Personalized Decision Making for Biopsies in Prostate Cancer Active Surveillance Programs

Medical Decision Making
XX(X):3–19

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 program, 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 follow-up visit-specific and patient-specific, cumulative risk of cancer progression. When this risk at a visit is above a certain threshold, we schedule a biopsy at that visit. 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. In comparison to the currently practiced schedules, the personalized approach saves one to seven burdensome biopsies per patient, depending upon the time of cancer progression 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 program.

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

Keywords

Active surveillance, biopsy, joint models, personalized medical decisions, prostate cancer

* *

Introduction

Prostate cancer is the second most frequently diagnosed cancer in men worldwide[1]. However, many of the diagnosed tumors are clinically insignificant (over-diagnosed)[2]. 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 cancer progression examination technique used in AS. When a patient's biopsy Gleason grading becomes larger than 6 (positive biopsy), AS is stopped and the patient is advised treatment for cancer progression[3]. However, biopsies are invasive, painful, and prone to medical complications[4,5]. Hence, they are conducted intermittently until a positive biopsy. Consequently, at the time of a positive biopsy, cancer progression is observed with a delay of an 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 of conducting a biopsy requires a fine compromise between the burden of a biopsy, and the potential duration of the delay in the detection of cancer progression.

[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

Corresponding author:
Anirudh Tomer, Erasmus MC, t.a.v. Anirudh Tomer / kamer flex Na-2823, 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: 3728

In AS, a delay in detection of cancer progression around twelve to fourteen months is assumed to be unlikely to substantially increase risk for adverse downstream outcomes[6,7]. However, for biopsies there is little consensus on the time gap between them[8]. Majority of the AS programs focus on minimizing only the delay in detection of cancer progression, by scheduling biopsies annually (most frequent schedule) for all patients. A drawback of annual biopsies, and other fixed/heuristic schedules[6], 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*) after inclusion in AS, but for a large proportion of patients who do not progress until late (*slow progressing*) in AS (see Figure 1), 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 (PRIAS)[9], 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 one, four, seven, and ten, and every five years thereafter. Despite this effort in PRIAS, a patient may get scheduled for four to ten biopsies over a period of ten years. Consequently, patients may not always comply with the biopsy schedule[3]. This can lead to the original problem of delayed detection of cancer progression, and reduce the effectiveness of AS.

This article is motivated by the need to better balance the number of biopsies (more are burdensome), and the delay in detection of cancer progression (less is beneficial), than practiced currently (see Figure 2). 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. Personalized decision making has received much interest in the literature, especially for the screening of various cancers, by exploiting Markov decision process models[11–13]. Similar models have also been used for personalizing the pre-diagnosis prostate cancer screening strategy[14,15]. However, in post-diagnosis AS programs, to the best of our knowledge, only fixed/heuristic approaches[6,7], or PSA doubling time[3] have been employed for deciding the time of biopsies.

In this work, we make a patient-specific decision of biopsy at a patient's pre-scheduled follow-up visit (logistical considerations) for DRE and PSA measurements. In comparison to the work referenced above, we make

