

Personalized Biopsies in Prostate Cancer Active Surveillance[☆]

Anirudh Tomer, MSc^{a,*}, Daan Nieboer, MSc^b, Monique J. Roobol, PhD^c,
Anders Bjartell, PhD^d, Dimitris Rizopoulos, PhD^a, Movember Foundations
Global Action Plan Prostate Cancer Active Surveillance (GAP3)
consortium^e

^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, Skne University Hospital, Malm, Sweden*

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

Abstract

Background: Low-risk prostate cancer patients enrolled in active surveillance (AS) undergo repeat biopsies. Treatment often advised when biopsy Gleason ≥ 7 (GS7). Many patients never obtain GS7, yet undergo biopsies frequently.

Objective: Better balance the number of biopsies and time delay in detection of GS7.

Design, Setting, and Participants: World's largest AS study PRIAS,

[☆]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), d.rizopoulos@erasmusmc.nl (Dimitris Rizopoulos, PhD)

7813 patients, 1134 obtained GS7, 100 medical centers in 17 countries, common study protocol.

Outcome Measurements, and Statistical Analysis: Prostate-specific antigen (PSA) measured every six months, repeat biopsies every one to three years. Bayesian joint model fitted to the observed time of GS7 and longitudinal PSA measurements. Risk predictions for GS7 externally validated in five largest AS cohorts from GAP3 database. Personalized risk based biopsy schedules developed using GS7 predictions. Total biopsies, time of biopsies and expected time delay in detection of GS7 calculated for various schedules to compare them.

Results and Limitations: Roughly 50% patients did not obtain GS7 in first 10 years in PRIAS. PSA velocity was a stronger predictor of GS7 with Hazard Ratio (increase from 1st to 3rd quartile): 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 GS7 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 risk based biopsy schedules as alternative to fixed schedules. To assist patients in biopsy decisions we provided total and time of biopsies, and expected time delay in detection of GS7, for fixed and personalized schedules. Personalized schedules update with more patient data over follow-up.

Patient Summary: Personalized biopsy schedules are a novel alternative to fixed biopsy schedules (e.g., yearly biopsies). They utilize a patient's PSA

and biopsy history to decide best time of biopsies. Personalized schedules can offer a better balance between number of biopsies and delay in detection of Gleason upgrade than yearly biopsies.

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

1. Introduction

Patients with low- and very low-risk screening-detected localized prostate cancer (PCa) are usually advised active surveillance (AS) instead of immediate radical treatment [1]. In AS, PCa progression is routinely monitored via prostate-specific antigen (PSA), digital rectal examination, and the Gleason score from repeat prostate biopsies. Since Gleason score is the strongest indicator of cancer-related outcomes, patients are commonly advised curative treatment when Gleason ≥ 7 (GS7) is detected [2].

Biopsies are conducted periodically, and hence GS7 is always detected with a time delay. The smaller this delay, the larger is the window of opportunity for curative treatment. For timely detection of GS7, many AS programs use fixed biopsy schedules (e.g., annual or biennial biopsies) for all patients [3, 4]. However, many AS patients do not require any biopsy in the first ten years of follow-up (see PRIAS and John’s Hopkins AS cohorts in Figure 1). Biopsies are also invasive, painful and prone to medical complications. Biopsy burden combined with patient non-compliance [5] to frequent biopsies, has raised concerns regarding the optimal biopsy schedule [6, 7].

A simple alternative to frequent biopsies is infrequent biopsies. However, studies suggest not reducing biopsy frequency beyond 24 months, to have a sufficient window of opportunity for curative treatment [6, 9]. Although, biopsy every 24 months still leads to five unnecessary biopsies over ten years for 50% patients (Figure 1). A promising alternative to fixed biopsies is biopsies based on personalized risk of GS7. For example, among the two patients shown in Figure 2, patient B is a more suitable candidate for biopsy than patient A, because his risk of GS7 is much higher.

