Personalized Decision Making for Biopsies in Prostate Cancer Active Surveillance Programs

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

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 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.

* *

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 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]. 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 cancer 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 Biomedical Data Sciences, Leiden University Medical Center, the Netherlands

Corresponding author:
Anirudh Tomer, Erasmus MC, t.a.v. Anirudh Tomer / kamer flex N a-2823, P O B ox 2 040, 3 000 CA Rotterdam, the Netherlands.
Email: a.tomer@erasmusmc.nl

Currently there is no consensus on the time gap between biopsies in AS[5]. Majority of the programs focus on minimizing only the delay, by scheduling biopsies annually for all patients. Annual biopsies, and other fixed/heuristic schedules[6] do not account for the difference in cancer progression speeds of patients. Although they may work well for patients who progress fast, but for slowly progressing patients, who are also the majority of the patients in AS, 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 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, and the delay in detection of 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 the pre-AS cancer screening. Their model used baseline patient characteristics as well as the latest discretized PSA level 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 progression, and PSA and DRE measurements. We then use it separately
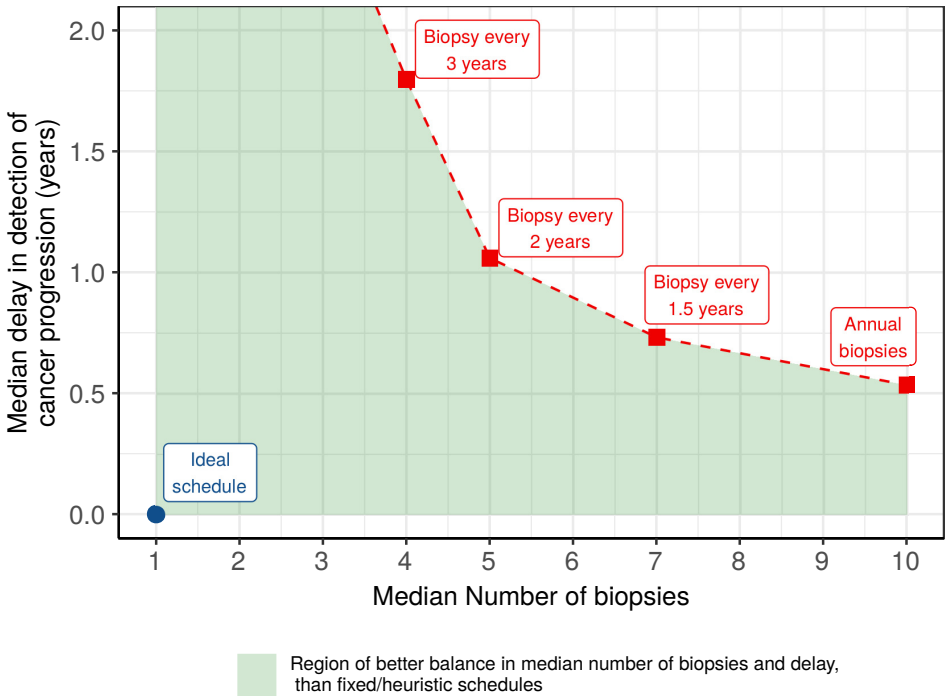
Figure 1. Median number of biopsies, and median delay in detection of cancer progression, due to various in-practice fixed/heuristic biopsy schedules (red squares), over a follow-up of ten years. Using personalized decision making for biopsies, we intend to better balance (green region) the number of biopsies and the delay, than in-practice schedules. An ideal biopsy schedule (blue circle) will schedule only one biopsy, exactly at the true time of cancer progression.

for each patient at each follow-up visit to develop a cancer progression risk profile, based on their observed data. If that risk is higher than a certain threshold, our method proposes a biopsy at the same follow-up visit. Since there is no clear consensus on choice of risk thresholds, we do not only use fixed risk 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 fixed/heuristic schedules, and PRIAS schedule, 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.
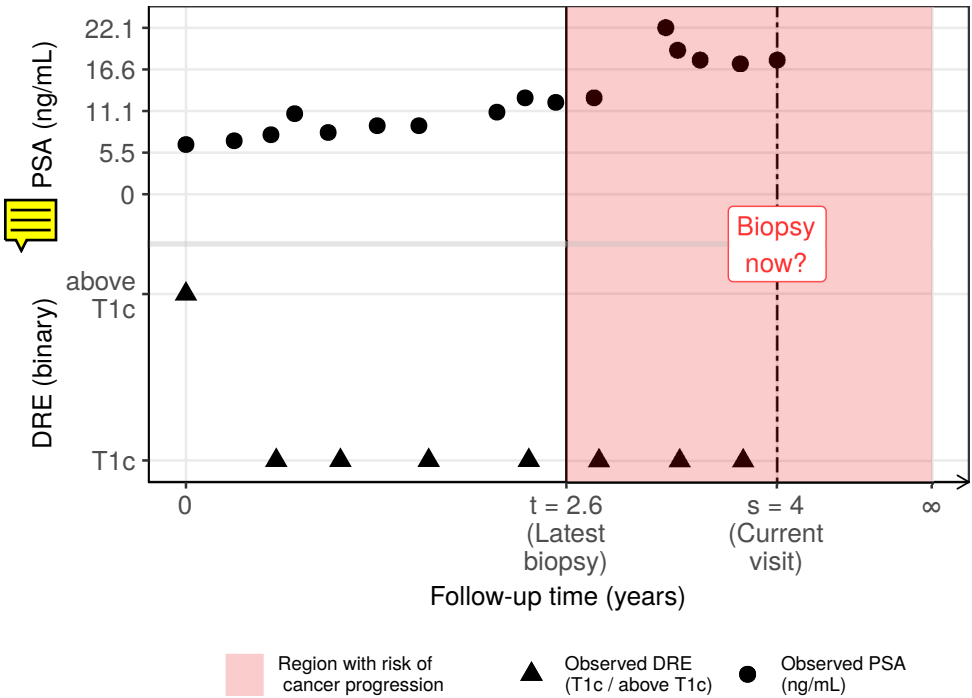
Figure 2. Illustration of the decision making problem: Available data of a patient j, who had his latest (negative) biopsy at $t = 2.6$ years. The shaded region shows the time period in which the patient is at the risk of cancer progression. His current follow-up visit is at $s = 4$ years. Using his entire history of DRE $\mathcal{Y}_{dj}(s)$ and PSA $\mathcal{Y}_{pj}(s)$ measurements up to the visit s, and the time of the latest biopsy t, we intend to make a decision on scheduling a biopsy at s.

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 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 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 $\mathcal{D}_n = \{l_i, r_i, \boldsymbol{y}_{di}, \boldsymbol{y}_{pi}; i = 1, \dots, n\}$.

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 3):

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

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. The fixed effect parameters are denoted by $\{\beta_{0d}, \dots, \beta_{3d}\}$, 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 3):

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

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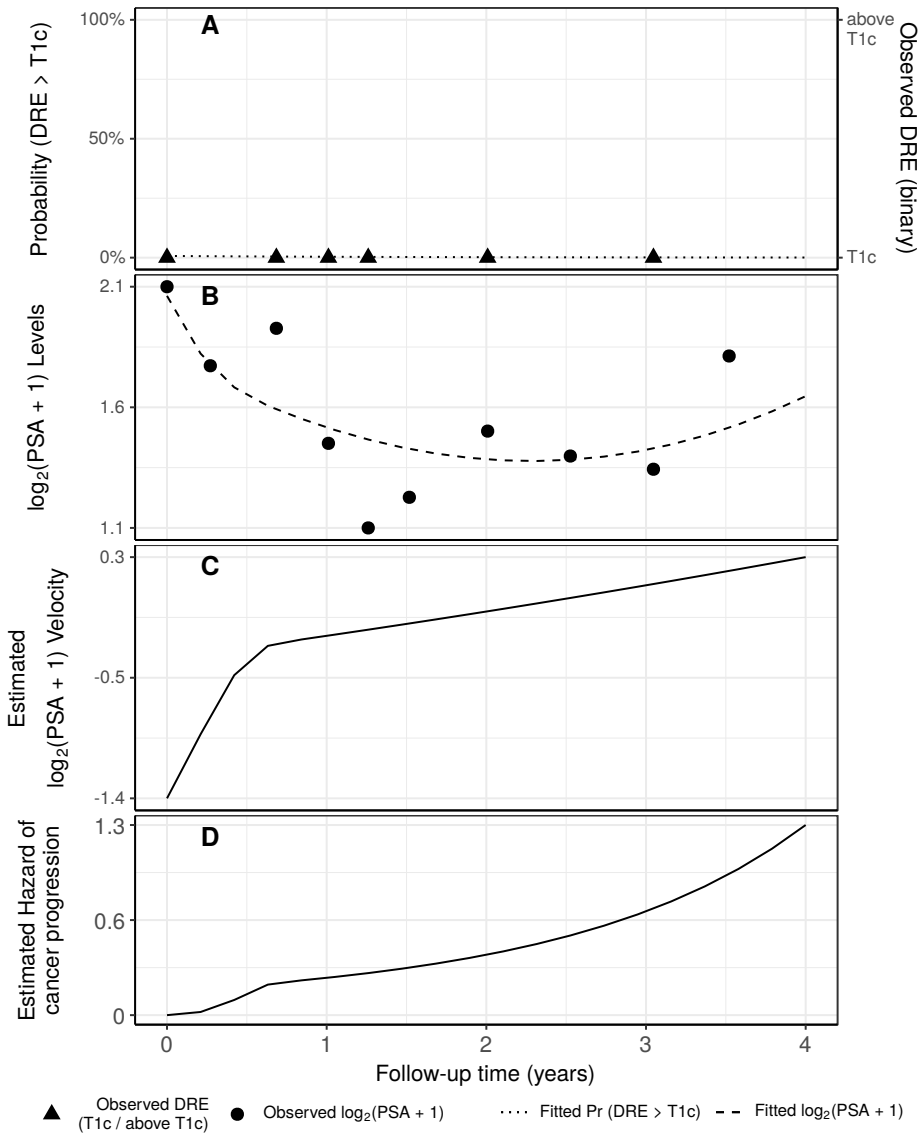
Figure 3. Illustration of the joint model fitted to the PRIAS dataset. **Panel A:** shows the observed DRE scores and the fitted probability of obtaining a DRE score greater than T1c (Equation 1) for . **Panel B:** shows the observed and fitted $\log_2(\text{PSA} + 1)$ levels (Equation 2). **Panel C:**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

