



Shared Decision Making of Burdensome Surveillance Tests Using Personalized Schedules and Their Burden and Benefit

Journal:	<i>Statistics in Medicine</i>
Manuscript ID	Draft
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Tomer, Anirudh; Erasmus MC, Department of Biostatistics Nieboer, Daan; Erasmus MC, Department of Public Health Roobol, Monique; Erasmus MC, Department of Urology Steyerberg, Ewout; LUMC, Department of Biomedical Data Sciences; Erasmus MC, Department of Public Health Rizopoulos, Dimitris; Erasmus MC, Department of Biostatistics
Keywords:	Chronic NCDs, Invasive diagnostic tests, Joint models, Personalized schedules, Shared decision making

SCHOLARONE™
Manuscripts

Shared Decision Making of Burdensome Surveillance Tests Using Personalized Schedules and Their Burden and Benefit

Anirudh Tomer^{*1}, Daan Nieboer², Monique J. Roobol³, Ewout W. Steyerberg^{2,4}, and Dimitris Rizopoulos¹

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

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

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

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

April 6, 2021

*Corresponding author: a.tomter@erasmusmc.nl

The authors gratefully acknowledge Nederlandse Organisatie voor Wetenschappelijk Onderzoek VIDI grant nr. 016.146.301, and Erasmus University Medical Center funding.

Abstract

Benchmark surveillance *tests* for detecting disease *progression* (e.g., biopsies, endoscopies) in early-stage chronic non-communicable diseases (e.g., cancer, lung diseases) are usually burdensome. For detecting progression timely, patients undergo invasive tests planned in a fixed one-size-fits-all manner (e.g., annually). We aim to present personalized test schedules based on the risk of disease progression, that optimize the burden (the number of tests) and the benefit (shorter time delay in detecting progression is better) better than fixed schedules, and enable shared decision making. Our motivation comes from the problem of scheduling biopsies in prostate cancer surveillance.

Using joint models for time-to-event and longitudinal data, we consolidate patients' longitudinal data (e.g., biomarkers) and results of previous tests, into individualized future cumulative-risk of progression. We then create personalized schedules by planning tests on future visits where the predicted cumulative-risk is above a *threshold* (e.g., 5% risk). We update personalized schedules with data gathered over follow-up. To find the optimal risk threshold, we minimize a utility function of the expected number of tests (burden) and expected time delay in detecting progression (shorter is beneficial) for different thresholds. We estimate these two in a patient-specific manner for following any schedule, by utilizing a patient's predicted risk profile. Patients/doctors can employ these quantities to compare personalized and fixed schedules objectively and make a shared decision of a test schedule.

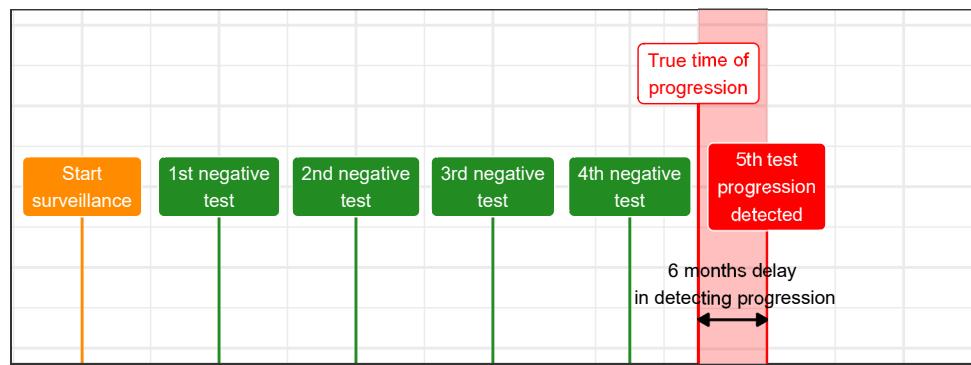
Keywords— Chronic NCDs, Invasive diagnostic tests, Joint models, Personalized schedules, Shared decision making

1 Introduction

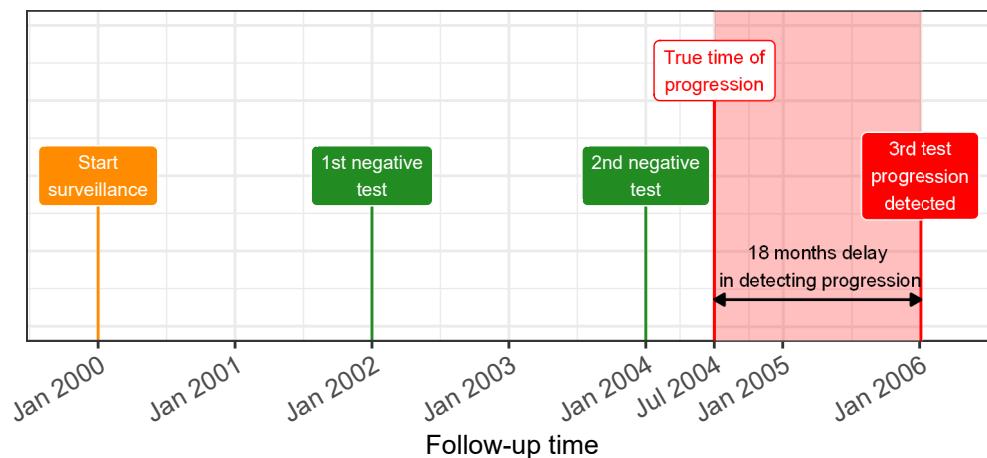
Chronic non-communicable diseases (e.g., cancer, lung, cardiovascular diseases) cause 60–70% of human deaths worldwide [1]. Often patients diagnosed with an early-stage disease undergo surveillance *tests* to detect disease *progression* timely. A progression is a non-terminal event, and usually a trigger for treatment and/or removal from surveillance. Benchmark tests used for confirming progression are usually *invasive*, e.g., biopsies in prostate cancer surveillance [2], endoscopies in Barrett’s esophagus [3], colonoscopies in colorectal cancer [4], and bronchoscopies in post lung transplant [5] surveillance. These invasive tests are repeated until progression is observed, typically as per a one-size-fits-all *fixed schedule*, e.g., biannually, [4, 5, 2]. A time gap between tests causes a time delay in detecting progression (Figure 1). A shorter delay in detecting progression (*benefit*) can provide a larger window of opportunity for curative treatment. However, with fixed schedules, this means conducting tests frequently. Frequent tests are *burdensome* as they may cause pain and/or severe medical complications [4, 6]. Consequently, patients may not always comply with frequent tests [2, 7]. In general, because fixed schedules do not differentiate between fast and slow/non-progressing patients, they impose disproportionate burden/benefits across the patient population.

The goal of this work is two-fold. First, to create *personalized* invasive test schedules that better optimize the burden (number of invasive tests) and benefit (shorter delay in detecting progression) of invasive tests than fixed schedules (Figure 1). To this end, we intend to use the patients' clinical data accumulated over the surveillance follow-up. This data includes baseline characteristics, previous test results, and longitudinal outcomes (e.g., biomarkers, medical imaging, physical examination). Second we develop a methodology to estimate the burden (number of invasive tests) and benefit (shorter time delay in detecting progression) of invasive test schedules. We intend to use these criteria for comparing test schedules objectively and subsequently enable shared decision making while choosing a test schedule.

1
2
3 **A Frequent tests - Shorter delay in detecting progression**



16
17 **B Infrequent tests - Longer delay in detecting progression**



34 **Figure 1: Goal: Finding the optimal tradeoff between the number of invasive tests**
 35 **(burden) and time delay in detecting progression (shorter is beneficial).** A progression
 36 is a non-terminal event in the surveillance of early-stage chronic non-communicable diseases. The
 37 true time of progression for the patient illustrated in this figure is July 2004. Since invasive
 38 tests are conducted repeatedly, progression is interval-censored and always observed with a delay.
 39 Frequent periodical invasive tests in **Panel A** lead to a shorter time delay in detecting progression
 40 than infrequent periodical invasive tests in **Panel B**. The interval-censored time of progression is
 41 Jan 2004–Jan 2005 in **Panel A** and between Jan 2004–Jan 2006 in **Panel B**.

The idea of personalizing the invasive test schedules is not new. In fact, currently, some surveillance protocols personalize test schedules using heuristic methods such as decision flowcharts [2, 3]. However, flowcharts discretize continuous outcomes, often exploit only the last measurement, ignore the measurement error in observed data, and plan only one test at a time. Alternatively, a complete personalized schedule of tests can be obtained using partially observable Markov decision processes or POMDPs [8, 9]. Although, POMDPs typically discretize continuous longitudinal outcomes to avoid the curse of dimensionality. Besides, in scenarios such as ours, where decisions (test/no test) and disease state (low-grade disease/progressed) are both binary, POMDPs may not be necessary either. The reason is that such POMDPs give the same optimal schedule, which can be alternatively obtained by just planning a test when the probability of transition from non-progressed to progressed state is more than a certain threshold [see 10, Equation 1].

Personalized schedules have also been obtained by optimizing an explicit utility function of the burden and/or benefit of a schedule. A challenge in this approach is quantifying burden and benefit. In our previous work [11], we quantified the burden and benefit as the time difference by which a future test undershoots (unnecessary test) or overshoots (delayed detection) the true progression time of a patient, respectively. These choices limited us to plan only one future test at a time. Others [12] proposed obtaining a complete test schedule by quantifying the burden of a test schedule as the expected number of tests and benefit as expected time delay in detecting progression. To obviate the issue that the number of tests and delay have different scales and units, they proposed scheduling tests when the risk of progression is above a threshold. Schedules based on risk threshold have also been proposed previously [10, 13]. The clinical interpretation of risk and the choice of risk threshold is not straightforward. In our previous work on risk-based test schedules [14] we, and others[15], motivated the choice of risk threshold to be based on measures of diagnostic accuracy (e.g., false positive rate, true positive rate). However, measures of diagnostic accuracy are not personalized criteria for choosing risk thresholds. Besides, a single risk-based test decision does not inform patients about future clinical consequences of continuing

1
2
3 on surveillance.
4

5 In this paper, we make two significant improvements over our own two previous works on the
6 same topic [11, 14]. First, instead of planning one test at a time, we derive full risk-based per-
7 sonalized test schedules. Thus, at any follow-up visit, patients know the time of all future tests
8 planned for them. The personalized schedules also dynamically update with new clinical data
9 over follow-up. Second, along with each schedule, we provide patients the clinical consequences
10 of following it. Namely, the expected number of tests required out of all planned tests to detect
11 progression and the expected time delay in detecting progression. There are three advantages of
12 using these two criteria for schedule selection instead of measures of diagnostic accuracy previ-
13 ously proposed by us [14] and others [15]. First, by using our proposed criteria we can evaluate
14 the performance of a complete schedule and not just a single test decision. Second, the proposed
15 criteria are easily-quantifiable surrogates for important clinical aspects such as the window of
16 opportunity for curative treatment, risk of adverse outcomes due to delayed detection of pro-
17 gression, financial costs of tests, risk of side-effects, and reduction in quality of life, etc. Third,
18 we calculate the expected number of tests and delay in a personalized manner, an improvement
19 over previous work by others [12]. Hence, for any schedule, fixed or personalized, patients can
20 objectively compare the clinical consequences of opting for them. This can enable shared decision
21 making of invasive test schedules.
22
23

24 The basic idea behind our new approach is as follows. We first develop a full specification of
25 the joint distribution of the patient-specific longitudinal outcomes and the time of progression.
26 To this end, we utilize joint models for time-to-event and longitudinal data [16, 17] because they
27 are inherently personalized. Specifically, joint models utilize patient-specific random effects [18]
28 to model longitudinal outcomes without discretizing them. Subsequently, we input clinical data
29 of a new patient into the fitted model to obtain their predicted patient-specific cumulative-risk
30 of progression at future visits. We then create personalized schedules by planning tests on future
31 visits where this predicted cumulative-risk is above a particular *threshold* (e.g., 5% risk). We
32
33

automate the choice of this threshold and the resulting schedule by optimizing a utility function of the expected number of tests and time delay in detecting progression for personalized schedules. To estimate these two quantities in a patient-specific manner we use patient's predicted risk profiles. Hence, patients/doctors can compare the consequences of opting for personalized versus fixed schedules objectively.

Our motivation comes from the task of scheduling biopsies in the world's largest prostate cancer surveillance study, called Prostate Cancer Research International Active Surveillance [2], or PRIAS. It has 7813 low/very-low grade cancer patients (1134 progressions, 104904 longitudinal measurements), many of whom are potentially over-diagnosed due to prostate-specific antigen (PSA) based screening [19]. To reduce subsequent over-treatment, in surveillance, serious treatments (e.g., surgery, radiotherapy) are delayed until progression is observed. Surveillance involves regular monitoring of a patient's PSA (ng/mL), digital rectal examination or DRE (tumor shape/size), and biopsy Gleason grade group [20]. Among these, a biopsy Gleason grade group ≥ 2 is the reference test for confirming progression. Most often, biopsies are scheduled annually [21]. However, such a frequent schedule can put an unnecessary burden on patients with slow/non-progressing cancers and cause non-compliance [2]. Since prostate cancer has the second-highest incidence among all cancers in males [22], individualized biopsy schedules can reduce the burden of biopsies in numerous patients worldwide.

The remaining paper is as follows. Section 2 introduces the joint modeling framework. The personalized scheduling methodology is described in Section 3, and demonstrated for prostate cancer surveillance patients in Section 4. In Section 5, we compare personalized and fixed schedules via a realistic simulation study based on a joint model fitted to the PRIAS dataset.

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

Let T_i^* denote the true time of disease progression for the i -th patient. Progression is always interval censored $l_i < T_i^* \leq r_i$ (Figure 1). Here, r_i and l_i denote the time of the last and second last invasive tests, respectively, when patients progress. In non-progressing patients, l_i denotes the time of the last test and $r_i = \infty$. Assuming K types of longitudinal outcomes, let \mathbf{y}_{ki} denote the $n_{ki} \times 1$ longitudinal response vector of the k -th outcome, $k \in \{1, \dots, K\}$. The observed data of all n patients is given by $\mathcal{A}_n = \{l_i, r_i, \mathbf{y}_{1i}, \dots, \mathbf{y}_{Ki}; i = 1, \dots, n\}$.

2.1 Longitudinal Sub-process

To model multiple longitudinal outcomes in a unified framework, a joint model employs individual generalized linear mixed sub-models [18]. Specifically, the conditional distribution of the k -th outcome \mathbf{y}_{ki} given a vector of patient-specific random effects \mathbf{b}_{ki} is assumed to belong to the exponential family, with linear predictor given by,

$$g_k[E\{\mathbf{y}_{ki}(t) | \mathbf{b}_{ki}\}] = m_{ki}(t) = \mathbf{x}_{ki}^\top(t)\boldsymbol{\beta}_k + \mathbf{z}_{ki}^\top(t)\mathbf{b}_{ki},$$

where $g_k(\cdot)$ denotes a known one-to-one monotonic link function, $y_{ki}(t)$ is the value of the k -th longitudinal outcome for the i -th patient at time t , and $\mathbf{x}_{ki}(t)$ and $\mathbf{z}_{ki}(t)$ are the time-dependent design vectors for the fixed $\boldsymbol{\beta}_k$ and random effects \mathbf{b}_{ki} , respectively. To model the correlation between different longitudinal outcomes, we link their corresponding random effects. Specifically, we assume that the vector of random effects $\mathbf{b}_i = (\mathbf{b}_{1i}^\top, \dots, \mathbf{b}_{Ki}^\top)^\top$ follows a multivariate normal distribution with mean zero and variance-covariance matrix W .

2.2 Survival Sub-process

In the survival sub-process, the hazard of progression $h_i(t)$ at a time t is assumed to depend on a function of patient and outcome-specific linear predictors $m_{ki}(t)$ and/or the random effects,

$$h_i\{t \mid \mathcal{M}_i(t), \mathbf{w}_i(t)\} = h_0(t) \exp \left[\boldsymbol{\gamma}^\top \mathbf{w}_i(t) + \sum_{k=1}^K f_k\{\mathcal{M}_{ki}(t), \mathbf{w}_i(t), \mathbf{b}_{ki}, \boldsymbol{\alpha}_k\} \right], \quad t > 0,$$

where $h_0(\cdot)$ denotes the baseline hazard, $\mathcal{M}_{ki}(t) = \{m_{ki}(s) \mid 0 \leq s < t\}$ is the history of the k -th longitudinal process up to t , and $\mathbf{w}_i(t)$ is a vector of exogenous, possibly time-varying covariates with regression coefficients $\boldsymbol{\gamma}$. Functions $f_k(\cdot)$, parameterized by vector of coefficients $\boldsymbol{\alpha}_k$, specify the features of each longitudinal outcome that are included in the linear predictor of the relative-risk model [23, 17, 24]. Some examples, motivated by the literature (subscripts k dropped for brevity), are,

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

These formulations of $f(\cdot)$ postulate that the hazard of progression at time t may depend on the underlying level $m_i(t)$ (e.g., PSA value in prostate cancer) or on both the level and velocity $m'_i(t)$ (e.g., PSA velocity) of the longitudinal outcome at t . Lastly, the baseline hazard $h_0(t)$ is modeled flexibly using P-splines [25]. The detailed specification of the baseline hazard, and the joint parameter estimation of the longitudinal and relative-risk sub-models using the Bayesian approach are presented in Supplementary A.

