

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. This risk was utilized to schedule personalized biopsies. Personalized and fixed schedules compared on number of biopsies and model estimated time delay in detection of reclassification for each schedule. Model validated externally in largest five AS cohorts of GAP3 database. Methodology implemented in a web-application.

Results and Limitations: Rate of reclassification in PRIAS was 35% at year 5 of follow-up. PSA velocity 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). Validation: Area under ROC curve for risk predictions between 0.55 and 0.75 for PRIAS, Johns Hopkins, Toronto, and Memorial Sloan Kettering AS cohorts. Model required recalibration for all external cohorts except Johns Hopkins cohort.

Conclusions: We used risk predictions of reclassification to schedule personalized biopsies for AS patients. To assist patients/doctors in shared

decision making of appropriate biopsy schedule, we provided them expected time delay in detection of reclassification, for personalized/fixed schedules. Our model is externally validated, and our methodology is available for multiple AS cohorts as a web-application.

Patient Summary: Personalized biopsy schedules are a novel alternative to fixed biopsy schedules. They rely on patient-specific risk of reclassification and can offer better balance between number of biopsies and time delay in detection of reclassification than current schedules.

Keywords: Active Surveillance, Biopsies, Personalized Medicine, Prostate Cancer, Shared Decision Making

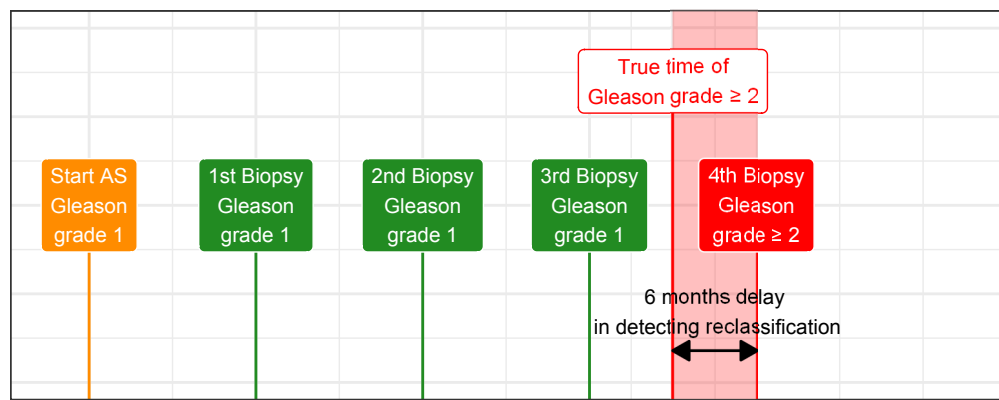
1. Introduction

Patients with low- and very low-risk screening-detected localized prostate cancer 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 the Gleason grade increases from grade 1 (Gleason 3+3) to 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 accumulated patient data (e.g., PSA, previous biopsy results) into risk estimates for reclassification. Existing calculators for risk of reclassifica-

A Biopsy every year



B Biopsy every 2 years

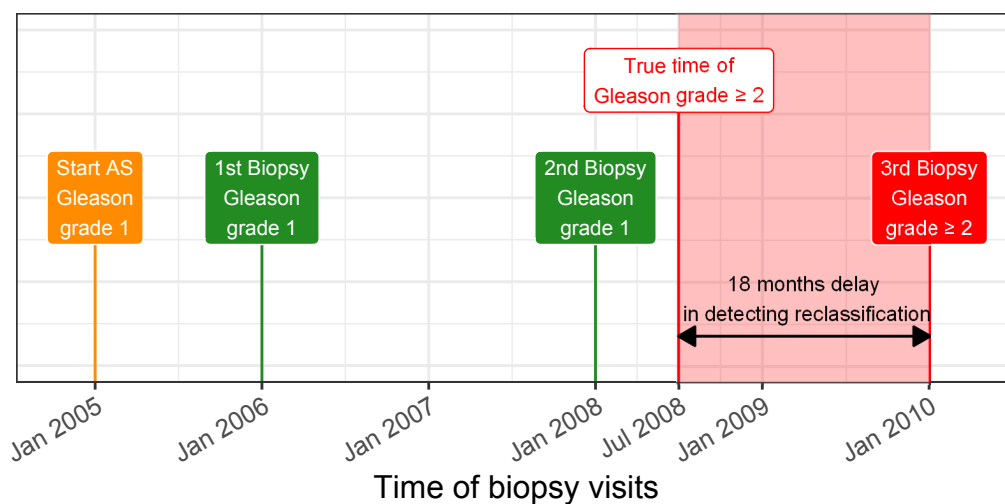


Figure 1: Trade-off between the number of biopsies and time delay in detecting reclassification (Increase in Gleason grade from 1 to 2 or higher): The true time of reclassification for the patient in this figure is July 2008. When biopsies are scheduled annually (**Panel A**), reclassification is detected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

A Should a biopsy be conducted at current visit?