it 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 (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 3):

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

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 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

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, respectively, taken up to the current visit s. 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^* (see Appendix A.4 for details):

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

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 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:

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

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, incorporating on each subsequent visit the new observed data, until a positive biopsy is observed. This leads to an entire personalized schedule of biopsies for each patient.

The choice of risk threshold dictates the schedule of biopsies. In this regard, an elementary 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.[21]), and is defined as:

$$
\begin{aligned}
F_1(t, s, \kappa) &= 2 \frac{\text{TPR}(t, s, \kappa) \, \text{PPV}(t, s, \kappa)}{\text{TPR}(t, s, \kappa) + \text{PPV}(t, s, \kappa)}, \\
\text{TPR}(t, s, \kappa) &= \Pr\big\{ R_j(s \mid t) > \kappa \mid t < T_j^* \le s \big\}, \\
\text{PPV}(t, s, \kappa) &= \Pr\big\{ t < T_j^* \le s \mid R_j(s \mid t) > \kappa \big\}.
\end{aligned}
\tag{6}
$$

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 $\kappa = \arg\max_{\kappa} F_1(t, s, \kappa)$.

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 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 the fitted joint models, we obtain the cancer progression risk profiles for each of the 500×250 test patients at each of their visits. We use the personalized risk profiles to make the decision of biopsy.

In this simulation study, we evaluate the following in-practice fixed/heuristic approaches for biopsies: biopsy annually, biopsy every one and a half years, biopsy every two years and biopsy every three years. For the personalized biopsy approach we evaluate three fixed risk thresholds: 5%, 10% and 15%, and a risk threshold chosen using F_1 score. Lastly, we also evaluate the PRIAS schedule of biopsies. We evaluate and compare these schedules on the basis of the number of biopsies they schedule and the corresponding delay in detection of cancer progression.

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 1.92 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. 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 5. 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[22], 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 6, 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 6), 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 6), 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 6 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[5,6] 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 5). 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 6), 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 6). 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 6) 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[23], 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

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.

23. 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.

A

Biopsy not recommended for patient j at year 4

B

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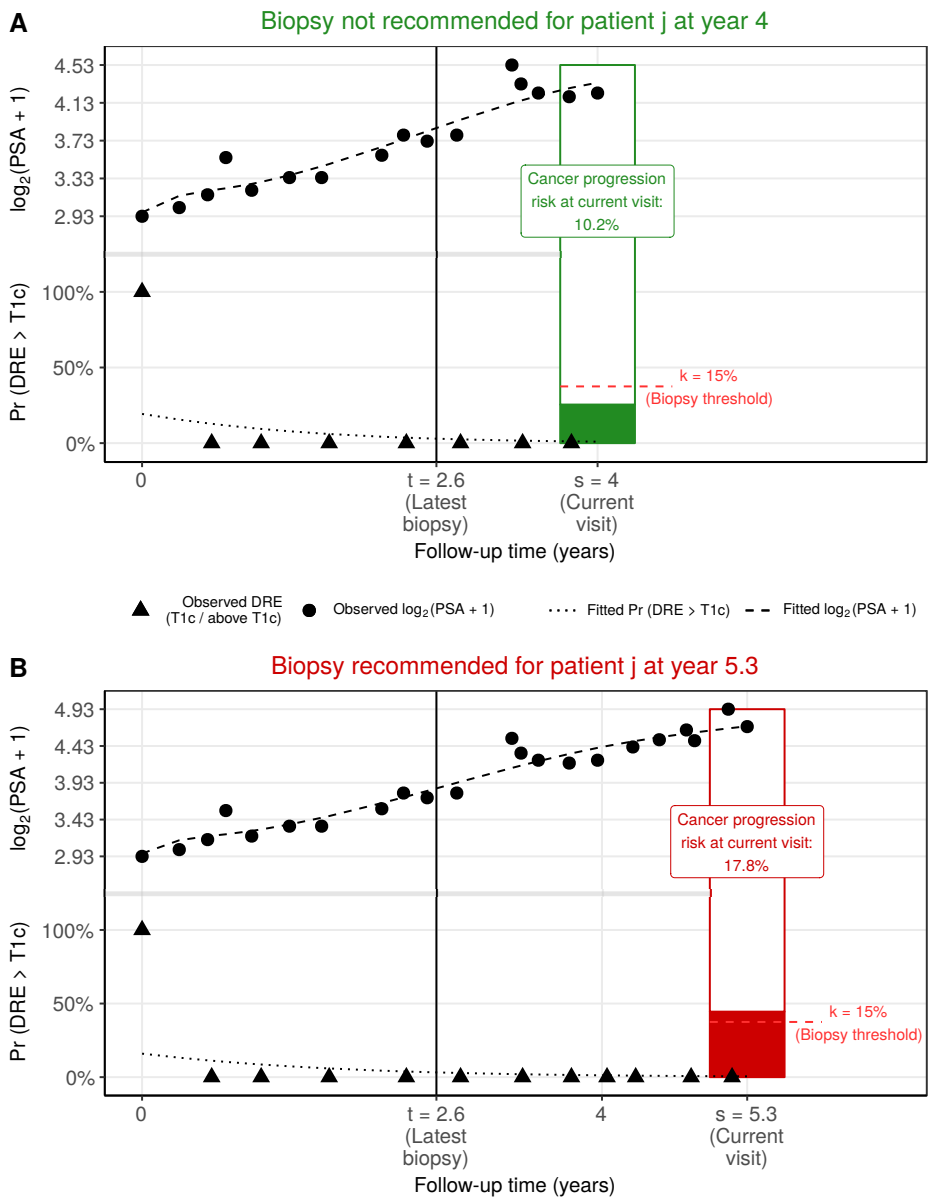
Figure 4. Illustration of personalized decision of biopsy for patient j at two different follow-up visits. Biopsy is recommended if the risk of cancer progression estimated from the joint model fitted to the observed data of the patient, is higher than the example risk threshold for biopsy ($\kappa =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

