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. Because of these reasons prostate cancer patients 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 study worldwide is the PRIAS study (www.prias-project.org) (Bokhorst et al., 2015). Patients 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.

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 Gleason reclassification within 1 year since its manifestation. The drawbacks of this schedule 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 both brachytherapy and 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 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 doubline time (PSA-DT) less than 10 years, repeat biopsy every year is advised. Despite this comparatively less rigorous schedule, it was found that the percentage of men receiving repeat biopsies decreased from 81% at year 1 to 60% in year 4, 53% in year and 33% in year 10 follow up Bokhorst et al., 2015. Some of the common reasons for non-compliance with the biopsy schedule were 'patient does not want biopsy', 'PSA stable', 'complications on last biopsy'and 'no signs of disease progression on previous biopsy'. These reasons indicate that current biopsy scheduling mechanisms require improvement.

Non compliance reduces the effectiveness of AS programs, as progression may get detected late. On the other hand even if patients comply with the schedule, be it annually or the schedule of PRIAS, it is not suitable for every patient. Patients whose cancers progress slowly, often end up having biopsies when they are not needed. For patients who have faster progressions, crude measures such as PSA-DT are used to decide if frequent biopsies should be performed. In contrast to the aforementioned schedules, the most useful biopsy schedule for a patient enumerated i will be the one with the least number of biopsies n_i^b and the smallest offset $O_i = T_i^o - T_i^*$ possible, where T_i^o is the time at which Gleason reclassification is observed and T_i^* is the actual time at which Gleason reclassification occurred. The search for the most useful biopsy schedule is the motivation behind this work. 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 Gleason reclassification and not disease progression or DRE crossing the threshold of cT2c.

To detect Gleason reclassification, in this work 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 health. 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 equispaced 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 interval possible. 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 also utilize joint models and are dynamic. i.e. at a time only one future visit is scheduled. More specifically, we have proposed two types of personalized schedules. One based on expected time of Gleason reclassification of a patient and the second based on the risk of Gleason reclassification. 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 are also used in the personalized schedules, and so is the population level information about hazard of Gleason reclassification.

To this end we chose to work with the joint model since it provides a framework to model the association between PSA measurements and the hazard of Gleason reclassification. Secondly, joint models utilize random effects to model this association, and therefore have an inherent subject specific nature. Further they 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, does not have any side effects, and 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 PSA-DT less than 3 years compared to men who had Gleason > 6 and PSA-DT larger than 3 years. i.e. Information from PSA could be useful in predicting reclassification. Lastly, as mentioned earlier some patients/doctors did not comply with the biopsy schedule because they

considered PSA to be stable. Using the PSA information in a more methodical manner can lead to an even 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 the efficacy of 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 a biopsy every year.

2 A framework for personalized biopsy schedule

The first step in creating a personalized schedule for prostate cancer patients is to come up with a model for Gleason scores, PSA levels and other subject 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 Gleason reclassification takes place. Since our interest also lies in time to Gleason reclassification, we model it as a time to event outcome. i.e. time to Gleason reclassification.

While univariate modeling of the two outcomes can be done as mentioned above, our goal is to also utilize the information from PSA levels in deciding the time for scheduling a biopsy. To achieve this, we jointly model time to Gleason reclassification and PSA levels using a joint model for time to event and longitudinal outcomes Rizopoulos, 2012.

2.1 Joint model for time to Gleason reclassification and PSA levels

Let T_i^* denote the true Gleason reclassification time and C_i denote the censoring time for the i^{th} patient. Let $T_i = \min(T_i^*, C_i)$ denote the observed Gleason reclassification time for the i^{th} patient and $\delta_i = I(T_i^* < C_i)$ is the event indicator. $I(\cdot)$ is the indicator function that takes the value 1 when $T_i^* < C_i$ and 0 otherwise. Let $\boldsymbol{y}_i$ denote the $n_i \times 1$ longitudinal outcome vector for the PSA levels of the i^{th} subject. The population of interest is all the patients enrolled in AS. For a sample of n patients from this population the complete data is denoted by $\mathcal{D}_n = \{T_i, \delta_i, \boldsymbol{y}_i; i = 1, ...n\}$.

Since both PSA levels and Gleason scores are affected by the state of prostate cancer, they are inherently correlated with each other. In the joint model between time to Gleason reclassification and PSA levels, this correlation between the two outcomes is modeled via a vector of time independent random effects $\boldsymbol{b}_i$. The same random effects are also used to model the correlation between the repeated measurements of the PSA levels. Given the observed PSA history, the censoring process is assumed to be non informative and independent of the time to Gleason reclassification process, as well as of future PSA measurements. With these assumptions, the joint distribution between the two outcomes is given by:

$$p(T_i, \delta_i, \boldsymbol{y}_i|\boldsymbol{b}_i; \theta) = p(T_i, \delta_i|\boldsymbol{b}_i; \theta)p(\boldsymbol{y}_i|\boldsymbol{b}_i; \theta), \text{where} \tag{1}$$

$$p(T_i, \delta_i|\boldsymbol{b}_i; \theta) = h_i(T_i|\mathcal{M}_i(T_i); \theta)^{\delta_i} e^{\int_0^{T_i} h_i(s|\mathcal{M}_i(s); \theta)\, ds}, \text{and} \tag{2}$$

$$p(\boldsymbol{y}_i|\boldsymbol{b}_i; \theta) = \prod_{j=1}^{n_i} p(y_{ij}|\boldsymbol{b}_i; \theta) \tag{3}$$

$\theta = (\theta_t^T, \theta_y^T, \theta_b^T)^T$ denotes the full parameter vector, with θ_t denoting the parameters for the time to Gleason reclassification outcome, θ_y the parameters for the PSA levels, and θ_b the parameters of the random-effects covariance matrix $\boldsymbol{D}$. $\mathcal{M}_i(t) = \{m_i(s), 0 \leq s \geq t\}$ denotes the PSA longitudinal history up to time t. It is important to note that $m_i(t) = \boldsymbol{x}_i^T(t)\boldsymbol{\beta} + \boldsymbol{z}_i^T(t)\boldsymbol{b}_i$ represent the true value of PSA at time t and not the observed $y_i(t)$. $\boldsymbol{x}_i^T(t)$ and $\boldsymbol{z}_i^T(t)$ are the time dependent

design vectors for the fixed effects $\boldsymbol{\beta}$ and random effects $\boldsymbol{b}_i$, respectively.

The hazard of Gleason reclassification at any time t, given by $h_i(t|\mathcal{M}_i(t);\theta)$ depends on the PSA history $\mathcal{M}_i(t)$ up to the time t. Joint models offer flexibility in modeling this dependence. In its simplest form, the hazard may depend on instantaneous value of PSA $m_i(t)$ at time t. It is also possible to consider that hazard of Gleason reclassification at time t may depend on PSA velocity $m_i'(t) = \dfrac{dm_i(t)}{dt}$ or $AUC(t) = \int_0^t m_i(s)\,ds$ at the same time. The fact that any functional form of dependence is possible, is evident from the following equation:

$$h_i(t|M_i(t);\theta) = h_0(t)e^{\gamma^T w_i + f\{M_i(t),\boldsymbol{b}_i,\alpha\}} \tag{4}$$

where the function $f(\cdot)$, parametrized by vector α specifies which features of longitudinal outcome process are are included in the linear predictor of the time to event model. For the case at hand, we consider that hazard of Gleason reclassification at any time depends on the PSA value as well as the PSA velocity at the same time. The corresponding hazard function is given by:

$$h_i(t|M_i(t);\theta) = h_0(t)e^{\gamma^T w_i + \alpha_1 m_i(t) + \alpha_2 m_i'(t)} \tag{5}$$

where, α_1 and α_2 are measures of how strongly hazard of Gleason reclassification depends on PSA value and PSA velocity, respectively. w_i is a vector of time independent covariates and γ are the corresponding parameters. $h_0(t)$ is the baseline hazard at time t, and is modeled flexibly using a B-splines approach. More specifically:

$$\log h_0(t) = \gamma_{h_0,0} + \sum_{q=1}^{Q} \gamma_{h_0,q} B_q(t,v) \tag{6}$$

where $B_q(t,v)$ denotes the q^{th} basis function of a B-spline with knots $v = v_1, ...v_Q$ and γ_{h_0} the vector of spline coefficients. 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 Fitting the joint model to PRIAS dataset

One of the goals of this work is to compare various scheduling methods for patients of the PRIAS study. We also compare the scheduling methods using a simulation study. For both of these goals, we require parameter estimates of the joint model between time to Gleason reclassification and PSA levels of PRIAS' patients. For the first goal the parameter estimates are used directly to schedule visits using personalized schedules. For the second goal, we require the parameter estimates of the joint model to generate data for a simulation study. For these reasons and to also introduce constructs such as dynamic survival probability, we next fit a joint model to the PRIAS data set.

The PRIAS data set contains information about 5943 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. Other information, namely, PSA levels and Gleason scores were measured at different time points on the basis of a predetermined schedule as mentioned earlier in the report. For the longitudinal analysis of PSA measurements we used $\log_2(PSA + 1)$ measurements instead of the raw data. The log transformation was done because the PSA scores took very large values at the onset of disease progression. This indicated that the underlying distribution for PSA scores was right skewed. Secondly at certain time points patients' PSA scores were measured to be 0, and thus we had to add a constant 1 to all PSA scores before log transformation. We found the same transformation in literature as well McGreevy et al., 2006; Sène et al., 2016. For the variable Age in the Cox model, we wanted to consider both the linear and quadratic effect. However doing so led many numerical instabilities while performing the joint model analysis. To avoid these pitfalls we had to center the Age variable around age 70 years. The longitudinal and survival submodels of the joint model we fitted is given by:

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

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. For the survival submodel the hazard function we fitted is given by:

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

The baseline hazard $h_0(t)$ is modeled as in equation 6. To fit the joint model we use the R package JMbayes Rizopoulos, 2014, which uses a Bayesian approach for parameter estimation. The parameter estimates for the survival submodel were estimated using Bayesian ridge methodology.

2.2.1 Parameter Estimates

The parameter estimates for the joint model we fitted to the PRIAS data set are shown in Table 1 and Table 2. Since the longitudinal evolution of $\log_2(PSA+1)$ 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 1. 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.

	Mean	Std. Dev	2.5%	97.5%	P
Intercept	2.717	0.008	2.701	2.733	<0.000
(Age - 70)	0.003	0.001	0.001	0.005	0.002
(Age - 70) × (Age - 70)	-0.001	1 $\times 10^{-4}$	-7 $\times 10^{-4}$	-4 $\times 10^{-4}$	<0.000
Spline: visitTimeYears[0, 0.5]	0.026	0.008	0.012	0.042	<0.000
Spline: visitTimeYears[0.5, 1.2]	0.208	0.013	0.184	0.233	<0.000
Spline: visitTimeYears[1.2, 2.5]	0.175	0.019	0.137	0.210	<0.000
Spline: visitTimeYears[2.5, 7]	0.309	0.028	0.256	0.366	<0.000
σ	0.273	0.001	0.271	0.275	<0.000

Table 1: Longitudinal submodel estimates for joint model.

