PRIAS Personalized Biopsy Schedules

Firstname1 Lastname1

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Suspendisse accumsan magna est, quis elementum leo laoreet eu. Donec sollicitudin elit non massa venenatis, in viverra dolor sagittis. Maecenas ac justo pulvinar, consectetur mauris hendrerit, vulputate lacus. Etiam tristique sapien quis sem commodo, et eleifend tortor viverra. In hac habitasse platea dictumst. Phasellus vel tempus risus, sit amet consectetur massa. Duis rutrum lectus eu ligula egestas iaculis. Sed condimentum, ipsum in dignissim condimentum, nisi turpis blandit massa, et aliquam magna ligula eget lacus. Donec ac eleifend nulla, quis cursus nisi. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Suspendisse accumsan magna est, quis elementum leo laoreet eu. Donec sollicitudin elit non massa venenatis, in viverra dolor sagittis. Maecenas ac justo pulvinar, consectetur mauris hendrerit, vulputate lacus. Etiam tristique sapien quis sem commodo, et eleifend tortor viverra. In hac habitasse platea dictumst. Phasellus vel tempus risus, sit amet consectetur massa. Duis rutrum lectus eu ligula egestas iaculis. We demonstrate that personalized schedules are a very promising method to decide biopsy times for prostate cancer patients.

1 Introduction

Prostate cancer is the development of cancer in the prostate gland. With increase in life expectancy and increase in number of screening tests, an increase in diagnosis of low grade prostate cancers has been observed. Majority of these cancers have good long-term survival and in many cases the prostate cancer is (over) diagnosed solely due of screening. i.e. it wouldn't have shown any malignant symptoms for a long time otherwise. To avoid overtreatment, patients diagnosed with prostate cancer are often motivated to join active surveillance (AS) programs instead of taking immediate treatment. The goal of AS programs is to routinely check 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 PRIAS (`www.prias-project.org`) (Bokhorst et al., 2015). Patients enrolled in PRIAS are closely monitored using serum prostate-specific antigen (PSA) levels, digital rectal examination (DRE) and repeat prostate biopsies. Biopsies are evaluated using the Gleason grading system. Gleason scores range between 2 and 10, with 10 corresponding to a very serious state of prostate cancer. Patients who join PRIAS have a Gleason score of 6 or less, DRE score of cT2c or less and a PSA of 10 ng/mL or less at the time of induction. Although a PSA doubling time(measured as the inverse of the slope of regression line through the base 2 logarithm of PSA values) of less than 3 years, DRE of cT3 or more, and a Gleason score more than 6 are indicators of prostate cancer progression, only DRE and Gleason scores are considered to be the gold standard in this regard (Bokhorst et al., 2016). If either the DRE or the Gleason score are found to be above the aforementioned threshold, then it is considered that the disease has progressed and the patient is removed from AS for further curative treatment. When the Gleason score becomes greater than 6, it is also known as Gleason reclassification (referred to as GR here onwards).

The reliability of Gleason score comes at a high cost. Biopsies are difficult to obtain, are painful and have serious side effects such as hematuria and sepsis for prostate cancer patients (Loeb et al., 2013). So much so, that PRIAS as well as majority of the AS programs around the world strongly adhere to the rule of not having more than 1 biopsy per year. Performing a biopsy every year has the advantage that it is possible to detect GR within 1 year since its occurrence. The drawbacks of this schedule though, are not only medical but also financial. Keegan et al., 2012 have shown that if a biopsy is performed every year then the costs of AS per head, at 10 years of follow-up

exceed the costs of treatment (brachytherapy or prostatectomy) at 6 and 8 years of follow-up, respectively. They also found that performing biopsy every other year led to 99% increase in savings (AS vs. primary treatment) per head over a period of 10 years compared to the scenario where biopsy is performed every year. Despite this, several AS studies schedule biopsies for every patient annually (Tosoian et al., 2011; Welty et al., 2015). For patients enrolled in PRIAS the schedule is comparatively less rigorous. One biopsy is performed at the time of induction, and the rest are scheduled at 1, 4, 7, 10 years and every 5 years thereafter. For patients who have a PSA doubling time (PSA-DT) less than 10 years, repeat biopsy every year is advised.

The biopsy schedule of PRIAS program is less rigorous than other programs, and yet PRIAS has a high non compliance rate for repeat biopsies. Bokhorst et al., 2015 reported that the percentage of men receiving repeat biopsies decreased from 81% at year 1 to 60% in year 4, 53% in year 7 and 33% in year 10 of follow up. Non compliance of biopsy schedule reduces the effectiveness of AS programs, as progression is detected late. On the other hand even if patients comply with the schedule, be it annually or the schedule of PRIAS, it may not be suitable for them. A patient whose cancer progresses slowly will often end up having biopsies when they are not needed. For a patient who has a faster progressing disease, crude measures such as PSA-DT are employed to decide if frequent biopsies are required. The fact that existing schedules require improvement is also evident in some of the reasons given by patients for non-compliance: 'patient does not want biopsy', 'PSA stable', 'complications on last biopsy'and 'no signs of disease progression on previous biopsy'.

Let us assume that we have a new patient j enrolled in the AS program and let T_j^* denote the actual time of GR for this patient. Let the schedule of biopsies for this patient be given by $\{T_{j0}^b, T_{j1}^b, \dots T_{jN_j^b}^b\}$, where $T_{jN_j^b}^b$ is the time at which GR is detected and no further biopsies are conducted. The total number of repeat biopsies conducted till GR is detected is denoted by N_j^b. Because of the periodical nature of biopsy schedules, we only know that $T_j^* \epsilon (T_{jN_j^b-1}^b, T_{jN_j^b}^b]$. The most useful biopsy schedule for a patient will be the one with the least number of biopsies N_j^b and the smallest offset $O_j = T_{jN_j^b}^b - T_j^*$ possible. i.e. the goal is to detect GR as early as possible with minimum possible number of biopsies. The search for the most useful biopsy schedule is the motivation behind this work. To this end, we have proposed alternative biopsy schedules, belonging to a class of schedules called personalized schedules. Personalized schedules are tailored separately for every patient and every disease. A simple example is the PRIAS schedule, which is personalized since it depends on the PSA-DT of the patient, an indicator of the state of disease. More sophisticated personalized schedules have been developed in the past. For e.g. Bebu and Lachin, 2017 have proposed Markov models based cost optimized personalized schedules. O'Mahony et al., 2015 have proposed cost optimized personalized equi-spaced screening intervals, using Microsimulation Screening Analysis (MISCAN) models. Parmigiani, 1998 have used information theory to come up with schedules for detecting time to event in the smallest possible time interval. Most of these methods however create an entire schedule in advance. Rizopoulos et al., 2016 have proposed dynamic personalized schedules for longitudinal markers using joint models for time to event and longitudinal data(Rizopoulos, 2012; Tsiatis and Davidian, 2004).

The personalized schedules we have proposed in this paper, utilize joint models and are dynamic. i.e. at a time only one future visit is scheduled, based on all the information gathered up to that point in time. More specifically, we have proposed two types of personalized schedules. One based on expected time of GR of a patient and the second based on the risk of GR. We have also analyzed an approach where the two types of personalized schedules are combined. Both types of schedules not only consider a patient's measurable attributes such as age, but also latent patient to patient variations in health, which cannot be measured directly. Results from previous repeat biopsies of the patient and PSA measurements as well as the population level information about hazard of GR, are used by the personalized schedules that we have proposed. It is important to note that a schedule for DRE measurements is not of interest since it is a non invasive procedure and has no serious medical implications. Thus the only event of interest is GR and not disease progression or DRE crossing the threshold of cT2c.

Using joint models to model the PSA measurements and risk of GR has the advantage that the association between the two is also modeled. More importantly, the association is modeled

via random effects, and therefore the models have an inherent patient specific nature. Secondly, joint models allow modeling the entire longitudinal history of PSA measurements, which is more sophisticated than PSA-DT. The use of PSA measurements in creating a personalized schedule is important because PSA is easy to measure, is cost effective and does not have any side effects. Secondly, in PRIAS Bokhorst et al., 2015 found that compliance rate for PSA measurements was as high as 91%. They also showed that there were more men who had a Gleason score greater than 6 as well as PSA-DT less than 3 years compared to men who had Gleason > 6 as well as PSA-DT larger than 3 years. i.e. Information from PSA was found to be indicative of GR. Lastly, some patients/doctors in PRIAS did not comply with the biopsy schedule because they considered PSA to be stable. However, if information from PSA is used in a methodical manner, it can lead to a more informative medical decision making process.