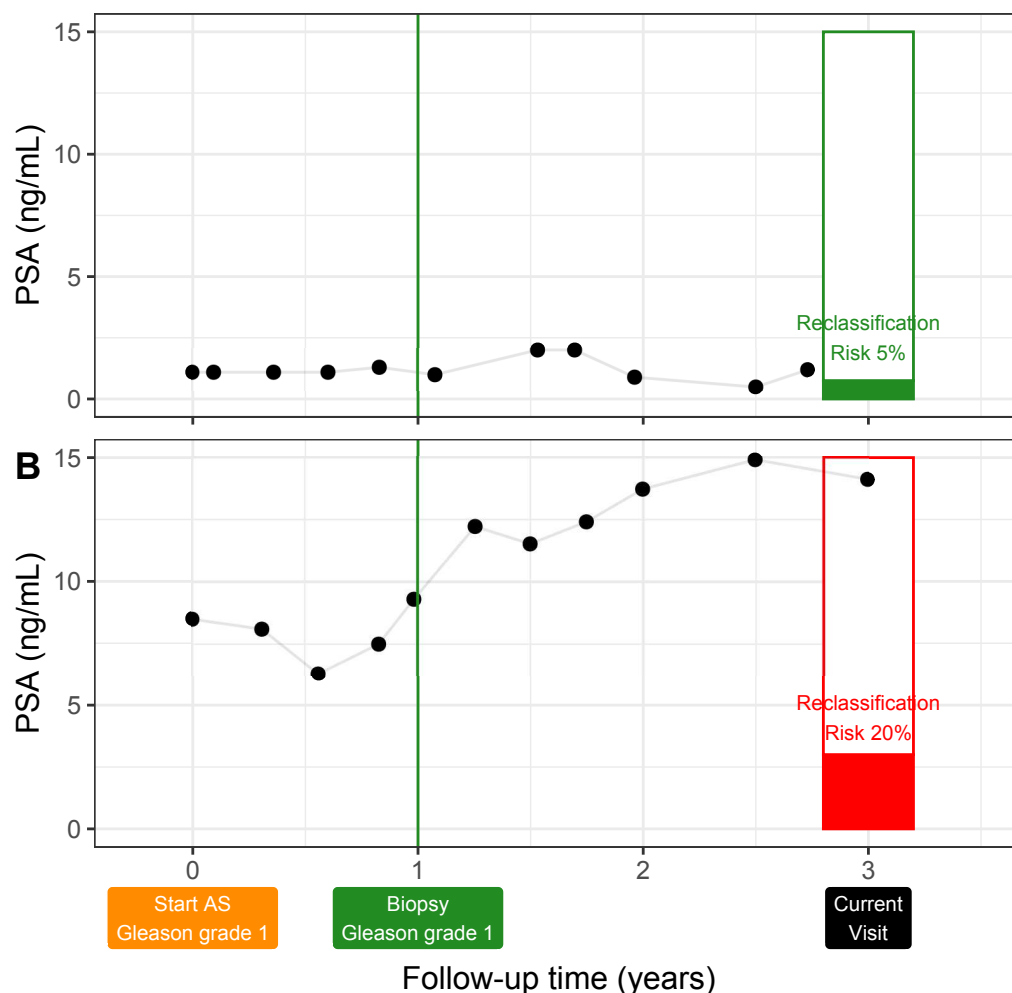


Figure 2: **Motivation for personalized risk-based decisions of biopsy:** Patient A (**Panel A**) and B (**Panel B**) had their latest biopsy at year one of follow-up (green vertical line). Patient A's prostate-specific antigen (PSA) profile remained stable until his current visit at year three, whereas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

tion [10, 11] use only the latest PSA measurement of a patient. In contrast, we intend to utilize 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 [12, 13, 14]. A joint model predicts risk of reclassification in a personalized manner. A subsequent challenge however, is translating risks into clinical decisions. For example, a 10% risk of reclassification can be perceived high/low depending upon the patient age. Patients may also weigh risks of reclassification 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 gives 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 largest give AS cohorts of the GAP3 database [15]. Thus, the web-application 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- and
 54 very-low risk prostate cancer diagnoses [3]. More than 100 medical centers
 55 from 17 countries contributed to PRIAS, using a common protocol (www.prias-project.org).
 56 Upon inclusion in AS, PSA was measured quarterly
 57 for the first two years of follow-up and semiannually thereafter. Biopsies were
 58 scheduled at year one, four, seven, and ten of follow-up. Additional yearly
 59 biopsies were scheduled when PSA doubling time was between zero and ten
 60 years.

