Biometrics 00, 1–24
December 0000

DOI: 10.1111/j.1541-0420.2005.00454.x

Personalized Schedules for Surveillance of Cancer Patients

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

[1]Department of Biostatistics, Erasmus University Medical Center, the Netherlands
[2]Department of Public Health, Erasmus University Medical Center, the Netherlands
[3]Department of Urology, Erasmus University Medical Center, the Netherlands
email: atomer@erasmusmc.nl

Summary: Cancer patients enrolled in active surveillance (AS) programs commonly undergo biopsies on a frequent basis for examination of cancer progression. AS programs employ a fixed schedule of biopsies for all patients. Such fixed and frequent schedules, schedule unnecessary biopsies for patients with slowly progressing cancer. Since biopsies have an associated risk of complications, such patients do not always comply with the schedule, which increases the risk of delayed detection of cancer progression. Motivated by the world's largest AS program, Prostate Cancer Research International Active Surveillance (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 historical biomarker levels and repeat biopsy results of a patient, to schedule the next biopsy. We also present methods to compare personalized schedules with existing biopsy schedules.

Key words: Active surveillance; Biopsy; Cancer; Joint models; Personalized medicine

1. Introduction

Cancer screening is a widely used practice to detect cancer before symptoms appear in otherwise healthy individuals. A major issue of screening programs that has been observed in many types of cancers is overdiagnosis (Esserman et al., 2014). To avoid subsequent overtreatment, patients diagnosed with low grade cancers are commonly advised to join active surveillance (AS) programs. The goal of AS is to routinely examine the progression of cancer and avoid serious treatments such as surgery, chemotherapy, or radiotherapy as long as they are not needed. To this end, AS includes, but is not limited to periodical evaluation of biomarkers pertaining to the cancer, physical examination, medical imaging, and biopsy.

The switch from AS to active treatment in many types of cancers (e.g., prostate cancer) is not done without consulting the biopsy results (Bokhorst et al., 2016). Biopsies can be reliable, but they are also painful, cause anxiety, and have an associated risk of complications such as inflammation, hematuria and sepsis (Loeb et al., 2013; Mueller et al., 1988). Because of this reason the schedule of biopsies has significant medical consequences for patients. A frequent schedule of biopsies may help detecting cancer earlier but the corresponding risk of complications will be high. And although such a schedule may work well for faster progressing cancer patients, but for slowly progressing cancer patients many unnecessary biopsies will be scheduled. Furthermore, patients do not always comply with such a schedule, as has been observed by Bokhorst et al. (2015). In the specific case of the world's largest AS program, Prostate Cancer Research International Active Surveillance (PRIAS) (Bokhorst et al., 2016), the compliance rate for biopsies steadily decreased from 81% at year one of follow up, to 60% at year four, 53% at year seven and 33% at year ten. Such non-compliance can lead to delayed detection of cancer progression, which reduces the effectiveness of AS programs.

This paper is motivated by the need to reduce the medical burden of biopsies while simultaneously avoiding late detection of cancer progression. In particular, we are motivated by the case of prostate cancer (PCa) patients, for whom several AS programs schedule biopsies with a gap of only one year (we refer to it as annual schedule hereafter) (Tosoian et al., 2011; Welty et al., 2015). Most AS programs for PCa strongly advise against scheduling biopsies more frequently than the annual schedule. The PRIAS schedule for biopsies is relatively lenient: one biopsy each is scheduled at year one of follow up, year four, year seven, year ten, and every five years thereafter. We intend to improve upon such fixed schedules by creating personalized schedules for biopsies. That is, a different schedule for every patient utilizing their periodically measured diagnostic information. In context of PCa, this information is serum prostate-specific antigen (PSA) levels (measured in ng/mL) and repeat biopsy results. Biopsies are graded using the Gleason score, which takes a value between 2 and 10, with 10 corresponding to the most serious state of the cancer. Patients enter AS only if their Gleason score is 6 or less. When a patient's Gleason score becomes greater than 6, also known as Gleason reclassification (referred to as GR hereafter), the patient switches from AS to active treatment. Hence, for AS programs it is of prime interest to detect GR soonest with least number of biopsies possible.

Personalized schedules for screening have received much

interest in the literature, especially in the medical decision making context. For diabetic retinopathy, cost optimized personalized schedules based on Markov models have been developed by Bebu and Lachin (2017). 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 the measurement of longitudinal biomarkers (Rizopoulos et al., 2016). In context of PCa, 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 the baseline characteristics as well as a discretized PSA level of the patient at the current check up time.

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 pre-scheduled check up time, we schedule biopsies at a per patient optimal future 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 repeat biopsies up to the current check up time. 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. We also present methods to evaluate and compare the various schedules for biopsies.

