Biometrics 00, 1–19 DOI: 10.1111/j.1541-0420.2005.00454.x
December 0000

# Personalized Schedules for Burdensome Surveillance Tests

**Anirudh Tomer[1,*], Daan Nieboer[2], Monique J. Roobol[3], Ewout W. Steyerberg[2,4], and Dimitris Rizopoulos[1]**

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

[2]Department of Public Health, Erasmus University Medical Center, the Netherlands

[3]Department of Urology, Erasmus University Medical Center, the Netherlands

[4]Department of Biomedical Data Sciences, Leiden University Medical Center, the Netherlands

*email: a.tomer@erasmusmc.nl

Summary: Commonly used gold standard surveillance *tests* (biopsies, endoscopies, etc.) for confirming disease *progression* in early-stage chronic non-communicable disease patients (cancer, cardiovascular, etc.) are invasive. For detecting progression timely (benefit), patients are exposed to numerous invasive tests repeatedly (burden) over their lifetime as per a fixed one-size-fits-all schedule. Motivated by this problem in the world's largest prostate cancer surveillance study PRIAS, we present disease progression risk-based personalized test schedules, that aim to better balance the number of invasive tests (burden), and time delay in detection of progression (less is beneficial) than fixed schedules.

Using joint models for time-to-event and longitudinal data, we first consolidate auxiliary longitudinal data (e.g., biomarkers) and results of previous invasive tests, into individualized future cumulative-risk of *progression*. Under this risk profile, we optimize a utility function of the number of tests and expected time delay in detection of progression to obtain a personalized schedule of invasive tests. Personalized schedules are updated as more patient data becomes available over follow-up. We assist patients/doctors in comparing the consequences of opting for personalized versus fixed schedules objectively. For this, we exploit a patient's cumulative-risk profile to estimate the expected time delay in detection of progression for following both personalized and fixed schedules, in a patient-specific manner. We implement our methodology in a web-application for real prostate cancer patients of the PRIAS study.

Key words: Chronic diseases; Invasive medical tests; Joint models; Personalized schedules; Prostate biopsy; Surveillance

This paper has been submitted for consideration for publication in *Biometrics*

## 1. Introduction

Chronic non-communicable diseases (e.g., cancer, renal, cardiovascular diseases, etc.) are the primary cause of human deaths worldwide (Alwan et al., 2010). In many patients diagnosed with an early stage disease, periodical surveillance *tests* are recommended to detect disease *progression*, a non-terminal event. Often the most accurate or gold standard surveillance tests are also invasive. For example, to confirm progression, biopsies are conducted repeatedly in prostate cancer (Bokhorst et al., 2015), endoscopies in Barrett's esophagus (Streitz et al., 1993), and colonoscopies in colorectal cancer (Krist et al., 2007). Repeat biopsies are also utilized to detect allograft deterioration in lung (McWilliams et al., 2008) and kidney transplant (Henderson et al., 2011) patients.

Usually, invasive tests are scheduled in a fixed manner, e.g., every six months. The frequency of tests in these fixed schedules varies between diseases (Henderson et al., 2011; Bokhorst et al., 2015; Krist et al., 2007) and cohorts. Although, due to the periodical nature of test schedules, progression is always detected with a time delay (Figure 1). This time delay can be reduced by scheduling invasive tests frequently. However, invasive tests are difficult to conduct, can lead to severe complications (Loeb et al., 2013; Krist et al., 2007), cause patient discomfort, and sometimes patients may not comply with frequent tests (Bokhorst et al., 2015). In this regard, fixed test schedules ignore the differences in speed of progression between patients, and impose an equal medical burden on all. Hence, the frequency of invasive tests holds important implications for patients.

[Figure 1 about here.]

In this paper, we aim to balance the number of invasive tests (burden) and the time delay in the detection of disease progression (less is beneficial) better than fixed schedules. For this purpose, we intend to create personalized test schedules that exploit patient-specific clinical data accumulated during follow-up. In surveillance, this data includes baseline characteristics

of patients; results from previous invasive tests; and auxiliary longitudinal outcomes such as biomarkers, physical examination, and medical imaging measurements, etc. Previous approaches for personalized schedules can be divided into three categories. First, heuristic methods such as decision making flowcharts, e.g., Bokhorst et al. (2015). However, flowcharts discretize continuous clinical outcomes, often utilize only the last measurement, and ignore the measurement error in observed outcomes. Second method is employing partially observable Markov decision processes (Alagoz et al., 2010; Steimle and Denton, 2017) for personalized test decisions. Although, the curse of dimensionality limits their application with continuous longitudinal outcomes. Third, personalized schedules obtained by optimizing an explicit utility function of the clinical parameters of interest (Bebu and Lachin, 2017; Rizopoulos et al., 2015), including our previous work on scheduling biopsies in prostate cancer (Tomer et al., 2019). In this work, we will employ the third approach.

Our methodology is as follows. First, we develop a full specification of the joint distribution of disease and patient-specific longitudinal outcomes, and the time of *progression*. We achieve this using joint models for time-to-event and longitudinal data (Tsiatis and Davidian, 2004; Rizopoulos, 2012). We use joint models because they are inherently personalized. Specifically, they exploit patient-specific random effects (Laird and Ware, 1982) to model longitudinal outcomes without discretizing them. We subsequently employ the fitted joint model for new patients, to estimate their patient-specific cumulative-risk of progression over their current and future follow-up visits. These risk predictions utilize their clinical data accumulated until their latest follow-up. We schedule invasive tests on all those future time points where a patient's conditional cumulative-risk of progression is equal to a certain threshold (e.g., 10% risk). We then automate the choice of this threshold and the resulting schedule. More specifically, we optimize a function of the number of tests in a schedule and the expected time delay in the detection of progression (Figure 1). We estimate this time delay in a patient-

specific manner for both fixed and personalized schedules. This can help patients/doctors to evaluate consequences of opting for personalized versus fixed schedules objectively.

This research is motivated by the problem of scheduling biopsies (Nieboer et al., 2018) in the world's largest prostate cancer active surveillance study PRIAS (Bokhorst et al., 2015). It has 7813 patients, 104,904 longitudinal measurements, and 1134 patients with cancer progression. These patients have low/very-low grade prostate cancer, often over-diagnosed due to prostate-specific antigen (PSA) based screening tests (Crawford, 2003). The goal of surveillance upon diagnosis is to delay serious treatments (e.g., surgery, chemotherapy, etc.) until cancer progresses further. For this purpose, patients are monitored continually via PSA (ng/mL) blood tests, digital rectal examination (DRE) for shape and size of the tumor, and biopsy Gleason grade group (Epstein et al., 2016). Since biopsy results are the strongest indicator of cancer-related outcomes, treatment is commonly advised upon observing an increase in a patient's biopsy Gleason grade group (cancer progression). The most commonly used schedule of biopsies in prostate cancer active surveillance is annual biopsies (Loeb et al., 2014). However, they lead to many unnecessary biopsies in slow/non-progressing patients (50% proportion in some cohorts). Biopsy burden combined with patient non-compliance to frequent biopsies (Bokhorst et al., 2015) has raised concerns regarding the optimal biopsy schedule. Since prostate cancer has the second-highest incidence among all cancers in males (Torre et al., 2015), biopsy schedules tailored for individual patients can reduce the overall burden of biopsies in a large number of patients worldwide.

The rest of the paper is as follows. Section 2 briefly introduces the joint modeling framework. In Section 3 we present the methodology for personalized schedules and then demonstrate them for biopsies in real PRIAS patients in Section 4. Lastly, in Section 5, we show the efficacy of personalized schedules via a realistic simulation study based on PRIAS patients.