*Prepared using **sagej.cls***

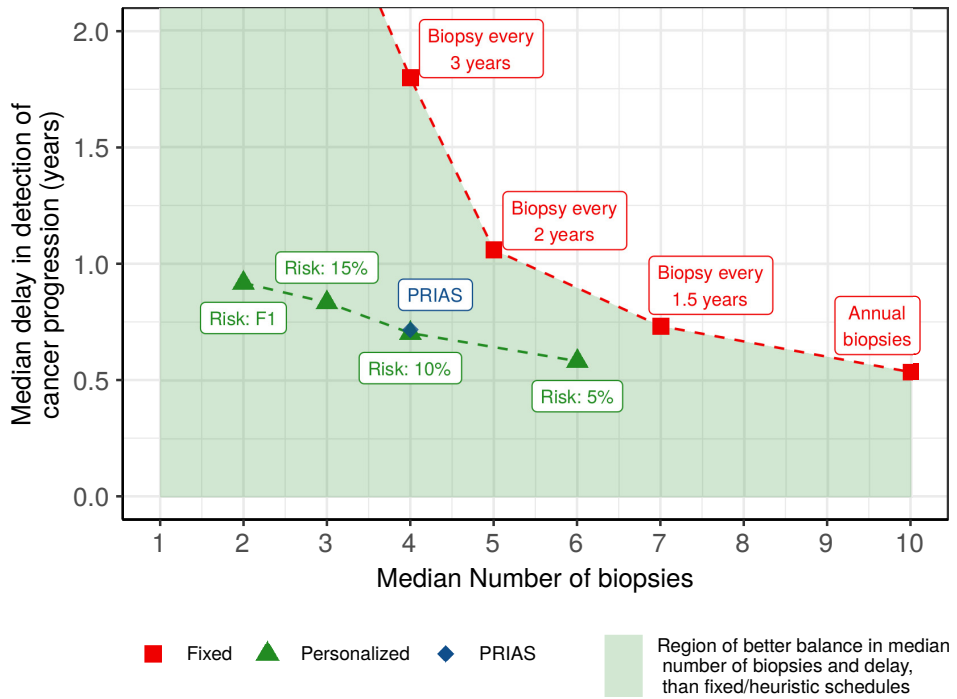**Figure 5.** Median number of biopsies, and median delay in detection of cancer progression, due to various in-practice fixed/heuristic biopsy schedules (red squares), PRIAS schedule (blue rhombus), and personalized schedules (green triangles), over a follow-up of ten years. The green shaded region depicts the region of better balance in number of biopsies and delay, than the in-practice fixed/heuristic schedules.
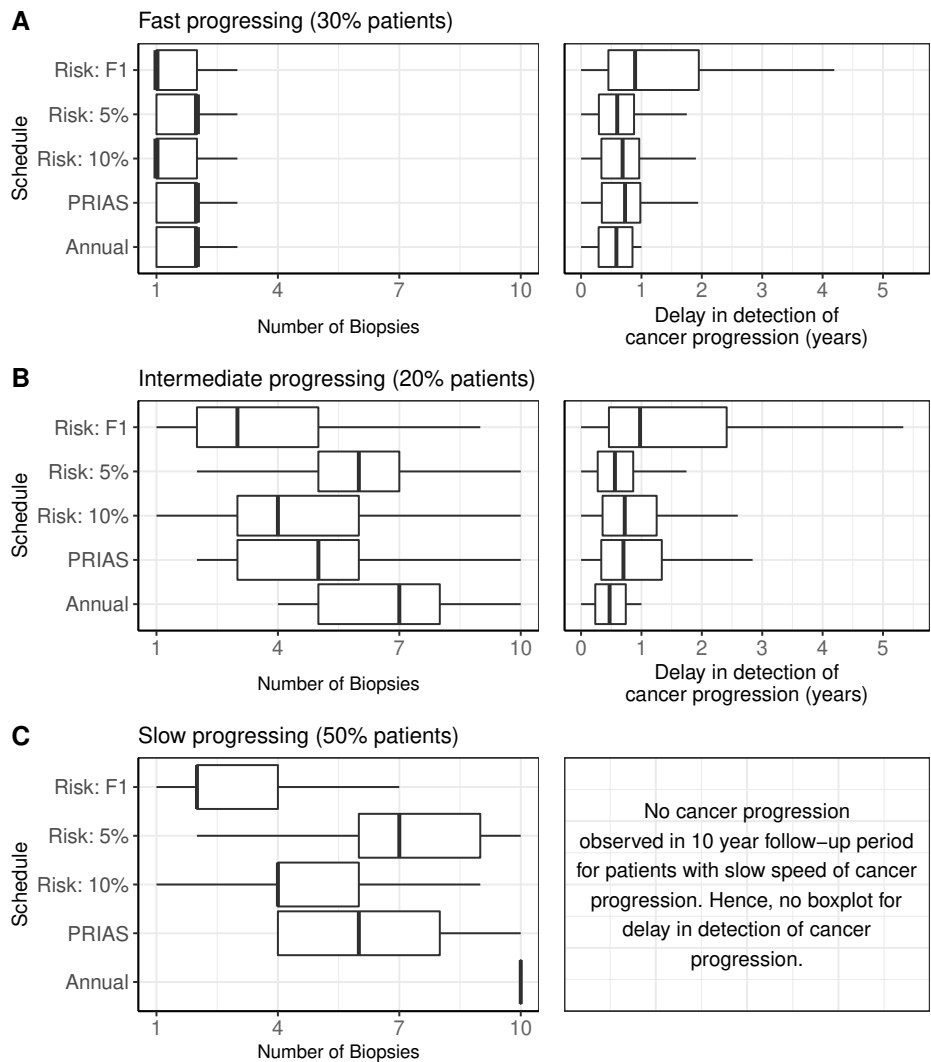
A

Fast progressing (30% patients)

B

Intermediate progressing (20% patients)

C

Slow progressing (50% patients)

Figure 6. 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).

Supplementary Materials for "Personalized Decision Making for Biopsies in Prostate Cancer Active Surveillance Programs"

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

[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

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

In this appendix section, we first provide a short introduction to the world's largest active surveillance (AS) program called Prostate Cancer Research International Active Surveillance, abbreviated as PRIAS (Bul et al., 2013), that we use to develop our methodology. We then present an introduction to the joint models for time-to-event and longitudinal data (Rizopoulos, 2012; Tsiatis and Davidian, 2004), that we fit to the PRIAS dataset. Lastly, we present the parameter estimation for our model using the Bayesian approach.

Appendix A.1   PRIAS Dataset

The PRIAS dataset consists of 5270 AS patients, of which 866 observe cancer progression. For each patient, prostate-specific antigen (PSA) measurements (ng/mL) are scheduled every 3 months for first 2 years and every 6 months thereafter. The DRE measurements are scheduled every 6 months. We use the DRE measurements after converting them on a binary scale, namely DRE > T1c and DRE ≤ T1c (Schröder et al., 1992). On average 5 DRE and 9 PSA measurements have been recorded per patient. Larger values for PSA and/or larger score for DRE, may indicate cancer progression. However, it is the occurrence of biopsy Gleason score larger than 6, that is commonly considered as cancer progression. In the PRIAS study, biopsies are scheduled at the following fixed follow-up times (measured since inclusion in AS): year 1, 4, 7, and 10, and every 5 years thereafter. An annual schedule of biopsies is prescribed to those patients who have a PSA doubling time between 0 and 10 years. The PSA doubling time at any point during follow-up is measured as

the inverse of the slope of the regression line through the base two logarithm of the observed PSA values.

In Figure 1, we show the survival probability of cancer progression estimated using the cancer progression times observed for the patients from the PRIAS program. Prostate cancer is a slowly progressing disease, which is evident by the 50% probability of not having a cancer progression by the end of the ten year follow-up period.
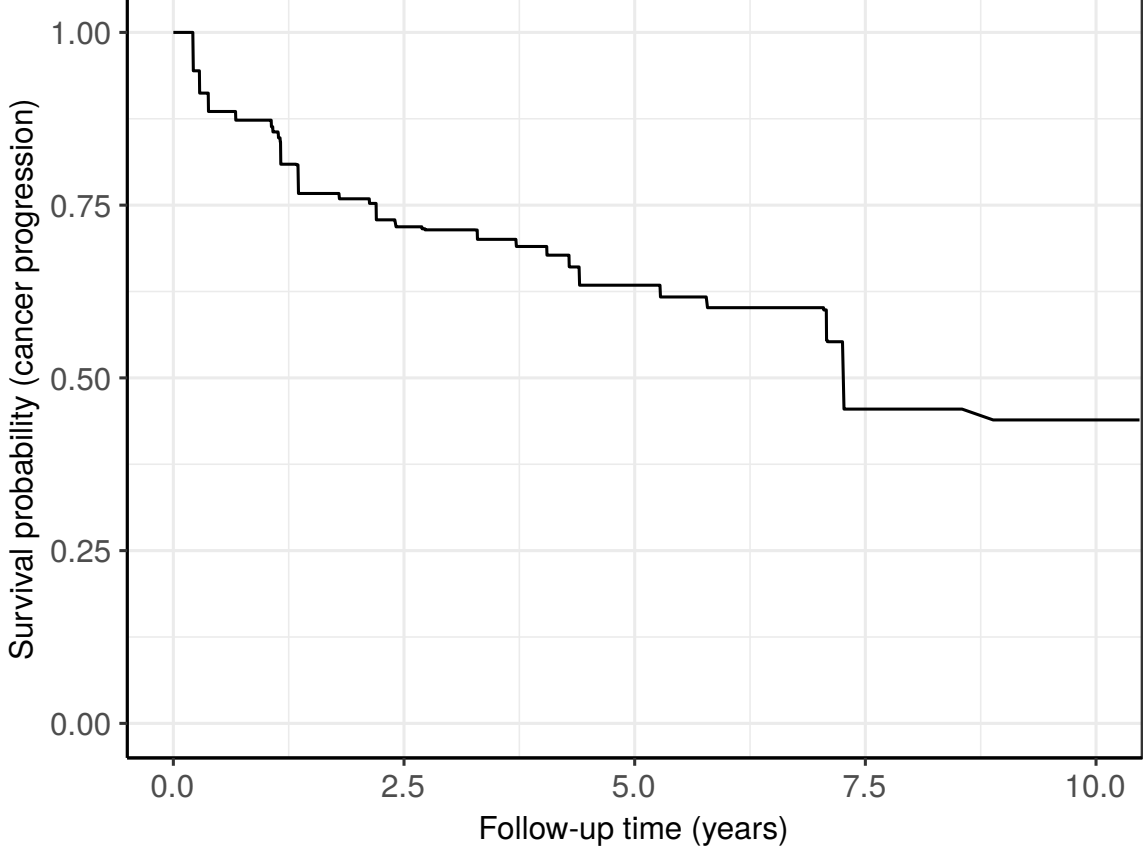