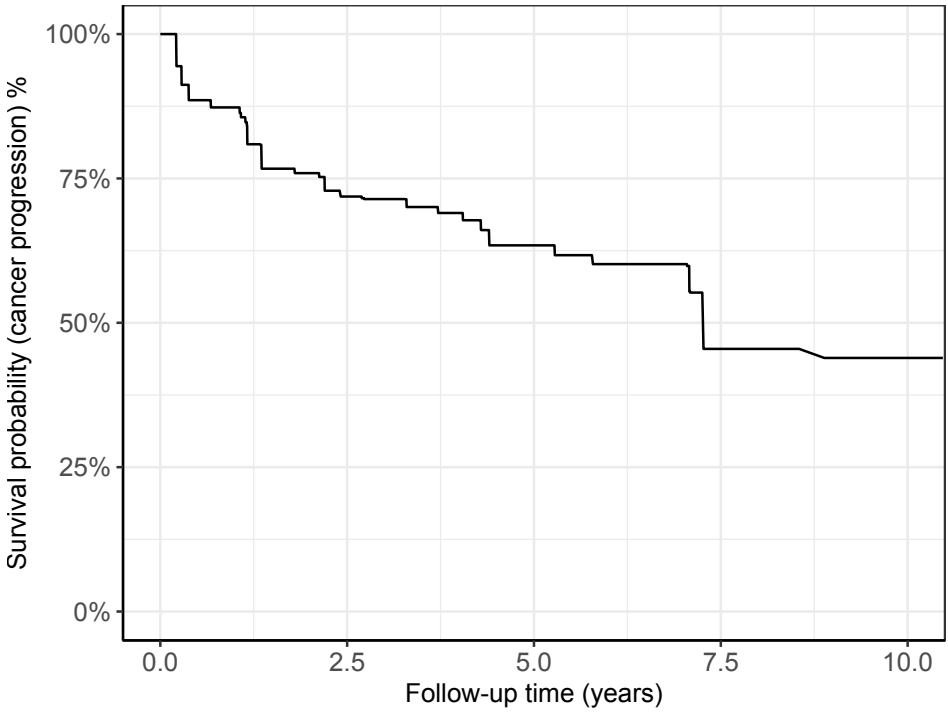

Figure 1. Estimated survival probability of cancer progression in AS: for patients in the Prostate Cancer Research International Active Surveillance (PRIAS) dataset. Nearly 50% patients (*slow progressing*) do not progress in the ten year follow-up period. Estimation is done using a nonparametric maximum likelihood estimate, of the distribution function for interval censored cancer progression times observed in PRIAS[10].

this decision using the entire history of DRE scores, PSA levels and the estimated rate of change of PSA (PSA velocity), and results of the latest biopsy of a patient up to the latest follow-up visit. We achieve this by utilizing joint models for time-to-event and longitudinal data[16,17]. Joint models consist of a longitudinal mixed effects sub-model for periodically measured outcomes such as DRE and PSA, and a relative risk sub-model for modeling the time of cancer progression. The association between these three outcomes, and especially the fact that the DRE and PSA measurements are missing after the detection of cancer progression, is modeled using patient-specific random-effects[18]. By estimating the parameters of the model jointly, we first obtain a full specification of the joint distribution of the time of cancer progression, and DRE and PSA measurements. We then use it at a patient's follow-up visit, to estimate the

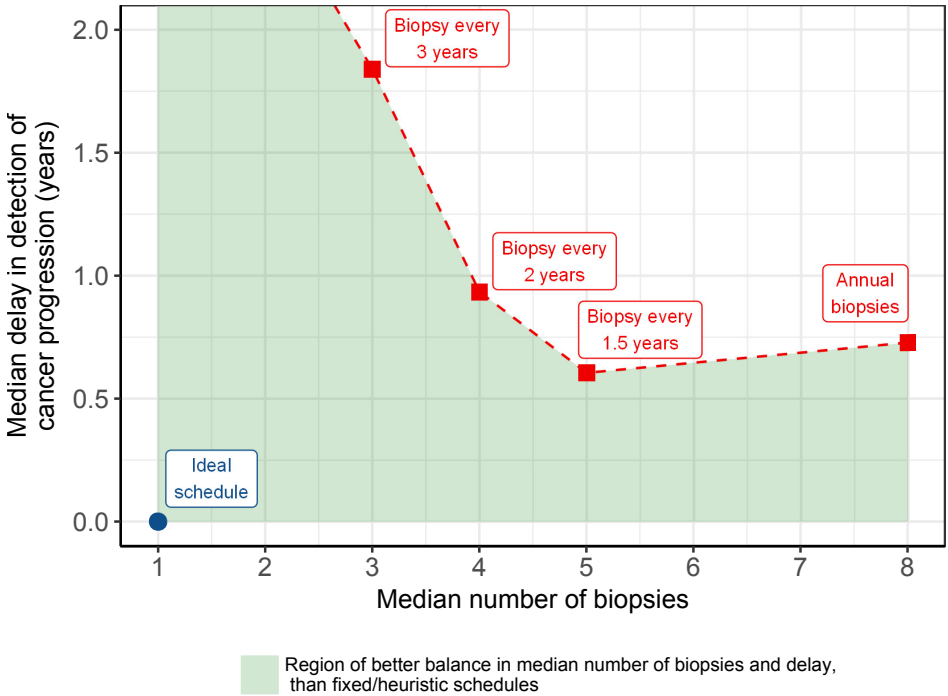

Figure 2. Burden-benefit frontier: Estimated median number of biopsies (more are burdensome), and median delay in detection of cancer progression (less is beneficial), due to various currently practiced fixed/heuristic biopsy schedules (red squares), over a follow-up of ten years. Estimation is based on cancer progression times generated using the survival probability curve in Figure 1. Using personalized decision making for biopsies, we intend to better balance the number of biopsies and the delay (green region), than currently practiced schedules. An ideal biopsy schedule (blue circle) will schedule only one biopsy, exactly at the true time of cancer progression.

patient-specific cumulative risk of observing cancer progression at that visit. If that risk is higher than a certain threshold, our method proposes a biopsy at the same follow-up visit. We exploit fixed risk thresholds used in standard clinical settings, as well as we propose a methodology to choose risk thresholds on the basis of their classification accuracy.