3 Personalized Schedule of Invasive Tests for Detecting Progression

3.1 Cumulative-risk of progression

Using the joint model fitted to the training data \mathcal{A}_n , we aim to derive a personalized schedule of invasive tests for a new patient j with true progression time T_j^* . To this end, our calculations exploit the *cumulative-risk* function. Let $t < T_j^*$ be the time of the last conducted test at which progression was not observed. Let $\{\mathcal{Y}_{1j}(v), \dots, \mathcal{Y}_{Kj}(v)\}$ denote the history of observed longitudinal data up to the current visit time v . The current visit can be after the last negative test, i.e., $v \geq t$ (e.g., PSA after negative biopsy in prostate cancer). The cumulative-risk of progression for patient j at future time u is then given by,

$$\begin{aligned} R_j(u | t, v) &= \Pr\{T_j^* \leq u | T_j^* > t, \mathcal{Y}_{1j}(v), \dots, \mathcal{Y}_{Kj}(v), \mathcal{A}_n\} \\ &= \int \int \Pr(T_j^* \leq u | T_j^* > t, \mathbf{b}_j, \boldsymbol{\theta}) p\{\mathbf{b}_j | T_j^* > t, \mathcal{Y}_{1j}(v), \dots, \mathcal{Y}_{Kj}(v), \boldsymbol{\theta}\} \\ &\quad \times p(\boldsymbol{\theta} | \mathcal{A}_n) d\mathbf{b}_j d\boldsymbol{\theta}, \quad u \geq t. \end{aligned} \tag{1}$$

The cumulative-risk function $R_j(\cdot)$ depends on patient-specific clinical data and the training dataset, via the posterior distribution of the random effects \mathbf{b}_j and posterior distribution of the vector of all parameters $\boldsymbol{\theta}$ of the fitted joint model, respectively. This cumulative-risk function is dynamic, in the sense that it automatically updates over time as more longitudinal data become available (Figure 2).

3.2 Personalized Test Decision Rule

In our previous works [11, 14], we used the cumulative-risk function in (1) to optimize loss functions inspired from Bayesian decision theory for deciding the time of an invasive test for the j -th patient. However, this approach assumed that only one test can be conducted for the patient.

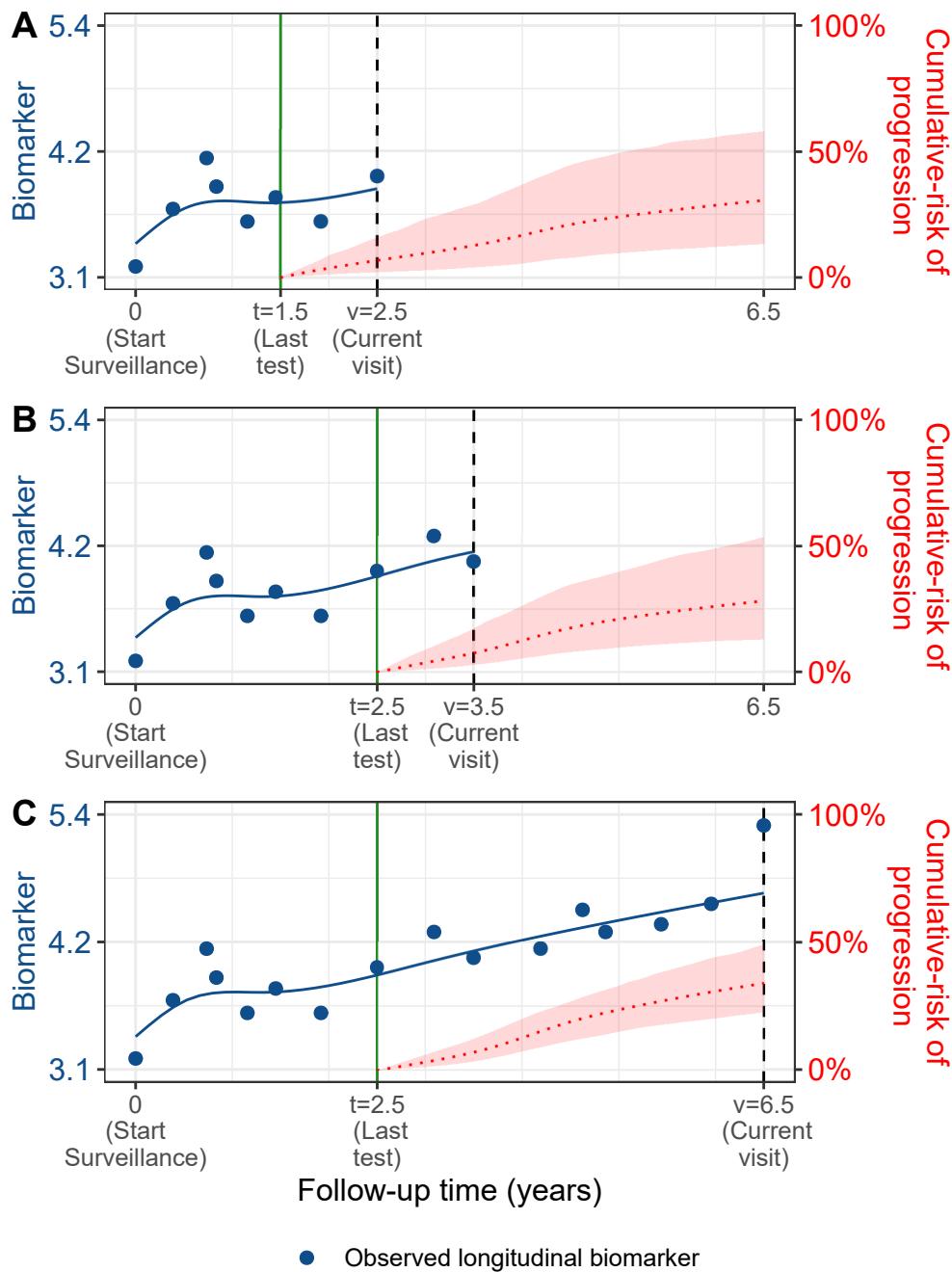


Figure 2: Cumulative-risk of progression updated dynamically over follow-up as more patient data is gathered. A single longitudinal outcome, namely, a continuous biomarker of disease progression, is used for illustration. **Panels A, and B:** are ordered by the time of the current visit v (dashed vertical black line) of a new patient. At each of these visits, we combine the accumulated longitudinal data (shown in blue circles), and time of the last negative invasive test t (solid vertical green line) to obtain the updated cumulative-risk profile $R_j(u | t, v)$ (dotted red line with 95% credible interval shaded) of the patient defined in (1). All values are illustrative.

Consequently, patients could foresee neither the future tests that may be required nor the clinical consequences of opting for such future tests. Hence, in this work we intend to exploit the whole cumulative-risk profile over time $R_j(\cdot)$ to develop full risk-based personalized schedule of invasive tests. In addition, we aim to use this patient-specific cumulative-risk function to estimate the burden (number of tests required) and benefit (time delay in detecting progression) of each schedule we develop for the patients.

Typically, the decision to undergo an invasive test is made on the same visit times on which longitudinal data (e.g., biomarkers) are measured. Let $U = \{u_1, \dots, u_L\}$ represent a schedule of such visits (e.g., biannual PSA measurement in prostate cancer). Here, $u_1 = v$ is also the current visit time. The maximum future visit time u_L can be chosen based on the available information in the training dataset \mathcal{A}_n . That is, tests for the new patient j are planned only up to a future visit time u_L at which a sufficient number of events in \mathcal{A}_n are available for making reliable risk predictions (e.g., up to the 80% or 90% percentile of progression times).

We propose to take the decision of conducting a test at a future visit time $u_l \in U$ if the cumulative-risk of progression at time u_l exceeds a certain risk threshold κ (Figure 3). In particular, the test decision at time u_l is given by,

$$Q_j^\kappa(u_l | t_l, v) = I\{R_j(u_l | t_l, v) \geq \kappa\}, \quad 0 \leq \kappa \leq 1, \quad (2)$$

where $I(\cdot)$ is the indicator function, $R_j(u_l | t_l, v)$ is the cumulative-risk of progression at the current decision time u_l , and $t_l < u_l$ is the time of the last test conducted before u_l . Thus, the future time at which a test will be planned, depends on both the threshold κ and the cumulative-risk of the patient. Moreover, when a test gets planned at time u_l , i.e., $Q_j^\kappa(u_l | t_l, v) = 1$, then the cumulative-risk profile is updated before making the next test decision at time u_{l+1} (Figure 3). Specifically, the cumulative-risk at time u_{l+1} is updated by setting the corresponding time of the last test $t_{l+1} = u_l$. This accounts for the possibility that progression may occur after time

1
2
3 $u_l < T_j^*$. Hence, the time of last test t_l is defined as,
4
5

$$6 \quad t_l = \begin{cases} 7 \quad t, & \text{if } l = 1, \\ 8 \quad u_{l-1}, & \text{if } l \geq 2 \text{ and } Q_j^\kappa(u_{l-1} | t_{l-1}, v) = 1, \\ 9 \quad t_{l-1}, & \text{if } l \geq 2 \text{ and } Q_j^\kappa(u_{l-1} | t_{l-1}, v) = 0. \end{cases}$$

10
11 We should note that in all future test decisions, we use only the observed longitudinal data up to
12
13 the current visit time v , i.e., $\{\mathcal{Y}_{1j}(v), \dots, Y_{Kj}(v)\}$.
14
15

16 3.3 Expected Number of Tests and Expected Time Delay in De- 17 18 tecting Progression

19
20 To facilitate shared-decision making of invasive tests, we translate our proposed decision rule, i.e.,
21
22 the choice of a specific risk threshold κ , into two clinically relevant quantities. First, the number
23
24 of tests (burden) we expect to perform for patient j , and second, if the patient progresses, the time
25
26 delay (shorter is beneficial) expected in detecting progression. To calculate these two quantities,
27
28 we first suppose that patient j does not progress between his last negative test at time t and the
29
30 maximum future visit time u_L . Under this assumption, the subset of future visit times in U on
31
32 which a test is planned using (2) results into a personalized schedule of future tests (Figure 3),
33
34 given by,

$$35 \quad \{s_1, \dots, s_{N_j}\} = \{u_l \in U : Q_j^\kappa(u_l | t_l, v) = 1\}, \quad N_j \leq L. \quad (3)$$

36
37 If patient j never progressed in the period $[t, u_L]$, as we initially supposed, all N_j tests in
38
39 $\{s_1, \dots, s_{N_j}\}$ will be conducted. However, fewer tests will be performed if the patient did progress
40
41 at some point $T_j^* < u_L$. We formally define the discrete random variable \mathcal{N}_j denoting the number
42
43
44
45
46
47

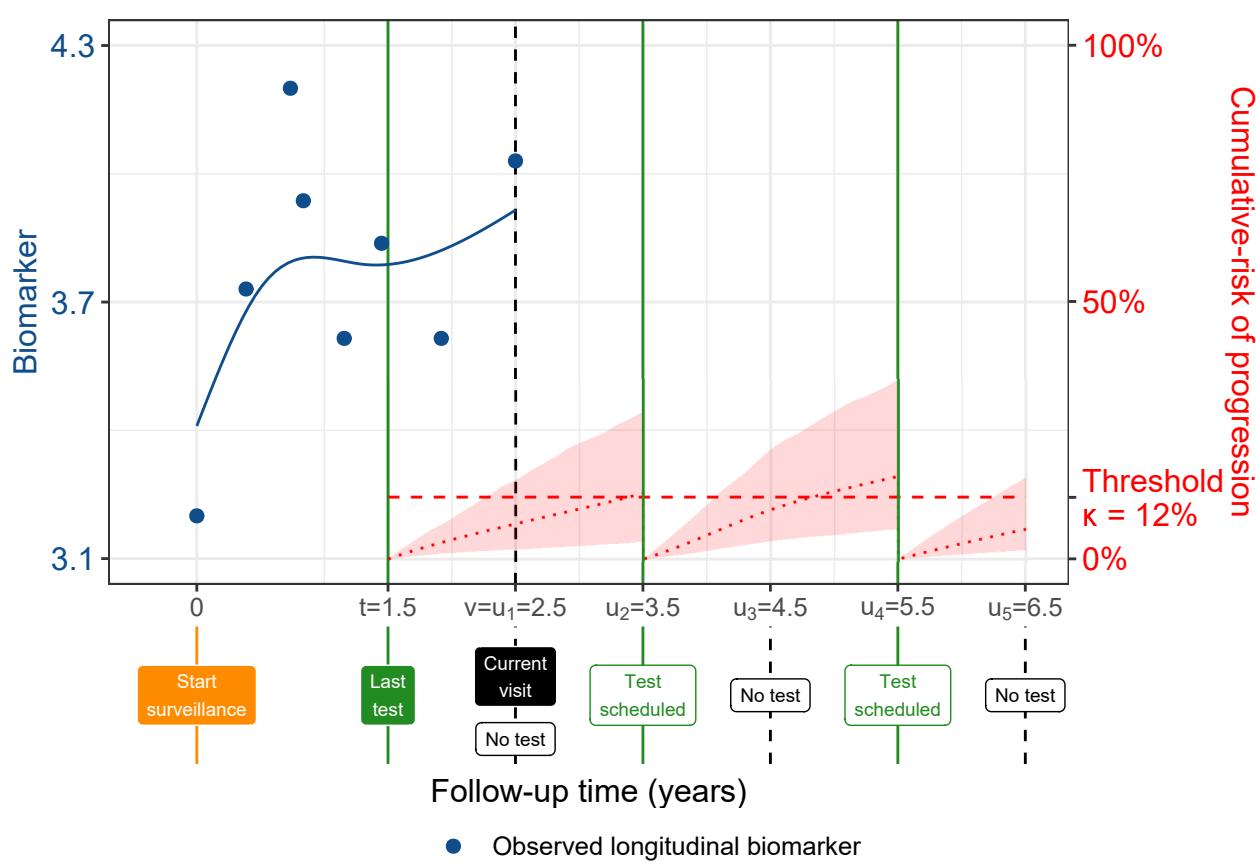


Figure 3: Successive personalized test decisions based on patient-specific cumulative-risk of progression (2). Time of current visit: $v = 2.5$ years (dashed vertical black line). Time of the last test on which progression was not observed: $t = 1.5$ years. Longitudinal data up to current visit: $\mathcal{Y}_j(v)$ is a continuous biomarker (blue circles). Example risk threshold: $\kappa = 0.12$ (12%). Grid of future visits on which future tests are planned: $U = \{2.5, 3.5, 4.5, 5.5, 6.5\}$ years. The cumulative-risk profiles $R_j(u_l | t_l, v)$ employed in (2) are shown with dotted red lines (95% credible intervals shaded), and are updated each time a test is planned (solid vertical green lines). Future test decisions $Q_j(u_l | t_l, v)$ defined in (2) are: $Q_j^\kappa(u_1 = 2.5 | t_1 = 1.5, v) = 0$, $Q_j^\kappa(u_2 = 3.5 | t_2 = 1.5, v) = 1$, $Q_j^\kappa(u_3 = 4.5 | t_2 = 3.5, v) = 0$, $Q_j^\kappa(u_4 = 5.5 | t_2 = 3.5, v) = 1$, and $Q_j^\kappa(u_5 = 6.5 | t_5 = 4.5, v) = 0$. All values are illustrative.

of performed tests in conjunction with the true progression time T_j^* as,

$$\mathcal{N}_j(S_j^\kappa) = \begin{cases} 1, & \text{if } t < T_j^* \leq s_1, \\ 2, & \text{if } s_1 < T_j^* \leq s_2, \\ \vdots & \\ N_j, & \text{if } s_{N_j-1} < T_j^* \leq s_{N_j}, \end{cases}$$

where $S_j^\kappa = \{s_1, \dots, s_{N_j}\}$ is the schedule of planned future tests. The expected number of future tests for patient j will be the expected value $E\{\mathcal{N}_j(S_j^\kappa)\}$, given by the expression,

$$E\{\mathcal{N}_j(S_j^\kappa)\} = \sum_{n=1}^{N_j} n \times \Pr(s_{n-1} < T_j^* \leq s_n | T_j^* \leq s_{N_j}), \quad s_0 = t,$$

where

$$\Pr(s_{n-1} < T_j^* \leq s_n | T_j^* \leq s_{N_j}) = \frac{R_j(s_n | t, v) - R_j(s_{n-1} | t, v)}{R_j(s_{N_j} | t, v)}.$$

Similarly, we can define the expected time delay in detecting progression, under the assumption that progression occurs before u_L . Specifically, the random variable time delay is equal to the difference between the time of the test at which progression is observed and the true time of progression T_j^* , and is given by,

$$\mathcal{D}_j(S_j^\kappa) = \begin{cases} s_1 - T_j^*, & \text{if } t < T_j^* \leq s_1, \\ s_2 - T_j^*, & \text{if } s_1 < T_j^* \leq s_2, \\ \vdots & \\ s_{N_j} - T_j^*, & \text{if } s_{N_j-1} < T_j^* \leq s_{N_j}, \end{cases}$$

The expected time delay in detecting progression is the expected value of $\mathcal{D}_j(S_j^\kappa)$, given by the expression,

$$E\{\mathcal{D}_j(S_j^\kappa)\} = \sum_{n=1}^{N_j} \left\{ s_n - E(T_j^* | s_{n-1}, s_n, v) \right\} \times \Pr(s_{n-1} < T_j^* \leq s_n | T_j^* \leq s_N),$$

1
2
3 where $E(T_j^* | s_{n-1}, s_n, v)$ denotes the conditional expected time of progression for the scenario
4 $s_{n-1} < T_j^* \leq s_n$ and is calculated as the area under the corresponding survival curve,
5
6

$$E(T_j^* | s_{n-1}, s_n, v) = s_{n-1} + \int_{s_{n-1}}^{s_n} \Pr\{T_j^* \geq u | s_{n-1} < T_j^* \leq s_n, \mathcal{Y}_{1j}(v), \dots, \mathcal{Y}_{Kj}(v), \mathcal{A}_n\} du.$$

7 The personalized schedule in (3), and the corresponding personalized expected number of tests
8 and time delay, have the advantage of getting updated with newly collected data over follow-up.
9 Also, the expected number of tests and time delay can be calculated for any schedule, fixed or
10 personalized. Hence, patients/doctors can use them to compare different schedules. Although,
11 a fair comparison of time delays between different schedules for the same patient, requires a
12 compulsory test at a common horizon time point in all schedules.
13
14