61 We selected all 7813 patients who had Gleason grade 1 [2] at the time of
 62 inclusion in PRIAS (Table 1). Our primary event of interest is increase in
 63 this Gleason grade upon repeat biopsy, called *reclassification* (1134 patients).
 64 Reclassification is a trigger for treatment advice in PRIAS. Although, 2250
 65 patients were provided treatment on the basis of their PSA, or number of
 66 biopsy cores with cancer, or anxiety/other reasons. Our reasons for focusing
 67 solely on reclassification are, namely reclassification is strongly associated
 68 with cancer related outcomes, and other triggers for treatment vary between
 69 cohorts.

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)

70 2.2. Statistical Model

71 To create personalized biopsy schedules based on patient-specific risk of
 72 reclassification, we required a risk prediction model. Available data was
 73 patient age at inclusion in AS, longitudinally measured PSA, the timing
 74 of repeat biopsies and corresponding Gleason grades, and observed time of
 75 reclassification. Analysis of this data required modeling the within-patient
 76 correlation for PSA, the association between the Gleason grades and PSA
 77 profiles of a patient, and handling missing PSA measurements when a patient
 78 experienced reclassification. In such situations, a commonly used model is
 79 the joint model for time-to-event and longitudinal data [12, 13, 14].

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

92 2.3. Risk of Reclassification Based Personalized Biopsies

93 The key component in personalized schedules is the cumulative-risk of re-
 94 classification. Given, a patient’s accumulated PSA measurements and biopsy

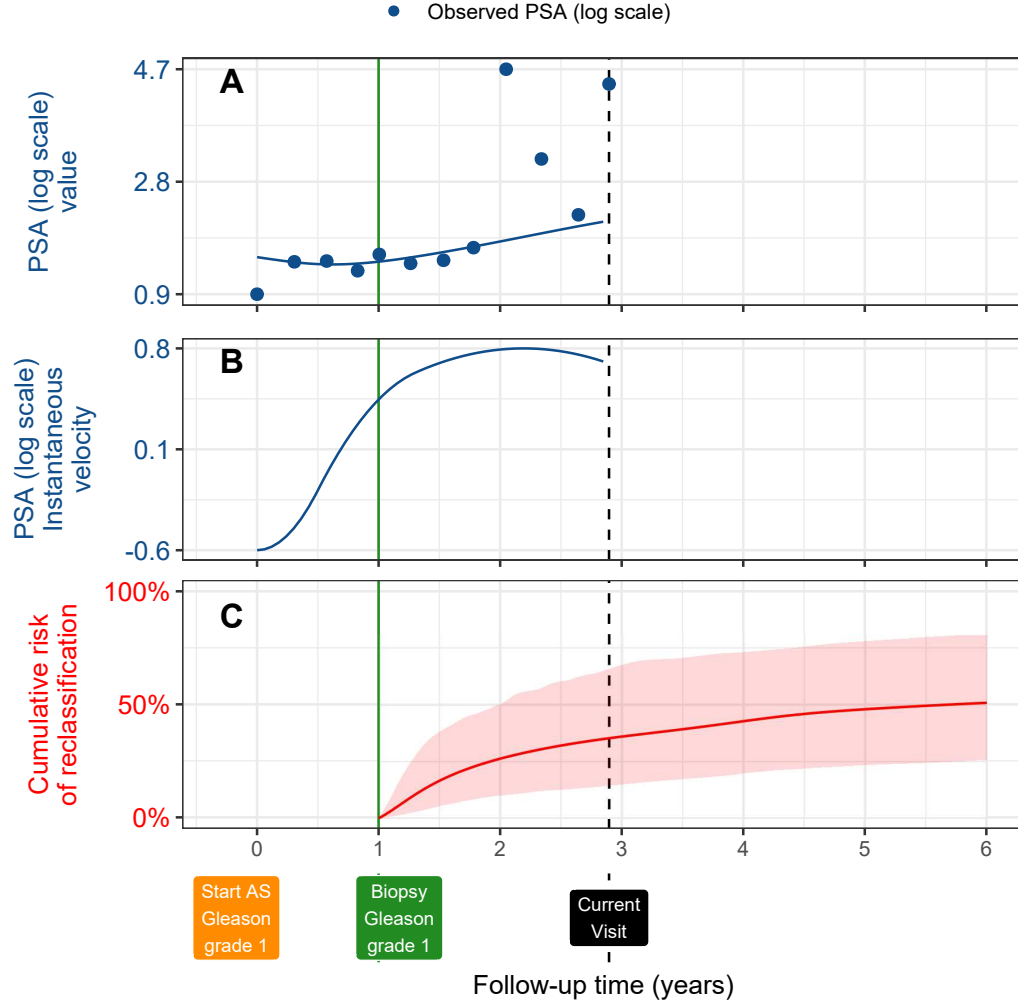


Figure 3: **Illustration of the joint model on a real PRIAS patient.** **Panel A:** Observed PSA (blue dots) and fitted PSA (solid blue line), log-transformed. **Panel B:** Estimated instantaneous velocity of PSA (log-transformed). **Panel C:** Predicted cumulative-risk of reclassification (95% credible interval shaded). Reclassification is defined as an increase in Gleason grade 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). The joint model estimated it by combining the fitted value and estimated instantaneous velocity of PSA (log scale), and time of the latest negative biopsy. Black dashed line at year 3 denotes the time of current visit.