The rest of the paper is organized as follows. Section 2 covers briefly the joint modeling framework in context of the problem at hand. Section 3 details the personalized scheduling approaches we have proposed in this paper. In Section 4 we demonstrate personalized schedules in a real world scenario by employing them for the patients from the PRIAS study. Lastly, in Section 5, we present the results from a simulation study we conducted, to compare personalized schedules with the schedule of PRIAS study, as well with the most aggressive biopsy schedule of doing annual biopsies.

2 Joint models for time to event and longitudinal outcomes

The first step in creating a personalized schedule for biopsies is to come up with a model for Gleason scores, PSA levels and other patient specific characteristics. In PRIAS, PSA levels are measured at the time of induction, every 3 months for the first 2 years in the study and then every 6 months thereafter. Thus PSA levels can be modeled as a longitudinal outcome. As mentioned earlier, patients in PRIAS have a Gleason score of 6 or less at the time of induction in the study, and patients are removed from AS the first time GR takes place. Since our interest lies in finding the time of GR, we model it as a time to event outcome. A joint model for time to event and longitudinal outcomes is used to model the association between the two types of outcomes. We next present a short introduction of the joint modeling framework we will use in this work.

2.1 Joint model specification

Let T_i^* denote the true event 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. The true time of event T_i^* cannot be observed directly and it is only known that it falls in an interval $(l_i, r_i]$, where $l_i = T_{iN_i^b-1}^b, r_i = T_{iN_i^b}^b$ if the event (GR in the current context) is observed, and $l_i = T_{iN_i^b}^b, r_i = \infty$ if patient drops out of AS. The latter is also known as right censoring. Further let $\boldsymbol{y}_i$ denote the $n_i \times 1$ longitudinal outcome vector of the i^{th} patient. The population of interest consists of all the patients enrolled in AS. For a sample of n patients from this population the observed data is denoted by $\mathcal{D}_n = \{l_i, r_i, \boldsymbol{y}_i; i = 1, \dots n\}$.

To model the evolution of the longitudinal measurements over time, the joint model utilizes generalized linear mixed effects model. The distribution of longitudinal outcome $\boldsymbol{y}_i$ conditional on the random effects $\boldsymbol{b}_i$ is assumed to be a member of the exponential family. The corresponding linear predictor is given by:

$$k[E\{y_i(t) \mid \boldsymbol{b}_i\}] = m_i(t) = \boldsymbol{x}_i^T(t)\boldsymbol{\beta} + \boldsymbol{z}_i^T(t)\boldsymbol{b}_i$$

where, $k(\cdot)$ denotes a known one-to-one monotonic link function and $y_i(t)$ denotes the value of the longitudinal outcome for patient i at time t. The row vector of design matrix for fixed effects is denoted by $\boldsymbol{x}_i(t)$ and for random effects is denoted by $\boldsymbol{z}_i(t)$. 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 $\boldsymbol{D}$.

To model the effect of longitudinal outcome on hazard of event, joint models utilize a relative risk sub-model. The hazard of event 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:

$$\begin{aligned} h_i(t \mid \mathcal{M}_i(t), \boldsymbol{w}_i) &= \lim_{\Delta t \to 0} Pr\{t \leq T_i^* < t + \Delta t \mid T_i^* \geq t, \mathcal{M}_i(t), \boldsymbol{w}_i\}/\Delta t \\ &= h_0(t) \exp[\boldsymbol{\gamma}^T \boldsymbol{w}_i + f\{M_i(t), \boldsymbol{b}_i, \boldsymbol{\alpha}\}] \end{aligned}$$

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

- Association between hazard of event at time t and longitudinal response at the same time point:
  $f\{M_i(t), \boldsymbol{b}_i, \boldsymbol{\alpha}\} = \alpha m_i(t)$
- Association between hazard of event at time t and longitudinal response, as well as slope of longitudinal response at the same time point:
  $f\{M_i(t), \boldsymbol{b}_i, \boldsymbol{\alpha}\} = \alpha_1 m_i(t) + \alpha_2 m_i'(t)$, with $m_i'(t) = \frac{dm_i(t)}{dt}$

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).

2.2 Parameter estimation

In this work, we estimate parameters of the joint model using Markov chain Monte Carlo (MCMC) methods under the Bayesian framework. Let $\boldsymbol{\theta}$ denote the vector of the parameters of the joint model. The joint model postulates that given the random effects, time to event and longitudinal responses taken over time are all mutually independent. Under this assumption the posterior distribution of the parameters is given by:

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{b} \mid \mathcal{D}_n) &\propto \prod_{i=1}^{n} p(l_i, r_i, \boldsymbol{y}_i \mid \boldsymbol{b}_i, \boldsymbol{\theta}) p(\boldsymbol{b}_i \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) \\ &\propto \prod_{i=1}^{n} \prod_{l=1}^{n_i} p(y_{il} \mid \boldsymbol{b}_i, \boldsymbol{\theta}) p(l_i, r_i \mid \boldsymbol{b}_i, \boldsymbol{\theta}) p(\boldsymbol{b}_i \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) \end{aligned}$$

where the likelihood contribution of longitudinal outcome is

$$p(y_{il} \mid \boldsymbol{b}_i, \boldsymbol{\theta}) = \exp\left\{ \frac{y_{il}\psi_{il}(\boldsymbol{b}_i) - c\{\psi_{il}(\boldsymbol{b}_i)\}}{a(\phi)} - d(y_{il}, \phi) \right\},$$

where $\psi_{il}(\boldsymbol{b}_i)$ and ϕ denote the natural and dispersion parameters in the exponential family, respectively, and $c(\cdot)$, $d(\cdot)$ and $a(\cdot)$ are known functions specifying the member of the exponential family. The likelihood contribution of the survival outcome is given by

$$p\{l_i, r_i \mid \boldsymbol{b}_i, \boldsymbol{\theta}\} = \exp\left\{ -\int_0^{l_i} h_i(s \mid \mathcal{M}_i(s), \boldsymbol{w}_i)\, ds \right\} - \exp\left\{ -\int_0^{r_i} h_i(s \mid \mathcal{M}_i(s), \boldsymbol{w}_i)\, ds \right\}$$

The integral in the likelihood function of the survival function does not have a closed-form solution, and therefore we used a 15-point Gauss–Kronrod quadrature rule to approximate it.

For the parameters of the longitudinal outcomes we use standard default priors. More specifically, independent normal priors with zero mean and variance 100 for the fixed effects $\boldsymbol{\beta}$ and inverse Gamma priors for scale parameters. For the variance–covariance matrix $\boldsymbol{D}$ of the random effects we take inverse Wishart prior with an identity scale matrix and degrees of freedom equal to the number q of the random effects. For the survival model parameters $\boldsymbol{\gamma}$ and the association parameters $\boldsymbol{\alpha}$, we use independent normal priors with zero mean and variance 100. For the penalized version of the B-spline approximation to the baseline hazard, we use the following prior for parameters γ_{h_0} (Lang and Brezger, 2004):

$$p(\gamma_{h_0} \mid \tau_h) \propto \tau_h^{\rho(\boldsymbol{K})/2} \exp\Big(-\frac{\tau_h}{2} \gamma_{h_0}^T \boldsymbol{K} \gamma_{h_0} \Big)$$

where τ_h is the smoothing parameter that takes a Gamma(1, 0.005) hyper-prior in order to ensure a proper posterior for γ_{h_0}, $\boldsymbol{K} = \Delta_r^T \Delta_r + 10^{-6} \boldsymbol{I}$, where Δ_r denotes the r^{th} difference penalty matrix, and $\rho(\boldsymbol{K})$ denotes the rank of $\boldsymbol{K}$.

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 the scheduling methods, let us assume that the a personalized schedule is to be created a new patient enumerated 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 their last biopsy performed at time t, and that the PSA measurements 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 each time results from repeat biopsy or a PSA measurement is available.

3.1 Posterior predictive distribution for time to GR

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

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

The posterior predictive distribution depends on the observed longitudinal history via the random effects $\boldsymbol{b_j}$. The posterior distribution of the parameters $\boldsymbol{\theta}$, denoted by $p(\boldsymbol{\theta} \mid \mathcal{D}_n)$ is obtained from the joint model fitted to the original data set of patients $\mathcal{D}_n$.

3.2 Loss functions

To find the time u of next biopsy, we use principles from statistical decision theory in a Bayesian setting (Berger, 1985; Robert, 2007). More specifically, we propose to choose future biopsy time u by minimizing the posterior expected loss $E_g[L(T_j^*, u)]$, where the expectation is taken w.r.t. the posterior predictive distribution $g(T_j^*)$.

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

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.

