

## Personalized schedules for prostate cancer biopsies

Anirudh Tomer<sup>1,\*</sup>, Daan Nieboer<sup>2</sup>, Monique J. Roobol<sup>3</sup>, Ewout W. Steyerberg<sup>2</sup>, and Dimitris Rizopoulos<sup>1</sup>

<sup>1</sup>Department of Biostatistics, Erasmus University Medical Center, the Netherlands

<sup>2</sup>Department of Public Health, Erasmus University Medical Center, the Netherlands

<sup>3</sup>Department of Urology, Erasmus University Medical Center, the Netherlands

\*email: atomer@erasmusmc.nl

**SUMMARY:** Low risk prostate cancer patients enrolled in active surveillance (AS) programs have to undergo biopsies on a frequent basis for examination of disease progression. Majority of the AS programs worldwide employ fixed schedules of biopsies for all patients. It has been found that such fixed and frequent schedules discourage to receive biopsies, and also bring a financial burden on the healthcare systems. Motivated by the world's largest AS program PRIAS, in this paper we present personalized schedules for biopsies to counter these problems. Using joint models for time to event and longitudinal data, our methods combine information from previous biopsy results and historical prostate-specific antigen (PSA) levels of a patient, to schedule the next biopsy. We also present criteria to compare the efficacy of personalized schedules with that of existing biopsy schedules, and a method to select the optimal schedule.

**KEY WORDS:** Personalized medicine; Prostate cancer; Active surveillance; Joint models.

### 1. Introduction

In this decade prostate cancer is the second most frequently diagnosed cancer (14% of all cancers) in males worldwide, with nearly 67% of all prostate cancer cases reported in developed countries (Torre et al., 2015). The increase in diagnosis of low grade prostate cancers has been attributed to increase in life expectancy and increase in number of screening programs (Potosky et al., 1995). A major issue of screening programs that also has been established in other types of cancers (e.g., breast cancer) is over-diagnosis. To avoid overtreatment, patients diagnosed with low grade prostate cancer are often motivated to join active surveillance (AS) programs. The goal of AS is to routinely examine the progression of prostate cancer and avoid serious treatments such as surgery or chemotherapy as long as they are not needed.

Currently the largest AS program worldwide is Prostate Cancer Research International Active Surveillance, also known as PRIAS (Bokhorst et al., 2016). Patients enrolled in PRIAS are closely monitored using serum prostate-specific antigen (PSA) levels, digital rectal examination (DRE) and repeat prostate biopsies. Results from biopsies are graded on a scale called Gleason, which takes values between 2 and 10, with 10 corresponding to a serious state of the disease. At the time of induction in PRIAS, patients must have a Gleason score of 6 or less, DRE score of cT2c or less and a PSA of 10 ng/mL or less. In PRIAS, a PSA doubling time (PSA-DT) of less than 3 years indicates prostate cancer progression, where PSA-DT is measured as the inverse of the slope of regression line through the base 2 logarithm of PSA values. However until either DRE or Gleason are observed to be higher than the aforementioned thresholds, patients are not removed from AS for curative treatment (Bokhorst et al., 2016). When the

Gleason score becomes greater than 6, it is also known as Gleason reclassification (referred to as GR hereafter).

Biopsies are reliable but they are also difficult to conduct, cause pain and have serious side effects such as hematuria and sepsis (Loeb et al., 2013). Due to these reasons, majority of the AS programs worldwide strongly advise at most one biopsy per year. Conducting a biopsy annually (we refer to it as annual schedule hereafter) has the advantage that GR can be detected within one year of its occurrence. The drawbacks of annual schedule although, are not only medical but also financial. Keegan et al. (2012) have shown that annual schedules can cost more than the treatment (brachytherapy or prostatectomy) to AS programs, and if biopsies were to be conducted every other year, then up to 28% increase in savings per head from AS over treatment could be achieved. Despite this, several AS programs employ the annual schedule (Tosoian et al., 2011; Welty et al., 2015).

A drawback of existing biopsy schedules is that they conduct unnecessary biopsies for patients with slowly progressing disease. This has lead to a high non-compliance rate for repeat biopsies (Bokhorst et al., 2015). In PRIAS some of the reasons reported for non-compliance were: 'patient does not want biopsy', 'complications on last biopsy' and 'no signs of disease progression on previous biopsy'. Non-compliance reduces the effectiveness of AS programs because progression is detected late. The PRIAS schedule and the compliance rates are: one biopsy each at year 1 (81%), year 4 (60%), year 7 (53%) and year 10 (33%). After year 10 biopsies are conducted every five years. An exception is made, if at any time a patient has PSA-DT less than 10 years, wherein annual schedule is prescribed. The reason to use PSA-DT in the scheduling process is that it was found to be indicative of GR in PRIAS (Bokhorst et al., 2015). It is important to note that measurement of PSA has a

high compliance rate of 91% in PRIAS. This is because PSA is easy to obtain and measuring PSA does not lead to any side effects.

This paper is motivated by the need to reduce the burden of biopsies and most optimally find the onset of GR. To this end, we intend to create personalized schedules for biopsies which improve upon the PRIAS and annual schedule. That is, a different schedule for every patient utilizing his recorded information. Personalized schedules for screening have received much interest in the literature, especially in the medical decision making context. For diabetic retinopathy, Markov models have been used by Bebu and Lachin (2017) to develop cost optimized personalized schedules. For breast cancer, personalized mammography screening policy based on the prior screening history and personal risk characteristics of women, using partially observable Markov decision process (MDP) models have been proposed by Ayer, Alagoz, and Stout (2012). MDP models have also been used to develop personalized screening policies for cervical cancer (Akhavan-Tabatabaei, Sánchez, and Yeung, 2017) and colorectal cancer (Erenay, Alagoz, and Said, 2014). Another type of model called joint model for time to event and longitudinal data (Tsiatis and Davidian, 2004; Rizopoulos, 2012) has also been used to create personalized schedules, albeit for longitudinal biomarkers (Rizopoulos et al., 2016). In context of prostate cancer, Zhang et al. (2012) have used partially observable MDP models to personalize the decision of (not) deferring a biopsy to the next checkup time during the screening process. The decision is based on a discretized PSA level of the patient at the current check up time, and the baseline characteristics.

Our work differs from the above referenced work in certain aspects. Firstly, the schedules we propose in this paper, account for the latent between-patient heterogeneity. We achieve this using joint models, which are inherently patient-specific because they utilize random effects. Secondly, joint models allow a continuous time scale and utilize the entire history of PSA levels. Lastly, instead of making a binary decision of (not) deferring a biopsy to the next check up time, we schedule biopsies at a per patient optimal future time, utilizing the historical PSA measurements and repeat biopsy results of the patient up to the current check up time. To this end, using joint models we first obtain a full specification of the joint distribution of PSA levels and time of GR. We then use it to define a patient-specific posterior predictive distribution of the time of GR given the observed PSA measurements and previous biopsies. Using the general framework of Bayesian decision theory, we propose a set of loss functions which are minimized to find the optimal time of conducting a biopsy. These loss functions yield us two categories of personalized schedules, those based on expected time of GR and those based on the risk of GR. We also analyze an approach where the two types of schedules are combined. To compare the proposed personalized schedules with the PRIAS and annual schedule we conduct a simulation study, and then discuss various criteria for evaluating the efficacy of each schedule, and a method to choose the most suitable one.

The rest of the paper is organized as follows. Section 2 briefly covers the joint modeling framework. Section 3 details the personalized scheduling approaches we have proposed in this paper. In Section 4 we discuss criteria for evaluation of the

efficacy of a schedule and the choice of the optimal schedule. In Section 5 we demonstrate the personalized schedules by employing them for the patients from the PRIAS program. Lastly, in Section 6, we present the results from a simulation study we conducted to compare personalized schedules with PRIAS and annual schedule.

