Personalized Decision Making for Biopsies in Prostate Cancer Active Surveillance Programs

Anirudh Tomer[1,*], Dimitris Rizopoulos[1], Daan Nieboer[2], Frank-Jan Drost[3],
Monique J. Roobol[3], and Ewout W. Steyerberg[2,4]

[1]Department of Biostatistics, Erasmus University Medical Center, the Netherlands
[2]Department of Public Health, Erasmus University Medical Center, the Netherlands
[3]Department of Urology, Erasmus University Medical Center, the Netherlands
[4]Department of Biomedical Data Sciences, Leiden University Medical Center, the Netherlands
*email: a.tomer@erasmusmc.nl

Abstract

Background. Low-risk prostate cancer patients enrolled in active surveillance programs commonly undergo biopsies for examination of cancer progression. Biopsies are conducted as per a fixed and frequent schedule (e.g., annual biopsies). Since biopsies are burdensome, patients do not always comply with the schedule, which increases the risk of delayed detection of cancer progression.

Objective. Our aim is to better balance the number of biopsies (burden) and the delay in detection of cancer progression (less is beneficial), by personalizing the decision of conducting biopsies.

Data Sources. We use patient data of the world's largest active surveillance program (PRIAS). It enrolled 5270 patients, had 866 cancer progressions, and an average of nine prostate-specific antigen (PSA) and five digital rectal examination (DRE) measurements per patient.

Methods. Using joint models for time-to-event and longitudinal data, we model the historical DRE and PSA measurements, and biopsy results of a patient at each follow-up visit. This results in a visit and patient-specific cumulative risk of cancer progression. If this

risk is above a certain threshold, we schedule a biopsy. We compare this personalized approach with the currently practiced biopsy schedules via an extensive and realistic simulation study, based on a replica of the patients from the PRIAS program.

Results. The personalized approach saved a median of six biopsies (median: 4, IQR: 2–5), compared to the annual schedule (median: 10, IQR: 3–10). However, the delay in detection of progression (years) is similar for the personalized (median: 0.7, IQR: 0.3–1.0) and the annual schedule (median: 0.5, IQR: 0.3–0.8).

Conclusions. We conclude that personalized schedules provide substantially better balance in the number of biopsies per detected progression for men with low-risk prostate cancer.

1 Introduction

Prostate cancer is the second most frequently diagnosed cancer in men worldwide (Torre et al., 2015). In prostate cancer screening programs, many of the diagnosed tumors are clinically insignificant (over-diagnosed) (Etzioni et al., 2002). To avoid further over-treatment, patients diagnosed with low-grade prostate cancer are commonly advised to join active surveillance (AS) programs. In AS, invasive treatments such as surgery are delayed until cancer progresses. Cancer progression is routinely monitored via serum prostate-specific antigen (PSA) measurements, a protein biomarker; digital rectal examination (DRE) measurements, a measure of the size and location of the tumor; and biopsies.

While larger values for PSA and/or DRE, may indicate cancer progression, biopsies are the most reliable cancer progression examination technique used in AS. When a patient's biopsy Gleason score becomes larger than 6 (positive biopsy, cancer progression detected), AS is stopped and the patient is advised treatment (Bokhorst et al., 2015). However, biopsies are invasive, painful, and prone to medical complications (Ehdaie et al., 2014; Fujita et al., 2009). Hence, they are conducted intermittently until a positive biopsy. Consequently, at the time of a positive biopsy, cancer progression may be observed with a delay of unknown duration. This delay is defined as the difference between the time of the positive biopsy and the unobserved true time of cancer progression. Thus, the decision to conduct biopsies requires a compromise between the burden of biopsy and the potential delay in the detection of cancer progression.

In AS, a delay in the detection of cancer progression around 12 to 14 months is assumed to be unlikely to substantially increase the risk of adverse downstream outcomes (Carvalho, Heijnsdijk, and Koning, 2017; Inoue et al.,

2018). However, for biopsies, there is little consensus on the time gap between them (Bruinsma et al., 2016; Loeb et al., 2014; Nieboer et al., 2018). Many AS programs focus on minimizing the delay in the detection of cancer progression, by scheduling biopsies annually for all patients. A drawback of annual biopsies, and other currently practiced fixed/heuristic schedules (Bruinsma et al., 2016; Loeb et al., 2014; Nieboer et al., 2018), is that they ignore the large variation in the time of cancer progression of AS patients. While they may work well for patients who progress early (*fast progressing*) in AS, but for a large proportion of patients who do not progress, or progress late (*slow progressing*) in AS, many unnecessary burdensome biopsies are scheduled. To mediate the burden between the *fast* and *slow progressing* patients, the world's largest AS program, Prostate Cancer Research International Active Surveillance (Bokhorst et al., 2016) (PRIAS), schedules annual biopsies only for patients with a low PSA doubling time (Bokhorst et al., 2015). For everyone else, PRIAS schedules biopsies at following fixed follow-up times: year one, four, seven, and ten, and every five years thereafter. Despite this effort in PRIAS, patients may get scheduled for four to ten biopsies over a period of ten years. Therefore, compliance for biopsies is low in PRIAS (Bokhorst et al., 2015). This can lead to a delay in the detection of cancer progression, and reduce the effectiveness of AS.

We aim to better balance the number of biopsies (more are burdensome), and the delay in the detection of cancer progression (less is beneficial), than currently practiced schedules. We intend to achieve this by personalizing the decision to conduct biopsies (see Figure 1). These decisions are made at a patient's pre-scheduled follow-up visits for DRE and PSA measurements. To develop the personalized decision making methodology, we utilize the data of the patients enrolled in the PRIAS study. We model this data and develop the personalized approach using joint models for time-to-event and longitudinal data (Rizopoulos, 2012; Tsiatis and Davidian, 2004). In order to compare the personalized approach with current schedules, we conduct an extensive simulation study based on a replica of the patients from the PRIAS program.

