Biometrics 00, 1–10
December 0000

DOI: 10.1111/j.1541-0420.2005.00454.x

This is an Example of an Article Title

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: A key word; But another key word; Still another key word; Yet another key word.

1. Introduction

Prostate-specific antigen (PSA) based screening for prostate cancer has led to over-diagnosis of low and very low-grade prostate cancers (Crawford, 2003; Newcomer et al., 1997). Many such cancers remain asymptomatic during a patient's lifetime. Hence, (over-) treatment upon diagnosis for low-grade cancers is nowadays substituted by prostate cancer active surveillance (AS). The goal of AS is to monitor patients continually and advise treatment only upon observing signs of cancer progression. To this end, tumors are evaluated periodically via PSA (ng/mL) blood test; digital rectal examination (DRE) for shape and size of the tumor; and biopsy Gleason grade group (Epstein et al., 2016), a pathology report. Among these, the biopsy Gleason grade group is the strongest indicator of cancer-related outcomes. Consequently, a commonly used trigger for treatment in AS is Gleason upgrade, defined as increase in repeat biopsy Gleason grade group from group 1 to 2 or higher.

Most AS programs are about a decade old. Hence, the midterm pros and cons of AS have become apparent only recently. Among these, a crucial and contentious issue is the use of yearly biopsies for all patients (Loeb et al., 2014). Yearly biopsies help to detect Gleason upgrade timely (maximum delay of one year, Figure **??**). However, biopsies are painful, prone to complications, and recovery can take months (Loeb et al., 2013). Yearly biopsies may be advantageous for patients who progress fast. Although, they also schedule many unnecessary biopsies for slow and/or non-progressing patients (50% proportion in some AS programs). Moreover, patients do not always comply with such schedules (Bokhorst et al., 2015). This may lead to delayed detection of Gleason upgrade and reduce the effectiveness of AS.

Our aim is to balance the number of biopsies (burden) and the time delay in the detection of Gleason upgrade (less is beneficial), better than fixed schedules. For this purpose, we intend to create personalized biopsy schedules that utilize patient-specific accumulated clinical data (age, historical PSA and DRE, and biopsy results). Previous alternatives to yearly biopsies can be divided into three categories. First, heuristic schedules such as biennial biopsies (Inoue et al., 2018). Their drawback is that they ignore the difference in rate of Gleason upgrade between cohorts and/or patients (Web Figure...), and also over the follow-up period within the same cohort and/or patient. Second, personalized biopsy decisions employing partially observable Markov decision processes (POMDP); for prostate cancer (Zhang et al., 2012; Barnett et al., 2018), list for other cancers (Alagoz et al., 2010). Although, the efficacy of POMDP's is limited by the choice of reward function. Choosing rewards is difficult because even for the simplest of the POMDP's, infinite possible reward sets result into the same decision (Vickers and Elkin, 2006). Third, personalized schedules can also be obtained by optimizing a loss function of clinical parameters of interest (Bebu and Lachin, 2017; Rizopoulos et al., 2015), including our previous work (Tomer et al., 2019). However, our method scheduled only one biopsy at a time, and thus ignored the total burden of biopsies in a schedule. In addition, we utilized certain loss functions that inherently took clinically unacceptable high risks (e.g., 50%) of Gleason upgrade.

Our methodology in this work is as follows. First, we obtain a full specification of the joint distribution of the longitudinal PSA and DRE measurements, and time of Gleason upgrade (Section 2). We achieve this using joint models for time-to-event and longitudinal data (Tsiatis and Davidian, 2004; Rizopoulos, 2012). They are inherently personalized because they exploit patient-specific random effects to model all observed data. We fit our model to the data of the world's largest AS cohort, PRIAS, with 7813 patients from 100 medical centers worldwide, employing a common study protocol (Bokhorst et al., 2016). Subsequently, we use the fitted joint model to predict the personalized cumulative-risk of Gleason upgrade for new patients, over their whole follow-up period (Section 3). The predictions utilize their complete longitudinal PSA and DRE values, and previous biopsy re-

sults. We then propose biopsies on all those future follow-up visits where a patient's conditional cumulative-risk of Gleason upgrade is above a certain threshold (e.g., 10% risk). We automate the choice of this threshold by optimizing a loss function of the two relevant features of a biopsy schedule, namely, the number of biopsies, and the expected time delay in detection of Gleason upgrade. The latter is also estimated in a patient-specific manner, for both fixed and personalized schedules. Thus patients and doctors can compare schedules before making a choice.