## 2. Joint model for time to event and longitudinal outcomes

We start with the definition of the joint modeling framework that will be used to fit a model to the available dataset, and then to plan biopsies for future patients. Let  $T_i^*$  denote the true GR time for the  $i$ -th patient enrolled in an AS program. Let the vector of times at which biopsies are conducted for this patient be denoted by  $T_i^b = \{T_{i0}^b, T_{i1}^b, \dots, T_{iN_i^b}^b; T_{ij}^b < T_{ik}^b, \forall j < k\}$ , where  $N_i^b$  are the total number of biopsies conducted. Because of the periodical nature of biopsy schedules,  $T_i^*$  cannot be observed directly and it is only known to fall in an interval  $(l_i, r_i]$ , where  $l_i = T_{iN_i^b-1}^b, r_i = T_{iN_i^b}^b$  if GR is observed, and  $l_i = T_{iN_i^b}^b, r_i = \infty$  if patient drops out of AS before GR is observed. Further let  $\mathbf{y}_i$  denote the  $n_i \times 1$  vector of PSA levels for the  $i$ -th patient. For a sample of  $n$  patients the observed data is denoted by  $\mathcal{D}_n = \{l_i, r_i, \mathbf{y}_i; i = 1, \dots, n\}$ .

The longitudinal outcome of interest, namely PSA level is continuous in nature and thus to model it the joint model utilizes a linear mixed effects model (LMM) of the form:

$$\begin{aligned} y_i(t) &= m_i(t) + \varepsilon_i(t) \\ &= \mathbf{x}_i^T(t)\boldsymbol{\beta} + \mathbf{z}_i^T(t)\mathbf{b}_i + \varepsilon_i(t), \end{aligned}$$

where  $\mathbf{x}_i(t)$  denotes the row vector of the design matrix for fixed effects and  $\mathbf{z}_i(t)$  denotes the same for random effects. Correspondingly the fixed effects are denoted by  $\boldsymbol{\beta}$  and random effects by  $\mathbf{b}_i$ . The random effects are assumed to be normally distributed with mean zero and  $q \times q$  covariance matrix  $\mathbf{D}$ . The true and unobserved PSA level at time  $t$  is denoted by  $m_i(t)$ . Unlike  $y_i(t)$ , the former is not contaminated with the measurement error  $\varepsilon_i(t)$ . The error is assumed to be normally distributed with mean zero and variance  $\sigma^2$ , and is independent of the random effects  $\mathbf{b}_i$ .

To model the effect of PSA on hazard of GR, joint models utilize a relative risk sub-model. The hazard of GR for patient  $i$  at any time point  $t$ , denoted by  $h_i(t)$ , depends on a function of subject specific linear predictor  $m_i(t)$  and/or the random effects:

$$\begin{aligned} h_i(t | \mathcal{M}_i(t), \mathbf{w}_i) &= \lim_{\Delta t \rightarrow 0} \frac{\Pr\{T_i^* \in [t, t + \Delta t) | T_i^* \geq t, \mathcal{M}_i(t), \mathbf{w}_i\}}{\Delta t} \\ &= h_0(t) \exp[\boldsymbol{\gamma}^T \mathbf{w}_i + f\{M_i(t), \mathbf{b}_i, \boldsymbol{\alpha}\}], \quad t > 0, \end{aligned}$$

where  $\mathcal{M}_i(t) = \{m_i(v), 0 \leq v \leq t\}$  denotes the history of the underlying PSA levels up to time  $t$ . The vector of baseline covariates is denoted by  $\mathbf{w}_i$ , and  $\boldsymbol{\gamma}$  are the corresponding parameters. The function  $f(\cdot)$  parametrized by vector  $\boldsymbol{\alpha}$  specifies the functional form of PSA levels (Brown, 2009; Rizopoulos, 2012; Taylor et al., 2013; Rizopoulos et al., 2014) that is used in the linear predictor of the relative risk model. Some functional forms relevant to the problem at hand are

the following:

$$\begin{cases} f\{M_i(t), \mathbf{b}_i, \boldsymbol{\alpha}\} = \alpha m_i(t), \\ f\{M_i(t), \mathbf{b}_i, \boldsymbol{\alpha}\} = \alpha_1 m_i(t) + \alpha_2 m'_i(t), \quad \text{with } m'_i(t) = \frac{dm_i(t)}{dt}. \end{cases}$$

These formulations of  $f(\cdot)$  postulate that the hazard of GR at time  $t$  may be associated with the underlying level of PSA  $m_i(t)$  at  $t$ , or with both the level and velocity of the PSA  $m'_i(t)$  at  $t$ . Lastly,  $h_0(t)$  is the baseline hazard at time  $t$ , and is modeled flexibly using P-splines. More specifically:

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

where  $B_q(t, \mathbf{v})$  denotes the  $q$ -th basis function of a B-spline with knots  $\mathbf{v} = v_1, \dots, v_Q$  and vector of spline coefficients  $\gamma_{h_0}$ . 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  $\gamma_{h_0}$  are penalized using a differences penalty (Eilers and Marx, 1996).

For the estimation of joint model's parameters we use a Bayesian approach. The details of the estimation method are presented in Web Appendix A of the supplementary material.

### 3. Personalized schedules for repeat biopsies

Once a joint model for GR and PSA levels is obtained, the next step is to use it to create personalized schedules for biopsies. To elucidate personalized schedules, let us assume that a personalized schedule is to be created for a new patient  $j$ , who is not present in the original sample of patients  $\mathcal{D}_n$ . Further let us assume that this patient did not have a GR at his last biopsy performed at time  $t$ , and that the PSA levels are available up to a time point  $s$ . The goal is to find the optimal time  $u \geq \max(t, s)$  of the next biopsy.

#### 3.1 Posterior predictive distribution for time to GR

Let  $\mathcal{Y}_j(s)$  denote the history of PSA levels taken up to time  $s$  for patient  $j$ . The information from PSA history and repeat biopsies is manifested by the posterior predictive distribution  $g(T_j^*)$ , given by (conditioning on baseline covariates  $\mathbf{w}_i$  is dropped for notational simplicity hereafter):

$$\begin{aligned} g(T_j^*) &= p\{T_j^* | T_j^* > t, \mathcal{Y}_j(s), \mathcal{D}_n\} \\ &= \int p\{T_j^* | T_j^* > t, \mathcal{Y}_j(s), \boldsymbol{\theta}\} p(\boldsymbol{\theta} | \mathcal{D}_n) d\boldsymbol{\theta} \\ &= \int \int p\{T_j^* | T_j^* > t, \mathbf{b}_j, \boldsymbol{\theta}\} p\{\mathbf{b}_j | T_j^* > t, \mathcal{Y}_j(s), \boldsymbol{\theta}\} p(\boldsymbol{\theta} | \mathcal{D}_n) d\mathbf{b}_j d\boldsymbol{\theta} \end{aligned} \quad (1)$$

The distribution  $g(T_j^*)$  depends on the observed longitudinal history of patient  $j$  via the random effects  $\mathbf{b}_j$ , and on the information from the original dataset  $\mathcal{D}_n$  via the posterior distribution of the parameters  $p(\boldsymbol{\theta} | \mathcal{D}_n)$ , where  $\boldsymbol{\theta}$  denotes the vector of all parameters.