2 Methods

2.1 Study Population

To develop our methodology we use the data of prostate cancer patients from the world's largest AS study called PRIAS (Bokhorst et al., 2016) (see Table 1). More than 100 medical centers from 17 countries worldwide con-
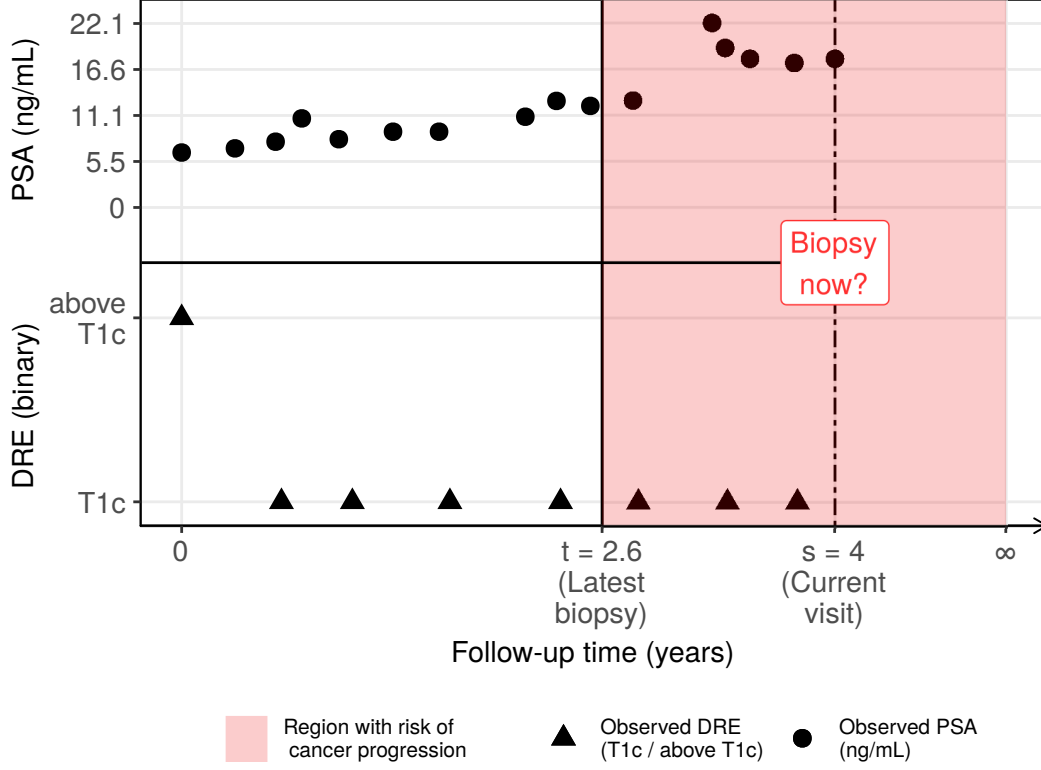
Figure 1: **The personalized decision making problem:** Available data of a patient j, who had his latest negative biopsy at $t = 2.6$ years. The shaded region shows the time period in which the patient is at risk of cancer progression. His current pre-scheduled follow-up visit for measurement of DRE and PSA is at $s = 4$ years. Using his entire history of DRE $\mathcal{Y}_{dj}(s)$ and PSA $\mathcal{Y}_{pj}(s)$ measurements up to the current visit s, and the time of the latest biopsy t, we intend to make a decision on scheduling a biopsy at the current visit.

tribute to the collection of data, utilizing a common study protocol and a web-based tool, both available at www.prias-project.org. We use data collected over a period of ten years, between December 2006 (beginning of PRIAS study) and December 2016. The primary event of interest is cancer progression detected upon a positive biopsy. The time of cancer progression is interval censored because biopsies are scheduled periodically. Biopsies are scheduled as per the PRIAS protocol (see Introduction). There are three types of competing events, namely death, removal of patients from AS on the basis of their observed DRE and PSA measurements, and loss to follow-up. We assume these three types of events to be censored observations (see

4

Table 1: **Summary statistics for the PRIAS dataset**. The primary event of interest is cancer progression. A DRE measurement equal to T1cSchröder et al., 1992 indicates a clinically inapparent tumor which is not palpable or visible by imaging, while tumors with DRE > T1c are palpable. The abbreviation IQR means interquartile range.

Data	Value
Total patients	5270
Cancer progression (primary event)	866
Loss to follow-up (anxiety or unknown)	685
Removal on the basis of PSA and DRE	464
Death (unrelated to prostate cancer)	61
Death (related to prostate cancer)	2
Median Age (years)	70 (IQR: 65–75)
Total PSA measurements	46015
Median number of PSA per patient	7 (IQR: 7–12)
Median PSA value (ng/mL)	5.6 (IQR: 4.0–7.5)
Total DRE measurements	25606
Median number of DRE per patient	4 (IQR: 3–7)
DRE = T1c (%)	23538/25606 (92%)

Appendix A.5 for details). However, our model allows removal of patients to depend on observed longitudinal data and baseline covariates of the patient. Under the aforementioned assumption of censoring, Figure 2 shows the cumulative risk of cancer progression over the study follow-up period.

For all patients, PSA measurements (ng/mL) are scheduled every 3 months for the first 2 years and every 6 months thereafter. The DRE measurements are scheduled every 6 months. We use the DRE measurements as DRE = T1c versus DRE > T1c. A DRE measurement equal to T1c (Schröder et al., 1992) indicates a clinically inapparent tumor which is not palpable or visible by imaging, while tumors with DRE > T1c are palpable.

Data Accessibility: The PRIAS database is not openly accessible. However, access to the database can be requested on the basis of a study proposal approved by the PRIAS steering committee. The website of the PRIAS program is www.prias-project.org.
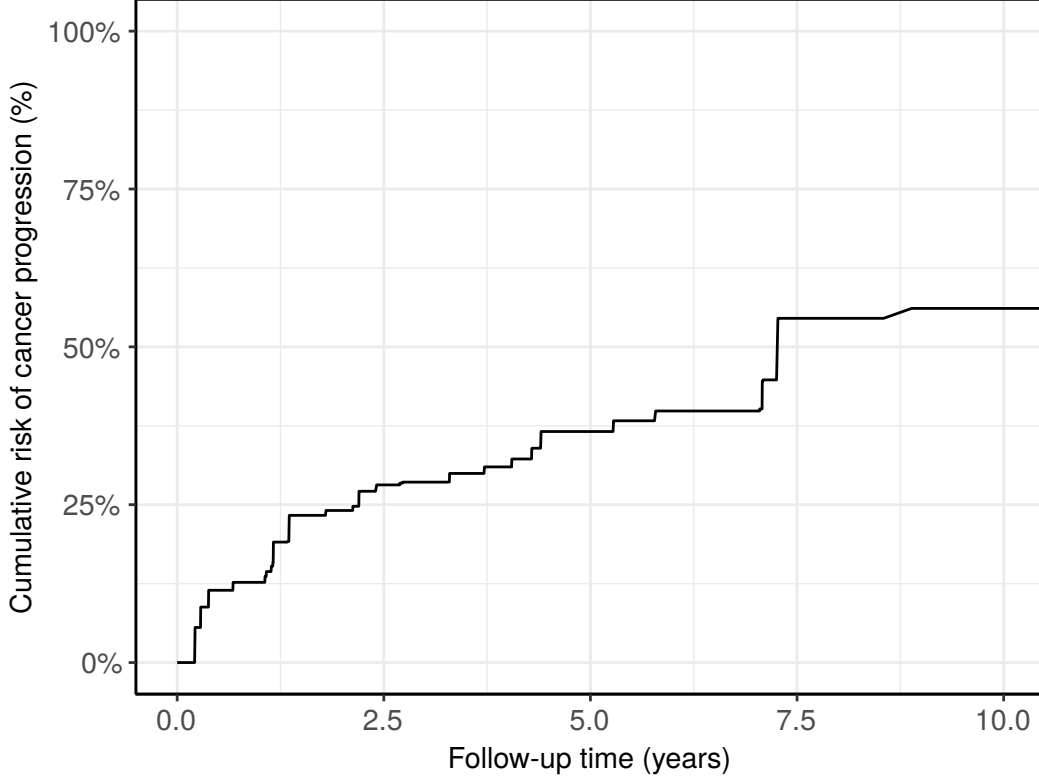
Figure 2: **Estimated cumulative risk of cancer progression in AS** for patients in the Prostate Cancer Research International Active Surveillance (PRIAS) dataset. Nearly 50% patients (*slow progressing*) do not progress in the ten year follow-up period. Cumulative risk is estimated using non-parametric maximum likelihood estimation (Turnbull, 1976), to account for interval censored cancer progression times observed in the PRIAS dataset. Censoring includes death, removal from AS on the basis of observed longitudinal data, and patient dropout.

