Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–11

DOI: 10.1177/0962280216660127
smm.sagepub.com

SAGE

A two-stage model in a Bayesian framework to estimate a survival endpoint in the presence of confounding by indication

Carine Bellera,[1,2,3,4] Cécile Proust-Lima,[2,5,6] Lawrence Joseph,[7] Pierre Richaud,[8] Jeremy Taylor,[9] Howard Sandler,[10] James Hanley[7] and Simone Mathoulin-Pélissier[1,2,3,4]

Abstract

Background: Biomarker series can indicate disease progression and predict clinical endpoints. When a treatment is prescribed depending on the biomarker, confounding by indication might be introduced if the treatment modifies the marker profile and risk of failure.

Objective: Our aim was to highlight the flexibility of a two-stage model fitted within a Bayesian Markov Chain Monte Carlo framework. For this purpose, we monitored the prostate-specific antigens in prostate cancer patients treated with external beam radiation therapy. In the presence of rising prostate-specific antigens after external beam radiation therapy, salvage hormone therapy can be prescribed to reduce both the prostate-specific antigens concentration and the risk of clinical failure, an illustration of confounding by indication. We focused on the assessment of the prognostic value of hormone therapy and prostate-specific antigens trajectory on the risk of failure.

Methods: We used a two-stage model within a Bayesian framework to assess the role of the prostate-specific antigens profile on clinical failure while accounting for a secondary treatment prescribed by indication. We modeled prostate-specific antigens using a hierarchical piecewise linear trajectory with a random changepoint. Residual prostate-specific antigens variability was expressed as a function of prostate-specific antigens concentration. Covariates in the survival model included hormone therapy, baseline characteristics, and individual predictions of the prostate-specific antigens nadir and timing and prostate-specific antigens slopes before and after the nadir as provided by the longitudinal process.

Results: We showed positive associations between an increased prostate-specific antigens nadir, an earlier changepoint and a steeper post-nadir slope with an increased risk of failure. Importantly, we highlighted a significant benefit of hormone therapy, an effect that was not observed when the prostate-specific antigens trajectory was not accounted for in the survival model.

Conclusion: Our modeling strategy was particularly flexible and accounted for multiple complex features of longitudinal and survival data, including the presence of a random changepoint and a time-dependent covariate.

[1]Clinical and Epidemiological Research Unit, Institut Bergonié, Comprehensive Cancer Center, F-33000 Bordeaux, France
[2]INSERM CIC-EC 14.01 (Clinical Epidemiology), Bordeaux, France
[3]INSERM, ISPED, Centre INSERM U1219 Bordeaux Population Health Center, Epicene Team, F-33000 Bordeaux, France
[4]Univ. Bordeaux, ISPED, Centre INSERM U1219 Bordeaux Population Health, Epicene Team, F-33000 Bordeaux, France
[5]INSERM, ISPED, Centre INSERM U1219 Bordeaux Population Health Center, Biostatistics Team, F-33000 Bordeaux, France
[6]Univ. Bordeaux, ISPED, Centre INSERM U1219 Bordeaux Population Health, Biostatistics Team, F-33000 Bordeaux, France
[7]Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Quebec, Canada
[8]Department of Radiotherapy, Comprehensive Cancer Center, Institut Bergonié, France
[9]Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA
[10]Department of Radiation Oncology, Cedars-Sinai Medical Center, Los Angeles, USA

Corresponding author:
Carine Bellera, INSERM U1219, Team EPICENE (Epidemiology of Cancer and Environmental Exposures), F-33000 Bordeaux, France.
Email: c.bellera@bordeaux.unicancer.fr

Keywords
Two-stage model, Bayesian analysis, hierarchical model, changepoint, simulations, MCMC, prostate cancer, PSA

1 Introduction

Medical studies can generate both repeated measurements of a variable marker and event history data, in which times of clinical events are recorded. The observed biomarker series can be an important indicator of disease progression and a predictor of clinical endpoints. Examples include the monitoring of CD4 counts in relationship with time to AIDS or measurements of serum prostate-specific antigens (PSA) for the surveillance of prostate cancer patients after treatment.