95 results, our joint model predicts the cumulative-risk of reclassification at his
 96 current as well as future visit times (Panel C, Figure 3). This cumulative-
 97 risk is also updated as more patient data becomes available over follow-up
 98 (Figure 5, Supplementary B).

99 In PRIAS, patient PSA is measured every 6 months. If during a PSA
 100 visit, a patient’s predicted cumulative-risk of reclassification is more than a
 101 certain threshold (e.g., 10%) we schedule an immediate biopsy. Since our
 102 model predicted his cumulative-risk at his future follow-up visits as well, we
 103 can schedule future biopsies too. We achieve this by repeatedly applying the
 104 same risk threshold rule at each future follow-up visit (Supplementary C). We
 105 maintain a minimum gap of one year between consecutive biopsies (PRIAS
 106 recommendation). Example personalized schedules based on 5% and 10%
 107 risk thresholds are shown in Panel B, Figure 4.

108 The choice of the risk threshold in the personalized schedule dictates the
 109 *consequences* of following that schedule. *Consequences* are, the timing and
 110 the total number of biopsies, and the expected delay in detecting reclassifica-
 111 tion. Our model also estimated these *consequences* in a personalized manner
 112 (Panel B,C in Figure 4, and Figure 9–11 in Supplementary C) given any
 113 schedule of biopsies. Thus patients can compare fixed schedules with differ-
 114 ent risk personalized schedules before making a choice. Due to the limited
 115 follow-up period of PRIAS, we are only able to schedule biopsies during the
 116 first six years of follow-up of a patient (Table 12, Supplementary C).

117 2.4. Model Validation

118 We validated our model internally using the PRIAS cohort, as well as
 119 externally using largest five AS cohorts of the GAP3 database [15]. These