2.2 A Bivariate Joint Model for the Longitudinal PSA, and DRE Measurements, and Time of Cancer Progression

Let T_i^* denote the true cancer progression time of the i-th patient included in PRIAS. Since biopsies are conducted periodically, T_i^* is observed with interval censoring $l_i < T_i^* \leq r_i$. When progression is observed for the patient at his latest biopsy time r_i, then l_i denotes the time of the second latest biopsy. Otherwise, l_i denotes the time of the latest biopsy and $r_i = \infty$. Let $\boldsymbol{y}_{di}$ and $\boldsymbol{y}_{pi}$ denote his observed DRE and PSA longitudinal measure-

ments, respectively. The observed data of all n patients is denoted by $\mathcal{D}_n = \{l_i, r_i, \boldsymbol{y}_{di}, \boldsymbol{y}_{pi}; i = 1, \ldots, n\}$.

In our joint model, the patient-specific DRE and PSA measurements over time are modeled using a bivariate generalized linear mixed effects sub-model. The sub-model for DRE is given by (see Panel A, Figure 3):

$$
\begin{aligned}
\text{logit}\big[\text{Pr}\{y_{di}(t) > \text{T1c}\}\big] = {} & \beta_{0d} + b_{0di} + (\beta_{1d} + b_{1di})t \\
& + \beta_{2d}(\text{Age}_i - 70) \\
& + \beta_{3d}(\text{Age}_i - 70)^2
\end{aligned}
\tag{1}
$$

where, t denotes the follow-up visit time, and Age_i is the age of the i-th patient at the time of inclusion in AS. We have centered the Age variable around the median age of 70 years for better convergence during parameter estimation. However, this does not change the interpretation of the parameters corresponding to the Age variable. The fixed effect parameters are denoted by $\{\beta_{0d}, \ldots, \beta_{3d}\}$, and $\{b_{0di}, b_{1di}\}$ are the patient specific random effects. With this definition, we assume that the patient-specific log odds of obtaining a DRE measurement larger than T1c remain linear over time.

The mixed effects sub-model for PSA is given by (see Panel B, Figure 3):

$$
\begin{aligned}
\log_2\big\{y_{pi}(t) + 1\big\} = {} & m_{pi}(t) + \varepsilon_{pi}(t), \\
m_{pi}(t) = {} & \beta_{0p} + b_{0pi} + \sum_{k=1}^{4}(\beta_{kp} + b_{kpi})B_k(t, \mathcal{K}) \\
& + \beta_{5p}(\text{Age}_i - 70) + \beta_{6p}(\text{Age}_i - 70)^2,
\end{aligned}
\tag{2}
$$

where, $m_{pi}(t)$ denotes the underlying measurement error free value of $\log_2(\text{PSA} + 1)$ transformed (Lin et al., 2000; Pearson et al., 1994) measurements at time t. We model it non-linearly over time using B-splines (De Boor, 1978). To this end, our B-spline basis function $B_k(t, \mathcal{K})$ has 3 internal knots at $\mathcal{K} = \{0.1, 0.7, 4\}$ years, and boundary knots at 0 and 5.42 years (95-th percentile of the observed follow-up times). This specification allows fitting the $\log_2(\text{PSA} + 1)$ levels in a piecewise manner for each patient separately. The internal and boundary knots specify the different time periods (analogously pieces) of this piecewise nonlinear curve. The fixed effect parameters are denoted by $\{\beta_{0p}, \ldots, \beta_{6p}\}$, and $\{b_{0pi}, \ldots, b_{4pi}\}$ are the patient specific random effects. The error $\varepsilon_{pi}(t)$ is assumed to be t-distributed with three degrees of freedom (see Appendix B.1) and scale σ, and is independent of the random effects.

To account for the correlation between the DRE and PSA measurements of a patient, we link their corresponding random effects. More specifically, the

complete vector of random effects $\boldsymbol{b}_i = (b_{0di}, b_{1di}, b_{0pi}, \ldots, b_{4pi})^T$ is assumed to follow a multivariate normal distribution with mean zero and variance-covariance matrix $\boldsymbol{D}$.

To model the impact of DRE and PSA measurements on the risk of cancer progression, our joint model uses a relative risk sub-model. More specifically, the hazard of cancer progression $h_i(t)$ at a time t is given by (see Panel D, Figure 3):

$$
\begin{aligned}
h_i(t) = h_0(t) \exp \Big(&\gamma_1(\text{Age}_i - 70) + \gamma_2(\text{Age}_i - 70)^2 \\
&+ \alpha_{1d}\text{logit}\big[\text{Pr}\{y_{di}(t) > \text{T1c}\}\big] \\
&+ \alpha_{1p}m_{pi}(t) + \alpha_{2p}\frac{\partial m_{pi}(t)}{\partial t}\Big),
\end{aligned} \tag{3}
$$

where, γ_1, γ_2 are the parameters for the effect of age. The parameter α_{1d} models the impact of log odds of obtaining a DRE > T1c on the hazard of cancer progression. The impact of PSA on the hazard of cancer progression is modeled in two ways: a) the impact of the error free underlying PSA value $m_{pi}(t)$ (see Panel B, Figure 3), and b) the impact of the underlying PSA velocity $\partial m_{pi}(t)/\partial t$ (see Panel C, Figure 3). The corresponding parameters are α_{1p} and α_{2p}, respectively. Lastly, $h_0(t)$ is the baseline hazard at time t, and is modeled flexibly using P-splines (Eilers and Marx, 1996). The detailed specification of the baseline hazard $h_0(t)$, and the joint parameter estimation of the two sub-models using the Bayesian approach (R package **JMbayes**, see Rizopoulos, 2016) are presented in Appendix A of the supplementary material.

2.3   Personalized Decisions for Biopsy

Let us assume that a decision of conducting a biopsy is to be made for a new patient j shown in Figure 1, at his current follow-up visit time s. Let $t \le s$ be the time of his latest negative biopsy. Let $\mathcal{Y}_{dj}(s)$ and $\mathcal{Y}_{pj}(s)$ denote his observed DRE and PSA measurements up to the current visit, respectively. From the observed measurements we want to extract the underlying measurement error free trend of $\log_2(\text{PSA} + 1)$ values and velocity, and the log odds of obtaining DRE > T1c. We intend to combine them to inform us when the cancer progression is to be expected (see Figure 4), and to further guide the decision making on whether to conduct a biopsy at the current follow-up visit. The combined information is given by the following posterior predictive distribution $g(T_j^*)$ of his time of cancer progression $T_j^* > t$ (see Appendix A.4 for details):

$$
g(T_j^*) = p\big\{T_j^* \mid T_j^* > t, \mathcal{Y}_{dj}(s), \mathcal{Y}_{pj}(s), \mathcal{D}_n\big\}. \tag{4}
$$

8

The distribution $g(T_j^*)$ is not only patient-specific, but also updates as extra information is recorded at future follow-up visits.

A key ingredient in the decision of conducting a biopsy for patient j at the current follow-up visit time s is the personalized cumulative risk of observing a cancer progression at time s (illustrated in Figure 4). This risk can be derived from the posterior predictive distribution $g(T_j^*)$ (Rizopoulos, 2011), and for $s \geq t$ it is given by:

$$R_j(s \mid t) = \Pr\{T_j^* \leq s \mid T_j^* > t, \mathcal{Y}_{dj}(s), \mathcal{Y}_{pj}(s), \mathcal{D}_n\}. \qquad (5)$$

A simple and straightforward approach to decide upon conducting a biopsy for patient j at the current follow-up visit would be to do so if his personalized cumulative risk of cancer progression at the visit is higher than a certain threshold $0 \leq \kappa \leq 1$. For example, as shown in Panel B of Figure 4, biopsy at a visit may be scheduled if the personalized cumulative risk is higher than 10% (example risk threshold). This decision making process is iterated over the follow-up period, incorporating on each subsequent visit the newly observed data, until a positive biopsy is observed. Subsequently, an entire personalized schedule of biopsies for each patient can be obtained.