Following external beam radiation therapy (EBRT) in patients with prostate cancer, the PSA levels decrease up to a nadir value and then start to rise again. A sharp PSA rise following the initial PSA decline is used as an indicator of treatment failure and clinical failure is expected to follow. A salvage hormonal treatment (HT) can be initiated, which is usually more effective when provided as early as possible, that is, even before clinical failure is observed. As such, PSA is both an intermediate variable and a time-dependent confounder in the relation between HT and the risk of failure, an illustration of confounding by indication.[1,2] This type of bias should be accounted for when estimating the prognostic value of PSA and the potential benefit of HT.

Assessment of prognostic factors of clinical failure is particularly relevant for the clinicians to improve patients' surveillance and management after EBRT. Post-treatment characteristics, such as the PSA nadir and PSA slopes, as well as HT, could help refine prognostic models based on baseline factors only. Aside from the well-known baseline prognostic factors such as baseline PSA, Gleason score and T-stage, studies have also focused on characteristics of the post-EBRT PSA trajectory. Although interassay coefficients of variation tend to be larger at lower PSA values,[3,4] most studies did not account for the PSA variability. Survival models were fitted with the PSA nadir or the post-nadir slope included as exploratory variables, where the PSA nadir was defined as the lowest observed PSA concentration and the post-nadir PSA slope was calculated by fitting least squares regression lines to the observed PSA series.[5–11]

The two-stage modeling approach allows estimation of the regression coefficients in a time-dependent Cox model, while addressing the limitations with the knowledge of the true marker trajectory.[12] In the first stage, the longitudinal process is modeled using a repeated measures component model, such as a random effects model. In the second stage, estimated characteristics of the longitudinal marker trajectory, such as slopes, are included as covariates in a survival model to assess their prognostic value.

Our aim was to highlight the flexibility of a two-stage model fitted within a Bayesian Markov Chain Monte Carlo (MCMC) framework. We applied this model to assess the prognostic value of the PSA profile (level and timing of the nadir; pre- and post-nadir slopes) as well as HT on the risk of clinical failure following EBRT in the presence of confounding by indication. We first present the longitudinal hierarchical PSA model that we developed earlier.[13] This model was particularly flexible since it allowed us to account for the presence of a random changepoint as well as the modeling of the residual variability as a function of the PSA concentration. We next extend the longitudinal model to a two-stage model by using estimated parameters of the longitudinal process as covariates in a Cox proportional hazards model to assess prognostic factors of clinical failure including baseline characteristics, PSA trajectory, and HT.

2 The PSA data sets

We worked with a dataset of 2384 men included in three cohorts: University of Michigan, Ann Arbor, MI, USA (UM)[14]; Radiation Therapy Oncology Group (RTOG 9406)[15,16]; and William Beaumont Hospital, Detroit, MI, USA (WBH).[17] All eligible cases had clinically localized prostate cancer of clinical stage T1 to T4, were node and metastasis negative, and were treated with external beam radiation therapy (RT). Patients with baseline or planned HT were ineligible. All patients were required to have one year of follow-up without clinical recurrence or salvage HT and at least two PSA measurements before the end of follow-up. All PSA measures were collected after RT until the end of follow-up (minimum time to clinical recurrence or lost to follow-up) or initiation of salvage HT was analyzed. Complete PSA, clinical failure, and salvage therapy histories were individually assessed for data consistency. All PSA measures were logarithmically transformed (base 2) to satisfy normality assumptions for inclusion in the statistical models. Prognostic factors included individual baseline characteristics such as initial T-stage (T1–T2 versus T3–T4), Gleason score (scores < 7, $= 7$, > 7), and pre-treatment $\log_2$PSA level (continuous

variable), as well as radiation dose received (defined as the sum of all fractional doses given during therapy). To reduce biases resulting from follow-up variations between cohorts, a common clinical failure criterion was applied as any of the following: distant metastases, nodal recurrence, or any palpable or biopsy-detected local recurrence three years after radiation; any local recurrence within three years of RT if the most previous PSA was > 2 ng/ml; and death from prostate cancer. This definition was intended to allow for the possibility of residual local disease. In patients with more than one clinical recurrence, only the first event was used. Additional details can be found in Proust-Lima et al.[18]

3 Two-stage estimation

In the first stage, we described the post-EBRT $\log_2$PSA profile using a piecewise linear model with a random changepoint, based on PSA data observed before the initiation of HT. In the second stage, estimated characteristics of the longitudinal process as well as baseline factors and HT were included as covariates in a survival model to assess prognostic factors of clinical failure.

