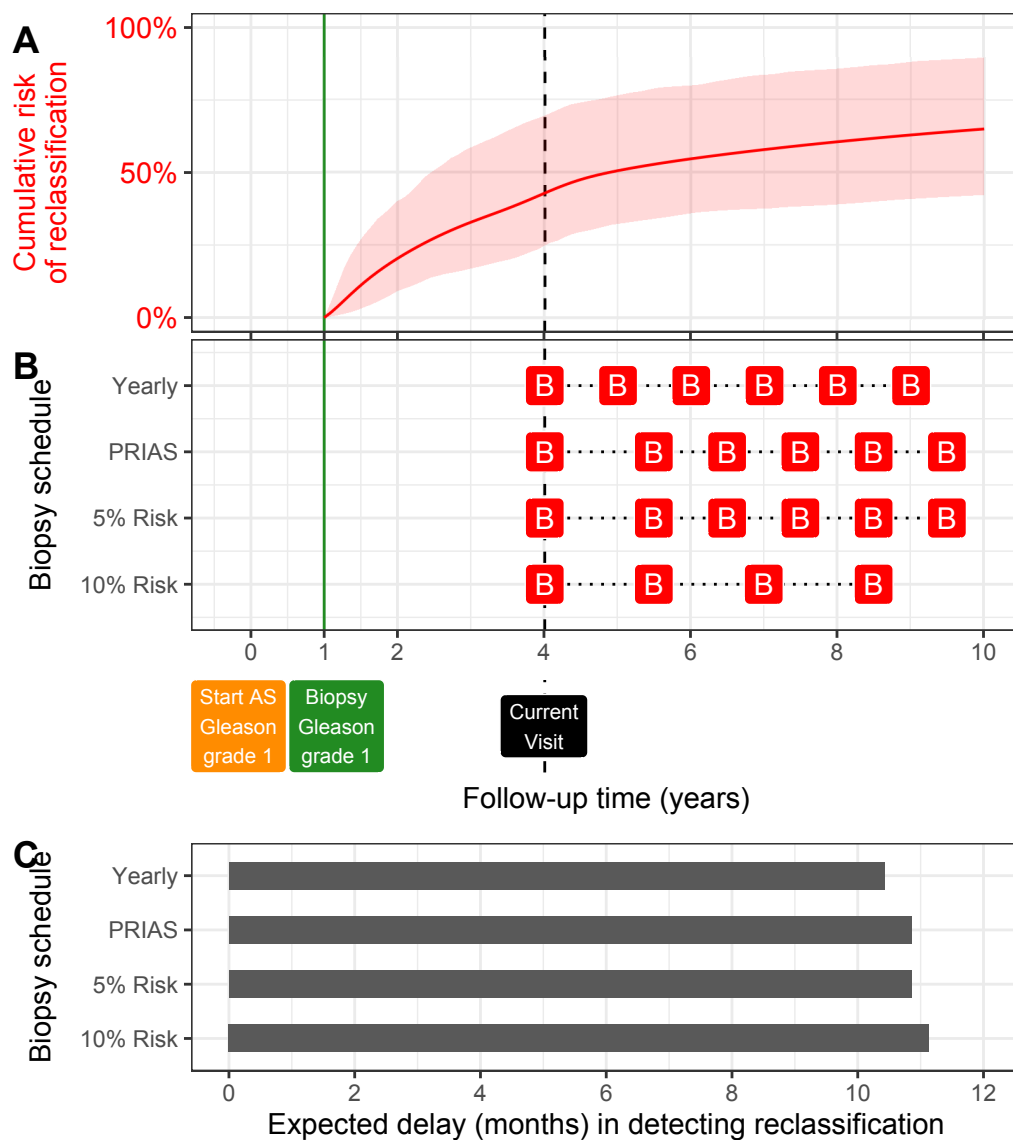


Figure 4: **Illustration of personalized and fixed schedules of biopsies.** The PSA profile of this patient is shown in Figure 3. **Panel A:** Predicted cumulative-risk of reclassification (95% credible interval shaded). **Panel B:** Personalized and fixed schedules of biopsies, with a red 'B' indicating a scheduled biopsy. **Panel C:** Expected delay in detecting reclassification for different schedules. Green vertical line at year 1 denotes time of latest negative biopsy. Black dashed line at year 4 denotes time of current visit.

