Biometrics 00, 1–16 DOI: 10.1111/j.1541-0420.2005.00454.x
December 0000

Personalized Schedules for Burdensome Surveillance Tests in Chronic Diseases

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: This is the summary for this paper.

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). After diagnosis, in many such patients, surveillance tests are performed periodically to detect disease *progression*, a non-terminal event. Often the most accurate or gold standard surveillance tests are also invasive. For example, to diagnose 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. Test frequency 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 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 detection of *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. This data includes baseline characteristics of patients;

results from previous invasive tests; and longitudinal biomarker, physical examination, and medical imaging measurements, etc. Previous approaches for personalized schedules may 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 latest data point, and ignore the measurement error in observed outcomes. Second, personalized test decisions employing partially observable Markov decision processes (Alagoz et al., 2010; Steimle and Denton, 2017). Although, their application with continuous outcomes is limited by the curse of dimensionality. Third, personalized schedules obtained by optimizing a loss function of 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.

First, we develop a full specification of the joint distribution of patient-specific longitudinal clinical outcomes and 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 over their current and future follow-up visits. These risk predictions utilize their clinical data accumulated until their latest follow-up. We then schedule invasive tests on all those future follow-up visits where a patient's conditional cumulative-risk of progression is above a certain threshold (e.g., 10% risk). We also 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. We estimate this delay in a patient-specific manner for both fixed and personalized schedules, thus facilitating shared-decision making of invasive test schedules.

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, 104904 longitudinal measurements, and 1134 patients with cancer progression. Patients in PRIAS have low and very-low grade prostate cancer, often over-diagnosed due to prostate-specific antigen (PSA) based screening tests (Crawford, 2003). The goal of surveillance is to delay serious treatments (e.g., surgery, chemotherapy, etc.) until cancer progression is observed. 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). The latter is the strongest indicator of cancer-related outcomes. Consequently, treatment is commonly advised upon observing an increase in a patient's biopsy Gleason grade group (cancer progression). Currently, the most common biopsy schedule of yearly biopsies (Loeb et al., 2014) leads 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^*. It 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 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. 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 $\boldsymbol{b}_i = (\boldsymbol{b}_{1i}^T, \ldots, \boldsymbol{b}_{Ki})^T$ 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:

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

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 WebAppendix 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 $j-$th patient. Lastly, $v \geqslant t$ denotes the time of the current follow-up visit.

3.1 *Cumulative-risk of progression*

First we consolidate 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 (Figure 2). It is given by,

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

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,

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

where s_n is the time of the n-th test in the personalized test schedule S_j^{κ}. The conditional cumulative-risk of progression 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 (true time T_j^*) may not have occurred until the time of the previously scheduled test $s_{n-1} < T_j^*$. However, the contribution of the observed longitudinal data $\{\mathcal{Y}_{1j}(v), \ldots, \mathcal{Y}_{Kj}(v)\}$ does not change while scheduling subsequent tests. Since the schedule is risk based, it is also updated as more patient data becomes available over follow-up.

[Figure 3 about here.]

3.3 *Risk Threshold* κ

The risk threshold κ controls the timing and the total number of invasive tests in the schedule S_j^κ. Through the timing of tests, κ also indirectly affects the time delay (Figure 1) that may occur in the detection of progression if this schedule is followed. Hence, κ 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 κ (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 κ_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^κ is followed. Additional consequences of following a particular schedule, such as life years saved, or quality-adjusted life years saved, can also be accommodated in Equation (5). However, this first requires setting a point of optimality in a higher dimensional Euclidean space of the aforementioned consequences.

Certain patients may have preferences for the maximum number of invasive tests conducted upon 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 or 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). Furthermore, the calculation of time delay is not limited to personalized schedules (Equation 4) only. Rather, for any schedule S of N invasive tests, the personalized expected 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 biopsy 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 clinical data consists of longitudinal PSA (continuous; ng/mL) and DRE (binary; tumor palpable or not) measurements (Section 1), 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 joint models that can be utilized for creating personalized biopsy schedules in future PRIAS patients.

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 yearly 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.

