

Supplementary Materials for “Should a Biopsy be Conducted on Follow-up Visits in Prostate Cancer Active Surveillance?: A Personalized Biopsy Scheduling Approach”

Author^{1,*}, Author², Author³, Author^{2,4}, and Author¹

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 Medical Statistics and Bioinformatics, Leiden University Medical Center, the Netherlands

*email: a.tomer@erasmusmc.nl

Appendix A Joint Model for Time-to-Event and Longitudinal Data

In this appendix section, we provide a short introduction of the PRIAS dataset, and present the joint modeling framework and the corresponding parameter estimation using the Bayesian approach. The PRIAS dataset consists of 5270 AS patients. For each patient, prostate-specific antigen (PSA) measurements (ng/mL) are scheduled every 3 months for first 2 years and every 6 months thereafter. The digital rectal examination (DRE) measurements which are binary with two levels, namely $DRE > T1c$ and $DRE \leq T1c$ (Schröder et al., 1992), are scheduled every 6 months. Larger values for PSA and/or larger score for DRE, may indicate cancer progression. Biopsies are scheduled at following fixed follow-up times (measured since inclusion in AS): year 1, 4, 7, and 10, and every 5 years thereafter. An annual schedule is prescribed to those patient’s who have a PSA doubling time between 0 and 10 years. PSA doubling time is measured as the inverse of the slope of the regression line through the base two logarithm of the observed PSA values.

Let T_i^* denote the true cancer progression time for the i -th AS patient. Let T_i^R and T_i^{R-1} denote the time of his latest and second latest biopsies, respectively. Since biopsies are conducted periodically, T_i^* cannot be observed directly and it is only known to fall in an interval $l_i < T_i^* \leq r_i$, where $l_i = T_i^{R-1}$, $r_i = T_i^R$ if progression is observed at the latest biopsy, and $l_i = T_i^R$, $r_i = \infty$ if progression is not observed yet. Further let $\mathbf{y}_{1i}, \mathbf{y}_{2i}$ denote the $n_{1i} \times 1$ and $n_{2i} \times 1$ vectors of the DRE and PSA longitudinal measurements, respectively. For a sample of n patients the observed data is denoted by $\mathcal{D}_n = \{l_i, r_i, \mathbf{y}_{1i}, \mathbf{y}_{2i}; i = 1, \dots, n\}$.

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

$$\begin{aligned} \text{logit}[\Pr\{y_{1i}(t) > T1c\}] &= \beta_{01} + b_{01i} + (\beta_{11} + b_{11i})t \\ &+ \beta_{21}(\text{Age}_i - 70) + \beta_{31}(\text{Age}_i - 70)^2 \end{aligned} \quad (1)$$

where, Age_i is the age of the i -th patient at the time of inclusion in AS, $\beta_{\cdot 1}$ are the fixed effect parameters, and $b_{\cdot 1i}$ are the random effect parameters. An example model fit for DRE is shown in Panel C of Figure 1. For the i -th patient, the sub-model for PSA is given by:

$$\begin{aligned} \log_2 \{y_{2i}(t) + 1\} &= \beta_{02} + b_{02i} + \sum_{k=1}^4 (\beta_{k2} + b_{k2i})B_k(t, \mathcal{K}) \\ &+ \beta_{52}(\text{Age}_i - 70) + \beta_{62}(\text{Age}_i - 70)^2 + \varepsilon_{2i}(t), \end{aligned} \quad (2)$$