The choice of the risk threshold dictates the schedule of biopsies and has to be made on each subsequent follow-up visit of a patient. In this regard, a straightforward approach is choosing a fixed risk threshold, such as 5% or 10% risk, at all follow-up visits. Fixed risk thresholds may be chosen by patients and/or doctors according to how they weigh the relative harms of doing an unnecessary biopsy versus a missed cancer progression (e.g., 10% threshold means a 1:9 ratio) if the biopsy is not conducted (Vickers and Elkin, 2006). An alternative approach is that at each follow-up visit a unique threshold is chosen on the basis of its classification accuracy. More specifically, given the time of latest biopsy t of patient j, and his current visit time s we find a visit-specific biopsy threshold κ, which gives the highest cancer progression detection rate (true positive rate, or TPR) for the period $(t, s]$. However, we also intend to balance for unnecessary biopsies (high false positive rate), or a low number of correct detections (high false negative rate) when the false positive rate is minimized. An approach to mitigating these issues is to maximize the TPR and positive predictive value (PPV) simultaneously. To this end, we utilize the F_1 score, which is a composite of both TPR and PPV (estimated as in Rizopoulos et al., 2017 Rizopoulos, Molenberghs, and

Lesaffre, 2017), and is defined as:

$$
\begin{aligned}
\mathrm{F}_1(t, s, \kappa) &= 2\frac{\mathrm{TPR}(t, s, \kappa)\,\mathrm{PPV}(t, s, \kappa)}{\mathrm{TPR}(t, s, \kappa) + \mathrm{PPV}(t, s, \kappa)}, \\
\mathrm{TPR}(t, s, \kappa) &= \mathrm{Pr}\big\{R_j(s \mid t) > \kappa \mid t < T_j^* \le s\big\}, \\
\mathrm{PPV}(t, s, \kappa) &= \mathrm{Pr}\big\{t < T_j^* \le s \mid R_j(s \mid t) > \kappa\big\},
\end{aligned} \tag{6}
$$

where, $\mathrm{TPR}(t, s, \kappa)$ and $\mathrm{PPV}(t, s, \kappa)$ are the time dependent true positive rate and positive predictive value, respectively. These values are unique for each combination of the time period $(t, s]$ and the risk threshold κ that is used to discriminate between the patients whose cancer progresses in this time period versus the patients whose cancer does not progress. The same holds true for the resulting F_1 score denoted by $\mathrm{F}_1(t, s, \kappa)$. The F_1 score ranges between 0 and 1, where a value equal to 1 indicates perfect TPR and PPV. Thus the highest F_1 score is desired in each time period $(t, s]$. This can be achieved by choosing a risk threshold κ which maximizes $\mathrm{F}_1(t, s, \kappa)$. That is, during a patient's visit at time s, given that his latest biopsy was at time t, the visit-specific risk threshold to decide a biopsy is given by $\kappa = \arg\max_\kappa \mathrm{F}_1(t, s, \kappa)$. The criteria on which we evaluate the personalized schedules based on fixed and visit-specific risk thresholds is the total number of biopsies scheduled, and the delay in detection of cancer progression (details in Results).

2.4 Simulation Study

Although the personalized decision making approach is motivated by the PRIAS study, it is not possible to evaluate it directly on the PRIAS dataset. This is because the patients in PRIAS have already had their biopsies as per the PRIAS protocol. In addition, the true time of cancer progression is interval or right censored for all patients, making it impossible to correctly estimate the delay in detection of cancer progression due to a particular schedule. To this end, we conduct an extensive simulation study to find the utility of personalized, PRIAS, and fixed/heuristic schedules. For a realistic comparison, we simulate patient data from the joint model fitted to the PRIAS dataset. The simulated population has the same ten year follow-up period as the PRIAS study. In addition, the estimated relations between DRE and PSA measurements, and the risk of cancer progression are retained in the simulated population.

From this population, we first sample 500 datasets, each representing a hypothetical AS program with 1000 patients in it. We generate a true cancer progression time for each of the 500×1000 patients and then sample a set of DRE and PSA measurements at the same follow-up visit times as given

10

in PRIAS protocol. We then split each dataset into training (750 patients) and test (250 patients) parts, and generate a random and noninformative censoring time for the training patients. We next fit a joint model of the specification given in Equations (1), (2), and (3) to each of the 500 training datasets and obtain MCMC samples from the 500 sets of the posterior distribution of the parameters.

In each of the 500 hypothetical AS programs, we utilize the corresponding fitted joint models to develop cancer progression risk profiles for each of the 500×250 test patients. We make the decision of biopsies for patients at their pre-scheduled follow-up visits for DRE and PSA measurements (see Study Population), on the basis of their estimated personalized cumulative risk of cancer progression. These decisions are made iteratively until a positive biopsy is observed. A recommended gap of one year between consecutive biopsies (Bokhorst et al., 2015) is also maintained. Subsequently, for each patient, an entire personalized schedule of biopsies is obtained.

We evaluate and compare both personalized and currently practiced schedules of biopsies in this simulation study. Comparison of the schedules is based on the number of biopsies scheduled and the corresponding delay in the detection of cancer progression. We evaluate the following currently practiced fixed/heuristic schedules: biopsy annually, biopsy every one and a half years, biopsy every two years and biopsy every three years. We also evaluate the biopsy schedule of the PRIAS program (see Introduction). For the personalized biopsy schedules, we evaluate schedules based on three fixed risk thresholds: 5%, 10%, and 15%, corresponding to a missed cancer progression being 19, 9, and 5.5 times more harmful than an unnecessary biopsy (Vickers and Elkin, 2006), respectively. We also implement a personalized schedule where for each patient, visit-specific risk thresholds are chosen using F_1 score.

3   Results

From the joint model fitted to the PRIAS dataset, we found that both $\log_2\{PSA + 1\}$ velocity, and log odds of having DRE > T1c were significantly associated with the hazard of cancer progression. For any patient, an increase in $\log_2\{PSA + 1\}$ velocity from -0.03 to 0.16 (first and third quartiles of the fitted velocities, respectively) corresponds to a 1.94 fold increase in the hazard of cancer progression. Whereas, an increase in odds of DRE > T1c from -6.650 to -4.356 (first and third quartiles of the fitted log odds, respectively) corresponds to a 1.40 fold increase in the hazard of cancer progression. Detailed results pertaining to the fitted joint model are presented in Appendix B.

3.1 Comparison of Various Approaches for Biopsies

From the simulation study, we obtain the number of biopsies and the delay in detection of cancer progression for each of the 500×250 test patients using different schedules. Figure 5 shows that the personalized and PRIAS approaches fall in the region of better balance between the median number of biopsies and the median delay than fixed/heuristic schedules. We next evaluate these schedules on the basis of both median and interquartile range (IQR) of the number of biopsies and delay (see Figure 6). For brevity, only the most widely used annual and PRIAS schedules, the proposed personalized approach with fixed risk thresholds of 5% and 10%, and visit-specific threshold chosen using F_1 score are discussed next (see Appendix C for remaining).