21 3.4 How to Select the Risk Threshold κ

22 The risk threshold κ controls the timing and the total number of invasive tests in the personalized
23 schedule S_j^κ . Through the timing and the total number of planned tests, κ also indirectly affects
24 the potential time delay (Figure 1) in detecting progression if a particular schedule is followed.
25 Hence, κ should be chosen while balancing both the number of invasive tests (burden) and the
26 time delay in detecting progression (shorter is beneficial).
27
28

29 To facilitate the choice of κ in practice, following our developments in the previous section, we
30 translate the different choices for threshold κ into the expected number of tests and time delay.
31 In particular, for a patient j having data available up to his current visit time v , we can construct
32 a bi-dimensional Euclidean space of his expected total number of tests (x-axis) and expected
33 time delay in detecting progression (y-axis), for different personalized test schedules obtained by
34 varying $\kappa \in [0, 1]$, e.g., Figure 4.
35
36

37 The ideal schedule for j -th patient is the one in which only one test is conducted, at exactly the
38 true time of progression T_j^* . In other words, the time delay will be zero. If we weigh the expected
39 number of tests and time delay as equally important, then we can select as the optimal threshold
40

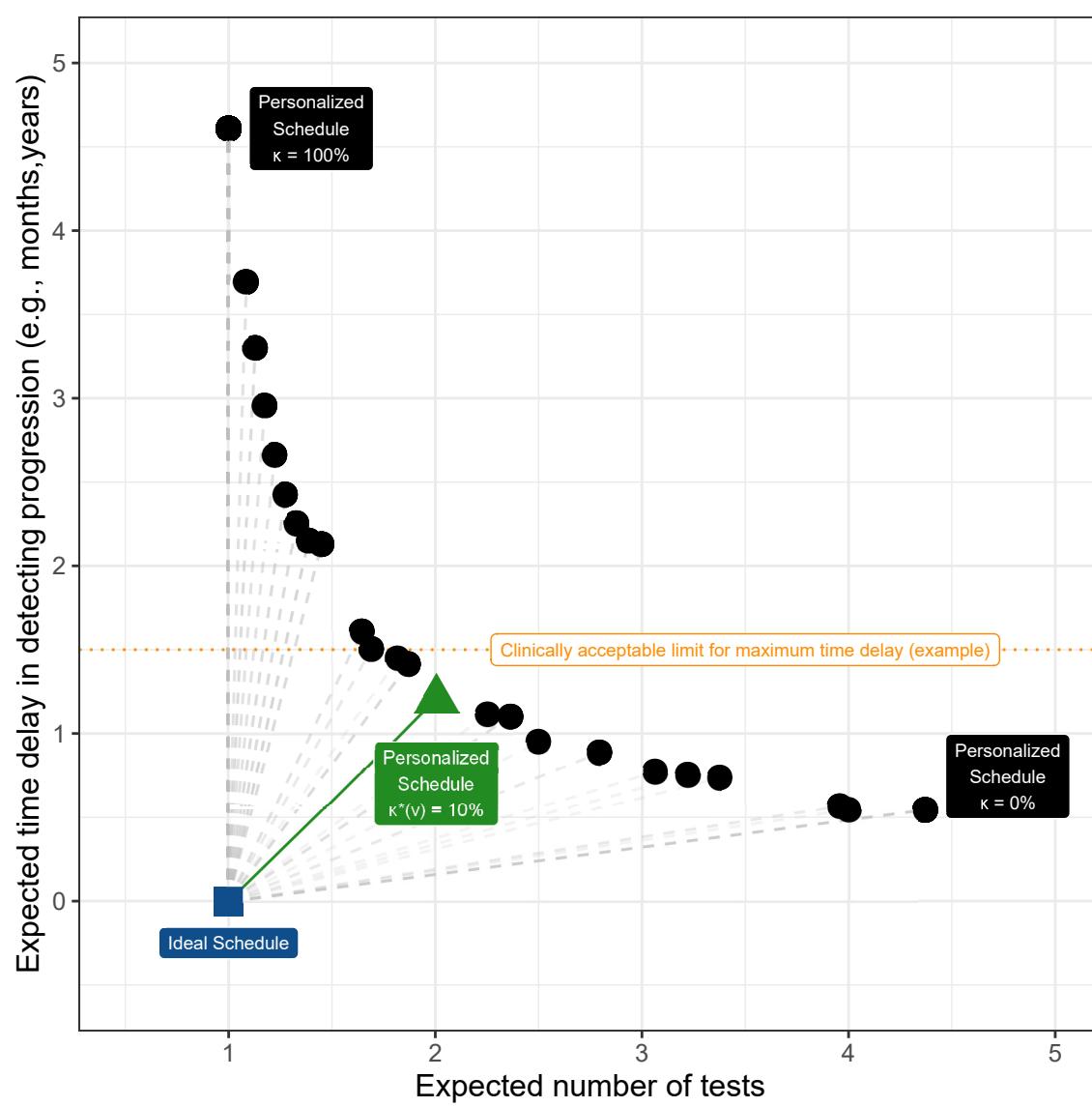


Figure 4: Optimal current-visit time v specific risk threshold $\kappa^*(v)$ obtained using (4) for the patient shown in Figure 3. Ideal schedule of tests: point (1,0) shown as a blue square. It plans exactly one invasive test at the true time of progression T_j^* of a patient. Hence, the time delay in detecting progression is zero. Various personalized schedules based on a grid of thresholds $\kappa \in [0, 1]$ are shown with black circles. Higher thresholds lead to fewer tests, but also higher expected time delay. The personalized schedule based on $\kappa^*(v) = 9.5\%$ threshold (green triangle) has the least Euclidean distance (solid green line) to the ideal schedule. It is also possible to optimize the least distance under a certain clinically acceptable limit on the time delay (dotted horizontal orange line).

at current visit time v , the threshold $\kappa^*(v)$ which minimizes the Euclidean distance between the ideal schedule, i.e., point $(1, 0)$ and the set of points representing the different personalized schedules S_j^κ corresponding to various $\kappa \in [0, 1]$, i.e.,

$$\kappa^*(v) = \arg \min_{0 \leq \kappa \leq 1} \sqrt{\left[E\{\mathcal{N}_j(S_j^\kappa)\} - 1 \right]^2 + \left[E\{\mathcal{D}_j(S_j^\kappa)\} - 0 \right]^2}. \quad (4)$$

In certain scenarios, patients/doctors may be apprehensive about undergoing more than a maximum expected number of future tests, or having an expected time delay higher than certain months. For such purposes, the Euclidean distance in (4) can be optimized under constraints on the expected number of tests or expected time delay (Figure 4). Doing so alleviates two problems, namely, that the time delay and the number of tests have different units of measurement, and that in (4) they are weighted equally [26].

We considered shorter delays in detecting progression as the benefit of repeated tests. However, in the literature, decision-theoretic measures such as quality-adjusted life-years/expectancy (QALY/QALE) gained [27] have also been used to quantify the benefit of testing. Optimizing (4) with QALE needs, setting the optimal point in a Euclidean space with QALE as a dimension, and obtaining expected QALEs for different schedules. For estimating the expected QALE in a personalized manner, a mathematical definition of QALE in terms of time delay \mathcal{D}_j in detecting progression [28] is required.

4 Application of Personalized Schedules in Prostate Cancer Surveillance

We next demonstrate personalized schedules for scheduling biopsies in prostate cancer active surveillance. To this end, we use results from a joint model fitted to the PRIAS dataset introduced in Section 1. The model definition (Supplementary B) utilized a linear mixed sub-model for

1
2
3 biannually measured PSA (continuous: log-transformed from ng/mL), and a logistic mixed sub-
4 model for biannually measured DRE (binary: tumor palpable or not). In the survival sub-model,
5 fitted PSA value, fitted instantaneous PSA velocity (defined in Section 2.2), and log-odds of having
6 a DRE indicating a palpable tumor, were included as time-dependent predictors. The model
7 parameters were estimated under the Bayesian framework using the R package **JMbayes** [29], and
8 are presented in Supplementary B. We next briefly present the key results relevant for personalized
9 scheduling.
10
11

12 First, the cause-specific cumulative-risk of cancer progression at the maximum study period of
13 ten years was 50% (Supplementary Figure 1). This indicates that many patients may not require
14 all of the yearly biopsies they are usually prescribed. Since personalized schedules are risk-based,
15 their overall performance is dependent on the predictive accuracy and discrimination capacity
16 of the fitted model. In this regard, the model had a moderate time-dependent area under the
17 receiver operating characteristic curve or AUC [30] over the follow-up period (between 0.61 and
18 0.68). The time-dependent mean absolute prediction error or MAPE [30] was moderate to large
19 (between 0.08 and 0.24) and decreased rapidly after year one of the follow-up. Thus, personalized
20 schedules based on this model may work better after year one with more follow-up data. Details
21 on AUC and MAPE are provided in Supplementary B.
22
23

33 4.1 Personalized Biopsy Schedules for a Demonstration Patient

34

35 We utilized the joint model fitted to the PRIAS dataset to schedule biopsies in a demonstration
36 prostate cancer patient shown in Figure 5. The time of his last negative biopsy was $t = 3.5$ years,
37 and the time of the current visit was $v = 5$ years. We made biopsy decisions over his future visits
38 for PSA measurement $U = \{u_1 = 5, u_2 = 5.5, \dots, u_L = 10\}$ years using four different schedules.
39 Two of the fixed schedules are annual biopsy schedule and the PRIAS schedule. The PRIAS
40 schedule has compulsory biopsies at year one, four, seven, and ten of follow-up, and additional
41

1
2
3 annual biopsies if PSA doubling-time [2] is high. Remaining two schedules are personalized,
4 namely, with a fixed threshold $\kappa = 10\%$ risk, and an automatically chosen current visit time v
5 specific risk $\kappa^*(v)$ (Section 3.4). Since the demonstration patient's time of last negative biopsy
6 $t = 3.5$ is after year one of follow-up, a time delay in detecting progression up to three years may
7 not lead to adverse downstream outcomes [31].
8
9

10 The cumulative-risk of progression of the demonstration patient increases 3% yearly on average,
11 up to 19% at the maximum study period of ten years. Hence, the patient may progress slowly.
12 Consequently, risk-based personalized approaches plan fewer biopsies than the annual schedule
13 (Panel B, Figure 5). Also, the time delay in detecting progression for personalized schedules
14 (Panel D, Figure 5) is below the safe limit of three years mentioned earlier. Thus, personalized
15 schedules can be a suitable alternative to the annual schedule.
16
17
18
19
20
21
22
23

24 5 Simulation Study 25

26 Although we evaluated personalized schedules for a demonstration patient, we also intend to
27 analyze and compare personalized and fixed schedules in a full cohort. Our criteria for comparison
28 of schedules are the total number of invasive tests planned (burden), and the actual time delay
29 in detecting progression (shorter is beneficial) for each schedule. Due to the periodical nature
30 of schedules, the actual time delay in detecting progression cannot be observed in real-world
31 surveillance. Hence, instead, we compare personalized versus fixed schedules via an extensive
32 simulated randomized clinical trial in which each hypothetical patient undergoes each schedule.
33 To keep our simulation study realistic, we employ the prostate cancer active surveillance scenario.
34 Specifically, our simulated population is generated using the joint model fitted to the PRIAS
35 cohort (Supplementary B).
36
37
38
39
40
41
42
43
44
45
46
47
48

49
50
51
52
53
54
55
56
57
58
59
60

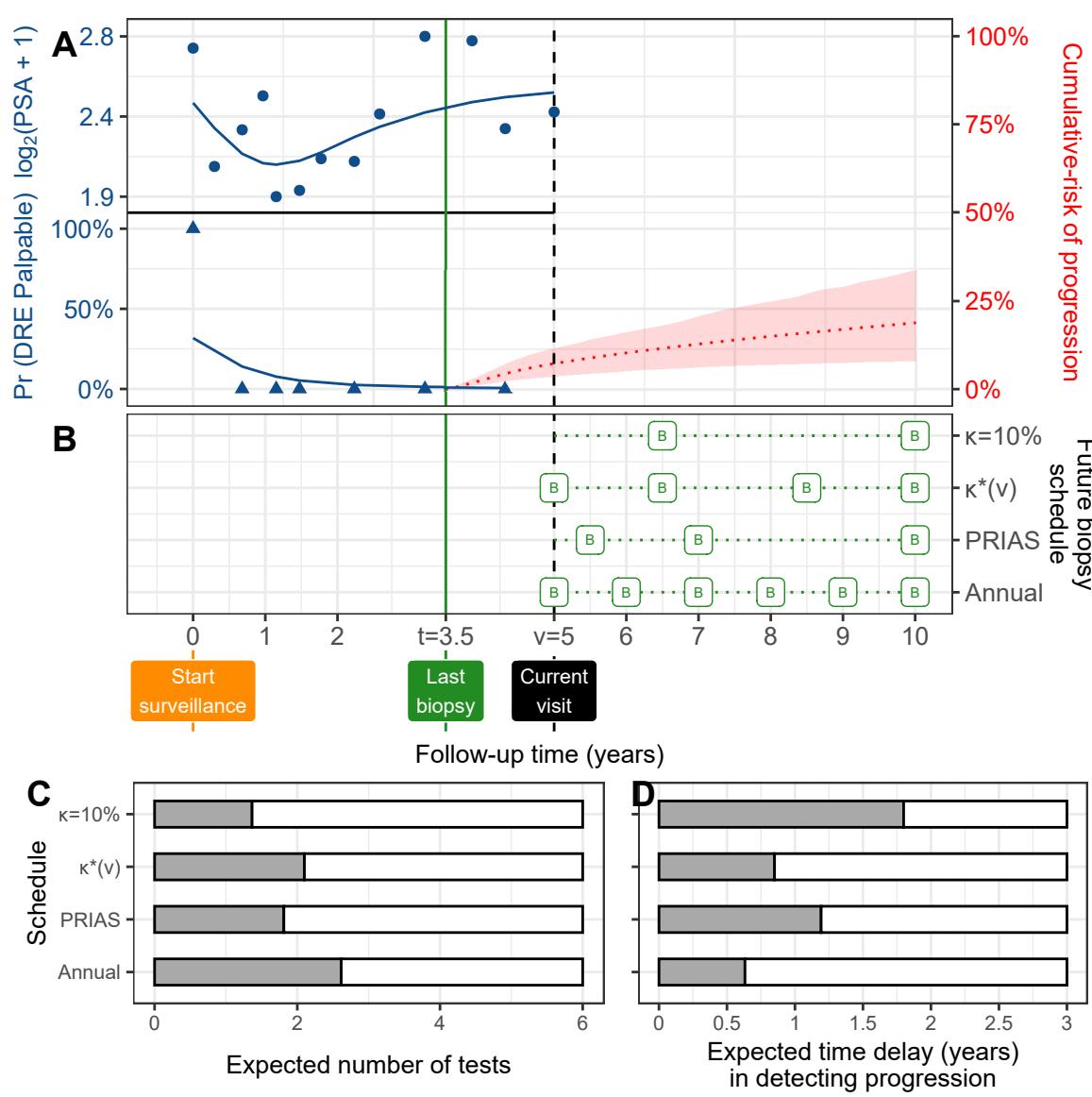


Figure 5: Personalized schedules for a demonstration prostate cancer patient.

Panel A: Time of current visit: $v = 5$ years (black dashed line). Time of last negative biopsy: $t = 3.5$ years (vertical green solid line). Longitudinal data: $\log_2(\text{PSA} + 1)$ transformed PSA (observed: blue dots, fitted: solid blue line), and binary DRE (observed: blue triangles, fitted probability: solid blue line). Cumulative-risk profile: dotted red line (95% credible interval shaded). **Panel B:** Biopsy indicated with a 'B', and $\kappa = 10\%$ and $\kappa^*(v)$ are personalized biopsy schedules using a risk threshold of 10%, and a visit time v specific automatically chosen threshold (4), respectively. PRIAS and Annual denote the PRIAS biopsy schedule (Section 4.1) and annual biopsy schedule. **Panel C,D:** For all schedules we calculate the expected number of tests and expected time delay in detecting progression if the patient progresses before year ten. Since a recommended minimum gap of one year is maintained between biopsies, maximum possible number of tests are six. A delay in detecting progression of up to three years may not lead to adverse outcomes [31].

5.1 Simulation Setup

From the simulation population, we first sample 500 datasets, each representing a hypothetical prostate cancer surveillance program with 1000 patients in it. We sample longitudinal DRE and PSA measurements biannually (PRIAS protocol) for each of the 500×1000 patients and then generate a true cancer progression time for them. We split each dataset into training (750 patients) and test (250 patients) parts, and generate a random and non-informative censoring time for the training patients. All training and test 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 (Supplementary B), to each of the 500 training datasets and retrieve 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 obtain the cumulative-risk of progression in each of the 500×250 test patients. These cumulative-risk profiles are further used to create personalized biopsy schedules for the test patients.

For each test patient, we conduct hypothetical biopsies using two fixed (PRIAS and annual schedule) and three personalized biopsy schedules. Personalized schedules are based on, a fixed risk threshold $\kappa = 10\%$, an optimal current visit time v specific threshold $\kappa^*(v)$ chosen via (4), and an optimal threshold obtained under the constraint that expected time delay in detecting progression is less than 0.75 years (9 months), denoted $\kappa^*\{v \mid E(\mathcal{D}) \leq 0.75\}$. The choice of 0.75 years delay constraint is arbitrary and is only used to illustrate that applying the constraint limits the average delay at 0.75 years. Successive personalized biopsy decisions are made only on the standard PSA follow-up visits, utilizing clinical data accumulated only until the corresponding current visit time (2). We maintain a minimum recommended gap of one year between consecutive prostate biopsies [2] as well. Biopsies are conducted until progression is detected, or the maximum follow-up period at year ten (horizon) is reached. The actual time delay in detecting progression is equal to the difference in time at which progression is detected and the actual (simulated) time

of progression of a patient.

