Personalized Decision Making for Biopsies in Prostate Cancer Active Surveillance Programs

Medical Decision Making
XX(X):3–18

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

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

Prepared using sagej.cls

Abstract

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

Objective. Motivated by the world's largest AS study, Prostate Cancer Research International Active Surveillance (PRIAS), our aim is to better balance the number of biopsies (burden) and the delay in detection of cancer progression (benefit). We intend to achieve this by personalizing the decision of conducting biopsies.

Methods. Using joint models for time-to-event and longitudinal data, we jointly model the observed prostate-specific antigen levels, digital rectal examination scores, and the latest biopsy results of a patient at each follow-up visit. This results in a visit and patient-specific, cancer progression risk profile. Using this personalized risk profile, we make the decision of conducting biopsy at a visit. We compare this personalized approach with the in-practice fixed biopsy schedules via an extensive and realistic simulation study, based on a replica of the patients from the PRIAS study.

Results. In comparison to the fixed schedules, the personalized approach saves one to seven biopsies per patient, depending upon the cancer progression speed of the patient. Despite this reduction in the number of biopsies, the delay in the detection of cancer progression for the personalized approach remains comparable with that of the biopsy schedule of the PRIAS study.

Conclusions. We conclude that the personalized schedules better balance the number of biopsies per detected cancer progression.

Keywords

Active surveillance, biopsy, joint models, personalized medicine, prostate cancer

* *

Introduction

Prostate cancer is the second most frequently diagnosed cancer in men worldwide[1]. The increase in diagnosis of low-grade prostate cancer has been attributed to increase in life expectancy and increase in the number of screening programs[2]. An issue of prostate cancer screening programs is overdiagnosis. To avoid further over-treatment, patients diagnosed with low-grade prostate cancer are commonly advised to join active surveillance (AS) programs. In AS, serious treatments such as surgery, chemotherapy, or radiotherapy are delayed until cancer progresses. Cancer progression is routinely examined via serum prostate-specific antigen (PSA) levels: a protein biomarker, digital rectal examination (DRE) score: a measure of the size and location of the tumor, medical imaging, and biopsies etc.

While larger values for PSA and/or larger score for DRE, may indicate cancer progression, biopsies are the most reliable prostate cancer progression examination technique used in AS. When a patient's biopsy Gleason grading becomes larger than 6 (definition of cancer progression in this work), AS is stopped and the patient is advised treatment for cancer progression[3]. However, biopsies are invasive, painful, and prone to medical complications[4]. Hence, they are conducted intermittently. This leads to a delay in the detection of cancer progression. The delay is equal to the difference between the time of the positive biopsy and the unobserved true time of progression. Hence, the decision of conducting a biopsy requires a fine compromise between the number of biopsies (more is burden) and the delay in detection of progression (less is benefit).

[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 Medical Statistics and Bioinformatics, Leiden University Medical Center, The Netherlands

Corresponding author:
Anirudh Tomer, Erasmus MC, t.a.v. Anirudh Tomer / kamer Na-2701, PO Box 2040, 3000 CA Rotterdam, the Netherlands.
Email: a.tomer@erasmusmc.nl
*Financial support for this study was provided Netherlands Organization for Scientific Research's VIDI grant nr. 016.146.301, and Erasmus MC funding. The funding agreement ensured the authors independence in designing the study, interpreting the data, writing, and publishing the report.
Word Count: 4046

Currently there is no consensus on the frequency of biopsies in AS[5]. Majority of the programs focus on minimizing only the delay, by scheduling biopsies annually. Annual biopsies, and other heuristic schedules[6] which do not account for the difference in cancer progression speeds of patients, may work well for patients who progress fast or for patients for whom the initial biopsy Gleason score was incorrect. However, for slowly progressing patients (majority of AS patients) many unnecessary burdensome biopsies are scheduled. To mediate the burden between fast and slow progressing patients, the world's largest AS program, Prostate Cancer Research International Active Surveillance (PRIAS)[7], schedules annual biopsies only for patients with a small PSA doubling time[3]. For everyone else, PRIAS schedules biopsies at following fixed follow-up times: year 1, 4, 7, and 10, and every 5 years thereafter. Despite this effort, in PRIAS over a period of 10 years a patient may get scheduled for 4 to 10 biopsies. Consequently, patients may not always comply with the schedule[3]. This can lead to the original problem of delayed detection of prostate cancer progression, and reduce the effectiveness of AS.

This article is motivated by the need to better balance the number of biopsies and the delay in detection of prostate cancer progression, than in practice currently (see Figure 1). We intend to achieve this by personalizing the decision of conducting biopsies at follow-up visits. To this end, we utilize the data of the patients of the PRIAS study (see Figure 2). Personalized decision making has received much interest in the literature, especially for screening of various cancers by utilizing Markov decision process models[8–10]. In case of prostate cancer, Zhang et al.[11] personalized the decision of biopsies using these models, to avoid over-diagnosis during screening. Their model used baseline patient characteristics as well as the latest PSA level (discretized) of the patient.