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 methods for evaluation and selection of a 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 S be the schedule of biopsies prescribed to this patient. The corresponding vector of time of biopsies is denoted by $T_i^S = \{T_{i0}^S, T_{i1}^S, \ldots, T_{iN_i^S}^S; T_{ij}^S < T_{ik}^S, \forall j < k\}$, where N_i^S 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 $l_i = T_{iN_i^S-1}^S, r_i = T_{iN_i^S}^S$ if GR is observed, and $l_i = T_{iN_i^S}^S, r_i = \infty$ if GR is not observed yet. Further let $\boldsymbol{y}_i$ 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 $\mathcal{D}_n = \{l_i, r_i, \boldsymbol{y}_i; i = 1, \ldots, n\}$.

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:

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

where $\boldsymbol{x}_i(t)$ denotes the row vector of the design matrix for fixed effects and $\boldsymbol{z}_i(t)$ denotes the same for random effects. Correspondingly the fixed effects are denoted by $\boldsymbol{\beta}$ and random effects by $\boldsymbol{b}_i$. The random effects are assumed to be normally distributed with mean zero and $q \times q$ covariance matrix $\boldsymbol{D}$. 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 σ^2, and is independent of the random effects $\boldsymbol{b}_i$.

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:

$$\begin{aligned} h_i(t \mid \mathcal{M}_i(t), \boldsymbol{w}_i) &= \lim_{\Delta t \to 0} \frac{\Pr\{T_i^* \epsilon [t, t+\Delta t) \mid T_i^* \geqslant t, \mathcal{M}_i(t), \boldsymbol{w}_i\}}{\Delta t} \\ &= h_0(t) \exp\left[\boldsymbol{\gamma}^T \boldsymbol{w}_i + f\{M_i(t), \boldsymbol{b}_i, \boldsymbol{\alpha}\}\right], \quad t > 0, \end{aligned}$$

where $\mathcal{M}_i(t) = \{m_i(v), 0 \leqslant v \leqslant t\}$ denotes the history of the underlying PSA levels up to time t. The vector of baseline covariates is denoted by $\boldsymbol{w}_i$, 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:

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

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

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

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).

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. Let us assume that a personalized schedule is to be created for a new patient j, who is not present in the original sample $\mathcal{D}_n$ of patients. 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 $u > \max(t, s)$ 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 $\boldsymbol{w}_i$ is dropped for notational simplicity hereafter):

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

The distribution $g(T_j^*)$ depends on the observed longitudinal history of patient j via the random effects $\boldsymbol{b_j}$, and on the information from the original dataset $\mathcal{D}_n$ via the posterior distribution of the parameters $p(\boldsymbol{\theta} \mid \mathcal{D}_n)$, 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 $E_g\big\{L(T_j^*, u)\big\}$, where the expectation is taken with respect to $g(T_j^*)$. The former is given by:

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

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.

Given the medical burden of biopsies, ideally 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 the absolute loss function $L(T_j^*, u) = \big|T_j^* - u\big|$ 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:

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

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:

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

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

$$\pi_j(u \mid t, s) = \Pr\big\{T_j^* \geqslant u \mid T_j^* > t, \mathcal{Y}_j(s), D_n\big\}, \quad u \geqslant t. \tag{4}$$

For ease of readability we denote $\pi_j^{-1}(0.5 \mid t, s)$ 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 $1 - \kappa, \kappa \epsilon [0, 1]$ on the risk of GR. This approach would be successful if κ 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:

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

where k_1, k_2 are constants parameterizing the loss function. The posterior expected loss $E_g\big\{L_{k_1,k_2}(T_j^*, u)\big\}$ obtains its minimum at $u = \pi_j^{-1}\big\{k_1/(k_1 + k_2) \mid t, s\big\}$ (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 PRIAS, within the first 10 years, 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 $\pi_j(u \mid t, s)$:

$$E_g(T_j^*) = t + \int_t^\infty \pi_j(u \mid t, s)\mathrm{d}u. \tag{6}$$

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 $p(T_j^* \mid T_j^* > t, \boldsymbol{b_j}, \boldsymbol{\theta})$, as compared to the estimation of $\pi_j(u \mid t, s)$ (Rizopoulos, 2011). The former approach was thus computationally faster.

As mentioned earlier, selection of the optimal biopsy time based on $E_g(T_j^*)$ alone will not be practically useful when the $\mathrm{var}_g(T_j^*)$ is large, which is given by:

$$\mathrm{var}_g(T_j^*) = 2\int_t^\infty (u-t)\pi_j(u \mid t, s)\mathrm{d}u - \Big\{\int_t^\infty \pi_j(u \mid t, s)\mathrm{d}u\Big\}^2. \tag{7}$$

Since a closed form solution is not available for the variance expression, it is estimated similar to the estimation of $E_g(T_j^*)$. The variance depends both on last biopsy time t and PSA history $\mathcal{Y}_j(s)$. The impact of the observed information on variance is demonstrated in Section 5.2.

For schedules based on dynamic risk of GR, the value of κ 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 5%, then $\kappa = 0.95$.