138 3. Results

139 3.1. Joint Model Results

140 **Overall cumulative-risk:** In PRIAS, the probability of experiencing
 141 reclassification within the first five and ten years was 33% and 42%, respec-
 142 tively (cumulative-risk plot in Appendix A). That is, many patients do not
 143 require any biopsy in the first ten years.

144 **Effect of age at inclusion in AS:** For every ten year increase in patient
 145 age, the adjusted hazard ratio of reclassification is 1.45 (95%CI: 1.30–1.63).

146 **Effect of PSA:** When PSA value (log scale) increases from 2.36 to
 147 3.07 (25-th and 75-th percentile of fitted PSA), the adjusted hazard ratio
 148 of reclassification is 0.99 (95%CI: 0.89–1.11). When estimated instantaneous
 149 PSA (log scale) velocity increases from -0.09 to 0.31 (25-th and 75-th per-
 150 centile of estimated velocity), the adjusted hazard ratio of reclassification is
 151 2.47 (95%CI: 1.93–2.99). Hence, instantaneous PSA (log scale) velocity is a
 152 stronger predictor for reclassification than PSA value (log scale). Detailed
 153 parameter estimates are in Appendix A.4.

154 3.2. Validation Results

155 **Discrimination (AUC):** Since AS studies are longitudinal in nature,
 156 the area under the receiver operating characteristic curve (AUC) changes
 157 over time. This time-varying AUC is shown in Panel A, Figure 5. The AUC
 158 is only calculated until year five (95-percentile of observed reclassification
 159 times in PRIAS) of follow-up.

160 **Calibration:** The calibration plots in Panel B, Figure 5 indicate that
 161 the model predictions are well calibrated for all five external cohorts.

Detailed results for model validation are available in Appendix B.

3.3. *Personalized Schedule Results*

Various personalized and fixed biopsy schedules for demo patients are shown in Figure 4

In addition, we scheduled biopsies only for the first ten years follow-up because of limited follow-up period of the training dataset PRIAS. A compulsory biopsy was done scheduled year ten of follow-up in all schedules for meaningful comparison of their expected delays in detection of GS7. and Appendix C's Figure 6, 7, 8 and 9. The biopsies denoted by 'B' show that personalized schedules schedule fewer biopsies than fixed schedules. At the same time the expected time delay in detection of GS7 is less than an year for personalized schedules. We have implemented this approach in a web-application (https://emcbiostatistics.shinyapps.io/prias_biopsy_recommender/, and Appendix D) for practical use.