3.1 Longitudinal process

Following EBRT and before initiation of HT, the PSA levels decrease to a nadir value and then start to rise again at varying rates between individuals. These rates are reasonably constant for a single man, with close to exponential patterns before and after the nadir.[19,20] The base 2 logarithm transformation leads to a piecewise linear pattern, with the reciprocal of the post-nadir $\log_2$PSA growth rate equivalent to the PSA doubling time (PSA*dt*), a variable of particular interest to clinicians. Similarly, the negative of the reciprocal of the $\log_2$PSA decline rate before the nadir corresponds to the PSA half-life. To illustrate the piecewise linear trajectory typically observed, we have plotted the longitudinal PSA observations for 20 men (Figure 1).

We used a hierarchical piecewise linear model to describe the longitudinal process,[13] an approach particularly flexible to account for the presence of a random changepoint and the wide between-subjects variations in PSA trajectories.[21] In the presence of confounding by indication, the longitudinal process is estimated based on the PSA data available prior to HT. Adjusting for post-HT PSA data would yield an estimate of the HT effect beyond that

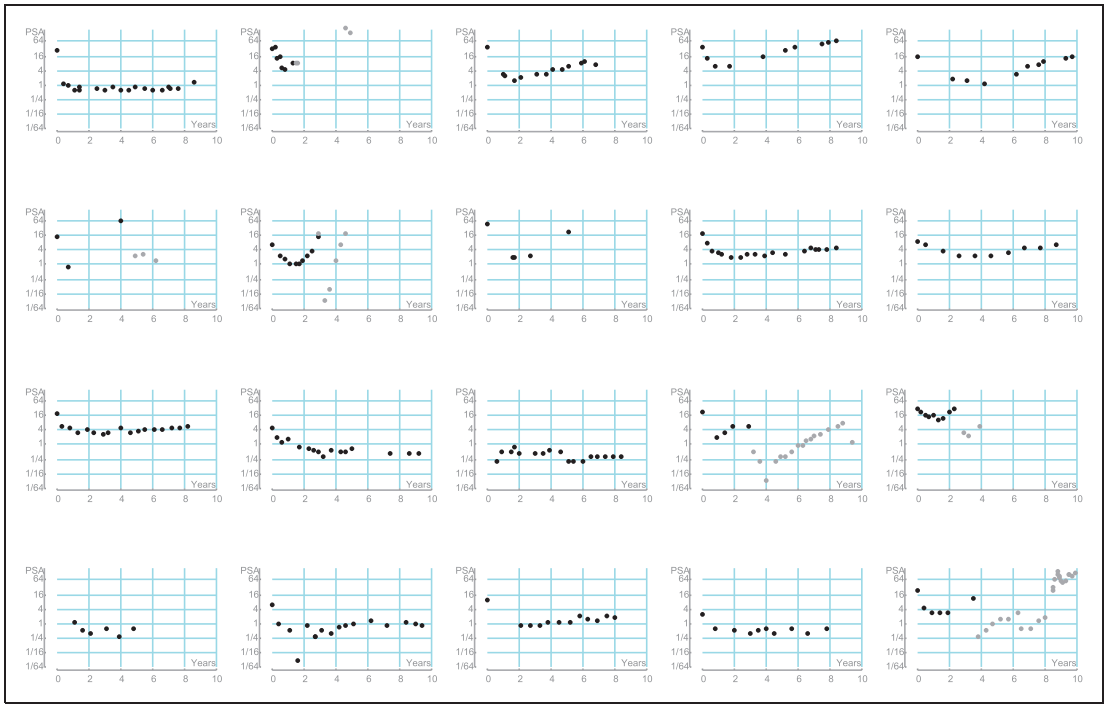

Figure 1. PSA trajectories for 20 men ($\log_2$ scale) following initiation of radiotherapy. PSA: prostate-specific antigen. Observations are shown in black before initiation of hormonotherapy, and in grey thereafter.