4.1 *Fitting the Joint Model to the PRIAS Dataset*

We first fit a joint model to the accumulated clinical data, with $\log_2(\text{PSA} + 1)$ transformed PSA (Lin et al., 2000; Pearson et al., 1994), and DRE as longitudinal measurements, and cancer progression as the event. In the PSA linear mixed-effects sub-model, we utilize B-splines (De Boor, 1978) to allow non-linearly evolution of PSA over follow-up. 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 resulting parameter estimates showed that $\log_2(\text{PSA}+1)$ velocity was the strongest indicator of cancer progression (adjustedhazard ratio...bla bla). The detailed specification of the joint model and the resulting parameter estimates are in Web-Appendix.

4.2 *Personalized Schedules for a Demonstration Patient*

We utilize the fitted joint model to schedule biopsies in a real PRIAS patient (Figure 5). To this end, in line with the protocols of most prostate cancer surveillance cohorts, we first schedule a compulsory biopsy at year one of follow-up (Nieboer et al., 2018). This biopsy promises early detection of progression for patients misdiagnosed as low-grade cancer patients or patients who chose surveillance despite having a higher grade at diagnosis. We also maintain a recommended minimum gap of one year between consecutive biopsies (Bokhorst et al., 2016). Consequently, we schedule personalized biopsies starting from year two. The added benefit of this approach is that due to the clinical data accumulated over two years, we can make reasonably accurate predictions of the cumulative-risk of progression. The follow-up period of PRIAS is limited currently. Hence, we are able to estimate the cumulative-risk of progression (Equation 3) and subsequently schedule biopsies only until year ten of follow-up.

In Figure 5 we can see

[Figure 5 about here.]

5. Simulation Study

Although we demonstrated personalized schedules for a real patient, we also intend to analyze and compare the efficacy of personalized and fixed schedules in a full study cohort. For this purpose, the ideal option of a real world randomized clinical trial is not possible. This is because, in all surveillance programs the true time of progression remains interval-censored.

Second, multiple types of invasive tests cannot be scheduled for the same patient for both practical and ethical reasons. Hence, we conduct an extensive simulation study instead. To keep it realistic, we employ the prostate cancer active surveillance scenario. More specifically, we utilize the joint model fitted to the PRIAS cohort to generate simulated cohorts that are replicas of PRIAS. The simulated population has the same ten year follow-up period as the PRIAS study.

5.1 *Simulation Setup*

From this population, we first sample 500 datasets, each representing a hypothetical surveillance program with 1000 patients in it. We generate a true cancer progression time for each of the 500×1000 patients and then sample a set of 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. We next fit a joint model of the specification given in Equations (**??**), (**??**), and (2) 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 AS programs, we utilize the corresponding fitted joint models to develop cancer progression risk profiles for each of the 500×250 test patients. We make the decision of biopsies for patients at their pre-scheduled follow-up visits for DRE and PSA measurements, on the basis of their estimated personalized cumulative risk of cancer progression. These decisions are made iteratively until a positive biopsy is observed. A recommended gap of one year between consecutive biopsiesBokhorst et al. (2015) is also maintained. Subsequently, for each patient, an entire personalized schedule of biopsies is obtained.

We evaluate and compare both personalized and currently practiced schedules of biopsies in this simulation study. Comparison of the schedules is based on the number of biopsies

scheduled and the corresponding delay in the detection of cancer progression. We evaluate the following currently practiced fixed/heuristic schedules: biopsy annually, biopsy every one and a half years, biopsy every two years and biopsy every three years. We also evaluate the biopsy schedule of the PRIAS program. For the personalized biopsy schedules, we evaluate schedules based on three fixed risk thresholds: 5%, 10%, and 15%, corresponding to a missed cancer progression being 19, 9, and 5.5 times more harmful than an unnecessary biopsy Vickers and Elkin (2006), respectively. We also implement a personalized schedule where for each patient, visit-specific risk thresholds are chosen using F_1 score.