In cases where κ cannot be chosen on the basis of the input of the physician or the patients, we propose to automate the choice of κ. More specifically, we propose to choose a threshold κ 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 $\pi_j(t + \Delta t \mid t, s) \leqslant \kappa$ and a control if $\pi_j(t + \Delta t \mid t, s) > \kappa$ (Rizopoulos, 2016; Rizopoulos, Molenberghs, and Lesaffre, 2017). The time window Δt can be automatically chosen as $\arg\max_{\Delta t} \mathrm{AUC}(t, \Delta t, s)$, 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 one year from those who do not, so that they can be provided immediate attention (only in exceptional cases a biopsy within an year of the last one). Thus, in this work we use a Δt of one year. As for the choice of the binary classification accuracy measure, we require a measure which is in line with the goal to focus on patients whose true time of GR falls in the time window Δt. To this end, the measure which combines both sensitivity and precision is F_1 score. It is defined as:

$$\mathrm{F}_1(t, \Delta t, s) = 2\frac{\mathrm{TPR}(t, \Delta t, s)\mathrm{PPV}(t, \Delta t, s)}{\mathrm{TPR}(t, \Delta t, s) + \mathrm{PPV}(t, \Delta t, s)}, \quad \mathrm{F}_1 \epsilon [0, 1],$$
$$\mathrm{TPR}(t, \Delta t, s) = \Pr\big\{\pi_j(t + \Delta t \mid t, s) \leqslant \kappa \mid T_j^* \epsilon (t, t + \Delta t]\big\},$$
$$\mathrm{PPV}(t, \Delta t, s) = \Pr\big\{T_j^* \epsilon (t, t + \Delta t] \mid \pi_j(t + \Delta t \mid t, s) \leqslant \kappa\big\}.$$

where $\mathrm{TPR}(\cdot)$ and $\mathrm{PPV}(\cdot)$ denote time dependent true positive rate (sensitivity) and positive predictive value (precision), respectively. The estimation for both is similar to the estimation of $\mathrm{AUC}(t, \Delta t, s)$ given by Rizopoulos et al. (2017). Since a high F_1 score is desired, the optimal value of κ is $\arg\max_\kappa \mathrm{F}_1(t, \Delta t, s)$. In this work we compute the latter using a grid search approach. That is, first F_1 is computed using the available dataset over a fine grid of κ values in the interval $[0, 1]$, and then κ 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 $u^{new} > \max(u, s)$. This process is iteratively repeated until GR is detected. To aid in medical decision making, we elucidate this process via an algorithm in Figure 1. Since it is medically required 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 schedule and to compare it with other schedules. To this end, we first present the methods to evaluate the biopsy schedules and then discuss the choice of a schedule.

4.1 *Evaluation of Schedules*

We evaluate a schedule S using two criteria, namely the number of biopsies $N_j^S \epsilon [1, \infty]$ a schedule conducts for the j-th patient to detect GR, and the offset $O_j^S \epsilon [0, \infty]$ by which it overshoots the true GR time T_j^*. The offset O_j^S is defined as $O_j^S = T_{jN_j^S}^S - T_j^*$, where $T_{jN_j^S}^S \geqslant T_j^*$ is the time at which GR is detected. Our interest lies in the joint distribution $p(N_j^S, O_j^S)$ of the number of biopsies and the offset. Given the medical burden of biopsies, ideally only one biopsy which leads to a zero offset should be conducted. That is, a schedule with a low mean number of biopsies $E(N_j^S)$ as well a low mean offset $E(O_j^S)$ is desired. It is also desired that a schedule has low variance of the number of biopsies $\mathrm{var}(N_j^S)$, as well as low variance of the offset $\mathrm{var}(O_j^S)$, so that the schedule works similarly for most patients. Quantiles of $p(N_j^S)$ may also be of interest. For example, a schedule which conducts less than two biopsies to detect GR in 95% of the patients may be preferred.