#### 3.2 Loss functions

To find the time  $u$  of the next biopsy, we use principles from statistical decision theory in a Bayesian setting (Berger, 1985; Robert, 2007). More specifically, we propose to choose  $u$  by

minimizing the posterior expected loss  $E_g\{L(T_j^*, u)\}$ , where the expectation is taken with respect to  $g(T_j^*)$ . The former is given by:

$$E_g\{L(T_j^*, u)\} = \int_t^\infty L(T_j^*, u) p\{T_j^* | T_j^* > t, \mathcal{Y}_j(s), \mathcal{D}_n\} dT_j^*.$$

Various loss functions  $L(T_j^*, u)$  have been proposed in literature (Robert, 2007). The ones we utilize, and the corresponding motivations are presented next.

One of the reasons, patients did not comply with the existing PRIAS schedule was 'complications on a previous biopsy'. Therefore, it is required to have as less biopsies as possible. In the ideal case only one biopsy, performed at the exact time of GR is sufficient. Hence, neither a time which overshoots the true GR time  $T_j^*$ , nor a time which undershoots is preferred. In this regard, the squared loss function  $L(T_j^*, u) = (T_j^* - u)^2$  and absolute loss function  $L(T_j^*, u) = |T_j^* - u|$  have the properties that the posterior expected loss is symmetric on both sides of  $T_j^*$ . Secondly, both loss functions have well known solutions available. The posterior expected loss for the squared loss function is given by:

$$\begin{aligned} E_g\{L(T_j^*, u)\} &= E_g\{(T_j^* - u)^2\} \\ &= E_g\{(T_j^*)^2\} + u^2 - 2uE_g(T_j^*). \end{aligned} \quad (2)$$

The posterior expected loss in (2) attains its minimum at  $u = E_g(T_j^*)$ , the expected time of GR. The posterior expected loss for the absolute loss function is given by:

$$\begin{aligned} E_g\{L(T_j^*, u)\} &= E_g(|T_j^* - u|) \\ &= \int_u^\infty (T_j^* - u) g(T_j^*) dT_j^* + \int_t^u (u - T_j^*) g(T_j^*) dT_j^*. \end{aligned} \quad (3)$$

The posterior expected loss in (3) attains its minimum at the median of  $g(T_j^*)$ , given by  $u = \pi_j^{-1}(0.5 | t, s)$ , where  $\pi_j^{-1}(\cdot)$  is the inverse of dynamic survival probability  $\pi_j(u | t, s)$  of patient  $j$  (Rizopoulos, 2011). It is given by:

$$\pi_j(u | t, s) = \Pr\{T_j^* \geq u | T_j^* > t, \mathcal{Y}_j(s), \mathcal{D}_n\}, \quad u \geq t. \quad (4)$$

For ease of readability we denote  $\pi_j^{-1}(0.5 | t, s)$  as median( $T_j^*$ ) hereafter.

Even though the mean or median time of GR may be obvious choices from a statistical perspective, from the viewpoint of doctors or patients, it could be more intuitive to make the decision for the next biopsy by placing a cutoff  $1 - \kappa, \kappa \in [0, 1]$  on the risk of GR. This approach would be successful if  $\kappa$  can sufficiently well differentiate between patients who will obtain GR in a given period of time, and those who will not. This approach is also useful when patients are apprehensive about delaying biopsies beyond a certain risk cutoff. In this regard, a biopsy can be scheduled at a time point  $u$  such that the dynamic risk of GR is higher than a certain threshold  $1 - \kappa$ , beyond  $u$ . To this end, the posterior expected loss for the following multilinear loss function can be minimized to find the optimal  $u$ :

$$L_{k_1, k_2}(T_j^*, u) = \begin{cases} k_2(T_j^* - u), & k_2 > 0 \quad \text{if } T_j^* > u, \\ k_1(u - T_j^*), & k_1 > 0 \quad \text{otherwise.} \end{cases} \quad (5)$$

where  $k_1, k_2$  are constants parameterizing the loss function.

The posterior expected loss  $E_g\{L_{k_1, k_2}(T_j^*, u)\}$  obtains its minimum at  $u = \pi_j^{-1}\{k_1/(k_1 + k_2) \mid t, s\}$  (Robert, 2007). The choice of the two constants  $k_1$  and  $k_2$  is equivalent to the choice of  $\kappa = k_1/(k_1 + k_2)$ .

**3.2.1 A hybrid approach.** In practice, for some patients we may not have sufficient information, resulting in high variance of  $g(T_j)$ , which in turn would make using a measure of central tendency such as mean or median time of GR unreliable (i.e., overshooting the true  $T_j^*$  by a big margin). In such occasions, the approach based on dynamic risk of GR could be more robust. This consideration leads us to a hybrid approach, namely, to select  $u$  using dynamic risk of GR based approach when the spread of  $g(T_j^*)$  is large, while using  $E_g(T_j^*)$  or median( $T_j^*$ ) when the spread of  $g(T_j^*)$  is small. What constitutes a large spread will be application-specific. In the PRIAS schedule, the maximum possible delay in detection of GR is three years. Thus we propose that if the difference between the 0.025 quantile of  $g(T_j^*)$ , and  $E_g(T_j^*)$  or median( $T_j^*$ ) is more than three years then proposals based on dynamic risk of GR be used instead.

### 3.3 Estimation

Since there is no closed form solution available for  $E_g(T_j^*)$ , for its estimation we utilize the following relationship between  $E_g(T_j^*)$  and  $\pi_j(u \mid t, s)$ :

$$E_g(T_j^*) = t + \int_t^\infty \pi_j(u \mid t, s) du. \quad (6)$$

There is no closed form solution available for the integral in (6), and hence we approximate it using Gauss-Kronrod quadrature. We preferred this approach over Monte Carlo methods to estimate  $E_g(T_j^*)$  from  $g(T_j^*)$ , because sampling directly from  $g(T_j^*)$  involved an additional step of sampling from the distribution  $p(T_j^* \mid T_j^* > t, \mathbf{b}_j, \boldsymbol{\theta})$ , as compared to the estimation of  $\pi_j(u \mid t, s)$  (Rizopoulos, 2011). The former approach was thus computationally faster.

For schedules based on dynamic risk of GR, the value of  $\kappa$  dictates the biopsy schedule and thus its choice has important consequences. In certain cases it may be chosen on the basis of doctor's advice or the amount of risk that is acceptable to the patient. For example, if the maximum acceptable risk is 75%, then  $\kappa = 0.25$ .

In cases where  $\kappa$  cannot be chosen on the basis of the input of the physician or the patients, we propose to automate the choice of  $\kappa$ . More specifically, we propose to choose a  $\kappa$  for which a binary classification accuracy measure (López-Ratón et al., 2014), discriminating between cases and controls, is maximized. In PRIAS, cases are patients who experience GR and the rest are controls. However, a patient can be in control group at some time  $t$  and in the cases at some future time point  $t + \Delta t$ , and thus time dependent binary classification is more relevant. In joint models, a patient  $j$  is predicted to be a case if  $\pi_j(t + \Delta t \mid t, s) \leq \kappa$  and a control if  $\pi_j(t + \Delta t \mid t, s) > \kappa$  (Rizopoulos, 2016; Rizopoulos, Molenberghs, and Lesaffre, 2017). The time window  $\Delta t$  can be automatically chosen as  $\arg \max_{\Delta t} AUC(t, \Delta t, s)$ , the latter being a measure of discriminative capability of the model (Rizopoulos, 2016; Rizopoulos et al., 2017). However such a time window may not be clinically relevant at all. In AS programs at any point

in time, it is of interest to identify patients who may obtain GR in the next 1 year from those who do not, so that they can be provided immediate attention (in exceptional cases a biopsy within an year of the last one). Thus, in this work we use a  $\Delta t$  of 1 year.