5.2 *Results*

6. Discussion

Put your final comments here. Key points:

1. heuristic schedules are burdensome 2. they do not account for cohort to cohort variation 3. doctors want to utilize information 4. new MRI is coming up....biopsies are still needed though 5. we need to combine information and use it rather than use single items such as flowchart 6. decisions need to be more informative: 7. refer to what we did earlier

Metholodology key points: 1. simple rule is threshold based biopsy 2. how to choose that threshold 3. minimize squared distance 4. constrain time delay to one year 5. or just go by threshold: do not take too much risk 6. or stick to a fixed threshold 7. interpret each threshold by the two pieces of information 8. issues with cutoff choice methods

Figure (all should be clear in black and white and in color): 1. figure of delay explanation 2. figure of dynamic risk 3. figure of scheduling biopsies 4. figure of biopsy schedule in demo patient 5. figure of distance explaining what we did 6. simulation results boxplot

link to web-application

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.

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.

Vickers, A. J. and Elkin, E. B. (2006). Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making* **26,** 565–574.

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`.

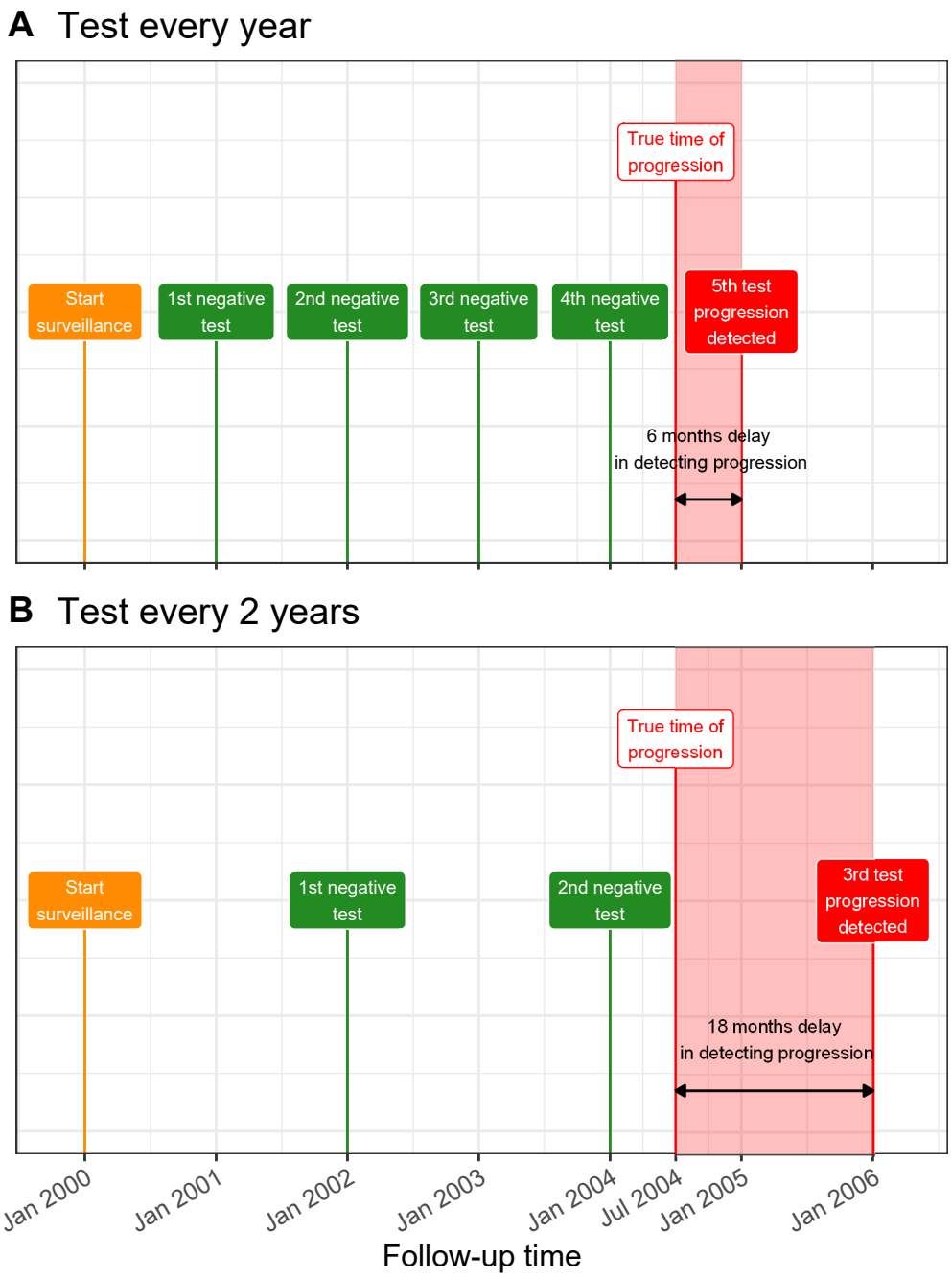

Figure 1. Trade-off between the number of invasive tests and time delay in detecting progression (non-terminal event of interest): The true time of progression for this patient July 2004. More frequent invasive tests in **Panel A** lead to a smaller time delay in detection of progression than less frequent 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**.