We evaluate the personalized risk based biopsy approach, and the currently practiced fixed/heuristic, PRIAS biopsy schedules on the basis of their utility for the patients. That is, the number of biopsies they schedule, and the corresponding delay incurred in the detection of cancer progression. To compare the aforementioned schedules on these criteria,

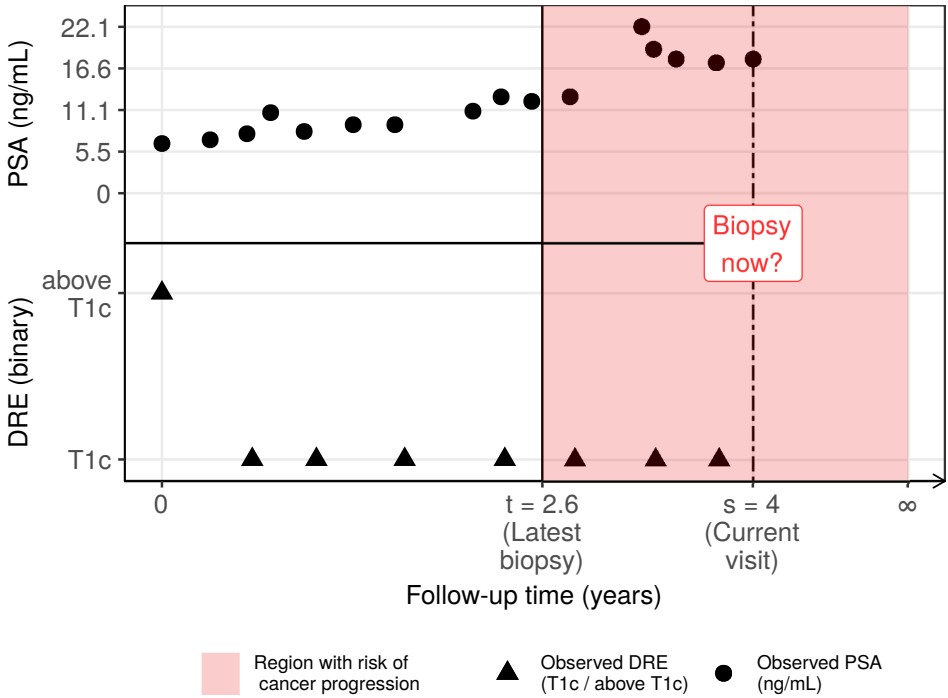

Figure 3. The personalized decision making problem: Available data of a patient j, who had his latest negative biopsy at $t = 2.6$ years. The shaded region shows the time period in which the patient is at risk of cancer progression. His current pre-scheduled follow-up visit for measurement of DRE and 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.

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 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 prostate cancer patients from the world's largest AS study called PRIAS[9]. More than 100 medical centers from 17 countries worldwide contribute to the collection of data, utilizing a common study protocol and a web-based tool, both available at www.prias-project.org. We use the data collected between December 2006 (beginning of the study) and December 2016. It consists of 5270 patients. Cancer progression is observed in 866 patients. 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 (ordinal scale) are scheduled every 6 months. We use the DRE measurements after converting them to a binary scale, namely DRE $>$ T1c and DRE $=$ T1c. A DRE score of T1c[19] indicates a clinically inapparent tumor which is not palpable or visible by imaging. Tumors with DRE $>$ T1c are large enough to be palpable. 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 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 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 bivariate generalized linear mixed effects sub-model. The sub-model for DRE is given by (see Panel A, Figure 4):