## 2. Joint Model for Time-to-Progression and Longitudinal Outcomes

Let the true time of disease progression for the $i$-th patient be $T_i^*$. Progression is always observed with interval censoring $l_i < T_i^* \leqslant r_i$ (Figure 1). In patients who obtain progression, $r_i$ and $l_i$ denote the time of their latest and second latest invasive tests. Otherwise, $l_i$ denotes the time of their latest test and $r_i = \infty$. Assuming $K$ auxiliary longitudinal outcomes, let $\boldsymbol{y}_{ki}$ denote the $n_{ki} \times 1$ longitudinal response vector of the $k$-th outcome, $k\epsilon\{1, \ldots, K\}$. The observed data of all $n$ patients is given by $\mathcal{A}_n = \{l_i, r_i, \boldsymbol{y}_{1i}, \ldots \boldsymbol{y}_{Ki}; i = 1, \ldots, n\}$.

To accommodate longitudinal outcomes of different types in a unified framework, the joint model consists of a generalized linear mixed-effects sub-model (Laird and Ware, 1982). In particular, the conditional distribution of $\boldsymbol{y}_{ki}$ given a vector of patient-specific random effects $\boldsymbol{b}_{ki}$ is assumed to be a member of the exponential family, with linear predictor given by,

$$g_k\big[E\{y_{ki}(t) \mid \boldsymbol{b}_{ki}\}\big] = m_{ki}(t) = \boldsymbol{x}_{ki}^T(t)\boldsymbol{\beta}_k + \boldsymbol{z}_{ki}^T(t)\boldsymbol{b}_{ki}, \tag{1}$$

where $g_k(\cdot)$ denotes a known one-to-one monotonic link function, $y_{ki}(t)$ denotes the value of the $k$-th longitudinal outcome for the $i$-th patient at time $t$, and $\boldsymbol{x}_{ki}(t)$ and $\boldsymbol{z}_{ki}(t)$ denote the time-dependent design vectors for the fixed $\boldsymbol{\beta}_k$ and random effects $\boldsymbol{b}_{ki}$, respectively. To account for the association between the different longitudinal outcomes, we link their corresponding random effects. More specifically, the complete vector of random effects $\boldsymbol{b}_i = (\boldsymbol{b}_{1i}^T, \ldots, \boldsymbol{b}_{Ki}^T)^T$ is assumed to follow a multivariate normal distribution with mean zero and variance-covariance matrix $W$.

For the survival process, we assume that the hazard of progression $h_i(t)$ at a time $t$ depends on a function of the patient and outcome-specific linear predictors $m_{ki}(t)$ and/or the random effects. More specifically,

$$h_i\big\{t \mid \mathcal{M}_i(t), \boldsymbol{w}_i\big\} = h_0(t) \exp\Big[\boldsymbol{\gamma}^T\boldsymbol{w}_i + \sum_{k=1}^{K}\sum_{l=1}^{L_k} f_{kl}\big\{\mathcal{M}_{ki}(t), \boldsymbol{w}_i, \boldsymbol{b}_{ki}, \boldsymbol{\alpha}_{kl}\big\}\Big], \quad t > 0, \tag{2}$$

where $h_0(\cdot)$ denotes the baseline hazard function, $\mathcal{M}_{ki}(t) = \{m_{ki}(s) \mid 0 \leqslant s < t\}$ denotes the history of the $k$-th longitudinal process up to $t$, and $\boldsymbol{w}_i(t)$ is a vector of exogenous, possibly

time-varying, covariates with corresponding regression coefficients $\boldsymbol{\gamma}$. Functions $f_{kl}(\cdot)$, parameterized by vector of coefficients $\boldsymbol{\alpha_{kl}}$, specify which features of each longitudinal outcome are included in the linear predictor of the relative-risk model (Brown, 2009; Rizopoulos, 2012; Taylor et al., 2013). Some examples, motivated by the literature (subscripts $k$ and $l$ dropped for brevity), are:

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

These formulations of $f(\cdot)$ postulate that the hazard of progression at time $t$ may be associated with the underlying level $m_i(t)$ of the longitudinal outcome at $t$, or with both the level and velocity $m_i'(t)$ (e.g., PSA value and velocity in prostate cancer) of the outcome at $t$. Lastly, $h_0(t)$ is the baseline hazard at time $t$, and is modeled flexibly using P-splines (Eilers and Marx, 1996). The detailed specification of the baseline hazard $h_0(t)$, and the joint parameter estimation of the longitudinal and relative-risk sub-models using the Bayesian approach are presented in Web-Appendix A.

## 3. Personalized Schedule of Invasive Tests for Detecting Progression

We intend to develop a personalized schedule of invasive tests for a new patient $j$, not present in training dataset $\mathcal{A}_n$. Tests are conducted ~~only~~ until progression is detected (Figure 1). Let $T_j^*$ be the true time of progression, and $t < T_j^*$ be the time of the latest test on which progression was not detected for the jth patient. Lastly, $v \geqslant t$ denotes the time of the current follow-up visit.

### 3.1 *Cumulative-risk of progression*

First we ~~consolidate~~ the history of observed longitudinal outcomes $\{\mathcal{Y}_{1j}(v), \ldots, \mathcal{Y}_{Kj}(v)\}$ until the current visit time $v$, and the previous negative test result $T_j^* > t$ ~~into~~ a patient-specific

cumulative-risk of progression at future time $u$ (Figure 2). ~~It is given by,~~

$$
\begin{aligned}
R_j(u \mid t, v) &= p\big\{T_j^* \leqslant u \mid T_j^* > t, \mathcal{Y}_{1j}(v), \ldots, \mathcal{Y}_{Kj}(v), \mathcal{A}_n\big\} \\
&= \int\!\!\int p(T_j^* \leqslant u \mid T_j^* > t, \boldsymbol{b}_j, \boldsymbol{\theta}) p\big\{\boldsymbol{b}_j \mid T_j^* > t, \mathcal{Y}_{1j}(v), \ldots, \mathcal{Y}_{Kj}(v), \boldsymbol{\theta}\big\} \\
&\quad \times p(\boldsymbol{\theta} \mid \mathcal{A}_n) \mathrm{d}\boldsymbol{b}_j \mathrm{d}\boldsymbol{\theta}, \quad u \geqslant t. \qquad (3)
\end{aligned}
$$

The personalized cumulative-risk function $R_j(\cdot)$ depends on the observed longitudinal data $\{\mathcal{Y}_{1j}(v), \ldots, \mathcal{Y}_{Kj}(v)\}$, and the training dataset $\mathcal{A}_n$ via the posterior distribution of patient-specific random effects $\boldsymbol{b}_j$, and posterior distribution of the vector of joint model parameters $\boldsymbol{\theta}$, respectively. The risk also dynamically updates as more longitudinal data becomes available over follow-up (Panel B and C, Figure 2).

[Figure 2 about here.]

### 3.2 *Schedule of Invasive Tests*

Our aim is to employ the cumulative-risk function ~~in Equation~~ (3) to develop a personalized schedule of invasive tests, starting from the current visit time $v$ until a maximum horizon time $h$. For this purpose we utilize a simple and straightforward approach of scheduling invasive tests on all those time points where the conditional cumulative-risk of progression is above a certain threshold $0 \leqslant \kappa \leqslant 1$, (e.g., 15% risk in Figure 3). More specifically,

$$
S_j^{\kappa} = \big\{s_1, \ldots, s_N \mid R_j(s_n \mid s_{n-1}, v) = \kappa, s_0 = t\big\}, \quad 1 \leqslant n \leqslant N, \qquad (4)
$$

where $s_n$ is the time of the $n$-th test in the personalized test schedule $S_j^{\kappa}$. The conditional cumulative-risk of progression at time $s_n$ denoted by $R_j(s_n \mid s_{n-1}, v)$ is defined as in Equation (3). It is called 'conditional' because each successive $n$-th test at future time $s_n$ is scheduled by accounting for the possibility that progression may not have occurred until the time of the previously scheduled test $s_{n-1}$. That is, $T_j^* > s_{n-1}$. However, the contribution of the observed longitudinal data $\{\mathcal{Y}_{1j}(v), \ldots, \mathcal{Y}_{Kj}(v)\}$ does not change while scheduling these

consecutive tests. Since the schedule is risk based, it also gets updated as more patient data becomes available over follow-up.

