Personalized Decision Making for Prostate Cancer Biopsies in Active Surveillance Programs

Medical Decision Making
XX(X):3–17

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

Au Thor[1] and Aut Hor[2]

Prepared using sagej.cls

Abstract

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

Objective. Motivated by the world's largest AS study, Prostate Cancer Research International Active Surveillance (PRIAS), our aim is to better balance the number of biopsies and the delay in detection of cancer progression. 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 historical 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 posterior predictive distribution of the time of cancer progression. Using this distribution we personalize the decision of conducting biopsy at a visit. We compare the personalized approach with the fixed biopsy schedules via an extensive and realistic simulation study based on an exact replica of the population 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 a reduction in the number of biopsies, the delay in the detection of cancer progression for the personalized approach remains comparable with that of the biopsy schedule of the PRIAS study.

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

Keywords

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

* *

Introduction

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

Although larger values for PSA and/or larger score for DRE, may indicate cancer progression, the biopsies are the most reliable prostate cancer progression examination technique used in AS. When a patient's biopsy Gleason grading becomes larger than 6, AS is stopped and patient is advised for treatment of cancer progression[3]. Because biopsies are invasive they are conducted intermittently, until cancer progression is detected. Consequently, progression is always detected with a delay, equal to the difference between the time of the positive biopsy and the unobserved true time of progression. Biopsies are also painful, and prone to medical complications[4]. Hence, the decision of conducting a biopsy requires a fine compromise between the burden incurred due to biopsies and the delay in detection of progression. However, currently there is no consensus on the best time interval for subsequent repeat biopsies[5]. Many AS programs focus on minimizing only the delay, by scheduling biopsies

[1]Department of Biostatistics, Erasmus University Medical Center, the Netherlands
[2]Department of XXXX, XXX XXXX XXXX XXXX, the Netherlands

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

annually[6]. Annual biopsies may work well for patients who progress fast, but for slowly progressing patients many unnecessary burdensome biopsies are scheduled. To improve the share of burden between fast and slow progressing patients, the world's largest AS program Prostate Cancer Research International Active Surveillance (PRIAS)[7], schedules annual biopsy only if at a follow-up visit a patient has a PSA doubling time between 0 and 10 years. PSA doubling time is measured as the inverse of the slope of the regression line through the base two logarithm of the observed PSA levels. For everyone else, PRIAS schedules biopsies at following fixed follow-up times: year 1, 4, 7, and 10, and every 5 years thereafter. Despite this effort, in PRIAS over a period of 10 years a patient may get scheduled for 4 to 10 biopsies. Consequently, patients may not always comply with the schedule[3]. This can lead to the original problem of delayed detection of prostate cancer progression, and reduce the effectiveness of AS.

This article is motivated by the need to better balance the number of biopsies and the delay in detection of prostate cancer progression, than in practice currently. We intend to achieve this by personalizing the decision of conducting biopsies at follow-up visits. Personalized decision making has received much interest in the literature, especially for various cancers. For example, Markov decision process (MDP) models have been used to create personalized screening schedules for breast cancer[8], cervical cancer[9], and colorectal cancer[10]. In the specific case of prostate cancer, Zhang et al.[11] have used partially observable MDP models to personalize the decision of (not) deferring a biopsy to the next checkup time during the screening process. This decision is based on the baseline characteristics as well as a discretized PSA level of the patient at the current screening-visit time.

In order to personalize the decision of biopsies in an AS setting, we utilize the data of the patients of the PRIAS study (see Figure 1 for illustration). In comparison to the work referenced above, we do not base the decision of biopsy only on the discretized current PSA level of a patient, but instead we use the entire history of PSA levels, DRE scores, and results of the latest biopsy. To this end, we employ joint models for time-to-event and longitudinal data[12,13]. Since joint models use random effects[14] to model the between-patient heterogeneity, they are inherently patient-specific. Using joint models we first obtain a full specification

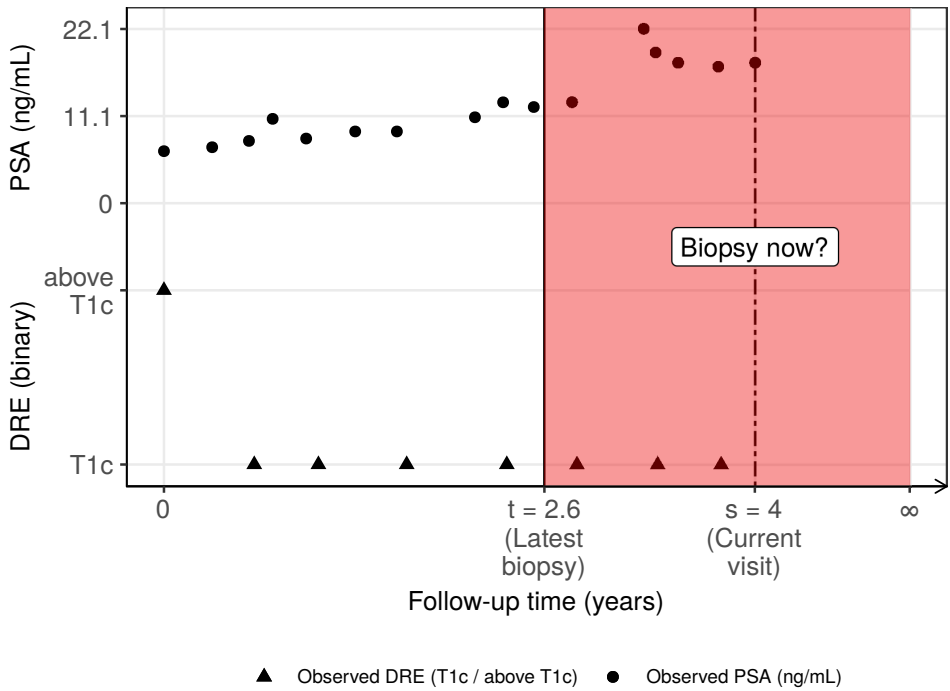

Figure 1. Available data from a patient who had his latest biopsy at $t = 2.6$ years, at which cancer progression was not detected. The latest follow-up visit of the patient is at $s = 4$ years. The shaded region shows the time period in which the patient is at the risk of cancer progression. We utilize the entire history of PSA and DRE measurements, along with the time of latest biopsy up to a follow-up visit to make the decision of biopsy at that visit.

of the joint distribution of the time of cancer progression, and PSA and DRE measurements. We then use it separately for each patient at each follow-up visit to define a visit and patient-specific posterior predictive distribution of the time of cancer progression, given the observed PSA and DRE measurements, and time of the latest biopsy up to that visit. We calculate the patient's risk of cancer progression at the follow-up visit using this distribution. If the risk is higher than a certain threshold our method schedules a biopsy at the same follow-up visit. Since there is no clear consensus on such risk thresholds we not only use fixed thresholds suggested by urologists, but also present a methodology to automate the choice of thresholds.

In order to evaluate our methodology we conduct an extensive simulation study. In the simulation study we compare the personalized

approach with the annual and PRIAS schedules. For a realistic comparison, we utilize an exact replica of the population of the PRIAS patients, generated using the 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 detail 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 are observed to have 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 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 the dataset has 5 DRE and 9 PSA measurements available per patient. In order to identify cancer progression, biopsies are scheduled as per the PRIAS protocol (see Introduction).

Joint Model for Time-to-Event and Longitudinal Data

Let T_i^* denote the true cancer progression time for the i-th patient in PRIAS. Let T_i^R and T_i^{R-1} denote the time of his latest and second latest biopsies, respectively. Since biopsies are conducted periodically, T_i^* cannot be observed directly and it is only known to fall in an interval $l_i < T_i^* \leq r_i$, where $l_i = T_i^{R-1}, r_i = T_i^R$ if progression is observed at the latest biopsy, and $l_i = T_i^R, r_i = \infty$ if progression is not observed yet. Further let $\boldsymbol{y}_{1i}$, and $\boldsymbol{y}_{2i}$ denote the $n_{1i} \times 1$, and $n_{2i} \times 1$ vectors of the DRE, and PSA longitudinal measurements, respectively. For a sample of n patients the observed data is denoted by $\mathcal{D}_n = \{l_i, r_i, \boldsymbol{y}_{1i}, \boldsymbol{y}_{2i}; i = 1, \ldots, n\}$.