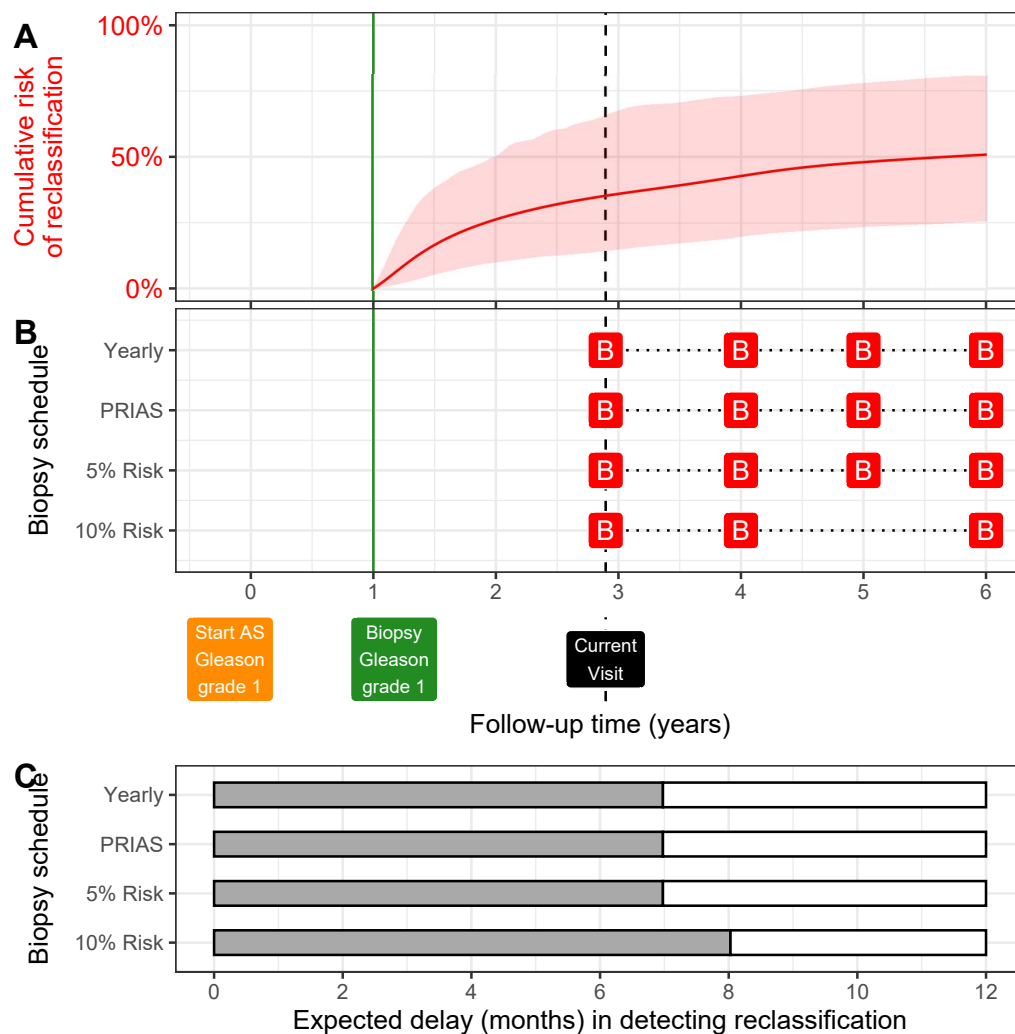


Figure 4: **Illustration of personalized and fixed schedules of biopsies.** The PSA profile of this patient is shown in Figure 3. **Panel A:** Predicted cumulative-risk of reclassification (95% credible interval shaded). **Panel B:** Personalized and fixed schedules of biopsies, with a red 'B' indicating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

were, namely 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). We assessed our model’s ability to discriminate between patients who observe reclassification versus patients who do not observe reclassification, using the area under the receiver operating characteristic curve or AUC [19]. We evaluated the prediction accuracy of our model visually using calibration plots [20, 21], and quantitatively via mean absolute prediction error [19]. Due to the longitudinal nature of AS studies, the AUC and prediction error varies over follow-up (Supplementary B.1). We recalibrated our model’s baseline hazard of reclassification in those external cohorts (Supplementary B.1) where our model was miscalibrated.

2.5. *Web-Application*

We implemented our methodology in a web-application https://emcbiostatistics.shinyapps.io/prias_biopsy_recommender/. It utilizes the 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 a limited follow-up period. This limit varies between cohorts according to their current study period (Table 12, Supplementary C). The web-application allows comparison of the *consequences* of following these schedules: personalized schedules based on 5%, 10%, and 15% risk threshold; annual biopsies; biennial biopsies; and PRIAS schedule.

144 3. Results

145 In PRIAS, the rate of reclassification within the first five years of follow-
 146 up was 35%. This rate was capped at a maximum of 50% in external GAP3
 147 cohorts (Panel B, Figure 5). That is, many patients do not require any biopsy
 148 in the first five years.

149 In the fitted joint model, when patient age increased from 61 to 71
 150 years (25-th to 75-th percentile), the adjusted hazard ratio of reclassification
 151 was 1.45 (95%CI: 1.30–1.63). When fitted PSA value (log scale) increased
 152 from 2.36 to 3.07 (25-th to 75-th percentile), the adjusted hazard ratio was
 153 0.99 (95%CI: 0.89–1.11). When estimated instantaneous PSA (log scale) ve-
 154 locity increased from -0.09 to 0.31 (25-th to 75-th percentile), the adjusted
 155 hazard ratio was 2.47 (95%CI: 1.93–2.99). Hence, instantaneous velocity of
 156 PSA (log scale) was a stronger predictor for reclassification than its value.
 157 Detailed parameter estimates are in Supplementary A.2.

158 The time-varying AUC and calibration of our model in different cohorts
 159 are shown in Panel A and Panel B of Figure 5, respectively. The AUC
 160 achieves a moderate level in all cohorts. It fluctuates roughly around 0.63
 161 over time. In terms of calibration, our model seems well calibrated only for
 162 the Johns Hopkins AS cohort. However, we resolved this issue by recalibrat-
 163 ing the baseline hazard of our model separately for each cohort (Figure 6,
 164 Supplementary B). The resulting risk predictions for individual patients from
 165 this recalibrated model and from separately fitted new joint models for each
 166 cohort were similar (Figure 7, Supplementary B). Comprehensive discussion
 167 of validation results is in Supplementary B.