For the survival submodel, the parameter estimates in Table 2 show that both the $\log_2(PSA+1)$ value and $\log_2(PSA+1)$ velocity are associated with time to Gleason reclassification. The effect is quite strong and if at any given time point the PSA becomes approximately 4 times of its value then the hazard of Gleason reclassification becomes 1.5 times of the original. This is valid under the condition that the $\log_2(PSA+1)$ velocity remains the same. The effect of $\log_2(PSA+1)$ velocity is far stronger, but it is not interpretable easily. Lastly, for the effect of Age on hazard we can say that it can be safely ignored for all practical purposes.

2.2.2 Dynamic survival probability

Since patients' PSA levels and Gleason scores are periodically measured, the entire PSA and repeat biopsy history can be used to periodically update the predictions about about time to

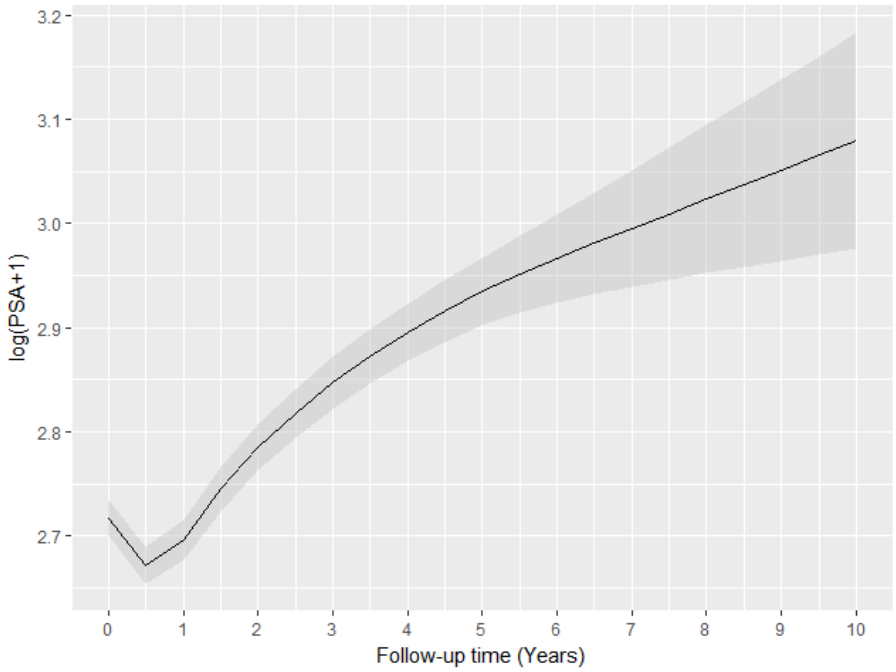


Figure 1: Fitted evolution of $\log_2(PSA+1)$ over a period of 10 years, for a patient who was inducted in AS at the Age of 70 years.

Variable	Mean	Std. Dev	2.5%	97.5%	P
Age - 70	0.036	0.007	0.023	0.050	<0.000
(Age - 70) × (Age - 70)	-0.002	0.001	-0.004	2×10^{-4}	0.016
$\log_2(PSA+1)$	0.184	0.093	0.016	0.369	0.032
Slope: $\log_2(PSA+1)$	1.937	0.278	1.420	2.525	<0.000

Table 2: Survival submodel estimates for joint model.

Gleason reclassification. More specifically, let us assume that a new patient numbered j, not present in the original sample of patients $\mathcal{D}_n$, did not have a Gleason reclassification at their last biopsy, performed at time t. Further, let us assume that the PSA measurements are available for the patient j up to a time point $s > t$. Then combining these two pieces of information, the distribution for time to Gleason reclassification for this patient is given by:

$$p(T_j^* | T_j^* > t, \mathcal{Y}_j(s), \mathcal{D}_n; \theta^*) \tag{8}$$

where $\mathcal{Y}_j(s) = \{y_j(r); 0 \leq r \leq s\}$ denotes the history of PSA measurements done up to time s. The personalized schedules we propose in this work depend on this dynamic distribution for time to Gleason reclassification. As more PSA measurements are taken and repeat biopsies are performed, this distribution is accordingly updated. The survival probabilities based on this distribution are called dynamic survival probabilities Rizopoulos, 2011. The dynamic survival probability at any time point $u > s$ is given by:

$$\pi_j(u|t) = Pr(T_j^* \geq u | T_j^* > t, \mathcal{Y}_j(s), D_n; \theta) \tag{9}$$

As an example we present the dynamic survival probability of patient 2362 from the PRIAS data set. Patient 2362 had his last repeat biopsy at 3.78 years at which the Gleason reclassification did not happen. The patient was later censored at 5.12 years. Using all this information up to 5.12 years, the dynamic survival probability at any time point $u > 3.78$ for this patient is given by the formula $Pr(T_{2362}^* \geq u | T_{2362}^* > 3.78, \mathcal{Y}_{2362}(5.12), D_n; \theta)$. Figure 2 shows the dynamic survival probability for patient 2362 at time points between 5.12 and 8.12 years.

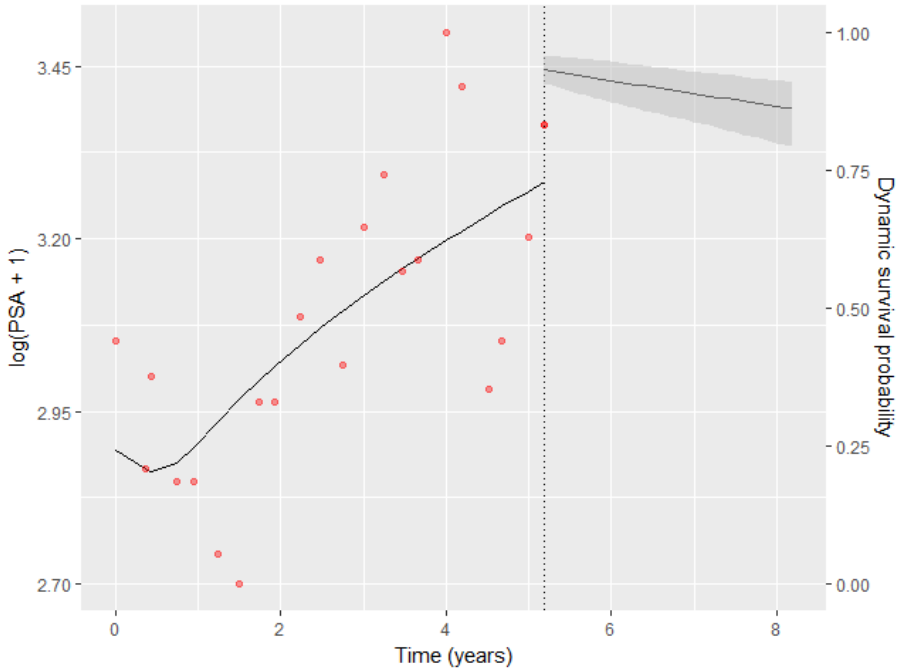


Figure 2: Dynamic survival probability for patient with ID 2362 in the PRIAS data set.

3 Personalized scheduling approaches

Once a joint model for Gleason reclassification and PSA levels is obtained, the next step is to use it to create personalized schedules for biopsies. In this section we present the various personalized biopsy scheduling approaches and their motivation. The personalized schedules that we propose are dynamic in nature and thus at any given time, only 1 future biopsy is scheduled. The age of the patient and entire PSA, repeat biopsy history up to that time point is considered while computing the time of next biopsy. To elucidate the scheduling methods, we use patient j, introduced in detail in section 2.2.2, as a case.

3.1 Conditional expected time to Gleason reclassification

Given that the patient j hadn't had Gleason reclassification up to time t, an estimate of true Gleason reclassification time, with zero bias and least squared loss is the conditional expected Gleason reclassification time: $E(T_j^*|T_j^* > t, \mathcal{Y}_j(s), D_n; \theta)$. The aforementioned two properties are also the motivation for using conditional expected Gleason reclassification time as the time of next biopsy.

$$E(T_j^*|T_j^* > t, \mathcal{Y}_j(s), D_n; \theta) = t + \int_t^\infty \pi_j(u|t)\, du \tag{10}$$

where, $\pi_j(u|t)$ is the dynamic survival probability (Section 2.2.2) for patient j at time u. While expected time to Gleason reclassification is unbiased, it is more useful when the variance of time to Gleason reclassification, given by $Var(T_j^*|T_j^* > t, \mathcal{Y}_j(s), \mathcal{D}_n; \theta)$, is less. The variance is given by:

$$Var(T_j^*|T_j^* > t, \mathcal{Y}_j(s), \mathcal{D}_n; \theta) = 2\int_t^\infty (u-t)\pi_j(u|t)\, du - \left(\int_t^\infty \pi_j(u|t)\, du\right)^2 \tag{11}$$

3.2 Dynamic risk of failure

When time to Gleason reclassification has a large variance (Eq. 11) then the offset (Section 1) based on using conditional expected time to Gleason reclassification can vary wildly from patient to patient. In such a scenario a doctor/patient may want to perform biopsies earlier than the expected time to Gleason reclassification. This can be achieved by not crossing the risk of Gleason

reclassification beyond a certain threshold.

The personalized scheduling approaches based on dynamic risk of reclassification, schedule biopsies at time points $u > t$ such that the dynamic risk of failure $1 - \pi_j(u|t)$ is higher than a certain threshold $1 - \kappa$ beyond u. Or in other words the dynamic survival probability $\pi_j(u|t)$ is below a threshold κ beyond u. The choice of the threshold $\kappa \epsilon [0, 1]$ can be on the basis of amount of risk a patient or doctor is willing to take. It is also possible to automate the choice of κ. One such way is to choose a κ for which a binary classification accuracy measure is maximized. In case of joint models, time dependent binary classification is more relevant because a patient can be in control group at some time t_a and it can be in the cases at some future time point $t_b > t_a$. Based on Rizopoulos, 2014 we consider a subject j to be a case if $\pi_j(t + \Delta t|t) \leq \kappa$ and a control if $\pi_j(t + \Delta t|t) > \kappa$. The time window Δt can be either chosen on a clinical basis (such as 1 year in PRIAS) or it can be chosen at a point where $AUC(t, \Delta t)$ Rizopoulos, 2014 is largest. i.e. we chose a Δt for which the model has the most discriminative capability at time t. Various binary classification accuracy measures can be maximized to select the cutoff κ. The ones we use in this report are:

- Max Sensitivity: $\arg\max_{\kappa \epsilon [0,1]}$ Senstivity,
  where sensitivity is $P(\pi_j(t + \Delta t|t) \leq \kappa | T_j^* \epsilon (t, t + \Delta t])$.
- Youden J Statistic: Sensitivity + Specificity - 1,
  where the specificity is defined as $P(\pi_j(t + \Delta t|t) > \kappa | T_j^* > t + \Delta t)$.
- Accuracy (ACC): $\frac{TP(t) + TN(t)}{TP(t) + FP(t) + TN(t) + FN(t)}$, where TP(t), FP(t), TN(t) and FN(t) are the number of true positives, false positives, true negatives and false negatives at time point t.
- F1 Score: $\frac{2TP(t)}{2TP(t) + FP(t) + FN(t)}$, and it is derived by taking harmonic mean of precision and sensitivity.