[Figure 3 about here.]

### 3.3 *Risk Threshold* $\kappa$

The risk threshold $\kappa$ controls the timing and the total number of invasive tests in the schedule $S_j^\kappa$. Through the timing of tests, $\kappa$ also indirectly affects the time delay (Figure 1) that may occur in the detection of progression if this schedule is followed. Hence, $\kappa$ should be chosen while balancing both the number of invasive tests (burden), and the time delay in the detection of progression (less is beneficial).

Consider the bi-dimensional Euclidean space of the total number of invasive tests (x-axis) and the corresponding expected time delay in detection of progression (y-axis) for schedules associated with various $\kappa$ (Figure 4). An ideal schedule of tests will have only one test planned exactly at the true time of progression $T_j^*$ of a patient. In other words it will lead to a zero time delay. This schedule is shown at the point of optimality (1, 0) in Figure 4. Subsequently, a threshold $\kappa_a$ can be chosen automatically by minimizing the Euclidean distance between the point $(1,\ 0)$ and the set of points representing various schedules corresponding to each $0 \leqslant \kappa \leqslant 1$. That is,

$$\kappa_a = \arg\min_\kappa \sqrt{\left(|S_j^\kappa| - 1\right)^2 + \left\{D_j(S_j^\kappa \mid t, v) - 0\right\}^2}, \quad 0 \leqslant \kappa \leqslant 1, \tag{5}$$

where, $D_j(S_j^\kappa \mid t, v)$ denotes the expected time delay in ~~detection of~~ progression ~~(estimation in Section 3.4)~~ if schedule $S_j^\kappa$ is followed. Additional consequences of following a particular schedule, such as (quality-adjusted) life years saved, can also be accommodated in ~~Equation~~ (5). This can be achieved by first setting a point of optimality in a higher dimensional Euclidean space of the aforementioned consequences, and then minimizing the Euclidean distance to the point of optimality.

Certain patients may have preferences for the maximum number of invasive tests scheduled

for them. Others may be apprehensive about having an expected time delay higher than a certain number of months. In this regard, the Euclidean distance in ~~Equation~~ (5) can be minimized under constraints on the number of tests and/or expected time delay (Figure 4). An additional benefit of this approach is that it alleviates the issue of time delay and the number of tests having different units of measurement (Cook and Wong, 1994).

[Figure 4 about here.]

### 3.4 *Expected Time Delay in Detection of Progression*

We estimate the expected time delay $D_j(S_j^{\kappa} \mid t, v)$ in ~~Equation~~ (5) in a patient-specific manner as well. That is, two patients may opt to follow the same schedule, but they will expect different time delays. To this end, we utilize the personalized cumulative-risk profile estimated in ~~Equation~~ (3). Also, the calculation of time delay is not limited to personalized schedules only. Rather, for any schedule $S$ consisting of $N$ invasive tests, the personalized expected time delay for the $j$-th patient is given by,

$$
\begin{aligned}
D_j(S \mid t, v) &= \sum_{n=1}^{N} R_j(s_n \mid s_{n-1}, v) d_j(s_n, s_{n-1}, v), \quad s_0 = t \\
d_j(s_n, s_{n-1}, v) &= s_n - s_{n-1} - E(T_j^* \mid s_{n-1}, s_n, v), \\
E(T_j^* \mid s_{n-1}, s_n, v) &= \int_{s_{n-1}}^{s_n} p\Big\{T_j^* \geqslant u \mid s_{n-1} < T_j^* \leqslant s_n, \mathcal{Y}_{1j}(v), \ldots, \mathcal{Y}_{Kj}(v), \mathcal{A}_n\Big\} \mathrm{d}u, \qquad (6)
\end{aligned}
$$

where $s_n$ is the time of the $n$-th invasive test in schedule $S$, and $E(T_j^* \mid s_{n-1}, s_n, v)$ is the conditional expected progression time, given that the patient obtains progression between the two consecutive tests conducted at times $s_{n-1}$ and $s_n$. This expected progression time is used to calculate individual time delays $d_j(s_n, s_{n-1}, v)$, under the scenario that patient obtains progression in the time interval $s_{n-1} < T_j^* \leqslant s_n$. Subsequently, we obtain the overall personalized expected time delay $D_j(S \mid t, v)$ as a weighted average of the individual delays $d_j(s_n, s_{n-1}, v)$. Hence, the personalized expected time delay $D_j(S \mid t, v)$ should be interpreted as the delay if $T_j^* \leqslant s_N$. That is, if the patient obtains progression before the last invasive

test in the schedule $S$. Personalized expected delay can assist patients and doctors in shared decision making of an appropriate test schedule. Although, in order to have a fair comparison of expected delay between different schedules for the same patient, a compulsory test at a common horizon time point should be planned in all schedules.

## 4. Demonstration of Personalized Schedules

We return to the prostate cancer active surveillance dataset, PRIAS, described in Section 1. The current PRIAS protocol for biopsies is fixed biopsies at year one, four, seven, and ten of follow-up, and every five years after that. Additional annual biopsies are scheduled if a patient's PSA doubling-time (Bokhorst et al., 2015) is high. The PSA is measured as per a fixed schedule, quarterly for the first two years, and semi-annually after that. The DRE is also measured semi-annually. The dataset is described in more detail in Web-Appendix B.

The clinical data that we intend to use consists of longitudinal PSA (continuous: ng/mL) and DRE (binary: tumor palpable or not) measurements, patient age at baseline, history of biopsies, and interval-censored times of cancer progression. The event of interest is cancer progression. We aim to use the accumulated clinical data to build a joint model that can be utilized for creating personalized biopsy schedules in future PRIAS patients.

### 4.1 *Fitting the Joint Model to the PRIAS Dataset*

We fit a joint model with $\log_2(\text{PSA}+1)$ transformed PSA (Lin et al., 2000; Pearson et al., 1994), and DRE as longitudinal outcomes, and cancer progression as the event (Web-Appendix B for exact specification). For PSA we utilize a linear mixed-effects sub-model wherein PSA profiles are modeled non-linearly over follow-up using B-splines (De Boor, 1978). For DRE, we utilize a logistic mixed-effects sub-model. To link the longitudinal sub-models for the PSA and DRE with the relative-risk sub-model for cancer progression, we include three features of the longitudinal outcomes in the relative-risk sub-model. Specifically,

the hazard of cancer progression at time $t$ depends on the fitted instantaneous $\log_2(\text{PSA}+1)$ value at time $t$, the estimated instantaneous $\log_2(\text{PSA}+1)$ velocity at $t$, and fitted log-odds of having a DRE indicating a palpable tumor at $t$. We estimated the parameters of our model under the Bayesian framework using the R package **JMbayes** (Rizopoulos, 2016).

The follow-up period of PRIAS is limited currently. Hence, our joint model is able to predict the cumulative-risk of progression only until the year ten of follow-up. The cumulative-risk of progression at year ten in PRIAS is 50% (see Web-Figure...). We found that the strongest predictor for progression in our model is $\log_2(\text{PSA}+1)$ velocity. Specifically, for an increase in fitted $\log_2(\text{PSA}+1)$ velocity from -0.03 to 0.15 the adjusted hazard ratio of progression was 1.6 (95%CI: 1.45–1.78). Detailed parameter estimates are in Web-Appendix B.

### 4.2 *Personalized Schedules for a Demonstration Patient*

We utilized the joint model fitted to the PRIAS dataset to schedule biopsies in a real PRIAS patient (Figure 5), starting from his current visit at year five, until year ten of follow-up. The cumulative-risk of progression of this patient at his current visit is 6% whereas at ten years it is 16.5%. Thus, the patient is predicted to progresses slowly. Consequently, risk based personalized schedules in Panel B of Figure 5 planned much less biopsies than the standard annual schedule. At the same time, the expected time delay in detection of progression in personalized schedules is also less than one year (maximum delay possible with annual schedule). It is important to note that for a fair comparison of expected time delay, we scheduled a compulsory biopsy at the horizon of ten years in all schedules. In addition, we maintained a recommended minimum gap of one year between consecutive biopsies (Bokhorst et al., 2016).