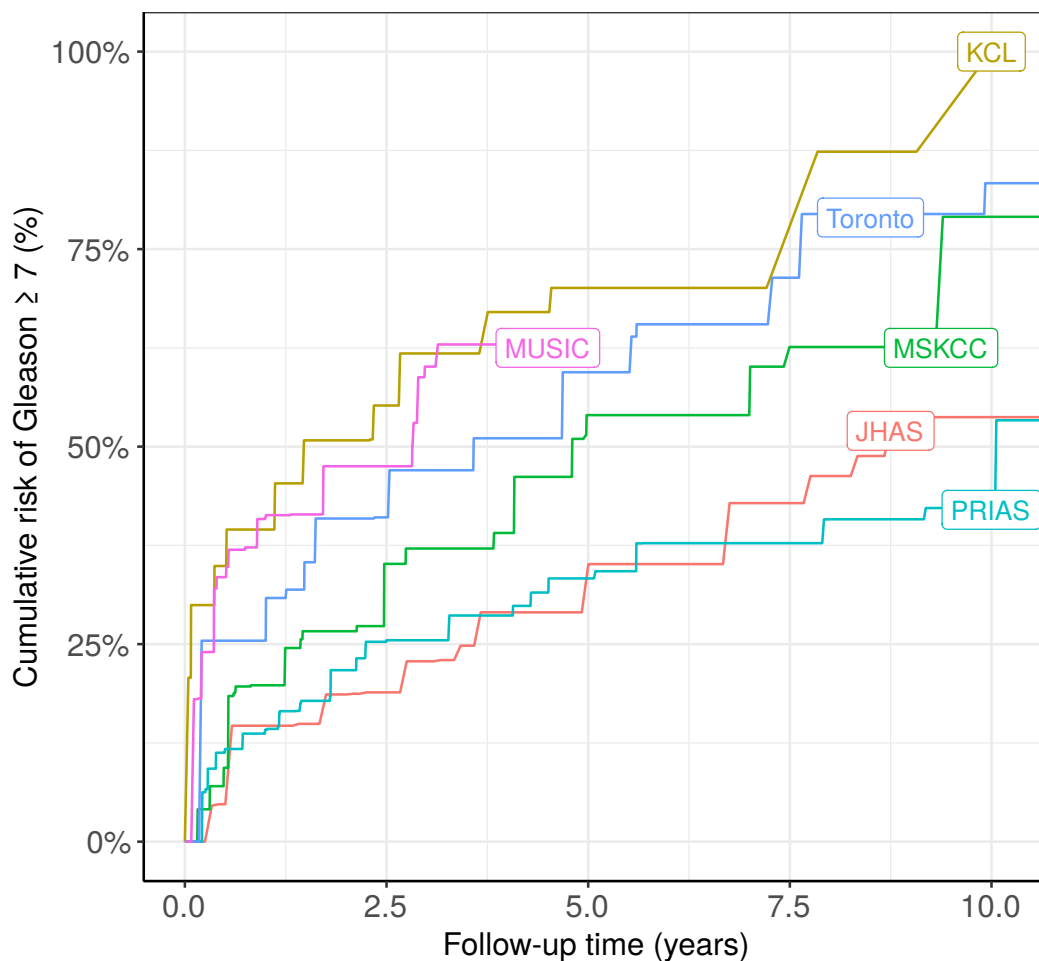


Figure 1: **Estimated cumulative risk of having Gleason ≥ 7 (GS7)** in the world's largest AS cohort PRIAS, and five of the largest AS cohorts part of the GAP3 database [8]. Abbreviations are *JHAS*: Johns Hopkins Active Surveillance, *PRIAS*: Prostate Cancer International Active Surveillance, *Toronto*: University of Toronto Active Surveillance, *MSKCC*: Memorial Sloan Kettering Cancer Center Active Surveillance, *KCL*: King's College London Active Surveillance. In the world's largest AS cohort PRIAS and in JHAS, roughly 50% of patients do not obtain GS7 in the first ten years. That is, ideally no biopsies should be prescribed to 50% of the patients in first ten years of AS.

26 The first challenge in developing risk-based schedules is consolidating ob-
 27 served patient data (e.g., PSA, previous biopsy results) into GS7 risk esti-
 28 mates. Previousy, studies have utilized latest value of PSA to predict the
 29 Gleason score [10, 11]. However, in AS the entire trajectory of PSA of a
 30 patient is available. To accomodate such longitudinal PSA data, a suitable
 31 model is joint model for time-to-event and longitudinal data [12, 13, 14]. A
 32 subsequent challenge is to translate risk estimates for GS7 into clinical deci-
 33 sions is challenging. For example, a 10% risk can be perceived as high/low de-
 34 pending upon the patient’s age. Patients may also weigh the risk of GS7 with
 35 the potential consequences of another biopsy. Two relevant consequences are
 36 the timing and total number of biopsies (burden), and the time delay in the
 37 detection of GS7 (smaller is better). These consequences vary between the
 38 patients, and also over the follow-up period for the same patient.

39 The goal of this work was to assist patients and doctors in making bet-
 40 ter decisions of biopsies than fixed and frequent biopsies. We intended to
 41 achieve this by providing the patients risk based personalized schedules of
 42 biopsies, and to allow them to compare the consequences of each schedule
 43 before making a decision. To this end, we took three steps. First we fitted
 44 a prediction (joint) model to the world’s largest AS dataset, PRIAS [2]. We
 45 then externally validated the model predictions in five largest AS cohorts
 46 that are part of the GAP3 database. Lastly, we utilized the personalized
 47 GS7 risk predictions to calculate the timing and total number of biopsies,
 48 and the time delay in the detection of GS7 for risk-based and fixed biopsy
 49 schedules.