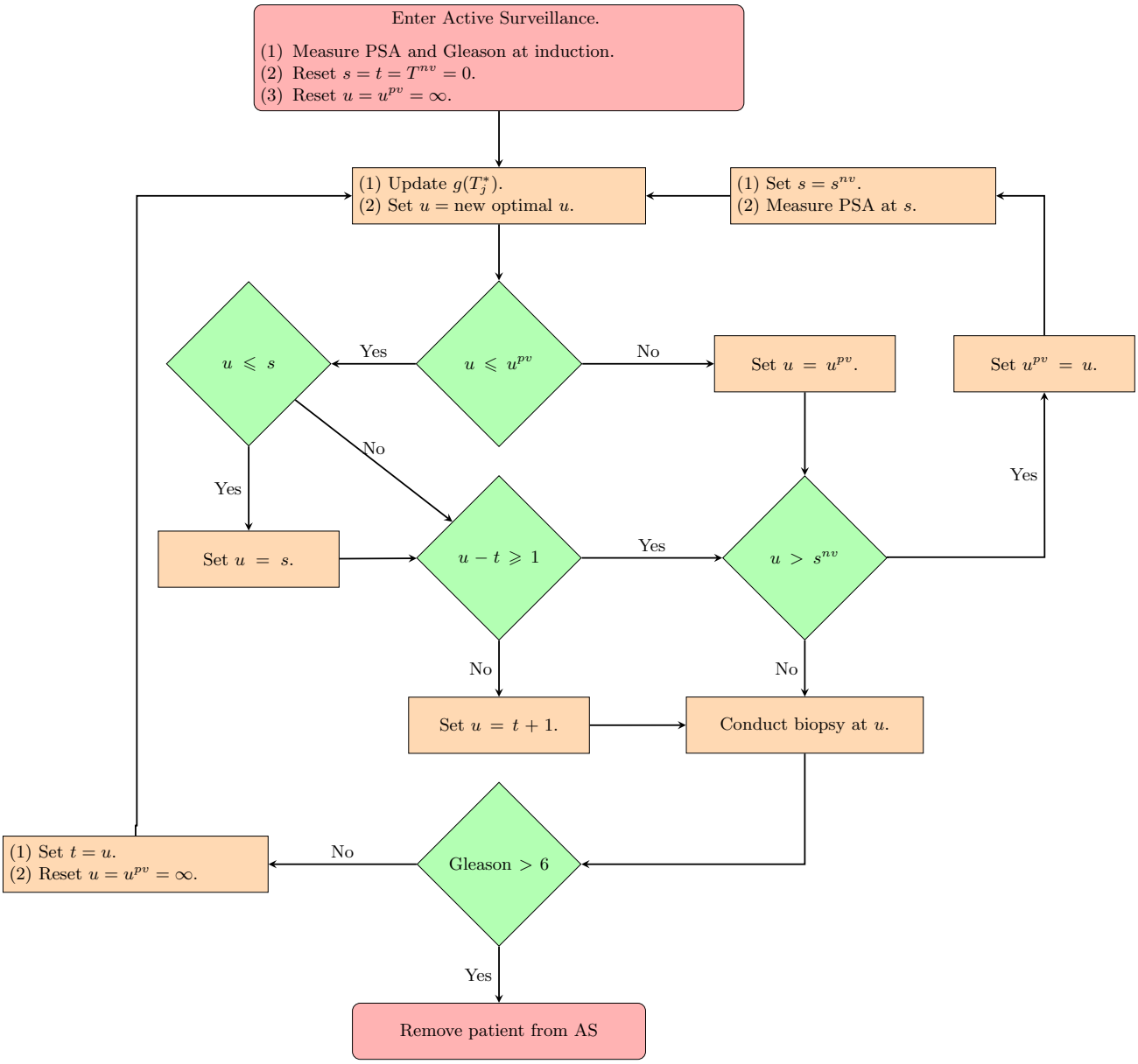

Figure 1. Algorithm for creating a personalized schedule for patient j. t denotes the time of the latest biopsy. s denotes the time of the latest available PSA measurement. u denotes the proposed personalized time of biopsy. u^{pv} denotes the time at which a repeat biopsy was proposed on the last visit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

4.2 *Finding a Suitable Schedule*

Given the multiple criteria for evaluation of a schedule, the next step is to use them to select a schedule. Using principles from compound optimal designs (Läuter, 1976) we propose to choose a schedule S which minimizes a loss function of the following form:

$$L(S) = \sum_{r=1}^{R} \eta_r \mathcal{R}_r(N_j^S). \tag{8}$$

where $\mathcal{R}_r(\cdot)$ is an evaluation criteria based on either the number of biopsies or the offset (for brevity of notation, only N_j^S is used in the equation above). Some examples of $\mathcal{R}_r(\cdot)$ are mean, median, variance and quantile function. Constants $\eta_1, \ldots, \eta_R$, where $\eta_r \epsilon [0, 1]$ and $\sum_{r=1}^{R} \eta_r = 1$, are weights to differentially weigh-in the contribution of each of the R criteria. An example loss function is:

$$L(S) = \eta_1 E(N_j^S) + \eta_2 E(O_j^S). \tag{9}$$

The choice of η_1 and η_2 is not easy, because biopsies have associated medical risks 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 η_1 and η_2 there exists a constant $C > 0$ for which mini-

mization of loss function in (9) is equivalent to minimization of the loss function subject to the constraint that $E(O_j^S) < C$. That is, a suitable schedule is the one which conducts the least number of biopsies to detect GR while simultaneously guaranteeing 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 (8), a schedule can be chosen by minimizing $\mathcal{R}_R(\cdot)$ under the constraint $\mathcal{R}_r(\cdot) < C_r; r = 1, \ldots, R-1$.