[Figure 5 about here.]

### 4.3 *Web-Application*

Question for Dimitris: should we implement a new version of our web-application which supports any kind of use-case, not just PRIAS. Current issue is that the joint model objects are quite large sometimes. Loading these objects can be difficult. Either way, I only plan to keep two tabs in this web-app. One for risk predictions and another for side by side comparison of personalized schedules versus fixed frequency schedule (like camera and phone comparison websites do).

## 5. Simulation Study

Although we demonstrated personalized schedules for a real patient, we also intend to analyze and compare personalized and fixed schedules in a full study cohort. We evaluate schedules on the basis of the total number of invasive tests planned, and the actual time delay in detection of progression for each schedule. However, due to the periodical nature of schedules the actual time delay in detection of progression cannot be observed in real world surveillance. Hence, instead we compare personalized versus fixed schedules via an extensive simulated randomized clinical trial in which each hypothetical patient undergoes each schedule. To keep our simulation study realistic, we employ the prostate cancer active surveillance scenario. More specifically, our simulated population is manifested by the fitted joint model obtained using the PRIAS cohort (Web-Appendix B).

### 5.1 *Simulation Setup*

From the simulation population, we first sample 500 datasets, each representing a hypothetical prostate cancer active surveillance program with 1000 patients in it. We generate a true cancer progression time for each of the $500 \times 1000$ patients and then sample a set of longitudinal DRE and PSA measurements at the same follow-up visit times as given in the PRIAS protocol. We then split each dataset into training (750 patients) and test (250

patients) parts, and generate a random and noninformative censoring time for the training patients. All test and training patients also observe Type-I censoring at year ten of follow-up (current study period of PRIAS). We next fit a joint model of the same specification as the model fitted to PRIAS (Web-Appendix B), to each of the 500 training datasets and obtain MCMC samples from the 500 sets of the posterior distribution of the parameters. In each of the 500 hypothetical surveillance programs, we utilize the corresponding fitted joint models to develop the profiles for cumulative-risk of progression in each of the $500 \times 250$ test patients. This cumulative-risk is further used to create personalized biopsy schedules for the test patients. For each test patient we conduct biopsies using personalized biopsy schedules based on two fixed risk thresholds, namely, $\kappa = 5\%$ and $\kappa = 10\%$, and an automatically chosen $\kappa_a$ (Equation 5). We also conduct the currently practiced PRIAS and annual biopsy schedules. We plan biopsies only on the standard PSA follow-up visits (Section 4) utilizing accumulated clinical data until that visit. Also, we maintain a minimum recommended gap of one year between consecutive prostate biopsies (Bokhorst et al., 2015). Biopsies are conducted until progression is detected, or the maximum follow-up period at year ten (horizon) is reached. The actual time delay in detection of progression is equal to the time at which progression is detected minus the actual (simulated) time of progression of a patient. ~~Although, a compulsory biopsy is conducted at year ten for all schedules, for a fair comparison between their corresponding time delays in detection of progression.~~

### 5.2 *Results*

Since the simulated cohorts are based on PRIAS, roughly only 50% of the patients progress in the ten year study period. While, we are able to calculate total number of biopsies scheduled in all $500 \times 250$ test patients, but the time delay in detection of progression is available only for those patients who progress in ten years (*progressing*). Hence, we show the simulation

results separately for *progressing* and *non-progressing* patients in Panel A, and Panel B of Figure 6, respectively.

For *progressing* patients (Panel A, Figure 6), we note that .....

The patients who are at the most advantage with the personalized schedules are the *non-progressing* patients (Panel B, Figure 6). For all of these patients, annual schedule leads to 10 (unnecessary) biopsies. The schedule of the PRIAS program schedules a median of six biopsies (IQR: 4–8). In comparison,....

[Figure 6 about here.]

## 6. Discussion

In this paper, we presented a methodology to create personalized schedules for burdensome surveillance *tests*, to detect disease *progression* in early-stage chronic non-communicable diseases. To this end, we utilized the framework of joint models for time-to-event and longitudinal data. Our approach first combines a patient's auxiliary longitudinal data (e.g., biomarkers) and results from previous invasive tests to estimate the patient-specific cumulative-risk of disease progression over his current and future follow-up time period. Then, using this risk profile, we schedule future invasive tests whenever the patient's conditional cumulative-risk of progression is predicted to be above a certain threshold. We select this risk threshold automatically in a personalized manner, by optimizing a utility function of the patient-specific consequences of choosing a particular risk threshold based schedule. These consequences are namely, the number of invasive tests for a particular schedule, and the expected time delay in detection of progression if that schedule is followed. Last, we calculate this expected time delay in a personalized manner for both personalized and fixed schedules, to assist patients/doctors in making a more informed decision of choosing a test schedule.

The use of joint models gives our schedules certain advantages. First, joint models utilize

individualized random-effects, making our schedules inherently personalized. Second, the patient-specific risk of progression employed by the proposed personalized schedules is estimated by utilizing all observed longitudinal and clinical data of a patient. In addition, the continuous longitudinal outcomes are not discritized which is commonly a case in Markov Decision Process based (Alagoz et al., 2010; Steimle and Denton, 2017), and flowchart based test schedules. The third and biggest advantage of our schedules, however, is that they update as more patient data becomes available over follow-up. Last, although we specifically discuss the use of personalized schedules in disease surveillance, they are generic for use under a screening setting as well.

Since our schedules are risk based, we proposed a utility function to automate the choice of a risk threshold based schedule. The utility function that we proposed focused only on two aspects of a schedule, namely the burden and the benefit. In this regard, we chose the number of invasive tests in a schedule (burden), and expected time delay in detection of progression (less is beneficial) because they are easy to interpret and are critical in making decision of an invasive test. Since we calculate the expected time delay in a patient-specific manner for both personalized and fixed schedules, patients/doctors can compare and choose schedules according to their preferences for the burden-benefit ratio. Additional measures such as (quality adjusted) life years saved can also be easily added in our utility function.

I am yet to add a paragraph discussing simulation study quickly. This discussion is useful because it shows the extent to which personalized schedules can benefit patients.

There are certain limitations of our work. First, in practice, most cohorts observe Type-I right censoring. Hence, the cumulative-risk profiles of patients, and calculation of expected time delay in detection of progression is only possible up to the time of Type-I censoring. This problem can only be resolved as more follow-up data becomes available over time. We proposed a joint model which assumes all events other than progression to be non-informative

censoring. Alternative models that account for competing risks may lead to better results as they estimate absolute, and not cause-specific, risk of progression. However, the methodology for scheduling biopsies need not change. Many surveillance tests are imperfect and prone to inter-observer variation (e.g., biopsy Gleason grade). Models which account for inter-observer variation in diagnostic tests (Balasubramanian and Lagakos, 2003) will be interesting to investigate further.

## Acknowledgments

The first and last authors would like to acknowledge support by Nederlandse Organisatie voor Wetenschappelijk Onderzoek (the national research council of the Netherlands) VIDI grant nr. 016.146.301, and Erasmus University Medical Center funding. Part of this work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative. The authors also thank the Erasmus University Medical Center's Cancer Computational Biology Center for giving access to their IT-infrastructure and software that was used for the computations and data analysis in this study.

## References

Alagoz, O., Ayer, T., and Erenay, F. S. (2010). Operations research models for cancer screening. *Wiley encyclopedia of operations research and management science* .

Alwan, A., MacLean, D. R., Riley, L. M., d'Espaignet, E. T., Mathers, C. D., Stevens, G. A., and Bettcher, D. (2010). Monitoring and surveillance of chronic non-communicable diseases: progress and capacity in high-burden countries. *The Lancet* **376,** 1861–1868.