3.2.1 Expected and median time of GR

One of the reasons, patients did not comply with the existing PRIAS schedule was 'complications on a previous biopsy '. Therefore, it makes sense to have as less biopsies as possible. In the ideal case only 1 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 $L(T_j^*, u) =\mid T_j^* - u \mid$ 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:

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

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

$$\begin{aligned} E_g[L(T_j^*, u)] &= E_g[\mid T_j^* - u \mid] \\ &= \int_u^\infty (T_j^* - u) g(T_j^*)\, dT_j^* + \int_t^u (u - T_j^*) g(T_j^*)\, dT_j^* \end{aligned} \tag{3}$$

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

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

3.2.2 Dynamic risk of GR

In a practical scenario it is possible that a doctor or a patient may not want to exceed a certain risk of GR $1 - \pi_j(u \mid t, s)$ since the last biopsy. This may be also useful in the cases where variance of $g(T_j^*)$ is high, rendering expected time of GR, or any other measure of central tendency of $g(T_j^*)$ unsuitable. The personalized scheduling approach based on dynamic risk of GR, schedules the next biopsy at a time point u such that the dynamic risk of GR is higher than a certain threshold $1 - \kappa,\ \kappa \epsilon [0, 1]$ beyond u. Or in other words the dynamic survival probability $\pi_j(u \mid t, s)$ is below a threshold κ beyond u. To this end, the posterior expected loss for the following multilinear loss function can be minimized to find the optimal u:

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

where $k_1 > 0$, $k_2 > 0$ are constants parameterizing the loss function. The posterior expected loss function $E_g[L_{k_1,k_2}(T_j^*, u)]$ obtains its minimum at $u = \pi_j^{-1}\left\{\frac{k_1}{k_1+k_2}\right\}$ (Robert, 2007). The choice of k_1, k_2 is equivalent to the choice of κ. More specifically, $\kappa = \frac{k_1}{k_1+k_2}$.

3.3 A mixed approach between $E_g[T_j^*]$ and dynamic risk of GR

When the variance $Var_g[T_j^*]$ of $g(T_j^*)$ is small, then $E_g[T_j^*]$ is practically very useful. However when the variance is large, there may not be a clear central tendency of the distribution and $E_g[T_j^*]$ will either overshoot T_j^* or undershoot it by a big margin. In the latter case more biopsies will be required until GR is detected at some time point $T_{jN_j^b} > T_j^*$. The overshooting margin can be measured as an offset $O_j = T_{jN_j^b} - T_j^*$. The maximum acceptable O_j in PRIAS is 3 years, which corresponds to the time gap between biopsies of the PRIAS fixed schedule. When $Var_g[T_j^*]$ is large, the proposals based on $E_g[T_j^*]$ can have a large O_j. Thus we propose that if the difference between the 0.025 quantile and $E_g[T_j^*]$ is more than 3 years then proposals based on dynamic risk of GR be used instead. We call this approach a mixed approach.

3.4 Estimation

3.4.1 Estimation of $E_g[T_j^*]$ and $Var_g[T_j^*]$

Since there is no closed form solution available for $E_g[T_j^*]$, for its estimation we utilize the following relationship between expected time of GR and dynamic survival probability:

$$E_g[T_j^*] = t + \int_t^\infty \pi_j(u \mid t, s)\, du$$

There is no closed form solution available for the integral and hence we approximate it using Gauss-Kronrod quadrature. We preferred this approach over Monte Carlo methods to estimate $E_g[T_j^*]$ from the posterior predictive distribution $g(T_j^*)$. This was done 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). Thus the latter approach was computationally faster. As mentioned earlier, a limitation of expected time of GR is that it is practically useful only when the variance of predictive distribution $g(T_j^*)$ is small. The variance is given by:

$$\begin{aligned} Var_g[T_j^*] &= E_g[T_j^{*2}] - E_g[T_j^*]^2 \\ &= 2\int_t^\infty (u-t)\pi_j(u \mid t, s)\, du - \left(\int_t^\infty \pi_j(u \mid t, s)\, du\right)^2 \end{aligned} \tag{6}$$

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 **??**.

3.4.2 Estimation of κ

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 e.g. if maximum acceptable risk is 75% then $\kappa = 0.25$, and correspondingly all $k_1, k_2 \mid k_1 = \frac{k_2}{3}$ can be used in equation (5) to calculate u.

While a doctor's advice can be invaluable, it is also possible to automate the choice of κ. We propose to choose a κ for which a binary classification accuracy measure (López-Ratón et al., 2014; Sokolova and Lapalme, 2009), 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_a and in the cases at some future time point $t_b > t_a$, 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, 2014). The time window Δt can be either chosen on a clinical basis, or such that uncertainty in estimation of $\pi_j(t + \Delta t \mid t, s)$ is below a certain threshold or it can even be chosen such that $AUC(t, \Delta t, s)$ (Rizopoulos, 2014) is largest. i.e. Δt for which the model has the most discriminative capability at time t. The binary classification accuracy measures we maximize to select the threshold κ are the following (the binary classification measures are functions of $t, \Delta t, s$, however the notation is dropped for readability):

- Youden's index: $J = \text{Sensitivity} + \text{Specificity} - 1$,
  where sensitivity is defined as $Pr(\pi_j(t+\Delta t \mid t, s) \leq \kappa \mid T_j^* \epsilon (t, t+\Delta t])$ and specificity is defined as $Pr(\pi_j(t + \Delta t \mid t, s) > \kappa \mid T_j^* > t + \Delta t)$ (Rizopoulos, 2014). Let $k_1 = FP \cdot TP - FN \cdot TN$ and $k_2 = (TP + FN)(FP + TN) - k_1$, where TP, FP, TN and FN are the number of true positives, false positives, true negatives and false negatives at time point t. The optimal k_1, k_2 or equivalently the κ is obtained from $\arg\max_{k_1,k_2} J$.
- F1 Score: $F1 = \frac{2TP}{2TP+FP+FN}$. In this case if $k_1 = 2TP$ and $k_2 = FP+FN$, then $\arg\max_{k_1,k_2} F1$ gives the optimal k_1, k_2 or equivalently the κ.

3.5 Algorithm

The algorithm in Figure 1 elucidates the process of creating a personalized schedule for patient j. In the algorithm:

1. t denotes the time of the latest biopsy.
2. s denotes the time of the latest available PSA measurement.
3. u denotes the personalized biopsy time based on $g(T_j^*)$.
4. u^{pv} denotes the time at which a repeat biopsy was proposed at the last visit to the hospital.
5. T^{nv} denotes the time of the next visit for PSA measurement.

Since PRIAS and most AS programs strongly advise against conducting more than 1 biopsy per year, the algorithm adjusts the optimal time u of biopsy in case the last biopsy was within an year.

4 Personalized schedules for patients in PRIAS

To demonstrate how the personalized scheduling algorithm described in Section **??** works, we apply it to the patients in the PRIAS dataset. To this end, we divide the PRIAS data set into training(5264 patients) and demonstration data sets (3 patients). We fit a joint model to the training data set and then use it to create a personalized schedule for patients in demonstration data set. We fit the joint model using the R package JMbayes (Rizopoulos, 2014), which uses the Bayesian methodology to estimate the model parameters.

4.1 Fitting the joint model to PRIAS dataset

The training data set contains information about 5938 prostate cancer patients who satisfied the conditions for enrollment in AS. For every patient the age at the time of induction in AS was recorded. PSA was measured every 3 months for first 2 years and every 6 months thereafter. To detect GR, biopsies were conducted at different time points on the basis of a predetermined schedule as well as PSA-DT as described in Section 1. For the longitudinal analysis of PSA measurements we used $\log_2 PSA$ measurements instead of the raw data. The log transformation was done because the PSA scores took very large values around the time of disease progression. This indicated that the underlying distribution for PSA scores was right skewed. The longitudinal sub-model of the joint model we fit is given by:

$$\begin{aligned}
\log_2 PSA(t) &= m_i(t) + \varepsilon_i(t), \\
m_i(t) &= (\beta_0 + b_{i0}) + \beta_1(Age - 70) + \beta_2(Age - 70)^2 \\
&\quad + \sum_{k=1}^{4} \beta_{k+2} B_k(t, \mathcal{K}) + b_{i1} B_7(t, 0.5) + b_{i2} B_8(t, 0.5) \\
\varepsilon_i(t) &\sim N(0, \sigma^2),
\end{aligned}$$

where, the measurement error $\varepsilon_i(t)$ is assumed normally distributed with mean zero and variance σ^2, and is independent of the random effects $\boldsymbol{b}_i$. The evolution of PSA levels over time is modeled flexibly using B-splines. For the fixed effects part the spline consists of 3 internal knots. The internal knots are at $\mathcal{K} = \{0.5, 1.2, 2.5\}$ years, and boundary knots are at 0 and 7 years. For the random effects part there is only 1 internal knot at 0.5 years and the boundary knots are at 0 and 7 years. The choice of knots was based on exploratory analysis as well as on the basis of model selection criteria AIC and BIC. The variable Age was median centered to avoid numerical instabilities while estimating the parameters in the model. For the survival sub-model the hazard function we fitted is given by:

$$h_i(t) = h_0(t) \exp[\gamma_1(Age - 70) + \gamma_2(Age - 70)^2 \alpha_1 m_i(t) + \alpha_2 m_i'(t)] \tag{7}$$

where, α_1 and α_2 are measures of strength of association between hazard of GR and PSA value $m_i(t)$ and PSA velocity $m_i'(t)$, respectively. As mentioned earlier, in PRIAS study PSA-DT is used to decide the schedule of biopsies. However PSA-DT is computed using observed PSA values,