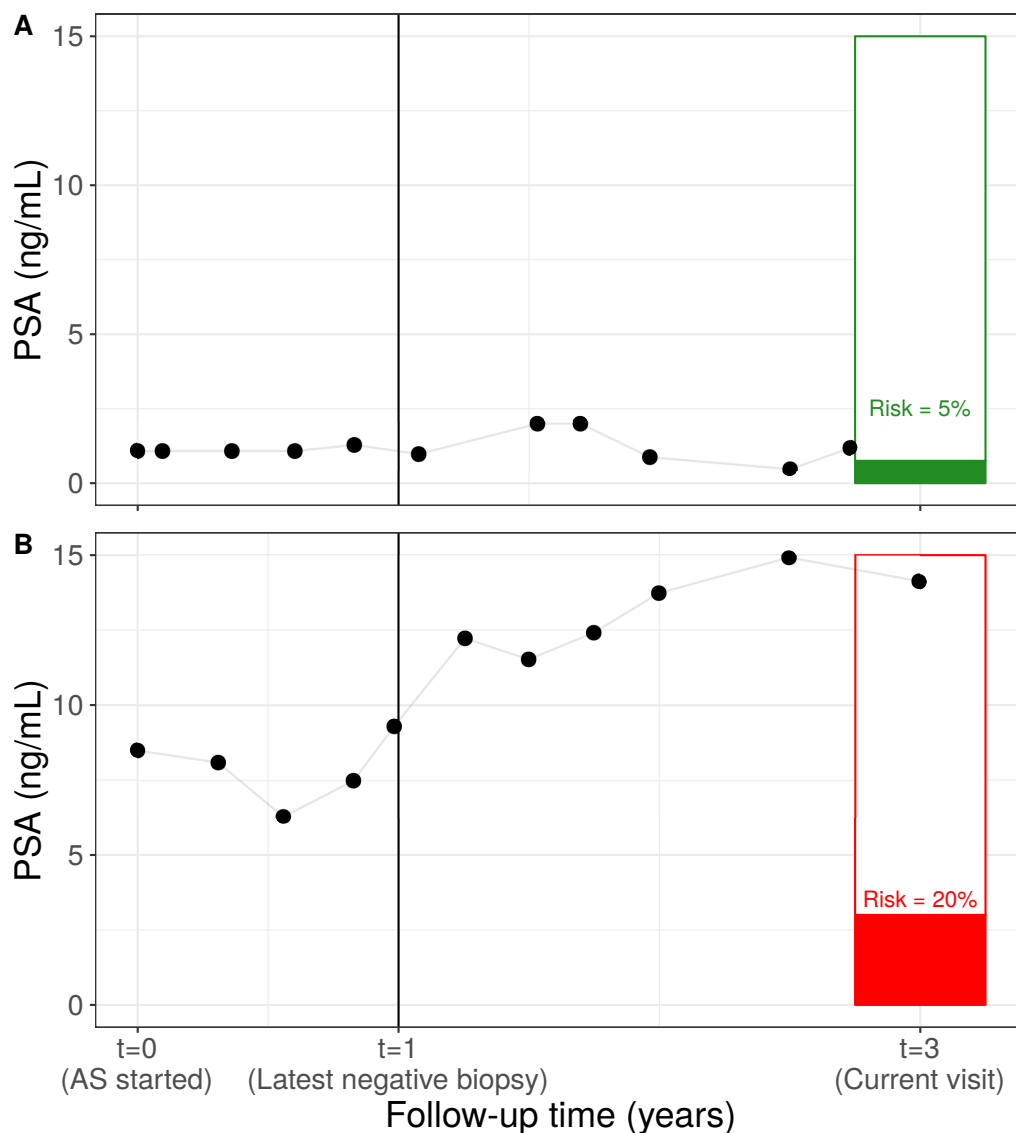


Figure 2: Motivation for personalized risk-based decisions of biopsy: Patient A and B had their latest biopsy at year one of follow-up. Patient A's prostate-specific antigen profile remained stable until the current visit time at year three, whereas patient B's profile has shown a rise. Consequently, patient B's estimated cumulative risk of GS7 at year three is higher than that of patient A, making patient B a more suitable candidate for biopsy than Patient A. Risk estimates in this figure are only illustrative.

50 2. Patients and Methods

51 2.1. Study Cohort

52 Prostate Cancer International Active Surveillance (PRIAS) is an ongoing
 53 prospective cohort study of men with low- and very-low risk PCa diagnoses
 54 [2]. More than 100 medical centers from 17 countries contribute
 55 in PRIAS, using a common study protocol (www.prias-project.org). We
 56 used the data collected between December 2006 (beginning of PRIAS study)
 57 and May 2019. The PSA was measured every three months until year two
 58 of follow-up and every six months thereafter. Biopsy schedule was year one,
 59 four, seven, and ten, and additional yearly biopsies when PSA doubling time
 60 was between zero and ten years. The primary event used in this work is
 61 Gleason ≥ 7 (GS7) because it is commonly used as a trigger for treatment
 62 advice. It was observed in 1134 patients. However, 2250 patients were provided
 63 treatment (see Table 1). Treatment in absence of GS7 may have been
 64 advised on the basis of PSA, number of biopsy cores with cancer, anxiety,
 65 or other reasons. We focused only on GS7 because of its strong association
 66 with cancer-related outcomes. Due to the periodical nature of biopsies, the
 67 time of GS7 was only available as a time interval in which GS7 occurred.

Table 1: **Patient characteristics for the PRIAS dataset.** The primary event of interest is Gleason ≥ 7 . IQR: interquartile range, PSA: prostate-specific antigen.

Characteristic	Value
Total patients	7813
Gleason ≥ 7 (primary event)	1134
Treatment	2250
Watchful waiting	334
Loss to follow-up	250
Death (unrelated to prostate cancer)	95
Death (related to prostate cancer)	2
Median age at diagnosis (years)	61 (IQR: 66–71)
Median follow-up period per patient (years)	1.8 (IQR: 0.9–3.99)
Total PSA measurements	65798
Median number of PSA measurements per patient	6 (IQR: 4–12)
Median PSA value (ng/mL)	5.7 (IQR: 4.1–7.7)
Total biopsies	15563
Median number of biopsies per patient	2 (IQR: 1–3)

68 2.2. Statistical Methods

69 Our aim was to develop a model for predicting the time of GS7. The
 70 available data for each patient were, age at the start of AS, all observed PSA
 71 measurements, and the history of biopsies. We wanted to account for the
 72 correlation between the PSA measurements of the same patient, and also
 73 their correlation with the time of GS7. An additional complication was that
 74 the PSA values were missing once a patient obtained GS7. A commonly
 75 used model to handle these issues is the joint model for time-to-event and
 76 longitudinal data [14, 12, 13].

77 The joint model we utilized, exploited patient-specific random effects [15]
 78 to act as a common source of correlation between the PSA and time of GS7
 79 outcomes (see Figure 3 and Appendix A.2's Figure 1). Random effects also
 80 represented the underlying state of PCa, and were included in both the linear
 81 mixed effects sub-model for $\log_2\{\text{PSA} + 1\}$ transformed measurements (see
 82 Appendix A.5), and the relative risk sub-model (similar to Cox model) for
 83 time of GS7. In the sub-model for PSA, random effects non-linearly modeled
 84 the evolution of PSA over time. Simultaneously, in the relative risk model
 85 random effects were used indirectly by including fitted $\log_2\{\text{PSA} + 1\}$ value
 86 and velocity as time dependent covariates. The $\log_2\{\text{PSA} + 1\}$ velocity was
 87 mathematically derived from fitted $\log_2\{\text{PSA} + 1\}$ values. Consequently, the
 88 $\log_2\{\text{PSA} + 1\}$ velocity also changed non-linearly over follow-up.