Balasubramanian, R. and Lagakos, S. W. (2003). Estimation of a failure time distribution based on imperfect diagnostic tests. *Biometrika* **90,** 171–182.

Bebu, I. and Lachin, J. M. (2017). Optimal screening schedules for disease progression with application to diabetic retinopathy. *Biostatistics* **19,** 1–13.

Bokhorst, L. P., Alberts, A. R., Rannikko, A., Valdagni, R., Pickles, T., Kakehi, Y., Bangma, C. H., Roobol, M. J., and PRIAS study group (2015). Compliance rates with the Prostate Cancer Research International Active Surveillance (PRIAS) protocol and disease reclassification in noncompliers. *European Urology* **68,** 814–821.

Bokhorst, L. P., Valdagni, R., Rannikko, A., Kakehi, Y., Pickles, T., Bangma, C. H., Roobol, M. J., and PRIAS study group (2016). A decade of active surveillance in the PRIAS study: an update and evaluation of the criteria used to recommend a switch to active treatment. *European Urology* **70,** 954–960.

Brown, E. R. (2009). Assessing the association between trends in a biomarker and risk of event with an application in pediatric HIV/AIDS. *The Annals of Applied Statistics* **3,** 1163–1182.

Cook, R. D. and Wong, W. K. (1994). On the equivalence of constrained and compound optimal designs. *Journal of the American Statistical Association* **89,** 687–692.

Crawford, E. D. (2003). Epidemiology of prostate cancer. *Urology* **62,** 3–12.

De Boor, C. (1978). *A practical guide to splines*, volume 27. Springer-Verlag New York.

Eilers, P. H. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* **11,** 89–121.

Epstein, J. I., Egevad, L., Amin, M. B., Delahunt, B., Srigley, J. R., and Humphrey, P. A. (2016). The 2014 international society of urological pathology (isup) consensus conference on gleason grading of prostatic carcinoma. *The American journal of surgical pathology* **40,** 244–252.

Henderson, L., Nankivell, B., and Chapman*, J. (2011). Surveillance protocol kidney transplant biopsies: their evolving role in clinical practice. *American Journal of Transplantation* **11,** 1570–1575.

Krist, A. H., Jones, R. M., Woolf, S. H., Woessner, S. E., Merenstein, D., Kerns, J. W.,

Foliaco, W., and Jackson, P. (2007). Timing of repeat colonoscopy: disparity between guidelines and endoscopists recommendation. *American journal of preventive medicine* **33,** 471–478.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* pages 963–974.

Lin, H., McCulloch, C. E., Turnbull, B. W., Slate, E. H., and Clark, L. C. (2000). A latent class mixed model for analysing biomarker trajectories with irregularly scheduled observations. *Statistics in Medicine* **19,** 1303–1318.

Loeb, S., Carter, H. B., Schwartz, M., Fagerlin, A., Braithwaite, R. S., and Lepor, H. (2014). Heterogeneity in active surveillance protocols worldwide. *Reviews in urology* **16,** 202–203.

Loeb, S., Vellekoop, A., Ahmed, H. U., Catto, J., Emberton, M., Nam, R., Rosario, D. J., Scattoni, V., and Lotan, Y. (2013). Systematic review of complications of prostate biopsy. *European urology* **64,** 876–892.

McWilliams, T. J., Williams, T. J., Whitford, H. M., and Snell, G. I. (2008). Surveillance bronchoscopy in lung transplant recipients: risk versus benefit. *The Journal of Heart and Lung Transplantation* **27,** 1203–1209.

Nieboer, D., Tomer, A., Rizopoulos, D., Roobol, M. J., and Steyerberg, E. W. (2018). Active surveillance: a review of risk-based, dynamic monitoring. *Translational andrology and urology* **7,** 106–115.

Pearson, J. D., Morrell, C. H., Landis, P. K., Carter, H. B., and Brant, L. J. (1994). Mixed-effects regression models for studying the natural history of prostate disease. *Statistics in Medicine* **13,** 587–601.

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

Rizopoulos, D. (2016). The R package JMbayes for fitting joint models for longitudinal and

time-to-event data using MCMC. *Journal of Statistical Software* **72,** 1–46.

Rizopoulos, D., Taylor, J. M., Van Rosmalen, J., Steyerberg, E. W., and Takkenberg, J. J. (2015). Personalized screening intervals for biomarkers using joint models for longitudinal and survival data. *Biostatistics* **17,** 149–164.

Steimle, L. N. and Denton, B. T. (2017). Markov decision processes for screening and treatment of chronic diseases. In *Markov Decision Processes in Practice*, pages 189–222. Springer.

Streitz, J. J., Andrews, J. C., and Ellis, J. F. (1993). Endoscopic surveillance of barrett's esophagus. does it help? *The Journal of thoracic and cardiovascular surgery* **105,** 383–7.

Taylor, J. M., Park, Y., Ankerst, D. P., Proust-Lima, C., Williams, S., Kestin, L., Bae, K., Pickles, T., and Sandler, H. (2013). Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics* **69,** 206–213.

Tomer, A., Nieboer, D., Roobol, M. J., Steyerberg, E. W., and Rizopoulos, D. (2019). Personalized schedules for surveillance of low-risk prostate cancer patients. *Biometrics* **75,** 153–162.

Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., and Jemal, A. (2015). Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians* **65,** 87–108.

Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* **14,** 809–834.

## Supporting Information

Web Appendix A, referenced in Section **??**, is available with this paper at the Biometrics website on Wiley Online Library.

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

## Appendix

### *Title of appendix*

Put your short appendix here. Remember, longer appendices are possible when presented as Supplementary Web Material. Please review and follow the journal policy for this material, available under Instructions for Authors at `http://www.biometrics.tibs.org`.