For automatic selection of the threshold  $\kappa$ , we require a binary classification accuracy measure which is in line with the goal to focus on patients whose true time of GR falls in the time window  $\Delta t$ . To this end, the measure which combines both sensitivity and precision is  $F_1$  score. It is defined as:

$$F_1(t, \Delta t, s) = 2 \frac{\text{TPR}(t, \Delta t, s) \text{PPV}(t, \Delta t, s)}{\text{TPR}(t, \Delta t, s) + \text{PPV}(t, \Delta t, s)}, \quad F_1 \in [0, 1],$$

$$\text{TPR}(t, \Delta t, s) = \Pr\{\pi_j(t + \Delta t \mid t, s) \leq \kappa \mid T_j^* \in (t, t + \Delta t]\},$$

$$\text{PPV}(t, \Delta t, s) = \Pr\{T_j^* \in (t, t + \Delta t] \mid \pi_j(t + \Delta t \mid t, s) \leq \kappa\}.$$

where  $\text{TPR}(\cdot)$  and  $\text{PPV}(\cdot)$  denote time dependent true positive rate (sensitivity) and positive predictive value (precision), respectively. The estimation for both proceeds as in Rizopoulos et al. (2017). Since a high  $F_1$  score is desired, the optimal value of  $\kappa$  is  $\arg \max_{\kappa} F_1(t, \Delta t, s)$ . In this work we compute the latter using a grid search approach. That is,  $F_1$  is computed using the available dataset over a fine grid of  $\kappa$  values in the interval  $[0, 1]$  and  $\kappa$  corresponding to the highest  $F_1$  is chosen.

### 3.4 Algorithm

The aforementioned personalized schedules, schedule biopsy at a time  $u > \max(t, s)$ . However, if time  $u < T_j^*$ , then GR is not detected at  $u$  and at least one more biopsy is required at an optimal time  $u^{\text{new}} > \max(u, s)$ . This process is iteratively repeated until GR is detected. We elucidate this process via an algorithm in Figure 3.4. Since it is required by PRIAS that biopsies are conducted at a gap of at least one year, when  $u - t < 1$ , the algorithm postpones  $u$  to  $t + 1$ , because it is the time nearest to  $u$ , at which the one year gap condition is satisfied.

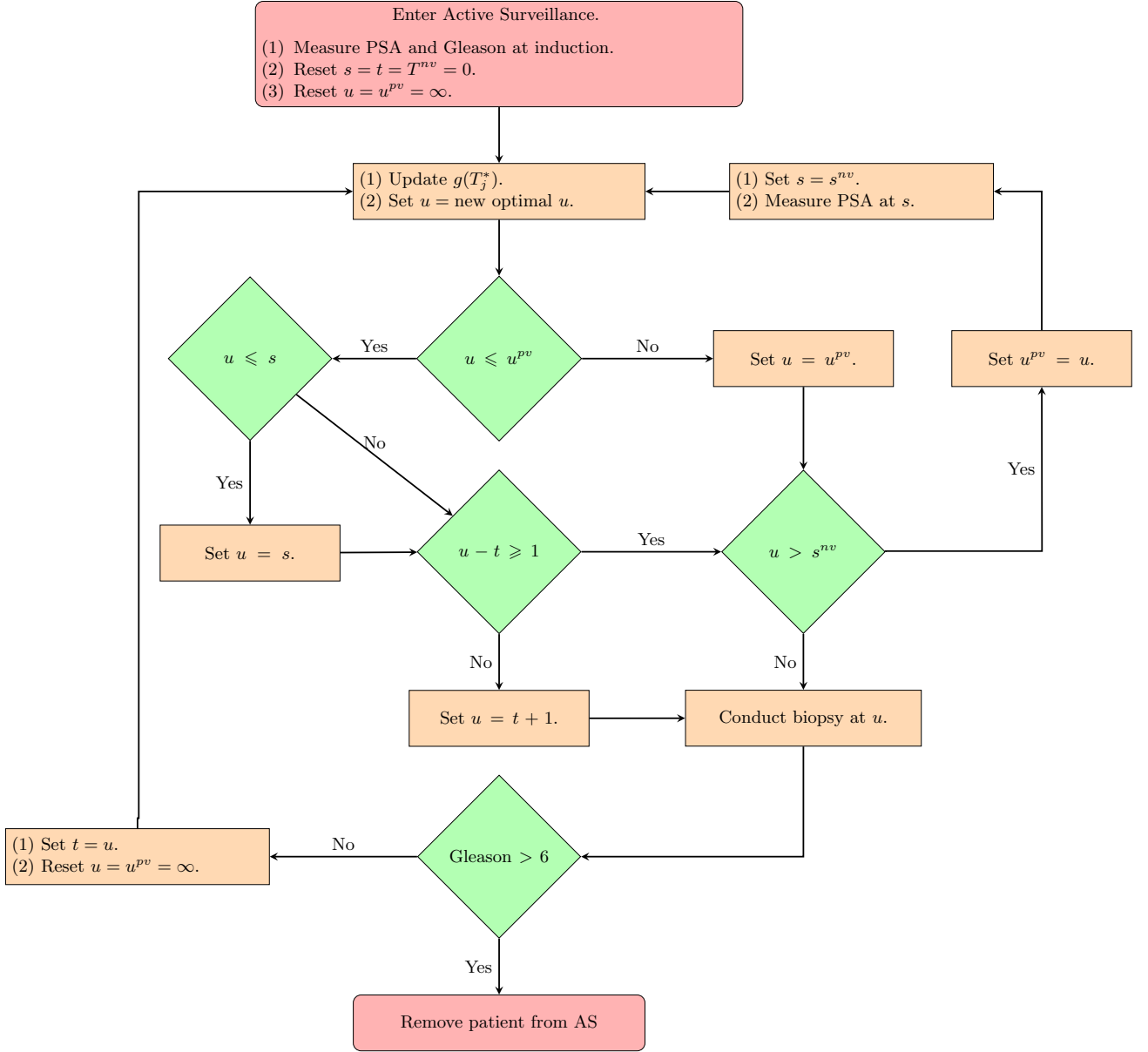
## 4. Choosing a schedule

Given a particular schedule  $S$  of biopsies, our next goal is to evaluate the efficacy of this schedule and to compare it with other schedules. To this end, we first present the criteria for evaluation of efficacy of biopsy schedules and then discuss the choice of the optimal schedule.

### 4.1 Evaluation of efficacy of schedules

The first criteria in the evaluation of efficacy of a schedule  $S$  is the number of repeat biopsies  $N^{bS}$  it conducts before GR is detected for a patient. More specifically, our interest lies in the marginal distribution  $p(N^{bS})$  of number of biopsies for the entire population of patients. Various measures of efficacy can be extracted from this distribution, such as the mean  $E(N^{bS})$ , or the variance  $\text{var}(N^{bS})$ . Given the medical and financial burden associated with biopsies, a small mean and small variance is desired. Quantiles of  $p(N^{bS})$  may also be of interest. For example, a schedule which takes less than 2 biopsies in 95% cases may be preferred.

The second criteria in evaluation of efficacy of a schedule  $S$  is the offset. The offset for a particular patient  $j$  can be defined



**Figure 1.** Algorithm for creating a personalized schedule for patient  $j$ .  $t$  denotes the time of the latest biopsy.  $s$  denotes the time of the latest available PSA measurement.  $u$  denotes the proposed personalized time of biopsy.  $u^{pv}$  denotes the time at which a repeat biopsy was proposed on the last visit to the hospital.  $T^{nv}$  denotes the time of the next visit for measurement of PSA.

as  $O_j^S = T_{jN_j^{bS}}^S - T_j^*$ , where  $N_j^{bS}$  is the number of biopsies required for patient  $j$  before GR is detected and  $T_{jN_j^{bS}}^S > T_j^*$  is the time at which GR is detected. Once again the interest lies in the marginal distribution  $p(O^S)$  of the offset for the entire population of patients. A small mean  $E(O^S)$  and small variance  $\text{var}(O^S)$  are desired.

