Biometrics 00, 1–5
December 0000

DOI: 10.1111/j.1541-0420.2005.00454.x

Personalized Schedules for Burdensome Surveillance Tests in Chronic Diseases

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

[1]Department of Biostatistics, Erasmus University Medical Center, the Netherlands
[2]Department of Public Health, Erasmus University Medical Center, the Netherlands
[3]Department of Urology, Erasmus University Medical Center, the Netherlands
[4]Department of Biomedical Data Sciences, Leiden University Medical Center, the Netherlands
email: a.tomer@erasmusmc.nl

Summary: This is the summary for this paper.

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

1. Introduction

Chronic non-communicable diseases (e.g., cancer, renal, cardiovascular diseases, etc.) are the primary cause of human deaths worldwide (Alwan et al., 2010). Early identification of unfavorable changes in disease state (referenced as *progression* hereafter) may prevent many of these deaths. To this end, surveillance tests are performed routinely for several diseases. Among all surveillance modalities, the most accurate or gold standard ones are often invasive. For example, biopsies, endoscopies, and colonoscopies are conducted repeatedly for diagnosing progression in prostate cancer (Bokhorst et al., 2016), Barrett's esophagus (Streitz et al., 1993), and colorectal cancer (Lieberman et al., 2012), respectively. Similarly, repeat biopsies are employed to detect allograft deterioration in lung (McWilliams et al., 2008) and kidney transplant (Henderson et al., 2011) patients.

Commonly, fixed schedules (e.g., every six months) are utilized for invasive tests. The frequency of tests varies between diseases and cohorts. Due to the periodical nature of schedules, progression is always detected with a time delay (see Figure). More frequent tests can lead to less time delay in detection of progression. However, invasive tests are difficult to conduct, are prone to severe complications, and cause patient discomfort. Furthermore, fixed schedules impose an equal medical burden, ignoring the speed of progression in patients. Hence, the frequency of invasive tests holds important implications for patients.

In this paper, we aim to balance the number of invasive tests (burden) and the time delay in detection of progression (less is beneficial) better than fixed schedules. For this purpose, we intend to create personalized test schedules that exploit patient-specific accumulated clinical data during follow-up. This data includes baseline characteristics of patients; results from previous invasive tests; and longitudinal biomarker, physical examination, and medical imaging measurements, etc. Previous approaches for personalized schedules may be divided into three categories. First, heuristic methods such as decision making flowcharts (citation here). Although, flowcharts discretize continuous clinical outcomes, often rely only the latest data measurement, and ignore the measurement error in observed outcomes. Second, personalized test decisions utilizing partially observable Markov decision processes (Alagoz et al., 2010; Steimle and Denton, 2017). However, their application with continuous outcomes is limited by the curse of dimensionality. Third, personalized schedules obtained by optimizing a loss function of clinical parameters of interest (Bebu and Lachin, 2017; Rizopoulos et al., 2015), including our previous work on scheduling biopsies in prostate cancer (Tomer et al., 2019). In this work, we will employ the third approach.

First, we develop a full specification of the joint distribution of patient-specific accumulated clinical data and time of *progression*. We achieve this using joint models for time-to-event and longitudinal data (Tsiatis and Davidian, 2004; Rizopoulos, 2012). We exploit joint models because they are inherently personalized. Specifically, they exploit patient-specific random effects (Laird and Ware, 1982) to model longitudinal outcomes without discretizing them. We subsequently employ the fitted joint model for new patients, to estimate their patient-specific cumulative-risk profile over the whole follow-up period. These risk predictions utilize their complete observed clinical data. We then schedule invasive tests on all those future follow-up visits where a patient's conditional cumulative-risk of progression is above a certain threshold (e.g., 10% risk). We automate the choice of this threshold and the resulting schedule. More specifically, we optimize a function of the number of tests in a schedule and the expected time delay in the detection of progression. We estimate this delay in a patient-specific manner for both fixed and personalized schedules, thus facilitating shared-decision of test schedules.