5. Demonstration of Personalized Schedules

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 three 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 three months for the first two years and every six 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 (Nieboer et al., 2015). The longitudinal sub-model of the joint model we fit is given by:

$$\log_2 \text{PSA}(t) = \beta_0 + \beta_1(Age - 70) + \beta_2(Age - 70)^2 + \sum_{k=1}^{4} \beta_{k+2} B_k(t, \mathcal{K}) + b_{i0} + b_{i1} B_7(t, 0.1) + b_{i2} B_8(t, 0.1) + \varepsilon_i(t). \tag{10}$$

where $B_k(t, \mathcal{K})$ denotes the k-th basis function of a B-spline with three internal knots at $\mathcal{K} = \{0.1, 0.5, 4\}$ years, and boundary knots at zero and seven years. The spline for the random effects consists of one internal knot at 0.1 years and boundary knots at zero and seven 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:

$$h_i(t) = h_0(t) \exp\left\{\gamma_1(Age-70) + \gamma_2(Age-70)^2 + \alpha_1 m_i(t) + \alpha_2 m_i'(t)\right\}. \tag{11}$$

where α_1 and α_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.

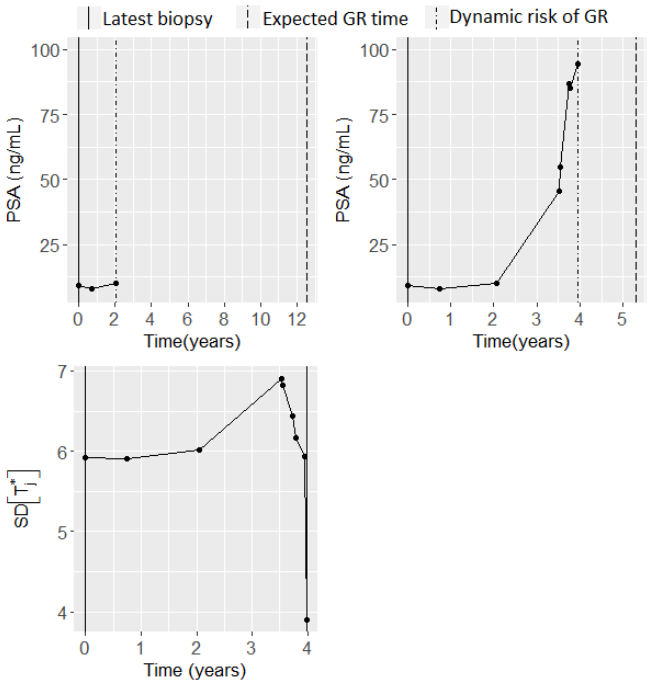

Figure 2. Top panel: Evolution of PSA, history of repeat biopsies and corresponding personalized schedules for patient 3174. Bottom Panel: History of repeat biopsies and $\text{SD}_g(T_j^*) = \sqrt{\text{var}_g(T_j^*)}$ over time for patient 3174.

5.2 *Demonstration for Patient 3174*

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. The evolution of PSA, repeat biopsy history and proposed times of biopsies for this patient are shown in the top panel of 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 to steep rise in $\log_2$ PSA velocity. More specifically, at year two the proposed biopsy time is 12.5 years whereas at year four it decreases to 5.3 years. On average, a biopsy scheduled using expected time of GR at year two should have a larger offset O_j^S compared to the same at year four. This is because the standard deviation of $g(T_j^*)$, given by $\text{SD}_g(T_j^*) = \sqrt{\text{var}_g(T_j^*)}$, is considerably lower at year four as shown in the bottom panel of Figure 2. In the figure it can be seen that the standard deviation also strongly depends on $\log_2$ PSA velocity. As for the schedules based on dynamic risk of GR, the threshold κ was automatically chosen using F_1 score, and was estimated to be between 1 and 0.9 at all time points. This value of κ corresponds to a time very close to the time of latest biopsy ($t = 0$). Hence the biopsies are scheduled much earlier than those based on expected time of GR.

The demonstration of personalized schedules for the two

other patients from the demonstration data set is presented in Web Appendix D of the supplementary material.

6. Simulation Study

The application of personalized schedules for patients from PRIAS demonstrated that these schedules adapt according to the historical data of each patient. However we could not perform a full scale comparison between personalized and PRIAS schedule, because the true time of GR was not known for the PRIAS patients. To this end, we conducted a simulation study comparing personalized schedules with PRIAS and annual schedule, whose details are presented next.