168 Various personalized and fixed biopsy schedules for a demonstration pa-

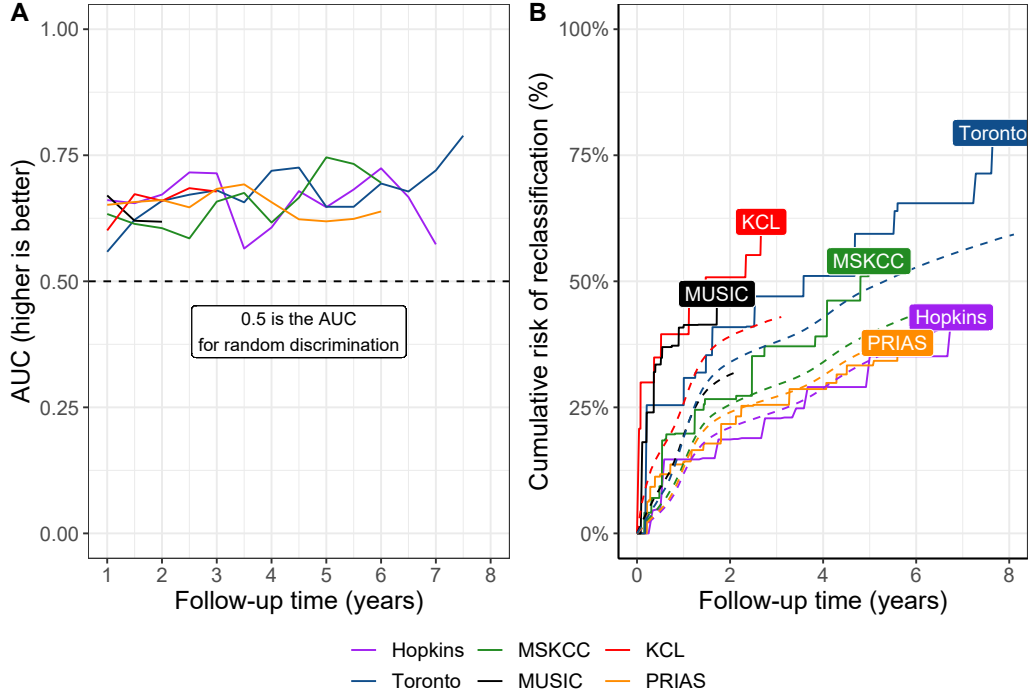


Figure 5: Model Validation Results. **Panel A:** time dependent area under the receiver operating characteristic curve or AUC (measure of discrimination). **Panel B:** calibration-at-large [20, 21], with solid lines depicting the non-parameteric estimate of the cumulative-risk of reclassification [22], and dashed lines showing the average cumulative-risk of reclassification obtained using the joint model fitted to the PRIAS dataset. 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.

169 tient in Figure 4 show that a personalized schedule based on 10% risk thresh-
170 old leads to one less biopsy than other schedules. At the same time, the cor-
171 responding time delay in detection of reclassification is expected to be only
172 one month more than other schedules. A compulsory biopsy was scheduled at
173 year six (maximum biopsy scheduling time of our model, Supplementary C)
174 in all schedules for a meaningful comparison across schedules. Schedules for
175 other demonstration patients are in Figure 9–11, Supplementary C.

176 4. Discussion

177 We developed a novel methodology and statistical model for personal-
 178 ized biopsy schedules in prostate cancer active surveillance (AS) patients.
 179 Personalized schedules utilize patient-specific risks of reclassification. Re-
 180 classification is defined as increase in Gleason grade [2] from grade 1 to
 181 2 or higher upon repeat biopsy. Reclassification risk calculators are not
 182 new [13, 23]. However, our work has four novel features. First, we developed
 183 a statistical model for developing personalized schedules using the world’s
 184 largest AS cohort PRIAS. Second, we created a methodology to estimate
 185 time delay in detection of reclassification (less is beneficial) in a personal-
 186 ized manner, given any biopsy schedule. Thus patients/doctors can compare
 187 schedules before making a choice. Third, we externally validated our model
 188 in the largest five AS cohorts of the GAP3 database [15]. Fourth, we imple-
 189 mented our methodology in a web-application [https://emcbiostatistics.](https://emcbiostatistics.shinyapps.io/prias_biopsy_recommender/)
 190 [shinyapps.io/prias_biopsy_recommender/](https://emcbiostatistics.shinyapps.io/prias_biopsy_recommender/) for PRIAS and externally val-
 191 idated cohorts.

192 Currently, in AS, either fixed schedules are used (e.g., annual biopsies), or
 193 PSA is used to trigger biopsies. Both approaches have been criticized [17, 6].
 194 We argue that current approaches have not exploited PSA fully and cor-
 195 rectly. For example, using observed PSA is incorrect as it has measure-
 196 ment error. Other approaches utilize only the latest PSA and/or when they
 197 utilize all PSA data they assume it changes over time linearly (constant
 198 PSA velocity). In contrast, our joint model builds a measurement error free
 199 patient-specific profile of PSA using all PSA measurements. It allows PSA to
 200 increase/decrease non-linearly over time (non-constant PSA velocity). Sub-

201 sequently, it consolidates underlying PSA value and velocity, previous biopsy
202 results, and baseline characteristics of a patient, to yield a single personalized
203 estimate for the risk of reclassification. Furthermore, this risk gets updated
204 as more patient data becomes available over follow-up. This is a more holistic
205 approach. Although we have not incorporated newer biomarkers and
206 magnetic resonance imaging (MRI) data, when this information becomes
207 available we can add them as predictors in our model. Decisions based on
208 combined information from multiple sources can yield better results than
209 based on MRI or PSA alone.