In comparison to the work referenced above, we do not base the decision of biopsy only on the latest PSA level of a patient, but instead we use the entire history, of DRE scores, PSA levels and the unobserved rate of change of PSA (PSA velocity), and results of the latest biopsy. To this end, we employ joint models for time-to-event and longitudinal data[12,13]. Joint models utilize patient-specific random effects[14] to model the observed data, and hence they are inherently personalized. Using joint models we first obtain a full specification of the joint distribution of the time of cancer

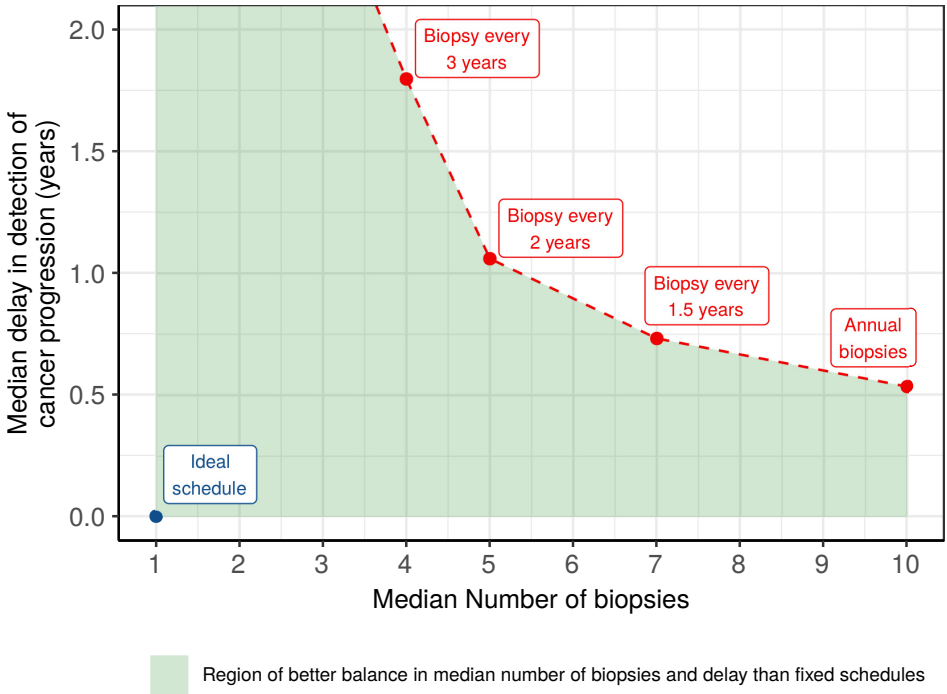

Figure 1. Median number of biopsies scheduled by various fixed scheduling approaches (in red) in practice (over a follow-up of 10 years), and the corresponding median delay in detection of cancer progression. An ideal schedule (in blue) will schedule only 1 biopsy, exactly at the true time of cancer progression. We intend to better balance the number of biopsies and the delay, than in practice currently, using personalized decision making for biopsies.

progression, and PSA and DRE measurements. We then use it separately for each patient at each follow-up visit to develop a cancer progression risk profile, based on their observed data. If the risk is higher than a certain threshold our method schedules a biopsy at the same follow-up visit. Since there is no clear consensus on choice of risk thresholds, we not only use fixed thresholds suggested by urologists, but also present a methodology to automate the choice of thresholds.

In order to compare the personalized approach, with the in-practice annual and PRIAS schedules, we conduct an extensive simulation study. For a realistic comparison, we simulate a replica of the population of the PRIAS patients, using the joint model fitted to the PRIAS dataset.

The rest of the article is structured as follows: The details of the joint modeling framework and biopsy decision making methodology are

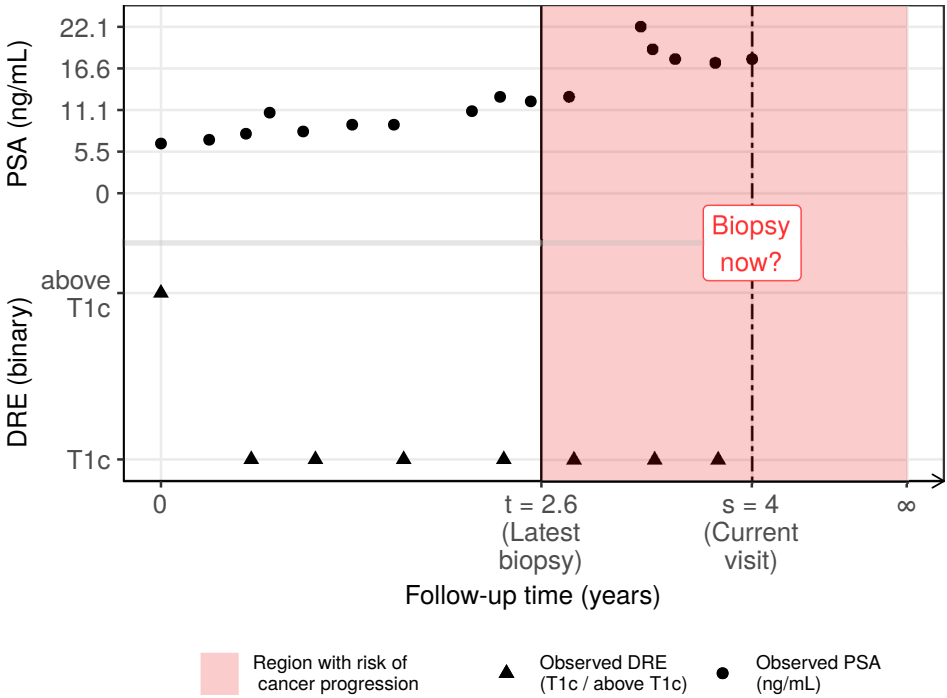

Figure 2. Illustration of the decision making problem: Available data of a patient, who had his latest (negative) biopsy at $t = 2.6$ years. The shaded region shows the time period in which the patient is at the risk of cancer progression. His current follow-up visit is at $s = 4$ years. Using the entire history of DRE $\mathcal{Y}_{dj}(s)$ and PSA $\mathcal{Y}_{pj}(s)$ measurements up to current visit time s, and the time of latest biopsy t, we intend to make a decision on scheduling a biopsy at s.

presented in the Methods section. The details of the simulation study and the corresponding results are presented in Methods and Results sections, respectively.