#### 4.2 Finding the most optimal schedule

Given the multiple measure of efficacy of a schedule, the next step is to find the optimal schedule. Using principles from compound optimal designs (Läuter, 1976) we propose

to choose a schedule  $S$  which minimizes the following loss function:

$$L(S) = \sum_{g=1}^G \lambda_g \mathcal{G}_g(N^{bS})^{d_g=1} \mathcal{G}_g(O^S)^{d_g=0}. \quad (7)$$

where  $\mathcal{G}_g(\cdot)$  is a measure of efficacy of  $S$  based on either number of biopsies or the offset, and  $d_g$  is the corresponding indicator for the choice. Some examples of  $\mathcal{G}_g(\cdot)$  are mean, median, variance and quantile function. Constants  $\lambda_1, \dots, \lambda_G$ , where  $\lambda_g \in [0, 1]$  and  $\sum_{g=1}^G \lambda_g = 1$ , are weights to differentially weigh-in the contribution of each of the  $G$  measures of efficacy.

An example loss function is:

$$L(S) = \lambda_1 E(N^{bS}) + \lambda_2 E(O^S). \quad (8)$$

Here choosing values for  $\lambda_1$  and  $\lambda_2$  is not easy, because biopsies have serious medical side effects and consequently the cost of an extra biopsy cannot be quantified or compared to a unit increase in offset easily. To obviate this problem we utilize the equivalence between compound and constrained optimal designs (Cook and Wong, 1994). More specifically, it can be shown that for any  $\lambda_1$  and  $\lambda_2$  there exists a constant  $C > 0$  for which minimization of loss function in (8) is equivalent to minimization of the same, subject to the constraint that  $E(O^S) < C$ . That is, the optimal schedule is the one with the least number of biopsies and an offset less than  $C$ . The choice of  $C$  now can be based on the protocol of AS program. In the more generic case in (7), the optimal solution can be found by minimizing  $\mathcal{G}_G(\cdot)$  under the constraint  $\mathcal{G}_g < C_g$ ;  $g = 1, \dots, G - 1$ .

## 5. Personalized schedules for patients in PRIAS

To demonstrate how the personalized schedules work, we apply them to the patients enrolled in PRIAS. To this end, we divide the PRIAS dataset into a training dataset with 5264 patients and a demonstration dataset with 3 patients who never experienced GR. We fit a joint model to the training dataset and then use it to create personalized schedules for patients in demonstration dataset. We fit the joint model using the R package Jmbayes (Rizopoulos, 2016), which uses the Bayesian methodology to estimate the model parameters.

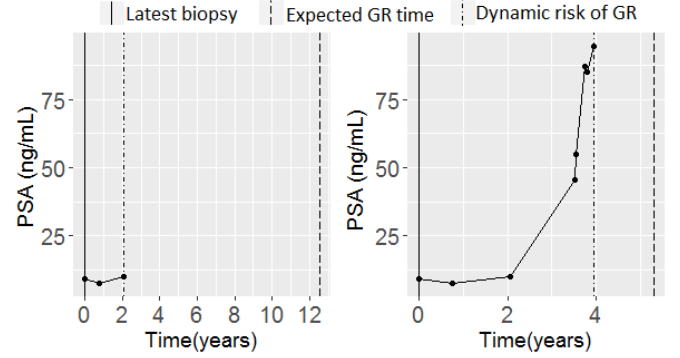
### 5.1 Fitting the joint model to PRIAS dataset

The training dataset contains age at the time of induction in PRIAS, PSA levels and the time interval in which GR is detected, for 5264 prostate cancer patients. PSA was measured at every 3 months for first 2 years and every 6 months thereafter. To detect GR, biopsies were conducted as per the PRIAS schedule (Section 1). For the longitudinal analysis of PSA we use  $\log_2$  PSA measurements instead of the raw data. This because the PSA scores take very large values around the time of disease progression, indicating that the underlying distribution for PSA is right skewed. The longitudinal sub-model of the joint model we fit is given by:

$$\log_2 \text{PSA}(t) = \beta_0 + \beta_1(\text{Age} - 70) + \beta_2(\text{Age} - 70)^2 + \sum_{k=1}^4 \beta_{k+2} B_k(t, \mathcal{K}) + b_{i0} + b_{i1} B_7(t, 0.1) + b_{i2} B_8(t, 0.1) + \varepsilon_i(t). \quad (9)$$

where  $B_k(t, \mathcal{K})$  denotes the  $k$ -th basis function of a B-spline with 3 internal knots at  $\mathcal{K} = \{0.1, 0.5, 4\}$  years, and boundary knots at 0 and 7 years. The spline for the random effects consists of 1 internal knot at 0.1 years and boundary knots at 0 and 7 years. The choice of knots was based on exploratory analysis as well as on model selection criteria AIC and BIC. Age of patients was median centered to avoid numerical instabilities during parameter estimation. For the relative risk sub-model the hazard function we fit is given by:

$$h_i(t) = h_0(t) \exp [\gamma_1(\text{Age} - 70) + \gamma_2(\text{Age} - 70)^2 + \alpha_1 m_i(t) + \alpha_2 m'_i(t)]. \quad (10)$$



**Figure 2.** PSA and repeat biopsy history, and corresponding personalized schedules for patient 3174.

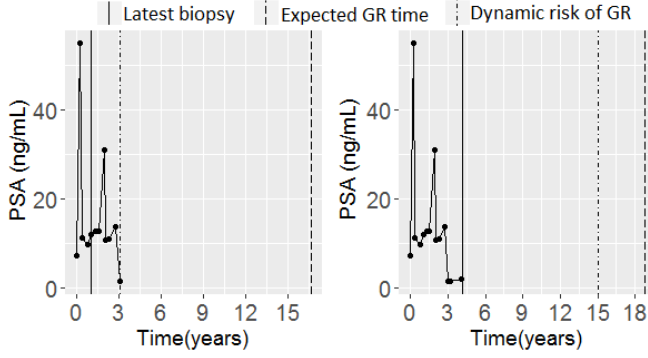
where  $\alpha_1$  and  $\alpha_2$  are measures of strength of the association between hazard of GR and  $\log_2$  PSA value  $m_i(t)$  and  $\log_2$  PSA velocity  $m'_i(t)$ , respectively. Since the PRIAS schedule depends only on the observed PSA values (via PSA-DT), the interval censoring observed in PRIAS is independent and non informative of the underlying health of the patient.

From the joint model fitted to the PRIAS dataset we found that only  $\log_2$  PSA velocity was strongly associated with hazard of GR. For any patient, a unit increase in  $\log_2$  PSA velocity led to an 11 times increase in the hazard of GR. The parameter estimates for the fitted joint model are presented in detail in Web Appendix C of the supplementary material.