89 The parameters of the two sub-models were estimated jointly using the
 90 R package **JMbayes** [16]. This package utilizes the Bayesian methodology
 91 to estimate model parameters.

92 2.3. *Assessment of Predictions of GS7*

93 We validated the risk predictions of GS7 from our model within the
 94 PRIAS dataset (internal validation), as well as in five of the largest AS
 95 cohorts part of the GAP3 database [8] (external validation). The external
 96 cohorts were University of Toronto AS (Toronto), Johns Hopkins AS (JHAS),
 97 Memorial Sloan Kettering Cancer Center AS (MSKCC), King’s College Lon-
 98 don AS (KCL), and Michigan Urological Surgery Improvement Collaborative
 99 AS (MUSIC). For validation, we utilized the area under the receiver operat-
 100 ing characteristic curve or AUC [17] as a measure of discrimination, and root
 101 mean squared prediction error or RMSPE [17] as a measure of calibration.
 102 Since AS studies are longitudinal in nature, we computed AUC and RMSPE
 103 in a time dependent manner, at a gap of every six months (follow-up schedule
 104 of PRIAS) until year five (95-percentile of the observed GS7 times in PRIAS)
 105 of follow-up.

106 2.4. *Personalized Schedule of Biopsies, and Its Consequences*

107 The key component in development of personalized schedules is the per-
 108 sonalized risk of GS7. This risk is predicted for new patients using the joint
 109 model fitted to the PRIAS dataset. For example, in Figure 4 we show a new
 110 patient whose cumulative-risk of GS7 (Panel B of Figure 4) is calculated at
 111 each of his follow-up visits since his latest negative biopsy. This risk is dy-
 112 namic in the sense that it is updated as more data is gathered over follow-up.
 113 Now, suppose that this patient does not intend to cross more than a certain
 114 cumulative-risk threshold (e.g., 10% risk) between his successive follow-up
 115 visits. To start with, this rule can be first applied at his current visit to
 116 schedule a biopsy. If a biopsy gets scheduled at his current visit, then his

117 cumulative risk profile is updated in order to account for the possibility of
 118 not finding GS7 during this biopsy. This process is repeated for each of the
 119 future follow-up visits, to obtain an entire personalized schedule of biopsies
 120 (Panel C of Figure 4).

121 To assist patients in making an informed choice for a schedule, be it
 122 personalized or fixed, we provided them patient-specific consequences of fol-
 123 lowing each schedule. To this end, we reused the cumulative-risk of GS7
 124 (Panel B of Figure 4). More specifically, for any schedule of biopsies we first
 125 split the cumulative-risk profile of the patient for each time period between
 126 subsequent biopsies. We then exploited these cumulative-risk profiles to ob-
 127 tain the expected time delay in detection of GS7 in the corresponding time
 128 period of the schedule (Equation 5, Appendix C). Lastly, we combined these
 129 time delays while accounting for the total risk of GS7 in the correspond-
 130 ing time period, to obtain the expected time delay in detection of GS7 for
 131 the schedule. Thus, patients had a method to compare various schedules in
 132 terms of the personalized burden (time and total biopsies), and personalized
 133 benefit (less delay in detection of GS7 is beneficial). We implemented this
 134 approach in a web-application.