We also compare κ chosen by these automatic selection methods against a fixed $\kappa = 0.85$. The choice of 0.85 was arbitrary, assuming a scenario where a patient/doctor does not want to exceed 15% risk of progression.

4 Personalized schedules for patients in PRIAS

As a first step in demonstrating how the personalized schedules work, we created a biopsy schedule for patients in PRIAS. To this end, we divided the PRIAS data set into training(5938 subjects) and demonstration data sets (5 subjects). We demonstrate that the biopsy schedule depend on subject specific traits and the evolution of PSA scores.

It can be seen in Figure 3 that the patient 3174 had a biopsy at the time of induction and none after that. The PSA for the patient increases rapidly after the 2nd year. In response to this rapid increase, the proposed biopsy times based on conditional expected failure time also decrease accordingly from 11 years to around 3.4 years. The change for risk based methods is not so drastic though. Further it can be seen that at the last visit for PSA, the proposed biopsy time is earlier than the last 5 PSA visits and similar is the case for the proposed biopsy time based on dynamic risk of failure. This is due to the fact that the last time of biopsy was time 0 (induction time) and thus the time to Gleason reclassification can take any value larger than 0. We discuss this issue in detail in section 5.1.

Figure 4 shows the PSA evolution and biopsy times for subject 911. It can be seen that this patient had 3 biopsies where Gleason reclassification did not happen. At year 2 when the patient's PSA increases rapidly the proposed failure times also decrease, whereas they increase over the next 1 year because the PSA also drops down in that time period. The fact that PSA velocity affects the biopsy times the most is also evident in the case of subject 2340, whose evolution is shown in Figure 5. Here the rate of change at each time point is not high, and even though the PSA

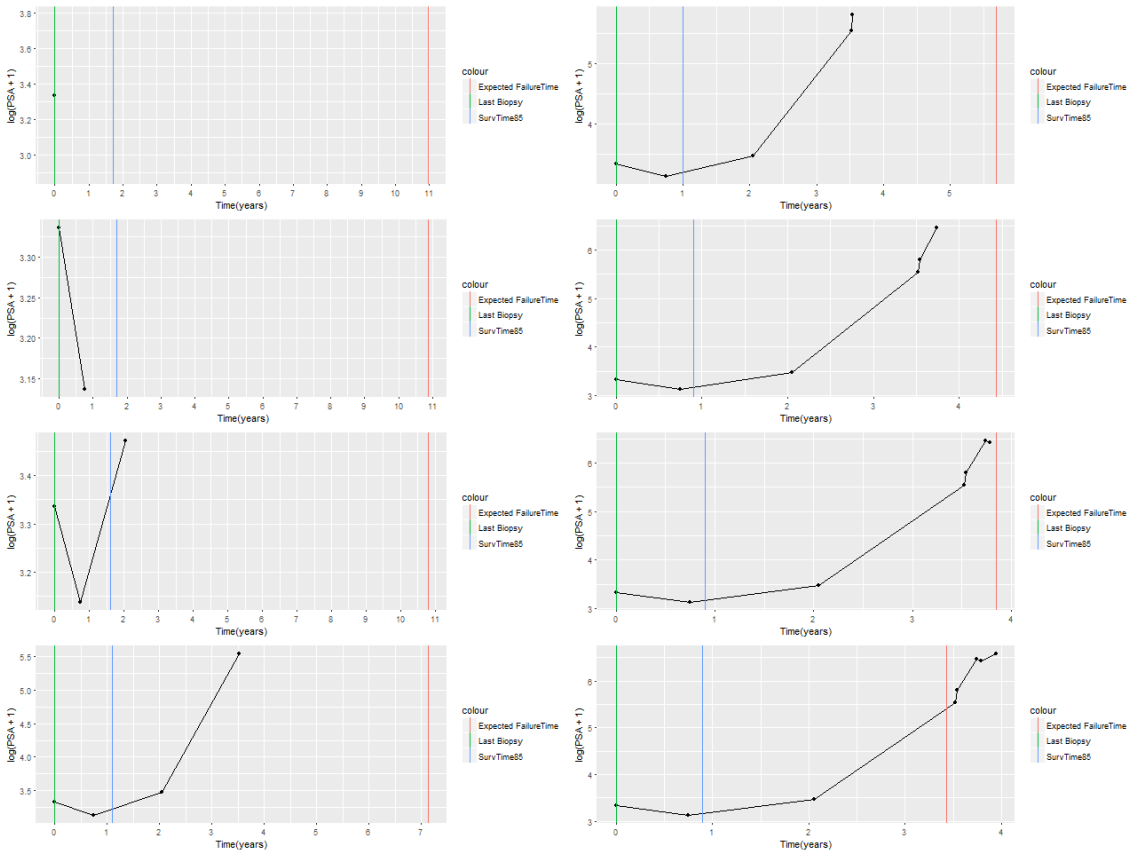


Figure 3: Proposed biopsy times for patient 3174 from PRIAS.

value reaches as high as 25 it has no effect on proposed biopsy times. This is in accordance with the estimated strength of association between PSA velocity/value and hazard of time to Gleason reclassification.

An interesting observation we made while creating these schedules was that the variance of time to Gleason reclassification (Eq. 11) was quite high, which essentially rules out the usefulness of conditional expected time to Gleason reclassification. Given the large difference in proposed biopsy times based on the former and methods based on dynamic risk of Gleason reclassification, one might conclude that the latter are more useful. However as we will see in the simulation study (Section 5) ahead, the usefulness of the two categories of methods depends on the distribution of time to Gleason reclassification.

5 Simulation study

For contrasting the personalized scheduling approaches against the schedule of PRIAS and the schedule of doing biopsy every year, we performed a simulation study. We used the parameters from the joint model fitted to the original PRIAS data set to generate 60 data sets with 1000 patients each. We employed a Weibull baseline hazard while generating Gleason reclassification times for the simulated patients. We divided the 60 data sets into groups of 10 each, and within each group the parameter values for the Weibull baseline hazard were kept the same, while across the group they were different. The reason to create such groups was to ensure the consistency of results within in each group. Using different parameter values for the baseline hazard across the groups ensured that we tested not only the scenarios in which patients have early progression times but also those in which patients have late progression times. This is important because how late/early patients are observed to have Gleason reclassification, partially depends on the induction criteria of the AS program.

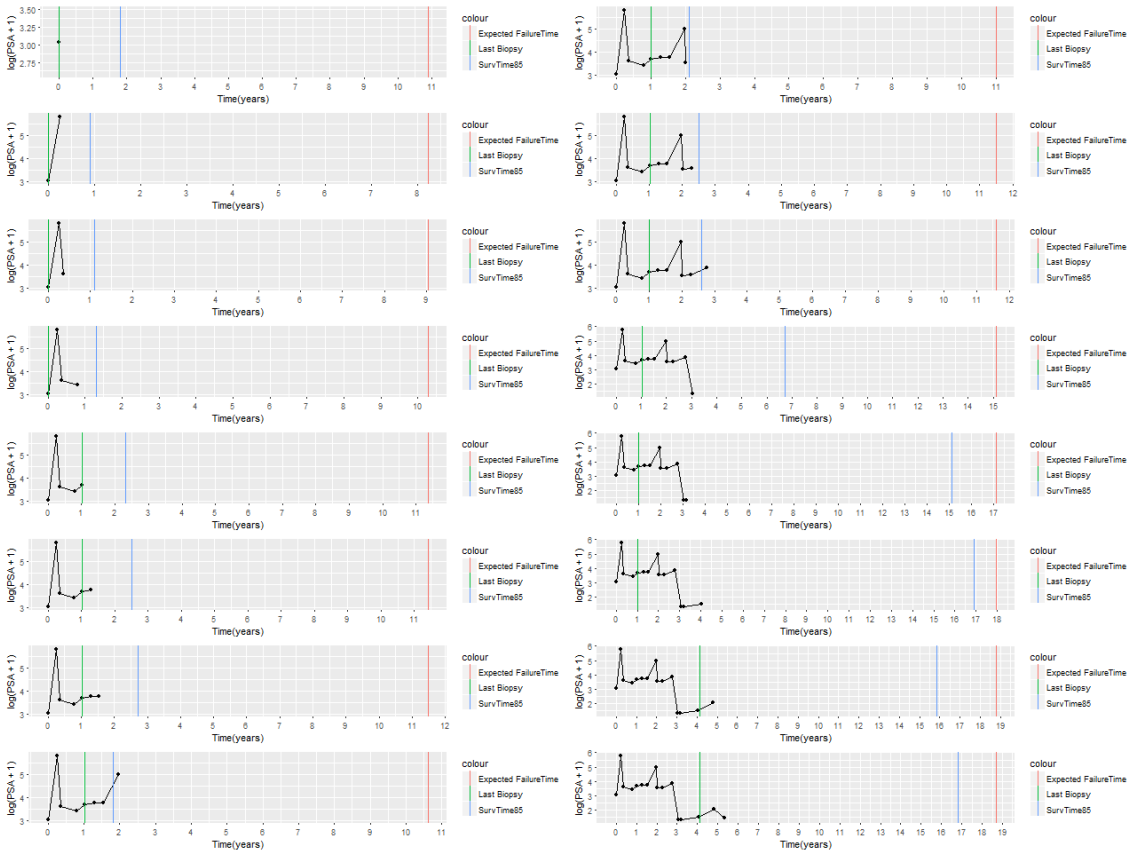


Figure 4: Proposed biopsy times for patient 3174 from PRIAS.

5.1 Simulation setup

We divided each of the 60 data sets into training (750 patients) and test (250 patients) parts. We fitted a joint model to the training data set and using it we generated biopsy schedules for patients in the test data set. Since we generate the true Gleason reclassification time for each of the patients in the test data set, we are able to calculate the offset for each patient and each scheduling method. While creating the personalized schedules we adhere to the PSA measurement schedule of PRIAS. On every visit for the PSA we calculate the time of conducting a biopsy using the scheduling methods given in section 3. We call it the proposed time of biopsy. We observed in section 4 that in many cases the proposed time of biopsy was postponed on each visit for PSA. In some cases two biopsies within a 1 year period were also proposed, which is not practical. Lastly, since the existence of PSA does not depend on Gleason reclassification, in some cases the proposed time was also earlier than the visit time. To avoid these problems altogether we use the following algorithm to select the time of biopsies from the proposed times.