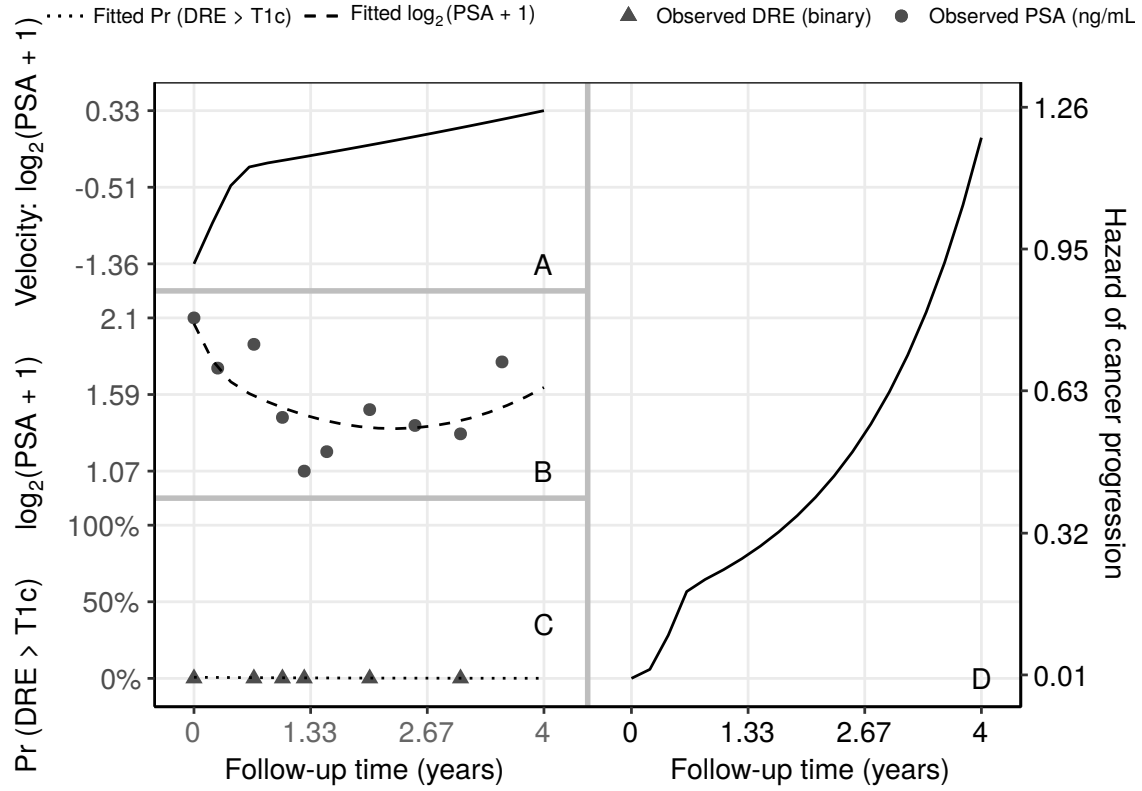


Figure 1: Illustration of the joint model fitted to the PRIAS dataset. Panel A shows the fitted $\log_2\{y_{2i}(t) + 1\}$ velocity (velocity can only be estimated, not observed) over time. Panel B shows the observed and fitted $\log_2\{y_{2i}(t) + 1\}$ levels. Panel C shows the observed DRE scores and the fitted probability of obtaining a DRE score greater than T1c level. The hazard function shown in Panel D, depends on the fitted log odds of having DRE > T1c, and the fitted $\log_2\{y_{2i}(t) + 1\}$ value and velocity.

where, $B_k(t, \mathcal{K})$ denotes the k -th basis function of a B-spline with three internal knots at $\mathcal{K} = \{0.1, 0.7, 4\}$ years, and boundary knots at 0 and 5.42 years (0.95 quantile of the observed follow-up times). The fixed effects parameters are denoted by $\beta_{.2}$ and random effects are denoted by $b_{.2i}$. The error $\varepsilon_{2i}(t)$ is assumed to be t-distributed with three degrees of freedom and scale σ (see Web Appendix C.1), and is independent of the random effects $b_{.2i}$. An example model fit for PSA is shown in Panel B of Figure 1. To account for the association between the DRE and PSA measurements, we link their corresponding random effects. More specifically, the complete vector of random effects $\mathbf{b}_i = (b_{.1i}, b_{.2i})^T$ is assumed to follow a multivariate normal distribution with mean zero and $q \times q$ variance-covariance matrix \mathbf{D} .