The patient-specific PSA and DRE measurements over time are modeled using a generalized linear mixed effects model. For the i-th patient, the

mixed effects sub-model for DRE is given by:

$$\begin{aligned}\text{logit}\big[\Pr\{y_{1i}(t) > \text{T1c}\}\big] &= \beta_{01} + b_{01i} + (\beta_{11} + b_{11i})t \\ &\quad + \beta_{21}(\text{Age}_i - 70) + \beta_{31}(\text{Age}_i - 70)^2 \end{aligned} \tag{1}$$

where, t denotes a specific time point in the AS follow-up, Age_i is the age of the i-th patient at the time of inclusion in AS. The fixed effect parameters are denoted by $\{\beta_{01}, \ldots, \beta_{31}\}$, and b_{01i}, b_{02i} are the patient specific random effects. With this definition, we assume that the log odds of obtaining a DRE score larger than T1c remain linear over time. An example model fit for DRE is shown in panel A of Figure 2. For the i-th patient, the mixed effects sub-model for PSA is given by:

$$\begin{aligned}\log_2\big\{y_{2i}(t) + 1\big\} &= m_{2i}(t) + \varepsilon_{2i}(t), \\ m_{2i}(t) &= \beta_{02} + b_{02i} + \sum_{k=1}^{4}(\beta_{k2} + b_{k2i})B_k(t, \mathcal{K}) \\ &\quad + \beta_{52}(\text{Age}_i - 70) + \beta_{62}(\text{Age}_i - 70)^2, \end{aligned} \tag{2}$$

where, $m_{2i}(t)$ denotes the underlying measurement error free value of $\log_2(\text{PSA} + 1)$ measurements at time t. To accommodate for a non-linear evolution of this value over the follow-up period in AS, we utilize B-splines[16]. In Equation (2), $B_k(t, \mathcal{K})$ denotes the k-th basis function of a B-spline with three internal knots at $\mathcal{K} = \{0.1, 0.7, 4\}$ years, and boundary knots at 0 and 5.42 years (0.95 quantile of the observed follow-up times). The fixed effect parameters are denoted by $\{\beta_{02}, \ldots, \beta_{62}\}$ and the patient specific random effects are denoted by $\{b_{02i}, \ldots, b_{42i}\}$. The error $\varepsilon_{2i}(t)$ 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. An example model fit for PSA is shown in panel B of Figure 2. To account for the association between the DRE and PSA measurements, we link their corresponding random effects. More specifically, the complete vector of random effects $\boldsymbol{b}_i = (b_{01i}, b_{02i}, b_{02i}, \ldots, b_{42i})^T$ is assumed to follow a multivariate normal distribution with mean zero and $q \times q$ variance-covariance matrix $\boldsymbol{D}$.

To model the impact of DRE and PSA measurements on the risk of cancer progression, we use a relative risk sub-model. More specifically,

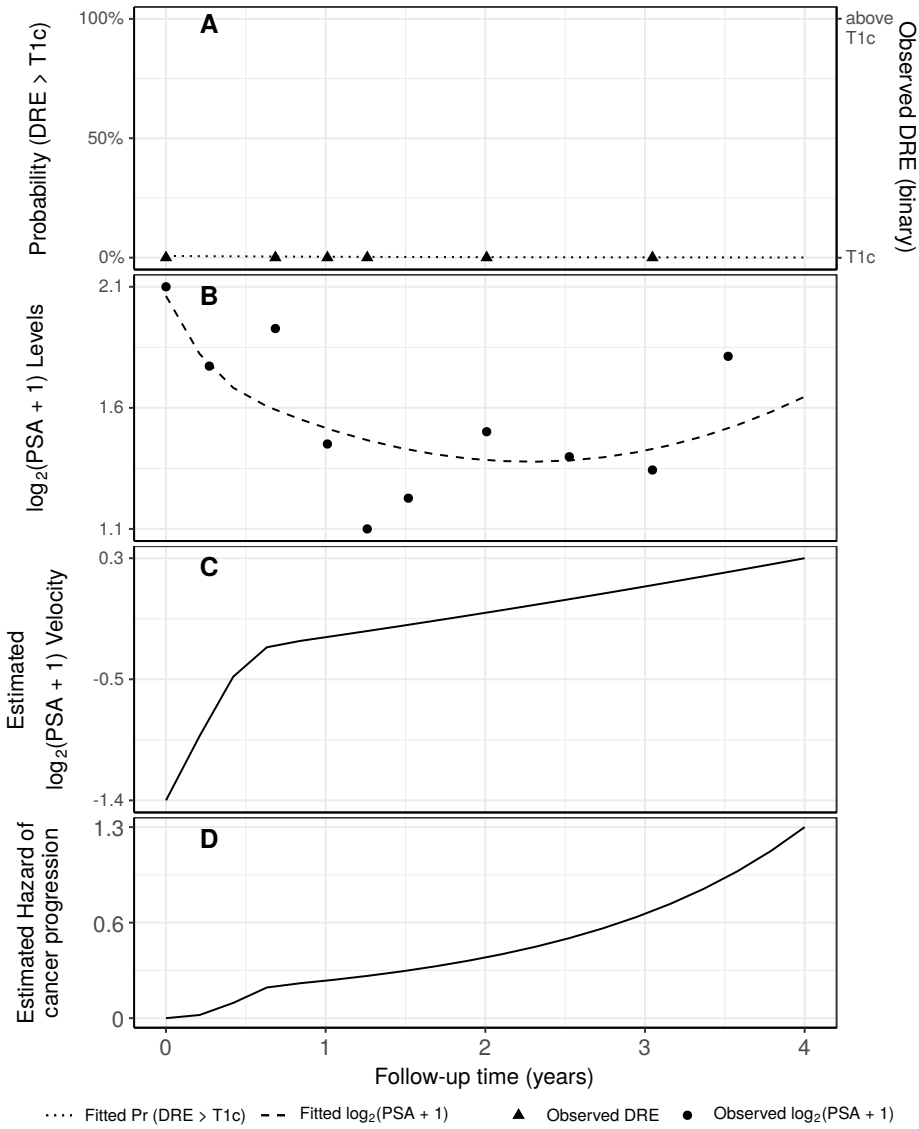

Figure 2. Illustration of the joint model fitted to the PRIAS dataset. **Panel A:** shows the observed DRE scores and the fitted probability of obtaining a DRE score greater than T1c (Equation 1). **Panel B:** shows the observed and fitted $\log_2(\text{PSA}+1)$ levels (Equation 2). **Panel C:** shows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

the hazard of cancer progression $h_i(t)$ at a time t is given by:

$$h_i(t) = h_0(t)\exp\Big(\gamma_1(\text{Age}_i - 70) + \gamma_2(\text{Age}_i - 70)^2 + \alpha_{11}\text{logit}\big[\Pr\{y_{1i}(t) > \text{T1c}\}\big] + \alpha_{21}m_{2i}(t) + \alpha_{22}\frac{\partial m_{2i}(t)}{\partial t}\Big), \quad (3)$$

where, γ_1, γ_2 are the coefficients for the effect of age. The parameter α_{11} 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, namely at any time t the effect of the instantaneous underlying value (dashed line in panel B of Figure 2) of PSA $m_{2i}(t)$ is given by α_{21}, and the effect of the instantaneous underlying velocity (panel C in Figure 2) of PSA $\partial m_{2i}(t)/\partial t$ is given by α_{22}. Lastly, $h_0(t)$ is the baseline hazard at time t, and is modeled flexibly using P-splines[17]. An example fitted hazard is shown in Panel D of Figure 2. The detailed specification of the baseline hazard $h_0(t)$, and parameter estimation using the Bayesian approach are presented in Appendix A of the supplementary material.

Personalized Decisions for Biopsy During Follow-up Visit

Let us assume that a decision of conducting a biopsy is to be made for a new patient j, who is not present in the PRIAS dataset. Let t be the time of his latest biopsy, and s denotes the current follow-up visit time. Let $\mathcal{Y}_{1j}(s)$ and $\mathcal{Y}_{2j}(s)$ denote the vector of all DRE and PSA measurements taken up to the current visit, respectively. We want to combine the observed DRE and PSA measurements of patient j, to inform us when the cancer progression is to be expected (see Figure 3). This will guide the decision making on whether to conduct a biopsy at the current follow-up visit. To this end, the combined observed information is given by the posterior predictive distribution $g(T_j^*)$ of the time of cancer progression T_j^*. It is given by:

$$\begin{aligned} g(T_j^*) &= p\big\{T_j^* \mid T_j^* > t, \mathcal{Y}_{1j}(s), \mathcal{Y}_{2j}(s), \mathcal{D}_n\big\} \\ &= \int\int p\big(T_j^* \mid T_j^* > t, \boldsymbol{b}_j, \boldsymbol{\theta}\big) \\ &\quad \times p\big\{\boldsymbol{b}_j \mid T_j^* > t, \mathcal{Y}_{1j}(s), \mathcal{Y}_{2j}(s), \boldsymbol{\theta}\big\} p\big(\boldsymbol{\theta} \mid \mathcal{D}_n\big)\mathrm{d}\boldsymbol{b}_j\mathrm{d}\boldsymbol{\theta}. \end{aligned}$$

The distribution $g(T_j^*)$ updates as extra information is recorded at follow-up visits. It is also unique for each patient as it depends on the historical data of the patient via the posterior distribution of the random effects $\boldsymbol{b}_j$.