Since patients have varying cancer progression speeds, the impact of each schedule also varies with it. In order to highlight these differences, we divide results for three types of patients, as per their time of cancer progression. They are *fast, intermediate,* and *slow progressing* patients. Although such a division may be imperfect and can only be done retrospectively in a simulation setting, we show results for these three groups for the purpose of illustration. Roughly 50% of the patients did not obtain cancer progression in the ten year follow-up period of the simulation study. We assume these patients to be *slow progressing* patients. We assume *fast progressing* patients are the ones with an initially misdiagnosed state of cancer (Cooperberg et al., 2011) or high-risk patients who choose AS instead of immediate treatment upon diagnosis. These are roughly 30% of the population, having a cancer progression time less than 3.5 years. We label the remaining 20% patients as *intermediate progressing* patients.

For *fast progressing* patients (Panel A, Figure 6), we note that the personalized schedules with a fixed 10% risk threshold and visit-specific threshold chosen using F_1 score, reduce one biopsy for 50% of the patients, compared to PRIAS and annual schedule. Despite this, the delay (years) is similar for the personalized schedule with fixed 10% risk threshold (median: 0.7, IQR: 0.3–1.0), and the commonly used annual (median: 0.6, IQR: 0.3–0.9) and PRIAS (median: 0.7, IQR: 0.3–1.0) schedules.

For *intermediate progressing* patients (Panel A, Figure 6), we note that the delay (years) due to personalized schedule with fixed 5% risk threshold (median: 0.6, IQR: 0.3–0.9) is comparable to that of annual schedule (median 0.5, IQR: 0.2–0.7). However, it schedules fewer biopsies (median: 6, IQR: 5–7) than the annual schedule (median: 7, IQR: 5–8). The delay (years) for PRIAS (median: 0.7, IQR: 0.3–1.3) and personalized schedule with fixed 10% risk (median: 0.7, IQR: 0.4–1.3) are similar, but the personalized approach

schedules one less biopsy for 50% of the patients. Although the approach with visit-specific risk threshold chosen using F_1 score schedules fewer biopsies than the 10% fixed risk approach, it also has a higher delay.

The patients who are at the most advantage with the personalized schedules are the *slow progressing* patients. These are a total of 50% patients who did not progress during the entire study. Hence, the delay is not available for these patients (Panel C of Figure 6). For all of these patients, annual schedule leads to 10 (unnecessary) biopsies. The schedule of the PRIAS program schedules a median of six biopsies (IQR: 4–8). In comparison, the biopsies scheduled by the personalized schedules using fixed 10% risk threshold (median: 4, IQR: 4–6) and visit-specific risk chosen using F_1 score (median: 2, IQR: 2–4), are much fewer.

Overall, we observed that the personalized schedule which uses a 10% risk threshold at all follow-up visits is dominant over the PRIAS schedule, biennial schedule of biopsies, and biopsies every one and a half years (see Appendix C for the latter two schedules). This personalized schedule not only schedules fewer biopsies than the aforementioned currently practiced schedules, but the delay in detection of cancer progression is also either equal or less. The personalized schedule which uses risk threshold chosen on the basis of classification accuracy (F_1 score) is dominant over the triennial schedule (see Appendix C) of biopsies. The personalized schedule which uses a 5% risk threshold schedules fewer biopsies than the annual schedule, while the delay is only trivially more than the annual schedule.

4 Discussion

We proposed a methodology which better balances the number of biopsies, and the delay in detection of cancer progression than the currently practiced biopsy schedules, for low-risk prostate cancer patients enrolled in active surveillance (AS) programs. The proposed methodology combines a patient's observed DRE and PSA measurements, and the time of the latest biopsy, into a personalized cancer progression risk function. If the cumulative risk of cancer progression at a follow-up visit is above a certain threshold, then a biopsy is scheduled. We conducted an extensive simulation study, based on a replica of the patients from the PRIAS program, to compare this personalized approach for biopsies with the currently practiced biopsy schedules. We found personalized schedules to be dominant over many of the current biopsy schedules (see Results).

The main reason for the better performance of personalized schedules is that they account for the variation in cancer progression rate between

patients, and also over time within the same patient. In contrast, the existing fixed/heuristic schedules ignore that roughly 50% of the patients never progress in the first ten years of follow-up (*slow progressing* patients) and do not require biopsies. The *fast progressing* patients require early detection. However, existing methods of identifying these patients, such as the use of PSA doubling time in PRIAS, inappropriately assume that PSA evolves linearly over time. Thus, they may not correctly identify such patients. The personalized approach, however, models the PSA profiles non-linearly. Furthermore, it appends information from PSA with information from DRE and previous biopsy results and combines them into a single a cancer progression risk function. The risk function is a finer quantitative measure than individual data measurements observed for the patients. In comparison to decision making with flowcharts, the risk as a single measure of a patient's underlying state of cancer may facilitate shared decision making for biopsies.

Existing work on reducing the burden of biopsies in AS primarily advocates less frequent heuristic schedules of biopsies (Inoue et al., 2018) (e.g., biopsies biennially instead of annually). To our knowledge, risk-based biopsy schedules have barely been explored yet in AS (Bruinsma et al., 2016; Nieboer et al., 2018). The part of our results pertaining to the fixed/heuristic schedules is comparable with corresponding results obtained in existing work (Inoue et al., 2018), even though the AS cohorts are not the same. Thus, we anticipate similar validity for the results pertaining to the personalized schedules.

A limitation of the personalized approach is that the choice of risk threshold is not straightforward, as different thresholds lead to different combinations of the number of biopsies and the delay in detection of cancer progression. An approach is to choose a risk threshold which leads to personalized schedule dominant (e.g., 10% risk) over the currently practiced schedules, for a given delay. Since personalized biopsy schedules are less burdensome, they may lead to better compliance. A second limitation is that the results that we presented are valid only in a 10 year follow-up period, whereas prostate cancer is a slow progressing disease. Thus more detailed results, especially for *slow progressing* patients cannot be estimated. However, very few AS cohorts have a longer follow-period than PRIAS (Bruinsma et al., 2016). In a screening setting often the ethno-racial background of the patient, as well as the history of cancer in first degree relatives are checked. Our model does not take into account either. The reason is that the history of cancer in relatives been found to be predictive of cancer progression only in African-American patients (Goh et al., 2013; Telang et al., 2017). This is also evident by the fact that PRIAS and many other surveillance programs do not utilize this information in their biopsy protocols (Bokhorst et al., 2016; Nieboer et al.,

2018). In addition, patients who have a higher risk of an aggressive form of cancer are usually not recommended active surveillance. Hence the proposed model is relevant only for low-risk prostate cancer patients eligible for active surveillance. An exception is the active surveillance patients who are old and/or have comorbid illnesses. Currently, such patients may be removed from active surveillance and are instead offered the less intensive watchful waiting (Bokhorst et al., 2016) option. It is also possible to model watchful waiting as a competing risk in our model. However, this falls outside the scope of the current work because cancer progression as detected via biopsy is the standard trigger for treatment advice. Lastly, our results are not valid when the patient data is missing not at random (MNAR).

There are multiple ways to extend the personalized decision making approach. For example, biopsy Gleason grading is susceptible to inter-observer variation (Coley et al., 2017). Thus accounting for it in our model will be interesting to investigate further. To improve the decision making methodology, future consequences of a biopsy can be accounted for in the model by combining Markov decision processes with joint models for time-to-event and longitudinal data. There is also a potential for including diagnostic information from magnetic resonance imaging (MRI), such as the volume of the prostate tumor as a longitudinal measurement in our model. The resulting predictions can be used to the decide the time of next MRI as well as to make a decision of biopsy. The same holds true for the quality of life measures as well. However, given the scarceness of both MRI and quality of life measurements in the dataset, including them in the current model may not be feasible. We intend to further validate our results in a multi-center AS cohort, and subsequently develop a web application to assist in making shared decisions for biopsies.