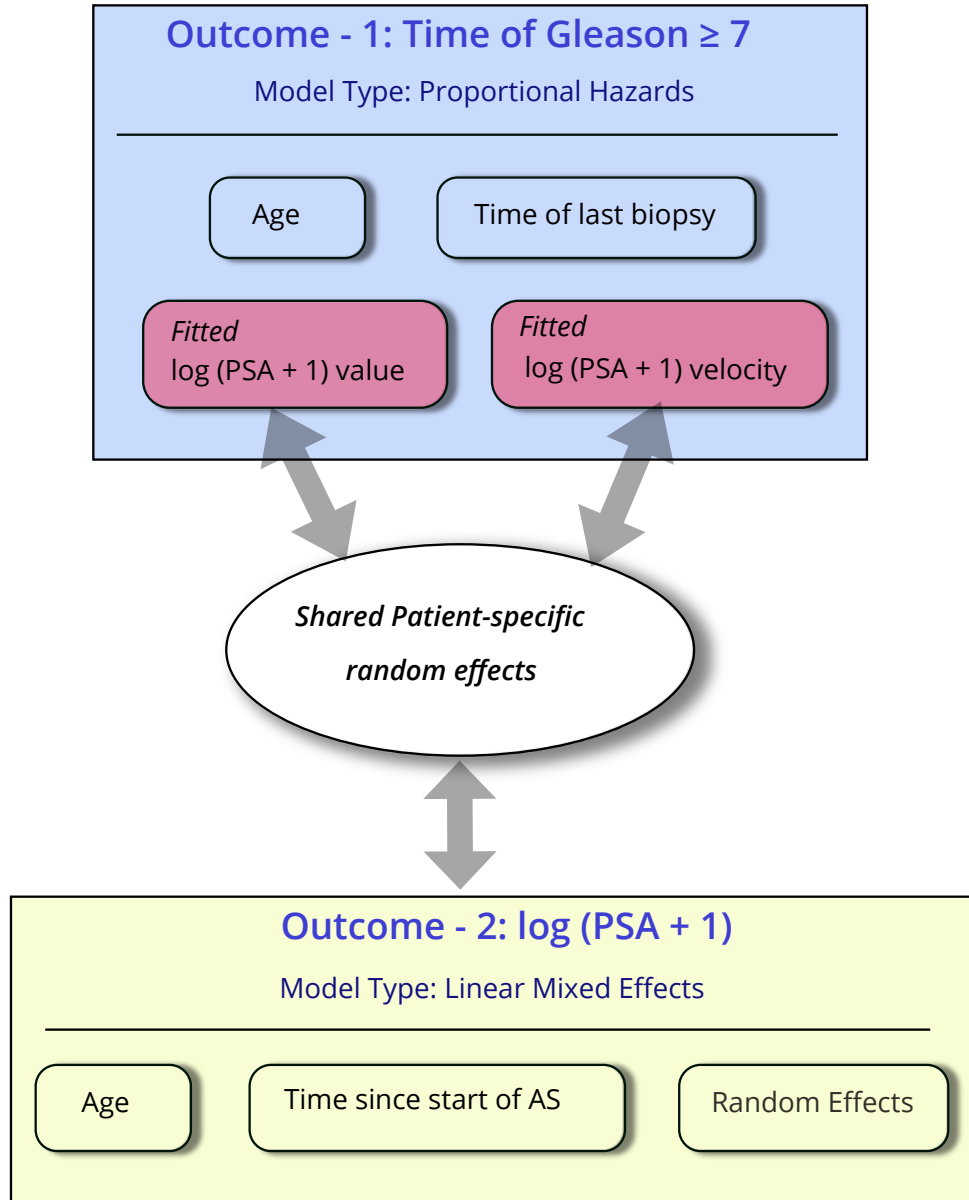


Figure 3: **Diagram of the joint model:** Patient-specific random effects (ellipse in center) are shared between sub-models for all outcomes pertaining to prostate cancer progression. The random effects model the correlation between the outcomes. In the linear mixed effects sub-model for $\log_2\{\text{PSA} + 1\}$ transformed PSA outcome (bottom rectangle), random effects are used as covariates. In the relative-risk sub-model (similar to Cox model) for time of Gleason ≥ 7 (GS7), random effects are utilized indirectly, by including fitted $\log_2\{\text{PSA} + 1\}$ value and velocities as covariates. Age of patient at baseline is included in both sub-models. Parameters of both sub-models are estimated jointly.

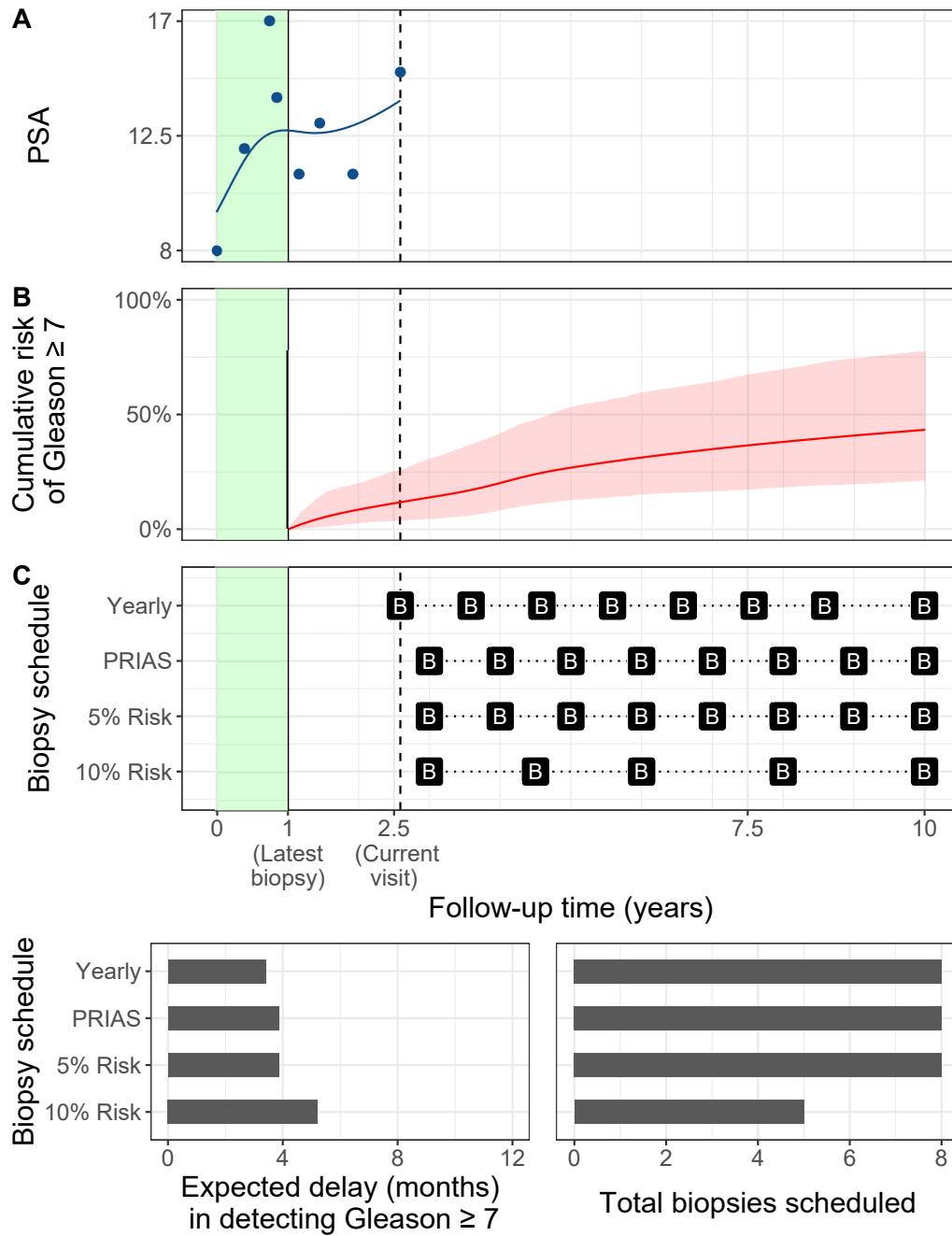


Figure 4: **Personalized and fixed schedules of biopsies for new patient.** **Panel A,B:** show the observed and fitted $\log_2(\text{PSA} + 1)$ measurements, and the dynamic cumulative-risk of Gleason ≥ 7 over the follow-up period. **Panel C** shows the personalized and fixed schedules of biopsies with a 'B' indicating the time of biopsy. In the bottom two panels, the various schedules are compared in terms of the number of biopsies they schedule, and the expected delay in detection of Gleason ≥ 7 if they are followed.