Figure 1: Estimated survival probability of cancer progression for patients in the PRIAS dataset. Estimation is done using a nonparametric maximum likelihood estimate (NPMLE), of the distribution function for interval censored cancer progression times (Turnbull, 1976).

Appendix A.2   Model Definition

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 $r_i = \infty$. Further, let $\boldsymbol{y}_{di}$, and $\boldsymbol{y}_{pi}$ denote the $n_{di} \times 1$, and $n_{pi} \times 1$ vectors of the 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 2):
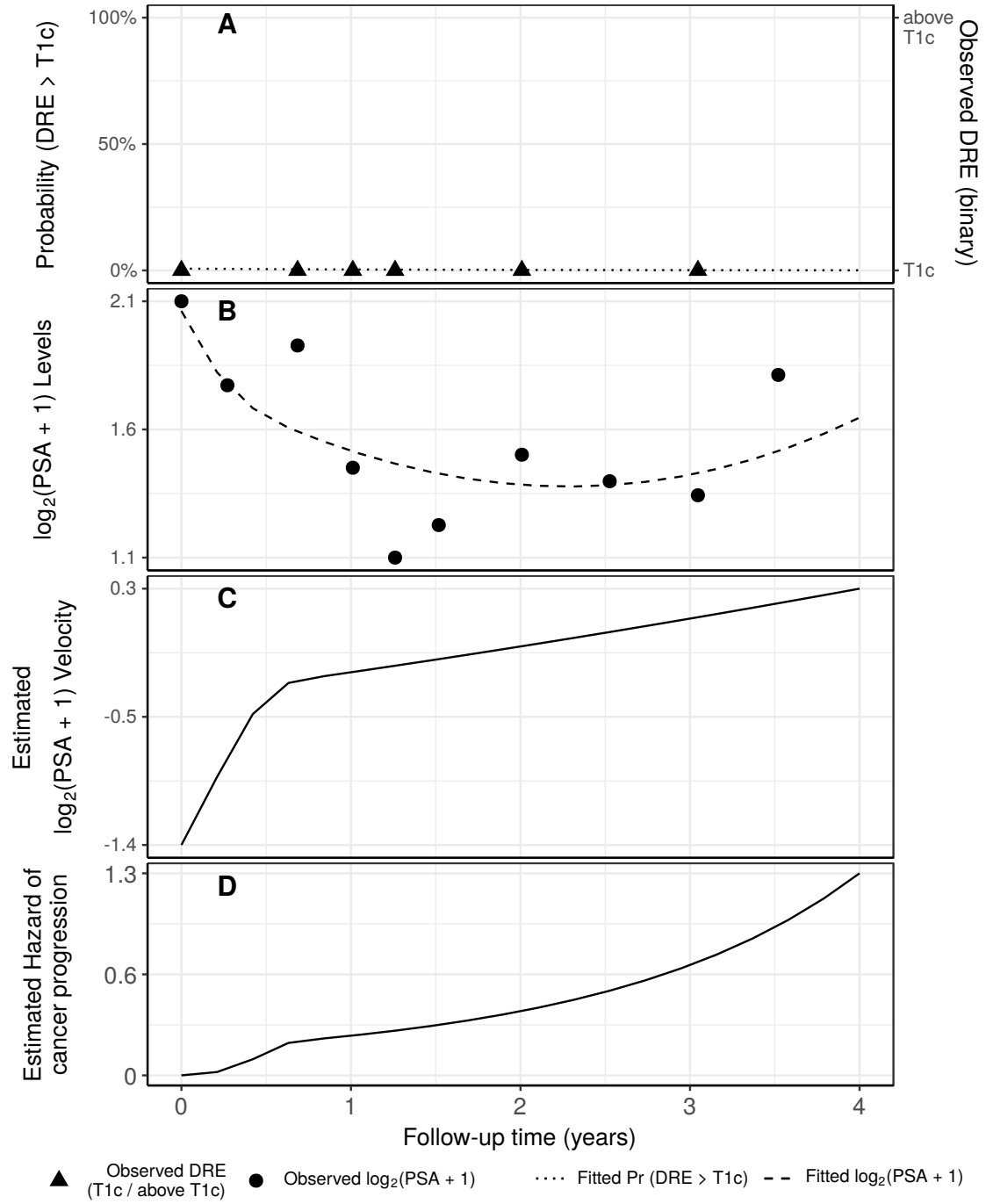
$$\text{logit}\big[\Pr\{y_{di}(t) > \text{T1c}\}\big] = \beta_{0d} + b_{0di} + (\beta_{1d} + b_{1di})t$$
$$+ \beta_{2d}(\text{Age}_i - 70) + \beta_{3d}(\text{Age}_i - 70)^2 \quad (1)$$

Figure 2: **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 2):

$$\log_2\left\{y_{pi}(t) + 1\right\} = m_{pi}(t) + \varepsilon_{pi}(t),$$

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

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 et al., 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). 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 2):

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

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 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 2), and b) the impact of the underlying PSA velocity $\partial m_{pi}(t)/\partial t$ (see Panel C, Figure 2). 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). More specifically:

$$\log h_0(t) = \gamma_{h_0,0} + \sum_{q=1}^{Q}\gamma_{h_0,q}B_q(t, \boldsymbol{v}),$$

where $B_q(t, \boldsymbol{v})$ denotes the q-th basis function of a B-spline with knots $\boldsymbol{v} = v_1, \ldots, v_Q$ and vector of spline coefficients γ_{h_0}. To avoid choosing the number and position of knots in the spline, a relatively high number of knots (e.g., 15 to 20) are chosen and the corresponding B-spline regression coefficients γ_{h_0} are penalized using a differences penalty (Eilers and Marx, 1996). An example fitted hazard is shown in panel D of Figure 2.

Appendix A.3   Parameter Estimation

We estimate the parameters of the joint model using Markov chain Monte Carlo (MCMC) methods under the Bayesian framework. Let $\boldsymbol{\theta}$ denote the vector of all

of the parameters of the joint model. The joint model postulates that given the random effects, the time to cancer progression, and the PSA and DRE measurements taken over time are all mutually independent. Under this assumption the posterior distribution of the parameters is given by:

$$p(\boldsymbol{\theta}, \boldsymbol{b} \mid \mathcal{D}_n) \propto \prod_{i=1}^{n} p(l_i, r_i, \boldsymbol{y}_{di}, \boldsymbol{y}_{pi}, \mid \boldsymbol{b}_i, \boldsymbol{\theta}) p(\boldsymbol{b}_i \mid \boldsymbol{\theta}) p(\boldsymbol{\theta})$$

$$\propto \prod_{i=1}^{n} p(l_i, r_i \mid \boldsymbol{b}_i, \boldsymbol{\theta}) p(\boldsymbol{y}_{di} \mid \boldsymbol{b}_i, \boldsymbol{\theta}) p(\boldsymbol{y}_{pi} \mid \boldsymbol{b}_i, \boldsymbol{\theta}) p(\boldsymbol{b}_i \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}),$$

$$p(\boldsymbol{b}_i \mid \boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^q \det(\boldsymbol{D})}} \exp(\boldsymbol{b}_i^T \boldsymbol{D}^{-1} \boldsymbol{b}_i),$$

where, the likelihood contribution of the DRE outcome, conditional on the random effects is:

$$p(\boldsymbol{y}_{di} \mid \boldsymbol{b}_i, \boldsymbol{\theta}) = \prod_{k=1}^{n_{di}} \frac{\exp\left[-\text{logit}\{\Pr(y_{dik} > \text{T1c})\} I(y_{dik} = \text{T1c}) \right]}{1 + \exp\left[-\text{logit}\{\Pr(y_{dik} > \text{T1c})\} \right]},$$

where $I(\cdot)$ is an indicator function which takes the value 1 if the k-th repeated DRE score $y_{dik} = \text{T1c}$, and takes the value 0 otherwise. The likelihood contribution of the PSA outcome, conditional on the random effects is:

$$p(\boldsymbol{y}_{pi} \mid \boldsymbol{b}_i, \boldsymbol{\theta}) = \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^{n_{pi}}} \exp\left(-\frac{\|\boldsymbol{y}_{pi} - \boldsymbol{m}_{pi}\|^2}{\sigma^2} \right),$$

The likelihood contribution of the time to cancer progression outcome is given by:

$$p(l_i, r_i \mid \boldsymbol{b}_i, \boldsymbol{\theta}) = \exp\left\{ -\int_0^{l_i} h_i(s)\mathrm{d}s \right\} - \exp\left\{ -\int_0^{r_i} h_i(s)\mathrm{d}s \right\}. \tag{4}$$

The integral in (4) does not have a closed-form solution, and therefore we use a 15-point Gauss-Kronrod quadrature rule to approximate it.