Methods

Study Population

To develop our methodology we use data of the patients of the PRIAS study (www.prias-project.org). The dataset consists of 5270 patients, of which 866 observe cancer progression. For each patient, PSA measurements (ng/mL) are scheduled every 3 months for the first 2 years and every 6 months thereafter. The DRE measurements (ordinal scale)

are scheduled every 6 months. We use the DRE measurements after converting them on a binary scale, namely DRE $>$ T1c and DRE $\leq$ T1c[15]. On average 5 DRE and 9 PSA measurements have been recorded per patient. In order to identify cancer progression, biopsies are scheduled as per the PRIAS protocol (see Introduction).

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

Let T_i^* denote the true cancer progression time of the i-th patient 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 measurements, 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 PSA and DRE measurements over time are modeled using a multivariate generalized linear mixed effects sub-model. The sub-model for DRE is given by (see Panel A, Figure 3):

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

where, t denotes the follow-up visit time, Age_i is the age of the i-th patient at the time of inclusion in AS. 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 log odds of obtaining a DRE score larger than T1c remain linear over time.

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

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

where, $m_{pi}(t)$ denotes the underlying measurement error free value of $\log_2(\text{PSA} + 1)$ transformed[16,17] measurements at time t. We model it

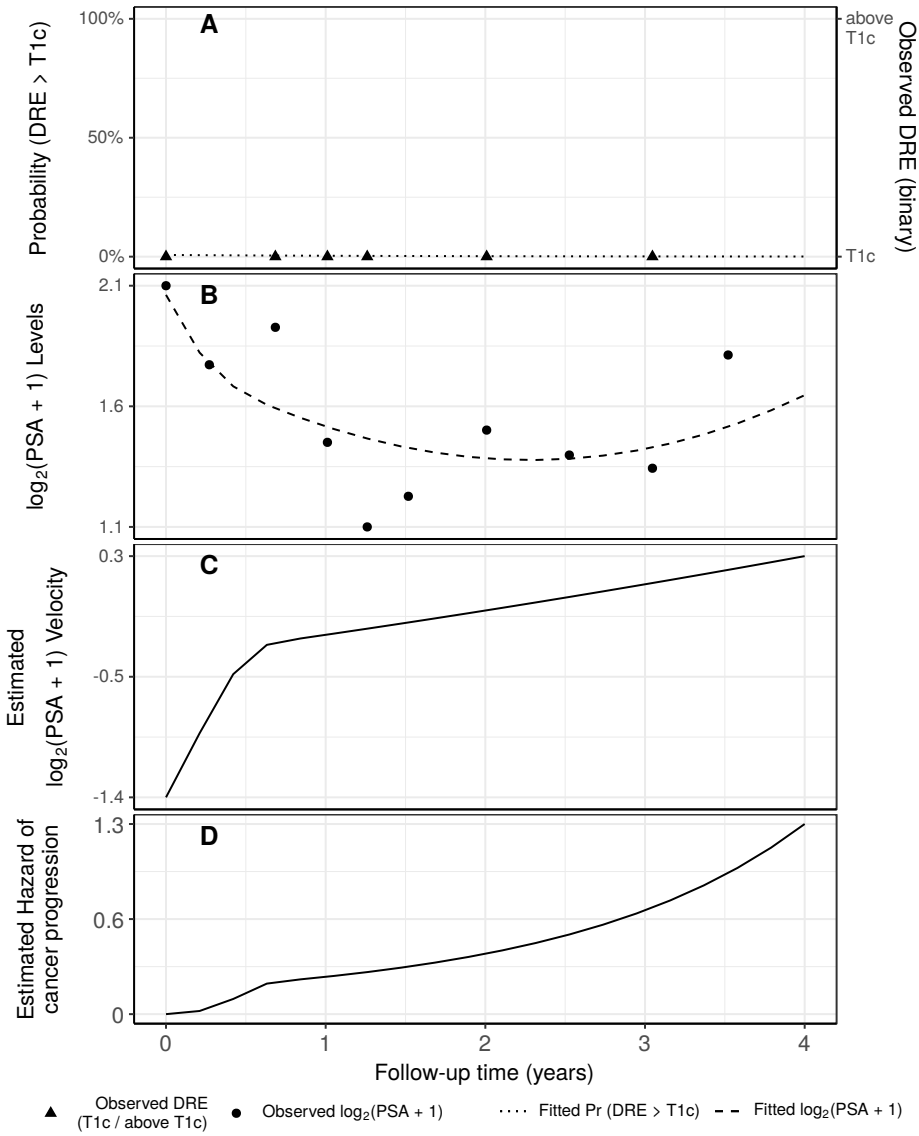

Figure 3. Illustration of the joint model fitted to the PRIAS dataset. **Panel A:** shows the observed DRE scores and the fitted probability of obtaining a DRE score greater than T1c (Equation 1) for . **Panel B:** shows the observed and fitted $\log_2(\text{PSA}+1)$ levels (Equation 2). **Panel C:** 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.

non-linearly over time using B-splines[18]. 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 (0.95 quantile of the observed follow-up times). The fixed effect model parameters are denoted by $\{\beta_{0p}, \ldots, \beta_{6p}\}$ and the patient specific random effects are denoted by $\{b_{0pi}, \ldots, b_{4pi}\}$. 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 the corresponding random effects. More specifically, the complete vector of random effects $\boldsymbol{b}_i = (b_{0di}, b_{0di}, b_{0pi}, \ldots, b_{4pi})^T$ 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):

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

where, γ_1, γ_2 are the coefficients for the effect of age. The parameter α_{1d} models the impact of log odds of obtaining 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[19]. The detailed specification of the baseline hazard $h_0(t)$, and the joint parameter estimation of the two sub-models using the Bayesian approach are presented in Appendix A of the supplementary material.