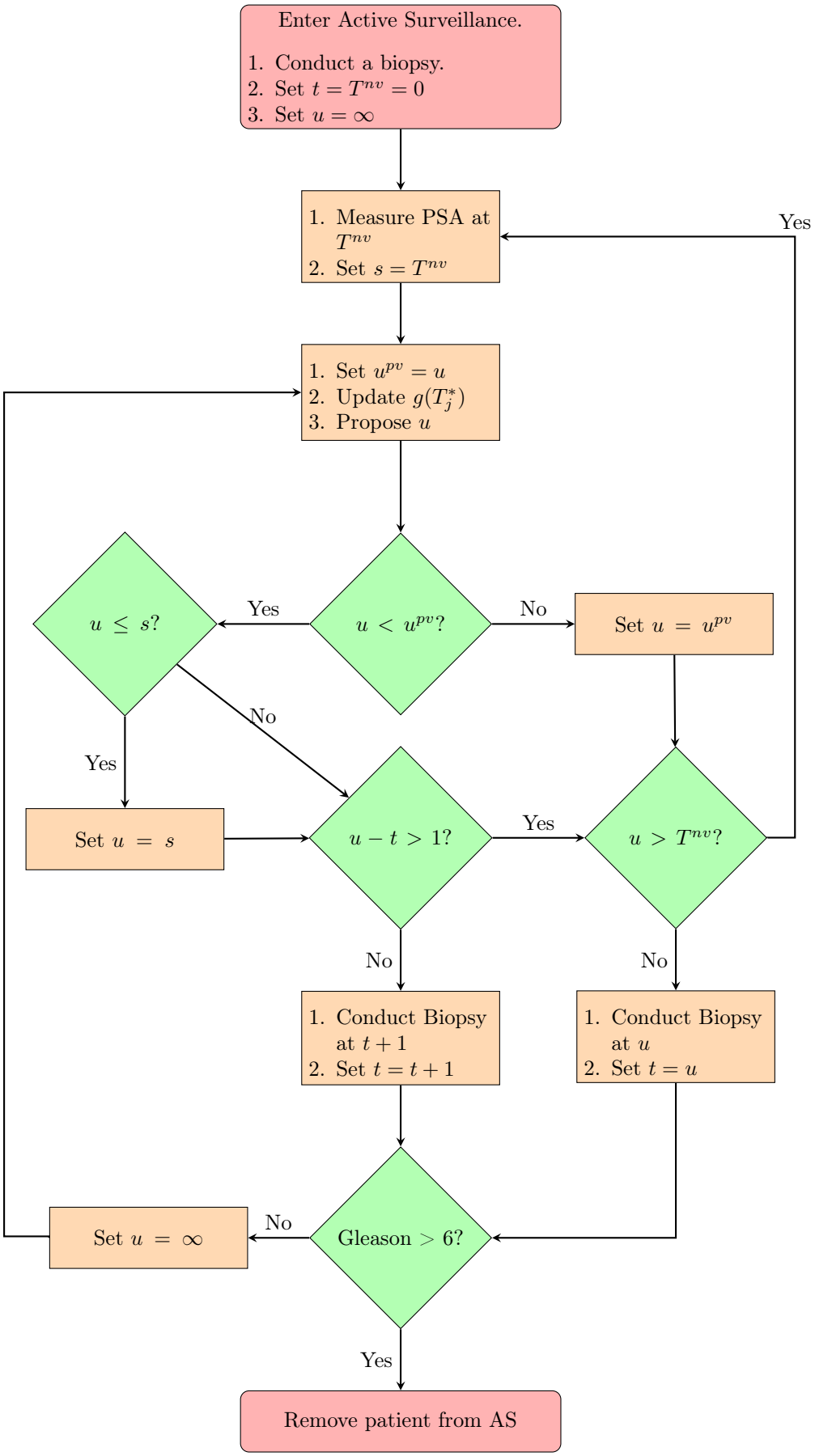


Figure 1: Algorithm for creating a personalized schedule for patient j.

and thus interval censoring observed in PRIAS is independent and non informative of underlying health of the patient. Lastly, to fit the joint model we use the R package JMbayes Rizopoulos, 2014, which uses a Bayesian approach for parameter estimation.

4.1.1 Parameter Estimates

The posterior parameter estimates $p(\boldsymbol{\theta} \mid \mathcal{D}^{PRIAS})$ for the joint model we fitted to the PRIAS data set are shown in Table 1 and Table 2. Since the longitudinal evolution of $\log_2 PSA$ is modeled with non-linear terms, the interpretation of the coefficients corresponding to time is not straightforward. In lieu of the interpretation we present the fitted evolution of PSA over a period of 10 years for a patient who is 70 years old in Figure 2. It can be seen that the after the first 6 months the PSA levels steadily increase over the follow up period. Since the model for PSA has only additive terms, this evolution remains same for all patients. The effect of Age only affects the baseline PSA score. However it is so small that it can be ignored for all practical purposes.

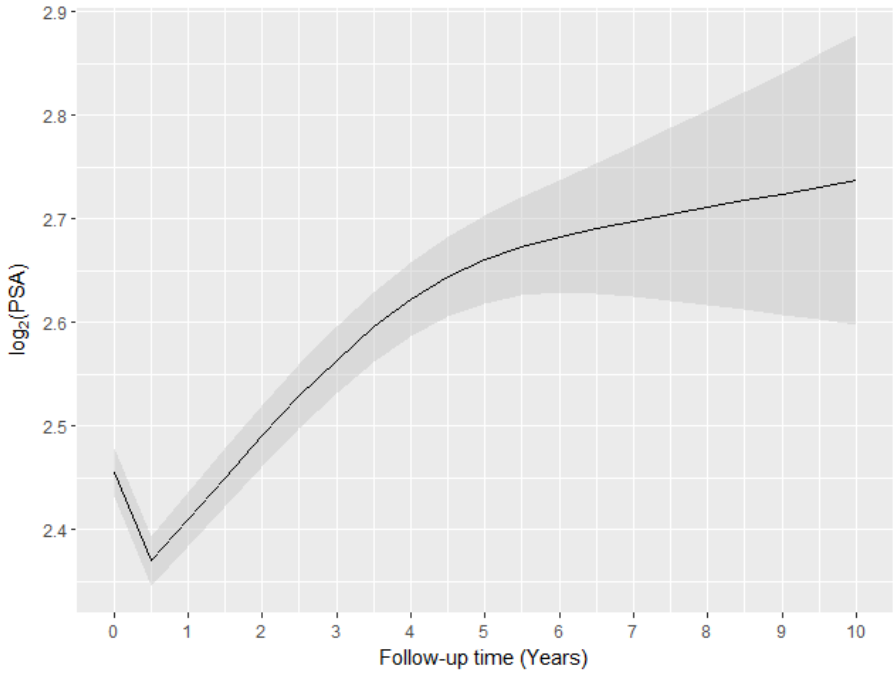


Figure 2: Fitted evolution of $\log_2 PSA$ over a period of 10 years, for a patient who was inducted in AS at the Age of 70 years.

	Mean	Std. Dev	2.5%	97.5%	P
Intercept	2.455	0.012	2.433	2.480	<0.000
(Age - 70)	0.003	0.001	4.9×10^{-4}	0.006	0.032
(Age - 70) × (Age - 70)	-0.001	1.4×10^{-4}	-0.001	-3.5×10^{-4}	<0.000
Spline: visitTimeYears[0, 0.5]	-0.006	0.012	-0.031	0.017	0.674
Spline: visitTimeYears[0.5, 1.2]	0.228	0.019	0.192	0.265	<0.000
Spline: visitTimeYears[1.2, 2.5]	0.140	0.029	0.088	0.197	<0.000
Spline: visitTimeYears[2.5, 7]	0.303	0.039	0.227	0.379	<0.000
σ	0.324	0.001	0.321	0.326	

Table 1: Longitudinal sub-model estimates for joint model.