A key ingredient in the decision of conducting a biopsy at the current follow-up visit time s, is the risk that the cancer has already progressed since the time of the last biopsy t (see Figure 3 for illustration). This risk can be derived from the posterior predictive distribution $g(T_j^*)$[18], and is given by:

$$R_j(s \mid t) = \Pr\{T_j^* \leq s \mid T_j^* > t, \mathcal{Y}_{1j}(s), \mathcal{Y}_{2j}(s), \mathcal{D}_n\}, \quad s \geq t.$$

A simple and straightforward approach to decide upon conducting a biopsy at a follow-up visit would be to do so when the risk of cancer progression at that visit is higher than a certain threshold $0 \leq \kappa \leq 1$. For example, as shown in panel B of Figure 3, biopsy at a follow-up visit may be scheduled if the risk is higher than 15% (example risk threshold). However, deciding upon a biopsy using the same risk threshold at all follow-up visits does not account for the time-varying prevalence of cancer progression in AS patients over the follow-up period.

As an alternative to fixed risk thresholds, we propose follow-up time dependent risk thresholds. Using information from the observed cancer progression times in the PRIAS dataset, at a follow-up visit the threshold is chosen on the basis of its ability to discriminate between patients who obtain cancer progression versus others. More specifically, given the time t of the latest biopsy we propose to choose a threshold κ for which a binary classification accuracy measure[19], discriminating between patients obtaining progression versus others, is maximized. In joint models, a patient j is predicted to have progression in the time period $(t, s]$ between the current visit and the last biopsy, if $R_j(s \mid t) > \kappa$[20,21]. Otherwise the patient is predicted to not have progression. Since we are interested in detecting progression, we choose F_1 score as the binary classification accuracy measure. It combines both time dependent true positive rate (TPR) and positive predictive value (PPV), and is defined as:

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

Since a high F_1 score is desired, the corresponding value of κ is $\arg\max_\kappa F_1(t, s, \kappa)$. The time dependent TPR and PPV are estimated from the joint model fitted to the PRIAS dataset[21].

Simulation Study

Although the personalized decision making approach is motivated by the PRIAS study, it is not possible to evaluate it on the PRIAS dataset. This is due to the fact that the PRIAS patients 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 compare personalized, PRIAS and annual schedules. For a realistic comparison, we simulate data from the joint model fitted to the PRIAS dataset. The simulated population satisfies the follow-up period of 10 years, as well as the percentage of patients who progress slow and fast. 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 with 1000 patients each. We generate a true cancer progression time for each of the patients, and then sample a set of PSA and DRE measurements at the same time points as given in PRIAS protocol. We then split each dataset into a training (750 patients) and a test (250 patients) part, and generate a random and noninformative censoring time for the training patients. We next fit a joint model of the specification given in Equation (1), (2) and (3) to each of the 500 training datasets and obtain MCMC samples from the 500 sets of the posterior distribution of the parameters. Using these fitted joint models, we obtain the posterior predictive distribution of time of cancer progression for each of the 500×250 test patients at each of their visits. While maintaining a gap of 1 year between consecutive biopsies, individually for each patient at each follow-up visit we make the decision of (not) conducting a biopsy.

In this simulation study, one biopsy is conducted for all patients at the beginning of the AS program and another one is conducted at the end of the follow-up period at 10 years. The rest of the biopsies are scheduled using the following methods: (abbreviated names in parenthesis): biopsy

every year (Annual), biopsy as per PRIAS protocol (PRIAS), personalized biopsy using a risk threshold of 5% (Risk: 5%), personalized biopsy using a risk threshold of 15% (Risk: 15%), and personalized biopsy using a risk threshold chosen on the basis of progression history of patients from the training dataset (Risk: F_1). This results into an entire personalized schedule for each patient. We compare the resulting biopsy schedules on two measures, namely the number of biopsies they schedule and the delay in detection of cancer progression incurred due the schedule. We define the delay as the difference between the time of the biopsy on which cancer progression is detected and the true time of cancer progression. Ideal numbers for these two measures are 1 biopsy and 0 years of delay.

Results

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

Model Fit

From the joint model fitted to the PRIAS dataset, we found that both $\log_2\{\text{PSA}+1\}$ velocity, and log odds of having DRE $>$ T1c were significantly associated with the hazard of cancer progression. For any patient, an increase in $\log_2\{\text{PSA}+1\}$ velocity from 0.03 to 0.16 (first and third quartiles of the fitted velocities, respectively) corresponds to a 1.92 fold increase in the hazard of cancer progression. Whereas, an increase in log odds of DRE $>$ T1c from -6.65 to -4.36 (first and third quartiles of the fitted log odds, respectively) corresponds to a 1.40 fold increase in the hazard of cancer progression. In terms of the predictive performance, we found that the area under the receiver operating characteristic curves (AUC)[21] was 0.59, 0.66, and 0.66 at years 1, 2, and 3 of followup, respectively. The AUC estimates for a joint model ignoring DRE measurements were 0.58, 0.64 and 0.60 at years 1, 2 and 3 of follow-up. Parameter estimates are presented in detail in Appendix B of the supplementary material.

Simulation Study

In the 500×750 training patients, we observed that for roughly 50% of the patients cancer progression did not take place in the 10 year follow-up. These could be seen as patients with a slow speed of cancer

progression. Roughly 30% of the patients obtain cancer progression within first 3.5 years. These could be high risk patients who choose AS instead of immediate treatment, or patients with an initially misdiagnosed state of cancer[22]. In this work we consider these patients as the fast progressing patients. We consider that the remaining 20% patients with cancer progression times between 3.5 and 10 years have an intermediate speed of cancer progression.

For faster progressing patients (30% of the total patients), the boxplots in panel A of Figure 4 shows the variation in number of biopsies, and the delay in detection of cancer progression, in years (time of last biopsy - true time of cancer progression) due to various biopsy schedules. We can see that the personalized schedules conduct a median of one biopsy compared to two biopsies for PRIAS and annual schedule. The performance of personalized schedule with automatically chosen threshold is similar to that of PRIAS schedule. Thus with personalized approach, one biopsy may get saved for faster progressing patients.

For patients with intermediate progression speed (20% of the total patients), the boxplots in panel B of Figure 4 shows the variation in number of biopsies, and the delay in detection of cancer progression due to various biopsy schedules. Firstly, we can see that personalized schedules with a small risk threshold such as 5% risk conduct many more biopsies than other personalized schedules. Consequently, their performance with respect to the delay in detection of progression is similar to that of annual schedule. However, personalized schedule with slightly higher risk (15%) and risk chosen automatically, schedule a median of 3 and 4 biopsies, respectively. This is despite the fact that the delay in detection of cancer progression due to these schedules is similar to that of the PRIAS schedule. However, the PRIAS schedule conducts more biopsies (median of 5 biopsies). Thus, the personalized approach may lead to one to two less biopsies for patients with intermediate speed of progression.

The patients who are at most advantage with the personalized schedules are the patients who progress slowly (50% of the total patients). Panel C of Figure 4 shows a boxplot of the number of biopsies conducted by various biopsy schedules for such patients. It can be seen that the annual schedule may lead to 10 unnecessary biopsies for everyone. The PRIAS schedule, schedules a median of 6 unnecessary biopsies. In comparison

the personalized schedules using 15% risk threshold and automatically chosen risk threshold, schedule only 3 and 4 biopsies, respectively.