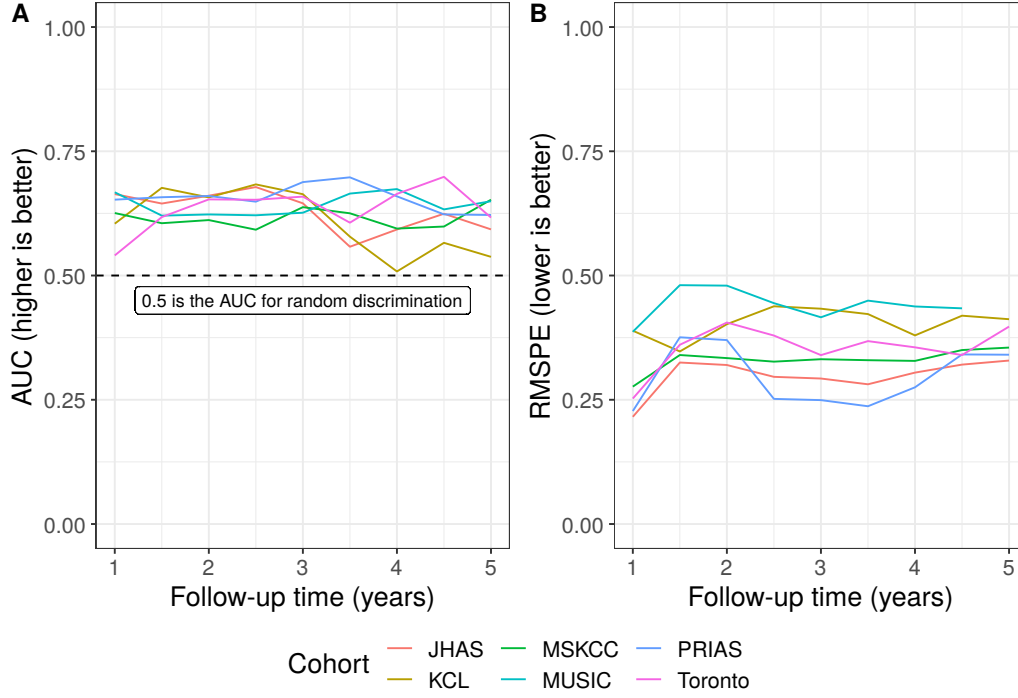


Figure 5: **Validation of predictions of Gleason ≥ 7 (GS7).** In **Panel A** we can see that the time dependent area under the receiver operating characteristic curve or AUC (measure of discrimination) is above 0.5 in PRIAS (internal validation), and in Toronto, JHAS, MSKCC, KCL, and MUSIC AS cohorts (external validation). In **Panel B** we can see that the time dependent root mean squared prediction error or RMSPE (measure of calibration) is similar for PRIAS, and JHAS and Toronto cohorts. The bootstrapped 95% confidence interval for these estimates are presented in Table 6 to Table 11 of Appendix B. Full names of Cohorts are *PRIAS*: Prostate Cancer International Active Surveillance, *Toronto*: University of Toronto Active Surveillance, *JHAS*: 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.

176 4. Discussion

177 We developed a novel methodology and model for personalized scheduling
 178 of biopsies in prostate cancer active surveillance (AS) patients. Personalized
 179 schedules utilize patient-specific risks of reclassification. Reclassification is
 180 defined as increase in biopsy Gleason grade [2] from grade 1 to 2 or higher.
 181 Calculators for risk of reclassification are not new [13, 20]. However, our work
 182 has four novel features. First, we personalized the risk of reclassification and
 183 used it to schedule biopsies in a personalized manner. Second, we developed
 184 a methodology that can calculate expected delay in detection of reclassifica-
 185 tion (less is beneficial) in a personalized manner, given any biopsy schedule.
 186 Thus patients and doctors can compare schedules before making a choice.
 187 Third, we implemented our methodology in a web-application [https://](https://emcbiostatistics.shinyapps.io/prias_biopsy_recommender/)
 188 emcbiostatistics.shinyapps.io/prias_biopsy_recommender/. Fourth,
 189 we validated our model in largest five AS cohorts from GAP3 database [15],
 190 and hence the web-application can be used by a large number of patients
 191 worldwide.

192 Currently biopsies are scheduled either in a fixed and frequent manner
 193 (e.g., annual biopsies), or PSA value/velocity/doubling-time is used as trig-
 194 ger for biopsies. These approaches have been criticized previously [17, 6].
 195 However, earlier approaches do not exploit PSA data fully and correctly.
 196 Specifically, they assume that PSA is observed without measurement error,
 197 and/or latest PSA is enough to decide biopsies, and/or PSA changes over
 198 time in a linearly. In contrast, our joint model builds a patient-specific profile
 199 of PSA using all PSA measurements. It also allows PSA and its velocity to
 200 change over time non-linearly. Subsequently, it consolidates these finer PSA

201 features, previous biopsy results, and baseline characteristics of a patient, to
202 yield a single personalized estimate for risk of reclassification. Furthermore,
203 the model updates this risk as more patient data is gathered over follow-up.
204 This is a more holistic approach.

205 A holistic model like ours allows incorporating newer biomarkers and mag-
206 netic resonance imaging (MRI) data. Such information is currently sparsely
207 available in PRIAS dataset. However, MRI data can be included as predic-
208 tor in our model in future. Decisions based on combined information from
209 multiple sources can yield better results than decisions based on MRI or PSA
210 alone.

211 Our model is not only useful for PRIAS patients, but also for a large num-
212 ber of patients from other cohorts. This is because we have recalibrated and
213 externally validated it in largest five AS cohorts from the GAP3 database
214 [15]. These are University of Toronto AS (Toronto-AS), Johns Hopkins AS
215 (JH-AS), Memorial Sloan Kettering Cancer Center AS (MSKCC-AS), King's
216 College London AS (KCL-AS), and Michigan Urological Surgery Improve-
217 ment Collaborative AS (MUSIC-AS). Extending our model and methodol-
218 ogy in smaller cohorts requires only recalibrating our model's baseline risk
219 of reclassification.