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

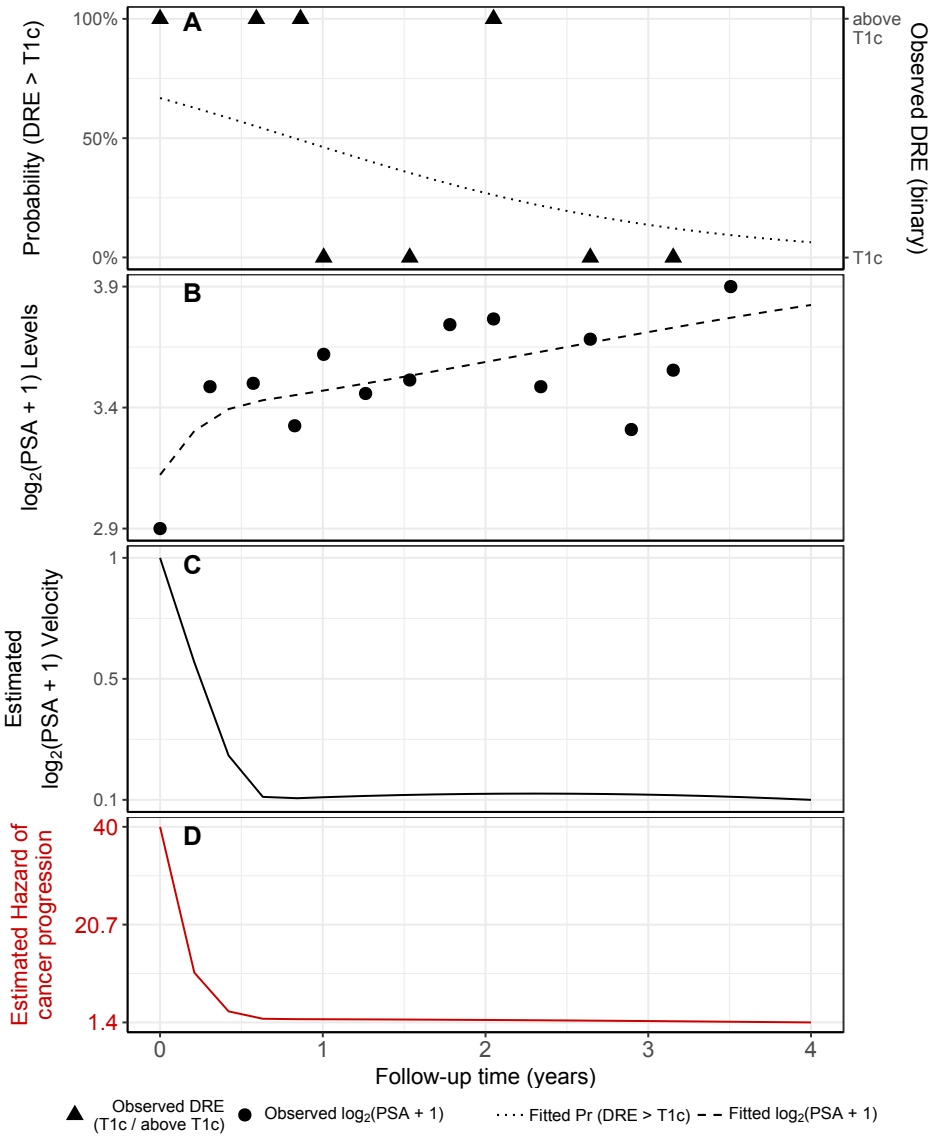

Figure 4. Illustration of the joint model fitted to the PRIAS dataset. **Panel A:** shows the observed DRE scores and the fitted probability of obtaining a DRE score greater than T1c (Equation 1) for . **Panel B:** shows the observed and fitted $\log_2(\text{PSA}+1)$ levels (Equation 2). **Panel C:** 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.

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. 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 score larger than T1c remain linear over time.

The mixed effects sub-model for PSA is given by (see Panel B, Figure 4):

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

where, $m_{pi}(t)$ denotes the underlying measurement error free value of $\log_2(\text{PSA}+1)$ transformed[20,21] measurements at time t. We model it non-linearly over time using B-splines[22]. 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). 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 4):

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

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 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 4), and b) the impact of the underlying PSA velocity $\partial m_{pi}(t)/\partial t$ (see Panel C, Figure 4). 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[23]. 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**[24]) are presented in Appendix A of the supplementary material.

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 3, at his current follow-up visit time s. Let $t \leq s$ 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 5), 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):

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

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 5). This risk can be derived from the posterior predictive distribution $g(T_j^*)$[25], and is given by:

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

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 5, biopsy at a visit may be scheduled if the personalized cumulative risk is higher than 15% (example risk threshold). This decision making process is iterated over the follow-up period, incorporating on each subsequent visit the new 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. In this regard, a straightforward approach is choosing risk thresholds utilized in standard clinical settings, e.g., 5% or 10% risk, at all follow-up visits. The number of biopsies a risk threshold may eventually lead to, is related to its accuracy of classification between patients who progress and non-progressions. The classification accuracy of a risk threshold varies over the follow-up period. This motivates an alternative approach, where at each follow-up visit a threshold with a better classification accuracy is chosen. More specifically, given the time of latest biopsy t, and the current visit time s of the new patient j, we are interested in a threshold κ, which gives the highest cancer progression detection rate (true positive rate, or TPR) for the period $(t, s]$. However, maximizing only for TPR may also lead to many unnecessary biopsy suggestions (high false positive rate, or FPR). Subsequently, balancing for the FPR, may further lead to a low number of correct detections (high false negative rate). These issues can be mitigated by maximizing the TPR and positive predictive value (PPV), simultaneously. To this end, we utilize the F_1 score, which is a composite of TPR and PPV (estimated as in Rizopoulos et al.[26]), and is defined as:

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

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, the threshold $\kappa = \arg\max_\kappa \mathrm{F}_1(t, s, \kappa)$.

It is important to note that the suggested F_1 score is not the utility of the corresponding personalized biopsy schedule for the patients. The utility on which we evaluate both fixed risk and F_1 score based personalized

schedules, is still the total number of biopsies scheduled, and the delay in detection of cancer progression (details 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 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 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 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 in it. We generate a true cancer progression time for each of the 500×1000 patients, and then sample a set of PSA and DRE measurements at the same follow-up visit times 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 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 visit and patient-specific estimated 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[3] 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 in this simulation study. Comparison of the schedules is based on the number of biopsies scheduled and the corresponding delay in 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 three fixed risk thresholds: 5%, 10% and 15%, and a risk threshold chosen using F_1 score.