Discussion

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

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

In the simulation study we found that the patients who never obtain cancer progression in the 10 year follow-up incur much burden due to currently used schedules. The PRIAS schedule, despite its effort to identify such patients using PSA doubling time, schedules a minimum of 4 biopsies and a median of 6 unnecessary biopsies for such patients. The annual schedule performs even worse by scheduling 10 unnecessary biopsies. In contrast, the personalized schedules that we proposed reduce it to a median of 3 to 4 biopsies depending upon the choice of the

risk threshold. The choice of the risk threshold is also important. A low risk threshold such as 5% risk may seem attractive to detect cancer progressions in time. However, it can work as worse as annual schedule in most situations, by scheduling unnecessary biopsies. A better idea is to either use a higher risk threshold or to automatically select it (see Methods). The personalized approach based on automatically chosen thresholds has the advantage of being generic for use in other active surveillance programs. In addition, it inherently accounts for the time-varying prevalence of cancer progression in AS patients over the follow-up period.

For patients with faster and intermediate speed of cancer progression, we found in our simulation study that the personalized approach with automatically chosen threshold conducts less biopsies than the PRIAS schedule but still leads to a similar delay in detection of cancer progression. For slow, intermediate, and fast progressing patients combined, the aforementioned personalized approach schedules 3.75 biopsies before detecting a cancer progression. The personalized approach using 15% risk schedules only 3 biopsies in total. These numbers are similar to the number of biopsies patients agree to undergo in PRIAS, if the non-compliance rates[3] are also accounted in the PRIAS schedule. Hence the personalized approach for biopsies better balances the number of biopsies per detected progression compared to the existing PRIAS and annual schedules.

However, a limitation of our approach is that at any given follow-up visit it cannot guarantee the total number of future biopsies required to detect cancer progression. This is due to the fact that the cancer progression cannot be foreseen with 100% accuracy. Hence, attempts to create an entire optimal schedule may not always work. Instead, such a guarantee can be given with a certain probability (see Figure 4). For example, if a 90% surety is required on the delay in detection of cancer progression being less than or equal to two years, then the personalized schedule with automatically chosen risk threshold schedules the least number of biopsies on average. Urologists may also choose a personalized biopsy approach suitable to patients with a certain speed of progression, if it is known in advance. In order to aid urologists in biopsy decision making process, we have developed a web application (examples included), hosted at `www.<insert-name-here>.com`.

Acknowledgements

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

Supplemental material

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

References

1. Torre LA, Bray F, Siegel RL et al. Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians* 2015; 65(2): 87–108.
2. Potosky AL, Miller BA, Albertsen PC et al. The role of increasing detection in the rising incidence of prostate cancer. *JAMA* 1995; 273(7): 548–552.
3. Bokhorst LP, Alberts AR, Rannikko A et al. Compliance rates with the Prostate Cancer Research International Active Surveillance (PRIAS) protocol and disease reclassification in noncompliers. *European Urology* 2015; 68(5): 814–821.
4. Ehdaie B, Vertosick E, Spaliviero M et al. The impact of repeat biopsies on infectious complications in men with prostate cancer on active surveillance. *The Journal of urology* 2014; 191(3): 660–664.
5. Loeb S, Carter HB, Schwartz M et al. Heterogeneity in active surveillance protocols worldwide. *Reviews in urology* 2014; 16(4): 202.
6. Welty CJ, Cowan JE, Nguyen H et al. Extended followup and risk factors for disease reclassification in a large active surveillance cohort for localized prostate cancer. *The Journal of Urology* 2015; 193(3): 807–811.
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. De Boor C, De Boor C, Mathématicien EU et al. *A practical guide to splines*, volume 27. Springer-Verlag New York, 1978.
17. Eilers PH and Marx BD. Flexible smoothing with B-splines and penalties. *Statistical Science* 1996; 11(2): 89–121.
18. Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* 2011; 67(3): 819–829.
19. López-Ratón M, Rodríguez-Álvarez MX, Cadarso-Suárez C et al. OptimalCutpoints: an R package for selecting optimal cutpoints in diagnostic tests. *Journal of Statistical Software* 2014; 61(8): 1–36.
20. 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.
21. Rizopoulos D, Molenberghs G and Lesaffre EM. Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biometrical Journal* 2017; 59(6): 1261–1276.
22. Cooperberg MR, Cowan JE, Hilton JF et al. Outcomes of active surveillance for men with intermediate-risk prostate cancer. *Journal of Clinical Oncology* 2011; 29(2): 228.

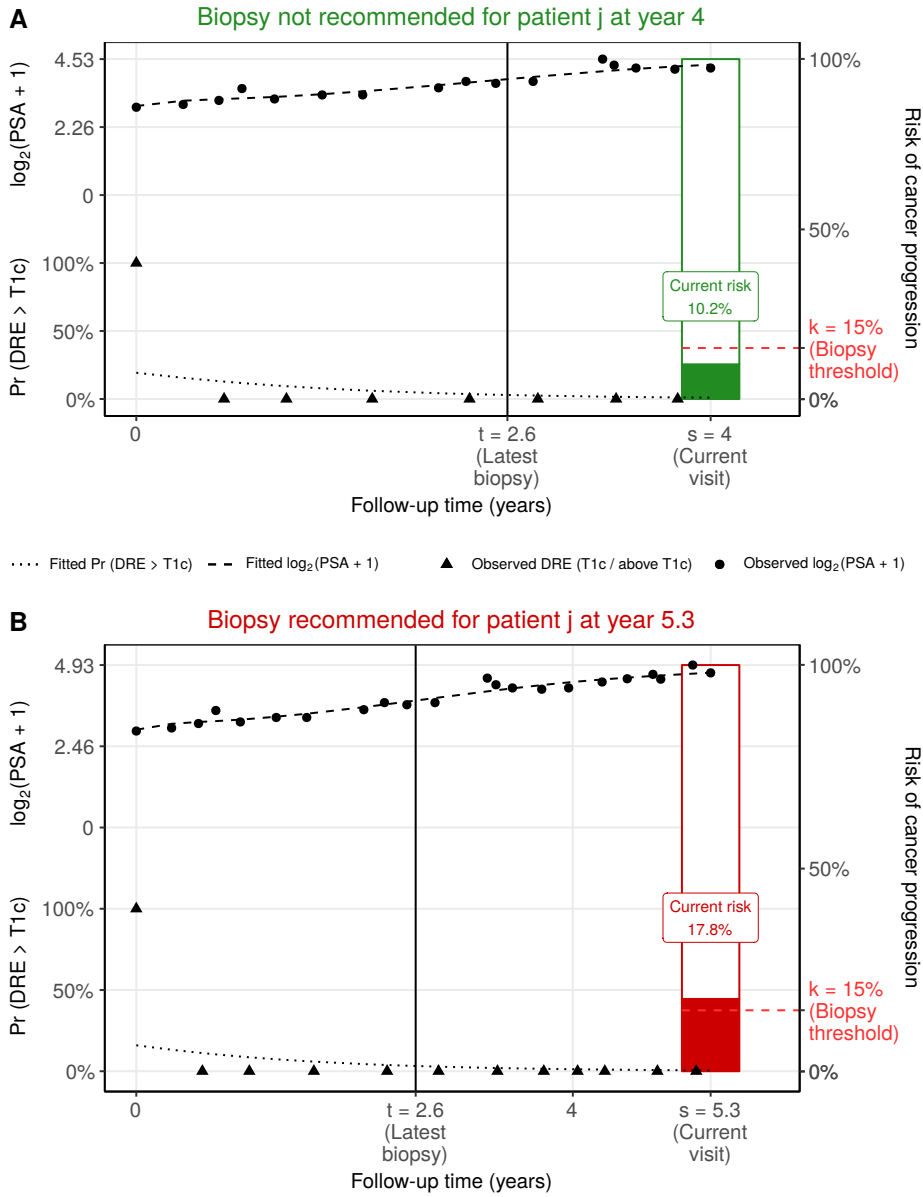

Figure 3. Illustration of personalized decision making for patient j at two different follow-up visits. Biopsy is recommended if the risk of cancer progression estimated from the joint model fitted to the PSA and DRE measurements of the patient, is higher than the example risk threshold for biopsy ($\kappa = 15\%$). **Panel A:** biopsy is not recommended for the patient j at the follow-up visit time $s = 4$ years, because his estimated risk of cancer progression (10.2%) is less than the biopsy risk threshold. **Panel B:** biopsy is recommended for the patient j at the follow-up visit time $s = 5.3$ years, because his estimated risk of cancer progression (17.8%) is more than the biopsy risk threshold.