210 Our model is useful for a large number of patients from PRIAS, as well as
211 from the largest five AS cohorts of the GAP3 database in which we validated
212 our model. These are the University of Toronto AS, Johns Hopkins AS,
213 Memorial Sloan Kettering Cancer Center AS, King’s College London AS, and
214 Michigan Urological Surgery Improvement Collaborative AS. We required
215 recalibration of our model’s baseline hazard of reclassification for all of these
216 cohorts except the Johns Hopkins cohort. This can be explained by the fact
217 that both PIRAS and Johns Hopkins cohorts have the same marginal rate
218 of reclassification over time (Panel B, Figure 5). Extending our model in
219 smaller cohorts requires only recalibrating our model.

220 Our work has important clinical implications. The rate of reclassification
221 after five years of follow-up was capped at 50% in all cohorts that we
222 evaluated (Figure 5). That is, a large number of patients do not require any
223 biopsy during the first five years of follow-up. Given the non-compliance and
224 burden of frequent biopsies [6], the availability of our methodology as a web-
225 application may encourage patients/doctors to consider personalized sched-

226 ules instead. To assist them in this decision making, the web-application
 227 provides an estimate of time delay in detection of reclassification for both
 228 personalized and fixed schedules, in a personalized manner. We hope this
 229 will objectively address patient apprehensions regarding adverse outcomes in
 230 AS.

231 This work has certain limitations. The proposed model is valid only for
 232 the first six years of follow-up in PRIAS, whereas reclassification may occur
 233 much later in many patients. This problem can be mitigated by refitting the
 234 model with new follow-up data in the future. While we focused only on re-
 235 classification, the number of positive biopsy cores can also be used to trigger
 236 treatment. We did not consider such additional criteria because they differ
 237 between cohorts [4], whereas, reclassification is commonly used. Although,
 238 reclassification is susceptible to inter-observer variation. Models which ac-
 239 count for this variation [13, 24] will be interesting to investigate further.
 240 However, the methodology for personalized scheduling, and for comparison
 241 of various schedules need not change.

242 5. Conclusions

243 We developed a novel methodology and statistical model for personalized
244 scheduling of biopsies in prostate cancer active surveillance (AS) patients.
245 Unlike fixed biopsy schedules, personalized schedules utilize a patient's risk
246 of reclassification to decide biopsies. They also update as more patient data
247 becomes available over follow-up. Our model is externally validated in largest
248 five AS cohorts of the GAP3 database. Our methodology is implemented
249 in a web-application ([https://emcbiostatistics.shinyapps.io/prias_
250 biopsy_recommender/](https://emcbiostatistics.shinyapps.io/prias_biopsy_recommender/)) and is accessible to large number of AS patients from
251 the validated cohorts. To assist patients/doctors in making a shared decision
252 of an appropriate biopsy schedule, the web-application provides expected
253 time delay in detection of reclassification (smaller is beneficial), and timing
254 and total number of biopsies (burden), for both personalized and currently
255 used fixed schedules.

256 **Author Contributions**

257 Anirudh Tomer had full access to all the data in the study and takes
 258 responsibility for the integrity of the data and the accuracy of the data anal-
 259 ysis.

260 *Study concept and design:* Tomer, Nieboer, Roobol, Bjartell, and Ri-
 261 zopoulos

262 *Acquisition of data:* Tomer, Nieboer, and Roobol

263 *Analysis and interpretation of data:* Tomer, Nieboer, and Rizopoulos

264 *Drafting of the manuscript:* Tomer, and Rizopoulos

265 *Critical revision of the manuscript for important intellectual content:* Tomer,
 266 Nieboer, Roobol, Bjartell, Steyerberg, and Rizopoulos

267 *Statistical analyses:* Tomer, Nieboer, Steyerberg, and Rizopoulos

268 *Obtaining funding:* Roobol, and Rizopoulos

269 *Administrative, technical or material support:* Nieboer

270 *Supervision:* Roobol, and Rizopoulos

271 *Other:* none

272 **Acknowledgments**

273 This work was supported by the Movember Foundation. The funder did
274 not play any role in the study design, collection, analysis or interpretation of
275 data, or in the drafting of this paper. The first and last authors would like
276 to acknowledge support by Nederlandse Organisatie voor Wetenschappelijk
277 Onderzoek (the national research council of the Netherlands) VIDI grant nr.
278 016.146.301, and Erasmus University Medical Center funding. Part of this
279 work was carried out on the Dutch national e-infrastructure with the sup-
280 port of SURF Cooperative. The authors also thank the Erasmus University
281 Medical Center’s Cancer Computational Biology Center for giving access to
282 their IT-infrastructure and software that was used for the computations and
283 data analysis in this study.