6.1 *Simulation Setup*

First we assume a population of patients enrolled in AS, with the same entrance criteria as that of PRIAS. The PSA and hazard of GR for patients from this population follow a joint model of the form postulated in Section 5.1, with parameters equal to the posterior mean of parameters estimated from the joint model fitted to PRIAS dataset (Web Appendix C of the supplementary material). We further assume that there are three equal sized subgroups G_1, G_2 and G_3 of patients in the population, differing in the baseline hazard of GR. This was done because we wanted to test the performance of different schedules for a population with a mixture of patients, namely those with faster progressing PCa, as well as those with slowly progressing PCa. For the three subgroups we use a Weibull distributed baseline hazard with the following shape and scale parameters (k, λ): $(1.5, 4)$, $(3, 5)$ and $(4.5, 6)$ for G_1, G_2 and G_3, respectively. The effect of these parameters is that the variance of time of GR is highest for G_1 and lowest for G_3, while the mean GR time is lowest in G_1 (faster progressing PCa) and highest in G_3 (slowly progressing PCa).

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 $\mathcal{D}^k = \{l_{ki}, r_{ki}, \boldsymbol{y}_{ki}; i = 1, \ldots, 750\}$, where $\boldsymbol{y}_{ki}$ 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.

We next fit a joint model of the specification given in (10) and (11) to each of the $\mathcal{D}^k, k = 1, \ldots, 500$, and obtain a MCMC sample from the posterior distribution $p(\boldsymbol{\theta} \mid \mathcal{D}^k)$. We then obtain $g(T^*_{kl})$ for each of the l-th test patient of the k-th data set and conduct hypothetical biopsies for him. For every patient we conduct biopsies using the following six types of schedules (abbreviated names in parenthesis): personalized schedules based on expected time of GR (Exp. GR time) and median time of GR (Med. GR time), personalized schedules based on dynamic risk of GR (Dyn. risk GR), a hybrid approach between median time of GR and dynamic risk of GR (Hybrid), PRIAS schedule and annual schedule. The biopsies are conducted iteratively in accordance with the algorithm in Figure 1.

To compare the aforementioned schedules we require estimates of the various criteria based on offset and number of biopsies conducted to detect GR (Section 4). To this end, we compute pooled estimates of each of the $E(N_j^S)$, $\text{var}(N_j^S)$, $E(O_j^S)$ and $\text{var}(O_j^S)$, as below:

$$\widehat{E(O_j^S)} = \frac{\sum_{k=1}^{500} n_k \widehat{E(O_k^S)}}{\sum_{k=1}^{500} n_k},$$

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

where n_k denotes the number of test patients, $\widehat{E(O_k^S)} = \sum_{l=1}^{n_k} O_{kl}^S / n_k$ is the estimated mean and $\widehat{\text{var}(O_k^S)} = \sum_{l=1}^{n_k} \left\{O_{kl}^S - \widehat{E(O_k^S)}\right\}^2 / (n_k - 1)$ is the estimated variance of the offset for the k-th simulation. The estimates for number of biopsies are obtained similarly.

6.2 *Results*

The pooled estimates of the aforementioned criteria are summarized in Table 1. In addition, mean offset is plotted against mean number of biopsies conducted to detect GR in Figure 3. From the figure it is evident that across the schedules there is an inverse relationship between $E(N_j^S)$ and $E(O_j^S)$. For example, the annual schedule conducts 5.2 biopsies on average to detect GR, 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 to detect GR, 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 $\text{SD}(O_j^S) = \sqrt{\text{var}(O_j^S)}$. However to achieve so, it conducts a wide range of number of biopsies from patient to patient, i.e., highest $\text{SD}(N_j^S) = \sqrt{\text{var}(N_j^S)}$. Schedules based on expected and median time of GR perform the opposite of annual schedule in terms of $\text{SD}(N_j^S)$ and $\text{SD}(O_j^S)$.

The PRIAS schedule conducts only 0.3 biopsies less than the 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 four criteria. The hybrid approach combines the benefits of methods with low $E(N_j^S)$ and $\text{SD}(N_j^S)$, and methods with low $E(O_j^S)$ and $\text{SD}(O_j^S)$. It conducts 1.5 less biopsies than the annual schedule on average and with a $E(O_j^S)$ of 9.7 months it detects GR within an year since its occurrence. Moreover, it has both $\text{SD}(N_j^S)$ and $\text{SD}(O_j^S)$ comparable to PRIAS.

The performance of each schedule differs for the three subgroups G_1, G_2 and G_3. The annual schedule remains the most consistent across subgroups in terms of the offset, but it conducts 2 extra biopsies for subgroup G_3 (slowly progressing PCa) than G_1 (faster progressing PCa). The performance of schedule based on expected time of GR is the most consistent

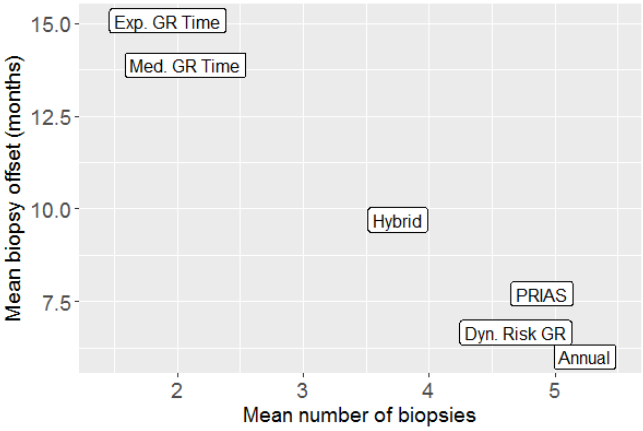

Figure 3. Estimated mean number of biopsies and mean offset (months) for the 6 schedules, using all 124781 patients.