Personalized Decisions for Biopsy During Follow-up Visit

Let us assume that a decision of conducting a biopsy is to be made for a new patient j shown in Figure 2, at his current follow-up visit time s.

Let $t \leq s$ be the time of his latest biopsy. Let $\mathcal{Y}_{dj}(s)$ and $\mathcal{Y}_{pj}(s)$ denote his observed DRE and PSA measurements taken up to the current visit s, 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 posterior predictive distribution $g(T_j^*)$ of his time of cancer progression T_j^*. It is given by:

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

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

A key ingredient in the decision of conducting a biopsy at the current follow-up visit time s, is the cumulative risk that the cancer has already progressed since the time of the last biopsy t (illustrated in Figure 4). This risk can be derived from the posterior predictive distribution $g(T_j^*)$[20], and is given by:

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

A simple and straightforward approach to decide upon conducting a biopsy at the current follow-up visit would be to do so if the 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 risk is higher than 15% (example risk threshold). This process is iterated over the follow-up period until a positive biopsy, while incorporating on each subsequent visit, the time of the latest negative biopsy, and updates in PSA and DRE profiles. This leads to a separate schedule of biopsies for each patient.

Ideally, thresholds should be chosen on the basis of their decision theoretic utility for patients. That is, an ideal threshold (see Figure 1) will minimize a bivariate function of the number of biopsies and the delay in detection of progression. However, this requires expressing number of biopsies (count) in units of time (delay), which is not easy. Alternatively, a threshold which minimizes expected delay is given by,

$\kappa = R_j\{E_g(T_j^*) \mid t\}$, where $E_g(T_j^*)$ is expected progression time. Since, cancer progression is a slowly progressing disease, most AS patients do not progress over a long follow-up period of 10 years. Consequently, it is impossible to compute $E_g(T_j^*)$ for patients without incorrectly extrapolating PSA, DRE profiles, and hazard of cancer progression, over follow-up periods with no data.

In this regard, an elementary approach is, using risk thresholds utilized in standard clinical settings, e.g., 5% or 10% risk, at all follow-up visits. However, a threshold's accuracy of classification between progressions and non-progressions may vary over time. This motivates an alternative approach, where at each follow-up visit, the threshold which gives the better classification accuracy is chosen. More specifically, given the time t of the latest biopsy of a patient, we are interested in a threshold κ with which we can detect maximum number of cancer progressions (high true positive rate, or TPR). However, this may lead to unnecessary biopsy suggestions for many patients (high false positive rate). This issue can be mitigated by maximizing for the positive predictive value, or PPV (also known as precision) simultaneously. To this end, we utilize the F_1 score, which is a composite of TPR and PPV (estimated as in Rizopoulos et al.[21]), and is defined as:

$$\begin{aligned}
\mathrm{F}_1(t,s,\kappa) &= 2\frac{\mathrm{TPR}(t,s,\kappa)\,\mathrm{PPV}(t,s,\kappa)}{\mathrm{TPR}(t,s,\kappa)+\mathrm{PPV}(t,s,\kappa)},\\
\mathrm{TPR}(t,s,\kappa) &= \Pr\big\{R_j(s \mid t) > \kappa \mid t < T_j^* \le s\big\},\\
\mathrm{PPV}(t,s,\kappa) &= \Pr\big\{t < T_j^* \le s \mid R_j(s \mid t) > \kappa\big\}.
\end{aligned} \tag{4}$$

The F_1 score ranges between 0 and 1, where a value of 1 signifies perfect TPR and PPV. Since a high F_1 score is desired, threshold $\kappa = \arg\max_\kappa \mathrm{F}_1(t,s,\kappa)$. It is important to note that F_1 score is not the utility of the threshold for the patients. The utility of both fixed thresholds, and threshold based on F_1 score is discussed in Results.

Simulation Study

Although the personalized decision making approach is motivated by the PRIAS study, it is not possible to evaluate it on the PRIAS dataset. This is due to the fact that 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 annual schedules. For a realistic comparison, we simulate data from the joint model fitted to the PRIAS dataset. The simulated population has the same follow-up period of 10 years as the PRIAS study. In addition the recovered relations between PSA and DRE 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 each. We generate a true cancer progression time for each of the patients, and then sample a set of PSA and DRE measurements at the same time points as given in PRIAS protocol. We then split each dataset into a training (750 patients) and a test (250 patients) part, and generate a random and noninformative censoring time for the training patients. We next fit a joint model of the specification given in Equation (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. Using these fitted joint models, we obtain the cancer progression risk profiles for each of the 500×250 test patients at each of their visits. For each patient at each follow-up visit we make the decision of (not) conducting a biopsy.

In this simulation study, during follow-up visits we make the decision of biopsies as per the following approaches (abbreviated names in parenthesis): biopsy every year (Annual), biopsy as per the PRIAS protocol (PRIAS), personalized biopsy using: a risk threshold of 5% (Risk: 5%), risk threshold of 15% (Risk: 15%), and risk threshold chosen automatically by maximizing F_1 score (Risk: Automatic). In addition, in each of the aforementioned approaches, one biopsy each is conducted at the time of inclusion in AS (year 0) and at the end (year 10) of the follow-up period. This results into an entire personalized schedule for each patient. We compare the resulting biopsy schedules on two measures, namely the number of biopsies they schedule and the delay in detection of cancer progression incurred due the schedule. We define the delay as the difference between the time of the biopsy on which cancer progression is detected and the true time of cancer progression. Ideal numbers for these two measures are 1 biopsy and 0 years of delay.