We use independent normal priors with zero mean and variance 100 for the fixed effects $\{\beta_{0d}, \ldots, \beta_{3d}, \beta_{0p}, \ldots, \beta_{6p}\}$, and inverse Gamma prior with shape and rate both equal to 0.01 for the parameter σ^2. For the variance-covariance matrix $\boldsymbol{D}$ of the random effects we take inverse Wishart prior with an identity scale matrix and degrees of freedom equal to 7 (number of random effects). For the relative risk model's parameters $\{\gamma_1, \gamma_2\}$ and the association parameters $\{\alpha_{1d}, \alpha_{1p}, \alpha_{2p}\}$, we use independent normal priors with zero mean and variance 100.

Appendix A.4   Personalized Posterior Predictive Distribution of Time of Cancer Progression

Let us assume a new patient j, for whom we need to make a personalized biopsy decision. Let his current follow-up visit time be s, latest time of biopsy be $t \leq s$, observed vectors of DRE and PSA measurements be $\mathcal{Y}_{dj}(s)$ and $\mathcal{Y}_{pj}(s)$, respectively. The combined information from the observed data is given by the following posterior predictive distribution $g(T_j^*)$ of his time of cancer progression T_j^*:

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

$$= \int \int p(T_j^* \mid T_j^* > t, \boldsymbol{b}_j, \boldsymbol{\theta})$$

$$\times p\{\boldsymbol{b}_j \mid T_j^* > t, \mathcal{Y}_{dj}(s), \mathcal{Y}_{pj}(s), \boldsymbol{\theta}\} p(\boldsymbol{\theta} \mid \mathcal{D}_n)\mathrm{d}\boldsymbol{b}_j\mathrm{d}\boldsymbol{\theta}.$$

The distribution $g(T_j^*)$ depends on observed data of the patient $\mathcal{Y}_{dj}(s)$ and $\mathcal{Y}_{pj}(s)$, as well information from the PRIAS dataset $\mathcal{D}_n$ via the posterior distribution of random effects $\boldsymbol{b}_j$ and posterior distribution of the vector of all parameters $\boldsymbol{\theta}$, respectively.

The distribution can be estimated as detailed in Rizopoulos, Molenberghs, and Lesaffre, (2017). However, majority of the prostate cancer patients do not progress in the ten year follow-up period of PRIAS (see Figure 1). Consequently, the personalized density function of cancer progression $g(T_j^*)$ can only be estimated for time points falling within the ten year follow-up.

Appendix B   Parameter Estimates from the Joint Model Fitted to the PRIAS Dataset

We fit a joint model to the PRIAS dataset using the R package **JMbayes** (Rizopoulos, 2016). The corresponding posterior parameter estimates are shown in Table 2 (longitudinal sub-model for DRE outcome), Table 3 (longitudinal sub-model for PSA outcome) and Table 4 (relative risk sub-model). The parameter estimates for the variance-covariance matrix $\boldsymbol{D}$ from the longitudinal sub-model are shown in the following Table 1:

Table 1:   Estimated variance-covariance matrix $\boldsymbol{D}$ of the random effects $\boldsymbol{b} = (b_{0d}, b_{1d}, b_{0p}, b_{1p}, b_{2p}, b_{3p}, b_{4p})$ (see Appendix A.2) from the joint model fitted to the PRIAS dataset. The variances of the random effects are highlighted along the diagonal of the variance-covariance matrix.

Random Effects	b_{0d}	b_{1d}	b_{0p}	b_{1p}	b_{2p}	b_{3p}	b_{4p}
b_{0d}	7.546	-0.564	-0.182	0.075	0.084	0.003	-0.019
b_{1d}	-0.564	1.379	0.081	0.119	0.165	0.266	0.219
b_{0p}	-0.182	0.081	0.208	0.031	0.034	0.068	0.014
b_{1p}	0.075	0.119	0.031	0.224	0.109	0.158	0.088
b_{2p}	0.084	0.165	0.034	0.109	0.293	0.324	0.238
b_{3p}	0.003	0.266	0.068	0.158	0.324	0.480	0.312
b_{4p}	-0.019	0.219	0.014	0.088	0.238	0.312	0.290

For the DRE mixed effects sub-model (see Equation 1) parameter estimates, in Table 2 we can see that the age of the patient trivially affects the baseline log odds of obtaining a DRE score larger than T1c. In Figure 3 we present the marginal evolution of log odds of obtaining a DRE larger than T1c, and the corresponding marginal probability, over a period of 10 years for a hypothetical AS patient who is included in AS at the age of 70 years. In addition, we present plots of observed DRE versus fitted probabilities of obtaining a DRE score larger than T1c, for nine randomly selected patients in Figure 4.

Table 2: Estimated mean and 95% credible interval for the parameters of the longitudinal sub-model (see Equation 1) for the DRE outcome.

Variable	Mean	Std. Dev	2.5%	97.5%	P
(Intercept)	-4.017	0.136	-4.270	-3.763	<0.000
$(\text{Age} - 70)$	0.058	0.009	0.041	0.075	<0.000
$(\text{Age} - 70)^2$	-0.001	0.001	-0.003	1.8×10^{-4}	0.076
visitTimeYears	-0.604	0.095	-0.794	-0.437	<0.000
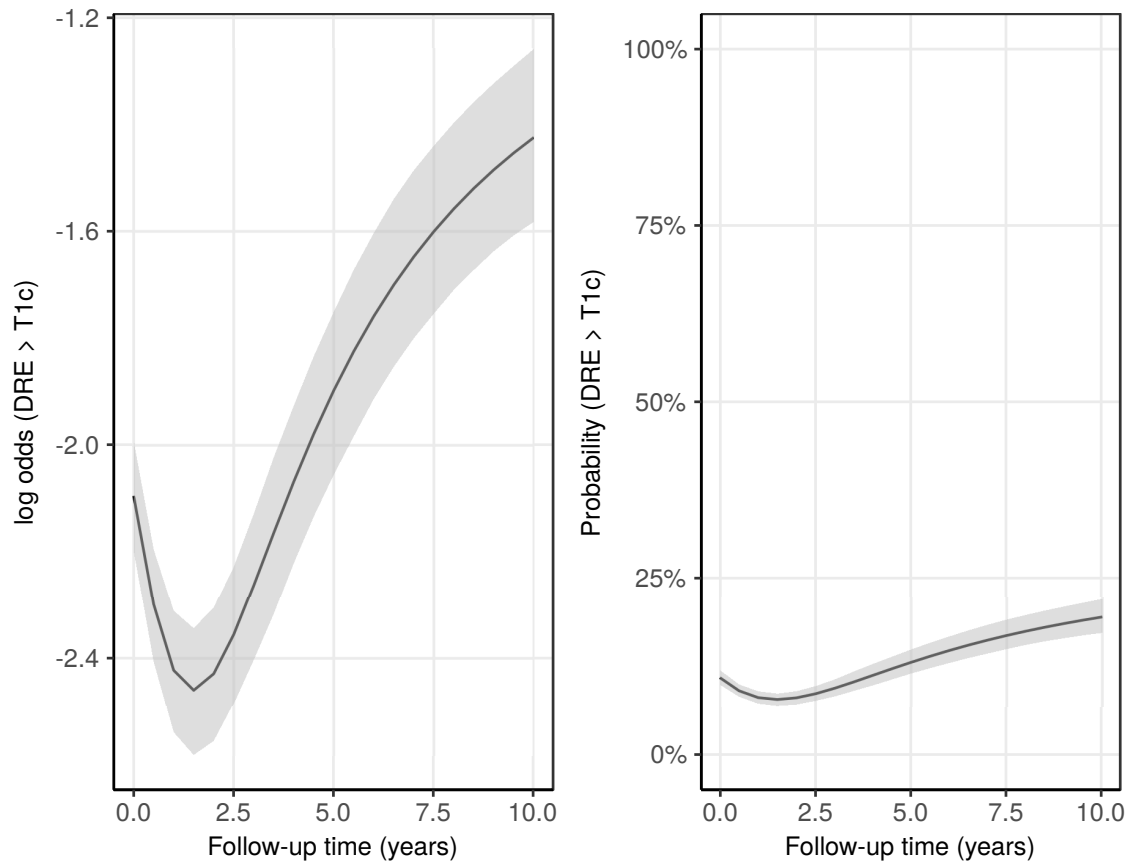
Figure 3: Fitted marginal evolution of the log odds of obtaining a DRE larger than T1c, and the corresponding marginal probability, with 95% credible interval. These results are for a hypothetical AS patient who is included in AS at the age of 70 years.