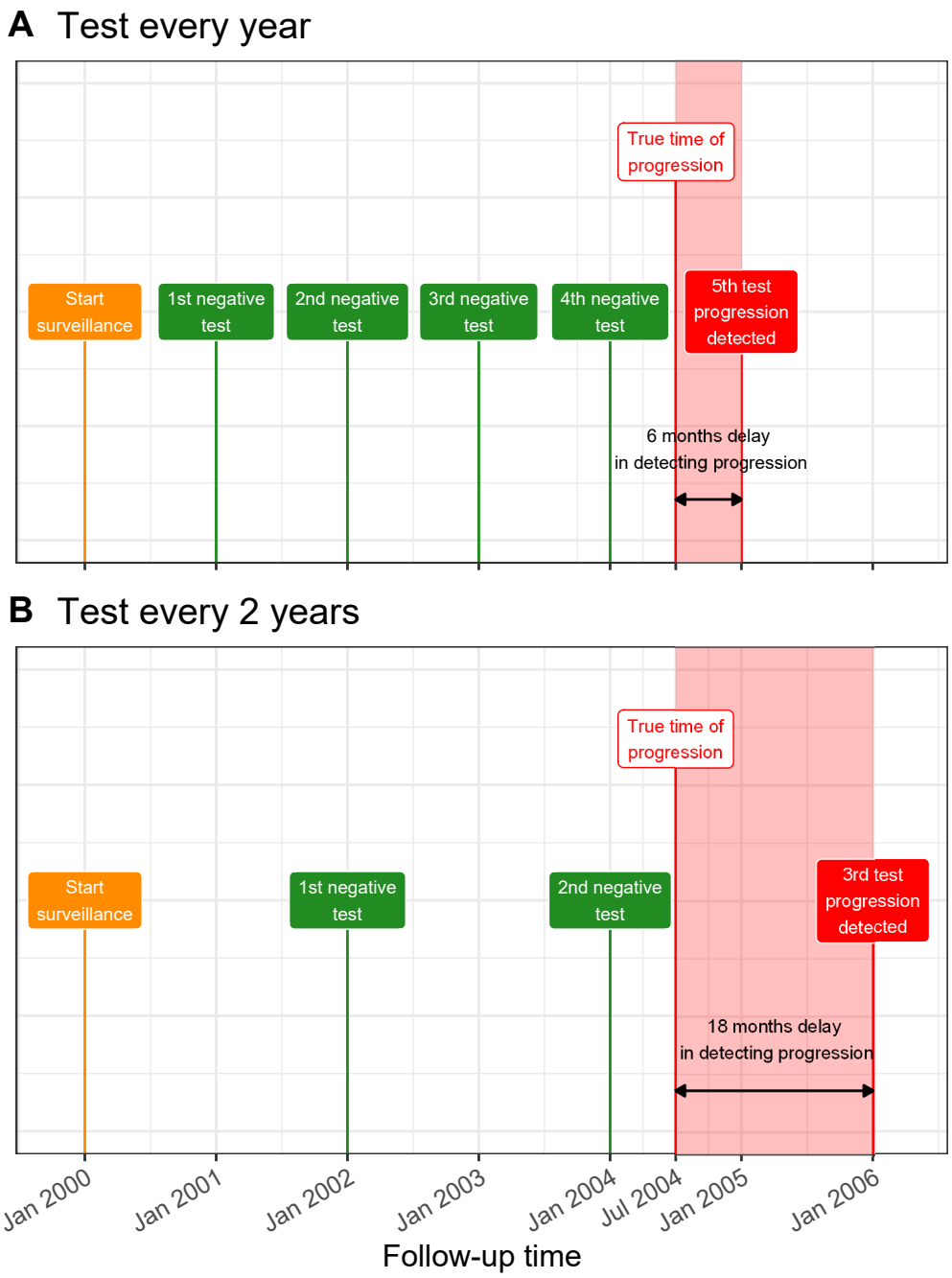


**Figure 1. Trade-off between the number of invasive tests and time delay in detecting progression (non-terminal event of interest):** The true time of progression for this patient July 2004. More frequent invasive tests in **Panel A** lead to a smaller time delay in detection of progression than less frequent invasive tests in **Panel B**. Since invasive tests are conducted periodically, the time of progression is observed as an interval. For example, between Jan 2004–Jan 2005 in **Panel A** and between Jan 2004–Jan 2006 in **Panel B**.

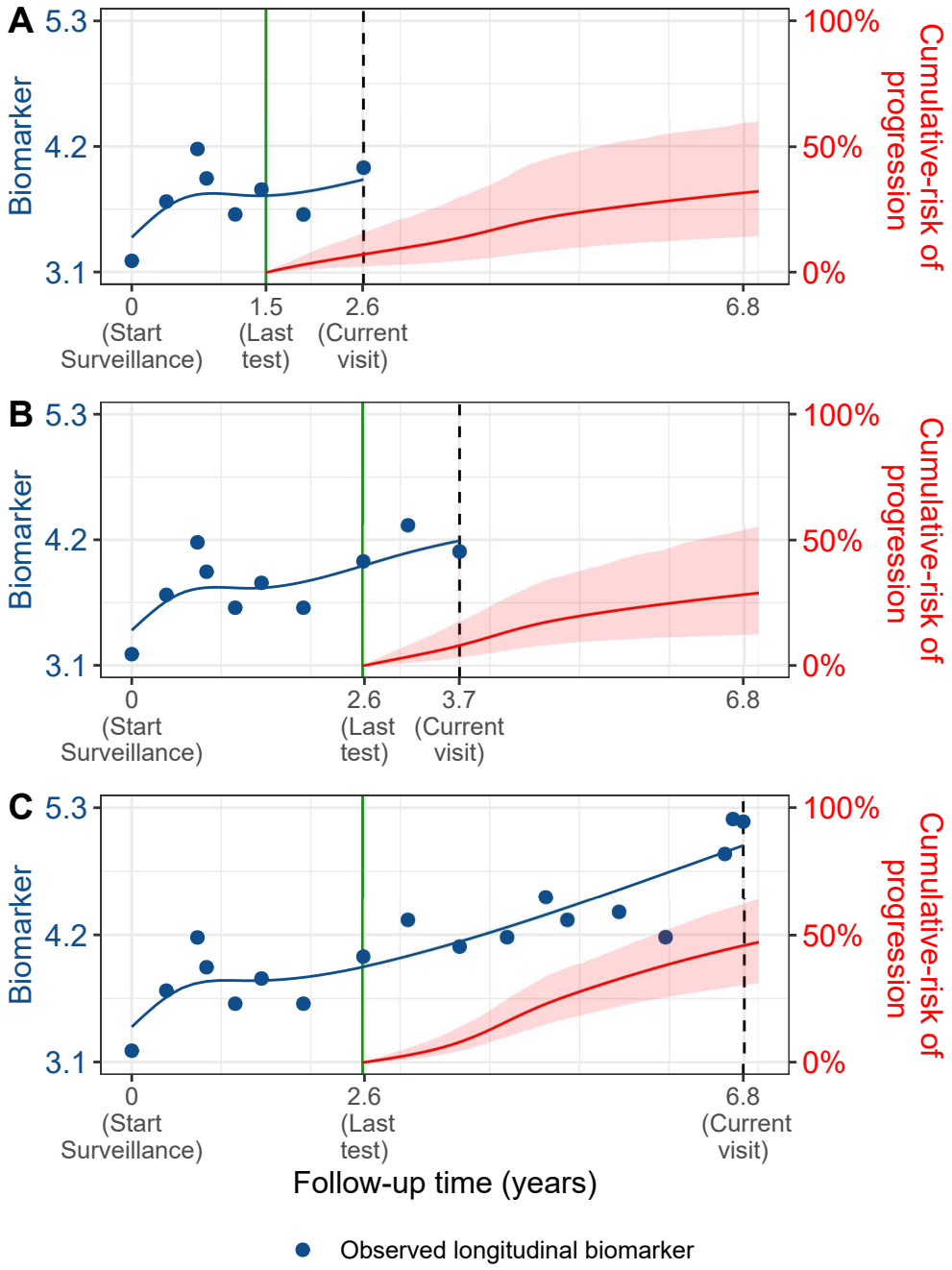


**Figure 2. Cumulative-risk of progression changing dynamically over follow-up** as more patient data is gathered. A single longitudinal outcome, namely, a continuous biomarker of disease progression, is used for illustration. **Panels A, B and C:** are ordered by the time of the current visit (dashed vertical black line) of a new patient. At each of these visits, we combine the accumulated longitudinal measurements (shown in blue), and last time of negative invasive test (solid vertical green line) to obtain the updated cumulative-risk profile (shown in red) of the patient. All values are illustrative.