Results

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 1.94 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. Detailed results pertaining to the fitted joint model are presented in Appendix B.

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 schedule, using 500×250 test patients. The corresponding median number of biopsies and delay are shown in Figure 6. The general trend is that more biopsies are required to have a smaller delay in detection. In addition, the personalized and PRIAS approaches seem to better balance the number of biopsies, and the delay, than the fixed/heuristic schedules. For brevity, we next detail the results for only the most widely used annual and PRIAS schedules, and compare them with personalized approach with risk thresholds of 5% and 10%, and threshold chosen using F_1 score. The complete set of results are presented in Appendix C.

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 do it only for the purpose of illustration. We assume that the *slow progressing* patients, are the 50% of the total population, having a cancer progression time after the ten year follow-up period of the study (see Figure 1 in Appendix A). We assume *fast progressing* patients, are the patients with an initially misdiagnosed state of cancer[27], or high risk patients who choose AS instead of immediate treatment. 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.

The boxplots in Figure 7, show the 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) due to various biopsy schedules, for these three types of patients. For *fast progressing* patients (Panel A, Figure 7), we can see that the personalized schedules with 10% risk and risk chosen using F_1 score, have a median of one biopsy compared to two biopsies for PRIAS and annual schedule. Despite this, the delay due to personalized schedule with 10% risk threshold is similar to that of PRIAS schedule.

For *intermediate progressing* patients (Panel A, Figure 7), we can see that the personalized schedule with a small risk threshold of 5% has a delay comparable to annual schedule. However, it schedules less biopsies than the annual schedule. The delay for PRIAS and personalized schedule with 10% risk is similar, but the personalized approach schedules less biopsies comparatively for at least 50% (median) of the patients.

The patients who are at most advantage with the personalized schedules are the *slow progressing* patients (50% of the total patients). In Panel C of Figure 7 we can see that the annual schedule may lead to 10 unnecessary biopsies for all such patients. The PRIAS schedule, schedules a median of 6 unnecessary biopsies. In comparison the personalized schedules using 10% risk threshold and risk chosen using F_1 score, schedule a median of only 2 and 4 biopsies, respectively.

Discussion

In this paper we proposed a personalized methodology for making decisions for prostate cancer biopsies in active surveillance (AS) programs. Our methodology uses the entire history of prostate-specific antigen (PSA) levels and digital rectal examination (DRE) scores, and results of the latest biopsy, of each patient at each follow-up visit, to make a decision of biopsy at that visit. To this end, it utilizes joint models for time-to-event and longitudinal data. Using joint models, our method combines the observed data of a patient, into a personalized cancer progression risk profile. It schedules a biopsy if the cumulative risk of cancer progression at a follow-up visit is above a certain threshold.

The proposed personalized method has the following advantages over the in-practice fixed/heuristic schedules[6,8] of biopsies. Firstly, it accounts for varying cancer progression speeds of each patient. In contrast fixed/heuristic schedules, use a common schedule for all patients, which leads to many unnecessary biopsies in the case of *slow progressing* (almost 50% of all patients, see Results) patients. Secondly, by providing urologists and patients an estimate of the risk of cancer progression, a more informed decision of biopsy can be made. The personalized method also has advantages over the in-practice PRIAS schedule of the world's largest AS program PRIAS. In PRIAS, the PSA profiles of patients are modeled linearly over time, which is contrary to observed PSA profiles of the patients (see Figure 6 in Appendix B). Consequently, PRIAS' method of scheduling more biopsies for only the *fast progressing* patients, may not always correctly identify *fast progressing* patients. To this end, we assume a non-linear profile for PSA, and utilize the complete observed data (historical PSA and DRE, time of latest biopsy) to develop the risk function (see Equation 5), which is finer quantitative measure.

We compared the personalized approach with the in-practice biopsy schedules, by conducting a realistic and extensive simulation study. We evaluated the biopsy schedules on the basis of the number of biopsies they schedule, and the corresponding delay in detection of cancer progression. From the simulation study we found that both the personalized and PRIAS schedules provide much better balance between the number of biopsies and delay in detection of progression, than the fixed/heuristic approaches (see Figure 6). Since, we conducted a simulation study, we are able to retrospectively check the impact of different schedules on patients with

different speeds of cancer progression. In this regard, we observed that the commonly used annual schedule puts the highest burden on the *slow progressing* patients (see Figure 7), by scheduling ten biopsies for each such patient. The PRIAS schedule, despite its effort to identify such patients using PSA doubling time, schedules a minimum of four and a median of six unnecessary biopsies for them. In contrast, the personalized biopsy decision making approach reduces it to a median of two to four biopsies depending upon the choice of the risk threshold.