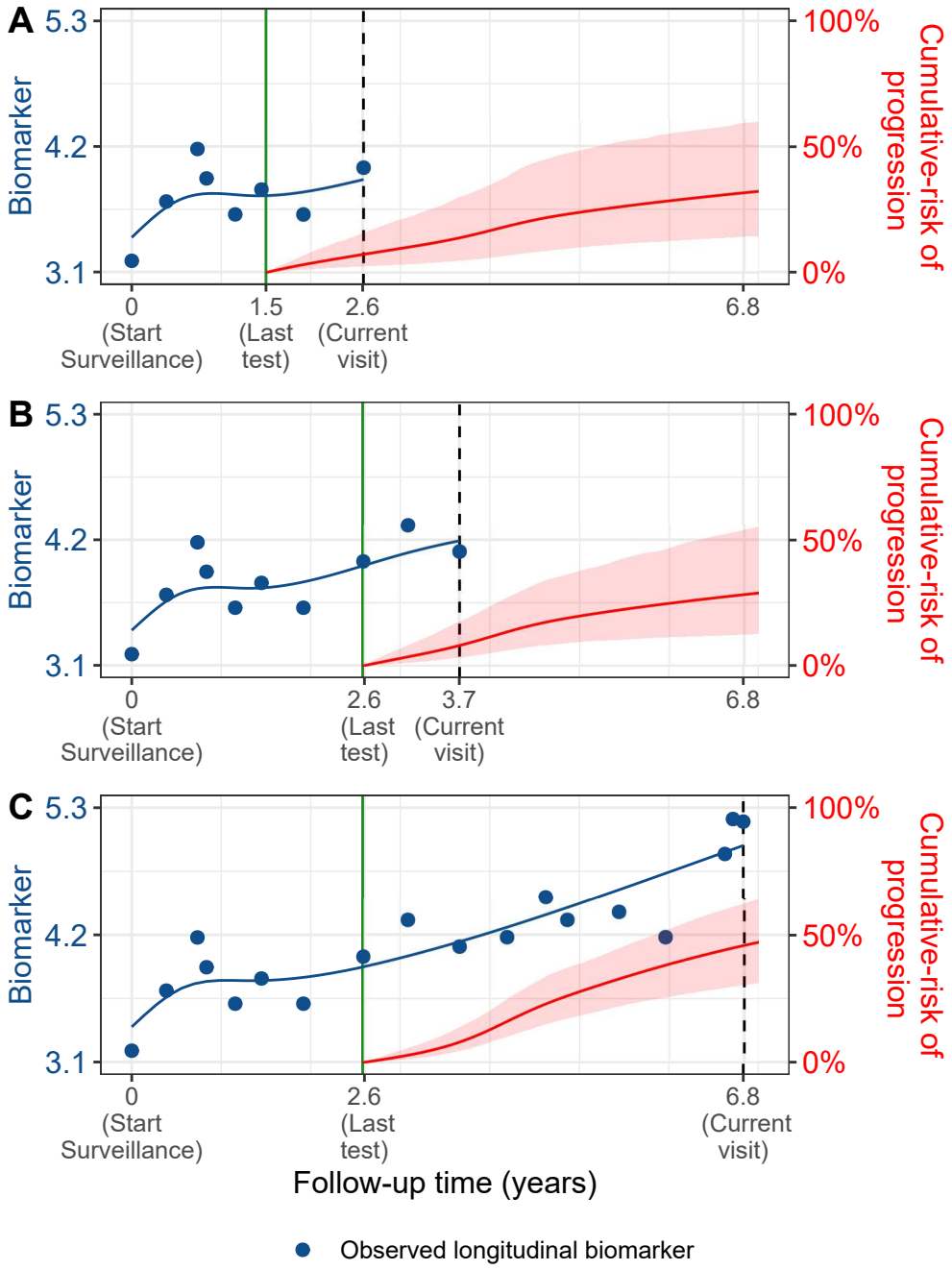

Figure 2. Cumulative-risk of progression changing dynamically over follow-up as more patient data is gathered. A single longitudinal outcome, namely, a continuous biomarker of disease progression, is used for illustration. **Panels A, B and C:** are ordered by the time of the current visit (dashed vertical 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.