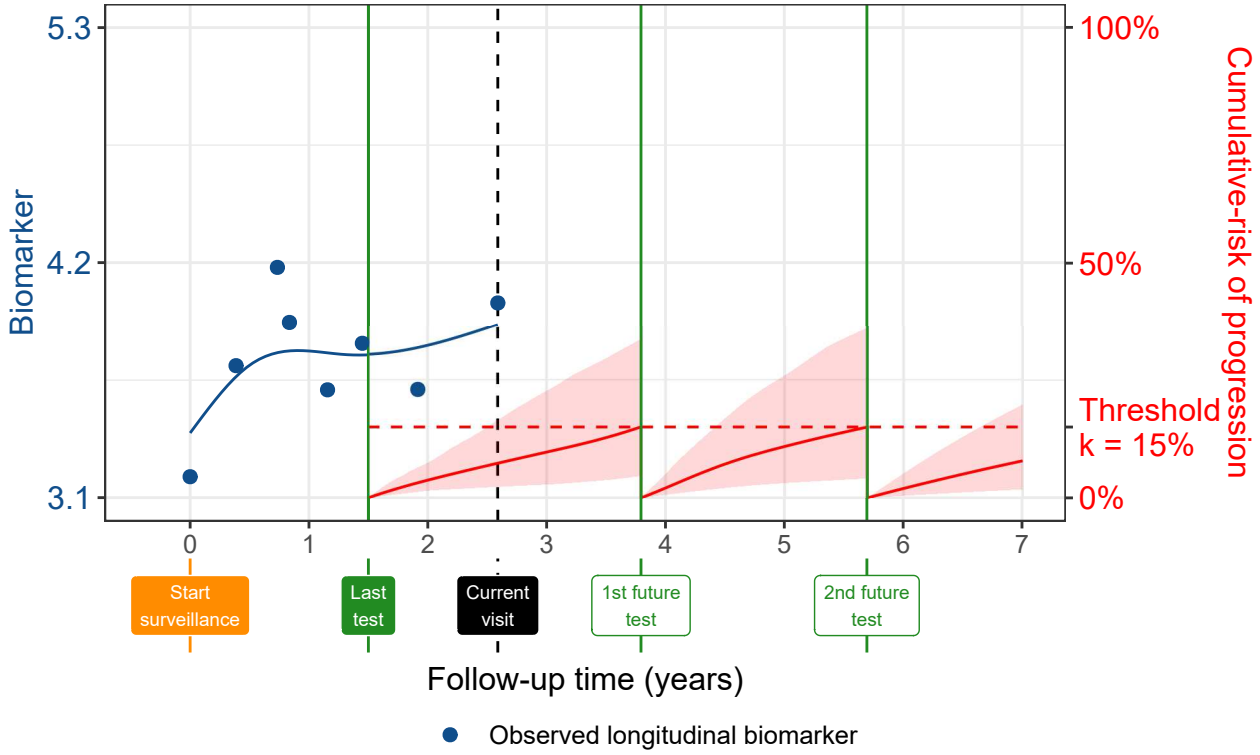


**Figure 3. Personalized Invasive Test Schedule Using Patient-specific Conditional Cumulative-risk of Progression**. A single longitudinal outcome, namely, a continuous biomarker (observed: blue dots, fitted: blue line) of disease progression is used for illustration. The last test on which progression was not observed was conducted at $t = 1.5$ years. The current visit time of the patient is $v = 2.6$ years. Two future invasive tests are scheduled, $S_j^\kappa = (3.8, 5.7)$ years, using a 15% risk threshold ($\kappa = 0.15$). The conditional cumulative-risk profiles $R_j(s_n \mid s_{n-1}, v)$ of Equation (4) are shown with red line (confidence interval shaded). It is called 'conditional' because, for example, the second test at future time 5.7 years, is scheduled after accounting for the possibility that progression (true time $T_j^*$) may not have occurred until the time of the previously scheduled test at $3.8 < T_j^*$ years. All values are illustrative.

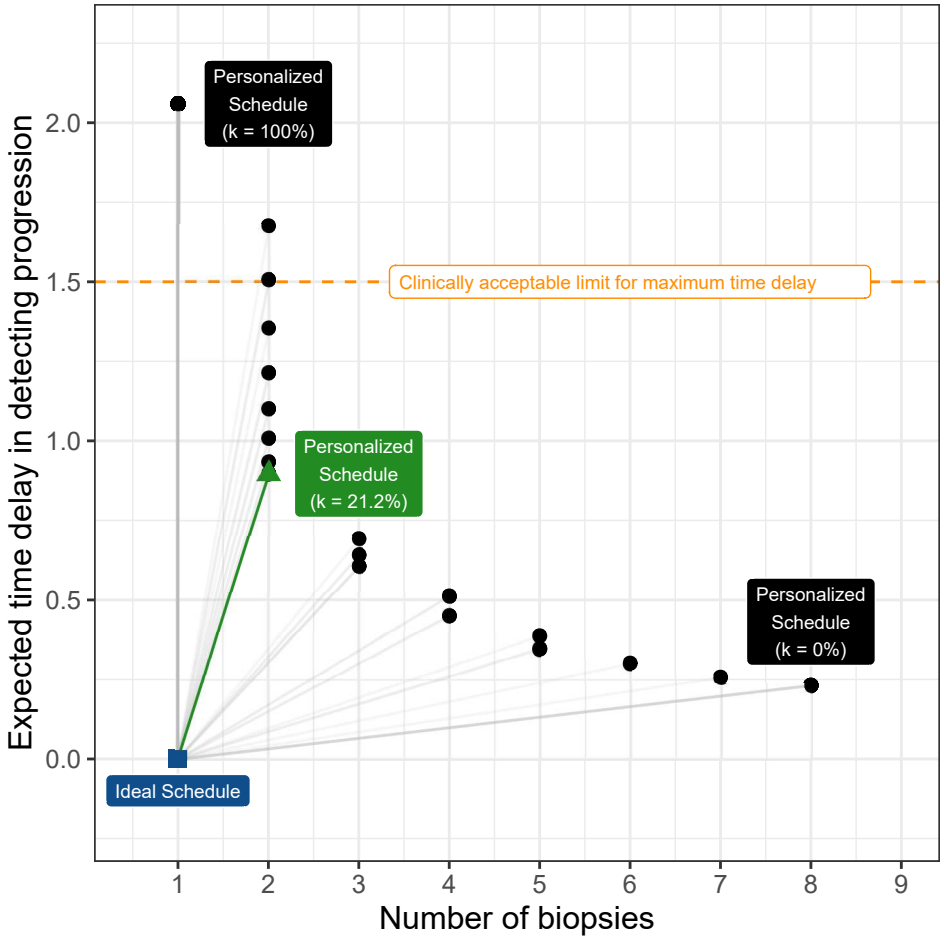


**Figure 4. Automatic choice of risk threshold $0 \leqslant \kappa \leqslant 1$ using Equation (5)**. The ideal schedule of tests at point (1,0) is shown as a blue square. It plans exactly one invasive test at the true time of progression $T_j^*$ of a patient and hence leads to a zero time delay in detection of progression. Personalized schedules based on a grid of thresholds chosen between $0 \leqslant \kappa \leqslant 1$ are shown with black circles. Higher thresholds lead to fewer tests, but also higher expected time delay. We propose to choose the personalized schedule based on $\kappa_a = 21.2\%$ threshold (green triangle). This is because it has the least Euclidean distance (shown with a green line) to the ideal schedule. It is also possible to find the least distance under a certain clinically acceptable limit on time delay (orange dashed line), or number of tests.