5.2 Simulation Results

In the simulation study, nearly 50% of the patients observed progression during the ten year study period (*progressing*) and 50% did not (*non-progressing*). While we can calculate the total number of biopsies scheduled in all 500×250 test patients, the actual time delay in detecting progression is available only for progressing patients. Hence, we show the simulation results separately for progressing and non-progressing patients (Figure 6).

Before discussing delay in detecting progression (Panel A, Figure 6), we note that mean delay up to 1.7 years in all patients [32], and up to three years in patients who progress after year one of follow-up [31], may not increase risks of adverse outcomes later. In this regard, the annual biopsies guarantee a maximum delay of one year in all patients. However, they also schedule the highest number of biopsies (Median 3, Inter-quartile range or IQR: 1–6). Much fewer biopsies are planned by the PRIAS schedule (Median 2, IQR: 1–4), but it also has a higher time delay (Median 0.74, IQR: 0.38–1.00 years). The personalized schedule based on optimal risk threshold $\kappa^*(v)$ schedules fewer biopsies than PRIAS and has a delay (Median 0.86, IQR: 0.46–1.26 years) slightly higher than PRIAS. The expected delay for risk threshold optimized with a constraint on expected delay $\kappa^*\{v | E(D) \leq 0.75\}$ is equal to 0.61 years, i.e., the constraint works as expected.

The simulated non-progressing patients (Panel B, Figure 6) gained the most with personalized schedules. The annual schedule plans 10 (unnecessary) biopsies for each such patient, and the PRIAS schedule plans a median of 6 (IQR: 4–8) biopsies. In contrast, the personalized schedule based on optimized risk threshold $\kappa^*(v)$ plans fewer biopsies consistently (Median 6, IQR: 6–7). The 10% threshold based schedule plans even fewer biopsies (Median 5, IQR: 4–6).

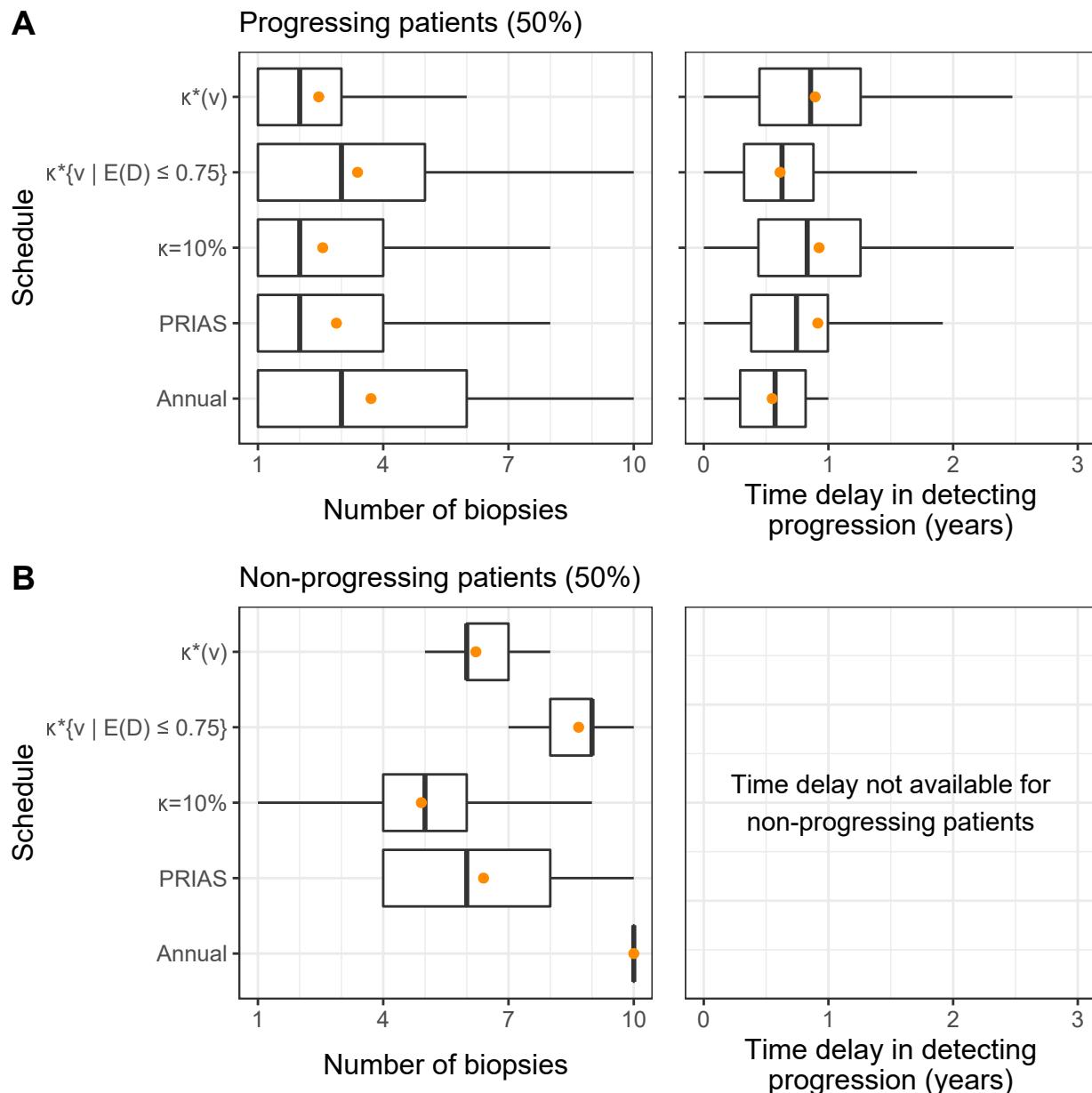


Figure 6: Number of biopsies and the time delay in detecting cancer progression for various biopsy schedules obtained via a simulation study. **Mean** is indicated by the orange circle. Time delay (years) is calculated as (time of positive biopsy - the actual simulated time of cancer progression). Biopsies are conducted²⁴ until cancer progression is detected. **Panel A:** simulated patients who obtained cancer progression in the ten year study period (progressing). **Panel B:** simulated patients who did not obtain cancer progression in the ten year study period (non-progressing). Types of schedules: $\kappa = 10\%$ and $\kappa^*(v)$ schedule a biopsy if the cumulative-risk of cancer progression at the current visit time v is more than 10%, and an automatically chosen threshold (4), respectively. Schedule $\kappa^*\{v | E(D) \leq 0.75\}$ is similar to $\kappa^*(v)$ except that the euclidean distance in (4) is minimized under the constraint that expected delay in detecting progression is at most 9 months (0.75 years). Annual corresponds to a schedule of yearly biopsies, and PRIAS corresponds to biopsies as per PRIAS protocol (Section 4).

6 Discussion

In this paper, we presented a methodology to create personalized schedules for burdensome diagnostic *tests* used to detect disease *progression* in early-stage chronic non-communicable disease *surveillance*. For this purpose, we utilized joint models for time-to-event and longitudinal data. Our approach first combines a patient's clinical data (e.g., longitudinal biomarkers) and previous invasive test results to estimate patient-specific cumulative-risk of disease progression over their current and future follow-up visits. We then plan future invasive tests whenever this cumulative-risk of progression is predicted to be above a certain threshold. We select the 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 (burden) planned in a schedule, and the expected time delay in detection of progression (shorter is beneficial) if the patient progresses. 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 and shared decision of choosing a test schedule.

Using joint models gives us certain advantages. First, since joint models employ random-effects, the corresponding risk-based schedules are inherently personalized. Second, to predict this patient-specific risk of progression, joint models utilize all observed longitudinal measurements of a patient. Also, the continuous longitudinal outcomes are not discretized, which is commonly a case in Markov Decision Process and flowchart-based test schedules. Third, personalized schedules update automatically with more patient data over follow-up. Fourth, we calculated the expected number of tests (burden) and expected time delay in detecting progression (shorter is beneficial) in a patient-specific manner. Using our methodology, these can be calculated for both personalized and fixed schedules. Thus, patients/doctors can compare risk-based and fixed schedules and make a shared decision of a test schedule according to their preferences for the expected burden-benefit ratio. Last, although this work concerns invasive test schedules in disease surveillance,

1
2
3 the methodology is generic for use under a screening setting as well.
4

5 Personalized schedules that we proposed require a risk threshold. We optimized the threshold
6 choice using a generic utility function based on the expected number of biopsies and time delay
7 in detecting progression. We used only these two measures because they are easy to interpret
8 but simultaneously critical for deciding the timing of invasive tests. Also, the time delay in
9 detecting progression is an easily-quantifiable surrogate for the window of opportunity for curative
10 treatment and additional benefits of observing progression early. Practitioners may extend/modify
11 our utility function by adding to/replacing time delay with commonly used decision-theoretic
12 measures such as quality-adjusted life-years/expectancy (QALY/QALE).
13
14

15 We evaluated personalized schedules in a full cohort via a realistic simulation of a randomized
16 clinical trial for prostate cancer surveillance patients. We observed that personalized schedules
17 reduced many unnecessary biopsies for non-progressing patients compared to the widely used
18 annual schedule. This happened at the cost of simultaneously having a slightly longer time
19 delay in detecting progression. Although, this delay should still be safe because it was almost
20 equal to the delay of the world's largest prostate cancer active surveillance program PRIAS's
21 schedule. The simulation study results are by no means the performance-limit of the personalized
22 schedules. Instead, models with higher predictive accuracy and discrimination capacity than the
23 PRIAS based model may lead to an even better balance between the number of tests and the
24 time delay in detecting progression. As for the practical usability of the PRIAS based model in
25 prostate cancer surveillance, despite the moderate predictive performance, we expect this model's
26 overall impact to be positive. There are two reasons for this. First, the risk of adverse outcomes
27 because of personalized schedules is quite low because of the low rate of metastases and prostate
28 cancer specific mortality in prostate cancer patients [2]. Second, studies [31, 32] have suggested
29 that after the confirmatory biopsy at year one of follow-up, biopsies may be done as infrequently
30 as every two to three years, with limited adverse consequences. In other words, longer delays in
31 detecting progression may be acceptable after the first negative biopsy.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 There are certain limitations to this work. First, in practice, most cohorts have a limited
4 study period. Hence, the cumulative-risk profiles of patients and resulting personalized schedules
5 can only be created up to the maximum study period. For this problem, the risk prediction model
6 should be updated with more follow-up data over time. The proposed joint model assumed all
7 events other than progression to be non-informative censoring. Alternative models that account
8 for competing risks may lead to better results as they estimate absolute and not the cause-specific
9 risk of progression. The detection of progression is susceptible to inter-observer variation, e.g.,
10 pathologists may grade the same biopsy differently. Progression is sometimes obscured due to
11 sampling error, e.g., biopsy results vary based on location and number of biopsy cores. Although
12 models that account for inter-observer variation [33] and sampling error [34] will provide better
13 risk estimates, the methodology for obtained personalized schedules can remain the same.
14
15
16
17
18
19
20
21
22
23

Acknowledgments

24
25
26 The first and last authors would like to acknowledge support by Nederlandse Organisatie voor
27 Wetenschappelijk Onderzoek (the national research council of the Netherlands) VIDI grant nr.
28 016.146.301, and Erasmus University Medical Center funding. Part of this work was carried out
29 on the Dutch national e-infrastructure with the support of SURF Cooperative. The authors
30 also thank the Erasmus University Medical Center's Cancer Computational Biology Center for
31 giving access to their IT-infrastructure and software that was used for the computations and data
32 analysis in this study. Last, we would like to thank the PRIAS consortium for enabling this
33 research project.
34
35
36
37
38
39
40
41
42
43
44
45
46
47

1 2 3 Data Availability 4 5

6 This simulation study utilized results from a statistical model fitted to the PRIAS dataset. The
7 PRIAS database is not openly accessible. However, access to the database can be requested on
8 the basis of a study proposal approved by the PRIAS steering committee. The website of the
9 PRIAS program is www.prias-project.org. Instructions for generating a synthetic dataset are
10 provided in the README file along with the source code.
11
12

13 14 15 Supplementary 16

17
18 Supplementary sections referenced in this paper are available in the file titled ‘supplementary.pdf’.
19 Source code is available at <https://github.com/anirudhtomer/PersonalizedSchedules> and is
20 also uploaded in a zip file along with the manuscript.
21
22

23 24 References 25 26

- 27
28 [1] World Health Organization WHO et al. *Global Status Report on Noncommunicable Diseases*
29 2014. Number WHO/NMH/NVI/15.1. World Health Organization, 2014.
30
31 [2] Leonard P Bokhorst, Arnout R Alberts, Antti Rannikko, Riccardo Valdagni, Tom Pickles,
32 Yoshiyuki Kakehi, Chris H Bangma, Monique J Roobol, and PRIAS study group. Compliance
33 rates with the Prostate Cancer Research International Active Surveillance (PRIAS) protocol
34 and disease reclassification in noncompliers. *European Urology*, 68(5):814–821, 2015.
35
36 [3] Bas Weusten, Raf Bisschops, Emanuel Coron, Mário Dinis-Ribeiro, Jean-Marc Dumonceau,
37 José-Miguel Esteban, Cesare Hassan, Oliver Pech, Alessandro Repici, Jacques Bergman,
38 et al. Endoscopic management of Barrett’s esophagus: European society of gastrointestinal
39 endoscopy (ESGE) position statement. *Endoscopy*, 49(02):191–198, 2017.
40
41

- [4] Alex H Krist, Resa M Jones, Steven H Woolf, Sarah E Woessner, Daniel Merenstein, J William Kerns, Walter Foliaco, and Paul Jackson. Timing of repeat colonoscopy: disparity between guidelines and endoscopists' recommendation. *American Journal of Preventive Medicine*, 33(6):471–478, 2007.
- [5] Tanya J McWilliams, Trevor J Williams, Helen M Whitford, and Gregory I Snell. Surveillance bronchoscopy in lung transplant recipients: risk versus benefit. *The Journal of Heart and Lung Transplantation*, 27(11):1203–1209, 2008.
- [6] Stacy Loeb, Annelies Vellekoop, Hashim U Ahmed, James Catto, Mark Emberton, Robert Nam, Derek J Rosario, Vincenzo Scattoni, and Yair Lotan. Systematic review of complications of prostate biopsy. *European Urology*, 64(6):876–892, 2013.
- [7] C.M.C. Le Clercq, B. Winkens, C.M. Bakker, E.T.P. Keulen, G.L. Beets, A.A.M. Masclinee, and S. Sanduleanu. Metachronous colorectal cancers result from missed lesions and non-compliance with surveillance. *Gastrointestinal Endoscopy*, 82(2):325–333.e2, 2015.
- [8] Oguzhan Alagoz, Turgay Ayer, and Fatih Safa Erenay. *Operations Research Models for Cancer Screening*. American Cancer Society, 2011.
- [9] Lauren N Steimle and Brian T Denton. *Markov Decision Processes for Screening and Treatment of Chronic Diseases*. Springer International Publishing, 2017.
- [10] Andrew J Vickers and Elena B Elkin. Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making*, 26(6):565–574, 2006.
- [11] Anirudh Tomer, Daan Nieboer, Monique J Roobol, Ewout W Steyerberg, and Dimitris Rizopoulos. Personalized schedules for surveillance of low-risk prostate cancer patients. *Biometrics*, 75(1):153–162, 2019.

- [12] Ionut Bebu and John M Lachin. Optimal screening schedules for disease progression with application to diabetic retinopathy. *Biostatistics*, 19(1):1–13, 2017.
- [13] Dimitris Rizopoulos, Jeremy MG Taylor, Joost Van Rosmalen, Ewout W Steyerberg, and Johanna JM Takkenberg. Personalized screening intervals for biomarkers using joint models for longitudinal and survival data. *Biostatistics*, 17(1):149–164, 2015.
- [14] Anirudh Tomer, Dimitris Rizopoulos, Daan Nieboer, Frank-Jan Drost, Monique J Roobol, and Ewout W Steyerberg. Personalized decision making for biopsies in prostate cancer active surveillance programs. *Medical Decision Making*, 39(5):499–508, 2019.
- [15] Yanqing Wang, Ying-Qi Zhao, and Yingye Zheng. Learning-based biomarker-assisted rules for optimized clinical benefit under a risk-constraint. *Biometrics*, 2019.
- [16] Anastasios A Tsiatis and Marie Davidian. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 14(3):809–834, 2004.
- [17] Dimitris Rizopoulos. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. CRC Press, 2012.
- [18] Charles E McCulloch and John M Neuhaus. Generalized linear mixed models. *Encyclopedia of Biostatistics*, 4, 2005.
- [19] Stacy Loeb, Marc A Bjurlin, Joseph Nicholson, Teuvo L Tammela, David F Penson, H Ballantine Carter, Peter Carroll, and Ruth Etzioni. Overdiagnosis and overtreatment of prostate cancer. *European Urology*, 65(6):1046–1055, 2014.
- [20] Jonathan I Epstein, Lars Egevad, Mahul B Amin, Brett Delahunt, John R Srigley, and Peter A Humphrey. The 2014 international society of urological pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma. *The American Journal of Surgical Pathology*, 40(2):244–252, 2016.