In this regard, a personalized approach with a 10% risk threshold has a similar delay in detection of cancer progression as PRIAS, and it schedules less biopsies (see Figure 7). While, it may seem attractive for clinical use, but in light of the non-compliance results[3], prescribing the same threshold to all patients may not be suitable. Patients have varying tolerance for the number of biopsies, and apprehensions about the delay in detection of cancer progression. Both of these must be considered while making the choice of a risk threshold. Furthermore, the choice should not be made only on the basis of point estimates such as median number of biopsies and/or median delay. Instead, measures of variance (see Figure 7) of both quantities should also be considered.

A limitation of our results is that, even if they are based on the world's largest AS program, other AS programs may differ in the patient characteristics[6]. In this regard, a simulation study based on a multi-center cohort is required. Currently, the follow-up of period of our study is ten years. More detailed results, especially for *slow progressing* patients, can be obtained using a cohort having a longer follow-up period. Since Biopsy Gleason grading is susceptible to inter-observer variation[28], accounting for it in our model will also be interesting to investigate further. There is also potential for including diagnostic information from magnetic resonance imaging (MRI) in our model. Lastly, considering quality of life measures while discussing utility of each personalized approach may also lead to better decision making. However, given the scarceness of such information in the dataset, including it in the current model may not be feasible.

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. Etzioni R, Penson DF, Legler JM et al. Overdiagnosis due to prostate-specific antigen screening: lessons from us prostate cancer incidence trends. *Journal of the National Cancer Institute* 2002; 94(13): 981–990.
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. Fujita K, Landis P, McNeil BK et al. Serial prostate biopsies are associated with an increased risk of erectile dysfunction in men with prostate cancer on active surveillance. *The Journal of urology* 2009; 182(6): 2664–2669.
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. de Carvalho TM, Heijnsdijk EA and de Koning HJ. Estimating the risks and benefits of active surveillance protocols for prostate cancer: a microsimulation study. *BJU international* 2017; 119(4): 560–566.
8. Loeb S, Carter HB, Schwartz M et al. Heterogeneity in active surveillance protocols worldwide. *Reviews in urology* 2014; 16(4): 202.
9. Bokhorst LP, Valdagni R, Rannikko A et al. A decade of active surveillance in the PRIAS study: an update and evaluation of the criteria used to recommend a switch to active treatment. *European Urology* 2016; 70(6): 954–960.
10. Turnbull BW. The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society Series B (Methodological)* 1976; : 290–295.
11. Ayer T, Alagoz O and Stout NK. A POMDP approach to personalize mammography screening decisions. *Operations Research* 2012; 60(5): 1019–1034.
12. 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.

13. 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.
14. Zhang J, Denton BT, Balasubramanian H et al. Optimization of prostate biopsy referral decisions. *Manufacturing & Service Operations Management* 2012; 14(4): 529–547.
15. Krahn MD, Mahoney JE, Eckman MH et al. Screening for prostate cancer: a decision analytic view. *Jama* 1994; 272(10): 773–780.
16. Tsiatis AA and Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* 2004; 14(3): 809–834.
17. Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. CRC Press, 2012. ISBN 9781439872864.
18. Laird NM and Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; : 963–974.
19. Schröder F, Hermanek P, Denis L et al. The tnm classification of prostate cancer. *The Prostate* 1992; 21(S4): 129–138.
20. 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.
21. 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.
22. De Boor C, De Boor C, Mathématicien EU et al. *A practical guide to splines*, volume 27. Springer-Verlag New York, 1978.
23. Eilers PH and Marx BD. Flexible smoothing with B-splines and penalties. *Statistical Science* 1996; 11(2): 89–121.
24. Rizopoulos D. The R package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC. *Journal of Statistical Software* 2016; 72(7): 1–46.
25. Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* 2011; 67(3): 819–829.
26. 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.
27. 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.
28. Carlson GD, Calvanese CB, Kahane H et al. Accuracy of biopsy Gleason scores from a large uropathology laboratory: use of a diagnostic protocol to minimize observer variability. *Urology* 1998; 51(4): 525–529.