For the survival sub-model, the parameter estimates in Table 2 show that only $\log_2 PSA$ velocity is strongly associated with hazard of GR. For any patient, a unit increase in $\log_2 PSA$ velocity corresponds to a 11 time increase in hazard of GR. The effect of $\log_2 PSA$ value and effect of Age on hazard of GR are so small that they can be safely ignored for all practical purposes.

Variable	Mean	Std. Dev	2.5%	97.5%	P
Age - 70	0.037	0.006	0.025	0.0490	<0.000
(Age - 70) × (Age - 70)	-0.001	0.001	-0.003	1.8×10^{-4}	0.104
$\log_2 PSA$	-0.049	0.064	-0.172	0.078	0.414
Slope: $\log_2 PSA$	2.407	0.319	1.791	3.069	<0.000

Table 2: Survival sub-model estimates for joint model.

4.2 Demonstration of personalized schedules

In this section, we demonstrate how the personalized scheduling algorithm adapts the time of performing a biopsy according to the PSA history and results from repeat biopsies. The 3 patients we have chosen for the demonstration data set are part of PRIAS program and they have had their repeat biopsies already. Hence a full scale comparison between PRIAS biopsy schedule and personalized scheduling algorithm's biopsy schedule is not possible.

The first patient of interest is patient nr. 3174 who was inducted in the PRIAS program at the age of 74 years. Between the last follow up and time of induction there was no repeat biopsy performed for this patient. Hence the predictive distribution $g(T_j^*)$ for this patient depends only on the PSA measurements. For this patient, the evolution of PSA, time of last biopsy and proposed times of biopsies are shown in Figure 3. It can be seen that the PSA remains stable for the first 2 years of follow ups, but increases rapidly after that for the next 2 years of follow up. Since the hazard of GR depends on PSA velocity (Table2), the schedule of biopsy based on personalized scheduling algorithm (Section 1), adjust the times of biopsy according to the steep rise in PSA profile. At 2 years the expected and median time of GR are 12.5 years and 15.2 years respectively, whereas at 4 years, they are 5.3 and 4 years respectively.

It is important to note that for patient nr. 3174, a biopsy scheduled using $E_g[T_j^*]$ at year 2 is not as useful as a biopsy scheduled using the same method at year 4. This because $Var_g[T_j^*]$ is considerably lower at year 4 as shown in Figure 4a. The variance doesn't depend much on number of PSA measurements, but rather on the PSA values. In the case at hand the variance drops quickly when PSA measurements increase sharply, which corresponds to PSA velocity $m_i'(t)$ being a strong predictor of the hazard of GR (Table 2).

The second patient of interest is patient nr. 911. Figure 5 shows the evolution of PSA, time of last biopsy and proposed biopsy times for this patient. Firstly, it can be seen that at between year 1.5 and year 2, the PSA rises sharply, and accordingly the personalized schedules based on expected time of GR preponed the proposed biopsy time from 14.2 years to 13.8 years. The change in median time of GR is trivial. The second observation is that between year 2 and year 3 the PSA decreases sharply and accordingly, the proposed biopsy times are postponed. More specifically the median time of GR increased from 15.2 years at year 2 to 17 years at year 3. Whereas the expected time of GR increased from 13.8 years to 16.6 years. It can also be seen that PSA remains stable up to to year 4. Lastly, because no GR is found at the repeat biopsy performed at 4.1 years, this further leads to postponing of the biopsy times, which become 18.7 and 19.9 for expected time of GR and median time of GR respectively. For patient nr. 911 the biopsies scheduled using $E_g[T_j^*]$ at year 4.1 are expected to be very close to the true GR time of the patient. This because the variance $Var_g[T_j^*]$ is quite low at year 4 as shown in Figure 4b. From the figure it is also evident that the variance decreases considerably each time information about T_j^* from a repeat biopsy is available.

The fact that PSA velocity affects the biopsy times is also evident in the case of patient 2340, whose PSA evolution is shown in Figure 6. We have purposefully ignored all the biopsies conducted for this patient in PRIAS, to isolate the effect of PSA history. It can be seen that PSA slowly rises to a value of 14 ng/mL over a period of 2.5 years. Correspondingly, there is little change in proposed biopsy times. On the other hand between 2.3 years and 3.3 years the PSA rises very sharply and correspondingly the expected time of GR decreases by an year from 12.5 years to 11.5 years in that period. The median time of GR remains the same though.

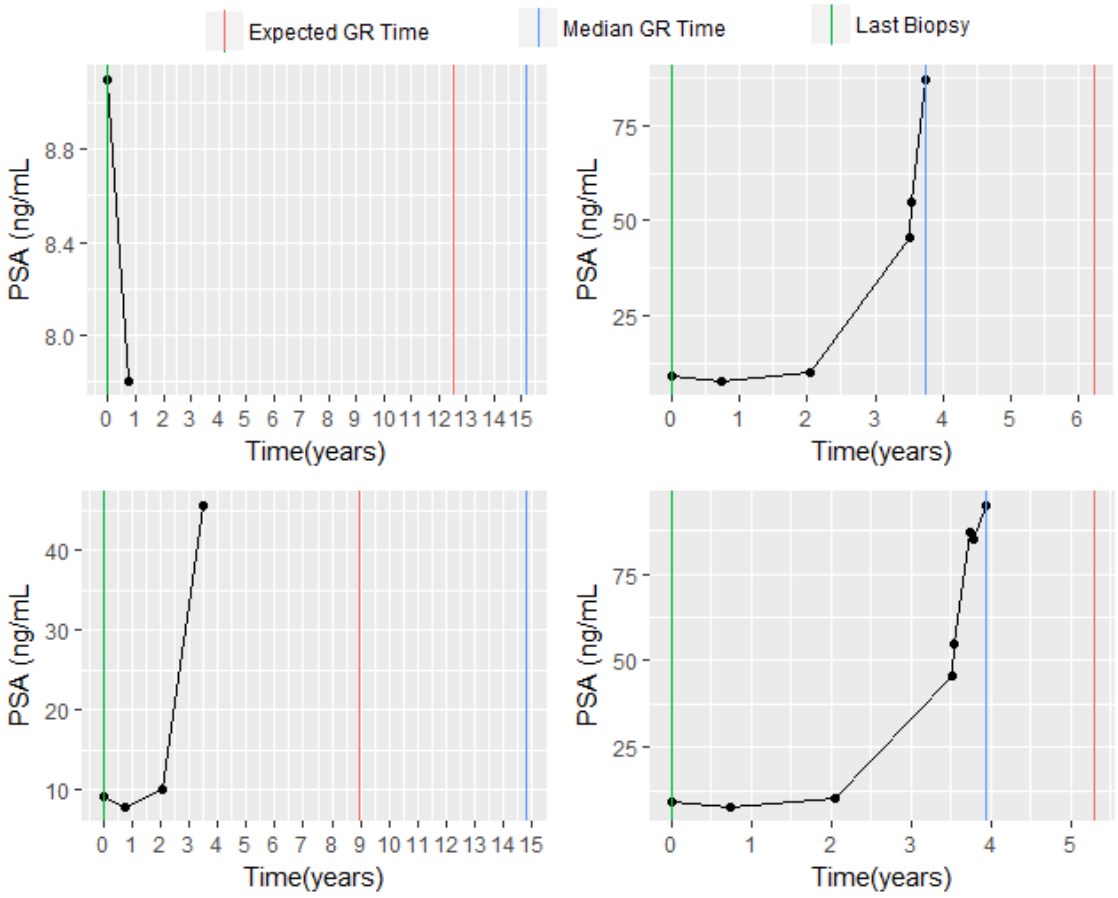
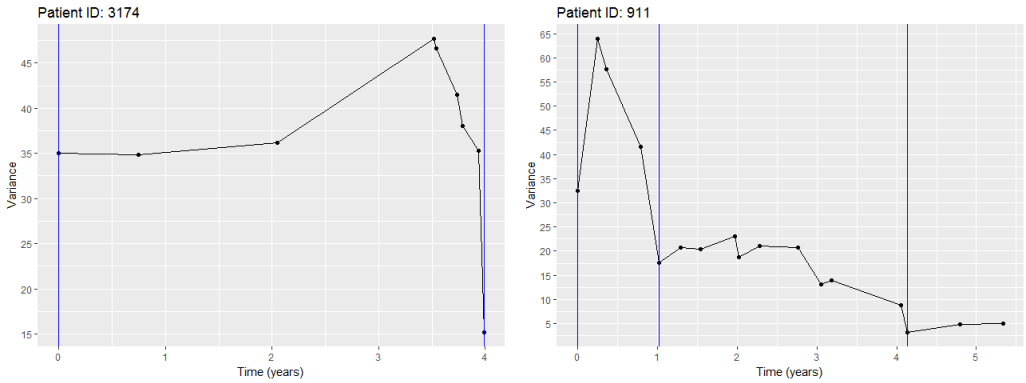


Figure 3: Proposed biopsy times for patient 3174 from PRIAS.