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

$$\begin{aligned}
 h_i(t) = h_0(t) \exp \big(& \alpha_{11} \text{logit}[\text{Pr}\{y_{1i}(t) > \text{T1c}\}] \\
 & + \alpha_{21} E[\log_2\{y_{2i}(t) + 1\}] + \alpha_{22} \frac{dE[\log_2\{y_{2i}(t) + 1\}]}{dt} \\
 & + \gamma_1(\text{Age}_i - 70) + \gamma_2(\text{Age}_i - 70)^2 \big), \quad (3)
 \end{aligned}$$

where, γ_1, γ_2 are the coefficients for the effect of Age. The parameter α_{11} models the impact of log odds of having DRE > T1c on the hazard of cancer progression. The impact of PSA on cancer progression is modeled in two ways, namely at any time t the effect of the instantaneous underlying value (dashed line in panel B of Figure 1) of PSA $E[\log_2\{y_{2i}(t) + 1\}]$ is given by α_{21} , and the effect of the instantaneous underlying velocity (panel A in Figure 1) of PSA $dE[\log_2\{y_{2i}(t) + 1\}]/dt$ is given by α_{22} . Lastly, $h_0(t)$ is the baseline hazard at time t , and is modeled flexibly using P-splines

(Eilers and Marx, 1996). 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 (Eilers and Marx, 1996). An example fitted hazard is shown in Panel D of Figure 1.

Appendix A.1 Parameter Estimation

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

$$\begin{aligned} p(\boldsymbol{\theta}, \mathbf{b} \mid \mathcal{D}_n) &\propto \prod_{i=1}^n p(l_i, r_i, \mathbf{y}_{1i}, \mathbf{y}_{2i} \mid \mathbf{b}_i, \boldsymbol{\theta}) p(\mathbf{b}_i \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) \\ &\propto \prod_{i=1}^n p(l_i, r_i \mid \mathbf{b}_i, \boldsymbol{\theta}) p(\mathbf{y}_{1i} \mid \mathbf{b}_i, \boldsymbol{\theta}) p(\mathbf{y}_{2i} \mid \mathbf{b}_i, \boldsymbol{\theta}) p(\mathbf{b}_i \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}), \\ p(\mathbf{b}_i \mid \boldsymbol{\theta}) &= \frac{1}{\sqrt{(2\pi)^q \det(\mathbf{D})}} \exp(\mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i), \end{aligned}$$

where the likelihood contribution of the DRE outcome, conditional on the random effects is:

$$p(\mathbf{y}_{1i} \mid \mathbf{b}_i, \boldsymbol{\theta}) = \prod_{k=1}^{n_{1i}} \frac{\exp \left[-\text{logit} \{ \Pr(y_{1ik} > \text{T1c}) \} I(y_{1ik} = \text{T1c}) \right]}{1 + \exp \left[-\text{logit} \{ \Pr(y_{1ik} > \text{T1c}) \} \right]},$$

where $I(\cdot)$ is an indicator function which takes the value 1 if the k -th repeated DRE score is T1c, and takes the value 0 otherwise. The likelihood contribution of the PSA outcome, conditional on the random effects is:

$$p(\mathbf{y}_{2i} \mid \mathbf{b}_i, \boldsymbol{\theta}) = \frac{1}{(\sqrt{2\pi}\sigma^2)^{n_{2i}}} \exp \left(-\frac{\|\mathbf{y}_{2i} - \mathbf{m}_{2i}\|^2}{\sigma^2} \right),$$

where \mathbf{m}_{2i} is the vector of underlying measurement error free values of $\log_2(\mathbf{y}_i + 1)$. The latter is given by the sum of the terms without the error term $\boldsymbol{\varepsilon}_{2i}$ in the right hand side of the Equation (2). The likelihood contribution of the time to cancer progression outcome is given by:

$$p(l_i, r_i \mid \mathbf{b}_i, \boldsymbol{\theta}) = \exp \left\{ -\int_0^{l_i} h_i(s) ds \right\} - \exp \left\{ -\int_0^{r_i} h_i(s) ds \right\}. \quad (4)$$

The integral in (4) 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 effects $\boldsymbol{\beta}$, and inverse Gamma prior with shape and rate both equal to 0.01 for the parameter σ^2 . 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 q (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.

References

- Eilers, Paul HC and Brian D Marx (1996). “Flexible smoothing with B-splines and penalties”. In: *Statistical Science* 11.2, pp. 89–121.
- Schröder, FH et al. (1992). “The TNM classification of prostate cancer”. In: *The Prostate* 21.S4, pp. 129–138.
- Tsiatis, Anastasios A and Marie Davidian (2004). “Joint modeling of longitudinal and time-to-event data: an overview”. In: *Statistica Sinica* 14.3, pp. 809–834.