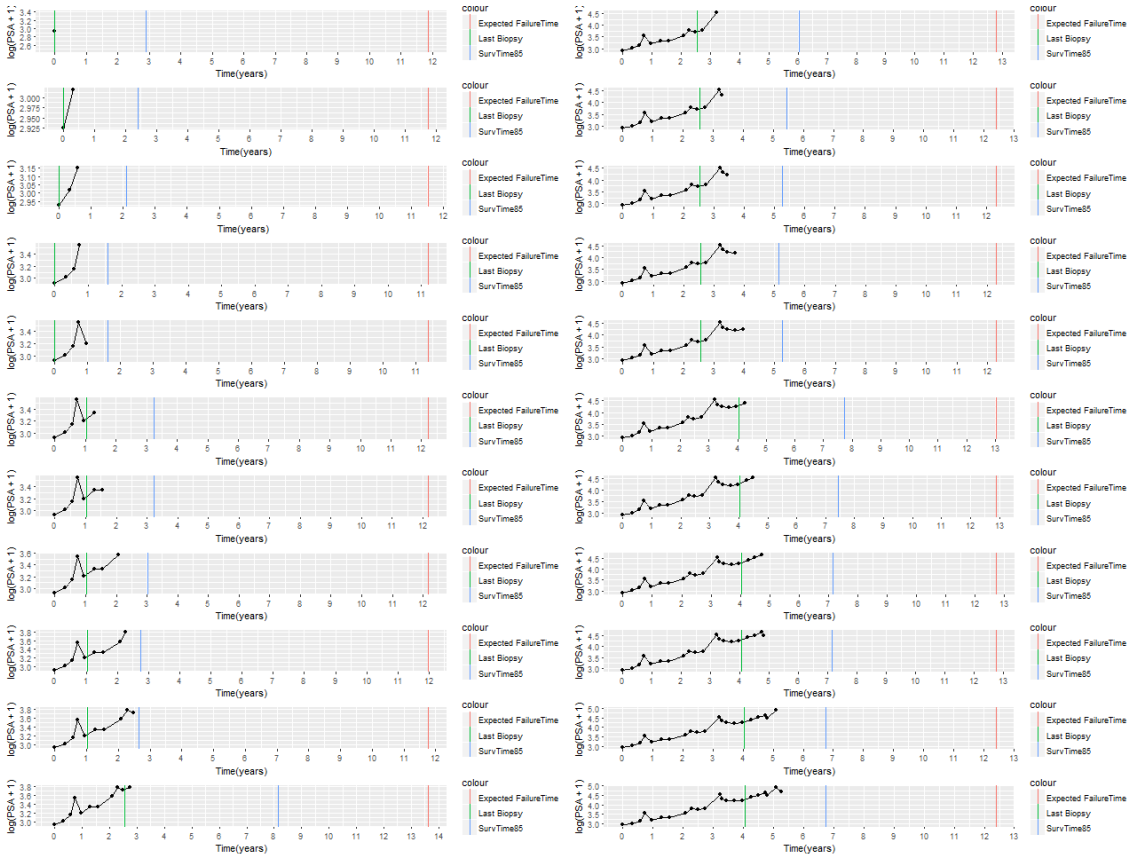
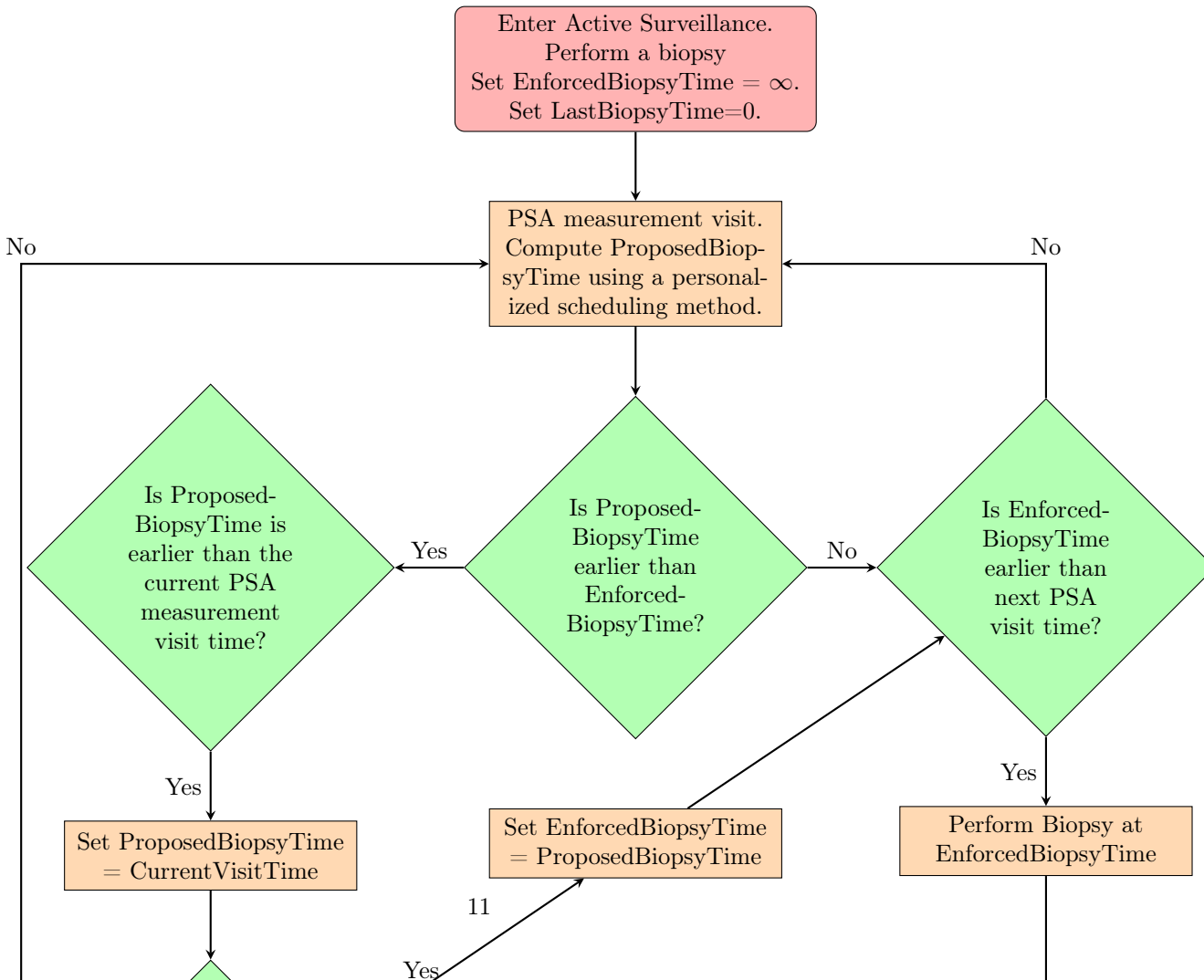


Figure 5: Proposed biopsy times for patient 3174 from PRIAS.



5.2 Results

Since the Gleason reclassification times varied widely across the simulation data sets, we discuss the results separately for the different scenarios. In total we have 6 scenarios corresponding to the 6 groups of 10 data sets. For brevity, we only chose 3 most dissimilar scenarios and further we present one set of results for each of the 3 scenarios. The complete set of results can be found at https://goo.gl/u6dg8G. At this moment, we will not compare our results against the PRIAS schedule, since the PRIAS schedule we used is fixed for all patients, whereas in reality it can be dynamic if the PSA doubling time is less than 10 years.

5.2.1 Scenario 1

In the first scenario the Gleason reclassification times for the patients are mostly between 7 and 11 years. The Gleason reclassification times of test patients from one of the data sets is shown in Figure 6. Figure 9 shows that, for this data set the approaches with the least number of biopsies, 2 in number, are conditional expected failure time, and dynamic risk with a threshold κ chosen such that accuracy (section 3.2) is maximized. The mean and median offset for the two approaches is around 9 months. However for a few outlying patients the offset is more than 24 months. If we compare this to the schedule of performing biopsy every year the mean offset is around 6 months and mean number of biopsies is 8.

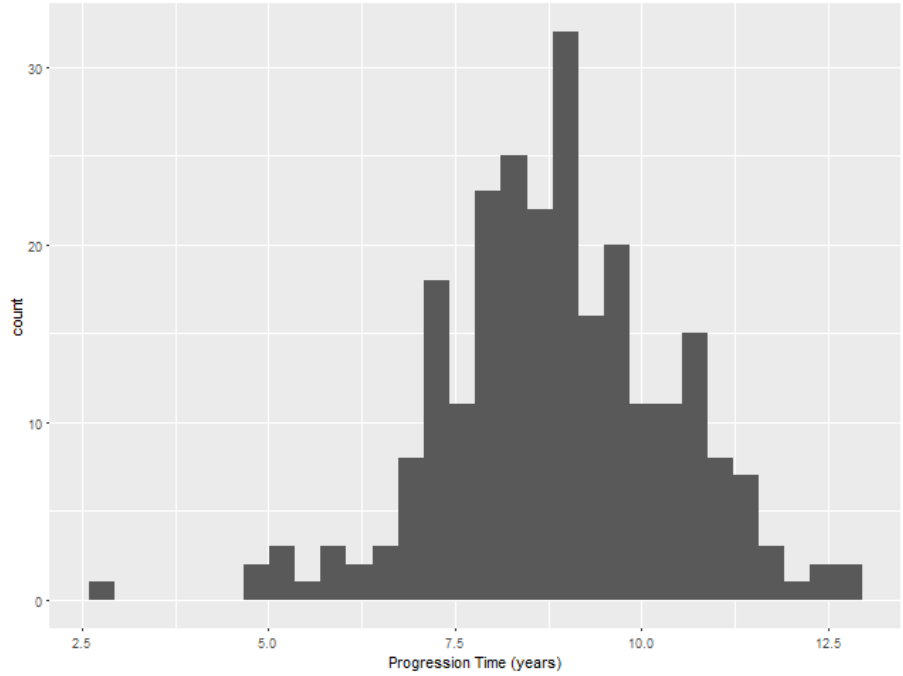


Figure 6: Gleason reclassification times for patients from the test data set of scenario 1.

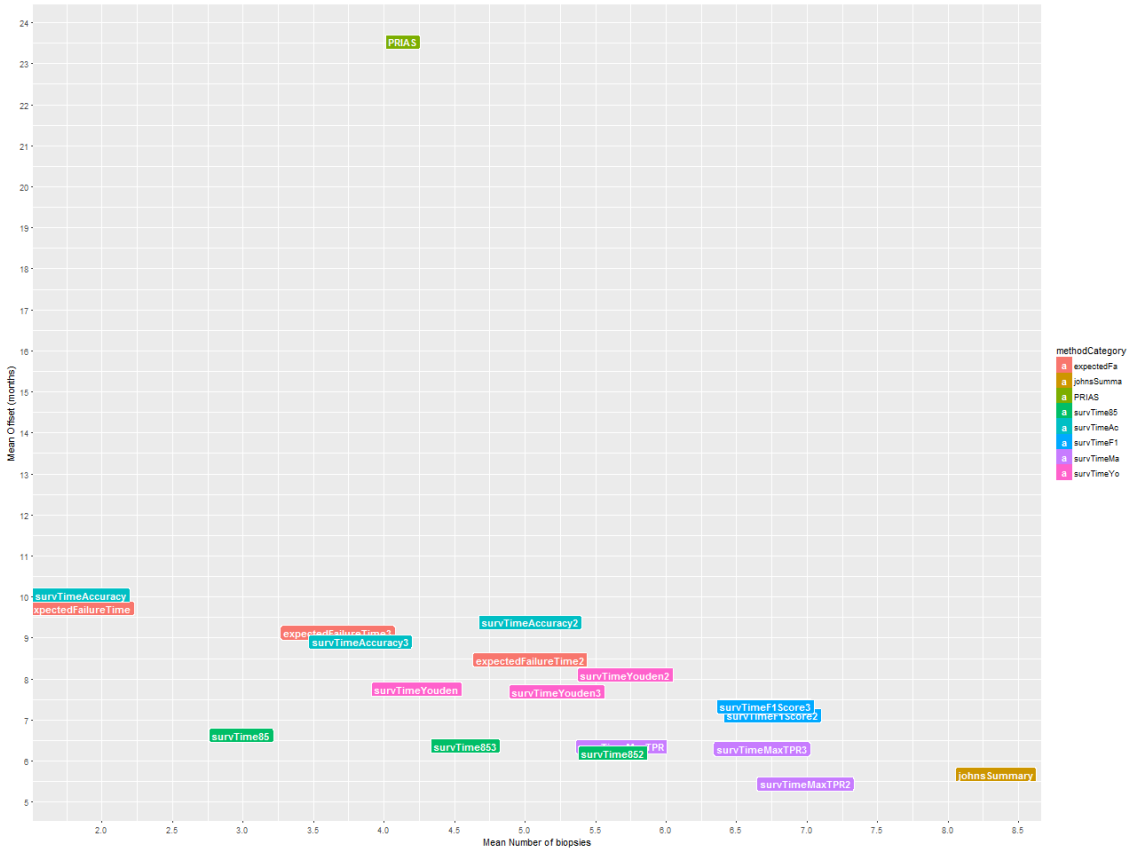


Figure 7: Mean number of biopsies against the mean offset (in years) for each of the approaches in scenario 1.