Acknowledgments

on the manuscript.

Supplementary material

Supplementary material for this article, containing Appendix A–D are available at https://github.com/anirudhtomer/prias/blob/master/report/decision_analytic/mdm/latex/supplementary.pdf
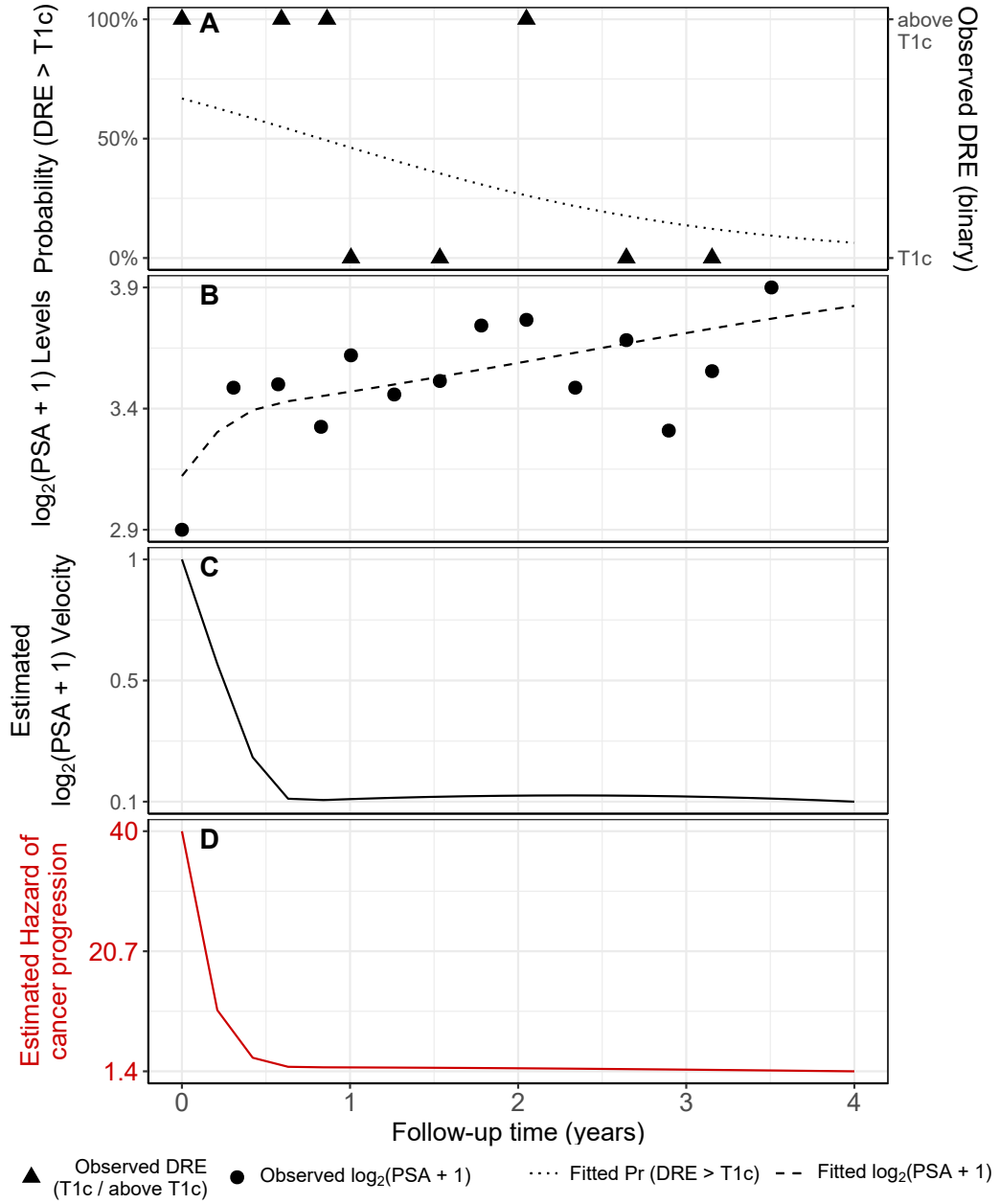
Figure 3: **Illustration of the joint model fitted to the PRIAS dataset.**
Panel A: shows the observed DRE measurements and the fitted probability
of obtaining DRE > T1c (Equation 1) for i-th patient. **Panel B:** shows the
observed and fitted $\log_2(\text{PSA} + 1)$ measurements (Equation 2). **Panel C:**
shows the estimated $\log_2(\text{PSA} + 1)$ velocity (velocity cannot be observed
directly) over time. The hazard function (Equation 3) shown in **Panel D**,
depends on the fitted log odds of having a DRE > T1c, and the fitted
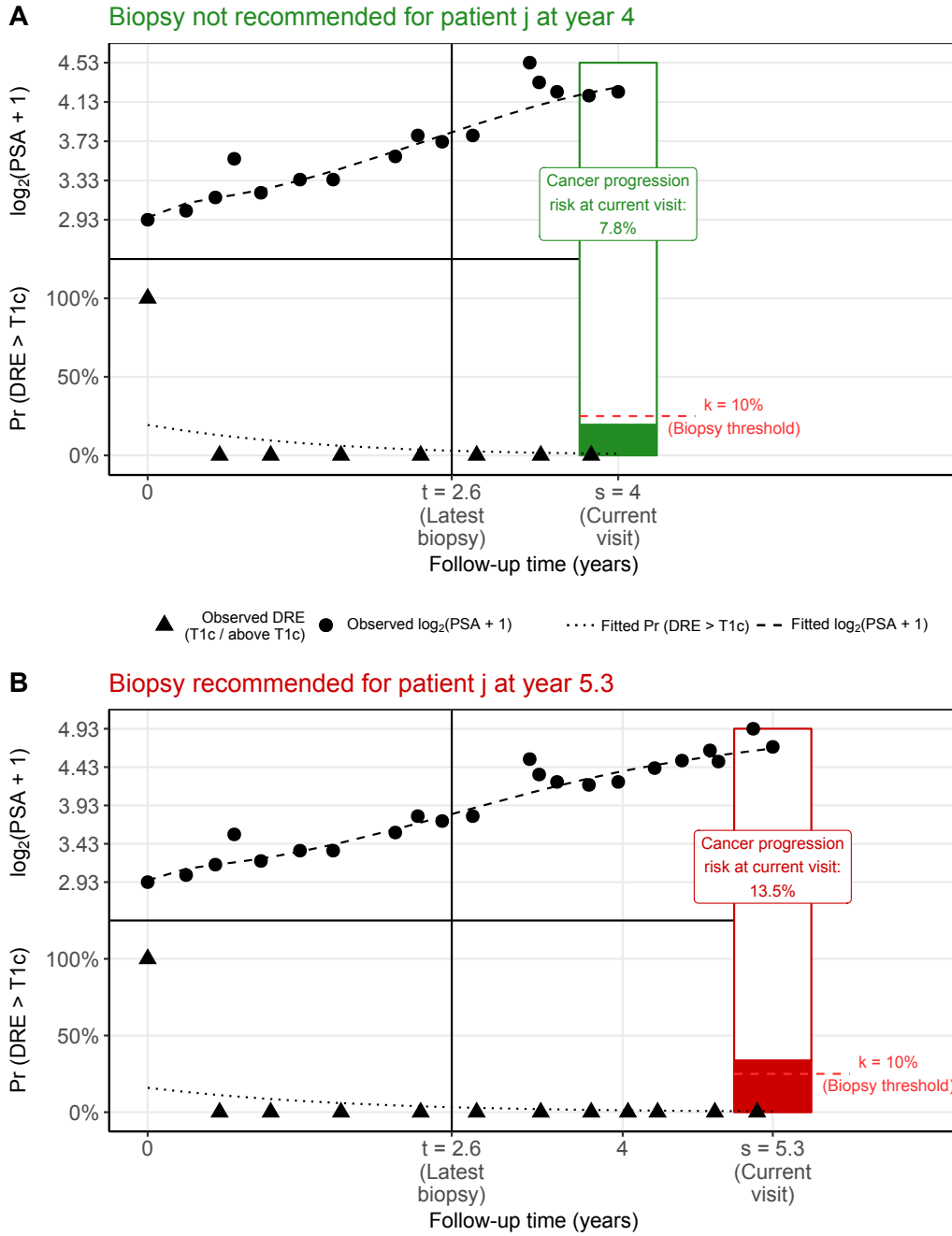$\log_2(\text{PSA} + 1)$ value and velocity.