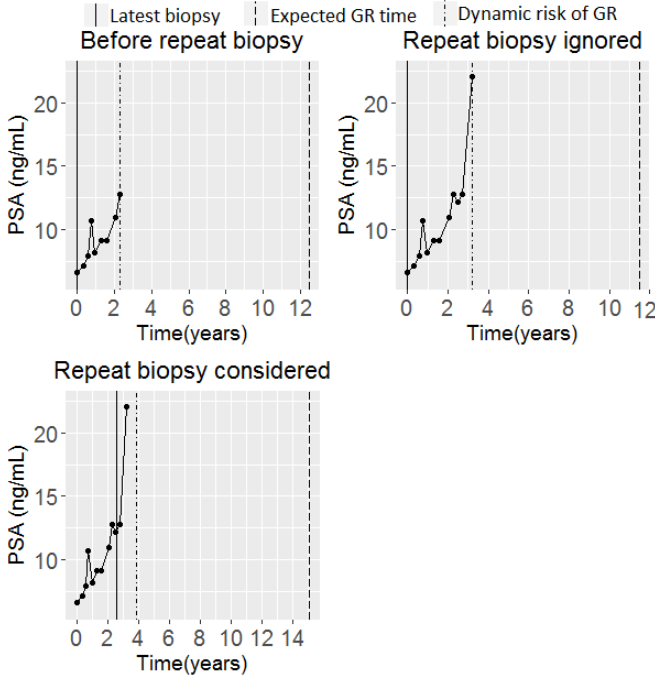
### 5.2 Demonstration of personalized schedules

Using the demonstration dataset, we next present the functioning of personalized schedules based on expected time of GR and dynamic risk of GR. The first patient of interest is patient 3174. Since no repeat biopsies were conducted for this patient in the time period we considered,  $g(T_j^*)$  depends only on the PSA levels. The evolution of PSA, repeat biopsy history and proposed times of biopsies are shown in Figure 2. It can be seen that the schedule of biopsy based on expected time of GR adjusts the times of biopsy according to the rise in hazard (which increases due steep rise in  $\log_2$  PSA velocity) after year 2. More specifically, at year 2 the proposed biopsy time is 12.5 years whereas at year 4 it decreases to 5.3 years. For schedules based on dynamic risk of GR, the optimal  $1 - \kappa$  value was found to be between 0 and 0.1 at all time points, because of the sharp rise in PSA values. This value of  $\kappa$  corresponds to a time very close to the time of latest biopsy (time 0). Hence the biopsies are scheduled much earlier than those based on expected time of GR.

The second patient of interest is patient 911, for whom the evolution of PSA, time of last biopsy and proposed biopsy times are shown in Figure 3. We can see the combined effect of decreasing PSA levels and a negative repeat biopsy on personalized schedules, between year 3 and year 4.5 for this patient. In accordance with the observed history of the patient, the proposed time of biopsy based on expected time of GR increases from 16.6 years to 18.7 years in this period. For dynamic risk of GR it increases from 3.2 years to 15 years, despite the optimal  $\kappa$  being equal to 0.98 for the entire period. Patient 2340 presents a case where information from PSA levels and repeat biopsies is conflicting. In Figure 4 we can see



**Figure 3.** PSA and repeat biopsy history, and corresponding personalized schedules for patient 911.



**Figure 4.** PSA and repeat biopsy history, and corresponding personalized schedules for patient 2340.

that the PSA for this patient becomes twice between year 2 and year 3.2. If only information from PSA is considered, then we can see that proposed time of biopsy based on expected time of GR is postponed from 12.5 to 11.5 years during this period. However, if we also take into account the negative result from the repeat biopsy at year 2.5, then the proposed time of biopsy is postponed from 12.5 years to 15 years. Thus more weight is given to a recent negative biopsy result than PSA, which is in accordance with the clinical practice. The proposed time of biopsy based on dynamic risk of GR is also postponed in light of the negative biopsy result.

The impact of evolution of PSA levels and repeat biopsy history on the variance of  $g(T_j^*)$  is presented in Web Appendix ??? of the supplementary material.

## 6. Simulation study

The application of personalized schedules for patients from PRIAS demonstrated that the schedules adapt according to the historical data of each patient. However we could not perform a full scale comparison between personalized and PRIAS schedules, because the true time of GR was not known for any of the PRIAS patients. To this end, we have performed a simulation study comparing personalized schedules based on expected time of GR, median time of GR and dynamic risk of GR, with a hybrid approach between median time of GR and dynamic risk of GR, PRIAS schedule and annual schedule. We employ these schedules for simulated patients enrolled in a hypothetical AS program, with the same entrance criteria as PRIAS.

### 6.1 Simulation setup

First we assume a population of patients enrolled in AS, whose PSA and hazard of GR follows a joint model of the form postulated in Section 5.1, with parameters equal to the posterior mean of parameters (Web Appendix C of the supplementary material) estimated from the joint model fitted to PRIAS dataset. To check the efficacy of personalized schedules in scenarios with a mixture of patients, with early as well as late failure times, we assume that the patient population belongs to three equal sized subgroups  $G_1, G_2, G_3$ . The failure times in each subpopulation are controlled by different Weibull distributed baseline hazards for each. The corresponding shape and scale parameters  $(k, \lambda)$  are: (1.5, 4), (3, 5) and (4.5, 6) for subgroups  $G_1, G_2$  and  $G_3$ , respectively. The effect of these parameters is that the variance in time of GR is highest for  $G_1$  and lowest for  $G_3$ , while the mean GR time is lowest in  $G_1$  and highest in  $G_3$ .

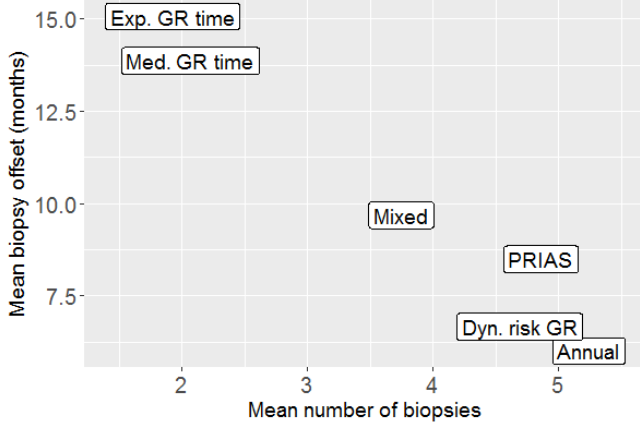
From this population we have sampled 500 datasets with 1000 patients each. Patients are randomly assigned to a subgroup. Further, each dataset is split into a training (750 patients) and a test (250 patients) part. The  $k$ -th simulated training dataset  $\mathcal{D}^k$  is given by  $\mathcal{D}^k = \{l_{ki}, r_{ki}, \mathbf{y}_{ki}; i = 1, \dots, 750\}$ , where  $\mathbf{y}_{ki}$  denote the PSA measurements for the  $i$ -th patient in  $\mathcal{D}^k$ . The frequency of PSA measurements is same as that in PRIAS. Other than simulating a true GR time  $T_{ki}^*$ , we also generate a random and non-informative censoring time  $C_{ki}$ . When  $T_{ki} < C_{ki}^*$ , then  $l_{ki} = r_{ki} = T_{ki}^*$ , otherwise  $l_{ki} = C_{ki}$  and  $r_{ki} = \infty$ . For the test patients, censoring time is not generated.

### 6.2 Estimation

We fit a joint model of the specification given in (9) and (10) to each of the  $\mathcal{D}^k, k = 1, \dots, 500$ , and obtain a MCMC sample from the posterior distribution  $p(\boldsymbol{\theta} | \mathcal{D}^k)$ . We next obtain  $g(T_{kj}^*)$  for the  $j$ -th test patient and conduct hypothetical biopsies iteratively in accordance with the algorithm in Figure 3.4. [TALK ABOUT DIFFERENT METHODS YOU COMPARE, AND HOW YOU COMPARE THEM....AND STILL IN ESTIMATION???

To compare schedules we require estimates of the various measures of efficacy of schedules (Section 4). To this end, we compute pooled estimates of each of the  $E[N^{bS}]$ ,  $\text{var}[N^{bS}]$ ,





**Figure 5.** Estimated mean number of biopsies and mean offset (months) for the 6 schedules, using all subgroups. Method names are abbreviated for better readability.