135 3. Results

136 For patients in the PRIAS dataset, the probability of obtaining reclassi-
 137 fication within the first five and ten years is 33% and 42%, respectively (see
 138 Figure 1). That is, ideally more than 50% of the patients do not require any
 139 biopsy in the first ten years. We next discuss the results from the joint model
 140 fitted to the PRIAS dataset. For every ten years increase in a patient age the
 141 corresponding adjusted hazard ratio of reclassification is 1.45 (95%CI: 1.30–
 142 1.63). For an increase in fitted $\log_2\{\text{PSA} + 1\}$ value from the first quartile
 143 of fitted values (2.36) to the third quartile (3.07), the corresponding adjusted
 144 hazard ratio of reclassification is 0.99 (95%CI: 0.89–1.11). On the other hand
 145 an increase in fitted $\log_2\{\text{PSA} + 1\}$ velocity from the first quartile of fitted
 146 velocity (-0.09) to the third quartile (0.31), the corresponding adjusted haz-
 147 ard ratio of reclassification is 2.47 (95%CI: 1.93–2.99). These results indicate
 148 that the velocity of $\log_2\{\text{PSA} + 1\}$ measurements is a stronger predictor of
 149 hazard of reclassification than the $\log_2\{\text{PSA} + 1\}$ value. Detailed parameter
 150 estimates are presented in Tables 2, 3 and 5 of Appendix A.4.

151 Using the joint model fitted to the PRIAS dataset we made risk predic-
 152 tions for GS7 in real PRIAS patients. As shown in Figure 4 of Appendix
 153 B, these risk estimates become more accurate as more data is gathered over
 154 follow-up. To check the accuracy of these risk predictions, we calculated
 155 the time dependent area under the receiver operating characteristic curves
 156 (AUC) as a measure of discrimination, and the root mean squared prediction
 157 error (RMSPE) as a measure of calibration. These are shown in Figure 5.
 158 For predictions within PRIAS (internal validation), the time-dependent AUC
 159 was between 0.62 and 0.69, and RMSPE between 0.23 and 0.37 over the

160 whole follow-up period. For validation in external cohorts, the AUC was
 161 similar to the AUC of PRIAS for all cohorts during the first three years of
 162 follow-up. The RMPSE however differed much more during the same pe-
 163 riod. The AS cohorts closest to PRIAS in terms of RMSPE were Johns
 164 Hopkins Active Surveillance and Memorial Sloan Kettering Cancer Center
 165 Active Surveillance. Detailed AUC and RMSPE results for all cohorts with
 166 95% bootstrapped confidence intervals are presented in Table 6 to Table 11
 167 of Appendix B.

168 Using the risk predictions for GS7, we developed personalized schedules of
 169 biopsy for real PRIAS patients. We maintained a minimum gap of one year
 170 between biopsies as advised by the PRIAS protocol. In addition, we sched-
 171 uled biopsies only for the first ten years follow-up because of limited follow-up
 172 period of the training dataset PRIAS. A compulsory biopsy was done sched-
 173 uled year ten of follow-up in all schedules for meaningful comparison of their
 174 expected delays in detection of GS7. Various personalized and fixed biopsy
 175 schedules for demo patients are shown in Figure 4 and Appendix C's Figure
 176 6, 7, 8 and 9. The biopsies denoted by 'B' show that personalized schedules
 177 schedule fewer biopsies than fixed schedules. At the same time the expected
 178 time delay in detection of GS7 is less than an year for personalized sched-
 179 ules. We have implemented this approach in a web-application ([https://](https://emcbiostatistics.shinyapps.io/prias_biopsy_recommender/)
 180 emcbiostatistics.shinyapps.io/prias_biopsy_recommender/, and Ap-
 181 pendix D) for practical use.