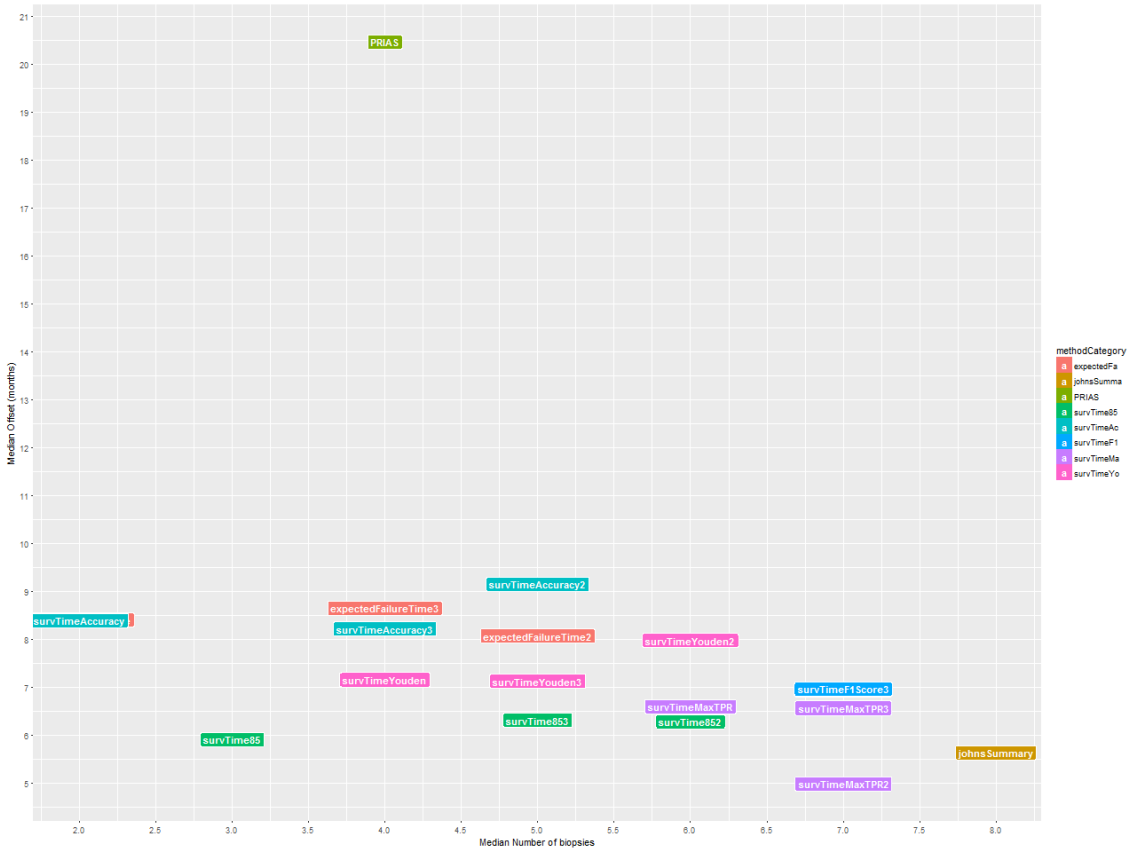


Figure 8: Median number of biopsies against the median offset (in years) for each of the approaches in scenario 1.

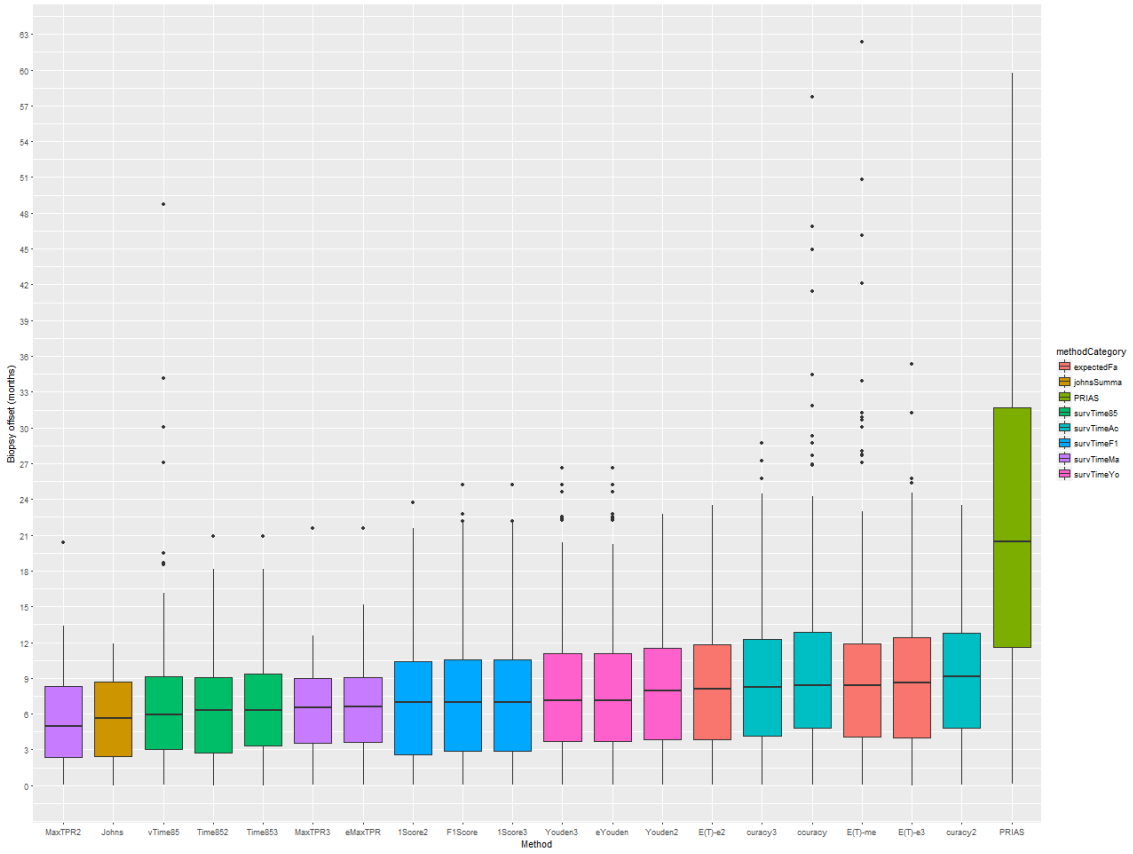


Figure 9: Boxplot for the offset corresponding to the various approaches in scenario 1.

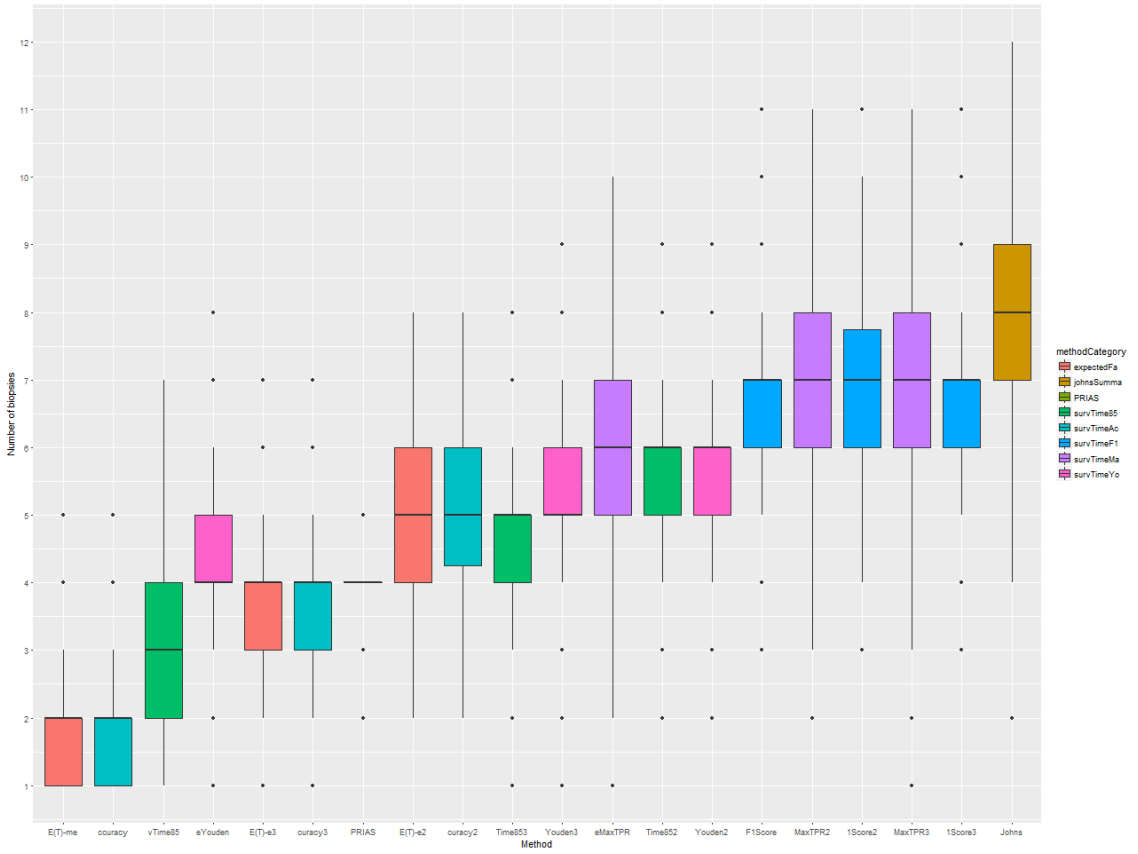


Figure 10: Boxplot for the number of biopsies corresponding to the various approaches in scenario 1.