$E[O^S]$  and  $\text{var}[O^S]$ , as below:

$$\widehat{E}[O^S] = \frac{\sum_{k=1}^{500} n_k \widehat{E}[O_k^S]}{\sum_{k=1}^{500} n_k},$$

$$\widehat{\text{var}}[O^S] = \frac{\sum_{k=1}^{500} (n_k - 1) \widehat{\text{var}}[O_k^S]}{\sum_{k=1}^{500} (n_k - 1)},$$

where  $n_k$  denotes the number of test patients,  $\widehat{E}[O_k^S] = \sum_{j=1}^{n_k} O_{kj}^S / n_k$  is the estimated mean and  $\widehat{\text{var}}[O_k^S] = \sum_{j=1}^{n_k} \{O_{kj}^S - \widehat{E}[O_k^S]\}^2 / n_k - 1$  is the estimated variance of the offset for the  $k$ -th simulation. The estimates for number of biopsies  $N^{bS}$  are obtained similarly.

### 6.3 Results

The pooled estimates of the various measures of efficacy are summarized in Table 1. In addition mean offset is plotted against mean number of biopsies in Figure 5. From the figure it is evident that there is an inverse relationship between  $E[N^{bS}]$  and  $E[O^S]$ . For example, the annual schedule conducts 5.2 biopsies on average, which is the highest among all schedules, however it has the least average offset of 6 months as well. On the other hand the schedule based on expected time of GR conducts only 1.9 biopsies on average, the least among all schedules, but it also has the highest average offset of 15 months. The schedule based on median time of GR performs similar to that based on expected time of GR. Since the annual schedule attempts to contain the offset within an year it has the least  $\text{var}[O^S]$ . However to achieve so, it conducts a wide range of number of biopsies from patient to patient, i.e. highest  $\text{var}[N^{bS}]$ . Schedules based on expected and median time of GR perform the opposite of annual schedule in terms of variance of  $N^{bS}$  and  $O^S$ .

The PRIAS schedule conducts only 0.4 biopsies less than annual schedule, but with a higher variance of offset, it does not guarantee early detection for everyone. If we compare the PRIAS schedule with dynamic risk of GR based schedule, we can see that the latter performs better than PRIAS schedule in all aspects. The hybrid approach combines the benefits of methods with low  $E[N^{bS}]$  and  $\text{var}[N^{bS}]$ , and methods with

**Table 1**

*Estimated mean and standard deviation of the number of biopsies and offset (months). Method names are abbreviated for consistency with Figure 5.*

a) All subgroups: 101823 patients				
Schedule	$E[N^{bS}]$	$E[O^S]$	$\text{SD}[N^{bS}]$	$\text{SD}[O^S]$
Annual	5.24	6.01	2.53	3.45
PRIAS	4.86	8.49	2.35	8.69
Exp. GR time	1.92	15.06	1.19	12.11
Med. GR time	2.07	13.87	1.42	11.80
Dyn. risk GR	4.69	6.66	2.20	4.37
Mixed	3.75	9.69	1.71	7.02
b) Subgroup $G_1$ : 33680 patients				
Schedule	$E[N^{bS}]$	$E[O^S]$	$\text{SD}[N^{bS}]$	$\text{SD}[O^S]$
Annual	4.33	6.02	3.14	3.44
PRIAS	4.05	7.98	2.87	8.08
Exp. GR time	1.72	21.65	1.47	14.77
Med. GR time	1.85	20.67	1.77	14.64
Dyn. risk GR	3.85	6.76	2.69	4.45
Mixed	3.24	10.24	2.17	7.73
c) Subgroup $G_2$ : 33907 patients				
Schedule	$E[N^{bS}]$	$E[O^S]$	$\text{SD}[N^{bS}]$	$\text{SD}[O^S]$
Annual	5.18	5.99	2.13	3.48
PRIAS	4.82	8.57	1.99	8.65
Exp. GR time	1.78	13.53	0.98	9.82
Med. GR time	1.90	12.31	1.16	9.43
Dyn. risk GR	4.63	6.66	1.82	4.34
Mixed	3.68	10.30	1.38	7.17
d) Subgroup $G_3$ : 34236 patients				
Schedule	$E[N^{bS}]$	$E[O^S]$	$\text{SD}[N^{bS}]$	$\text{SD}[O^S]$
Annual	6.2	6.01	1.77	3.46
PRIAS	5.70	8.92	1.73	9.27
Exp. GR time	2.27	10.11	0.99	7.53
Med. GR time	2.45	8.71	1.15	6.36
Dyn. risk GR	5.57	6.58	1.56	4.32
Mixed	4.31	8.54	1.27	5.91

low  $E[O^S]$  and  $\text{var}[O^S]$ . It conducts 1.5 less biopsies than annual schedule on average and with  $E[O^S]$  at 9.7 months it detects GR within an year since its occurrence. The variance in number of biopsies and offset are not only not high but also lesser than PRIAS schedule.

We next discuss the performance of these schedules for each of the 3 subgroups  $G_1, G_2$  and  $G_3$ . We observe that annual schedule remains the most consistent across subgroups in terms of the offset, but it varies the most in terms of number of biopsies, conducting 2 extra biopsies in subgroup  $G_3$  (highest mean, low variance of GR time) than in  $G_1$  (low mean, high variance of GR time). The performance of schedule based on expected time of GR is the most consistent in terms of number of biopsies but most inconsistent in terms of offset. It performs the best in subgroup  $G_3$ . For the dynamic risk of GR based schedule and the hybrid approach the dynamics are similar to that of the annual schedule. Unlike the latter two schedules,



the PRIAS schedule not only conducts more biopsies in  $G_3$  than  $G_1$  but also has a higher offset for  $G_3$  than  $G_1$ .

The choice of the optimal schedule using (7) depends on the chosen measures of efficacy. For example, schedule based on dynamic risk of GR is the most optimal if on average the least number of biopsies are to be conducted, while simultaneously making sure that at least 90% of the patients have an average offset less than an year. Schedule based on expected time of GR is most optimal if on average the least number of biopsies are to be conducted, while simultaneously making sure that at least 80% of the patients have an average offset less than 2 years. For at least 90% of the patients it detects GR within 3 years with 2 biopsies on average. We however recommend the hybrid approach, since it conducts only 3.8 biopsies on average while guaranteeing an offset of 24 months for 95% of the patients and 36 months for 99.9% of the patients. Besides if further cutoffs are required on variance of number of biopsies or offset they are not too high either for the hybrid approach.

## 7. Discussion

In this paper we presented personalized biopsy schedules for patients enrolled in AS programs. The problem at hand was that low risk prostate cancer patients enrolled in AS have to undergo repeat biopsies on a frequent basis for examination of disease progression, which causes medical side effects and also brings financial burden. Because of these issues, AS programs such as PRIAS were observing a high non-compliance for repeat biopsies. To approach these problems, we first used a joint model for time to event and longitudinal data to combine the information from PSA and repeat biopsy histories of patients in a more sophisticated manner than the existing PRIAS schedule. Using the fitted model, we proposed 2 different classes of personalized schedules for individual patients. They are schedules based on the central tendency of the distribution of time of GR of a patient, and schedules based on dynamic risk of GR. In addition we also proposed a combination (hybrid approach) of these 2 approaches. We then proposed criteria for evaluation of various schedules and a method to select the most optimal schedule.