Table 1
Estimated mean and standard deviation of the number of biopsies and offset (months).

a) All subgroups: 124781 patients				
Schedule	$E(N_j^S)$	$E(O_j^S)$	SD$[N_j^S]$	SD$[O_j^S]$
Annual	5.24	6.01	2.53	3.46
PRIAS	4.90	7.71	2.36	6.31
Exp. GR time	1.92	15.08	1.19	12.11
Med. GR time	2.06	13.88	1.41	11.80
Dyn. risk GR	4.69	6.66	2.19	4.38
Hybrid	3.75	9.70	1.71	7.25
b) Subgroup G_1: 41484 patients				
Schedule	$E(N_j^S)$	$E(O_j^S)$	SD$[N_j^S]$	SD$[O_j^S]$
Annual	4.32	6.02	3.13	3.44
PRIAS	4.07	7.44	2.88	6.11
Exp. GR time	1.72	21.65	1.47	14.75
Med. GR time	1.84	20.66	1.76	14.62
Dyn. risk GR	3.85	6.75	2.69	4.44
Hybrid	3.25	10.25	2.16	8.07
c) Subgroup G_2: 41423 patients				
Schedule	$E(N_j^S)$	$E(O_j^S)$	SD$[N_j^S]$	SD$[O_j^S]$
Annual	5.18	5.98	2.13	3.47
PRIAS	4.85	7.70	2.00	6.29
Exp. GR time	1.77	13.54	0.98	9.83
Med. GR time	1.89	12.33	1.16	9.44
Dyn. risk GR	4.63	6.66	1.82	4.37
Hybrid	3.68	10.32	1.37	7.45
d) Subgroup G_3: 41874 patients				
Schedule	$E(N_j^S)$	$E(O_j^S)$	SD$[N_j^S]$	SD$[O_j^S]$
Annual	6.20	6.02	1.76	3.46
PRIAS	5.76	7.98	1.71	6.51
Exp. GR time	2.27	10.09	0.99	7.47
Med. GR time	2.45	8.70	1.15	6.32
Dyn. risk GR	5.58	6.58	1.56	4.33
Hybrid	4.32	8.55	1.26	5.91

in terms of number of biopsies but it detects GR an year later on average in subgroup G_1 than G_3. For the dynamic risk of GR based schedule and the hybrid schedule 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 detects GR later in G_3 than G_1.

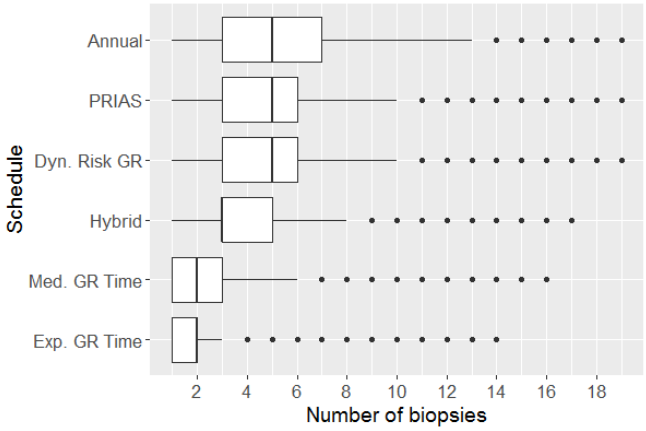

Figure 4. Boxplot showing variation in number of biopsies conducted by different methods, using all simulated patients.

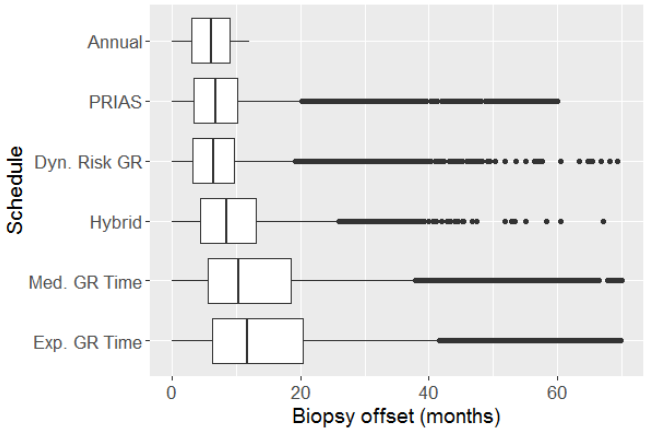

Figure 5. Boxplot showing variation in biopsy offset (months) for different methods, using all simulated patients. X-axis is trimmed at 70 months for visual clarity.