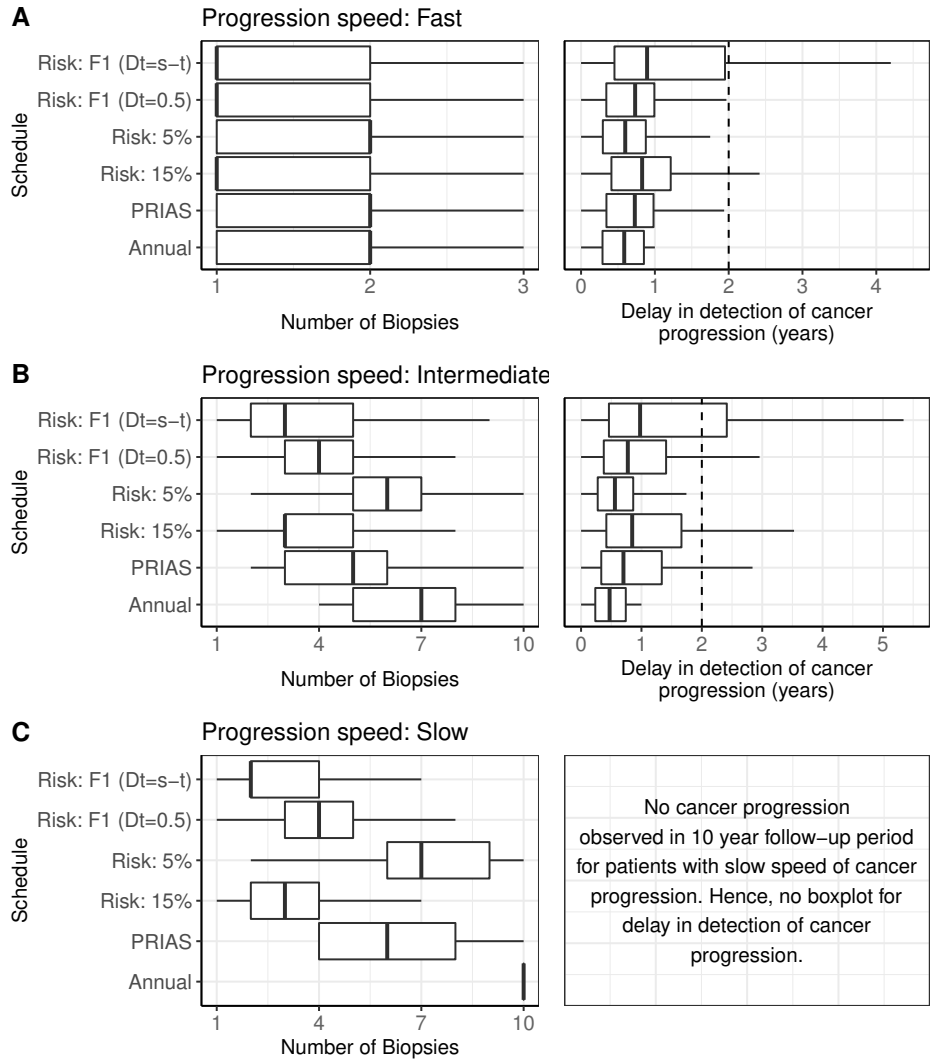

Figure 4. Boxplot showing variation in number of biopsies, and the delay in detection of cancer progression, in years (time of last biopsy - true time of cancer progression) for various biopsy schedules. Biopsies are conducted until cancer progression is detected. **Panel A:** results for simulated patients who had a faster speed of cancer progression, with progression times between 0 and 3.5 years. **Panel B:** results for simulated patients who had an intermediate speed of cancer progression, with progression times between 3.5 and 10 years. **Panel C:** results for simulated patients who did not have cancer progression in the 10 years of follow-up. **Types of personalized schedules:** Risk: 15% and Risk: 5% approaches, schedule biopsy if the risk of cancer progression at a visit is more than 15% and 5%, respectively. Risk: F1 (Dt=0.5) and Risk: F1 (Dt=s-t) work similar to Risk: 15% and Risk: 5%, except that the risk threshold for biopsy is chosen automatically by maximizing F_1 score (see Methods). The term Dt represents the time window (years) in which F_1 score is maximized, and s is the time of the current follow-up visit and t is the time of the latest biopsy. 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 Prostate Cancer Biopsies in Active Surveillance Programs"

Auth or[1,*], Aut hor[2], A uthor[3], Au thor[2,4], and Autho r[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 Medical Statistics and Bioinformatics, Leiden University Medical Center, the Netherlands
*email: a.tomer@erasmusmc.nl

Appendix A Joint Model for Time-to-Event and Longitudinal Data

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 $\leq$ T1c Schröder et al., 1992. On average the dataset has 5 DRE and 9 PSA measurements available 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 cancer progression. In 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.

Appendix A.2 Model Definition

Let T_i^* denote the true cancer progression time for the i-th patient in PRIAS. Let T_i^R and T_i^{R-1} denote the time of his latest and second latest biopsies, respectively. Since biopsies are conducted periodically, T_i^* cannot be observed directly and it is only known to fall in an interval $l_i < T_i^* \leq r_i$, where $l_i = T_i^{R-1}, r_i = T_i^R$ if progression is observed at the latest biopsy, and $l_i = T_i^R, r_i = \infty$ if progression is not observed yet. Further let $\boldsymbol{y}_{1i}$, and $\boldsymbol{y}_{2i}$ denote the $n_{1i} \times 1$, and $n_{2i} \times 1$ vectors of the DRE, and PSA longitudinal measurements, respectively. For a sample of n patients the observed data is denoted by $\mathcal{D}_n = \{l_i, r_i, \boldsymbol{y}_{1i}, \boldsymbol{y}_{2i}; i = 1, \ldots, n\}$.

The patient-specific PSA and DRE measurements over time are modeled using a generalized linear mixed effects model. For the i-th patient, the mixed effects sub-model for DRE is given by:

$$\begin{aligned} \text{logit}\big[\Pr\{y_{1i}(t) > \text{T1c}\}\big] &= \beta_{01} + b_{01i} + (\beta_{11} + b_{11i})t \\ &\quad + \beta_{21}(\text{Age}_i - 70) + \beta_{31}(\text{Age}_i - 70)^2 \end{aligned} \tag{1}$$

where, t denotes a specific time point in the AS follow-up, Age_i is the age of the i-th patient at the time of inclusion in AS. The fixed effect parameters are denoted by $\{\beta_{01}, \ldots, \beta_{31}\}$, and b_{01i}, b_{02i} are the patient specific random effects. With this definition, we assume that the log odds of obtaining a DRE score larger than T1c remain linear over time. An example model fit for DRE is shown in panel A of Figure 1. For the i-th patient, the mixed effects sub-model for PSA is given by:

$$\begin{aligned} \log_2\big\{y_{2i}(t) + 1\big\} &= m_{2i}(t) + \varepsilon_{2i}(t), \\ m_{2i}(t) &= \beta_{02} + b_{02i} + \sum_{k=1}^{4}(\beta_{k2} + b_{k2i})B_k(t, \mathcal{K}) \\ &\quad + \beta_{52}(\text{Age}_i - 70) + \beta_{62}(\text{Age}_i - 70)^2, \end{aligned} \tag{2}$$

where, $m_{2i}(t)$ denotes the underlying measurement error free value of $\log_2(\text{PSA} + 1)$ measurements at time t. To accommodate for a non-linear evolution of this value over the follow-up period in AS, we utilize B-splines (De Boor et al., 1978). In Equation (2), $B_k(t, \mathcal{K})$ denotes the k-th basis function of a B-spline with three internal knots at $\mathcal{K} = \{0.1, 0.7, 4\}$ years, and boundary knots at 0 and 5.42 years (0.95 quantile of the observed follow-up times). The fixed effect parameters are denoted by $\{\beta_{02}, \ldots, \beta_{62}\}$ and the patient specific random effects are denoted by $\{b_{02i}, \ldots, b_{42i}\}$. The error $\varepsilon_{2i}(t)$ 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. An example model fit for PSA is shown in panel B of Figure 1. To account for the association between the DRE and PSA measurements, we link their corresponding random effects. More specifically, the complete vector of random effects $\boldsymbol{b}_i = (b_{01i}, b_{02i}, b_{02i}, \ldots, b_{42i})^T$ is assumed to follow a multivariate normal distribution with mean zero and $q \times q$ variance-covariance matrix $\boldsymbol{D}$.