Results

We first discuss the results pertaining to the joint model fitted to the PRIAS dataset and then discuss results from the simulation study.

Model Fit

From the joint model fitted to the PRIAS dataset, we found that both $\log_2\{\text{PSA}+1\}$ velocity, and log odds of having DRE $>$ T1c were significantly associated with the hazard of cancer progression. For any patient, an increase in $\log_2\{\text{PSA}+1\}$ velocity from 0.03 to 0.16 (first and third quartiles of the fitted velocities, respectively) corresponds to a 1.92 fold increase in the hazard of cancer progression. Whereas, an increase in log odds of DRE $>$ T1c from -6.65 to -4.36 (first and third quartiles of the fitted log odds, respectively) corresponds to a 1.40 fold increase in the hazard of cancer progression. In terms of the predictive performance, we found that the area under the receiver operating characteristic curves (AUC)[21] was 0.65, 0.62, 0.75, 0.71 and 0.59 at years one, two, three, four, and five of followup, respectively. Parameter estimates are presented in detail in Appendix B of the supplementary material.

Simulation Study

Figure 5 shows the mean (obtained from 500 x 250 test patients) number of biopsies conducted by various biopsy schedules, plotted against the corresponding delay in detection of cancer progression in years (time of last biopsy - true time of cancer progression). The general trend is that more biopsies are required to have a smaller delay in detection.

Since patients have varying cancer progression speeds, the impact of different approaches for biopsies also varies with the speed. In order to highlight these differences we trichotomize patients into 3 categories (fast, intermediate, slow speed) as per their time of cancer progression. In the simulated patients, we observed that for roughly 50% of the patients cancer progression did not take place in the 10 year follow-up. These could be seen as patients with a slow speed of cancer progression. Roughly 30% of the patients obtain cancer progression within first 3.5 years. These could be high risk patients who choose AS instead of immediate treatment, or patients with an initially misdiagnosed state of cancer[22]. These patients can be considered as fast progressing patients. We consider that the remaining 20% patients with cancer progression times between 3.5 and

10 years have an intermediate speed of cancer progression. Although such a trichotomization may not be perfect and can only be done retrospectively in a simulation setting, we do it only for the purpose of illustration.

For faster progressing patients (30% of the total patients), the boxplots in panel A of Figure 6 shows the variation in the number of biopsies (of 500 x 250 test patients), and the delay in detection of cancer progression, in years (time of last biopsy - true time of cancer progression) due to various biopsy schedules. We can see that the personalized schedules conduct a median of one biopsy compared to two biopsies for PRIAS and annual schedule. The performance of personalized schedule with 15% risk threshold is similar to that of PRIAS schedule. Thus with personalized approach, one biopsy may get saved for faster progressing patients.

For patients with intermediate progression speed (20% of the total patients), the boxplots in panel B of Figure 6 shows the variation in number of biopsies, and the delay in detection of cancer progression due to various biopsy schedules. Firstly, we can see that personalized schedules with a small risk threshold such as 5% risk conduct many more biopsies than other personalized schedules. Consequently, their performance with respect to the delay in detection of progression is similar to that of annual schedule. However, personalized schedule with slightly higher risk (15%) and risk chosen automatically, schedule a median of 3 biopsies each. This is despite the fact that the delay in detection of cancer progression due to the schedule with 15% risk threshold is similar to that of the PRIAS schedule. However, the PRIAS schedule conducts more biopsies (median of 5 biopsies). Thus, the personalized approach may lead to two less biopsies for patients with intermediate speed of progression.

The patients who are at most advantage with the personalized schedules are the patients who progress slowly (50% of the total patients). Panel C of Figure 6 shows a boxplot of the number of biopsies conducted by various biopsy schedules for such patients. It can be seen that the annual schedule may lead to 10 unnecessary biopsies for everyone. The PRIAS schedule, schedules a median of 6 unnecessary biopsies. In comparison the personalized schedules using 15% risk threshold and automatically chosen risk threshold, schedule only 2 and 3 biopsies, respectively.

Discussion

BASED ON A SINGLE AS PROGRAM. We NEED MORE because of classification accuracy

AND RIsk automatic did not work best for delay, but more AS programs are needed to evluate it best.

Prostate cancer active surveillance (AS) programs schedule biopsies for patients to detect cancer progression. Biopsies are burdensome and hence each biopsy counts. However, currently there is no consensus on the best time interval for subsequent repeat biopsies[5]. In order to reduce the delay in detection of cancer progression many AS programs schedule biopsies annually which leads to many unnecessary biopsies for slowly progressing patients. The world's largest AS program PRIAS attempts to identify these patients using their prostate-specific antigen (PSA) profile. However, despite their methodology, compliance for biopsies is low[3]. With an aim to better balance the burden of biopsies on patients and the delay in detection of cancer patients, in this article, we presented a methodology for personalizing the biopsy decision making process in AS programs.

Our methodology utilizes joint models for time-to-event and longitudinal data. Existing approaches for scheduling biopsies either discard information from PSA and digital rectal examination (DRE), or use crude measures such as PSA doubling time. In contrast, our proposed methodology makes a separate decision of biopsy for each patient at each follow-up visit on the basis of finer measures such as patient specific underlying instantaneous PSA value, PSA velocity, probability of having DRE larger than level T1c, and time of the latest biopsy. Our method combines the aforementioned measures into a patient and visit specific cancer progression risk function. It schedules a biopsy if the risk of cancer progression crosses a certain threshold. We compared our approach with the existing annual and PRIAS schedules, by conducting a realistic and extensive simulation study, for a 10 year follow-up period.