- 1
2
3 [21] Stacy Loeb, H Ballentine Carter, Mark Schwartz, Angela Fagerlin, R Scott Braithwaite, and
4 Herbert Lepor. Heterogeneity in active surveillance protocols worldwide. *Reviews in Urology*,
5 16(4):202–203, 2014.
6
7 [22] Lindsey A Torre, Freddie Bray, Rebecca L Siegel, Jacques Ferlay, Joannie Lortet-Tieulent,
8 and Ahmedin Jemal. Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians*,
9 65(2):87–108, 2015.
10
11 [23] Elizabeth R. Brown. Assessing the association between trends in a biomarker and risk of event
12 with an application in pediatric HIV/AIDS. *The Annals of Applied Statistics*, 3(3):1163–1182,
13 2009.
14
15 [24] Jeremy M.G. Taylor, Yongseok Park, Donna P. Ankerst, Cecile Proust-Lima, Scott Williams,
16 Larry Kestin, Kyoungwha Bae, Tom Pickles, and Howard Sandler. Real-time individual
17 predictions of prostate cancer recurrence using joint models. *Biometrics*, 69(1):206–213,
18 2013.
19
20 [25] Paul HC Eilers and Brian D Marx. Flexible smoothing with B-splines and penalties. *Statistical
21 Science*, 11(2):89–121, 1996.
22
23 [26] R. Dennis Cook and Weng K. Wong. On the equivalence of constrained and compound
24 optimal designs. *Journal of the American Statistical Association*, 89(426):687–692, 1994.
25
26 [27] Franco Sassi. Calculating QALYs, comparing QALY and DALY calculations. *Health Policy
27 and Planning*, 21(5):402–408, 2006.
28
29 [28] Tiago M de Carvalho, Eveline AM Heijnsdijk, and Harry J de Koning. When should active
30 surveillance for prostate cancer stop if no progression is detected? *The Prostate*, 77(9):962–
31 969, 2017.

- [29] Dimitris Rizopoulos. The R package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC. *Journal of Statistical Software*, 72(7):1–46, 2016.
- [30] Dimitris Rizopoulos, Geert Molenberghs, and Emmanuel MEH Lesaffre. Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biometrical Journal*, 59(6):1261–1276, 2017.
- [31] Tiago M de Carvalho, Eveline AM Heijnsdijk, and Harry J de Koning. Estimating the risks and benefits of active surveillance protocols for prostate cancer: a microsimulation study. *BJU International*, 119(4):560–566, 2017.
- [32] Lurdes YT Inoue, Daniel W Lin, Lisa F Newcomb, Amy S Leonardson, Donna Ankerst, Roman Gulati, H Ballentine Carter, Bruce J Trock, Peter R Carroll, Matthew R Cooperberg, et al. Comparative analysis of biopsy upgrading in four prostate cancer active surveillance cohorts. *Annals of Internal Medicine*, 168(1):1–9, 2018.
- [33] Raji Balasubramanian and Stephen W Lagakos. Estimation of a failure time distribution based on imperfect diagnostic tests. *Biometrika*, 90(1):171–182, 2003.
- [34] R Yates Coley, Scott L Zeger, Mufaddal Mamawala, Kenneth J Pienta, and H Ballentine Carter. Prediction of the pathologic Gleason score to inform a personalized management program for prostate cancer. *European Urology*, 72(1):135–141, 2017.

1
2
3
4
5
6 **Supplementary Materials for ‘Shared**
7 **Decision Making of Burdensome Surveillance**
8 **Tests Using Personalized Schedules and**
9 **Their Burden and Benefit’**

10
11
12
13
14
15 Anirudh Tomer^{*1}, Daan Nieboer², Monique J. Roobol³, Ewout W.
16 Steyerberg^{2,4}, and Dimitris Rizopoulos¹

17
18 ¹Department of Biostatistics, Erasmus University Medical Center, the
19 Netherlands

20
21 ²Department of Public Health, Erasmus University Medical Center, the
22 Netherlands

23
24 ³Department of Urology, Erasmus University Medical Center, the
25 Netherlands

26
27 ⁴Department of Biomedical Data Sciences, Leiden University Medical
28 Center, the Netherlands

29
30 April 6, 2021

31
32

⁴³Corresponding author: a.tomter@erasmusmc.nl

33
34 The authors gratefully acknowledge Nederlandse Organisatie voor Wetenschappelijk Onderzoek VIDI grant
35 nr. 016.146.301, and Erasmus University Medical Center funding.

A Joint Model for Time-to-Progression and Longitudinal Outcomes

Let T_i^* denote the true time of disease progression for the i -th patient. Progression is always interval censored $l_i < T_i^* \leq r_i$. Here, r_i and l_i denote the time of the last and second last invasive tests, respectively, when patients progress. In non-progressing patients, l_i denotes the time of the last test and $r_i = \infty$. Assuming K types of longitudinal outcomes, let \mathbf{y}_{ki} denote the $n_{ki} \times 1$ longitudinal response vector of the k -th outcome, $k \in \{1, \dots, K\}$. The observed data of all n patients is given by $\mathcal{A}_n = \{l_i, r_i, \mathbf{y}_{1i}, \dots, \mathbf{y}_{Ki}; i = 1, \dots, n\}$.

A.1 Longitudinal Sub-process

To model multiple longitudinal outcomes in a unified framework, a joint model employs individual generalized linear mixed sub-models [1]. Specifically, the conditional distribution of the k -th outcome \mathbf{y}_{ki} given a vector of patient-specific random effects \mathbf{b}_{ki} is assumed to belong to the exponential family, with linear predictor given by,

$$g_k[E\{\mathbf{y}_{ki}(t) | \mathbf{b}_{ki}\}] = m_{ki}(t) = \mathbf{x}_{ki}^\top(t)\boldsymbol{\beta}_k + \mathbf{z}_{ki}^\top(t)\mathbf{b}_{ki},$$

where $g_k(\cdot)$ denotes a known one-to-one monotonic link function, $y_{ki}(t)$ is the value of the k -th longitudinal outcome for the i -th patient at time t , and $\mathbf{x}_{ki}(t)$ and $\mathbf{z}_{ki}(t)$ are the time-dependent design vectors for the fixed $\boldsymbol{\beta}_k$ and random effects \mathbf{b}_{ki} , respectively. To model the correlation between different longitudinal outcomes, we link their corresponding random effects. Specifically, we assume that the vector of random effects $\mathbf{b}_i = (\mathbf{b}_{1i}^\top, \dots, \mathbf{b}_{Ki}^\top)^\top$ follows a multivariate normal distribution with mean zero and variance-covariance matrix W .

A.2 Survival Sub-process

In the survival sub-process, the hazard of progression $h_i(t)$ at a time t is assumed to depend on a function of patient and outcome-specific linear predictors $m_{ki}(t)$ and/or the random effects,

$$h_i\{t \mid \mathcal{M}_i(t), \mathbf{w}_i(t)\} = h_0(t) \exp \left[\boldsymbol{\gamma}^\top \mathbf{w}_i(t) + \sum_{k=1}^K f_k\{\mathcal{M}_{ki}(t), \mathbf{w}_i(t), \mathbf{b}_{ki}, \boldsymbol{\alpha}_k\} \right], \quad t > 0,$$

where $h_0(\cdot)$ denotes the baseline hazard, $\mathcal{M}_{ki}(t) = \{m_{ki}(s) \mid 0 \leq s < t\}$ is the history of the k -th longitudinal process up to t , and $\mathbf{w}_i(t)$ is a vector of exogenous, possibly time-varying covariates with regression coefficients $\boldsymbol{\gamma}$. Functions $f_k(\cdot)$, parameterized by vector of coefficients $\boldsymbol{\alpha}_k$, specify the features of each longitudinal outcome that are included in the linear predictor of the relative-risk model [2, 3, 4]. Some examples, motivated by the literature (subscripts k dropped for brevity), are,

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

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 [5].

More specifically:

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

where $B_q(t, \mathbf{v})$ denotes the q -th basis function of a B-spline with knots $\mathbf{v} = v_1, \dots, v_Q$ and vector of spline coefficients γ_{h_0} . To avoid choosing the number and position of knots in the

spline, a relatively high number of knots (e.g., 15 to 20) are chosen and the corresponding B-spline regression coefficients γ_{h_0} are penalized using a differences penalty [5].

A.3 Parameter Estimation

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

$$\begin{aligned} p(\boldsymbol{\theta}, \mathbf{b} | \mathcal{D}_n) &\propto \prod_{i=1}^n p(l_i, r_i, \mathbf{y}_{1i}, \dots, \mathbf{y}_{Ki}, | \mathbf{b}_i, \boldsymbol{\theta}) p(\mathbf{b}_i | \boldsymbol{\theta}) p(\boldsymbol{\theta}) \\ &\propto \prod_{i=1}^n \prod_{k=1}^K p(l_i, r_i | \mathbf{b}_i, \boldsymbol{\theta}) p(\mathbf{y}_{ki} | \mathbf{b}_i, \boldsymbol{\theta}) p(\mathbf{b}_i | \boldsymbol{\theta}) p(\boldsymbol{\theta}), \\ p(\mathbf{b}_i | \boldsymbol{\theta}) &= \frac{1}{\sqrt{(2\pi)^{|W|} \det(\mathbf{D})}} \exp(\mathbf{b}_i^\top \mathbf{D}^{-1} \mathbf{b}_i), \end{aligned}$$

where, the likelihood contribution of the k -th longitudinal outcome vector \mathbf{y}_{ki} for the i -th patient, conditional on the random effects is:

$$p(\mathbf{y}_{ki} | \mathbf{b}_i, \boldsymbol{\theta}) = \prod_{j=1}^{n_{ki}} \exp \left[\frac{y_{kij}\psi_{kij}(\mathbf{b}_{ki}) - c_k\{\psi_{kij}(\mathbf{b}_{ki})\}}{a_k(\varphi)} - d_k(y_{kij}, \varphi) \right],$$

where n_{ki} are the total number of longitudinal measurements of type k for patient i . The natural and dispersion parameters of the exponential family are denoted by $\psi_{kij}(\mathbf{b}_{ki})$ and φ , respectively. In addition, $c_k(\cdot)$, $a_k(\cdot)$, $d_k(\cdot)$ are known functions specifying the member of the exponential family. The likelihood contribution of the time to progression outcome

is given by:

$$p(l_i, r_i | \mathbf{b}_i, \boldsymbol{\theta}) = \exp \left[- \int_0^{l_i} h_i \{s | \mathcal{M}_i(t), \mathbf{w}_i(t)\} ds \right] - \exp \left[- \int_0^{r_i} h_i \{s | \mathcal{M}_i(t), \mathbf{w}_i(t)\} ds \right]. \quad (1)$$

The integral in (1) does not have a closed-form solution, and therefore we use a 15-point Gauss-Kronrod quadrature rule to approximate it.

We use independent normal priors with zero mean and variance 100 for the fixed effect parameters of the longitudinal model. For scale parameters we use inverse Gamma priors. For the variance-covariance matrix \mathbf{D} of the random effects we take inverse Wishart prior with an identity scale matrix and degrees of freedom equal to the total number of random effects. For the relative risk model's parameters $\boldsymbol{\gamma}$ and the association parameters $\boldsymbol{\alpha}$, we use independent normal priors with zero mean and variance 100. However, when $\boldsymbol{\alpha}$ becomes high dimensional (e.g., when several functional forms are considered per longitudinal outcome), we opt for a global-local ridge-type shrinkage prior, i.e., for the s -th element of $\boldsymbol{\alpha}$ we assume:

$$\alpha_s \sim \mathcal{N}(0, \tau\psi_s), \quad \tau^{-1} \sim \text{Gamma}(0.1, 0.1), \quad \psi_s^{-1} \sim \text{Gamma}(1, 0.01). \quad (2)$$

The global smoothing parameter τ has sufficiently mass near zero to ensure shrinkage, while the local smoothing parameter ψ_s allows individual coefficients to attain large values. Other options of shrinkage or variable-selection priors could be used as well [6]. Finally, the penalized version of the B-spline approximation to the baseline hazard is specified using the following hierarchical prior for γ_{h_0} [7]:

$$p(\gamma_{h_0} | \tau_h) \propto \tau_h^{\rho(\mathbf{K})/2} \exp \left(- \frac{\tau_h}{2} \gamma_{h_0}^\top \mathbf{K} \gamma_{h_0} \right) \quad (3)$$

where τ_h is the smoothing parameter that takes a $\text{Gamma}(1, \tau_{h\delta})$ prior distribution, with a hyper-prior $\tau_{h\delta} \sim \text{Gamma}(10^{-3}, 10^{-3})$, which ensures a proper posterior distribution for

1
2
3 $\gamma_{h_0}[8]$, $\mathbf{K} = \Delta_r^\top \Delta_r + 10^{-6} \mathbf{I}$, with Δ_r denoting the r -th difference penalty matrix, and $\rho(\mathbf{K})$
4 denotes the rank of \mathbf{K} .
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

B Joint Model for the PRIAS Dataset Used in Simulation Study

B.1 Dataset

This work uses a joint model fitted to the PRIAS dataset (Table 1). The PRIAS database is not openly accessible. However, access to the database can be requested on the basis of a study proposal approved by the PRIAS steering committee. The website of the PRIAS program is www.priasm.org. We have presented the PRIAS based model’s definition and parameter estimates below.

Table 1: **Summary of the PRIAS dataset.** The primary event of interest is cancer progression (increase in biopsy Gleason grade group from grade group 1 to 2 or higher). Abbreviations: PSA is prostate-specific antigen; DRE is digital rectal examination, with level T1c [9] indicating a clinically inapparent tumor which is not palpable or visible by imaging, whereas tumors with DRE > T1c are palpable; IQR is interquartile range.

Characteristic	Value
Total patients	7813
<i>progression (primary event)</i>	1134
Treatment	2250
Watchful waiting	334
Lost to follow-up	203
Discontinued on request	46
Death (other)	95
Death (prostate cancer)	2
Total DRE measurements	37326
Total PSA measurements	67578
Total biopsies	15686
Median age at diagnosis (years)	66 (IQR: 61–71)
Median PSA (ng/mL)	5.7 (IQR: 4.1–7.7)
DRE = T1c (%)	34883/37326 (94%)
Median number of PSA per patient	6 (IQR: 4–12)
Median number of DRE per patient	4 (IQR: 2–7)
Median number of biopsies per patient	2 (IQR: 1–2)

B.2 Model Specification

Let T_i^* denote the true progression time of the i -th patient included in PRIAS. Since biopsies are conducted periodically, T_i^* is observed with interval censoring $l_i < T_i^* \leq r_i$. When progression is observed for the patient at his latest biopsy time r_i , then l_i denotes the time of the second latest biopsy. Otherwise, l_i denotes the time of the latest biopsy and $r_i = \infty$. Let \mathbf{y}_{di} and \mathbf{y}_{pi} denote his observed DRE (digital rectal examination) and PSA (prostate-specific antigen) longitudinal measurements, respectively. The observed data of all n patients is denoted by $\mathcal{A}_n = \{l_i, r_i, \mathbf{y}_{di}, \mathbf{y}_{pi}; i = 1, \dots, n\}$.

The patient-specific DRE and PSA measurements over time are modeled using a bivariate generalized linear mixed effects sub-model. The sub-model for DRE is given by:

$$\text{logit}[\Pr\{y_{di}(t) > \text{T1c}\}] = \beta_{0d} + b_{0di} + (\beta_{1d} + b_{1di})t + \beta_{2d}(\text{Age}_i - 65) + \beta_{3d}(\text{Age}_i - 65)^2 \quad (4)$$

where, t denotes the follow-up visit time, and Age_i is the age of the i -th patient at the time of inclusion in AS. The fixed effect parameters are denoted by $\{\beta_{0d}, \dots, \beta_{3d}\}$, and $\{b_{0di}, b_{1di}\}$ are the patient specific random effects. With this definition, we assume that the patient-specific log odds of obtaining a DRE measurement larger than T1c [9], i.e., palpable tumor, remain linear over time.

The mixed effects sub-model for PSA is given by:

$$\begin{aligned} \log_2 \{y_{pi}(t) + 1\} &= m_{pi}(t) + \varepsilon_{pi}(t), \\ m_{pi}(t) &= \beta_{0p} + b_{0pi} + \sum_{k=1}^3 (\beta_{kp} + b_{kpi}) B_k(t, \mathcal{K}) \\ &\quad + \beta_{4p}(\text{Age}_i - 65) + \beta_{5p}(\text{Age}_i - 65)^2, \end{aligned} \quad (5)$$

where, $m_{pi}(t)$ denotes the underlying measurement error free value of $\log_2(\text{PSA} + 1)$ transformed [10, 11] measurements at time t . We model it non-linearly over time using B-

splines [12]. To this end, the B-spline basis function $B_k(t, \mathcal{K})$ has two internal knots at $\mathcal{K} = \{0.75, 2.12\}$ years (33.34 and 66.67 percentile of observed follow-up times), and boundary knots at 0 and 6.4 years (95-th percentile of the observed follow-up times). The fixed effect parameters are denoted by $\{\beta_{0p}, \dots, \beta_{5p}\}$, and $\{b_{0pi}, \dots, b_{3pi}\}$ are the patient specific random effects. The error $\varepsilon_{pi}(t)$ is assumed to be t-distributed with three degrees of freedom (Appendix B.4) and scale σ , and is independent of the random effects.

To account for the correlation between the DRE and PSA measurements of a patient, their corresponding random effects are linked. Specifically, the complete vector of random effects $\mathbf{b}_i = (b_{0di}, b_{1di}, b_{0pi}, \dots, b_{3pi})^\top$ is assumed to follow a multivariate normal distribution with mean zero and variance-covariance matrix \mathbf{W} .

To model the impact of DRE and PSA measurements on the risk of progression, the joint model uses a relative risk sub-model. More specifically, the hazard of progression $h_i(t)$ at a time t is given by:

$$h_i(t) = h_0(t) \exp \left(\gamma_1(\text{Age}_i - 65) + \gamma_2(\text{Age}_i - 65)^2 + \alpha_{1d} \text{logit}[\Pr\{y_{di}(t) > \text{T1c}\}] + \alpha_{1p} m_{pi}(t) + \alpha_{2p} \frac{\partial m_{pi}(t)}{\partial t} \right), \quad (6)$$

where, γ_1, γ_2 are the parameters for the effect of age. The parameter α_{1d} models the impact of log odds of obtaining a DRE > T1c on the hazard of progression. The impact of PSA on the hazard of progression is modeled in two ways: a) the impact of the error free underlying PSA value $m_{pi}(t)$, and b) the impact of the underlying PSA velocity $\partial m_{pi}(t)/\partial t$. The corresponding parameters are α_{1p} and α_{2p} , respectively. Lastly, $h_0(t)$ is the baseline hazard at time t , and is modeled flexibly using P-splines [5].