This research is motivated by the problem of scheduling biopsies (Nieboer et al., 2018) in the world's largest prostate cancer active surveillance (AS) study PRIAS (Bokhorst

et al., 2016). It has 7813 patients from 100 medical centers worldwide, employing a common study protocol. These patients have low and very-low grade prostate cancer, often over-diagnosed due to prostate-specific antigen (PSA) based screening tests (Crawford, 2003). The goal of AS is to delay serious treatments (e.g., surgery, chemotherapy, etc.) until cancer progression is observed. For this purpose, patients are monitored continually via PSA (ng/mL) blood tests, digital rectal examination (DRE) for shape and size of the tumor, and biopsy Gleason grade group (Epstein et al., 2016). Since the latter is the strongest indicator of cancer-related outcomes, treatment is commonly advised upon observing an increase (cancer progression) in a patient's biopsy Gleason grade group. Currently, the most common biopsy schedule is yearly biopsies (Loeb et al., 2014). However, this leads to many unnecessary biopsies in slow/non-progressing patients (50% proportion in some cohorts). Biopsy burden combined with patient non-compliance to frequent biopsies (Bokhorst et al., 2015) has raised concerns regarding the optimal biopsy schedule. Prostate cancer is the second most frequently diagnosed cancer in males worldwide (Torre et al., 2015). Hence, biopsy schedules tailored for each AS patient can reduce the overall burden of biopsies in a large number of patients.

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

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

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

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

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

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

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

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

where $h_0(\cdot)$ denotes the baseline hazard function, $\boldsymbol{w}_i(t)$ is a vector of exogenous, possibly time-varying, covariates with corresponding regression coefficients $\boldsymbol{\gamma}$. Functions $f_{kl}(\cdot)$, parameterized by coefficients α_{kl}, specifies which components/features of each longitudinal outcome are included in the linear predictor of the relative risk model (Brown, 2009; Rizopoulos, 2012; Taylor et al., 2013). Some examples, motivated by the literature (subscripts k and l dropped for brevity), are

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

These formulations of $f_{kl}(\cdot)$ 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. The detailed specification of the baseline hazard, and parameter estimation using the Bayesian approach are presented in Web Appendix A of the supplementary material. Lastly, $h_0(t)$ is the baseline hazard at time t, and is modeled flexibly using P-splines (Eilers and Marx, 1996). The detailed specification of the baseline hazard $h_0(t)$, and the joint parameter estimation of the longitudinal and relative-risk sub-models using the Bayesian approach are presented in WebAppendix A.

3. Personalized Schedule for Biopsies

We intend to develop a personalized schedule of biopsies for a new patient j not present in training dataset $\mathcal{D}_n$. The schedule of longitudinal measurements remains fixed. Let T_j^* be the true time of his Gleason upgrade, $t < T_j^*$ be the time of his latest negative biopsy, and $s > t$ be the time of his latest visit for longitudinal measurements.

3.1 *Cumulative-risk of Gleason Upgrade*

The first step is to consolidate his observed clinical data, namely all longitudinal PSA $\mathcal{Y}_{pj}(s)$ and DRE $\mathcal{Y}_{dj}(s)$ measurements, and previous biopsy results $T_j^* > t$, into a patient-specific cumulative risk of Gleason upgrade. Since the current follow-up period of PRIAS is limited, we are able to estimate this risk only for the first ten years of follow-up. It is given by:

$$R_j(u \mid t, s) = p\Big\{T_j^* \leqslant u \mid T_j^* > t, \mathcal{Y}_{pj}(s), \mathcal{Y}_{dj}(s), \mathcal{D}_n\Big\}, \quad s \leqslant u \leqslant 10. \quad (3)$$

An advantage of this cumulative-risk is that it updates as more longitudinal or biopsy data becomes available over follow-up.

3.2 *Schedule of Biopsies*

Our aim is to employ this cumulative-risk function in the personalized biopsy schedule. However, in line with the protocols of most AS cohorts (Nieboer et al., 2018), we first schedule a compulsory biopsy at year one of follow-up. This promises early detection of Gleason upgrade for patients misdiagnosed as low-grade cancer patients, or patients who chose AS despite having a higher grade at diagnosis. We also maintain a recommended minimum gap of one year between consecutive biopsies (Bokhorst et al., 2016). Consequently, we schedule personalized biopsies starting from year two until year ten (Equation 3) of follow-up. The added benefit of this approach is that due to the longitudinal measurements accumulated over two years, and year one biopsy results, we are able to make reasonably accurate predictions of the cumulative-risk of Gleason upgrade.