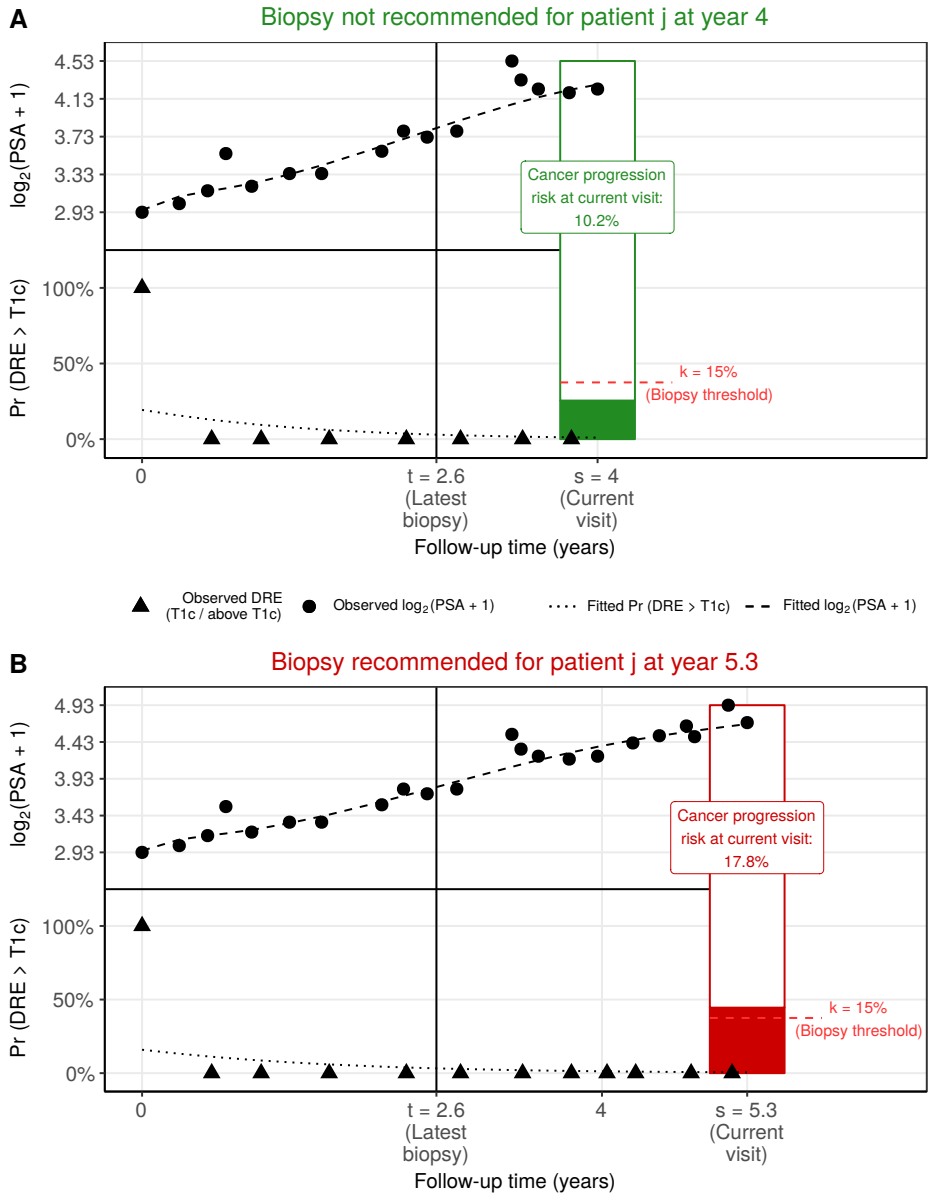

Figure 5. Illustration of personalized decision of biopsy for patient j at two different follow-up visits. Biopsy is recommended if the personalized cumulative risk of cancer progression estimated from the joint model fitted to the observed data of the patient, is higher than the example risk threshold for biopsy ($\kappa = 15\%$). **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 (10.2%) 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 (17.8%) is more than the threshold.