B.3 Parameter Estimates

Figure 1 shows the cumulative-risk of progression over the follow-up period. The posterior parameter estimates for the PRIAS based joint model are given in Table 4 (longitudinal sub-model for DRE outcome), Table 5 (longitudinal sub-model for PSA outcome) and Table 6 (relative risk sub-model). The parameter estimates for the variance-covariance matrix \mathbf{W} from the longitudinal sub-model are shown in Table 3. As described in Appendix A the baseline hazard of the joint model model utilized a cubic P-spline. The knots of this P-spline were placed at the following time points: 0.000, 0.000, 0.000, 0.000, 0.401, 0.801, 1.202, 1.603, 2.003, 2.404, 2.805, 3.205, 3.606, 4.007, 4.407, 4.808, 5.209, 12.542, 12.542, 12.542, 12.542. The parameters of the fitted spline function are given in Table 2.

We present plots of observed DRE versus fitted probabilities of obtaining a DRE measurement larger than T1c, for nine randomly selected patients in Figure 2. Similarly observed versus fitted PSA profiles for nine randomly selected patients are shown in Figure 3.

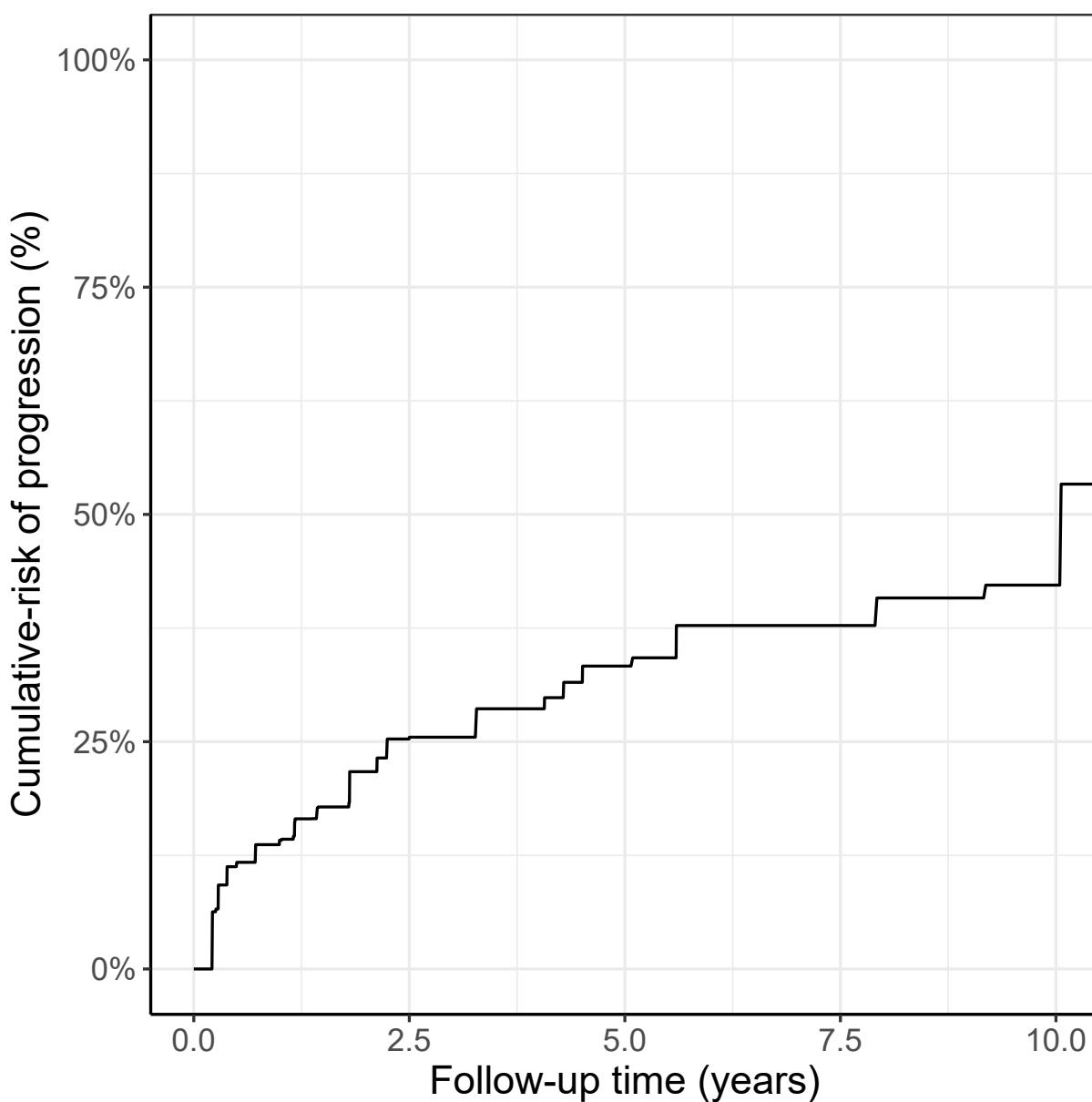


Figure 1: **Estimated cumulative-risk of cancer progression** for patients in the Prostate Cancer Research International Active Surveillance (PRIAS) dataset. Nearly 50% patients (*slow progressing*) do not progress in the ten year follow-up period. Cumulative risk is estimated using nonparametric maximum likelihood estimation [13], to account for interval censored progression times observed in the PRIAS dataset. Censoring includes death, removal from surveillance on the basis of observed longitudinal data, and patient dropout.

Table 2: Estimated parameters of the P-spline function utilized to model the baseline hazard $h_0(t)$ in joint model fitted to the PRIAS dataset. Parameters are named with the prefix ‘ps’ indicating P-spline parameter.

Variable	Mean	Std. Dev	2.5%	97.5%
ps1	-1.091	0.535	-2.286	-0.235
ps2	-2.113	0.271	-2.638	-1.591
ps3	-2.486	0.308	-3.095	-1.883
ps4	-2.083	0.311	-2.740	-1.483
ps5	-1.918	0.279	-2.460	-1.388
ps6	-2.620	0.265	-3.138	-2.140
ps7	-3.169	0.303	-3.796	-2.580
ps8	-3.416	0.340	-4.075	-2.823
ps9	-3.432	0.345	-4.103	-2.796
ps10	-3.223	0.352	-3.997	-2.573
ps11	-2.840	0.349	-3.577	-2.214
ps12	-2.481	0.350	-3.148	-1.762
ps13	-2.540	0.352	-3.206	-1.840
ps14	-2.841	0.321	-3.447	-2.212
ps15	-3.046	0.381	-3.853	-2.328
ps16	-3.113	0.701	-4.533	-1.796
ps17	-3.195	1.232	-5.894	-0.978

Table 3: Estimated variance-covariance matrix \mathbf{W} of the random effects $\mathbf{b} = (b_{0d}, b_{1d}, b_{0p}, b_{1p}, b_{2p}, b_{3p})$ from the joint model fitted to the PRIAS dataset.

Random Effects	b_{0d}	b_{1d}	b_{0p}	b_{1p}	b_{2p}	b_{3p}
b_{0d}	9.233	-0.183	-0.213	0.082	0.058	0.023
b_{1d}		-0.183	1.259	0.091	0.079	0.145
b_{0p}			-0.213	0.091	0.247	0.007
b_{1p}				0.082	0.079	0.007
b_{2p}					0.248	0.264
b_{3p}						0.189
						0.058
						0.145
						0.067
						0.018
						0.023
						0.109
						0.018
						0.327
						0.380

Table 4: Estimated mean and 95% credible interval for the parameters of the longitudinal sub-model (4) for the DRE outcome.

Variable	Mean	Std. Dev	2.5%	97.5%
(Intercept)	-4.407	0.151	-4.716	-4.113
(Age - 65)	0.057	0.009	0.039	0.075
(Age - 65) ²	-0.002	0.001	-0.004	0.000
year of visit	-1.089	0.113	-1.292	-0.866

Table 5: Estimated mean and 95% credible interval for the parameters of the longitudinal sub-model (5) for the PSA outcome.

Variable	Mean	Std. Dev	2.5%	97.5%
(Intercept)	2.687	0.007	2.674	2.701
(Age - 65)	0.008	0.001	0.006	0.010
(Age - 65) ²	-0.001	0.000	-0.001	0.000
Spline: [0.00, 0.75] years	0.199	0.009	0.181	0.217
Spline: [0.75, 2.12] years	0.293	0.012	0.269	0.316
Spline: [2.12, 6.4] years	0.379	0.014	0.352	0.406
σ	0.144	0.001	0.142	0.145

Table 6: Estimated mean and 95% credible interval for the parameters of the relative risk sub-model (6) of the joint model fitted to the PRIAS dataset.

Variable	Mean	Std. Dev	2.5%	97.5%
(Age - 65)	0.034	0.005	0.025	0.043
(Age - 65) ²	0.000	0.001	-0.001	0.001
logit{Pr(DRE > T1c)}	0.047	0.014	0.018	0.073
Fitted log ₂ (PSA + 1) value	0.024	0.076	-0.125	0.170
Fitted log ₂ (PSA + 1) velocity	2.656	0.291	2.090	3.236

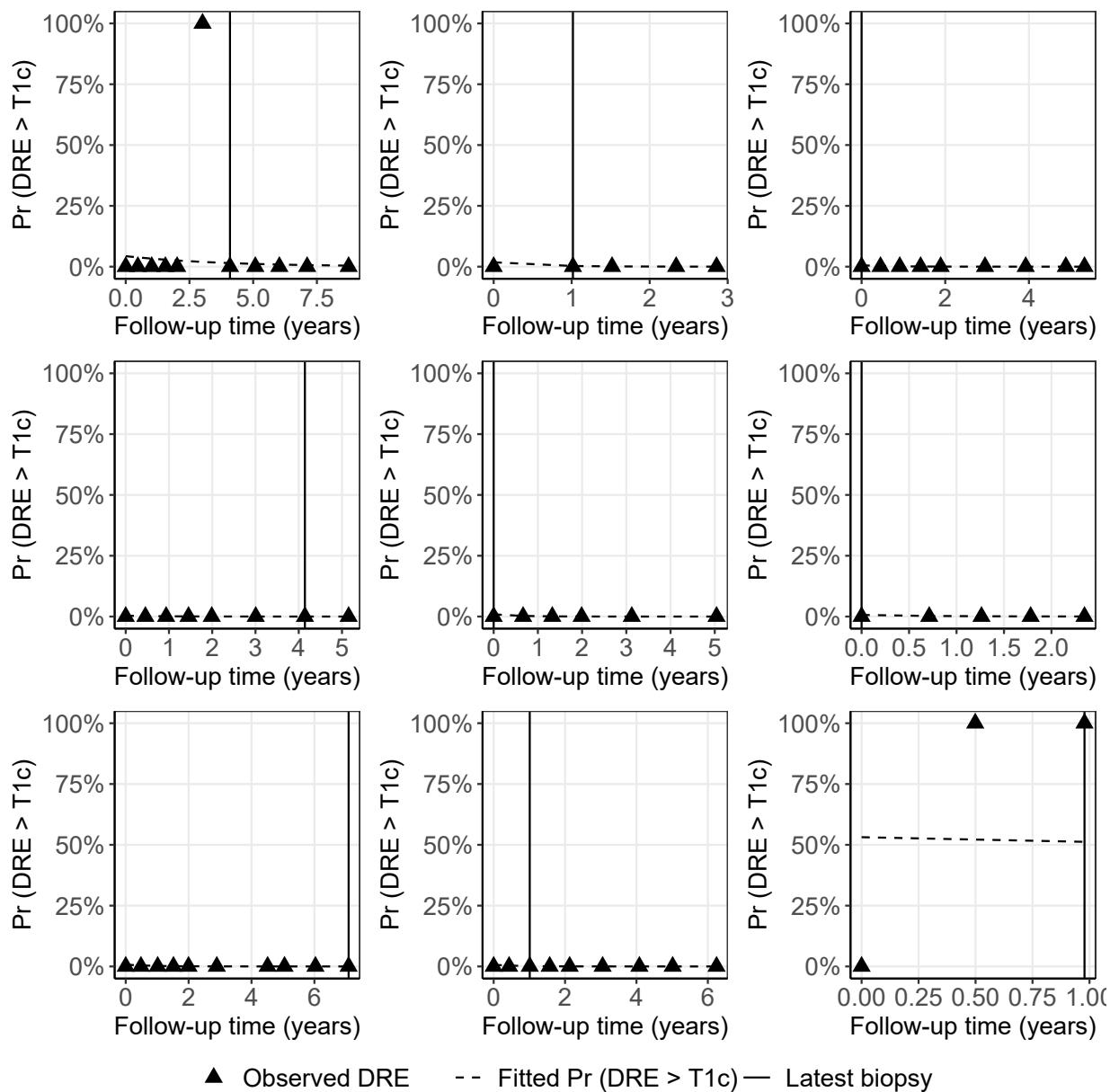


Figure 2: Observed DRE versus fitted probabilities of obtaining a DRE measurement larger than T1c, for nine randomly selected PRIAS patients. The fitted profiles utilize information from the observed DRE measurements, PSA measurements, and time of the latest biopsy. Observed DRE measurements plotted against 16.0% probability are equal to T1c. Observed DRE measurements plotted against 100% probability are larger than T1c.

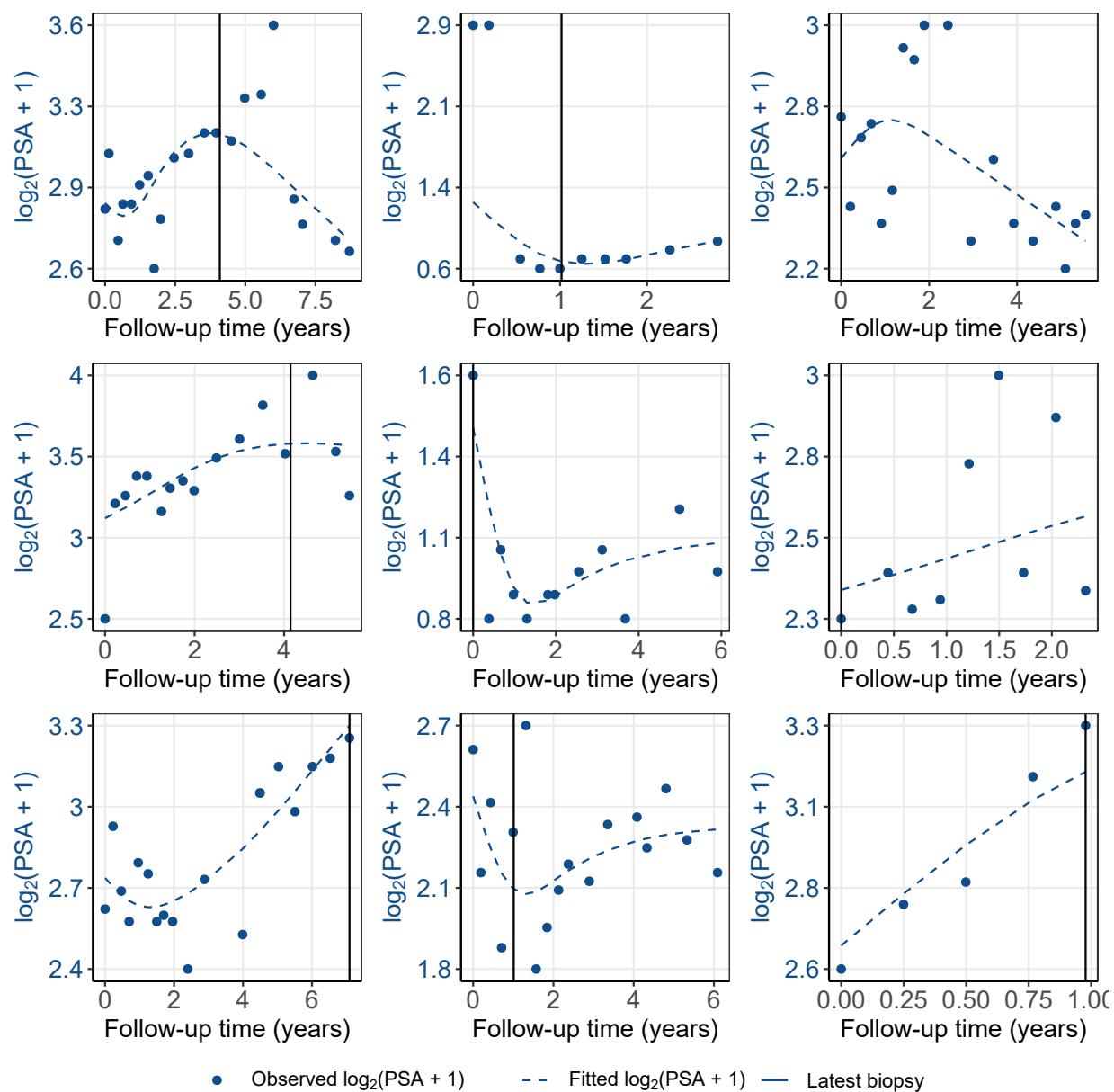


Figure 3: Fitted versus observed $\log_2(\text{PSA} + 1)$ profiles for nine randomly selected PRIAS patients. The fitted profiles utilize information from the observed PSA measurements, DRE measurements, and time of the latest biopsy.

Table 7: Data of the demonstration patient in Figure 5 of the main manuscript. Age of the patient at baseline was 60 years and time of last negative biopsy was 3.5 years. DRE: digital rectal examination.