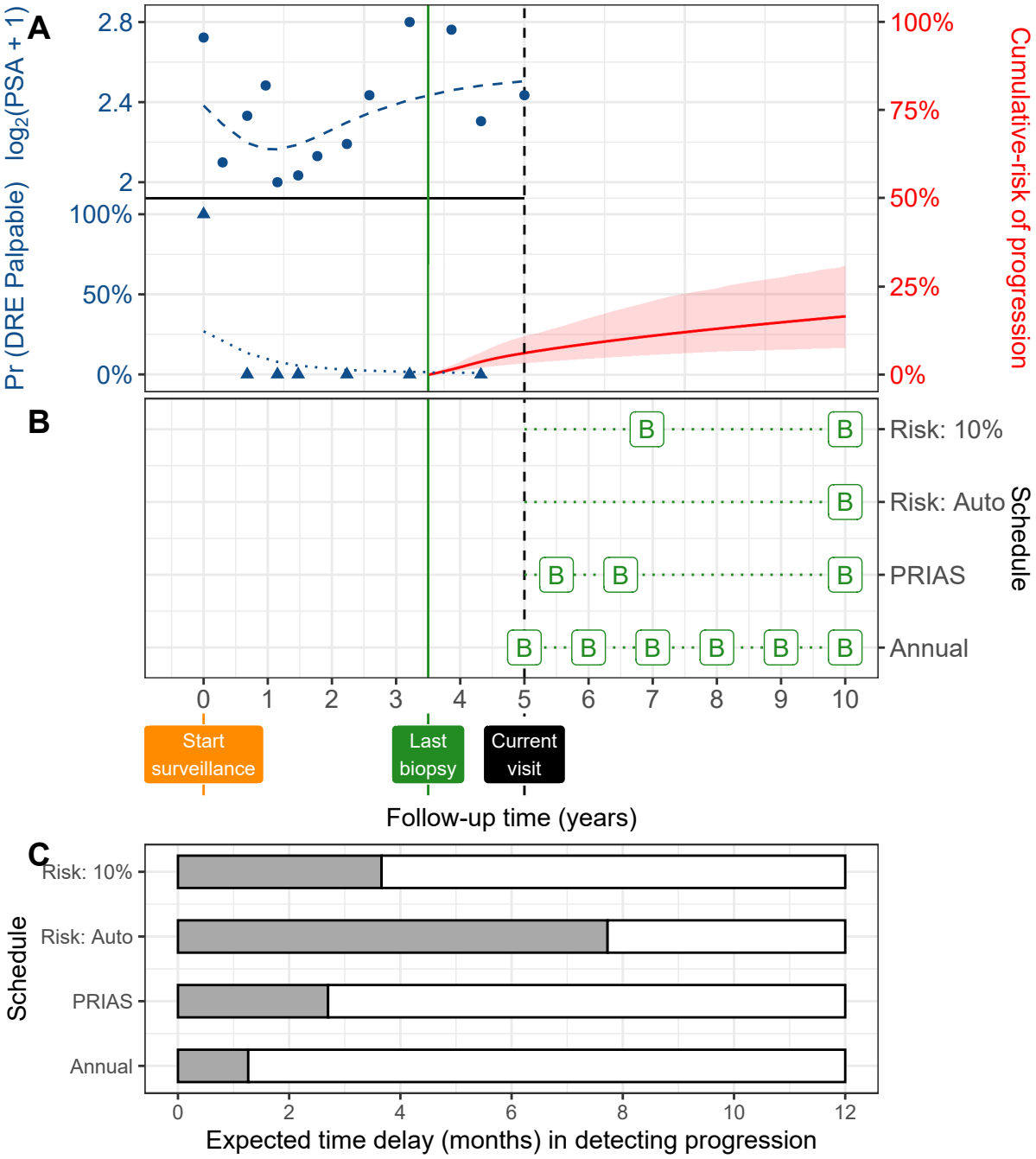


**Figure 5. Demonstration of personalized schedules for a real PRIAS patient:** In **Panel A**: Time of last negative biopsy is year 3.5 (vertical green solid line). Longitudinal data is repeated DRE (blue triangles) and PSA measurements (blue circles). Current visit is year five (vertical black dashed line). Estimated cumulative-risk profile is shown with a solid red line (95%CI is shaded). It is 6% at the current visit and 16.5% at year ten (horizon). In **Panel B**, we visualize different biopsy schedules, with a 'B' indicating a biopsy. **Risk: 10%** and **Risk: Auto** are personalized biopsy schedules using a fixed risk threshold $\kappa = 10\%$, and automatically chosen $\kappa_a$ (Equation 5), respectively. **PRIAS** and **Annual** denote the PRIAS biopsy schedule (paragraph 2 of Section 4) and annual biopsy schedule. **Panel C**: For personalized, PRIAS, and annual biopsy schedules we calculate the corresponding expected time delays in detection of progression (Equation 6). A compulsory biopsy at year ten was planned in all schedules for a fair comparison of expected time delay.

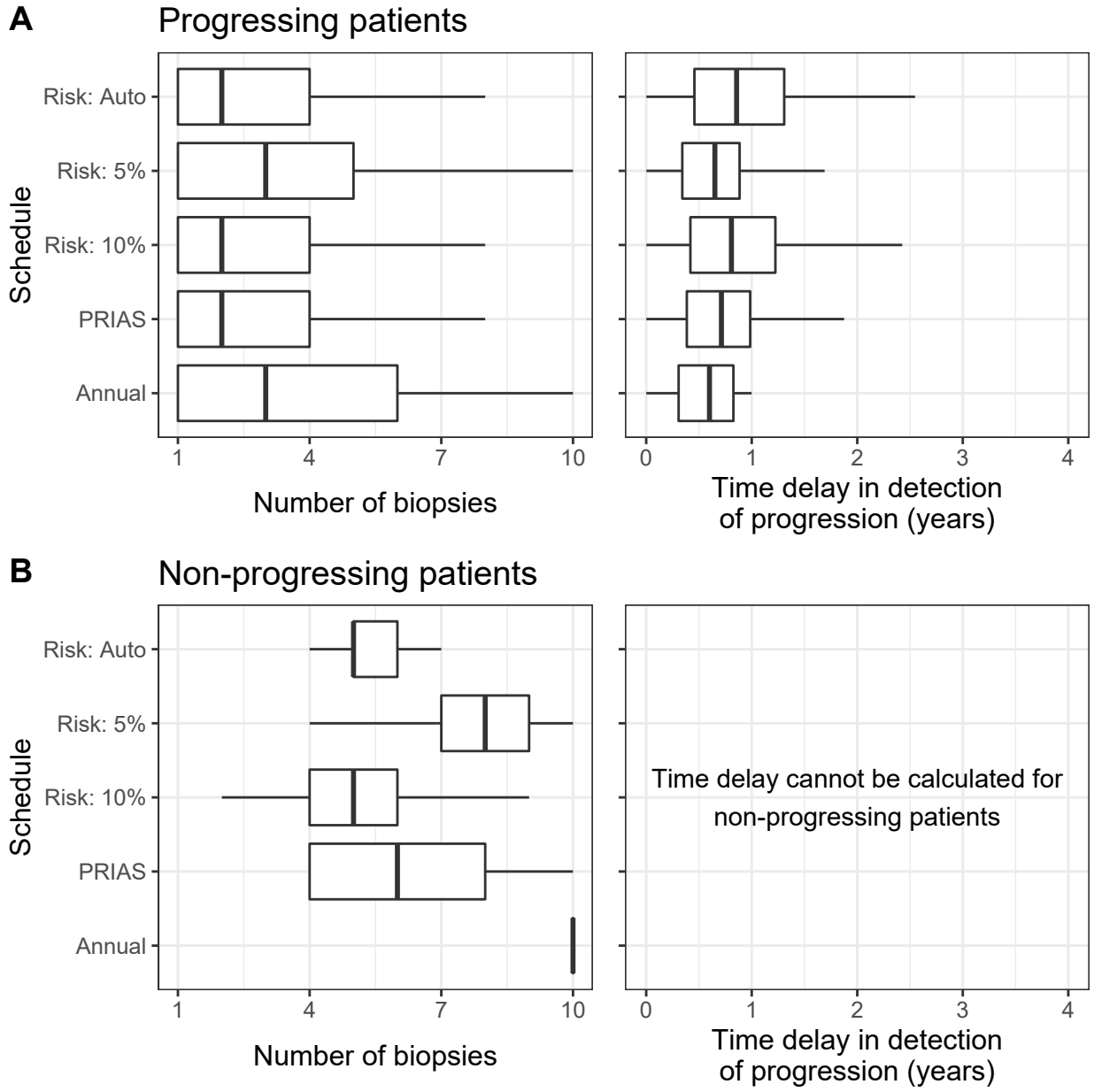


**Figure 6. Boxplot showing variation in the number of biopsies, and the time delay in detection of cancer progression** for various biopsy schedules. Time delay (years) is calculated as (time of positive biopsy - true time of cancer progression). Biopsies are conducted until cancer progression is detected. **Panel A:** results for simulated patients who obtained cancer progression in the ten year study period (*progressing*). **Panel B:** results for simulated patients who did not obtain cancer progression in the ten year study period (*non-progressing*). Types of personalized schedules: Risk: 10% and Risk: 5% approaches, schedule a biopsy if the cumulative-risk of cancer progression at a visit is more than 10% and 5%, respectively. Risk: Auto works similar as previous, except that a visit-specific risk threshold is chosen using Equation (5). Annual corresponds to a schedule of yearly biopsies and PRIAS corresponds to biopsies as per PRIAS protocol (Section 4).