Prepared using sagej.cls

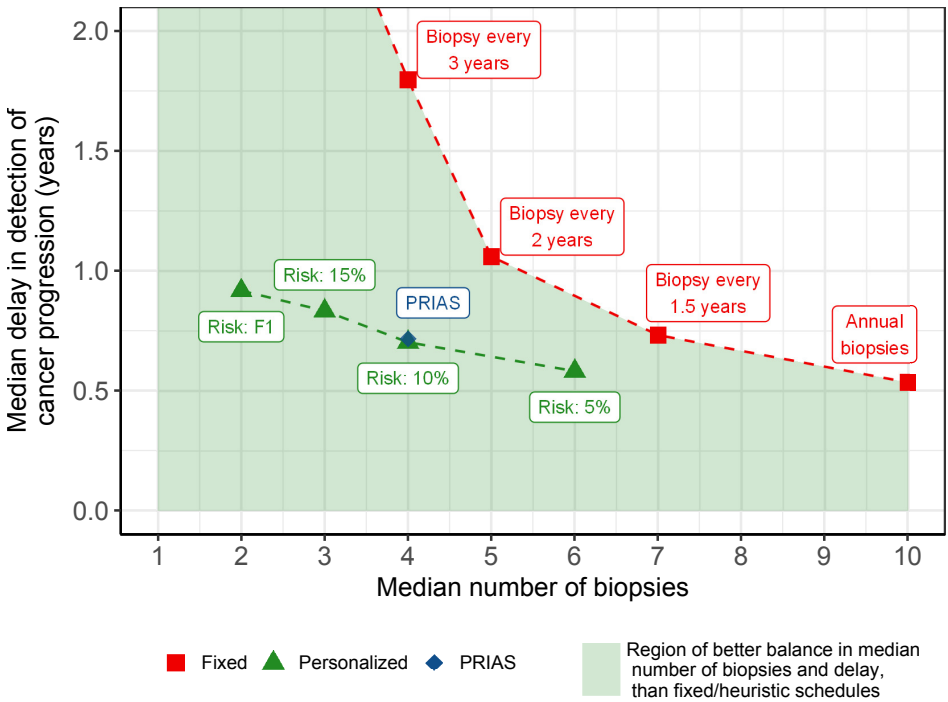

Figure 6. Simulation study results for burden-biopsy frontier: Estimated median number of biopsies, and median delay in detection of cancer progression, due to the currently practiced fixed/heuristic biopsy schedules (red squares) and PRIAS schedule (blue rhombus), and personalized schedules (green triangles), over a follow-up of ten years. The green shaded region depicts the region of better balance in number of biopsies and delay, than the currently practiced fixed/heuristic schedules. Estimation is based on results obtained from the simulation study we conducted.

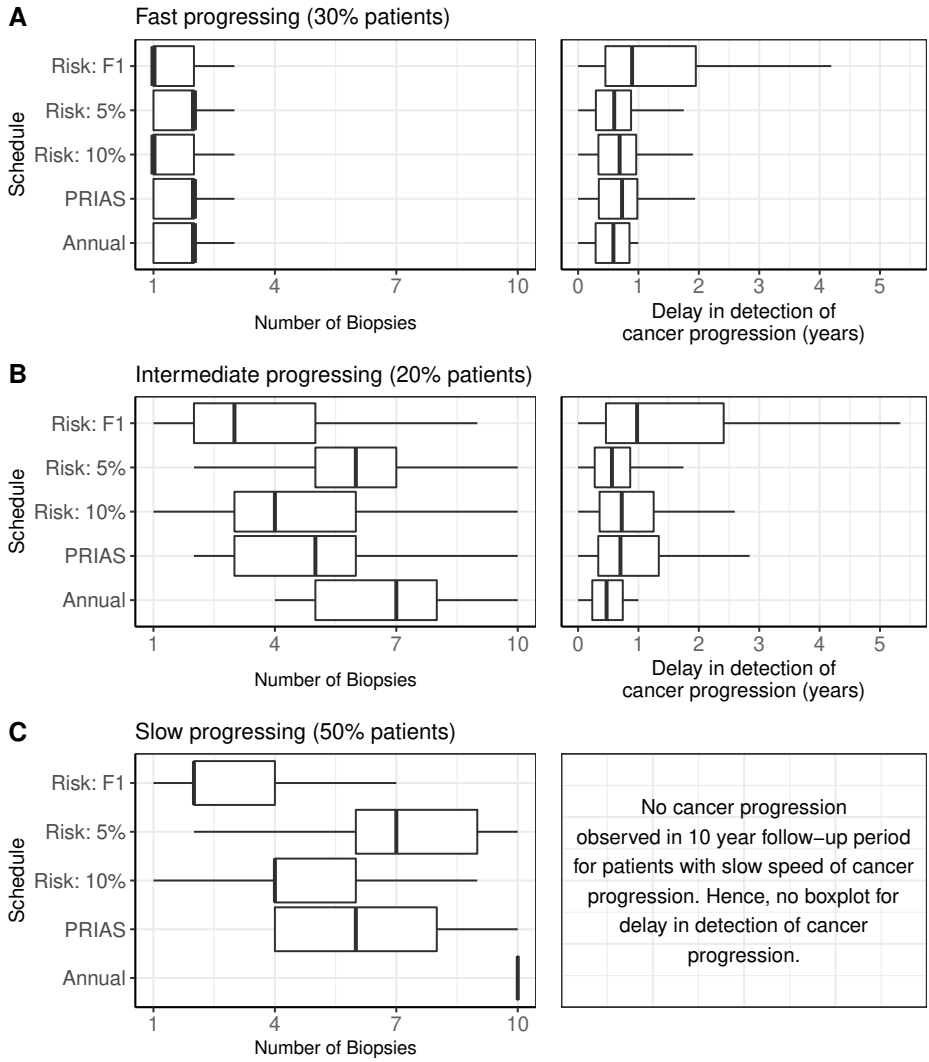

Figure 7. Boxplot showing variation in number of biopsies, and the delay in detection of cancer progression, in years (time of positive biopsy - true time of cancer progression) for various biopsy schedules. Biopsies are conducted until cancer progression is detected. **Panel A:** results for simulated 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 biopsy if the risk of cancer progression at a visit is more than 10% and 5%, respectively. Risk: F1 works similar as previous, except that the 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).