Visit time (years)	PSA	$\log_2(\text{PSA} + 1)$	DRE > T1c
0.00	5.7	2.77	1
0.30	3.2	2.09	-
0.68	4.0	2.30	0
0.97	4.6	2.50	-
1.15	2.9	1.92	0
1.47	3.0	1.95	0
1.77	3.3	2.14	-
2.23	3.5	2.12	0
2.58	4.4	2.39	-
3.21	6.1	2.84	0
3.86	5.9	2.81	-
4.32	3.9	2.31	0
5.00	4.4	2.41	-

B.4 Assumption of t-distributed (df=3) Error Terms

With regards to the choice of the distribution for the error term ε_p for the PSA measurements (5), we attempted fitting multiple joint models differing in error distribution, namely t-distribution with three, and four degrees of freedom, and a normal distribution for the error term. However, the model assumption for the error term were best met by the model with t-distribution having three degrees of freedom. The quantile-quantile plot of subject-specific residuals for the corresponding model in Panel A of Figure 4, shows that the assumption of t-distributed (df=3) errors is reasonably met by the fitted model.

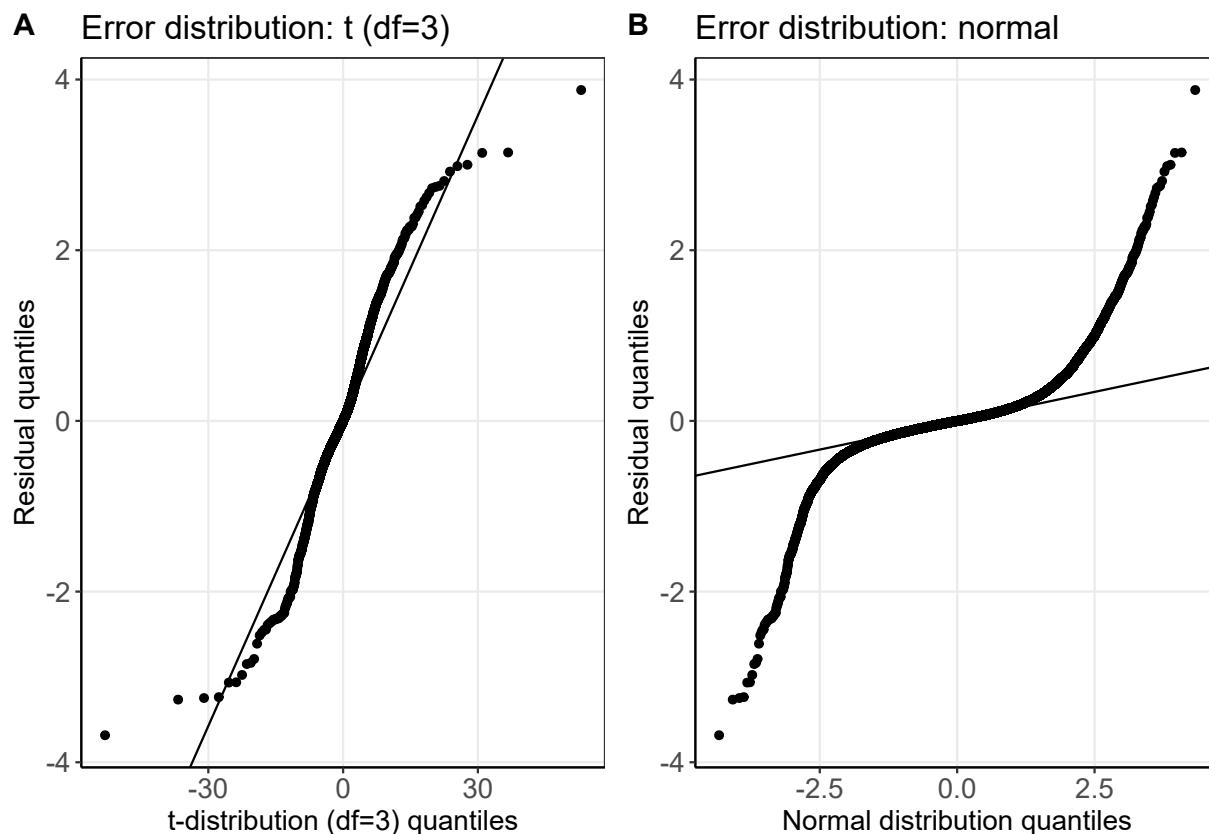


Figure 4: Quantile-quantile plot of subject-specific residuals from the joint models fitted to the PRIAS dataset. **Panel A:** model assuming a t-distribution ($df=3$) for the error term ε_p . **Panel B:** model assuming a normal distribution for the error term ε_p .

B.5 Predictive Performance of the PRIAS based Model

We calculate the predictive performance of the PRIAS based joint model using time-dependent area under the receiver operating characteristic curve or AUC (measure of discrimination), and the mean absolute prediction error or MAPE. Mathematical calculations for these in the joint modeling framework are detailed elsewhere [14]. Because these are temporal extensions of their standard versions [15] in a longitudinal setting, at every six months of follow-up (standard visit times in PRIAS), we calculated a unique AUC and MAPE for predicting risk of progression in the subsequent one year (recommended time gap between subsequent biopsies). For emulating a realistic situation, we calculated the AUC and MAPE at each follow-up using only the validation data available until that follow-up. For example, calculations for AUC and MAPE for the time interval year two to year three do not utilize data of patients who progressed before year two. The AUC and MAPE for our model are shown in Table 8.

Table 8: Follow-up time dependent, area under the receiver operating characteristic curves (AUC), and mean absolute prediction error (MAPE), with bootstrapped 95% confidence interval in brackets. The choice of year six as the maximum follow-up period is based on the reasoning that it is roughly the 95-percentile of observed follow-up times.

Follow-up period (years)	AUC (95% CI)	MAPE (95%CI)
0.0 to 1.0	0.658 [0.620, 0.693]	0.234 [0.229, 0.240]
0.5 to 1.5	0.648 [0.631, 0.663]	0.220 [0.213, 0.226]
1.0 to 2.0	0.624 [0.600, 0.644]	0.151 [0.147, 0.155]
1.5 to 2.5	0.649 [0.604, 0.704]	0.127 [0.118, 0.134]
2.0 to 3.0	0.683 [0.629, 0.729]	0.134 [0.121, 0.143]
2.5 to 3.5	0.681 [0.604, 0.739]	0.115 [0.105, 0.128]
3.0 to 4.0	0.647 [0.600, 0.710]	0.079 [0.073, 0.087]
3.5 to 4.5	0.630 [0.583, 0.668]	0.095 [0.089, 0.101]
4.0 to 5.0	0.614 [0.557, 0.659]	0.104 [0.098, 0.111]
4.5 to 5.5	0.615 [0.541, 0.702]	0.101 [0.088, 0.114]
5.0 to 6.0	0.617 [0.550, 0.713]	0.102 [0.086, 0.121]

C Simulation Study

In the simulation study, we evaluated the following biopsy schedules: biopsy every year (annual), biopsy according to the PRIAS schedule (PRIAS), personalized biopsy schedules based on two fixed risk thresholds, namely, $\kappa = 10\%$, and automatically chosen $\kappa^*(v)$ (Section 3 of main manuscript), and automatically chosen $\kappa^*\{v \mid E(D) \leq 0.75\}$ with a constraint of 0.75 years (9 months) on expected delay in detecting progression. The choice of 0.75 years delay constraint is arbitrary and is only used to illustrate that applying the constraint limits the average delay at 0.75 years. We compare all the aforementioned schedules on two criteria, namely the number of biopsies they schedule and the corresponding time delay in detection of cancer progression, in years (time of positive biopsy - true time of cancer progression). The corresponding results, using 500×250 test patients are presented in 9. 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 500×250 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.

Table 9: **Simulation study results for all patients:** Estimated mean (μ), median (Med), first quartile Q_1 , and third quartile Q_3 for number of biopsies (nb) and for the time delay (d) in detection of cancer progression in years, for various biopsy schedules. The delay is equal to the difference between the time of the positive biopsy and the simulated true time of progression. Types of schedules: $\kappa = 10\%$ and $\kappa^*(v)$ schedule a biopsy if the cumulative-risk of cancer progression at a visit is more than 10%, and an automatically chosen threshold, respectively. Schedule $\kappa^*\{v \mid E(D) \leq 0.75\}$ is similar to $\kappa^*(v)$ except that the euclidean distance is minimized under the constraint that expected delay in detecting progression is at most 9 months (0.75 years). Annual corresponds to a schedule of yearly biopsies, and PRIAS corresponds to biopsies as per PRIAS protocol.

Progressing patients (50%)

Schedule	Q_1^{nb}	μ^{nb}	Med ^{nb}	Q_3^{nb}	Q_1^d	μ^d	Med ^d	Q_3^d
Annual	1	3.71	3	6	0.29	0.55	0.57	0.82
PRIAS	1	2.88	2	4	0.38	0.92	0.74	1.00
$\kappa = 10\%$	1	2.55	2	4	0.45	1.00	0.85	1.33
$\kappa^*(v)$	1	2.46	2	3	0.45	0.89	0.86	1.26
$\kappa^*\{v \mid E(D) \leq 0.75\}$	1	3.39	3	5	0.32	0.61	0.63	0.88

Non-progressing patients (50%)

Annual	10	10.00	10	10	-	-	-	-
PRIAS	4	6.40	6	8	-	-	-	-
$\kappa = 10\%$	4	4.91	5	6	-	-	-	-
$\kappa^*(v)$	6	6.22	6	7	-	-	-	-
$\kappa^*\{v \mid E(D) \leq 0.75\}$	8	8.68	9	9	-	-	-	-

D Source Code

Source code URL: <https://github.com/anirudhtomer/PersonalizedSchedules>. It has also been uploaded as a zip file with the manuscript.

References

- [1] Charles E McCulloch and John M Neuhaus. Generalized linear mixed models. *Encyclopedia of Biostatistics*, 4, 2005.
- [2] Elizabeth R. Brown. 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(3):1163–1182, 2009.
- [3] Dimitris Rizopoulos. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. CRC Press, 2012.
- [4] Jeremy M.G. Taylor, Yongseok Park, Donna P. Ankerst, Cecile Proust-Lima, Scott Williams, Larry Kestin, Kyoungwha Bae, Tom Pickles, and Howard Sandler. Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics*, 69(1):206–213, 2013.
- [5] Paul HC Eilers and Brian D Marx. Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–121, 1996.
- [6] Eleni-Rosalina Andrinopoulou and Dimitris Rizopoulos. Bayesian shrinkage approach for a joint model of longitudinal and survival outcomes assuming different association structures. *Statistics in Medicine*, 35(26):4813–4823, 2016.
- [7] Stefan Lang and Andreas Brezger. Bayesian p-splines. *Journal of Computational and Graphical Statistics*, 13(1):183–212, 2004.
- [8] Astrid Jullion and Philippe Lambert. Robust specification of the roughness penalty

1
2
3 prior distribution in spatially adaptive bayesian p-splines models. *Computational*
4
5 *Statistics & Data Analysis*, 51(5):2542–2558, 2007.
6
7

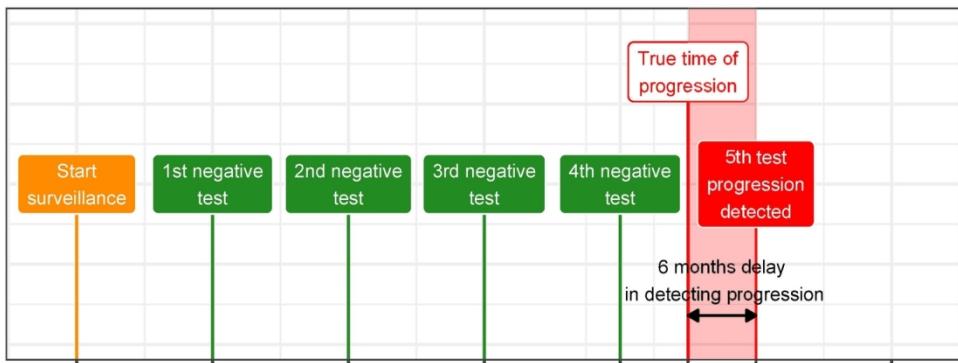
- 8 [9] FH Schröder, P Hermanek, L Denis, WR Fair, MK Gospodarowicz, and M Pavone-
9 Macaluso. The TNM classification of prostate cancer. *The Prostate*, 21(S4):129–138,
10 1992.
11
12 [10] Jay D Pearson, Christopher H Morrell, Patricia K Landis, H Ballentine Carter, and
13 Larry J Brant. Mixed-effects regression models for studying the natural history of
14 prostate disease. *Statistics in Medicine*, 13(5-7):587–601, 1994.
15
16 [11] Haiqun Lin, Charles E McCulloch, Bruce W Turnbull, Elizabeth H Slate, and Larry C
17 Clark. A latent class mixed model for analysing biomarker trajectories with irregularly
18 scheduled observations. *Statistics in Medicine*, 19(10):1303–1318, 2000.
19
20 [12] Carl De Boor. *A practical guide to splines*, volume 27. Springer-Verlag New York,
21 1978.
22
23 [13] Bruce W Turnbull. The empirical distribution function with arbitrarily grouped, cen-
24 sored and truncated data. *Journal of the Royal Statistical Society. Series B (Methodo-*
25 *logical)*, 38(3):290–295, 1976.
26
27 [14] Dimitris Rizopoulos, Geert Molenberghs, and Emmanuel MEH Lesaffre. Dynamic
28 predictions with time-dependent covariates in survival analysis using joint modeling
29 and landmarking. *Biometrical Journal*, 59(6):1261–1276, 2017.
30
31 [15] Ewout W Steyerberg, Andrew J Vickers, Nancy R Cook, Thomas Gerdts, Mithat Go-
32 nen, Nancy Obuchowski, Michael J Pencina, and Michael W Kattan. Assessing the
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 performance of prediction models: a framework for some traditional and novel mea-
4 sures. *Epidemiology (Cambridge, Mass.)*, 21(1):128, 2010.
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review

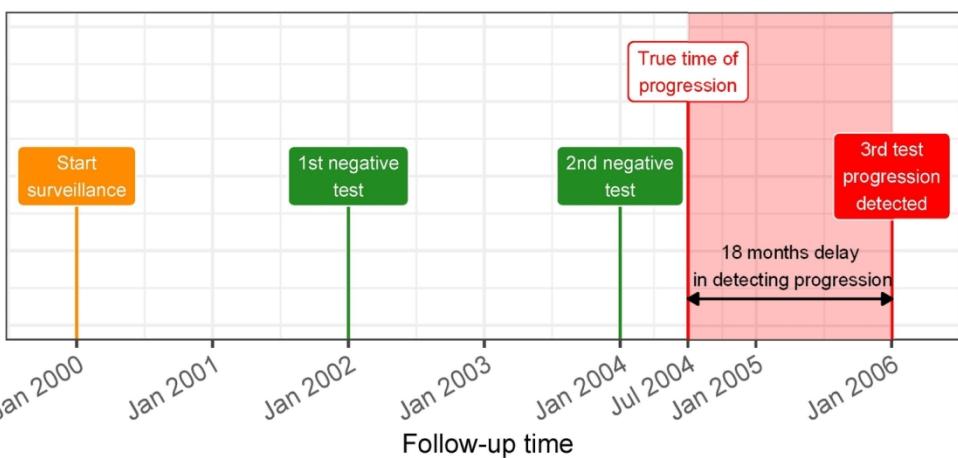
1
2
3
4
5
6

A Frequent tests - Shorter delay in detecting progression



21

B Infrequent tests - Longer delay in detecting progression



36

Figure 1

Goal: Finding the optimal tradeoff between the number of invasive tests (burden) and time delay in detecting progression (shorter is beneficial). A progression is a non-terminal event in the surveillance of early-stage chronic non-communicable diseases. The true time of progression for the patient illustrated in this figure is July 2004. Since invasive tests are conducted repeatedly, progression is interval-censored and always observed with a delay. Frequent periodical invasive tests in Panel A lead to a shorter time delay in detecting progression than infrequent periodical invasive tests in Panel B. The interval-censored time of progression is Jan 2004 - Jan 2005 in Panel A and between Jan 2004 - Jan 2006 in Panel B.