5.2.2 Scenario 2

Unlike the previous scenario, the Gleason reclassification times for the patients in Scenario 2 are mostly between 0 and 5 years. The Gleason reclassification times of test patients from one of the data sets is shown in Figure 11. Figure 14 shows that, for this data set the approach with the least number of biopsies is conditional expected failure time. The mean and median number of biopsies are 1.5 and 1 respectively. While the mean and median offset is around 18 months, there is considerable variation in the offset for the various patients. The first and third quartiles are at 12 months and 27 months respectively. The large variation in offset is due to the fact that time to Gleason reclassification has larger variance since it is based on a small history of the patient. In such a scenario one might be inclined towards doing a biopsy every year as they have the least offset. However that approach leads to a very large number of biopsies as shown in Figure 15. Certain approaches based on dynamic risk perform better here, both in terms of number of biopsies as well as the offset. In particular if the choice of κ can be based on maximization of the Youden index or a fixed κ can be chosen.

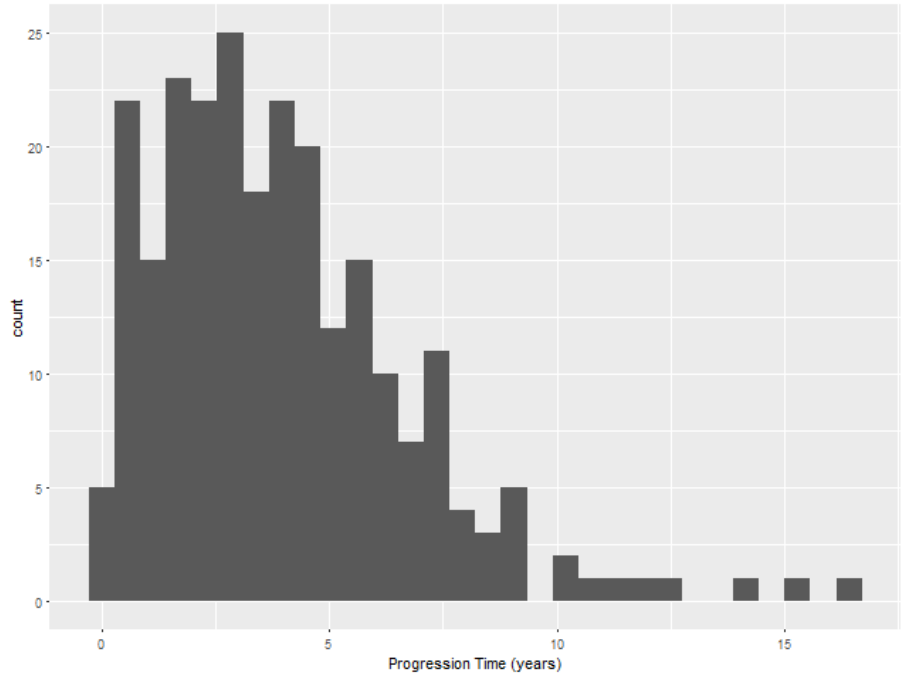


Figure 11: Gleason reclassification times for patients from the test data set of scenario 2.

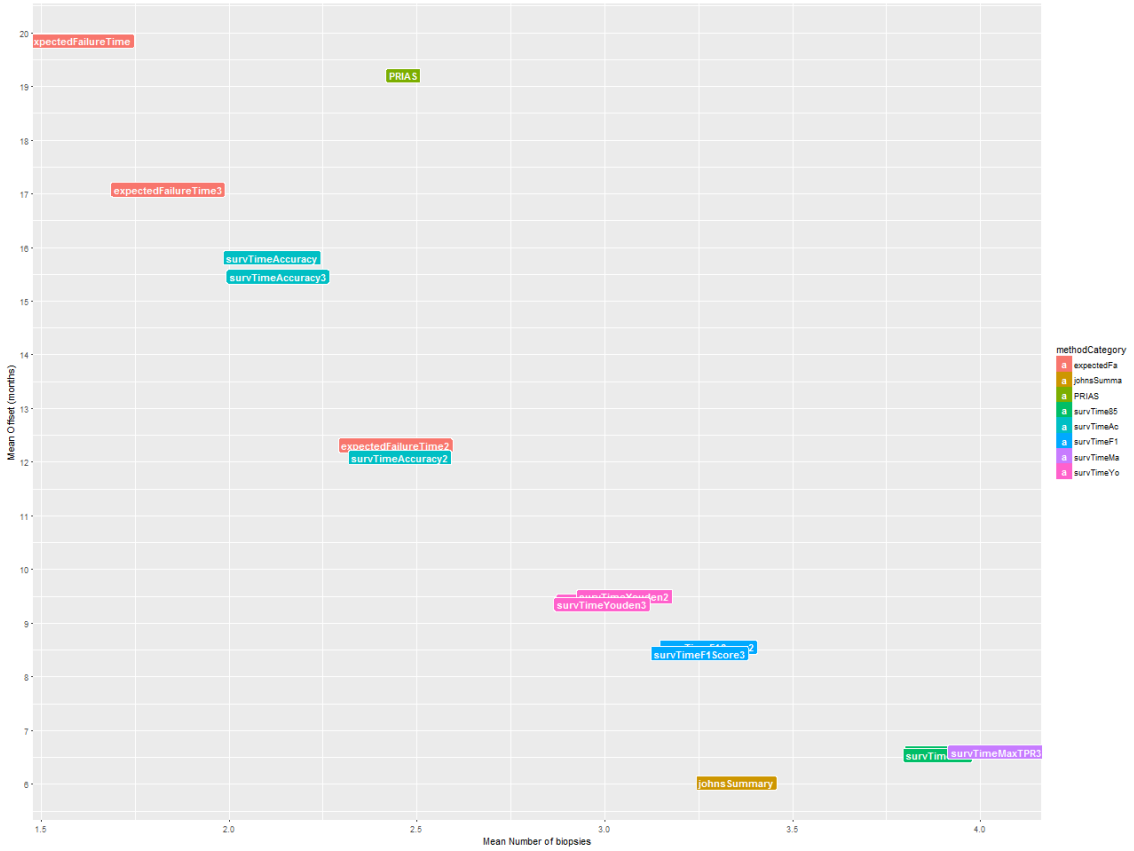


Figure 12: Mean number of biopsies against the mean offset (in years) for each of the approaches in scenario 2.

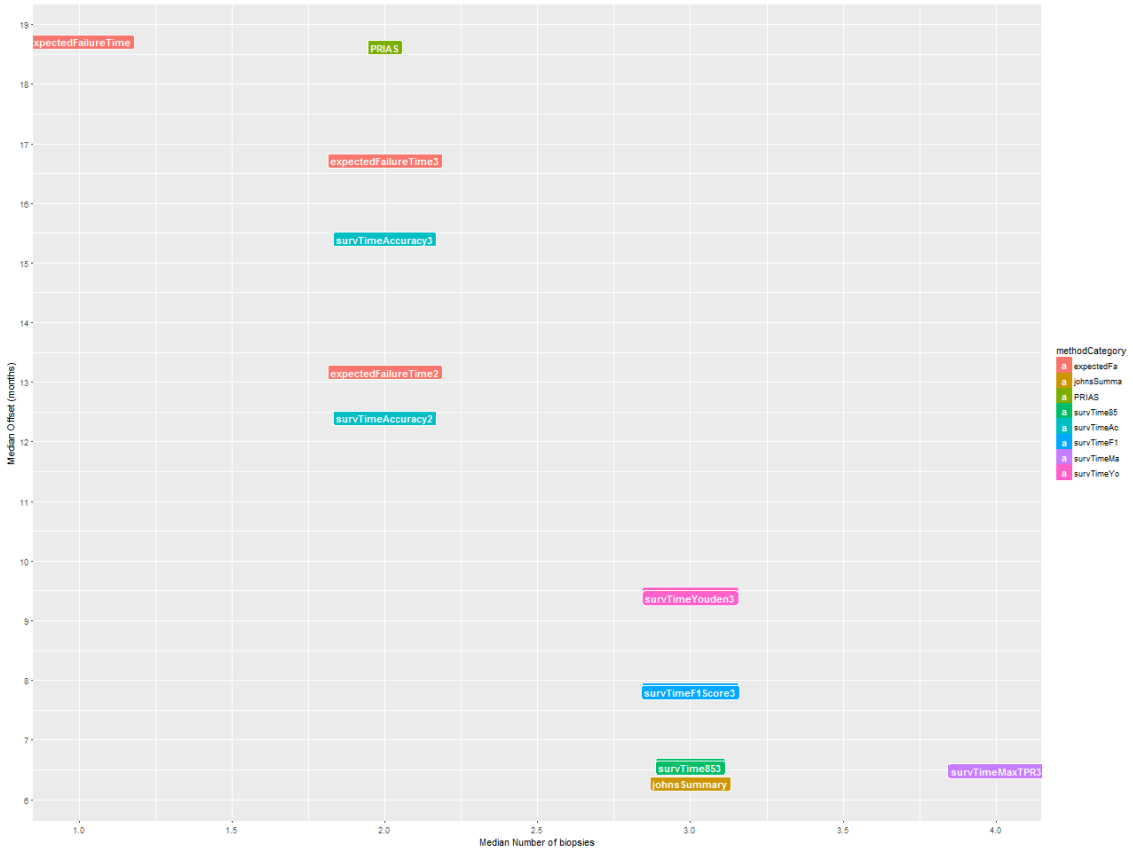


Figure 13: Median number of biopsies against the median offset (in years) for each of the approaches in scenario 2.