Figure 4: Observed DRE versus fitted probabilities of obtaining a DRE score larger than T1c, for nine randomly selected PRIAS patients. The fitted profiles utilize information from the observed DRE scores, PSA levels, and time of the latest biopsy. Observed DRE scores plotted against 0% probability are equal to T1c. Observed DRE scores plotted against 100% probability are larger than T1c.

For the PSA mixed effects sub-model parameter estimates (see Equation 2), in Table 3 we can see that the age of the patient trivially affects the 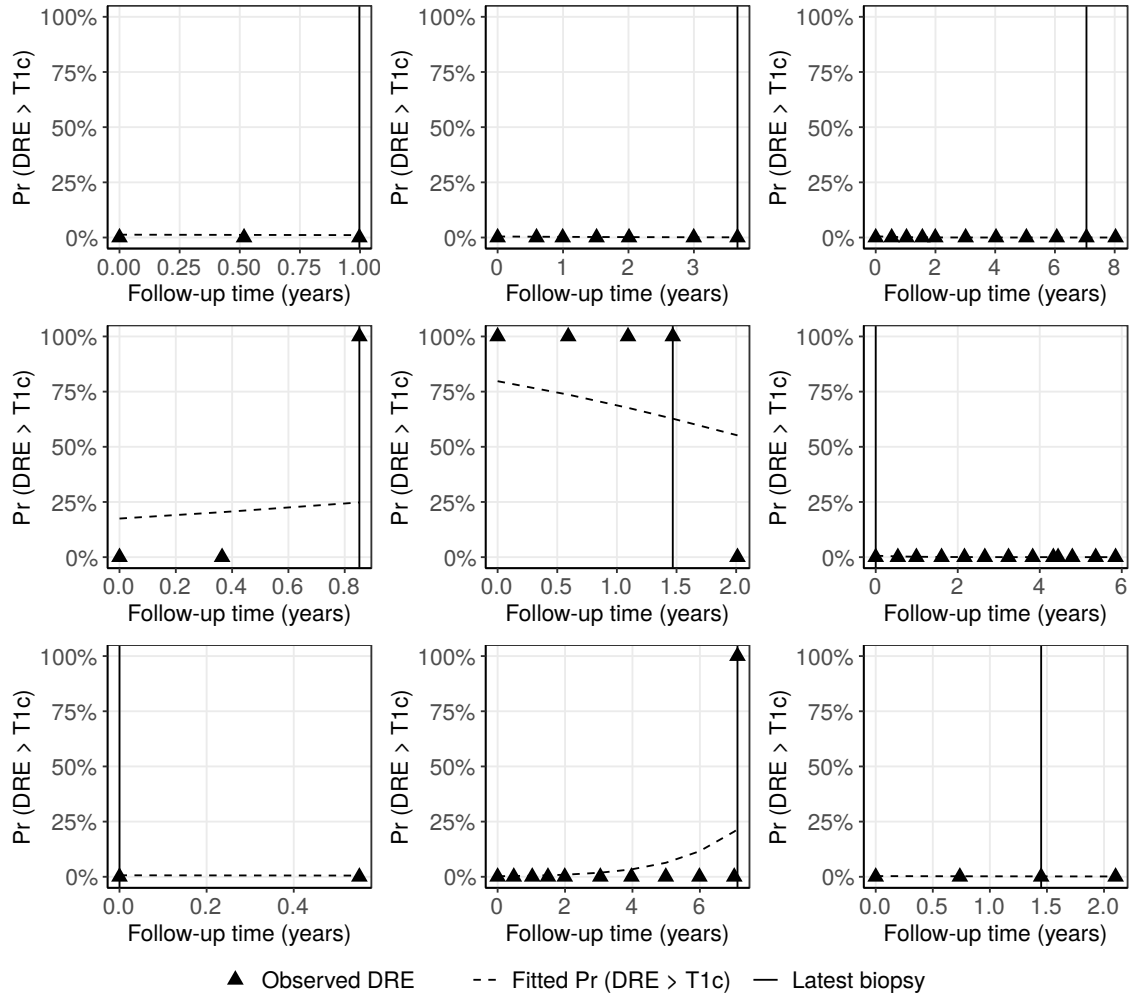baseline $\log_2(\text{PSA} + 1)$ level. Since the longitudinal evolution of $\log_2(\text{PSA} + 1)$ levels is modeled with non-linear terms, the interpretation of the coefficients corresponding to time is not straightforward. In lieu of the interpretation, in Figure 5 we present the fitted marginal evolution of $\log_2(\text{PSA} + 1)$ over a period of 10 years for a hypothetical patient who is included in AS at the age of 70 years. In addition, we present plots of observed versus fitted PSA profiles for nine randomly selected patients in Figure 6.

Table 3: Estimated mean and 95% credible interval for the parameters of the longitudinal sub-model (see Equation 2) for the PSA outcome.

Variable	Mean	Std. Dev	2.5%	97.5%	P
(Intercept)	2.701	0.008	2.686	2.716	<0.000
$(\text{Age} - 70)$	0.003	0.001	0.001	0.005	<0.000
$(\text{Age} - 70)^2$	-4.7×10^{-4}	9.8×10^{-5}	-6.6×10^{-4}	-2.7×10^{-4}	<0.000
Spline: $[0.00, 0.10]$ years	0.054	0.009	0.037	0.073	<0.000
Spline: $[0.10, 0.70]$ years	0.177	0.012	0.151	0.200	<0.000
Spline: $[0.70, 4.00]$ years	0.194	0.016	0.161	0.225	<0.000
Spline: $[4.00, 5.42]$ years	0.341	0.015	0.312	0.371	<0.000
σ	0.137	0.001	0.135	0.138	