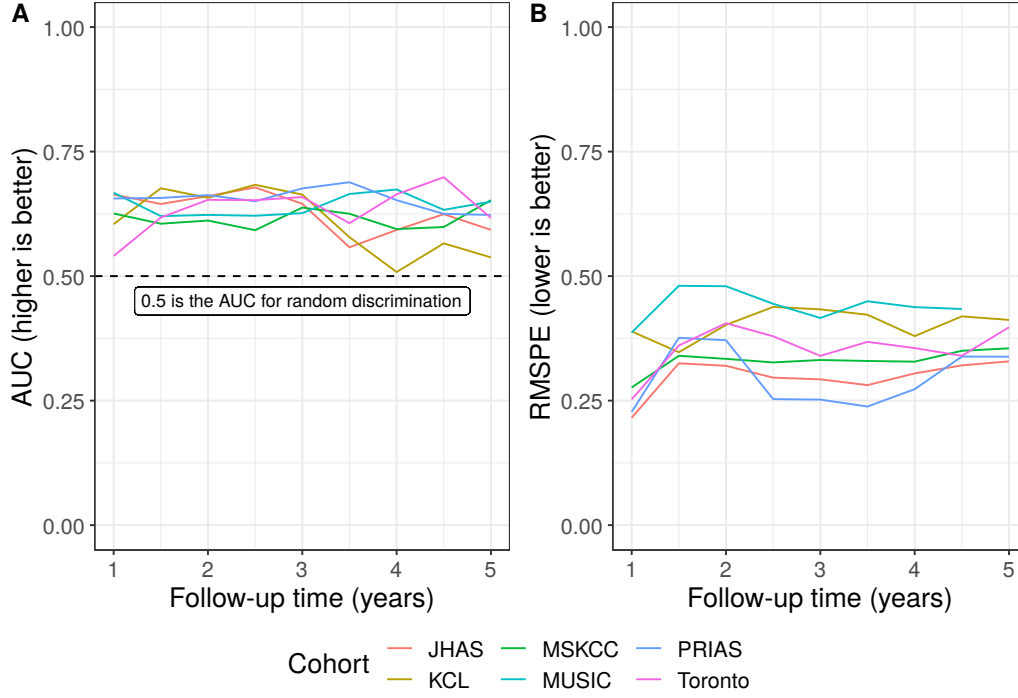


Figure 5: **Validation of predictions of Gleason ≥ 7 (GS7).** In **Panel A** we can see that the time dependent area under the receiver operating characteristic curve or AUC (measure of discrimination) is above 0.5 in PRIAS (internal validation), and in Toronto, JHAS, MSKCC, KCL, and MUSIC AS cohorts (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.

182 4. Discussion

183 We developed personalized schedules for repeat biopsies in PCa patients
 184 enrolled in AS programs. These schedules were based on a patient’s risk for
 185 having a Gleason ≥ 7 (GS7). Patient- and visit-specific risks of GS7 were
 186 estimated using their entire history of PSA and repeat biopsies, and baseline
 187 characteristics. Consequently, the personalized schedules were updated as
 188 more data was gathered over follow-up. Risk calculators for GS7 are not new
 189 [13, 18]. However, the novelty of our work is that we developed a methodology
 190 for scheduling personalized biopsies using those risks, and also a methodology
 191 to compare schedules, be it personalized or fixed, in simple terms of burden
 192 and benefit. More specifically, for each schedule we provided patients the
 193 times of biopsy and total biopsies (burden), and the time delay in detection of
 194 GS7 (less is beneficial) expected due to that schedule. We also implemented
 195 our methodology in a web-application.

196 The proposed joint model accounted for the complex correlation struc-
 197 ture that exists between longitudinal PSA measurements and time of GS7
 198 of a patient. It also accounted for PSA measurements that were missing in
 199 patients who obtained GS7. This model adjusts the risks of GS7 upon a neg-
 200 ative repeat biopsy. Thus complete patient information is consolidated into
 201 a single patient risk profile. Our model is fitted to the world’s largest PCa
 202 AS program, PRIAS. We also externally validated our model predictions for
 203 GS7 in five largest AS cohorts that are part of the GAP3 database [8]. We
 204 found that the AUC for predictions of GS7 over the follow-up period (Fig-
 205 ure 5) was similar in external cohorts and PRIAS (internal validation). The
 206 RMSPE however was similar to PRIAS only for Memorial Sloan Kettering

207 Cancer Center and Johns Hopkins cohorts. Given the large size of the latter
208 two cohorts, we expect that our model and the methodology will be useful
209 to a large number of AS patients. Extending our model and methodology in
210 other cohorts only requires fitting the model to their AS dataset.

211 The clinical implications of our work are as follows. The median survival
212 time for GS7 is more than ten years in PRIAS, and in some other cohorts
213 (Figure 1). That is, more than 50% of AS patients do not require any biopsy
214 during the first ten years of follow-up. We hope that our work will address
215 patient apprehensions regarding adverse outcomes in AS, in a more objective
216 manner. Many AS programs still utilize a rigorous schedule of yearly biopsies
217 [3]. However, with concerns about non-compliance and burden of biopsies
218 [5], the availability of our web based tool may encourage patients and doctors
219 to consider personalized schedules.

220 Our work has certain limitations. The proposed model is valid only for
221 the first ten years of follow-up in PRIAS, whereas GS7 may occur much later
222 in many patients. Due to this issue, we could schedule biopsies only for
223 the first ten years follow-up. In addition, the calibration and discrimination
224 measures of predictions were also less accurate in later follow-up periods.
225 These problems can be mitigated by refitting the model as more follow-up
226 data is gathered in PRIAS. While we focused only on GS7, it is susceptible to
227 inter-observer variation. Models which account for this variation [13, 19] will
228 be interesting to investigate further. However, the methodology to schedule
229 biopsies, and to estimate the consequences of following a schedule need not
230 change. There is also a potential for including diagnostic information from
231 novel biomarkers, quality of life measures, and magnetic resonance imaging

232 (MRI). Currently, this data is very sparsely available in the PRIAS dataset.
233 However, in future, adding this information in our model is trivial. This is
234 because modeling correlation for extra outcomes, mainly entails connecting
235 sub-models for the outcomes to shared random effects (see Figure 3). Our
236 model can also be used to schedule MRI scans, since they are expensive in
237 developing countries.

238 5. Conclusions

239 We developed a novel methodology for scheduling biopsies to detect Glea-
240 son ≥ 7 (GS7) in PCa patients enrolled in AS. Our methodology consolidates
241 a patient's entire history of PSA and repeat biopsies, and baseline character-
242 istics into risk profile of GS7 over his follow-up period. It then utilizes this
243 risk profile to schedule biopsies in a personalized manner. The personalized
244 schedule is updated as more patient data is gathered over follow-up. To as-
245 sist patients in making the choice of the best biopsy schedule, we provided
246 them personalized burden (time and total biopsies), and personalized bene-
247 fit (less time delay in detection of GS7 is beneficial), for both personalized
248 and currently used schedules. Lastly, we implemented this approach in a
249 web-application.