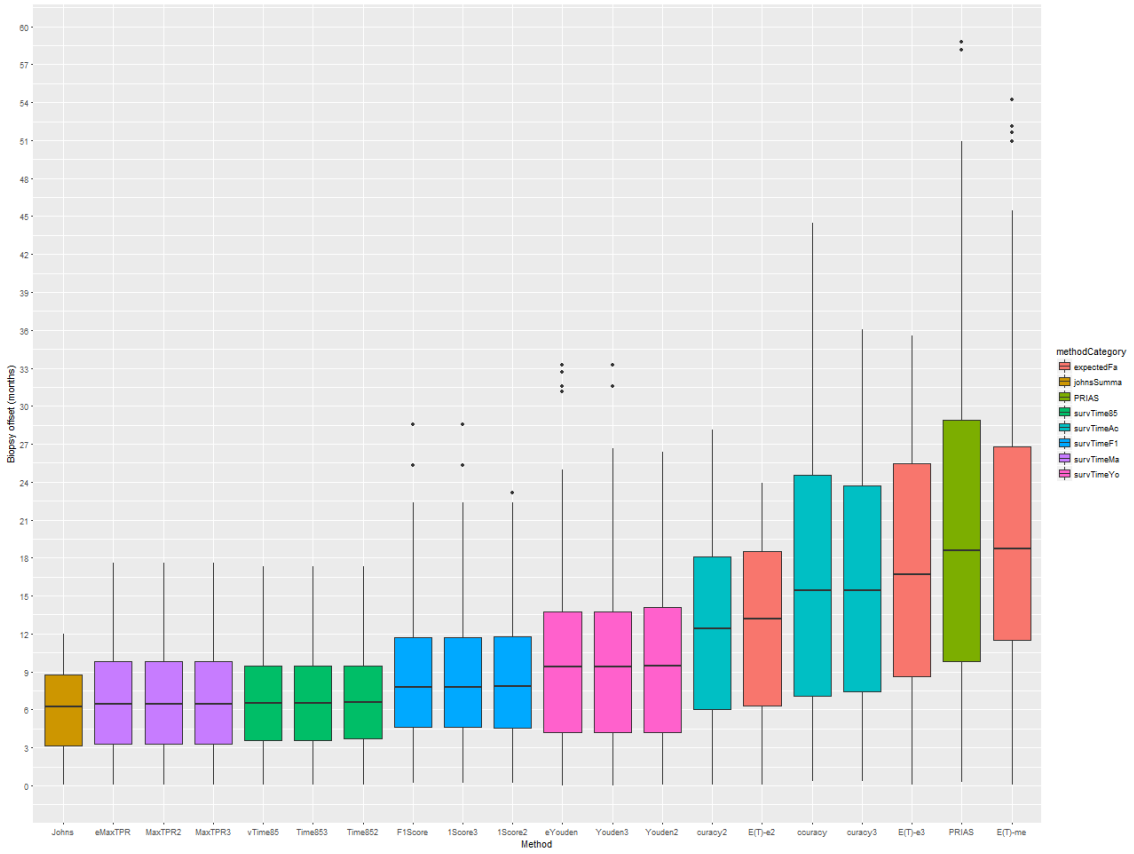


Figure 14: Boxplot for the offset corresponding to the various approaches in scenario 2.

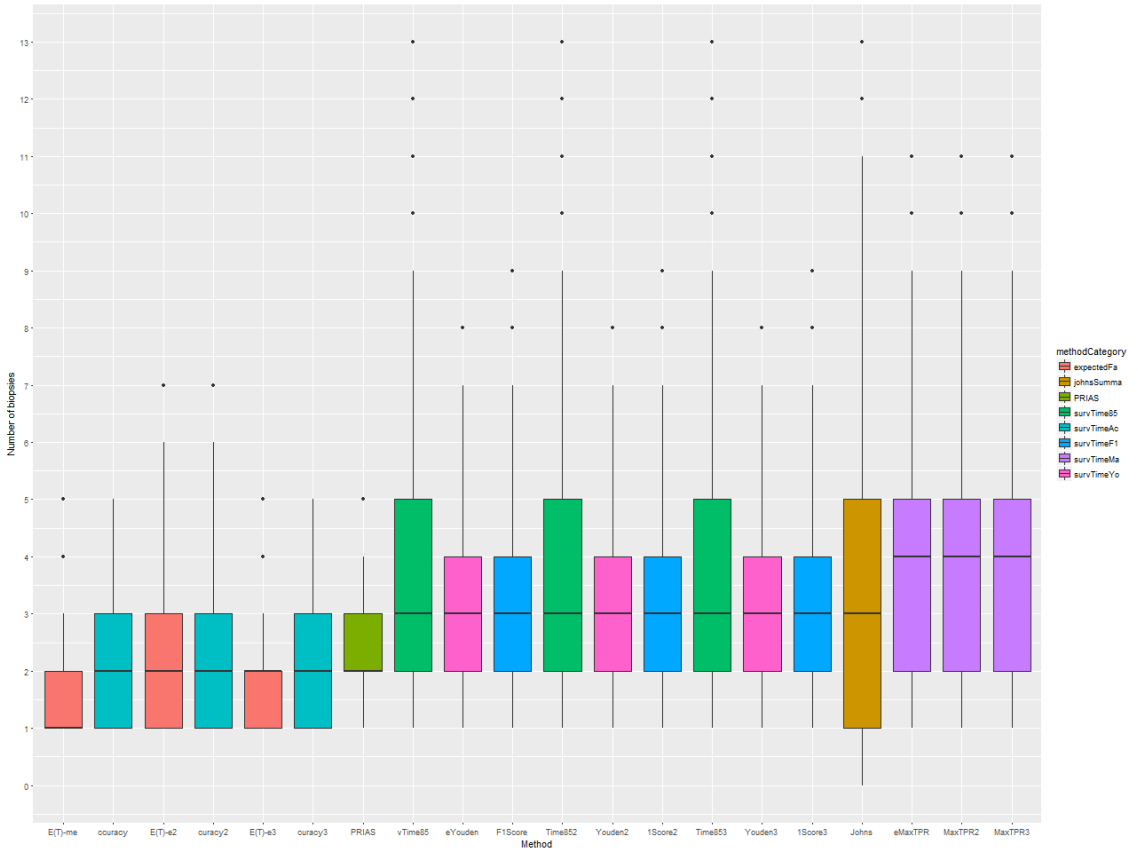


Figure 15: Boxplot for the number of biopsies corresponding to the various approaches in scenario 2.

5.2.3 Scenario 3

In scenario 3, the Gleason reclassification times for the patients are widely spread, mostly between 2 and 9 years. i.e. patients fail late as well early. The Gleason reclassification times of test patients from one of the data sets is shown in Figure 16. Figure 19 shows that, for this data set the approach with the least number of biopsies is conditional expected failure time. The mean and median number of biopsies are 1.5 and 1 respectively. This method also has considerable variation in the offset for the various patients. The first and third quartiles are at 9 months and 27 months respectively. A more detailed analysis revealed that the variation in offset was higher for subjects with earlier failure times. For e.g. the mean and median offset in subjects with reclassification times more than 4 years was nearly 13.5 months. For others it was nearly 39.5 months. This is once again is due to the fact that time to Gleason reclassification has larger variance since it is based on a small history of the patient. And thus usefulness of conditional expected failure time is questionable.

Dynamic risk based methods once again provide an alternative. If a fixed κ of 0.15 is used then although the offset will be low, there is very high variation in number of biopsies. i.e. people who have reclassifications early will benefit but people who have reclassifications later will have too many biopsies. If κ is chosen such that accuracy is maximized, and a biopsy is done whenever there is a gap of 2 years then a reasonable offset is obtained. The first and third quartiles are almost 4.5 and 18 months. The first and third quartiles for number of biopsies are 3 and 4.

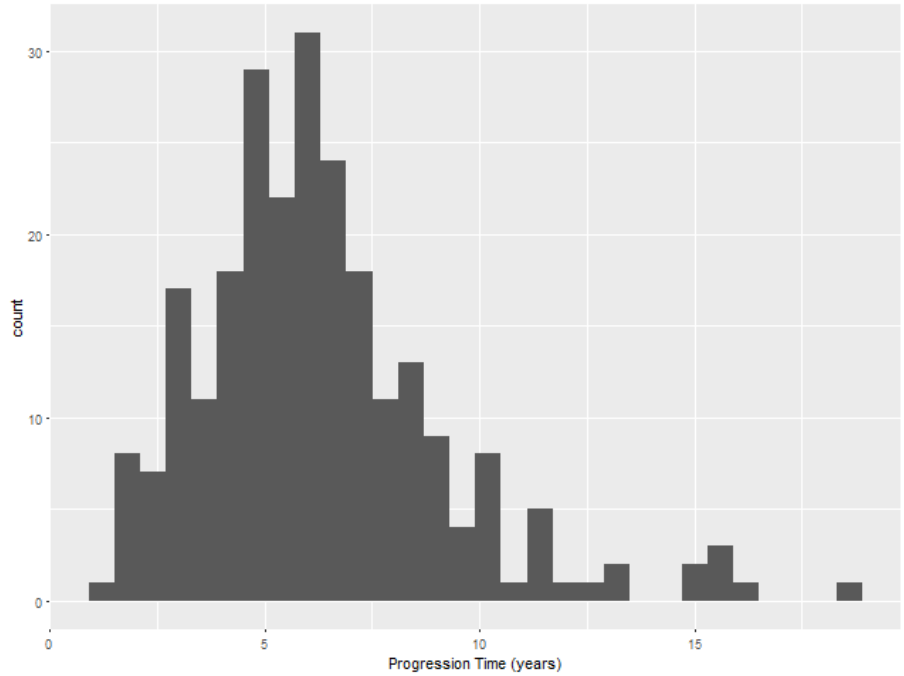


Figure 16: Gleason reclassification times for patients from the test data set of scenario 3.

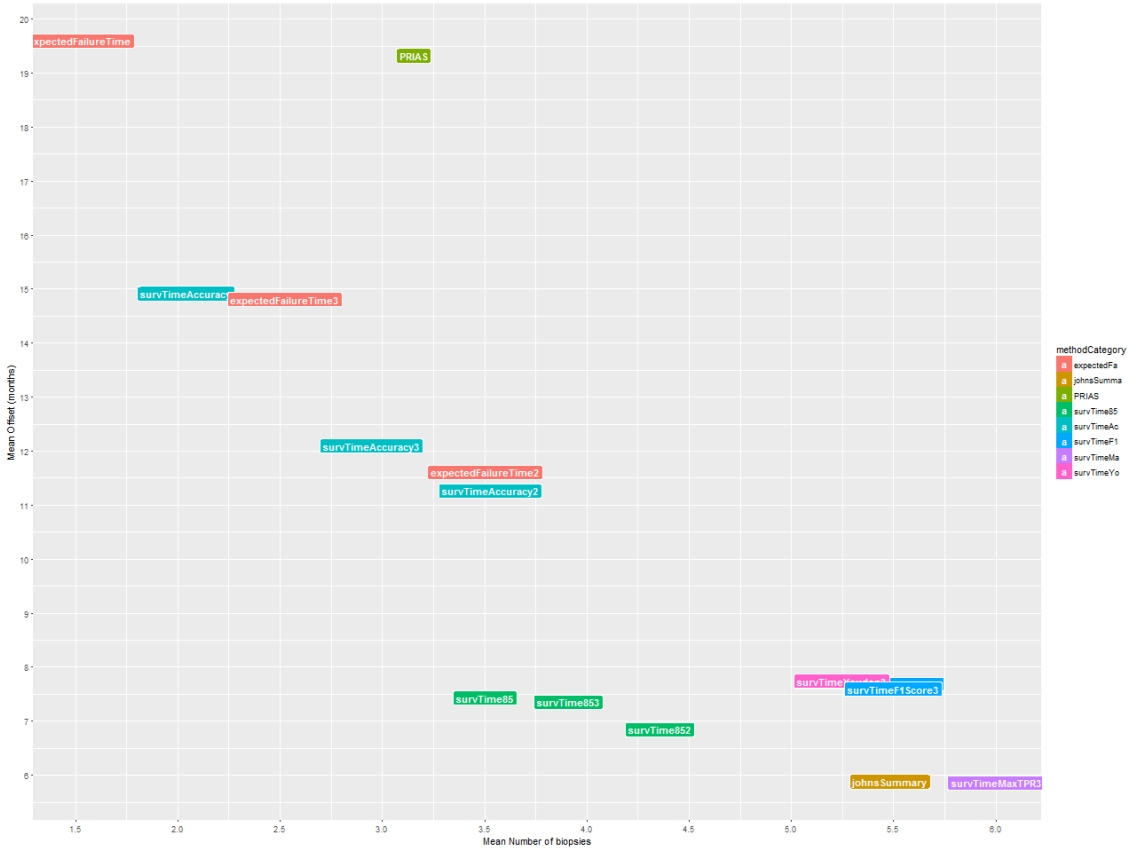


Figure 17: Mean number of biopsies against the mean offset (in years) for each of the approaches in scenario 3.