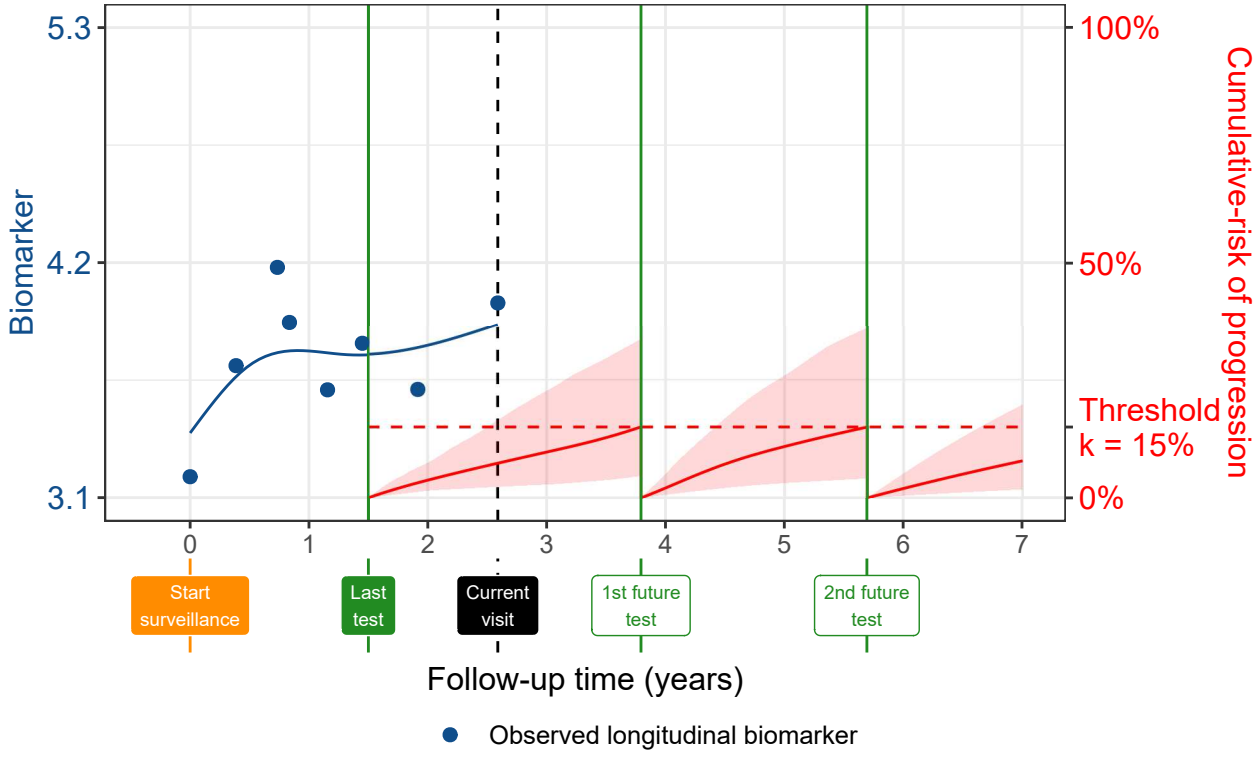

Figure 3. Personalized Invasive Test Schedule Using Patient-specific Conditional Cumulative-risk of Progression. A single longitudinal outcome, namely, a continuous biomarker (observed: blue dots, fitted: blue line) of disease progression is used for illustration. The last test on which progression was 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.

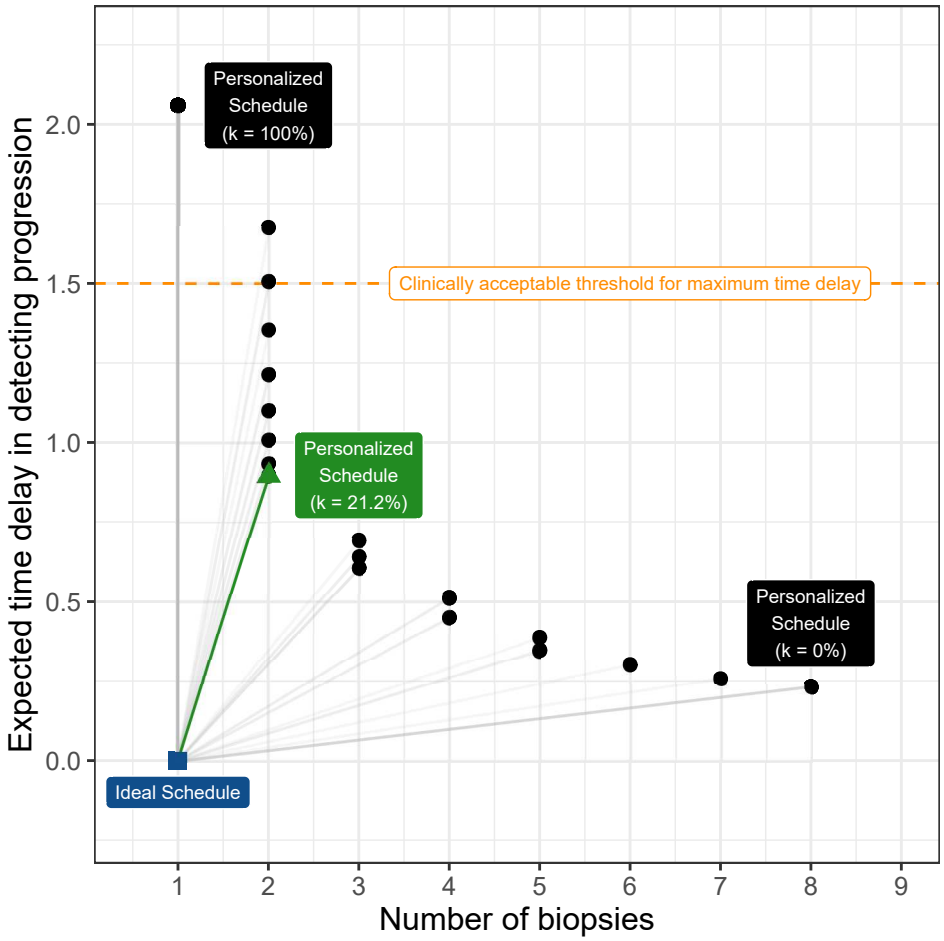

Figure 4. Automatic choice of risk threshold $0 \leqslant \kappa \leqslant 1$ using Equation (5). The ideal schedule of tests at point (1,0) is shown as a blue square. It plans exactly one biopsy at the true time of progression T_j^* of a patient and hence leads to a zero time delay in detection of progression. 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 biopsies, 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 threshold on time delay (orange dashed line), or number of biopsies.