220 Our work has important clinical implications. The median survival time
221 for reclassification is more than ten years in PRIAS, and in some other co-
222 horts (Figure 5). That is, more than 50% of AS patients may not require
223 any biopsy during the first ten years of follow-up. Given the concerns about
224 non-compliance and burden of frequent biopsies [6], the availability of our
225 web-application may encourage patients and doctors to consider personal-

226 ized schedules instead. For both personalized and fixed schedules, the web-
227 application also provides an estimate of delay in detection of reclassification.
228 We hope this will address patient apprehensions regarding adverse outcomes
229 in AS, in a more objective manner.

230 Our work has certain limitations. The proposed model is valid only for
231 the first ten years of follow-up in PRIAS, whereas reclassification may oc-
232 cur much later in many patients. In addition, our model predictions were
233 less accurate in later follow-up period due to lack of training data. These
234 problems can be mitigated by refitting the model with new follow-up data
235 in future. Although, we focused only on reclassification, an increase in num-
236 ber of positive biopsy cores can also act as a trigger for treatment. We did
237 not consider such additional triggers because they differ between cohorts [4].
238 Whereas, reclassification is a commonly used criteria. Reclassification is sus-
239 ceptible to inter-observer variation. Models which account for this variation
240 [13, 21] will be interesting to investigate further. However, the methodology
241 for personalized scheduling, and for comparison of various schedules need not
242 change.

243 5. Conclusions

244 We developed a novel methodology and model for personalized scheduling
245 of biopsies in prostate cancer active surveillance (AS) patients. Personalized
246 schedules utilize a patient's risk of reclassification to decide biopsies. They
247 are a promising alternative to fixed and frequent biopsy schedules. We have
248 made them available in a web-application [https://emcbiostatistics.shinyapps.](https://emcbiostatistics.shinyapps.io/prias_biopsy_recommender/)
249 [io/prias_biopsy_recommender/](https://emcbiostatistics.shinyapps.io/prias_biopsy_recommender/) for world's six largest AS cohorts. The
250 web-application also assists patients and doctors in choosing a suitable sched-
251 ule. This is because, it provides them estimates of personalized burden (time
252 and total biopsies), and personalized benefit (time delay in detection of re-
253 classification; lesser is beneficial), for both personalized and currently used
254 schedules.

255 Acknowledgments

256 This work was supported by the Movember Foundation. The funder did
 257 not play any role in the study design, collection, analysis or interpretation of
 258 data, or in the drafting of this paper. The first and last Authors would like
 259 to acknowledge support by Nederlandse Organisatie voor Wetenschappelijk
 260 Onderzoek (the national research council of the Netherlands) VIDI grant nr.
 261 016.146.301, and Erasmus University Medical Center funding.

- 262 1. Briganti A, Fossati N, Catto JW, Cornford P, Montorsi F, Mottet N,
 263 Wirth M, Van Poppel H. Active surveillance for low-risk prostate cancer:
 264 the European Association of Urology position in 2018. *European urology*
 265 2018;74(3):357–68.
- 266 2. Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA.
 267 The 2014 international society of urological pathology (isup) consensus
 268 conference on gleason grading of prostatic carcinoma. *The American*
 269 *journal of surgical pathology* 2016;40(2):244–52.
- 270 3. Bul M, Zhu X, Valdagni R, Pickles T, Kakehi Y, Rannikko A, Bjartell A,
 271 Van Der Schoot DK, Cornel EB, Conti GN, et al. Active surveillance for
 272 low-risk prostate cancer worldwide: the prias study. *European urology*
 273 2013;63(4):597–603.
- 274 4. Nieboer D, Tomer A, Rizopoulos D, Roobol MJ, Steyerberg EW. Active
 275 surveillance: a review of risk-based, dynamic monitoring. *Translational*
 276 *andrology and urology* 2018;7(1):106–15.