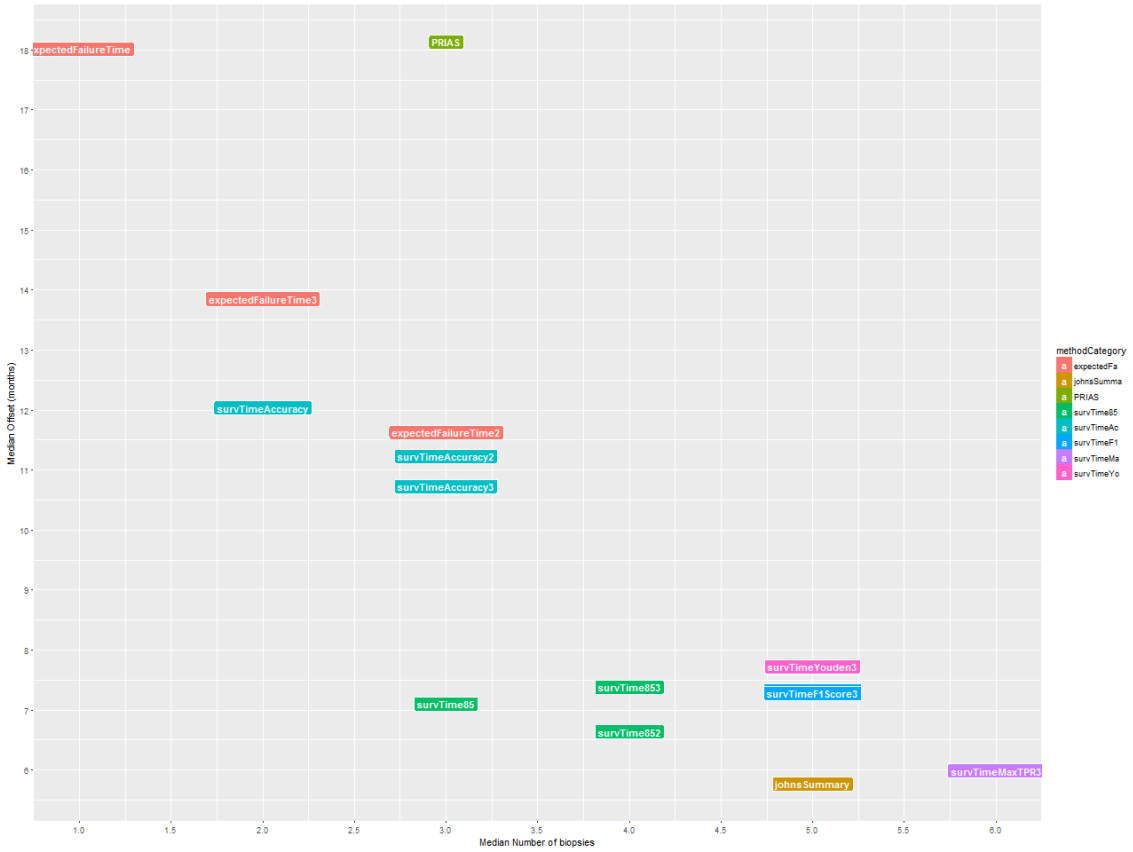


Figure 18: Median number of biopsies against the median offset (in years) for each of the approaches in scenario 3.

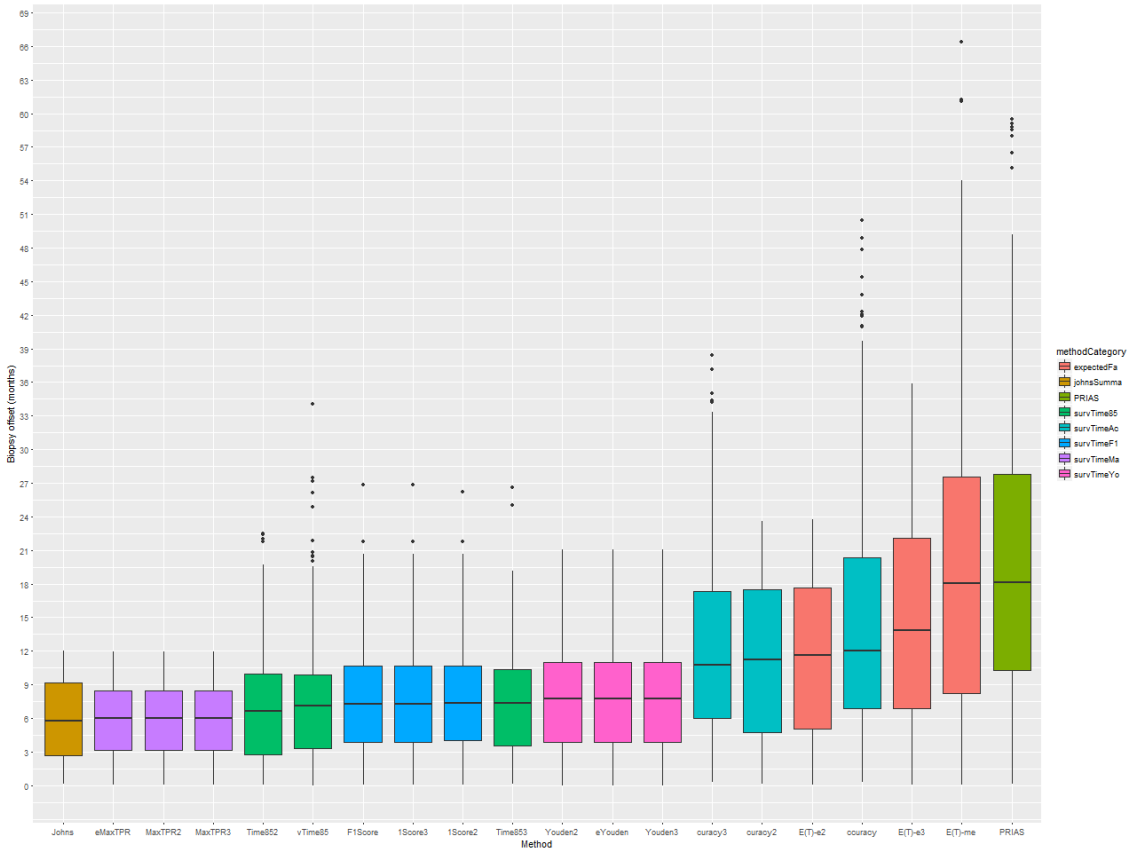


Figure 19: Boxplot for the offset corresponding to the various approaches in scenario 3.

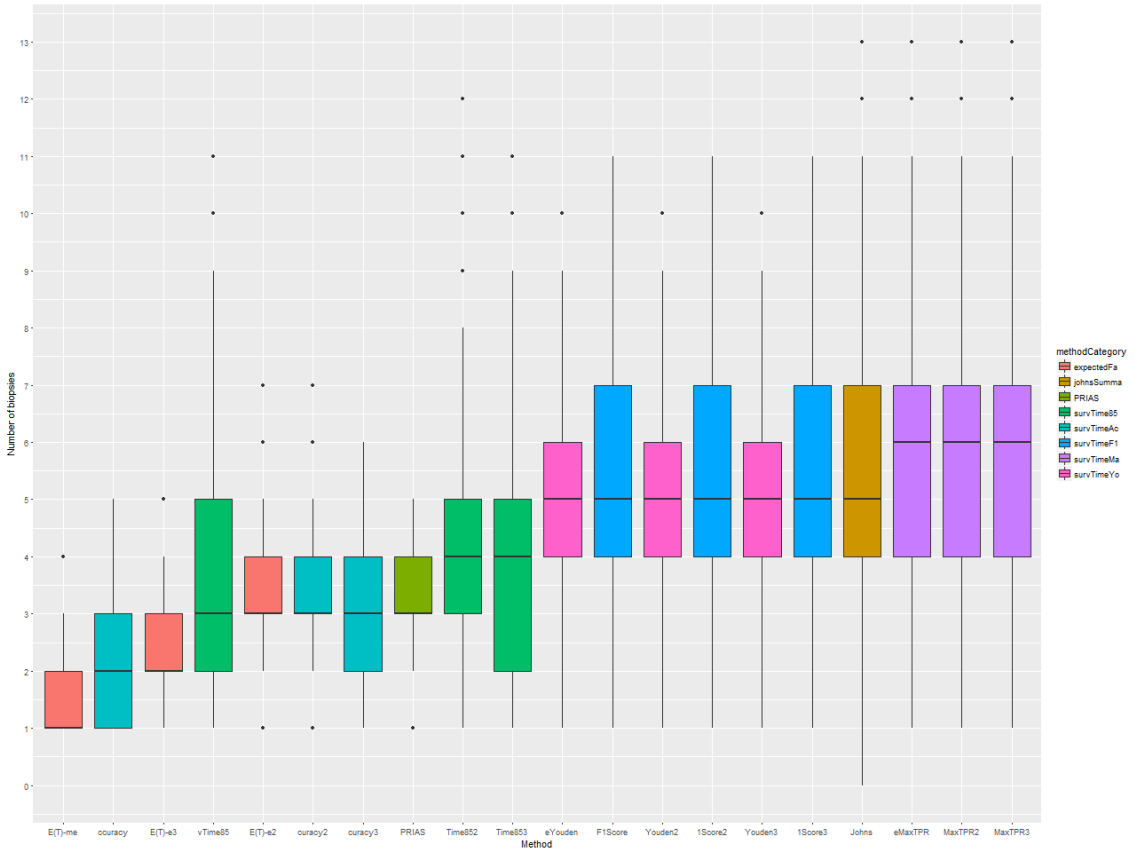


Figure 20: Boxplot for the number of biopsies corresponding to the various approaches in scenario 3.

5.3 A mixed approach: E(T) vs dynamic risk

While we did create biopsy schedules using conditional expected Gleason reclassification time, and dynamic risk of Gleason reclassification based approaches, we saw that no single approach works best in all scenarios. In particular, we observed that when the failure times are early dynamic risk based approaches perform better in terms of offset than conditional expected Gleason reclassification time. The latter works best when the Gleason reclassification times are later. This has motivated us to use a mixed approach, where we schedule biopsies using dynamic risk based approaches if the variance (Eq. 11) of time to Gleason reclassification is higher than a certain threshold, otherwise we use conditional expected Gleason reclassification time. This is still work in progress.

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.

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.

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.

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

McGreevy, Katharine et al. (2006). "Impact of race and baseline PSA on longitudinal PSA". In: *International journal of cancer* 118.7, pp. 1773–1776.

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.

Sène, Mbéry et al. (2016). "Individualized dynamic prediction of prostate cancer recurrence with and without the initiation of a second treatment: Development and validation". In: *Statistical methods in medical research* 25.6, pp. 2972–2991.

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.