We demonstrated using PRIAS dataset that the personalized schedules adjust the time of biopsy on the basis of results from historical biopsies and PSA, even when the two are not in concordance with each other. Secondly, we conducted a simulation study to compare various schedules. We observed that personalized schedules based on dynamic risk of GR performed better than PRIAS schedule in terms of both mean and variance of number of biopsies and offset. We also observed that the PRIAS schedule conducted more biopsies and had higher offset for patients having higher mean GR time. We prefer the schedule based on dynamic risk of GR over PRIAS schedule on the basis of these results. The schedules based on expected and median time of GR conducted only 2 biopsies on average, which is very promising compared to PRIAS and annual schedule which conducted 4.9 and 5.2 biopsies on average, respectively. In addition, for the former two schedules, at least 90% of the patients had an offset less than 36 months, which is the maximum possible offset in PRIAS. If a stronger restriction is prescribed for the offset, then we propose that the hybrid approach be used since it

neither conducts too many biopsies nor it has a very high offset.

While each of the personalized methods has their own disadvantages and advantages, they also offer multiple choices to the AS programs to choose one as per their requirements, instead of choosing a common fixed schedule for all patients. In this regard, there is potential to develop more personalized schedules. For example, using loss functions which asymmetrically penalize overshooting/undershooting the target GR time, can be interesting. Depending upon the requirements it is also possible to choose  $\kappa$  on the basis of binary classification accuracy measures which focus on non-cases as well. Although in this work we assumed that GR time was interval censored, in reality the Gleason scores are susceptible to inter-observer variation (Carlson et al., 1998). Models and schedules which account for error in measurement of time of GR, will be interesting to investigate further. Lastly, there is potential for including diagnostic information from Magnetic resonance imaging (MRI) or DRE. Unlike PSA levels such information may not always be continuous in nature, in which case our proposed methodology can be very easily extended by utilizing the framework of generalized linear mixed models.

## ACKNOWLEDGEMENTS

The authors 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.

## SUPPLEMENTARY MATERIALS

Web Appendix A and C, referenced in Section 2 and Section 5 respectively, and the derivation of Equation (6) in Web Appendix B, are available in the document supplementary\_material.pdf.

## REFERENCES

- Akhavan-Tabatabaei, R., Sánchez, D. M., and Yeung, T. G. (2017). A markov decision process model for cervical cancer screening policies in colombia. *Medical Decision Making* **37**, 196–211.
- Ayer, T., Alagoz, O., and Stout, N. K. (2012). A POMDP approach to personalize mammography screening decisions. *Operations Research* **60**, 1019–1034.
- Bebu, I. and Lachin, J. M. (2017). Optimal screening schedules for disease progression with application to diabetic retinopathy. *Biostatistics* doi:10.1093/biostatistics/kxx009.
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer Science & Business Media.
- Bokhorst, L. P., Alberts, A. R., Rannikko, A., Valdagni, R., Pickles, T., Kakehi, Y., Bangma, C. H., Roobol, M. J., study group, P., et al. (2015). Compliance rates with the Prostate Cancer Research International Active Surveillance (PRIAS) protocol and disease reclassification in noncompliers. *European Urology* **68**, 814–821.

- Bokhorst, L. P., Valdagni, R., Rannikko, A., Kakehi, Y., Pickles, T., Bangma, C. H., Roobol, M. J., study group, P., et al. (2016). A decade of active surveillance in the PRIAS study: an update and evaluation of the criteria used to recommend a switch to active treatment. *European Urology* **70**, 954–960.
- Brown, E. R. (2009). Assessing the association between trends in a biomarker and risk of event with an application in pediatric HIV/AIDS. *The Annals of Applied Statistics* **3**, 1163–1182.
- Carlson, G. D., Calvanese, C. B., Kahane, H., and Epstein, J. I. (1998). Accuracy of biopsy Gleason scores from a large uropathology laboratory: use of a diagnostic protocol to minimize observer variability. *Urology* **51**, 525–529.
- Cook, R. D. and Wong, W. K. (1994). On the equivalence of constrained and compound optimal designs. *Journal of the American Statistical Association* **89**, 687–692.
- Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* **11**, 89–121.
- Erenay, F. S., Alagoz, O., and Said, A. (2014). Optimizing colonoscopy screening for colorectal cancer prevention and surveillance. *Manufacturing & Service Operations Management* **16**, 381–400.
- Keegan, K. A., Dall’Era, M. A., Durbin-Johnson, B., and Evans, C. P. (2012). Active surveillance for prostate cancer compared with immediate treatment. *Cancer* **118**, 3512–3518.
- Läuter, E. (1976). Optimal multipurpose designs for regression models. *Mathematische Operationsforschung und Statistik* **7**, 51–68.
- Loeb, S., Vellekoop, A., Ahmed, H. U., Catto, J., Emberton, M., Nam, R., Rosario, D. J., Scattoni, V., and Lotan, Y. (2013). Systematic review of complications of prostate biopsy. *European Urology* **64**, 876–892.
- López-Ratón, M., Rodríguez-Álvarez, M. X., Cadarso-Suárez, C., Gude-Sampedro, F., et al. (2014). OptimalCutpoints: an R package for selecting optimal cutpoints in diagnostic tests. *Journal of Statistical Software* **61**, 1–36.
- Potosky, A. L., Miller, B. A., Albertsen, P. C., and Kramer, B. S. (1995). The role of increasing detection in the rising incidence of prostate cancer. *JAMA* **273**, 548–552.
- Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* **67**, 819–829.
- Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. CRC Press.
- Rizopoulos, D. (2016). The R package JMBayes for fitting joint models for longitudinal and time-to-event data using MCMC. *Journal of Statistical Software* **72**, 1–46.
- Rizopoulos, D., Hatfield, L. A., Carlin, B. P., and Takkenberg, J. J. (2014). Combining Dynamic Predictions From Joint Models for Longitudinal and Time-to-Event Data Using Bayesian Model Averaging. *Journal of the American Statistical Association* **109**, 1385–1397.
- Rizopoulos, D., Molenberghs, G., and Lesaffre, E. M. (2017). Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biometrical Journal* doi:10.1002/bimj.201600238.
- Rizopoulos, D., Taylor, J. M. G., Van Rosmalen, J., Steyerberg, E. W., and Takkenberg, J. J. M. (2016). Personalized screening intervals for biomarkers using joint models for longitudinal and survival data. *Biostatistics* **17**, 149–164.
- Robert, C. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media.
- Roobol, M. J. and Bangma, C. H. (2014). PRIAS. <https://www.prias-project.org/>. [Online; accessed 18-August-2014].
- Taylor, J. M., Park, Y., Ankerst, D. P., Proust-Lima, C., Williams, S., Kestin, L., Bae, K., Pickles, T., and Sandler, H. (2013). Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics* **69**, 206–213.
- Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., and Jemal, A. (2015). Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians* **65**, 87–108.
- Tosoian, J. J., Trock, B. J., Landis, P., Feng, Z., Epstein, J. I., Partin, A. W., Walsh, P. C., and Carter, H. B. (2011). Active surveillance program for prostate cancer: an update of the Johns Hopkins experience. *Journal of Clinical Oncology* **29**, 2185–2190.
- Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* **14**, 809–834.
- Welty, C. J., Cowan, J. E., Nguyen, H., Shinohara, K., Perez, N., Greene, K. L., Chan, J. M., Meng, M. V., Simko, J. P., Cooperberg, M. R., et al. (2015). Extended followup and risk factors for disease reclassification in a large active surveillance cohort for localized prostate cancer. *The Journal of Urology* **193**, 807–811.
- Zhang, J., Denton, B. T., Balasubramanian, H., Shah, N. D., and Inman, B. A. (2012). Optimization of prostate biopsy referral decisions. *Manufacturing & Service Operations Management* **14**, 529–547.

Received October 0000. Revised February 0000. Accepted March 0000.