To model the impact of DRE and PSA measurements on the risk of cancer progression, we use a relative risk sub-model. More specifically, the hazard of cancer progression $h_i(t)$ at a time t is given by:

$$\begin{aligned} h_i(t) = h_0(t) \exp\Big(&\gamma_1(\text{Age}_i - 70) + \gamma_2(\text{Age}_i - 70)^2 \\ &+ \alpha_{11}\text{logit}\big[\Pr\{y_{1i}(t) > \text{T1c}\}\big] + \alpha_{21}m_{2i}(t) + \alpha_{22}\frac{\partial m_{2i}(t)}{\partial t}\Big), \end{aligned} \tag{3}$$

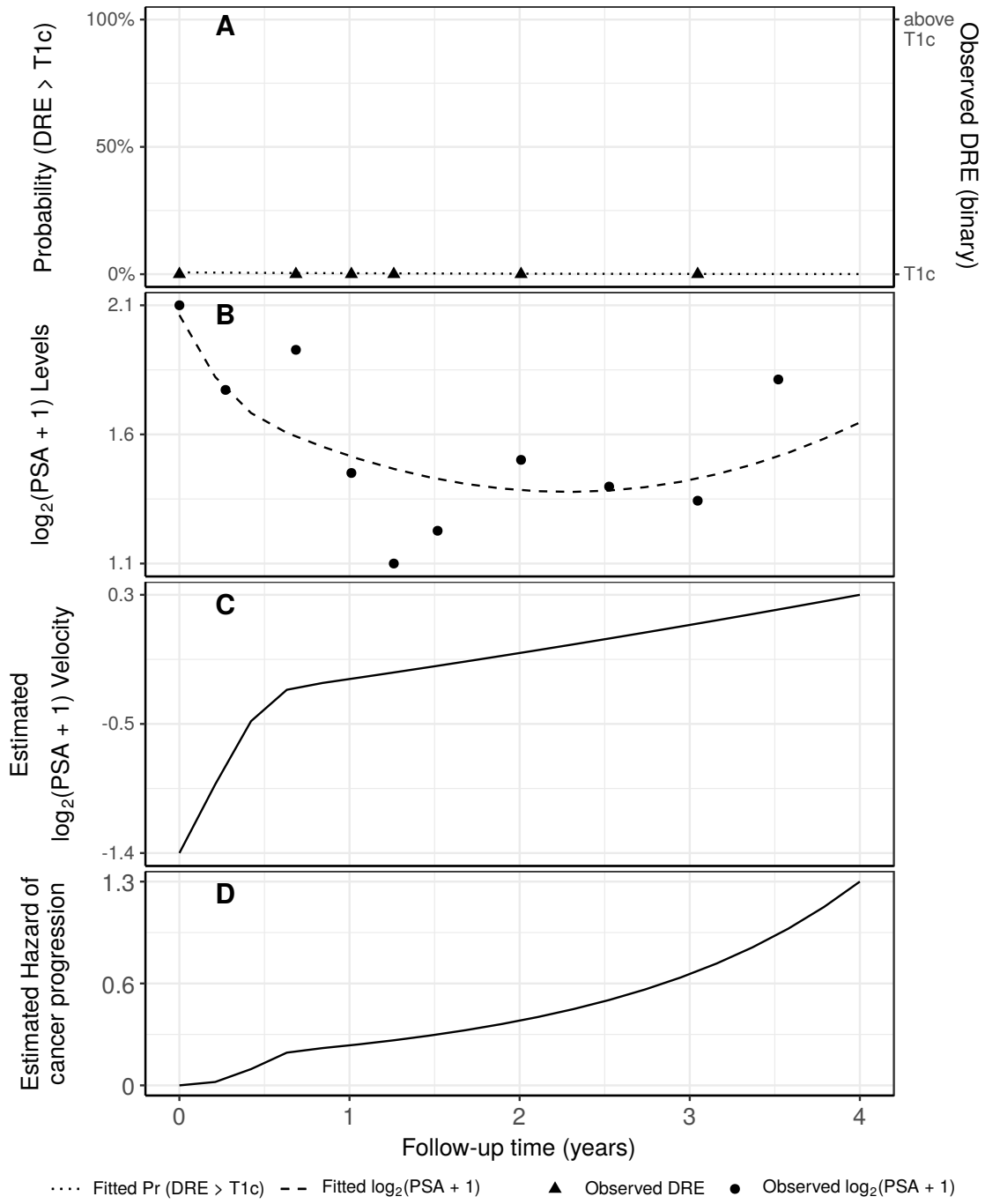

Figure 1: Illustration of the joint model fitted to the PRIAS dataset. **Panel A:** shows the observed DRE scores and the fitted probability of obtaining a DRE score greater than T1c (Equation 1). **Panel B:** shows the observed and fitted $\log_2(\text{PSA}+1)$ levels (Equation 2). **Panel C:** shows 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, γ_1, γ_2 are the coefficients for the effect of age. The parameter α_{11} 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, namely at any time t the effect of the instantaneous underlying value (dashed line in panel B of Figure 1) of PSA $m_{2i}(t)$ is given by α_{21}, and the effect of the instantaneous

underlying velocity (panel C in Figure 1) of PSA $\partial m_{2i}(t)/\partial t$ is given by α_{22}. 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 1.

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:

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{b} \mid \mathcal{D}_n) &\propto \prod_{i=1}^{n} p(l_i, r_i, \boldsymbol{y}_{1i}, \boldsymbol{y}_{2i}, \mid \boldsymbol{b}_i, \boldsymbol{\theta}) p(\boldsymbol{b}_i \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) \\ &\propto \prod_{i=1}^{n} p(l_i, r_i \mid \boldsymbol{b}_i, \boldsymbol{\theta}) p(\boldsymbol{y}_{1i} \mid \boldsymbol{b}_i, \boldsymbol{\theta}) p(\boldsymbol{y}_{2i} \mid \boldsymbol{b}_i, \boldsymbol{\theta}) p(\boldsymbol{b}_i \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}), \\ p(\boldsymbol{b}_i \mid \boldsymbol{\theta}) &= \frac{1}{\sqrt{(2\pi)^q \det(\boldsymbol{D})}} \exp(\boldsymbol{b}_i^T \boldsymbol{D}^{-1} \boldsymbol{b}_i), \end{aligned}$$

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

$$p(\boldsymbol{y}_{1i} \mid \boldsymbol{b}_i, \boldsymbol{\theta}) = \prod_{k=1}^{n_{1i}} \frac{\exp\Big[- \operatorname{logit}\big\{\Pr(y_{1ik} > \mathrm{T1c})\big\} I(y_{1ik} = \mathrm{T1c})\Big]}{1 + \exp\Big[- \operatorname{logit}\big\{\Pr(y_{1ik} > \mathrm{T1c})\big\}\Big]},$$

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

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

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

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

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 $\{\beta_{01}, \ldots, \beta_{31}, \beta_{02}, \ldots, \beta_{62}\}$, 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 q (number of random effects). For the relative risk model's parameters $\{\gamma_1, \gamma_2\}$ and the association parameters $\{\alpha_{11}, \alpha_{21}, \alpha_{22}\}$, we use independent normal priors with zero mean and variance 100.

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

The posterior parameter estimates for the joint model we fitted to the PRIAS dataset 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), and 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 $\boldsymbol{b} = (b_{01}, b_{11}, b_{02}, b_{12}, b_{22}, b_{32}, b_{42})$ (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_{01}	b_{11}	b_{02}	b_{12}	b_{22}	b_{32}	b_{42}
b_{01}	7.546	-0.564	-0.182	0.075	0.084	0.003	-0.019
b_{11}	-0.564	1.379	0.081	0.119	0.165	0.266	0.219
b_{02}	-0.182	0.081	0.208	0.031	0.034	0.068	0.014
b_{12}	0.075	0.119	0.031	0.224	0.109	0.158	0.088
b_{22}	0.084	0.165	0.034	0.109	0.293	0.324	0.238
b_{32}	0.003	0.266	0.068	0.158	0.324	0.480	0.312
b_{42}	-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 2 we present the marginal evolution of probability of obtaining a DRE score larger than T1c, 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 3.