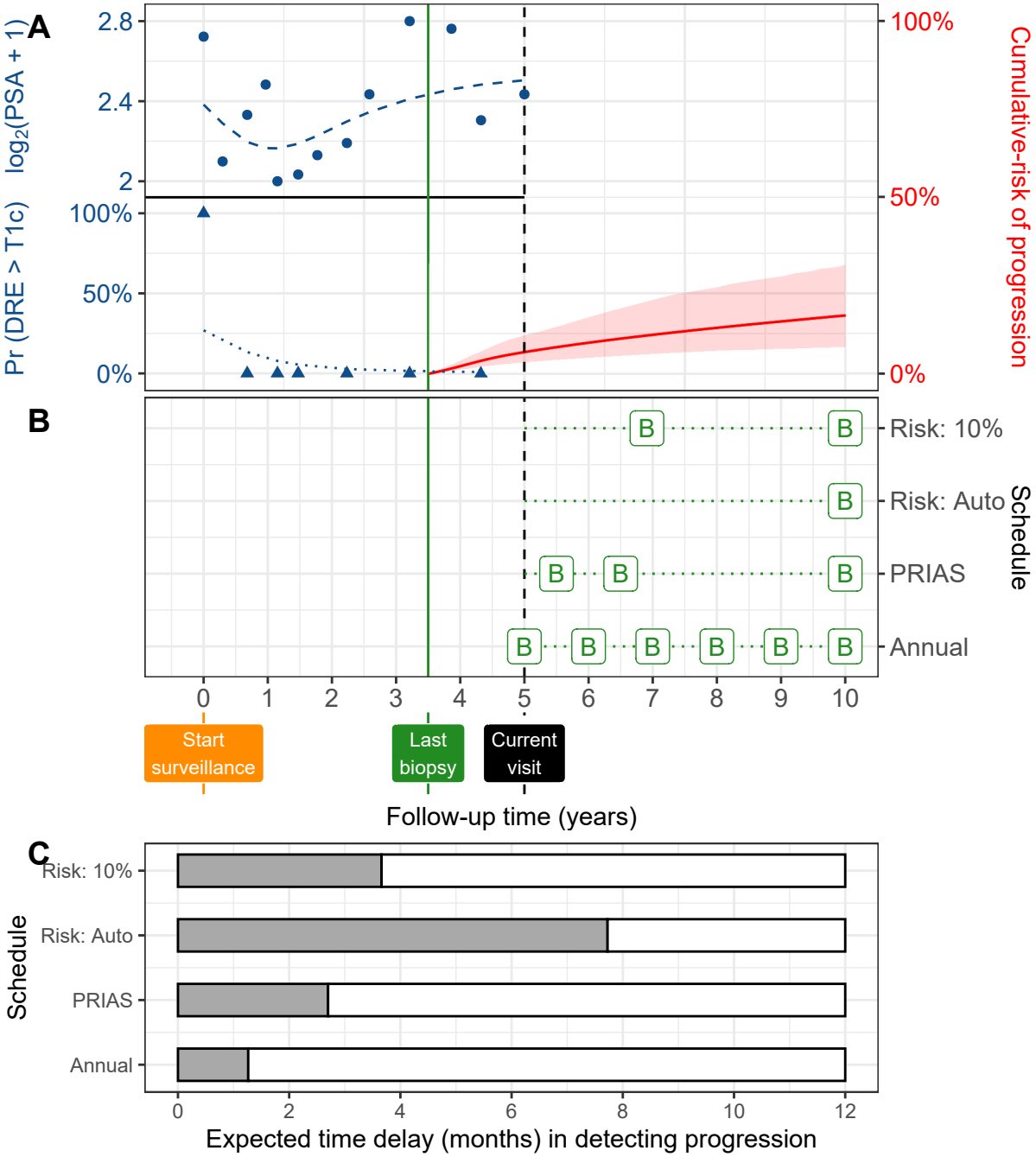

Figure 5. Demonstration as more patient data is gathered. A single longitudinal outcome, namely, a continuous biomarker of disease progression is used for illustration. **Panels A,B and C:** are ordered by the time of the current visit (dashed vertical black line) of a new patient. At each of these visits, 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.