(a) Variance of the predictive distribution $g(T_j^*)$ over a period of 4 years for patient 3174. Blue vertical lines indicate biopsies.

(b) Variance of the predictive distribution $g(T_j^*)$ over a period of 5 years for patient 911. Blue vertical lines indicate biopsies.

Figure 4: Variance of the predictive distribution $g(T_j^*)$

5 Simulation study

The application of personalized schedules for patients in PRIAS demonstrated that personalized schedules adapt according to the PSA and repeat biopsy history of each patient. However, since the patients in PRIAS have already had their biopsies as per the PRIAS schedule, we were not able to evaluate the efficacy of personalized schedules against the PRIAS schedule. To this end, we have performed a simulation study to compare 3 broad categories of schedules: Personalized schedules, PRIAS schedule and the schedule of annual biopsy. In the simulation study we simulate

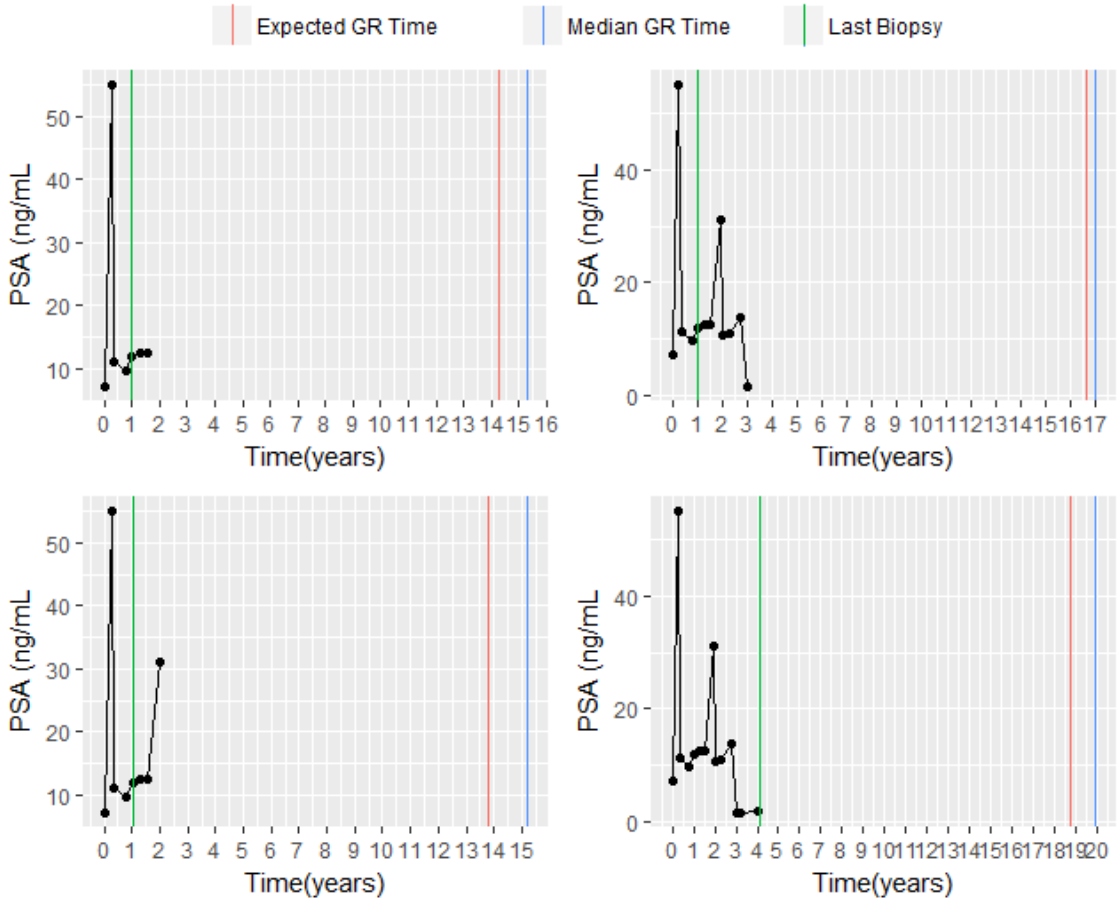


Figure 5: Proposed biopsy times for patient 911 from PRIAS.

the progression of patients enrolled in an AS program. The PSA measurements are measured periodically as per a fixed schedule, however the biopsies are conducted as per the scheduling algorithms until GR is detected. We next present the details of the simulation study and the biopsy schedule evaluation criteria.

5.1 Simulation setup

5.1.1 Patient population

For the simulation study we first select a population $\mathcal{P}$ of patients enrolled in AS. We assume that the PSA and hazard of GR for the patients from this population follows a joint model of the form postulated in Section 4.1. The population parameters $\boldsymbol{\theta}^{\mathcal{P}}$ are selected to be equal to the posterior mean of parameters $E[\boldsymbol{\theta} \mid \mathcal{D}^{PRIAS}]$ estimated from the joint model fitted to PRIAS data set (Section 4.1.1). To demonstrate the efficacy of personalized schedules for patients with early as well late failure times, we ensured that the patients in population $\mathcal{P}$ were from 3 equal sized sub-groups. The baseline hazard for each of the sub-group was assumed to be the hazard function for a Weibull distribution. The shape and scale parameters (k, λ) for the aforementioned Weibull distribution are $(1.5, 4)$, $(3, 5)$ and $(4.5, 6)$.

5.1.2 Simulated data sets

From the population $\mathcal{P}$ we randomly sample a total of 200 data sets with 1000 patients each. For each of the patients, the longitudinal profiles for PSA scores are generated as per the visiting schedule of PRIAS study. i.e. Every 3 months for first 2 years and every 6 months thereafter. The true GR reclassification times are also generated. We then divide each of the 200 simulated data sets into training (750 patients) and test (250 patients) parts. Further we generate random and non-informative censoring times for the patients in 200 training data sets $\mathcal{D}^1, \ldots \mathcal{D}^{200}$. The

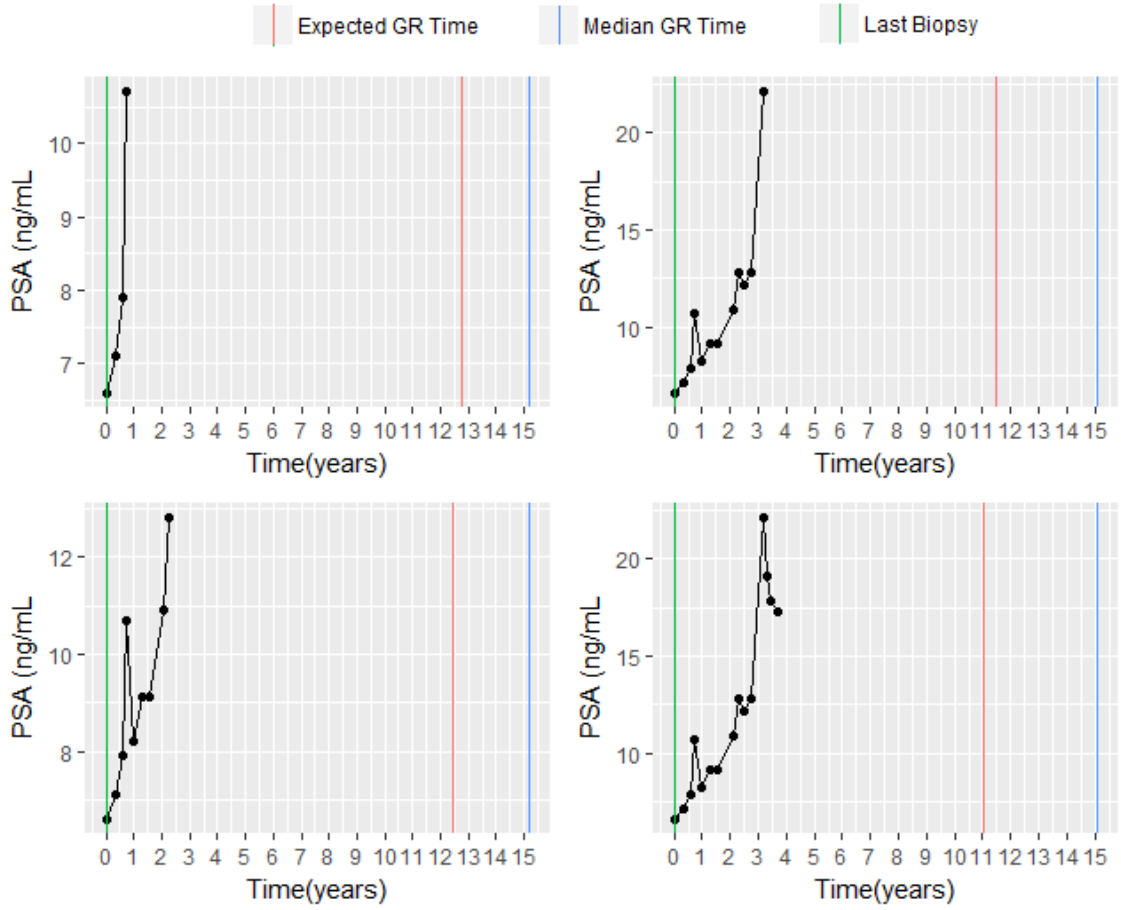


Figure 6: Proposed biopsy times for patient 2340 from PRIAS.