In Section 4 we briefly discuss the parameter estimates of the fitted joint model, and demonstrate the personalized schedules for real PRIAS patients. To compare the personalized and fixed schedules, we conduct an extensive and realistic simulation study based on a replica of the patients from the PRIAS cohort in Section 5. We discuss the advantages and limitations of our work in Section 6.

2. A Bivariate Joint Model for Time-to-Gleason Upgrade and Longitudinal PSA and DRE

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:

$$R_j(u \mid t, s) = p\Big\{T_j^* \leqslant u \mid T_j^* > t, \mathcal{Y}_{pj}(s), \mathcal{Y}_{dj}(s), \mathcal{D}_n\Big\}, \quad s \leqslant u \leqslant 10. \tag{1}$$

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 1) 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 $L = \{2, 2.5 \ldots 10\}$ 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:

$$B_j^\kappa = \Big\{b_{jk} \epsilon L \mid R_j(b_{jk} \mid b_{jk-1}, s) \geqslant \kappa \wedge (b_{jk} - b_{jk-1} \geqslant 1)\Big\}, \tag{2}$$

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 (1). 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 2 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:

$$\kappa_a = \arg\min_\kappa \sqrt{\left(|B_j^\kappa| - 1\right)^2 + \left\{D_j(B_j^\kappa \mid t, s) - 0\right\}^2}, \quad 0 \leqslant \kappa \leqslant 1, \tag{3}$$

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 (3) 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 3 in a patient-specific manner using personalized cumulative-risk profile estimated in Equation (1). 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 $D_j(B, t, s)$:

$$D_j(B,t,s) = \sum_{k=1}^{|B|} R_j(b_k \mid b_{k-1}, s) \times \Big\{ b_k - b_{k-1} - \int_{b_{k-1}}^{b_k} 1 - R_j(u \mid b_k, b_{k-1}, s)\, du \Big\},$$
$$R_j(u \mid b_k, b_{k-1}, s) = p\Big\{ T_j^* \leqslant u \mid b_{k-1} < T_j^* \leqslant b_k, \mathcal{Y}_{pj}(s), \mathcal{Y}_{dj}(s), \mathcal{D}_n \Big\}, \tag{4}$$

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

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* .

Barnett, C. L., Auffenberg, G. B., Cheng, Z., Yang, F., Wang, J., Wei, J. T., Miller, D. C., Montie, J. E., Mamawala, M., and Denton, B. T. (2018). Optimizing active surveillance strategies to balance the competing goals of early detection of grade progression and minimizing harm from biopsies. *Cancer* **124,** 698–705.

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.

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

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.

Inoue, L. Y., Lin, D. W., Newcomb, L. F., Leonardson, A. S., Ankerst, D., Gulati, R., Carter, H. B., Trock, B. J., Carroll, P. R., Cooperberg, M. R., et al. (2018). Comparative analysis of biopsy upgrading in four prostate cancer active surveillance cohorts. *Annals of internal medicine* **168,** 1–9.

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

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

Newcomer, L. M., Stanford, J. L., Blumenstein, B. A., and Brawer, M. K. (1997). Temporal trends in rates of prostate cancer: declining incidence of advanced stage disease, 1974 to 1994. *The Journal of urology* **158,** 1427–1430.

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., 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.

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.

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.

Zhang, J., Denton, B. T., Balasubramanian, H., Shah, N. D., and Inman, B. A. (2012). Optimization of prostate biopsy referral decisions. *Manufacturing & Service Operations Management* **14,** 529–547.

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`.