250 Acknowledgments

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

- 257 1. Briganti A, Fossati N, Catto JW, Cornford P, Montorsi F, Mottet N,
 258 Wirth M, Van Poppel H. Active surveillance for low-risk prostate cancer:
 259 the European Association of Urology position in 2018. *European urology*
 260 2018;74(3):357–68.
- 261 2. Bul M, Zhu X, Valdagni R, Pickles T, Kakehi Y, Rannikko A, Bjartell A,
 262 Van Der Schoot DK, Cornel EB, Conti GN, et al. Active surveillance for
 263 low-risk prostate cancer worldwide: the prias study. *European urology*
 264 2013;63(4):597–603.
- 265 3. Nieboer D, Tomer A, Rizopoulos D, Roobol MJ, Steyerberg EW. Active
 266 surveillance: a review of risk-based, dynamic monitoring. *Translational*
 267 *andrology and urology* 2018;7(1):106–15.
- 268 4. Loeb S, Carter HB, Schwartz M, Fagerlin A, Braithwaite RS, Lepor H.
 269 Heterogeneity in active surveillance protocols worldwide. *Reviews in*
 270 *urology* 2014;16(4):202–3.
- 271 5. Bokhorst LP, Alberts AR, Rannikko A, Valdagni R, Pickles T, Kakehi Y,
 272 Bangma CH, Roobol MJ, PRIAS study group . Compliance rates with

- 273 the Prostate Cancer Research International Active Surveillance (PRIAS)
274 protocol and disease reclassification in noncompliers. *European Urology*
275 2015;68(5):814–21.
- 276 6. Inoue LY, Lin DW, Newcomb LF, Leonardson AS, Ankerst D, Gulati R,
277 Carter HB, Trock BJ, Carroll PR, Cooperberg MR, et al. Comparative
278 analysis of biopsy upgrading in four prostate cancer active surveillance
279 cohorts. *Annals of internal medicine* 2018;168(1):1–9.
- 280 7. Bratt O, Carlsson S, Holmberg E, Holmberg L, Johansson E, Josefsson
281 A, Nilsson A, Nyberg M, Robinson D, Sandberg J, et al. The study of
282 active monitoring in sweden (sams): a randomized study comparing two
283 different follow-up schedules for active surveillance of low-risk prostate
284 cancer. *Scandinavian journal of urology* 2013;47(5):347–55.
- 285 8. Bruinsma SM, Zhang L, Roobol MJ, Bangma CH, Steyerberg EW,
286 Nieboer D, Van Hemelrijck M, consortium MFGAPPCASG, Trock B,
287 Ehdaie B, et al. The movember foundation’s gap3 cohort: a profile of
288 the largest global prostate cancer active surveillance database to date.
289 *BJU international* 2018;121(5):737–44.
- 290 9. de Carvalho TM, Heijnsdijk EA, de Koning HJ. Estimating the risks
291 and benefits of active surveillance protocols for prostate cancer: a mi-
292 crosimulation study. *BJU international* 2017;119(4):560–6.
- 293 10. Partin AW, Yoo J, Carter HB, Pearson JD, Chan DW, Epstein JI, Walsh
294 PC. The use of prostate specific antigen, clinical stage and gleason score

- 295 to predict pathological stage in men with localized prostate cancer. *The*
 296 *Journal of urology* 1993;150(1):110–4.
- 297 11. Makarov DV, Trock BJ, Humphreys EB, Mangold LA, Walsh PC, Ep-
 298 stein JI, Partin AW. Updated nomogram to predict pathologic stage of
 299 prostate cancer given prostate-specific antigen level, clinical stage, and
 300 biopsy gleason score (partin tables) based on cases from 2000 to 2005.
 301 *Urology* 2007;69(6):1095–101.
- 302 12. Tomer A, Nieboer D, Roobol MJ, Steyerberg EW, Rizopoulos D. Per-
 303 sonalized schedules for surveillance of low-risk prostate cancer patients.
 304 *Biometrics* 2019;75(1):153–62.
- 305 13. Coley RY, Zeger SL, Mamawala M, Pienta KJ, Carter HB. Prediction
 306 of the pathologic gleason score to inform a personalized management
 307 program for prostate cancer. *European urology* 2017;72(1):135–41.
- 308 14. Rizopoulos D. Joint Models for Longitudinal and Time-to-Event Data:
 309 With Applications in R. CRC Press; 2012. ISBN 9781439872864.
- 310 15. Laird NM, Ware JH, et al. Random-effects models for longitudinal data.
 311 *Biometrics* 1982;38(4):963–74.
- 312 16. Rizopoulos D. The R package JMBayes for fitting joint models for lon-
 313 gitudinal and time-to-event data using MCMC. *Journal of Statistical*
 314 *Software* 2016;72(7):1–46.
- 315 17. Rizopoulos D, Molenberghs G, Lesaffre EM. Dynamic predictions with
 316 time-dependent covariates in survival analysis using joint modeling and
 317 landmarking. *Biometrical Journal* 2017;59(6):1261–76.

- 318 18. Ankerst DP, Xia J, Thompson Jr IM, Hoeffler J, Newcomb LF, Brooks
319 JD, Carroll PR, Ellis WJ, Gleave ME, Lance RS, et al. Precision
320 medicine in active surveillance for prostate cancer: development of the
321 canary–early detection research network active surveillance biopsy risk
322 calculator. *European urology* 2015;68(6):1083–8.
- 323 19. Balasubramanian R, Lagakos SW. Estimation of a failure time distribu-
324 tion based on imperfect diagnostic tests. *Biometrika* 2003;90(1):171–82.