We exploit PRIAS cohort's fixed schedule of longitudinal measurements $L = \{2, 2.5 \dots 10\}$ between year two and ten, for the personalized biopsy schedule. More specifically, we schedule a biopsy at all those future visits where the conditional cumulative-risk of Gleason upgrade is larger than a certain threshold $0 \leqslant \kappa \leqslant 1$ (e.g., 10% risk). The resulting personalized schedule of biopsies B_j^κ is given by:

$$B_j^\kappa = \Big\{ b_{jk} \epsilon L \mid R_j(b_{jk} \mid b_{jk-1}, s) \geqslant \kappa \wedge (b_{jk} - b_{jk-1} \geqslant 1) \Big\}, \tag{4}$$

where b_{jk} is the time of the k-th biopsy for the j-th patient. The conditional cumulative-risk of Gleason upgrade denoted by $R_j(b_{jk} \mid b_{jk-1}, s)$ is defined as in Equation (3). In this risk the contribution of the observed longitudinal data $\mathcal{Y}_{pj}(s)$ and $\mathcal{Y}_{dj}(s)$ does not change while scheduling subsequent biopsies. However, the 'conditional' part here is that successive k-th biopsy at time b_{jk} is scheduled by accounting for the possibility that Gleason upgrade may not have occurred until the previously scheduled biopsy $T_j^* > b_{jk-1}$.

The personalized schedule Equation 4 is updated as more patient data becomes available over follow-up.

3.3 *Risk Threshold* κ

The risk threshold κ controls the timing and total number of biopsies in the schedule B_j^κ. Through the timing of biopsies, κ also indirectly affects the time delay that may occur in detection of Gleason upgrade. Hence, κ should be chosen while balancing both the number of biopsies (burden), and the delay in detection of Gleason upgrade (less is beneficial).

Consider the bi-dimensional Euclidean space of the total number of biopsies (x-axis) and the corresponding expected time delay in detection of Gleason upgrade (y-axis) for schedules associated with various κ (Figure...). An ideal schedule of biopsies will have only one biopsy planned exactly at the true time of Gleason upgrade T_j^* of a patient. In other words it will lead to a zero time delay. This schedule is shown at point (1, 0) in Figure. Subsequently, an appropriate threshold κ_a can be chosen by minimizing the Euclidean distance between the point (1,0) and the set of points representing various schedules corresponding to each $\kappa \epsilon [0, 1]$. That is:

$$\kappa_a = \arg\min_\kappa \sqrt{\big(|B_j^\kappa| - 1\big)^2 + \big\{D_j(B_j^\kappa \mid t, s) - 0\big\}^2}, \quad 0 \leqslant \kappa \leqslant 1 \tag{5}$$

where, $D_j(B_j^\kappa \mid t, s)$ denotes the expected time delay in detection of Gleason upgrade (estimation in Section 3.4) if schedule B_j^κ is followed.

Certain patients may have preferences for the maximum number of biopsies conducted upon them. Others may be apprehensive to have an expected delay higher than a certain number of months. In this regard, the Euclidean distance in Equation (5) can be minimized under constraints on the aforementioned criteria (see Figure..). For example, a reasonable constraint on expected time delay is one year, as it is also the maximum possible delay with the commonly used yearly schedule.

3.4 *Expected Time Delay in Detection of Gleason Upgrade*

We estimate the expected time delay $D_j(B_j^\kappa \mid t, s)$ in Equation 5 in a patient-specific manner using personalized cumulative-risk profile estimated in Equation (3). That is, two patients may opt to follow the same schedule, but they will expect different time delays. The calculation of delay is not limited to personalized schedules. In general, for any schedule of biopsies B, the personalized expected delay for j-th patient is given by $D_j(B, t, s)$:

$$D_j(B, t, s) = \sum_{k=1}^{|B|} R_j(b_k \mid b_{k-1}, s) \times \Big\{ b_k - b_{k-1} - \int_{b_{k-1}}^{b_k} 1 - R_j(u$$

$$R_j(u \mid b_k, b_{k-1}, s) = p\Big\{ T_j^* \leqslant u \mid b_{k-1} < T_j^* \leqslant b_k, \mathcal{Y}_{pj}(s), \mathcal{Y}_{dj}(s), \mathcal{D}_n \Big\}, \tag{6}$$

where b_k is the k-th biopsy in schedule B.