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
(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

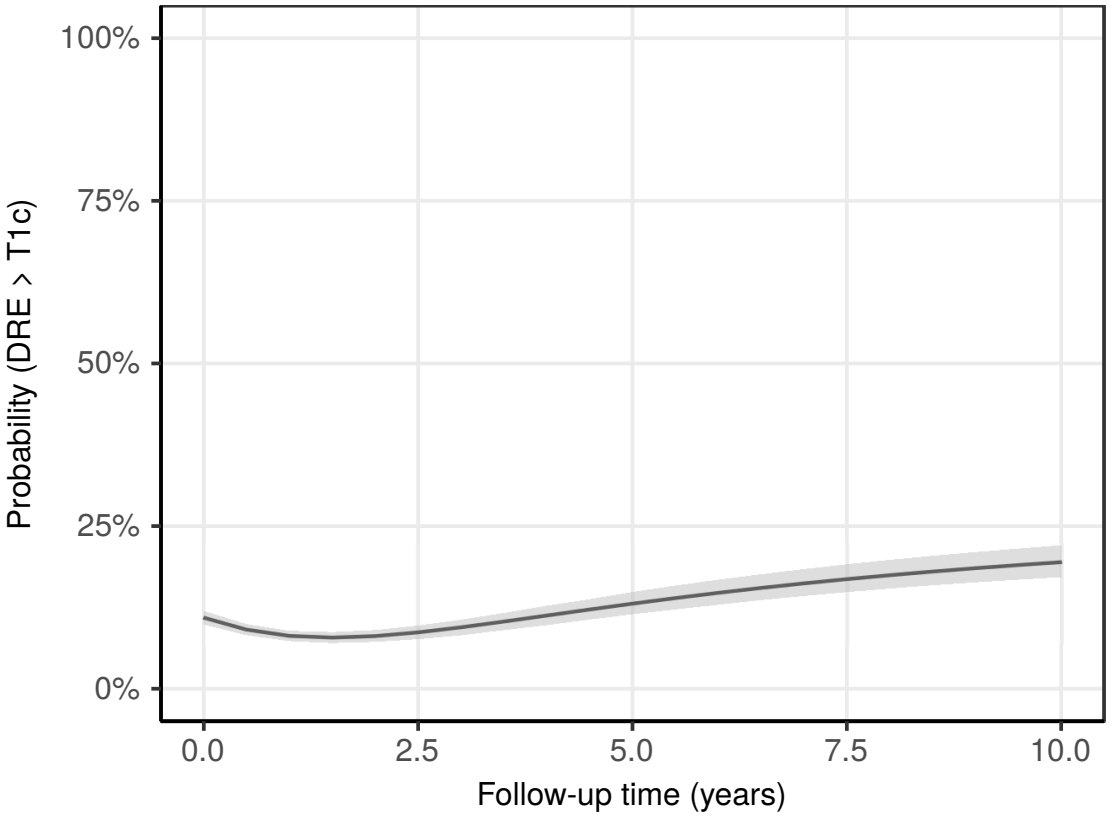

Figure 2: Fitted average probability of obtaining a DRE score larger than T1c with 95% credible interval, over a period of 10 years, for a hypothetical AS patient who is included in AS at the age of 70 years.

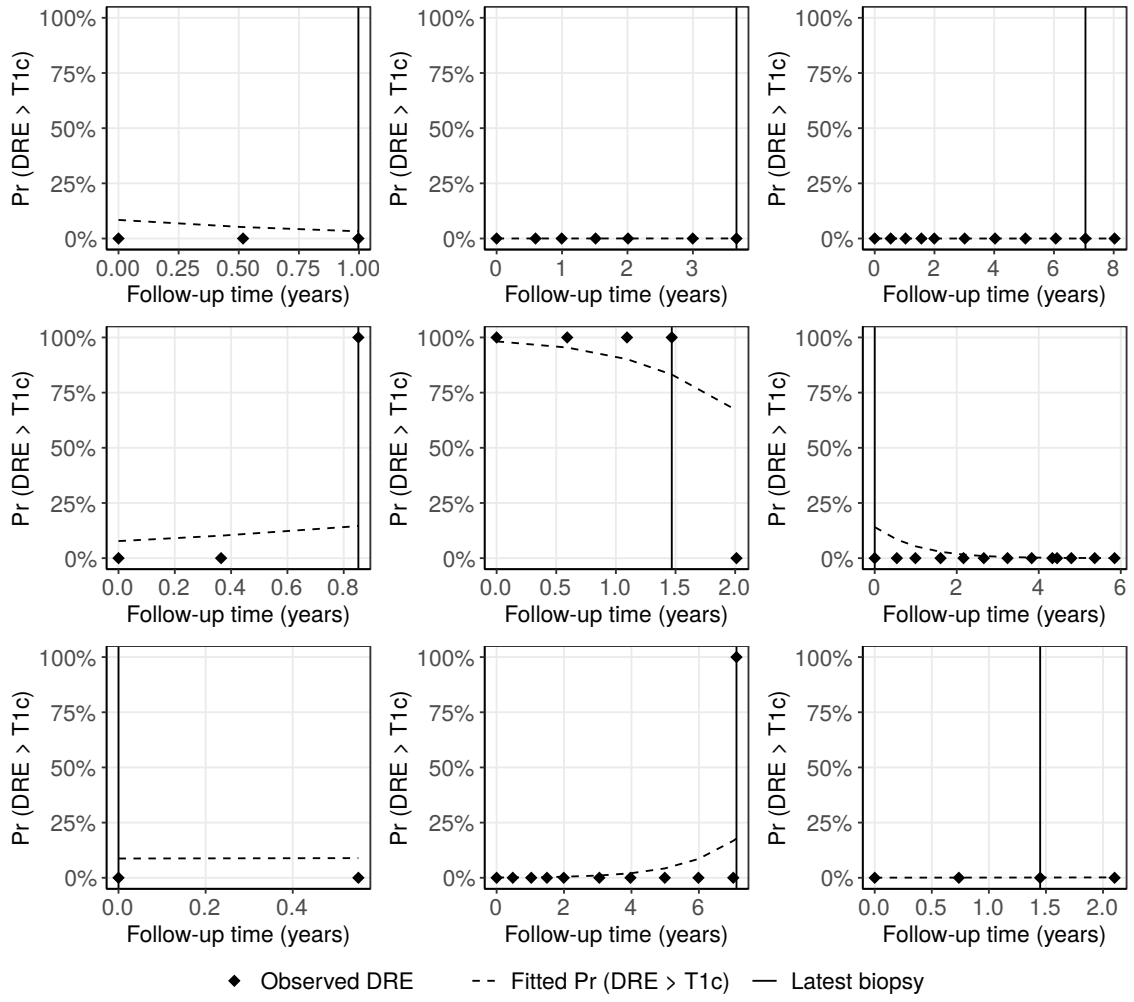

Figure 3: Observed DRE versus fitted probabilities of obtaining a DRE score larger than T1c, for nine randomly selected PRIAS patients. The fitted profiles utilize information from the observed DRE scores, PSA levels, and time of the latest biopsy. Observed DRE scores plotted against 0% probability are 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 4 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 5.

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 $\times 10^{-4}$	9.8 $\times 10^{-5}$	-6.6 $\times 10^{-4}$	-2.7 $\times 10^{-4}$	<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	

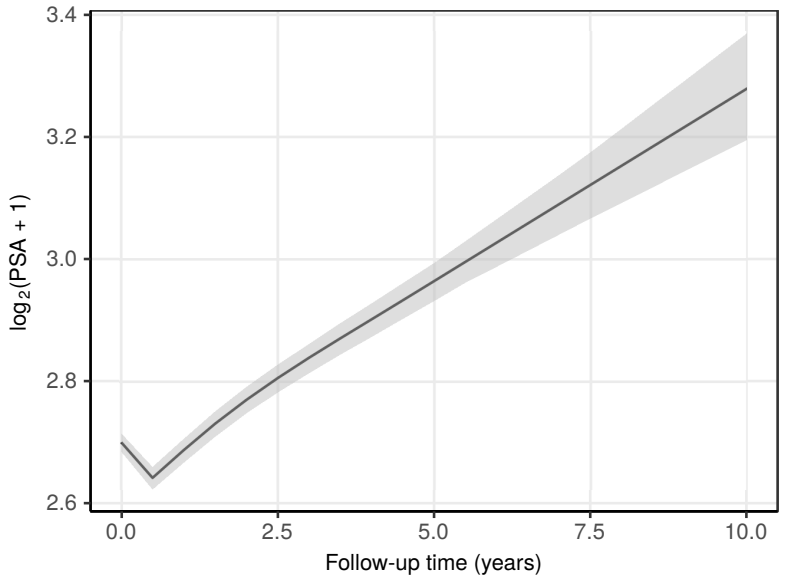

Figure 4: Fitted marginal evolution of $\log_2(\text{PSA}+1)$ levels over a period of 10 years with 95% credible interval, for a hypothetical patient who is included in AS at the age of 70 years.