- 277 5. Loeb S, Carter HB, Schwartz M, Fagerlin A, Braithwaite RS, Lepor H.
 278 Heterogeneity in active surveillance protocols worldwide. *Reviews in*
 279 *urology* 2014;16(4):202–3.
- 280 6. Bokhorst LP, Alberts AR, Rannikko A, Valdagni R, Pickles T, Kakehi Y,
 281 Bangma CH, Roobol MJ, PRIAS study group . Compliance rates with
 282 the Prostate Cancer Research International Active Surveillance (PRIAS)
 283 protocol and disease reclassification in noncompliers. *European Urology*
 284 2015;68(5):814–21.
- 285 7. Inoue LY, Lin DW, Newcomb LF, Leonardson AS, Ankerst D, Gulati R,
 286 Carter HB, Trock BJ, Carroll PR, Cooperberg MR, et al. Comparative
 287 analysis of biopsy upgrading in four prostate cancer active surveillance
 288 cohorts. *Annals of internal medicine* 2018;168(1):1–9.
- 289 8. Bratt O, Carlsson S, Holmberg E, Holmberg L, Johansson E, Josefsson
 290 A, Nilsson A, Nyberg M, Robinsson D, Sandberg J, et al. The study of
 291 active monitoring in sweden (sams): a randomized study comparing two
 292 different follow-up schedules for active surveillance of low-risk prostate
 293 cancer. *Scandinavian journal of urology* 2013;47(5):347–55.
- 294 9. de Carvalho TM, Heijnsdijk EA, de Koning HJ. Estimating the risks
 295 and benefits of active surveillance protocols for prostate cancer: a mi-
 296 crosimulation study. *BJU international* 2017;119(4):560–6.
- 297 10. Partin AW, Yoo J, Carter HB, Pearson JD, Chan DW, Epstein JI, Walsh
 298 PC. The use of prostate specific antigen, clinical stage and gleason score

- 299 to predict pathological stage in men with localized prostate cancer. *The*
300 *Journal of urology* 1993;150(1):110–4.
- 301 11. Makarov DV, Trock BJ, Humphreys EB, Mangold LA, Walsh PC, Ep-
302 stein JI, Partin AW. Updated nomogram to predict pathologic stage of
303 prostate cancer given prostate-specific antigen level, clinical stage, and
304 biopsy gleason score (partin tables) based on cases from 2000 to 2005.
305 *Urology* 2007;69(6):1095–101.
- 306 12. Tomer A, Nieboer D, Roobol MJ, Steyerberg EW, Rizopoulos D. Per-
307 sonalized schedules for surveillance of low-risk prostate cancer patients.
308 *Biometrics* 2019;75(1):153–62.
- 309 13. Coley RY, Zeger SL, Mamawala M, Pienta KJ, Carter HB. Prediction
310 of the pathologic gleason score to inform a personalized management
311 program for prostate cancer. *European urology* 2017;72(1):135–41.
- 312 14. Rizopoulos D. Joint Models for Longitudinal and Time-to-Event Data:
313 With Applications in R. CRC Press; 2012. ISBN 9781439872864.
- 314 15. Bruinsma SM, Zhang L, Roobol MJ, Bangma CH, Steyerberg EW,
315 Nieboer D, Van Hemelrijck M, consortium MFGAPPCASG, Trock B,
316 Ehdaie B, et al. The movember foundation’s gap3 cohort: a profile of
317 the largest global prostate cancer active surveillance database to date.
318 *BJU international* 2018;121(5):737–44.
- 319 16. Laird NM, Ware JH, et al. Random-effects models for longitudinal data.
320 *Biometrics* 1982;38(4):963–74.

- 321 17. Vickers AJ, Savage C, O'Brien MF, Lilja H. Systematic review of pre-
322 treatment prostate-specific antigen velocity and doubling time as pre-
323 dictors for prostate cancer. *Journal of Clinical Oncology* 2009;27(3):398.
- 324 18. Rizopoulos D. The R package JMBayes for fitting joint models for lon-
325 gitudinal and time-to-event data using MCMC. *Journal of Statistical*
326 *Software* 2016;72(7):1–46.
- 327 19. Rizopoulos D, Molenberghs G, Lesaffre EM. Dynamic predictions with
328 time-dependent covariates in survival analysis using joint modeling and
329 landmarking. *Biometrical Journal* 2017;59(6):1261–76.
- 330 20. Ankerst DP, Xia J, Thompson Jr IM, Hoefler J, Newcomb LF, Brooks
331 JD, Carroll PR, Ellis WJ, Gleave ME, Lance RS, et al. Precision
332 medicine in active surveillance for prostate cancer: development of the
333 canary–early detection research network active surveillance biopsy risk
334 calculator. *European urology* 2015;68(6):1083–8.
- 335 21. Balasubramanian R, Lagakos SW. Estimation of a failure time distribu-
336 tion based on imperfect diagnostic tests. *Biometrika* 2003;90(1):171–82.