In the simulation study we found that the patients who never obtain cancer progression in the 10 year follow-up incur much burden due to currently used schedules. The PRIAS schedule, despite its effort to identify such patients using PSA doubling time, schedules a minimum of 4 biopsies and a median of 6 unnecessary biopsies for such patients. The annual schedule performs even worse by scheduling 10 unnecessary

biopsies. In contrast, the personalized schedules that we proposed reduce it to a median of 2 to 3 biopsies depending upon the choice of the risk threshold. The choice of the risk threshold is also important. A low risk threshold such as 5% risk may seem attractive to detect cancer progressions in time. However, it can schedule as many biopsies as annual schedule does. An alternative is to either use a higher risk threshold or to automatically select it (see Methods). The personalized approach based on automatically chosen thresholds is useful when the manual choice of a risk thresholds is dilemmatic. Since it obtains thresholds from the observed data of patients, it is generic for usage in other scheduling applications.

On average the the personalized biopsy approach which uses 15% risk threshold, schedules only 3 biopsies per person than the 4.5 scheduled by PRIAS schedule. However, on average the personalized approach (15% risk) exceeds the delay in detection of cancer progression due to PRIAS schedule by only 3.4 months. The 3 biopsies that personalized approach schedules matches the number of biopsies patients agree to undergo in PRIAS, if the non-compliance rates[3] are also accounted for. Hence the personalized approach for biopsies better balances the number of biopsies per detected progression compared to the existing PRIAS and annual schedules.

A limitation of our approach however, is that at any given follow-up visit it cannot guarantee the total number of future biopsies required to detect cancer progression. This is due to the fact that the cancer progression cannot be foreseen with 100% accuracy. Hence, attempts to create an entire optimal schedule may not always work. Instead, such a guarantee can be given with a certain probability (see Figure 6). For example, if a 90% surety is required on the delay in detection of cancer progression being less than or equal to two years, then the personalized schedule with automatically chosen risk threshold schedules the least number of biopsies on average. Urologists may also choose a personalized biopsy approach suitable to patients with a certain speed of progression, if it is known in advance. In order to aid urologists in biopsy decision making process, we have developed a web application (examples included), hosted at `www.<insert-name-here>.com`.

Acknowledgements

The first and last authors would like to acknowledge support by the Netherlands Organization for Scientific Research's VIDI grant nr. 016.146.301, and Erasmus MC funding. The authors also thank the Erasmus MC Cancer Computational Biology Center for giving access to their IT-infrastructure and software that was used for the computations and data analysis in this study. Lastly, we thank Frank-Jan H. Drost from the Department of Urology, Erasmus University Medical Center, for helping us in accessing the PRIAS data set.

Supplemental material

Supplementary material for this article are available after references and figures in this document.

References

1. Torre LA, Bray F, Siegel RL et al. Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians* 2015; 65(2): 87–108.
2. Potosky AL, Miller BA, Albertsen PC et al. The role of increasing detection in the rising incidence of prostate cancer. *JAMA* 1995; 273(7): 548–552.
3. Bokhorst LP, Alberts AR, Rannikko A et al. Compliance rates with the Prostate Cancer Research International Active Surveillance (PRIAS) protocol and disease reclassification in noncompliers. *European Urology* 2015; 68(5): 814–821.
4. Ehdaie B, Vertosick E, Spaliviero M et al. The impact of repeat biopsies on infectious complications in men with prostate cancer on active surveillance. *The Journal of urology* 2014; 191(3): 660–664.
5. Loeb S, Carter HB, Schwartz M et al. Heterogeneity in active surveillance protocols worldwide. *Reviews in urology* 2014; 16(4): 202.
6. Inoue LY, Lin DW, Newcomb LF et al. Comparative analysis of biopsy upgrading in four prostate cancer active surveillance cohorts. *Annals of internal medicine* 2018; 168(1): 1–9.
7. Bul M, Zhu X, Valdagni R et al. Active surveillance for low-risk prostate cancer worldwide: the prias study. *European urology* 2013; 63(4): 597–603.
8. Ayer T, Alagoz O and Stout NK. A POMDP approach to personalize mammography screening decisions. *Operations Research* 2012; 60(5): 1019–1034.
9. Akhavan-Tabatabaei R, Sánchez DM and Yeung TG. A Markov decision process model for cervical cancer screening policies in Colombia. *Medical Decision Making* 2017; 37(2): 196–211.
10. Erenay FS, Alagoz O and Said A. Optimizing colonoscopy screening for colorectal cancer prevention and surveillance. *Manufacturing & Service Operations Management* 2014; 16(3): 381–400.
11. Zhang J, Denton BT, Balasubramanian H et al. Optimization of prostate biopsy referral decisions. *Manufacturing & Service Operations Management* 2012; 14(4): 529–547.
12. Tsiatis AA and Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* 2004; 14(3): 809–834.

13. Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. CRC Press, 2012. ISBN 9781439872864.
14. Laird NM and Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; : 963–974.
15. Schröder F, Hermanek P, Denis L et al. The tnm classification of prostate cancer. *The Prostate* 1992; 21(S4): 129–138.
16. Pearson JD, Morrell CH, Landis PK et al. Mixed-effects regression models for studying the natural history of prostate disease. *Statistics in Medicine* 1994; 13(5-7): 587–601.
17. Lin H, McCulloch CE, Turnbull BW et al. A latent class mixed model for analysing biomarker trajectories with irregularly scheduled observations. *Statistics in Medicine* 2000; 19(10): 1303–1318.
18. De Boor C, De Boor C, Mathématicien EU et al. *A practical guide to splines*, volume 27. Springer-Verlag New York, 1978.
19. Eilers PH and Marx BD. Flexible smoothing with B-splines and penalties. *Statistical Science* 1996; 11(2): 89–121.
20. Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* 2011; 67(3): 819–829.
21. Rizopoulos D, Molenberghs G and Lesaffre EM. Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biometrical Journal* 2017; 59(6): 1261–1276.
22. Cooperberg MR, Cowan JE, Hilton JF et al. Outcomes of active surveillance for men with intermediate-risk prostate cancer. *Journal of Clinical Oncology* 2011; 29(2): 228.