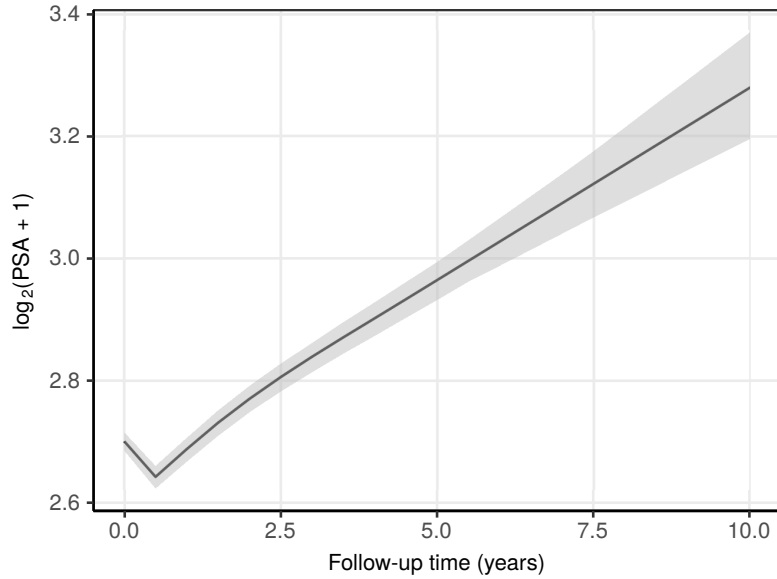Figure 5: Fitted marginal evolution of $\log_2(\text{PSA} + 1)$ levels over a period of 10 years with 95% credible interval, for a hypothetical patient who is included in AS at the age of 70 years.
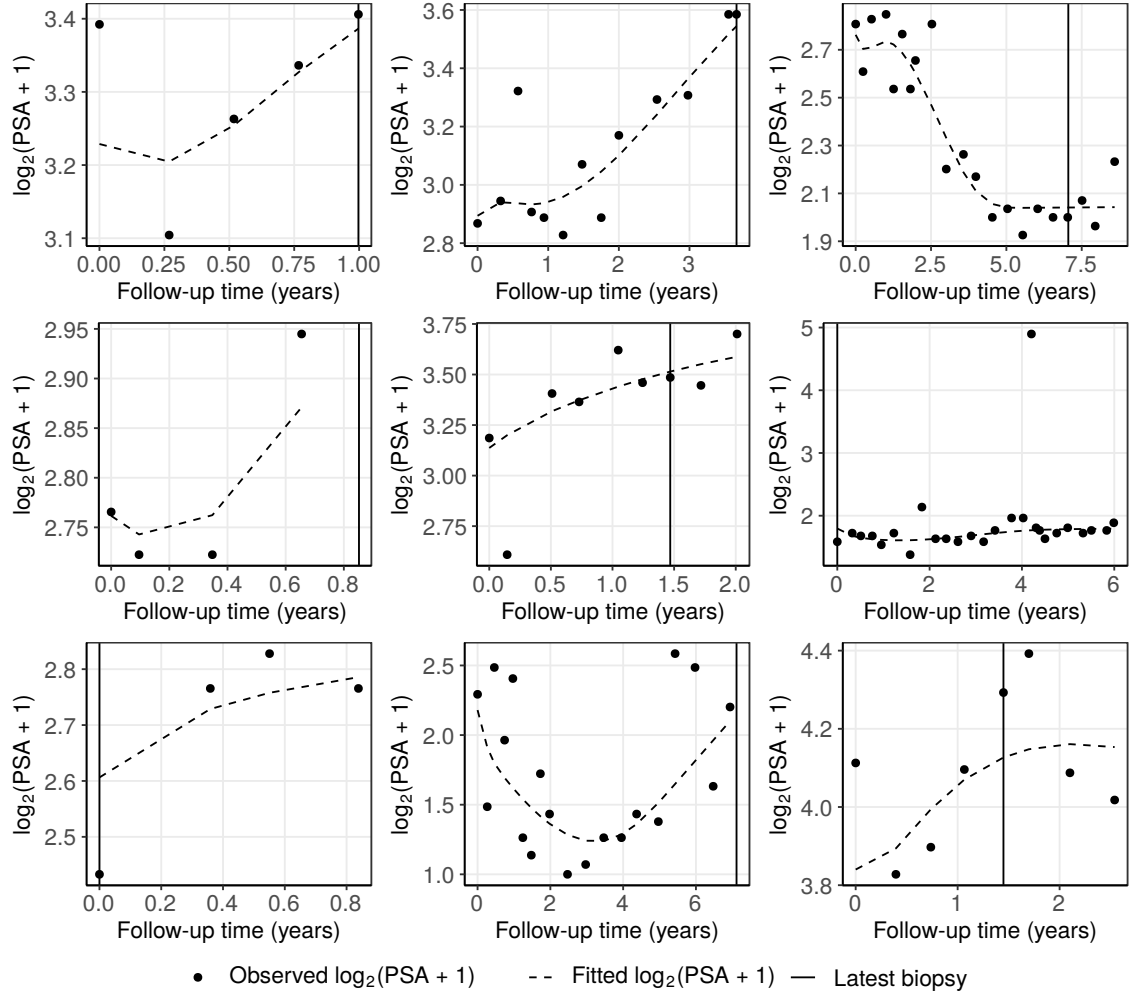
Figure 6: Fitted versus observed $\log_2(\text{PSA} + 1)$ profiles for nine randomly selected PRIAS patients. The fitted profiles utilize information from the observed PSA levels, DRE scores, and time of the latest biopsy.

For the relative risk sub-model (see Equation 3), the parameter estimates in Table 4 show that both $\log_2\{\text{PSA} + 1\}$ velocity, and the log odds of having DRE $>$ T1c were significantly associated with the hazard of cancer progression.

Table 4: Estimated mean and 95% credible interval for the parameters of the relative risk sub-model (see Equation 3) of the joint model fitted to the PRIAS dataset.

Variable	Mean	Std. Dev	2.5%	97.5%	P
$(\text{Age} - 70)$	0.012	0.006	2.3×10^{-4}	0.022	0.045
$(\text{Age} - 70)^2$	-0.001	0.001	-0.002	1.6×10^{-4}	0.095
$\text{logit}\{\text{Pr}(\text{DRE} > \text{T1c})\}$	0.147	0.017	0.115	0.183	<0.00
Fitted $\log_2(\text{PSA} + 1)$ value	0.104	0.078	-0.044	0.256	0.193
Fitted $\log_2(\text{PSA} + 1)$ velocity	3.396	0.564	2.376	4.475	<0.00

It is important to note that since age, $\log_2\{\text{PSA} + 1\}$ value and velocity, and log odds of DRE $>$ T1c are all measured on different scales, a comparison between the corresponding parameter estimates is not easy. To this end, in Table 5, we present the multiplicative change in hazard of cancer progression, for an increase in the aforementioned variable from their first to the third quartile. For example, an increase in log odds of DRE $>$ T1c, from -6.65 to -4.36 (fitted first and third quartiles) corresponds to a 1.40 fold increase in hazard of cancer progression. The interpretation for the rest is similar.

Table 5: Multiplicative increase in hazard of cancer progression, due to an increase in the model variables from their first quartile (Q_1) to the third quartile (Q_3). Except for age, the quartiles for all other variables are based on their fitted values.

Variable	Q_1	Q_3	Multiplicative increase in hazard
Age	65	75	1.13
$\log_2(\text{PSA} + 1)$ value	2.65	2.78	1.01
$\log_2(\text{PSA} + 1)$ velocity	-0.03	0.16	1.92
$\text{logit}\{\text{Pr}(\text{DRE} > \text{T1c})\}$	-6.65	-4.36	1.40

Appendix B.1   Assumption of t-distributed (df=3) Error Terms

With regards to the choice of the distribution for the error term ε_p for the PSA measurements (see Equation 2), we attempted fitting multiple joint models differing in error distribution, namely t-distribution with three, and four degrees of freedom, and a normal distribution for the error term. However, the model assumption for the error term were best met by the model with t-distribution having three degrees of freedom. The quantile-quantile plot of subject-specific residuals for the corresponding model in Panel A of Figure 7, shows that the assumption of t-distributed (df=3) errors is reasonably met by the fitted model.
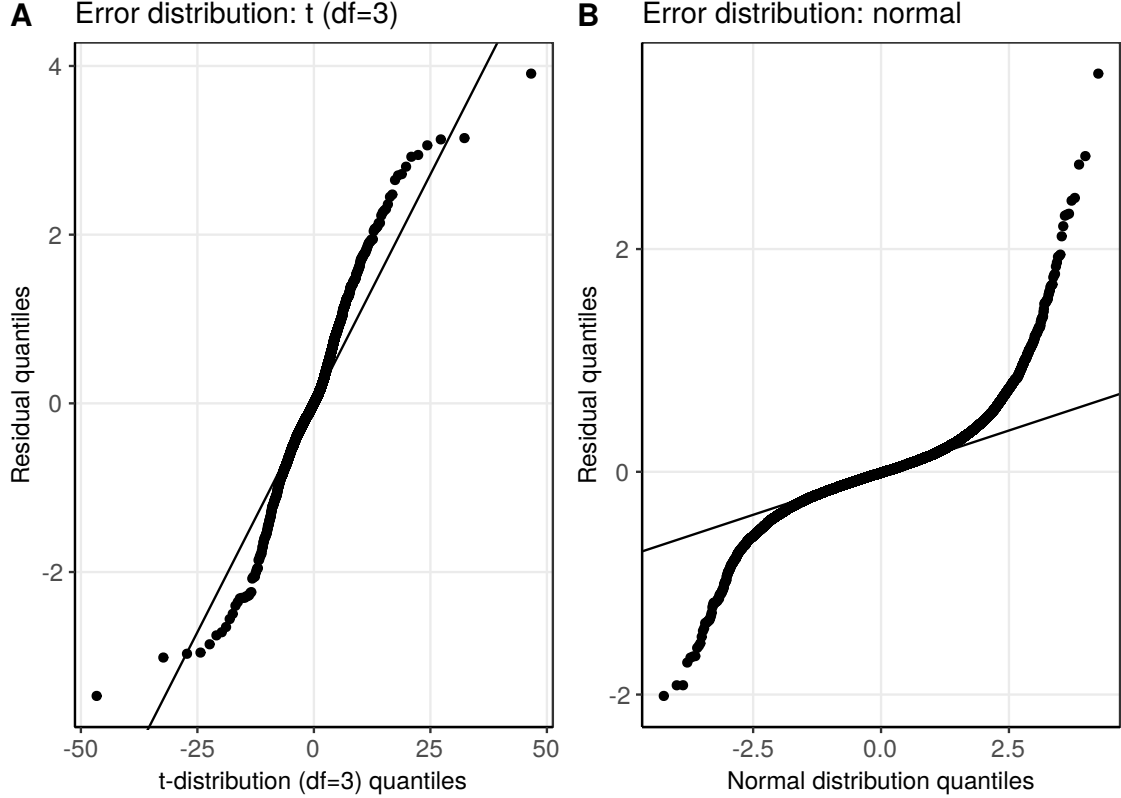

Figure 7: Quantile-quantile plot of subject-specific residuals from the joint models fitted to the PRIAS dataset. **Panel A**: model assuming a t-distribution (df=3) for the error term ε_p. **Panel B**: model assuming a normal distribution for the error term ε_p.