observed data in the k^{th} training data set is $\mathcal{D}^k = \{T_{ki}, \delta_{ki}, \boldsymbol{y}_{ki}; i = 1, \dots 750\}$, where $\boldsymbol{y}_{ki}$ denotes the PSA measurements for the i^{th} patient in the k^{th} training data set. $T_{ki} = \min(T^*_{ki}, C_{ki})$ denotes the observed GR time whereas the true GR time and censoring times are denoted by T^*_{ki} and C_{ki} respectively. $\delta_{ki} = I(T^*_{ki} < C_{ki})$ is the event indicator, where $I(\cdot)$ is an indicator function that takes the value 1 when $T^*_{ki} < C_{ki}$ and 0 otherwise. For the patients in the test data sets no censoring times are generated.

5.1.3 Personalized schedules for test patients

We create personalized biopsy schedules only for the patients in the test data set. To this end we first fit a joint model to the training data set to obtain posterior estimates of parameters $p(\boldsymbol{\theta} \mid \mathcal{D}^k)$, which are required for the posterior predictive distribution $g(T^*_{kj})$ of the j^{th} patient from the k^{th} test data set. To model the baseline hazard we use a P-splines approach (Section 2.1). While the posterior predictive distribution is sufficient for scheduling biopsies based on expected time of GR, the choice of time window Δt (Section 3.2.2) has to be made for scheduling biopsies on the basis of dynamic risk of GR. In PRIAS and in most AS programs biopsies are done at a gap of 1 to 3 years. A gap of 1 year between biopsies detects GR the earliest, and in worst case the detection of GR can be delayed by 1 year. Being a clinically relevant period of time to differentiate between patients who obtain GR and those who don't, we choose a Δt of 1 year. In schedules based on dynamic risk of GR, for a test patient j, we choose a value of κ which maximizes a binary classification measure (Section 3.4.2) at the last known repeat biopsy time t of the patient. To this end the binary classification measures are first computed over a fine grid of values in the interval $[0, 1]$ using the training data set and then the most optimal κ is chosen.

To create personalized schedules we employ the algorithm described in Section 3.5. The algorithm is run for 7 different settings, one each corresponding to the following: PRIAS schedule,

annual biopsy schedule, expected time of GR, median time of GR, dynamic risk of GR with κ chosen such that a) Youden index is maximized, b) F1 score is maximized. In addition to these a mixed approach is also employed where at any point in time.....

5.2 Evaluating efficacy of scheduling methods

For a particular biopsy scheduling method S, the first criteria in the evaluation of efficacy of the method is the number of repeat biopsies N^{bS} it takes before GR is detected for a patient from the population. The less the N^{bS} the better it is for patients. Our interest lies in the marginal distribution of N^{bS} for the population $\mathcal{P}$. Since In its simplest form, the mean $E[N^{bS}]$ and variance $Var[N^{bS}]$ of the aforementioned distribution indicate the performane of the scheduling algorithm. More specifically, a low mean as well as low variance is desired. Other quantiles of the distribution may also be used evaluation criteria. For e.g. a method which takes less than 2 (say) biopsies in 95% cases may be desriable.

The second criteria in evaluation of efficacy of a schedule is the offset. Given a schedule S, the offset for a particular patient j is defined as $O_j^S = T^S_{jN_j^{bS}} - T_j^*$, where N_j^{bS} is the number of biopsies required for patient j before GR is detected and $T^S_{jN_j^{bS}} > T_j^*$ is the time at which GR is detected by the scheduling mechanism S. Once again the interest lies in both the mean $E[O^S]$ and variance $Var[O^S]$ of the marginal distribution of offset.

5.2.1 Finding the most optimal schedule

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

$$L(S) = \sum_{g=1}^{G} \lambda_g \mathcal{G}_g(N^{bS})^{d_g=1} \mathcal{G}_g(O^S)^{d_g=0} \tag{8}$$

where $\mathcal{G}_g(\cdot)$ is either a function of number of biopsies or of the offset, and d_g is the corresponding indicator for this choice. Some examples of $\mathcal{G}_g(\cdot)$ are mean, median, variance and quantile function. Constants $\lambda_1, \dots \lambda_G$, where $\lambda_g \epsilon [0,1]$ and $\sum_{g=1}^{G} \lambda_g = 1$, are weights to differentially weigh-in the contribution of each of the G evaluation criteria manifested via the functions $\mathcal{G}_g(\cdot)$. An example loss function is:

$$L(S) = \lambda_1 E[N^{bS}] + \lambda_2 E[O^S] \tag{9}$$

The choice of λ_1 and λ_2 is not easy. This because biopsies have serious medical side effects and the cost of an extra biopsy cannot be quantified easily. To obviate this issue we utilize the equivalence between compound and constrained optimal designs (Cook and Wong, 1994). More specifically, it can be shown that for any λ_1 and λ_2 there exists a constant $C > 0$ for which minimization of loss function in Equation 9 is equivalent to minimization of the same, subject to the constraint that $E[O^S] < C$. i.e. 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. For e.g. in PRIAS the maximum gap between 2 repeat biopsies is 3 years. i.e. If GR is detected within 3 years of its occurance, it is acceptable by the AS program. For a more general scenario of the form in Equation 8, the solution can be found by minimizing $\mathcal{G}_G(\cdot)$ under the constraint $\mathcal{G}_g < C_g; g = 1, \dots G-1$.

In this work, we estimate $E[N^{bS}]$, $Var[N^{bS}]$, $E[O^S]$ and $Var[O^S]$ using pooled estimates of each from the 200 repetitions of the simulation study. The estimates are calculated separately for each of the 7 methods mentioned in Section 5.1.3. The pooled estimates for a scheduling method S are calculated as following:

$$\widehat{E[O^S]} = \frac{\sum_{k=1}^{200} n_k \widehat{E[O_k^S]}}{\sum_{k=1}^{200} n_k},$$
$$\widehat{Var[O^S]} = \frac{\sum_{k=1}^{200} (n_k - 1) \widehat{Var[O_k^S]}}{\sum_{k=1}^{200} (n_k - 1)},$$

where n_k are the number of test patients in the k^{th} simulation, $\widehat{E[O_k^S]} = \frac{\sum_{j=1}^{n_k} O_{kj}^S}{n_k}$ and $\widehat{Var[O_k^S]} = \frac{\sum_{j=1}^{n_k} (O_{kj}^S - \widehat{E[O_k^S]})^2}{n_k - 1}$ are the estimated mean offset and estimated variance of the offset for the k^{th} simulation, respectively. The estimates for number of biopsies N^{bS} are calculated similarly.

5.3 Results

The pooled estimates of the mean and variance of number of biopsies as well as offset from the simulation study are summarized in Table 3. We can see that schedules based on Expected time of GR and Median time of GR schedule on an average 2 biopsies before GR is detected and the variance in number of biopsies is very low. On the other hand PRIAS and annual biopsy schedule 4 or more biopsies for the same and the variance in number of biopsies is high. As expected the higher number of biopsies lead to a smaller offset as well. For e.g. annual biopsy schedule has an average offset of 6 months and has a low variance as well. The average offset for Expected and Median time of GR is approximately 13 to 14 months, which is acceptable time period in PRIAS. The variance of the offset is high though. In a nutshell methods which conduct high number of biopsies have less variance in offset.

	Total Patients	$E[N^{bS}]$	$Var[N^{bS}]$	$E[O^S]$	$Var[O^S]$
Annual	387	4.307	10.426	5.973	11.886
F1score	395	4.658	4.303	6.731	16.101
Expected Time of GR	379	1.997	1.405	13.764	115.228
Median Time of GR	395	2.056	1.670	14.271	136.769
Mixed.Approach	387	3.220	5.074	10.496	65.175
PRIAS	387	4.054	8.812	7.663	50.295
Youden	379	4.625	3.594	8.900	231.035

Table 3: Pooled estimates of mean and variance of number of biopsies and offset for the simulation study.