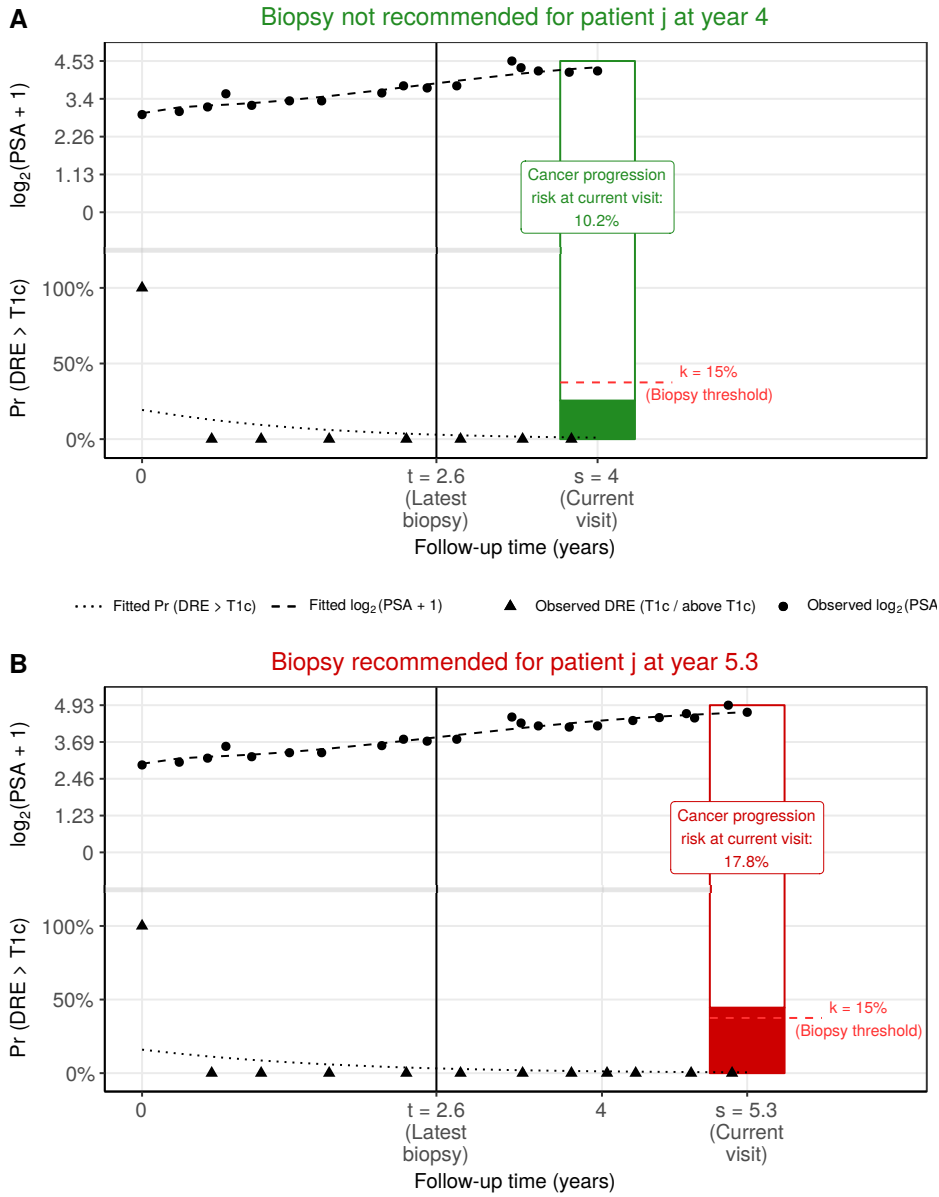

Figure 4. Illustration of personalized decision making for patient j at two different follow-up visits. Biopsy is recommended if the risk of cancer progression estimated from the joint model fitted to the PSA and DRE measurements of the patient, is higher than the example risk threshold for biopsy ($\kappa = 15\%$). **Panel A:** biopsy is not recommended for the patient j at the follow-up visit time $s = 4$ years, because his estimated risk of cancer progression (10.2%) is less than the biopsy risk threshold. **Panel B:** biopsy is recommended for the patient j at the follow-up visit time $s = 5.3$ years, because his estimated risk of cancer progression (17.8%) is more than the biopsy risk threshold.