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. 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.
4. 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.
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. 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)

- 306 protocol and disease reclassification in noncompliers. *European Urology*
307 2015;68(5):814–21.
- 308 7. Inoue LY, Lin DW, Newcomb LF, Leonardson AS, Ankerst D, Gulati R,
309 Carter HB, Trock BJ, Carroll PR, Cooperberg MR, et al. Comparative
310 analysis of biopsy upgrading in four prostate cancer active surveillance
311 cohorts. *Annals of internal medicine* 2018;168(1):1–9.
- 312 8. Bratt O, Carlsson S, Holmberg E, Holmberg L, Johansson E, Josefsson
313 A, Nilsson A, Nyberg M, Robinson D, Sandberg J, et al. The study of
314 active monitoring in sweden (sams): a randomized study comparing two
315 different follow-up schedules for active surveillance of low-risk prostate
316 cancer. *Scandinavian journal of urology* 2013;47(5):347–55.
- 317 9. de Carvalho TM, Heijnsdijk EA, de Koning HJ. Estimating the risks
318 and benefits of active surveillance protocols for prostate cancer: a mi-
319 crosimulation study. *BJU international* 2017;119(4):560–6.
- 320 10. Partin AW, Yoo J, Carter HB, Pearson JD, Chan DW, Epstein JI, Walsh
321 PC. The use of prostate specific antigen, clinical stage and gleason score
322 to predict pathological stage in men with localized prostate cancer. *The*
323 *Journal of urology* 1993;150(1):110–4.
- 324 11. Makarov DV, Trock BJ, Humphreys EB, Mangold LA, Walsh PC, Ep-
325 stein JI, Partin AW. Updated nomogram to predict pathologic stage of
326 prostate cancer given prostate-specific antigen level, clinical stage, and
327 biopsy gleason score (partin tables) based on cases from 2000 to 2005.
328 *Urology* 2007;69(6):1095–101.

- 329 12. Tomer A, Nieboer D, Roobol MJ, Steyerberg EW, Rizopoulos D. Per-
 330 sonalized schedules for surveillance of low-risk prostate cancer patients.
 331 *Biometrics* 2019;75(1):153–62.
- 332 13. Coley RY, Zeger SL, Mamawala M, Pienta KJ, Carter HB. Prediction
 333 of the pathologic gleason score to inform a personalized management
 334 program for prostate cancer. *European urology* 2017;72(1):135–41.
- 335 14. Rizopoulos D. Joint Models for Longitudinal and Time-to-Event Data:
 336 With Applications in R. CRC Press; 2012. ISBN 9781439872864.
- 337 15. Bruinsma SM, Zhang L, Roobol MJ, Bangma CH, Steyerberg EW,
 338 Nieboer D, Van Hemelrijck M, consortium MFGAPPCASG, Trock B,
 339 Ehdaie B, et al. The movember foundation’s gap3 cohort: a profile of
 340 the largest global prostate cancer active surveillance database to date.
 341 *BJU international* 2018;121(5):737–44.
- 342 16. Laird NM, Ware JH, et al. Random-effects models for longitudinal data.
 343 *Biometrics* 1982;38(4):963–74.
- 344 17. Vickers AJ, Savage C, O’Brien MF, Lilja H. Systematic review of pre-
 345 treatment prostate-specific antigen velocity and doubling time as pre-
 346 dictors for prostate cancer. *Journal of Clinical Oncology* 2009;27(3):398.
- 347 18. Rizopoulos D. The R package JMBayes for fitting joint models for lon-
 348 gitudinal and time-to-event data using MCMC. *Journal of Statistical*
 349 *Software* 2016;72(7):1–46.
- 350 19. 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.
20. Royston P, Altman DG. External validation of a cox prognostic
model: principles and methods. *BMC medical research methodology*
2013;13(1):33.
21. 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. *Epidemi-
ology (Cambridge, Mass)* 2010;21(1):128.
22. Turnbull BW. The empirical distribution function with arbitrarily
grouped, censored and truncated data. *Journal of the Royal Statisti-
cal Society Series B (Methodological)* 1976;38(3):290–5.
23. Ankerst DP, Xia J, Thompson Jr IM, Hoeffler 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.
24. Balasubramanian R, Lagakos SW. Estimation of a failure time distribu-
tion based on imperfect diagnostic tests. *Biometrika* 2003;90(1):171–82.