Personalized expected delay can assist patients and doctors in shared decision making of an appropriate biopsy schedule. Although, this delay should only be interpreted as the expected delay if the patient obtains Gleason upgrade before the last biopsy in the schedule. In order to have a fair comparison of expected delay between different schedules for the same patient, we schedule a compulsory biopsy at year ten (see Section 3.1) in all schedules, personalized or fixed.

4. Demonstration of Personalized Schedules

5. Simulation Study

Although we demonstrated our schedules for a real AS patient, we also intend to analyze the efficacy of our schedules in the whole PRIAS cohort. Comparing personalized schedules with fixed schedules using a real word RCT is not possible because of two reasons. First, patients in PRIAS have already had their biopsies. Second, we do not know the true time of Gleason upgrade for real patients. Instead, we utilize the joint model fitted to the PRIAS cohort to generate simulated cohorts that are replicas of PRIAS cohort.

Although the personalized decision making approach is motivated by the PRIAS study, it is not possible to evaluate it directly on the PRIAS dataset. This is because the patients in PRIAS have already had their biopsies as per the PRIAS protocol. In addition, the true time of cancer progression is interval or right censored for all patients, making it impossible to correctly estimate the delay in detection of cancer progression due to a particular schedule. To this end, we conduct an extensive simulation study to find the utility

of personalized, PRIAS, and fixed/heuristic schedules. For a realistic comparison, we simulate patient data from the joint model fitted to the PRIAS dataset. The simulated population has the same ten year follow-up period as the PRIAS study. In addition, the estimated relations between DRE and PSA measurements, and the risk of cancer progression are retained in the simulated population.

5.1 *Simulation Setup*

From this population, we first sample 500 datasets, each representing a hypothetical AS program with 1000 patients in it. We generate a true cancer progression time for each of the 500×1000 patients and then sample a set of DRE and PSA measurements at the same follow-up visit times as given in PRIAS protocol. We then split each dataset into training (750 patients) and test (250 patients) parts, and generate a random and noninformative censoring time for the training patients. We next fit a joint model of the specification given in Equations (**??**), (**??**), and (2) to each of the 500 training datasets and obtain MCMC samples from the 500 sets of the posterior distribution of the parameters.

In each of the 500 hypothetical AS programs, we utilize the corresponding fitted joint models to develop cancer progression risk profiles for each of the 500×250 test patients. We make the decision of biopsies for patients at their pre-scheduled follow-up visits for DRE and PSA measurements, on the basis of their estimated personalized cumulative risk of cancer progression. These decisions are made iteratively until a positive biopsy is observed. A recommended gap of one year between consecutive biopsiesBokhorst et al. (2015) is also maintained. Subsequently, for each patient, an entire personalized schedule of biopsies is obtained.

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

5.2 *Results*

6. Discussion

Put your final comments here. Key points:

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

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

Figure (all should be clear in black and white and in color): 1. time delay explanation figure 2. joint model figure 3. figure of biopsy schedule in demo patient 4. figure of distance explaining what we did 5. simulation results boxplot 6. figure of schedule markov decision like plot

link to web-application

Acknowledgments

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

References

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

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

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

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

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

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

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

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

Epstein, J. I., Egevad, L., Amin, M. B., Delahunt, B., Srigley,

J. R., and Humphrey, P. A. (2016). The 2014 international society of urological pathology (isup) consensus conference on gleason grading of prostatic carcinoma. *The American journal of surgical pathology* **40,** 244–252.

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

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

Lieberman, D. A., Rex, D. K., Winawer, S. J., Giardiello, F. M., Johnson, D. A., and Levin, T. R. (2012). Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the us multi-society task force on colorectal cancer. *Gastroenterology* **143,** 844–857.

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

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

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

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

Rizopoulos, D. (2016). The R package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC. *Journal of Statistical Software* **72,** 1–46.

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

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

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

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

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

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

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

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

Supporting Information

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

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

Appendix

Title of appendix

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