The choice of a suitable schedule using (8) depends on the chosen criteria for evaluation of schedules. For example, the schedule based on dynamic risk of GR is suitable if on average the least number of biopsies are to be conducted to detect GR, while simultaneously making sure that at least 90% of the patients have an average offset less than one year (Figure 4 and 5). The schedule based on expected time of GR is suitable if on average the least number of biopsies are to be conducted to detect GR, while simultaneously making sure that at least 90% of the patients have an average offset less than three years. If a stricter cutoff is required on offset the hybrid approach may be suitable, since it conducts only

3.8 biopsies on average while guaranteeing an offset of two years for 95% of the patients and three years 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 schedules for surveillance of cancer patients. The problem at hand was the following: Low risk cancer patients enrolled in AS have to undergo repeat biopsies on a frequent basis for examination of cancer progression, and since biopsies have an associated risk of complications, not all patients comply with the schedule of biopsies. This reduces the effectiveness of AS programs because cancer progression is detected late. To approach these problems we proposed personalized schedules based on joint models for time to event and longitudinal data. At any given point in time, the proposed personalized schedules utilize a patient's information from historical biomarker measurements (such as PSA in prostate cancer) and repeat biopsies conducted up to that time. We proposed two different classes of personalized schedules for individual patients. They are schedules based on expected and median 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 two approaches, which is useful in scenarios where variance of time of GR for a patient is high. We then proposed criteria for evaluation of various schedules and a method to select a suitable schedule.

We demonstrated using the PRIAS dataset that the personalized schedules adjust the time of biopsy on the basis of results from historical PSA measurements and repeat biopsies, even when the two are not in concordance with each other (Web Appendix D). Secondly, we conducted a simulation study to compare various schedules. We observed that the schedules based on expected and median time of GR conduct only two biopsies on average to detect GR, which is promising compared to PRIAS (4.9 biopsies) and annual schedule (5.2 biopsies). We also observed that the performance of the schedules depends on the true GR time of the patient. For example, in simulated patients who have a slowly progressing PCa (subgroup G_3), personalized schedule based on expected time of GR detects GR one year earlier on average compared to patients who have a faster progressing PCa (subgroup G_1), while conducting approximately the same number of biopsies for both subgroups. For subgroup G_1, the annual or PRIAS schedule may be preferred because they detect GR at 6 and 7.4 months since its occurrence, respectively. However for slowly progressing PCa patients they conduct up to 6 biopsies to detect GR at 6 and 8 months, respectively. In such scenarios, that is, where it is not known in advance if the patient will have a faster or slower progression of PCa, the hybrid approach provides an interesting alternative. This because, it conducts 1 less biopsy than annual schedule in faster progressing PCa patients while detecting GR at 10.3 months since its occurrence on average. Whereas, for slowly progressing PCa patients it conducts 2 less biopsies than the annual schedule while detecting GR at 8.6 months since its occurrence on average.

While each of the personalized schedules have their own advantages and disadvantages, 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, we proposed to choose schedules which conduct as less biopsies as possible while restricting the offset to be below a AS program specific threshold, or vice versa. There is also potential to develop more personalized schedules. For example, using loss functions which asymmetrically penalize overshooting/undershooting the target GR time can be interesting. Furthermore, for dynamic risk of GR based schedules, it is also possible to choose κ using other binary classification accuracy measures or as per patient's wish (Web Appendix E). Although in this work we assumed that the time of GR 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 extended by utilizing the framework of generalized linear mixed models.

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 cleaning the PRIAS data set.

Supplementary Materials

Web Appendix A, C, D and E referenced in Section 2, Section 5, and Section 7, respectively, and the derivation of Equation (6) and (7) 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 PRIAS, 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 PRIAS, 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.

Esserman, L. J., Thompson, I. M., Reid, B., Nelson, P., Ransohoff, D. F., Welch, H. G., Hwang, S., Berry, D. A., Kinzler, K. W., Black, W. C., et al. (2014). Addressing overdiagnosis and overtreatment in cancer: a prescription for change. *The lancet oncology* **15,** e234–e242.

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.

Mueller, P., Miketic, L., Simeone, J., Silverman, S., Saini, S., Wittenberg, J., Hahn, P., Steiner, E., and Forman, B. (1988). Severe acute pancreatitis after percutaneous biopsy of the pancreas. *American Journal of Roentgenology* **151,** 493–494.

Nieboer, D., Vergouwe, Y., Roobol, M. J., Ankerst, D. P., Kattan, M. W., Vickers, A. J., Steyerberg, E. W., and Group, P. B. C. (2015). Nonlinear modeling was applied thoughtfully for risk prediction: the Prostate Biopsy Collaborative Group. *Journal of clinical epidemiology* **68,** 426–434.

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.

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.

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.