6 Discussion

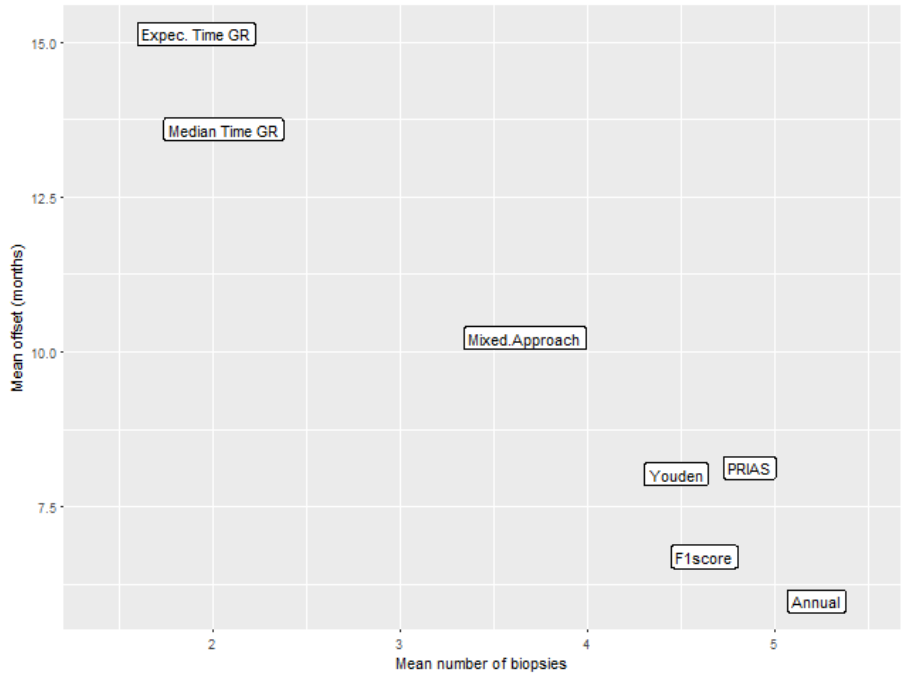
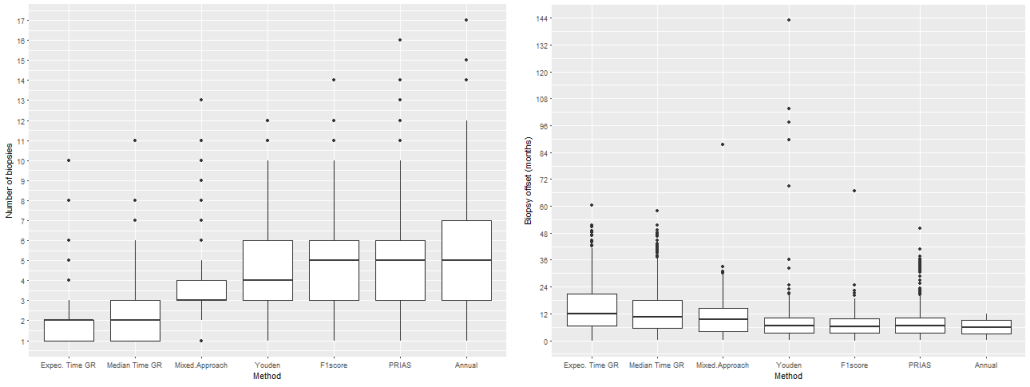


Figure 7: Estimated mean number of biopsies and mean offset (months) for the 7 scheduling methods.



(a) Boxplot for number of biopsies.

(b) Boxplot for offset (months)

Figure 8: Boxplot for number of biopsies and offset (months), for all the patients in the 200 simulated data sets.

References

Bebu, Ionut and John M. Lachin (2017). "Optimal screening schedules for disease progression with application to diabetic retinopathy". In: *Biostatistics*. DOI: 10.1093/biostatistics/kxv009.

Berger, James O (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer Science & Business Media.

Bokhorst, Leonard P et al. (2015). "Compliance rates with the Prostate Cancer Research International Active Surveillance (PRIAS) protocol and disease reclassification in noncompliers". In: *European urology* 68.5, pp. 814–821.

Bokhorst, Leonard 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". In: *European Urology* 70.6, pp. 954–960.

Brown, Elizabeth R (2009). "Assessing the association between trends in a biomarker and risk of event with an application in pediatric HIV/AIDS". In: *The annals of applied statistics* 3.3, p. 1163.

Cook, R Dennis and Weng Kee Wong (1994). "On the equivalence of constrained and compound optimal designs". In: *Journal of the American Statistical Association* 89.426, pp. 687–692.

Eilers, Paul HC and Brian D Marx (1996). "Flexible smoothing with B-splines and penalties". In: *Statistical science*, pp. 89–102.

Keegan, Kirk A et al. (2012). "Active surveillance for prostate cancer compared with immediate treatment". In: *Cancer* 118.14, pp. 3512–3518.

Lang, Stefan and Andreas Brezger (2004). "Bayesian P-splines". In: *Journal of computational and graphical statistics* 13.1, pp. 183–212.

Läuter, E (1976). "Optimal multipurpose designs for regression models". In: *Mathematische Operationsforschung und Statistik* 7.1, pp. 51–68.

Loeb, Stacy et al. (2013). "Systematic review of complications of prostate biopsy". In: *European urology* 64.6, pp. 876–892.

López-Ratón, Mónica et al. (2014). "OptimalCutpoints: an R package for selecting optimal cutpoints in diagnostic tests". In: *Journal of Statistical Software* 61.8, pp. 1–36.

O'Mahony, James F et al. (2015). "The Influence of Disease Risk on the Optimal Time Interval between Screens for the Early Detection of Cancer: A Mathematical Approach". In: *Medical Decision Making* 35.2, pp. 183–195.

Parmigiani, Giovanni (1998). "Designing observation times for interval censored data". In: *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 446–458.

Rizopoulos, Dimitris (2011). "Dynamic Predictions and Prospective Accuracy in Joint Models for Longitudinal and Time-to-Event Data". In: *Biometrics* 67.3, pp. 819–829.

— (2012). *Joint models for longitudinal and time-to-event data: With applications in R*. CRC Press.

— (2014). "The R package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC". In: *arXiv preprint arXiv:1404.7625*.

Rizopoulos, Dimitris et al. (2016). "Personalized screening intervals for biomarkers using joint models for longitudinal and survival data". In: *Biostatistics* 17.1, p. 149. DOI: 10.1093/biostatistics/kxv031. eprint: /oup/backfile/content_public/journal/biostatistics/17/1/10.1093_biostatistics_kxv031/3/kxv031.pdf.

Robert, Christian (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media.

Sokolova, Marina and Guy Lapalme (2009). "A systematic analysis of performance measures for classification tasks". In: *Information Processing & Management* 45.4, pp. 427–437.

Taylor, Jeremy MG et al. (2013). "Real-time individual predictions of prostate cancer recurrence using joint models". In: *Biometrics* 69.1, pp. 206–213.

Tosoian, Jeffrey J et al. (2011). "Active surveillance program for prostate cancer: an update of the Johns Hopkins experience". In: *Journal of Clinical Oncology* 29.16, pp. 2185–2190.

Tsiatis, Anastasios A and Marie Davidian (2004). "Joint modeling of longitudinal and time-to-event data: an overview". In: *Statistica Sinica*, pp. 809–834.

Welty, Christopher J et al. (2015). "Extended followup and risk factors for disease reclassification in a large active surveillance cohort for localized prostate cancer". In: *The Journal of urology* 193.3, pp. 807–811.

A Appendix Heading

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Suspendisse accumsan magna est, quis elementum leo laoreet eu. Donec sollicitudin elit non massa venenatis, in viverra dolor sagittis. Maecenas ac justo pulvinar, consectetur mauris hendrerit, vulputate lacus. Etiam tristique sapien quis sem commodo, et eleifend tortor viverra. In hac habitasse platea dictumst. Phasellus vel tempus risus, sit amet consectetur massa. Duis rutrum lectus eu ligula egestas iaculis. Sed condimentum, ipsum in dignissim condimentum, nisi turpis blandit massa, et aliquam magna ligula eget lacus. Donec ac eleifend nulla, quis cursus nisi. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Suspendisse accumsan magna est, quis elementum leo laoreet eu. Donec sollicitudin elit non massa venenatis, in viverra dolor sagittis. Maecenas ac justo pulvinar, consectetur mauris hendrerit, vulputate lacus. Etiam tristique sapien quis sem commodo, et eleifend tortor viverra. In hac habitasse platea dictumst. Phasellus vel tempus risus, sit amet consectetur massa. Duis rutrum lectus eu ligula egestas iaculis. Sed condimentum, ipsum in dignissim condimentum, nisi turpis blandit massa, et aliquam magna ligula eget lacus. Donec ac eleifend nulla, quis cursus nisi.