Appendix B.2 Predictive Performance of the Joint Model Fitted to the PRIAS dataset

We evaluate the predictive performance of our model using two measures: the area under the receiver operating characteristic curve, and the prediction error (Rizopoulos, Molenberghs, and Lesaffre, 2017). Given the longitudinal nature of the data at hand, in a joint model time dependent AUC and prediction errors are more relevant. More specifically, given the time of latest biopsy t and PSA, DRE measurements up to time s, we are interested in a medically relevant time frame $(t, s]$, within which the occurrence of cancer progression is of interest. In the case of prostate cancer, at any point in time it is of interest to identify patients who may have obtained cancer progression in the last one year ($s - t = 1$). Using data of the patients from the PRIAS study, we calculate the AUC and prediction error at the following s: year one, year two, year three, year four, and year five (95-percentile of observed cancer progression times) of follow-up in AS. The resulting AUC, and prediction are presented in Table 6.

Table 6: Area under the receiver operating characteristic curves (AUC), and prediction error, with 95% confidence interval in brackets.

Follow-up year	AUC	Prediction Error
1	0.651 [0.633, 0.663]	0.055 [0.052, 0.059]
2	0.621 [0.610, 0.640]	0.144 [0.140, 0.148]
3	0.748 [0.728, 0.770]	0.076 [0.075, 0.078]
4	0.710 [0.691, 0.736]	0.076 [0.072, 0.079]
5	0.592 [0.577, 0.614]	0.107 [0.103, 0.112]

Appendix C   Full Results of the Simulation Study

In the simulation study, we evaluate the following in-practice fixed/heuristic approaches (Inoue et al., 2018; Loeb et al., 2014) for biopsies: biopsy every year, biopsy every one and a half years, biopsy every two years and biopsy every three years. For the personalized biopsy approach we evaluate three fixed risk thresholds: 5%, 10% and 15%, and a risk threshold chosen using F_1 score. Lastly, we also evaluate the PRIAS schedule of biopsies. We compare all the aforementioned schedules on two criteria, namely the number of biopsies they schedule and the corresponding delay in detection of cancer progression, in years (time of positive biopsy - true time of cancer progression). The corresponding results, using 500×250 test patients are presented in Table 7.

Table 7: **Simulation study results for all patients**: Estimated first, second (median), and third quartiles for number of biopsies (Q_1^{nb}, Q_2^{nb}, Q_3^{nb}) and for the delay in detection of cancer progression (Q_1^{delay}, Q_2^{delay}, Q_3^{delay}), in years, for various biopsy schedules. The delay is equal to the difference between the time of the positive biopsy and the unobserved true time of progression. The results in the table are obtained from test patients of our simulation study.

In-practice schedules	Q_1^{nb}	Q_2^{nb}	Q_3^{nb}	Q_1^{delay}	Q_2^{delay}	Q_3^{delay}
Every year (Annual)	3	10	10	0.3	0.5	0.8
Every 1.5 years	2	7	7	0.4	0.7	1.1
Every 2 years	2	5	5	0.6	1.1	1.5
Every 3 years	1	4	4	1.1	1.8	2.3
PRIAS	2	4	6	0.3	0.7	1.0
Personalized approach						
Risk threshold: 5%	2	6	8	0.3	0.6	0.9
Risk threshold: 10%	2	4	5	0.3	0.7	1.0
Risk threshold: 15%	2	3	4	0.4	0.8	1.4
Risk using F_1 score	1	2	3	0.5	0.9	2.2

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). We assume *fast progressing* patients, are the patients with an initially misdiagnosed state of cancer (Cooperberg et al., 2011), 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. Table 8, Table 9, and Table 10 show the results for the *fast, intermediate,* and *slow progressing* patients, respectively.

Table 8: **Simulation study results for *fast progressing* patients (30% of all patients)**: Estimated first, second (median), and third quartiles for number of biopsies (Q_1^{nb}, Q_2^{nb}, Q_3^{nb}) and for the delay in detection of cancer progression (Q_1^{delay}, Q_2^{delay}, Q_3^{delay}), in years, for various biopsy schedules. The delay is equal to the difference between the time of the positive biopsy and the unobserved true time of progression. The results in the table are obtained from the *fast progressing* test patients of our simulation study.

In-practice schedules	Q_1^{nb}	Q_2^{nb}	Q_3^{nb}	Q_1^{delay}	Q_2^{delay}	Q_3^{delay}
Every year (Annual)	1	2	2	0.3	0.6	0.9
Every 1.5 years	1	1	2	0.4	0.8	1.2
Every 2 years	1	1	1	0.7	1.1	1.5
Every 3 years	1	1	1	1.5	2.0	2.5
PRIAS	1	2	2	0.3	0.7	1.0
Personalized approach						
Risk threshold: 5%	1	2	2	0.3	0.6	0.9
Risk threshold: 10%	1	1	2	0.3	0.7	1.0
Risk threshold: 15%	1	1	2	0.4	0.8	1.2
Risk using F_1 score	1	1	2	0.5	0.9	2.0

Table 9: **Simulation study results for *intermediate progressing* patients (20% of all patients)**: Estimated first, second (median), and third quartiles for number of biopsies (Q_1^{nb}, Q_2^{nb}, Q_3^{nb}) and for the delay in detection of cancer progression (Q_1^{delay}, Q_2^{delay}, Q_3^{delay}), in years, for various biopsy schedules. The delay is equal to the difference between the time of the positive biopsy and the unobserved true time of progression. The results in the table are obtained from the *intermediate progressing* test patients of our simulation study.

In-practice schedules	Q_1^{nb}	Q_2^{nb}	Q_3^{nb}	Q_1^{delay}	Q_2^{delay}	Q_3^{delay}
Every year (Annual)	5	7	8	0.2	0.5	0.7
Every 1.5 years	4	5	6	0.3	0.7	1.0
Every 2 years	3	4	4	0.4	1.0	1.5
Every 3 years	2	3	3	0.6	1.3	2.0
PRIAS	3	5	6	0.3	0.7	1.3
Personalized approach						
Risk threshold: 5%	5	6	7	0.3	0.6	0.9
Risk threshold: 10%	3	4	6	0.4	0.7	1.3
Risk threshold: 15%	3	3	5	0.4	0.8	1.7
Risk using F_1 score	2	3	5	0.5	1.0	2.4

Table 10: **Simulation study results for *slow progressing* patients (50% of all patients)**: Estimated first, second (median), and third quartiles for number of biopsies (Q_1^{nb}, Q_2^{nb}, Q_3^{nb}) and for the delay in detection of cancer progression (Q_1^{delay}, Q_2^{delay}, Q_3^{delay}), in years, for various biopsy schedules. The delay is equal to the difference between the time of the positive biopsy and the unobserved true time of progression. The results in the table are obtained from the *slow progressing* test patients of our simulation study. Since no cancer progression is observed in the ten year follow-up period for these patients, delay cannot be estimated, and hence is not reported.

In-practice schedules	Q_1^{nb}	Q_2^{nb}	Q_3^{nb}	Q_1^{delay}	Q_2^{delay}	Q_3^{delay}
Every year (Annual)	10	10	10			
Every 1.5 years	7	7	7			
Every 2 years	5	5	5			
Every 3 years	4	4	4			
PRIAS	4	6	8			
Personalized approach						
Risk threshold: 5%	6	7	9			
Risk threshold: 10%	4	4	6			
Risk threshold: 15%	2	3	4			
Risk using F_1 score	2	2	4			

Appendix D   Source Code

The R code for fitting the joint model to the PRIAS dataset, and for the simulation study, along with sample dataset, and instructions for running the code are available with this paper at the following link:

https://github.com/anirudhtomer/prias/tree/master/src/decision_analytic

References

Bul, Meelan et al. (2013). "Active surveillance for low-risk prostate cancer worldwide: the PRIAS study". In: *European urology* 63.4, pp. 597–603.

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

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

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

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, p. 202.

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 (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.

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)*, pp. 290–295.