17

Figure 4: **Illustration of personalized decision of biopsy** for patient j at two different follow-up visits. Biopsy is recommended if the personalized cumulative risk of cancer progression estimated from the joint model fitted to the observed data of the patient, is higher than the example risk threshold for biopsy ($\kappa = 10\%$). **Panel A:** biopsy is not recommended for the patient j at the follow-up visit time $s = 4$ years, because his estimated personalized cumulative risk of cancer progression (7.8%) is less than the threshold. **Panel B:** biopsy is recommended for the patient j at the follow-up visit time $s = 5.3$ years, because his estimated personalized cumulative risk of cancer progression (13.5%) is more than the threshold.
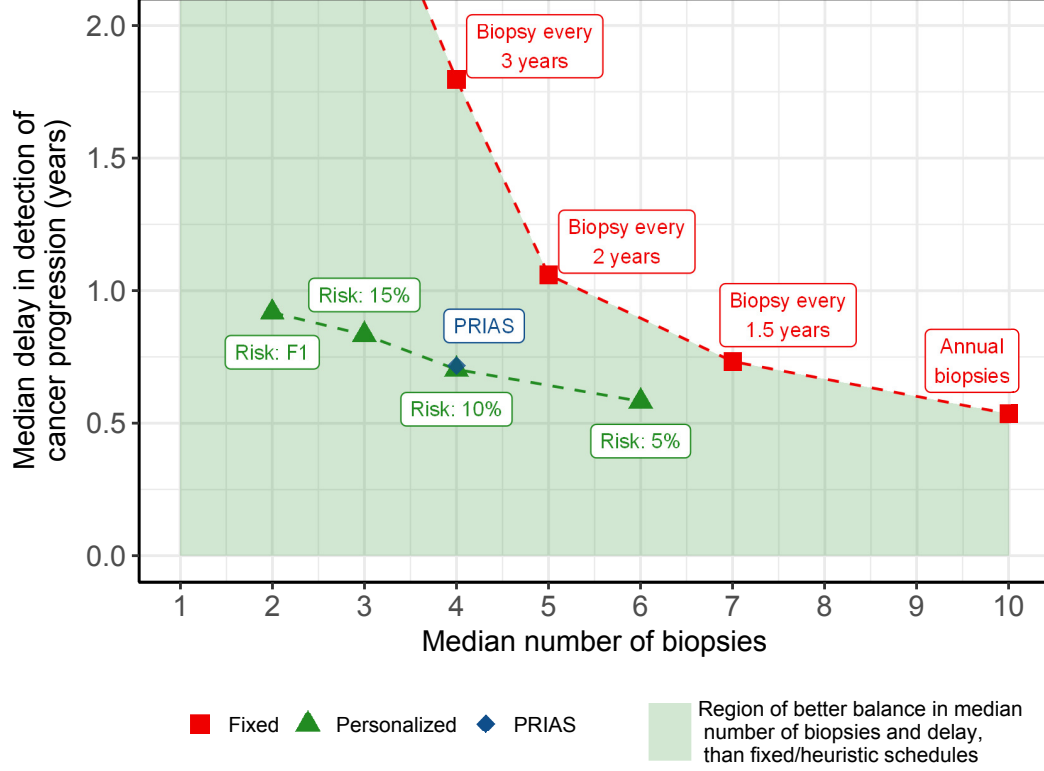
Figure 5: **Burden-biopsy frontier:** Median number of biopsies (X-axis), and median delay in detection of cancer progression (in years, Y-axis), estimated from the simulation study. Results for currently practiced fixed/heuristic biopsy schedules are shown by red squares, for PRIAS schedule by a blue rhombus, and for personalized schedules by green triangles. **Types of personalized schedules:** Risk: 15%, Risk: 10%, and Risk: 5% approaches, schedule a biopsy if the cumulative risk of cancer progression at a visit is more than 15%, 10%, and 5%, respectively. Risk: F1 works similar as previous, except that for each patient, a visit-specific risk threshold is chosen by maximizing F_1 score (see Methods). The green shaded region depicts the region of better balance in the median number of biopsies and median delay than the currently practiced fixed/heuristic schedules.
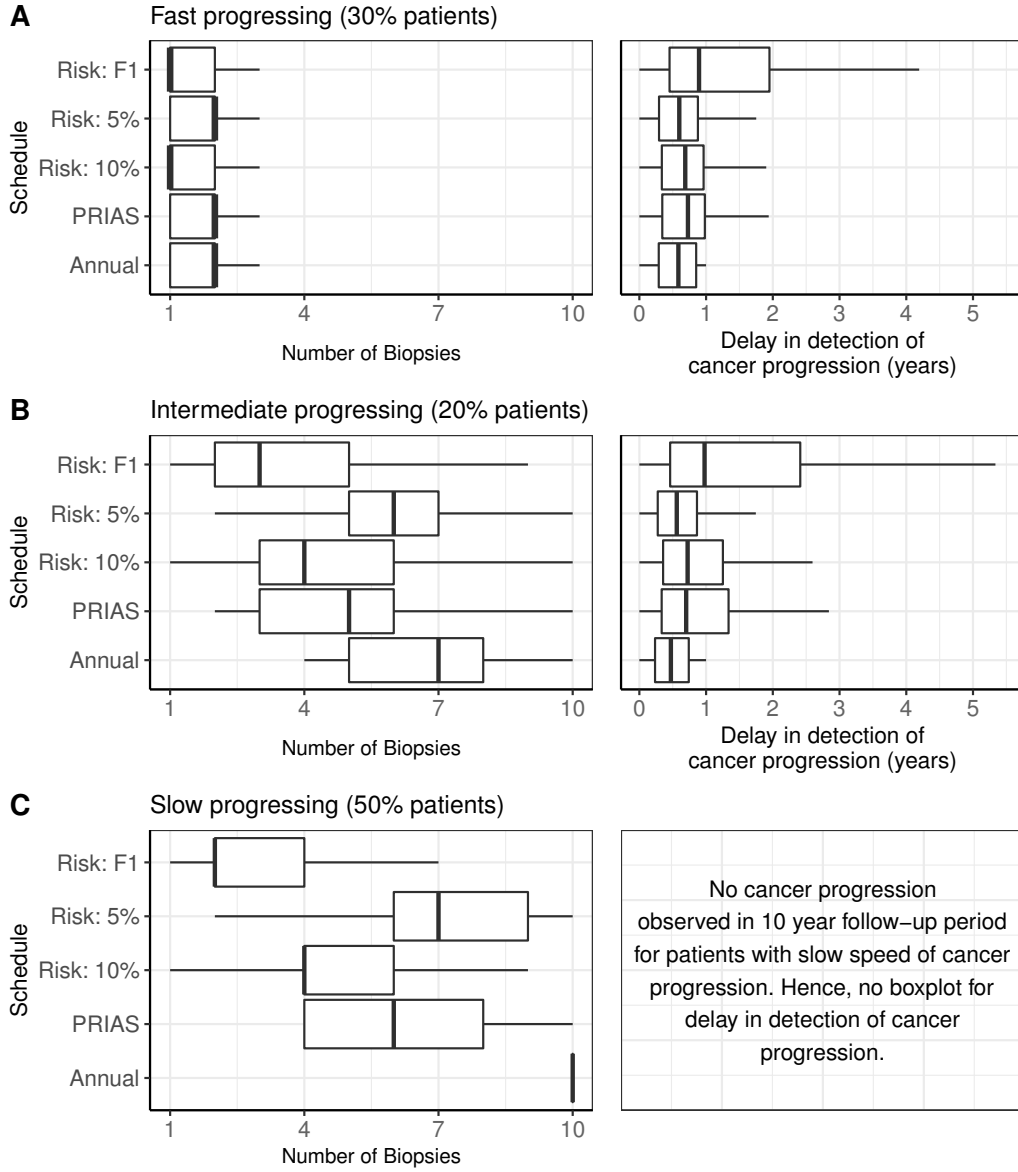
Figure 6: Boxplot showing variation in the number of biopsies, and the delay in detection of cancer progression, in years (time of positive biopsy - true time of cancer progression) for various biopsy schedules. Biopsies are conducted until cancer progression is detected. **Panel A:** results for simulated patients who had a faster speed of cancer progression, with progression times between 0 and 3.5 years. **Panel B:** results for simulated patients who had an intermediate speed of cancer progression, with progression times between 3.5 and 10 years. **Panel C:** results for simulated patients who did not have cancer progression in the ten years of follow-up. **Types of personalized schedules:** Risk: 10% and Risk: 5% approaches, schedule a biopsy if the cumulative risk of cancer progression at a visit is more than 10% and 5%, respectively. Risk: F1 works similar as previous, except that a visit-specific risk threshold is chosen by maximizing F_1 score (see Methods). Annual corresponds to a schedule of yearly biopsies and PRIAS corresponds to biopsies as per PRIAS protocol (see Introduction).