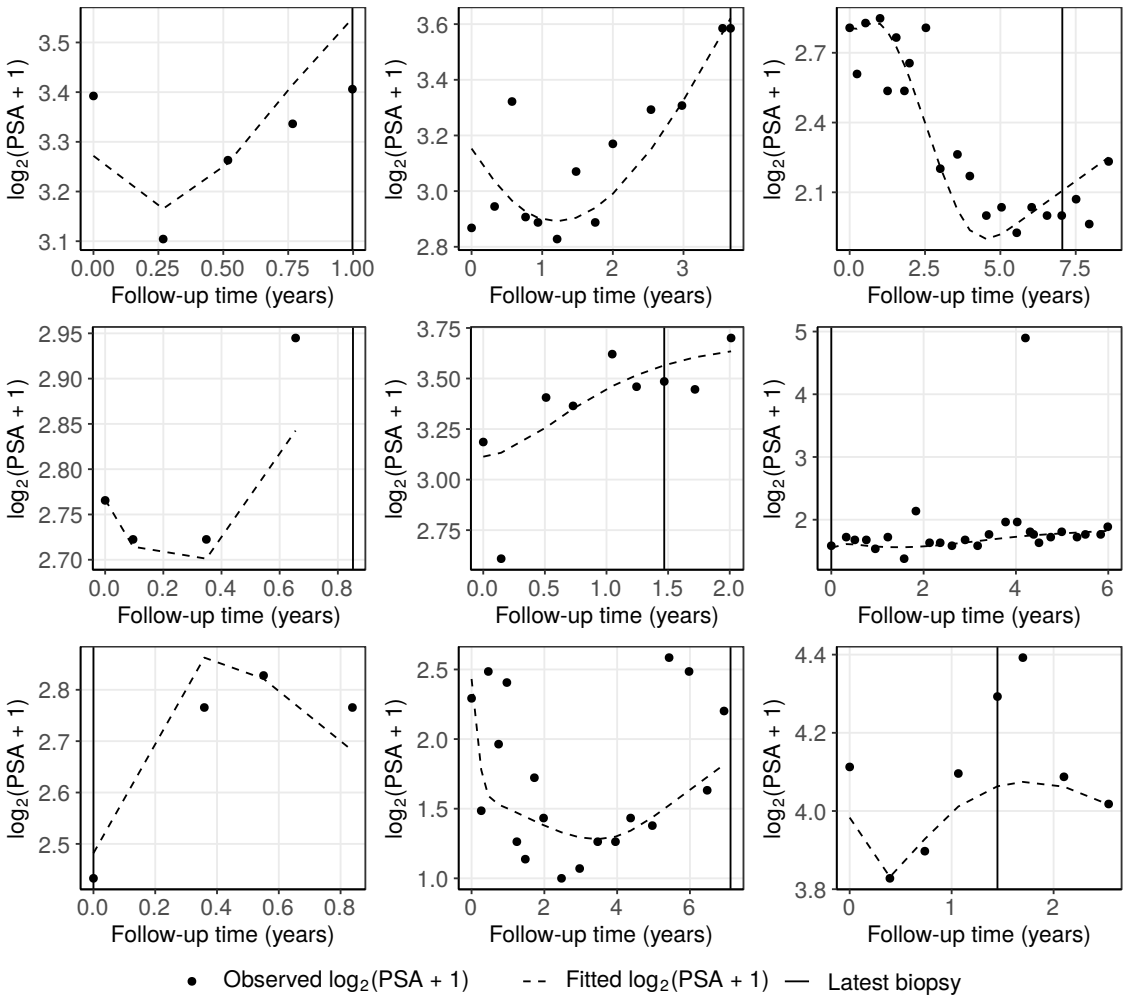

Figure 5: Fitted versus observed $\log_2(\text{PSA}+1)$ profiles for nine randomly selected PRIAS patients. The fitted profiles utilize information from the observed PSA levels, DRE scores, and time of the latest biopsy.

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. 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. An increase in age at the time of inclusion in AS from 65 years to 75 years (first and third quartiles of age in PRIAS dataset) corresponds to a 1.13 fold increase in the hazard of GR.

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
$\text{logit}\{\Pr(\text{DRE} > \text{T1c})\}$	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

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

With regards to the choice of the distribution for the error term ε_2 for the PSA measurements (see Equation 2), we attempted fitting multiple joint models differing in error distribution, namely t-distribution with three, four, and five degrees of freedom, and a normal distribution for the error term. However, the model assumptions were best met by the model with t-distribution having three degrees of freedom, for the error terms. The quantile-quantile plot of subject-specific residuals for the corresponding model in Figure 6, shows that the assumption of t-distributed (df=3) errors is reasonably met by the fitted model.

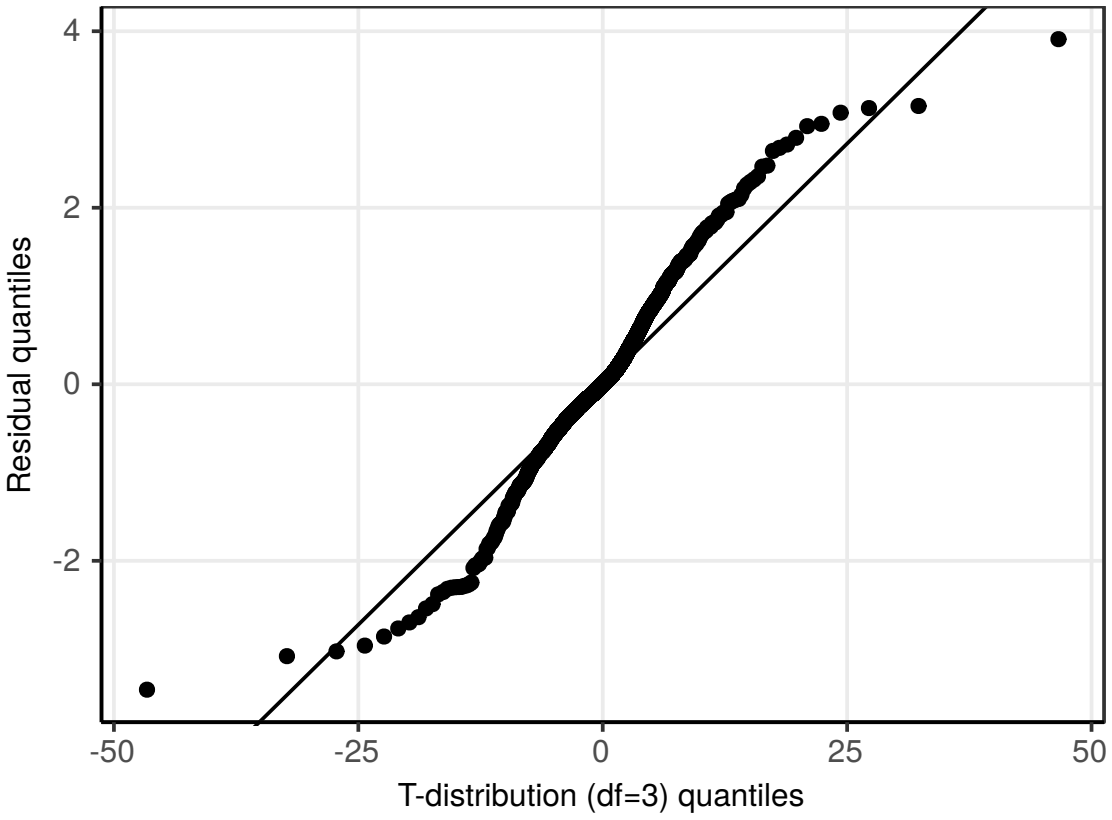

Figure 6: Quantile-quantile plot of subject-specific residuals from the joint model fitted to the PRIAS dataset.

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

To compare the predictive performance of a models having association between hazard of GR and value of longitudinal outcome values, versus a model having the association with both value and velocity, we calculate the area under the receiver operating characteristic curves, also called AUC (Rizopoulos, Molenberghs, and Lesaffre, 2017), for these models (with the only change that $\log_2$ PSA levels are used as the outcome). Since in a joint model time dependent AUC is more relevant, we calculate the AUC at year one, year two and year three of follow-up in AS. The time window for which the AUC is calculated is one year. The resulting AUC are presented in 5.

Table 5: Area under the receiver operating characteristic curves (AUC), and 95% confidence interval in brackets. AUC's are calculated for two joint models: first one having association between hazard of GR and longitudinal outcome's value as well as velocity, and second one having association with only longitudinal outcome's value (with the only change that $\log_2$ PSA levels are used as the outcome).

Year	value and velocity association	value association
1	0.613 [0.582, 0.632]	0.595 [0.565, 0.618]
2	0.648 [0.608, 0.685]	0.609 [0.568, 0.654]
3	0.593 [0.560, 0.638]	0.590 [0.536, 0.628]

Appendix C 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.

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.

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

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.