45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

133x127mm (300 x 300 DPI)

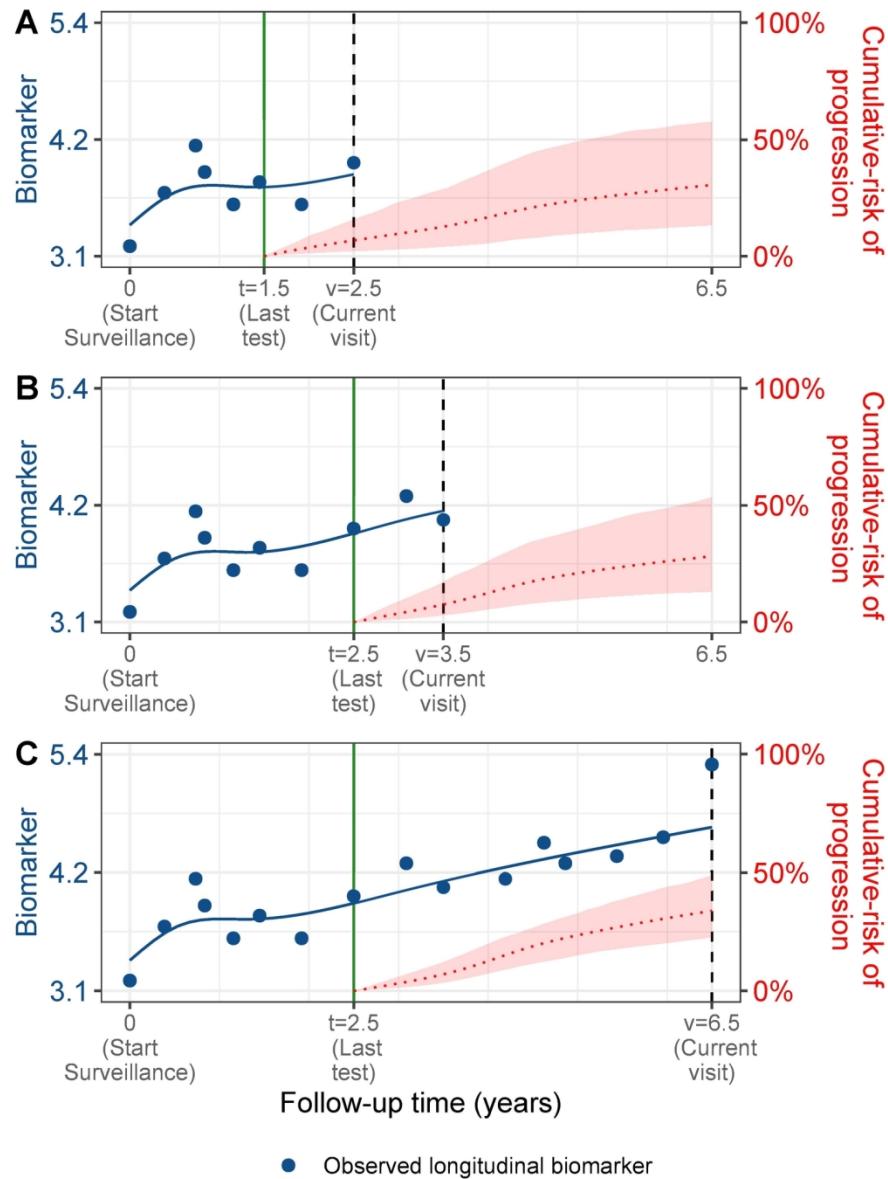


Figure 2

Cumulative-risk of progression updated 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, and B: are ordered by the time of the current visit v (dashed vertical black line) of a new patient. At each of these visits, we combine the accumulated longitudinal data (shown in blue circles), and time of the last negative invasive test t (solid vertical green line) to obtain the updated cumulative-risk profile $R_j(u | t, v)$ (dotted red line with 95% credible interval shaded) of the patient defined in (1). All values are illustrative.

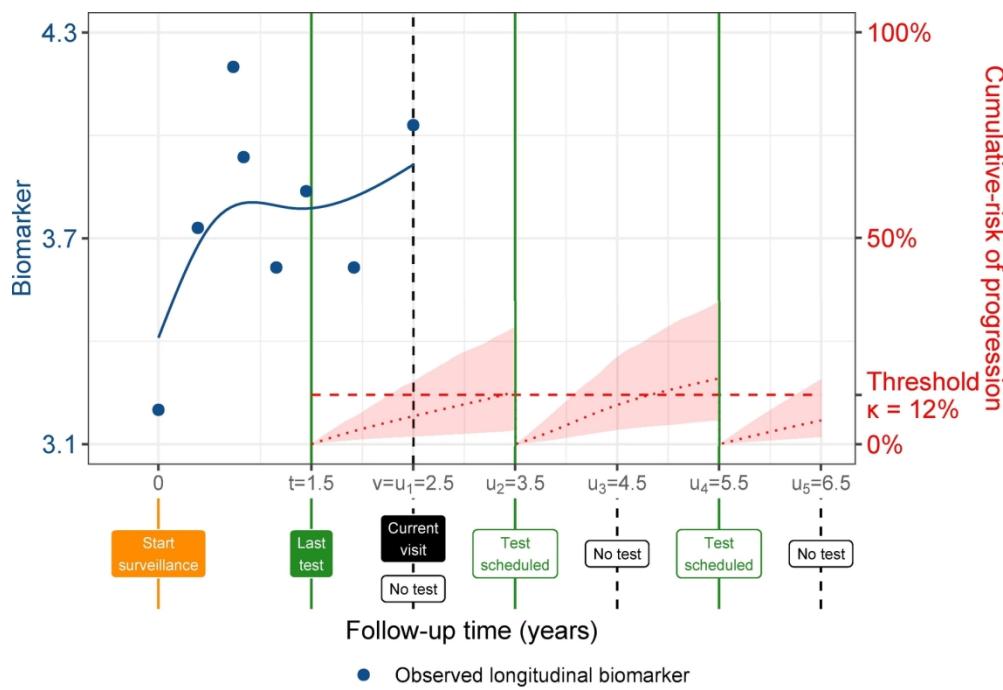


Figure 3

Successive personalized test decisions based on patient-specific cumulative-risk of progression (2). Time of current visit: $v=2.5$ years (dashed vertical black line). Time of the last test on which progression was not observed: $t=1.5$ years. Longitudinal data up to current visit: $Y_j(v)$ is a continuous biomarker (blue circles). Example risk threshold: $k=0.12$ (12%). Grid of future visits on which future tests are planned: $U = \{2.5, 3.5, 4.5, 5.5, 6.5\}$ years. The cumulative-risk profiles $R_j(u_l | t_l, v)$ employed in (2) are shown with dotted red lines (95% credible intervals shaded), and are updated each time a test is planned (solid vertical green lines). Future test decisions $Q_j(u_l | t_l, v)$ defined in (2) are: $Q_j^k(u_1=2.5 | t_1=1.5, v)=0$, $Q_j^k(u_2=3.5 | t_2=1.5, v)=1$, $Q_j^k(u_3=4.5 | t_2=3.5, v)=0$, $Q_j^k(u_4=5.5 | t_2=3.5, v)=1$, and $Q_j^k(u_5=6.5 | t_5=4.5, v)=0$. All values are illustrative.

165x114mm (300 x 300 DPI)

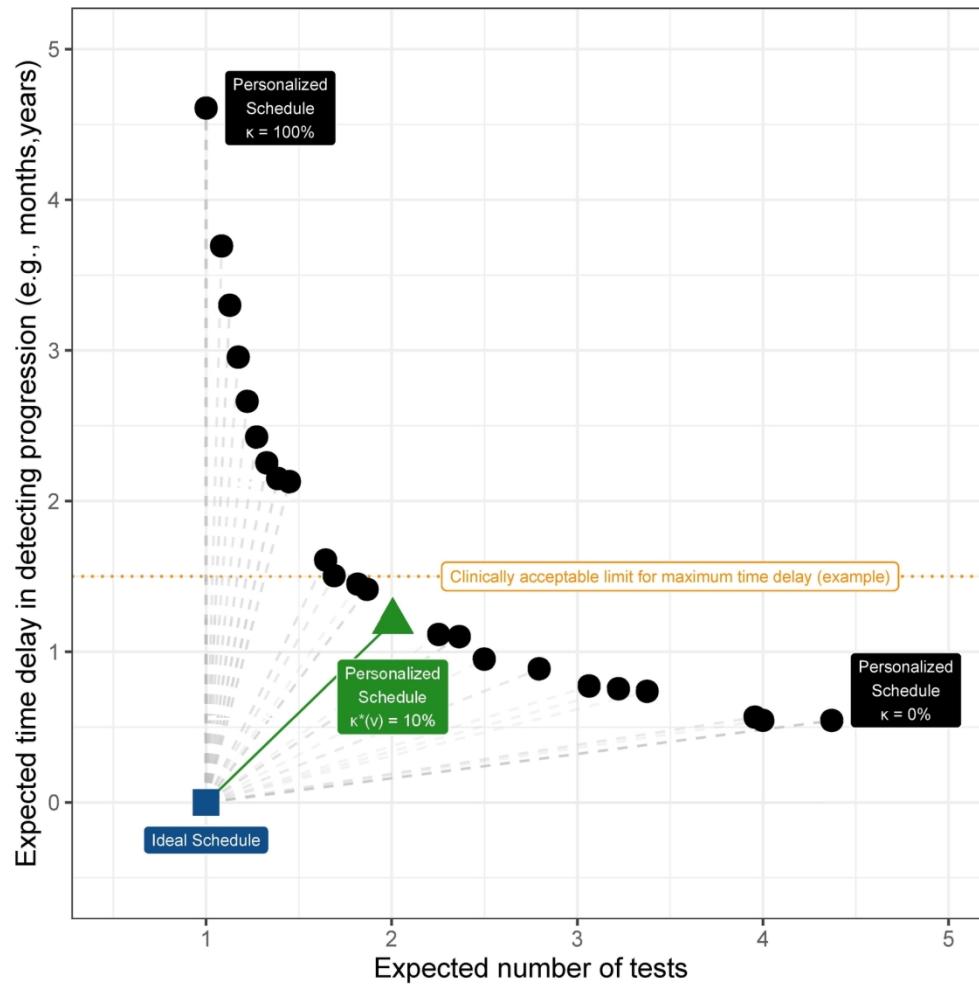


Figure 4

Optimal current-visit time v specific risk threshold $k^*(v)$ obtained using (4) for the patient shown in Figure 3. Ideal schedule of tests: point $(1,0)$ shown as a blue square. It plans exactly one invasive test at the true time of progression T^{*j} of a patient. Hence, the time delay in detecting progression is zero. Various personalized schedules based on a grid of thresholds $k \in [0,1]$ are shown with black circles. Higher thresholds lead to fewer tests, but also higher expected time delay. The personalized schedule based on $k^*(v)=9.5\%$ threshold (green triangle) has the least Euclidean distance (solid green line) to the ideal schedule. It is also possible to optimize the least distance under a certain clinically acceptable limit on the time delay (dotted horizontal orange line).

152x152mm (300 x 300 DPI)

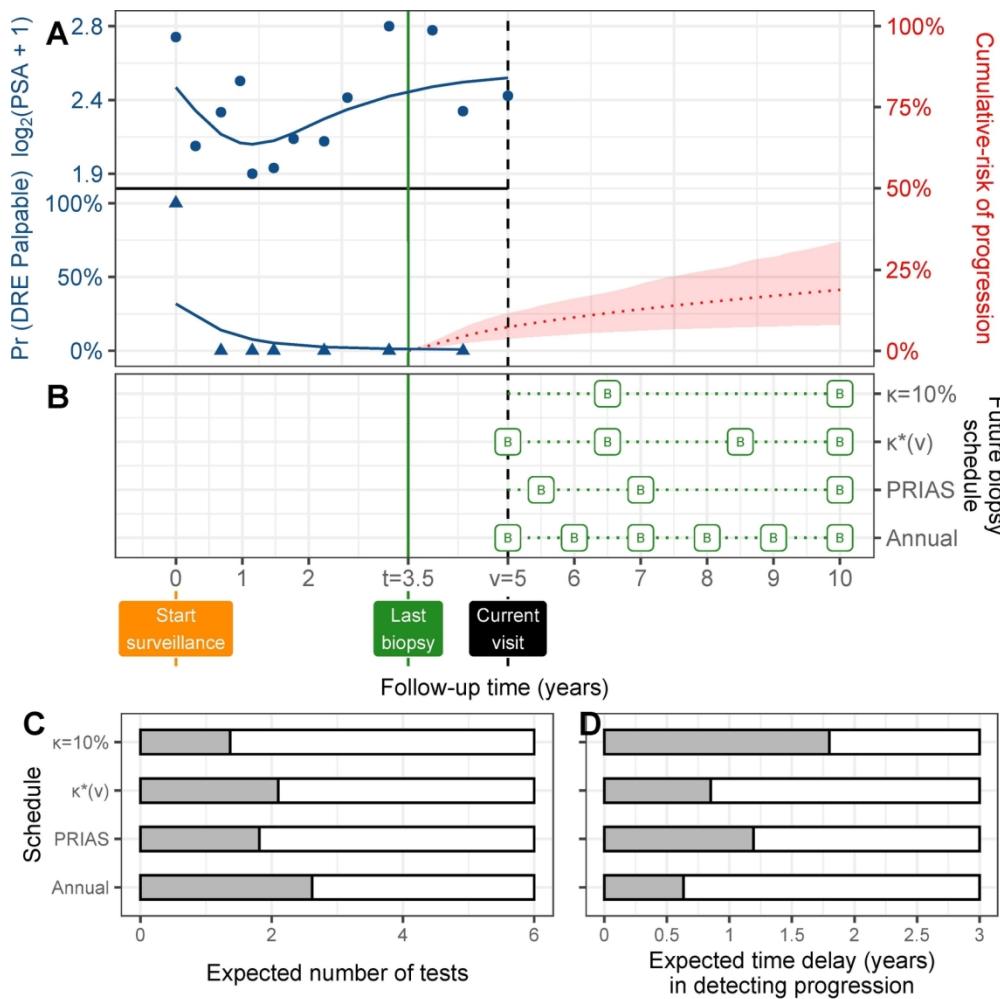


Figure 5

Personalized schedules for a demonstration prostate cancer patient. Panel A: Time of current visit: $v=5$ years (black dashed line). Time of last negative biopsy: $t=3.5$ years (vertical green solid line). Longitudinal data: $\log_2(\text{PSA} + 1)$ transformed PSA (observed: blue dots, fitted: solid blue line), and binary DRE (observed: blue triangles, fitted probability: solid blue line). Cumulative-risk profile: dotted red line (95% credible interval shaded). Panel B: Biopsy indicated with a 'B', and $k=10\%$ and $k^*(v)$ are personalized biopsy schedules using a risk threshold of 10%, and a visit time v specific automatically chosen threshold (4), respectively. PRIAS and Annual denote the PRIAS biopsy schedule (Section 4.1) and annual biopsy schedule. Panel C,D} For all schedules we calculate the expected number of tests and expected time delay in detecting progression if the patient progresses before year ten. Since a recommended minimum gap of one year is maintained between biopsies, maximum possible number of tests are six. A delay in detecting progression of up to three years may not lead to adverse outcomes [31].

152x152mm (300 x 300 DPI)

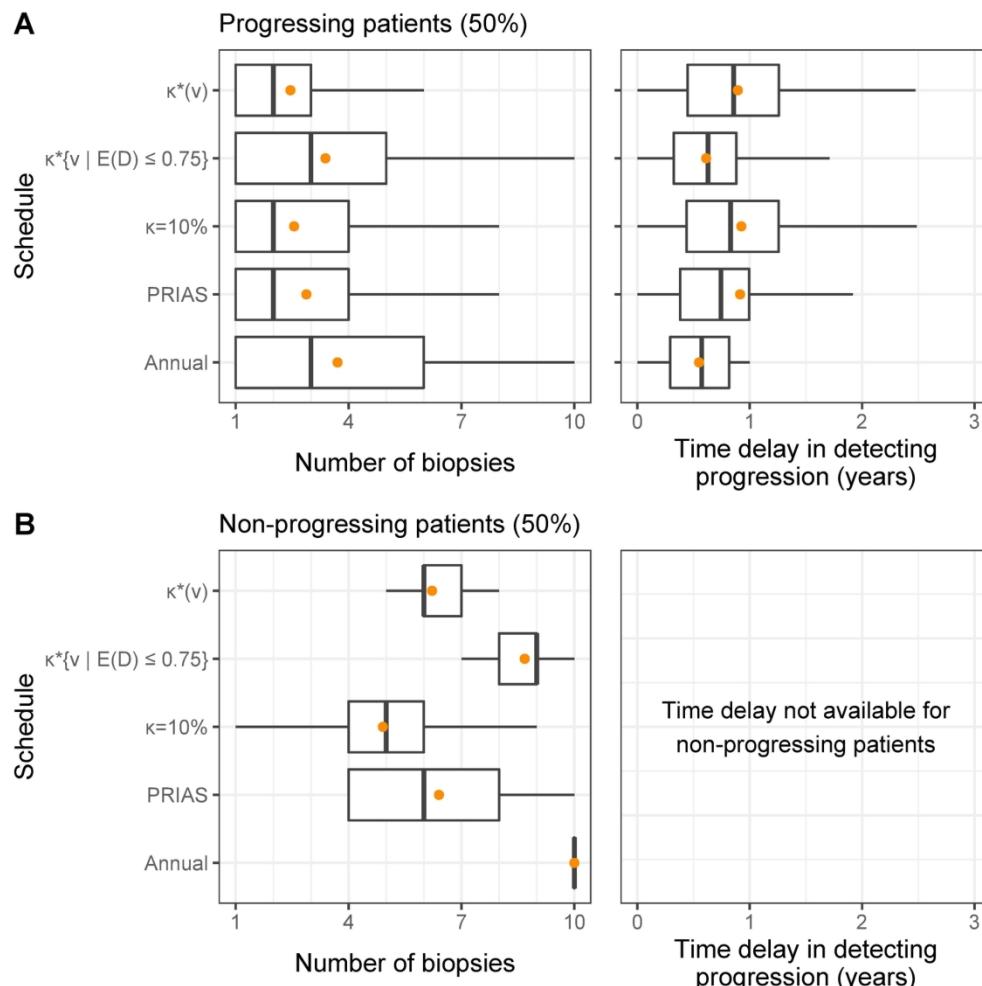


Figure 6

Number of biopsies and the time delay in detecting cancer progression for various biopsy schedules obtained via a simulation study. Mean is indicated by the orange circle. Time delay (years) is calculated as (time of positive biopsy - the actual simulated time of cancer progression). Biopsies are conducted until cancer progression is detected. Panel A: simulated patients who obtained cancer progression in the ten year study period (progressing). Panel B: simulated patients who did not obtain cancer progression in the ten year study period (non-progressing). Types of schedules: $k=10\%$ and $k^*(v)$ schedule a biopsy if the cumulative-risk of cancer progression at the current visit time v is more than 10%, and an automatically chosen threshold (4), respectively. Schedule $k^*[v | E(D) \leq 0.75]$ is similar to $k^*(v)$ except that the euclidean distance in (4) is minimized under the constraint that expected delay in detecting progression is at most 9 months (0.75 years). Annual corresponds to a schedule of yearly biopsies, and PRIAS corresponds to biopsies as per PRIAS protocol (Section 4).

165x165mm (300 x 300 DPI)