References

Bokhorst, Leonard P et al. (2015). "Compliance rates with the Prostate Cancer Research International Active Surveillance (PRIAS) protocol and disease reclassification in noncompliers". In: *European Urology* 68.5, pp. 814–821.

Bokhorst, Leonard P et al. (2016). "A decade of active surveillance in the PRIAS study: an update and evaluation of the criteria used to recommend a switch to active treatment". In: *European Urology* 70.6, pp. 954–960.

Bruinsma, Sophie M et al. (2016). "Active surveillance for prostate cancer: a narrative review of clinical guidelines". In: *Nature Reviews Urology* 13.3, pp. 151–167.

Carvalho, Tiago M de, Eveline AM Heijnsdijk, and Harry J de Koning (2017). "Estimating the risks and benefits of active surveillance protocols for prostate cancer: a microsimulation study". In: *BJU international* 119.4, pp. 560–566.

Coley, Rebecca Yates et al. (2017). "Prediction of the pathologic Gleason score to inform a personalized management program for prostate cancer". In: *European Urology* 72.1, pp. 135–141.

Cooperberg, Matthew R et al. (2011). "Outcomes of active surveillance for men with intermediate-risk prostate cancer". In: *Journal of Clinical Oncology* 29.2, pp. 228–234.

De Boor, Carl (1978). *A practical guide to splines*. Vol. 27. Springer-Verlag New York.

Ehdaie, Behfar et al. (2014). "The impact of repeat biopsies on infectious complications in men with prostate cancer on active surveillance". In: *The Journal of urology* 191.3, pp. 660–664.

Eilers, Paul HC and Brian D Marx (1996). "Flexible smoothing with B-splines and penalties". In: *Statistical Science* 11.2, pp. 89–121.

Etzioni, Ruth et al. (2002). "Overdiagnosis due to prostate-specific antigen screening: lessons from US prostate cancer incidence trends". In: *Journal of the National Cancer Institute* 94.13, pp. 981–990.

Fujita, Kazutoshi et al. (2009). "Serial prostate biopsies are associated with an increased risk of erectile dysfunction in men with prostate cancer on active surveillance". In: *The Journal of urology* 182.6, pp. 2664–2669.

Goh, Chee L et al. (2013). "Clinical implications of family history of prostate cancer and genetic risk single nucleotide polymorphism (SNP) profiles in an active surveillance cohort". In: *BJU international* 112.5, pp. 666–673.

Inoue, Lurdes YT et al. (2018). "Comparative Analysis of Biopsy Upgrading in Four Prostate Cancer Active Surveillance Cohorts". In: *Annals of internal medicine* 168.1, pp. 1–9.

Lin, Haiqun et al. (2000). "A latent class mixed model for analysing biomarker trajectories with irregularly scheduled observations". In: *Statistics in Medicine* 19.10, pp. 1303–1318.

Loeb, Stacy et al. (2014). "Heterogeneity in active surveillance protocols worldwide". In: *Reviews in urology* 16.4, pp. 202–203.

Nieboer, Daan et al. (2018). "Active surveillance: a review of risk-based, dynamic monitoring". In: *Translational andrology and urology* 7.1, pp. 106–115.

Pearson, Jay D et al. (1994). "Mixed-effects regression models for studying the natural history of prostate disease". In: *Statistics in Medicine* 13.5-7, pp. 587–601.

Rizopoulos, Dimitris (2011). "Dynamic Predictions and Prospective Accuracy in Joint Models for Longitudinal and Time-to-Event Data". In: *Biometrics* 67.3, pp. 819–829.

— (2012). *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R.* CRC Press.

— (2016). "The R Package JMbayes for Fitting Joint Models for Longitudinal and Time-to-Event Data Using MCMC". In: *Journal of Statistical Software* 72.7, pp. 1–46.

Rizopoulos, Dimitris, Geert Molenberghs, and Emmanuel MEH Lesaffre (2017). "Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking". In: *Biometrical Journal* 59.6, pp. 1261–1276.

Schröder, FH et al. (1992). "The TNM classification of prostate cancer". In: *The Prostate* 21.S4, pp. 129–138.

Telang, Jaya M et al. (2017). "Prostate cancer family history and eligibility for active surveillance: a systematic review of the literature". In: *BJU international* 120.4, pp. 464–467.

Torre, Lindsey A et al. (2015). "Global cancer statistics, 2012". In: *CA: A Cancer Journal for Clinicians* 65.2, pp. 87–108.

Tsiatis, Anastasios A and Marie Davidian (2004). "Joint modeling of longitudinal and time-to-event data: an overview". In: *Statistica Sinica* 14.3, pp. 809–834.

Turnbull, Bruce W (1976). "The empirical distribution function with arbitrarily grouped, censored and truncated data". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 38.3, pp. 290–295.

Vickers, Andrew J and Elena B Elkin (2006). "Decision curve analysis: a novel method for evaluating prediction models". In: *Medical Decision Making* 26.6, pp. 565–574.