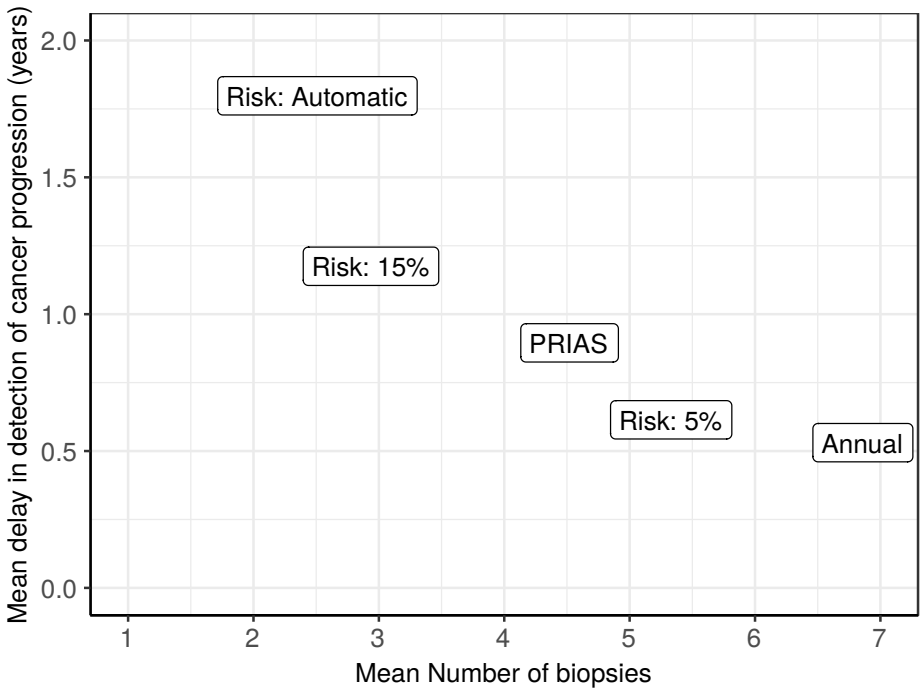

Figure 5. Mean number of biopsies (of 500 x 250 test patients) scheduled by various biopsy schedules, plotted against the corresponding delay in detection of cancer progression, in years (time of last biopsy - true time of cancer progression). Biopsies are conducted until cancer progression is detected. **Types of personalized schedules:** Risk: 15% and Risk: 5% approaches, schedule biopsy if the risk of cancer progression at a visit is more than 15% and 5%, respectively. Risk: Automatic works similar to Risk: 15% and Risk: 5%, except that the risk threshold for biopsy is chosen automatically 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).

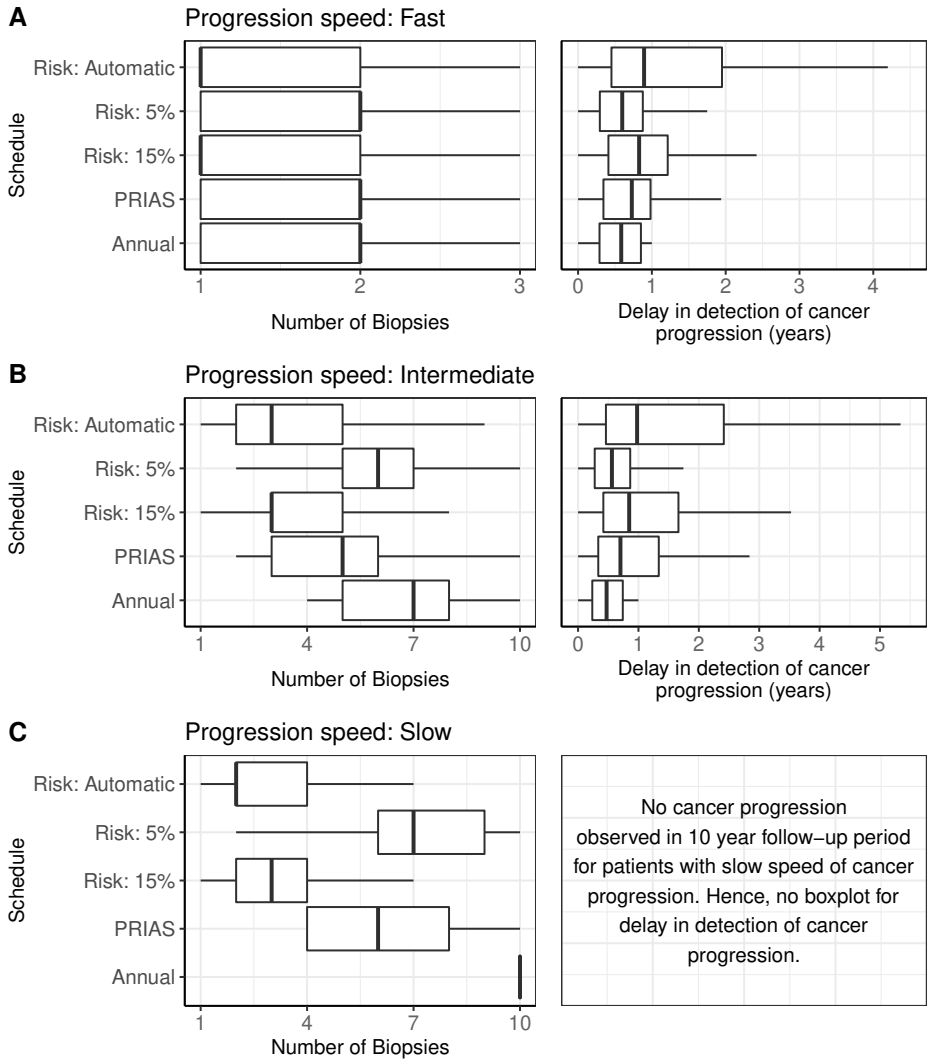

Figure 6. Boxplot showing variation in number of biopsies, and the delay in detection of cancer progression, in years (time of last biopsy - true time of cancer progression) for various biopsy schedules. Biopsies are conducted until cancer progression is detected. **Panel A:** results for simulated patients 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 10 years of follow-up. **Types of personalized schedules:** Risk: 15% and Risk: 5% approaches, schedule biopsy if the risk of cancer progression at a visit is more than 15% and 5%, respectively. Risk: Automatic works similar to Risk: 15% and Risk: 5%, except that the risk threshold for biopsy is chosen automatically 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).