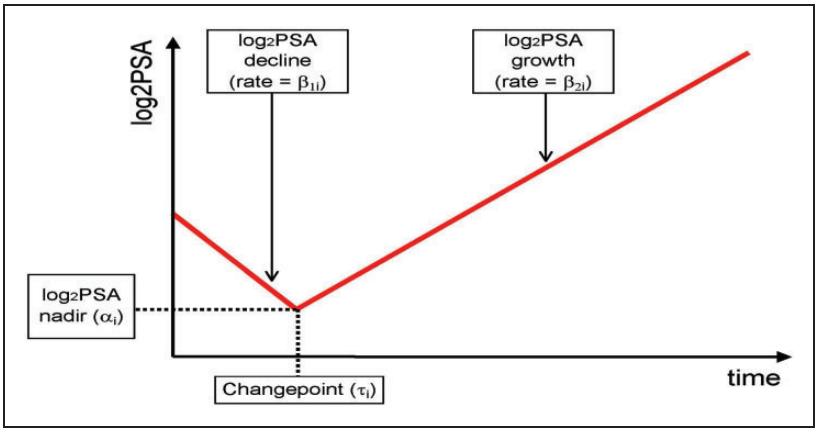

Figure 2. Individual piecewise linear model with four parameters.

due to the longitudinal process, and as such provide biased estimate of HT effect. We denote by α_i^*, β_{1i}^*, β_{2i}^* and τ_i^*, the $\log_2$PSA nadir, the $\log_2$PSA decline rate prior to the PSA nadir (the slope of the first line), the post-nadir $\log_2$PSA growth rate (the slope of the second line), and the changepoint (or location of the nadir in follow-up time), respectively, for the ith individual (Figure 2). For a given cohort, let $log_2\text{PSA}_{ij}$ be the $\log_2$PSA concentration for the jth measurement for the ith man. We assumed that the $log_2 PSA_{ij}$ were normally distributed, with expected value μ_{ij} and variance σ_{ij}^2, as well as an unstructured correlation to describe the series of j observations among each ith individual. The model was as follows

$$\log_2 \text{PSA}_{ij} \sim N(\mu_{ij}, \sigma_{ij}^2), \tag{1}$$

where

$$\mu_{ij} = \begin{cases} \alpha_i^* + \beta_{1i}^*(t_{ij} - \tau_i^*), & t_{ij} < \tau_i^*, \\ \alpha_i^* + \beta_{2i}^*(t_{ij} - \tau_i^*), & t_{ij} \geq \tau_i^* \end{cases} \tag{2}$$

We expressed the individual random parameters as linear functions of k covariates

$$\alpha_i^* = \alpha_i + \gamma_{11} \times x_{11}^{(i)} + \cdots + \gamma_{1k} \times x_{1k}^{(i)} \tag{3}$$

$$\beta_{1i}^* = \beta_{1i} + \gamma_{21} \times x_{21}^{(i)} + \cdots + \gamma_{2k} \times x_{2k}^{(i)} \tag{4}$$

$$\beta_{2i}^* = \beta_{2i} + \gamma_{31} \times x_{31}^{(i)} + \cdots + \gamma_{3k} \times x_{3k}^{(i)} \tag{5}$$

$$\tau_i^* = \tau_i + \gamma_{41} \times x_{41}^{(i)} + \cdots + \gamma_{4k} \times x_{4k}^{(i)} \tag{6}$$

where γ_{mn} for $m = 1$–4 and $n = 1$ to k, are regression parameters, and $x_{mn}^{(i)}$ for $m = 1$–4 and $n = 1$ to k are the k individual baseline covariates for the ith individual. We assumed normal distributions for the individual hierarchical parameters: $\alpha_i \sim N(\mu_\alpha, \sigma_\alpha^2)$, $\beta_{1i} \sim N(\mu_{\beta_1}, \sigma_{\beta_1}^2)$, $\beta_{2i} \sim N(\mu_{\beta_2}, \sigma_{\beta_2}^2)$, $\tau_i \sim N(\mu_\tau, \sigma_\tau^2)$. We assumed that the individual parameters α_i, β_{1i}, β_{2i} and τ_i were *a priori* uncorrelated, both within and between subjects, although they are related through the likelihood function. We assigned noninformative normal priors for the mean (intercept) parameters μ_α, μ_τ, μ_{β_1}, μ_{β_2} and noninformative uniform prior distributions on the standard deviation parameters σ_α, σ_τ, σ_{β_1}, and σ_{β_2}.[22]

Interassay coefficients of variation tend to be larger at lower PSA values.[3,4] We thus expressed the PSA variability σ_{ij}^2 as a function of the PSA concentration. Specifically, we modeled the logarithm of the precision as a linear function of the $\log_2$PSA level,[13] with thus the variance given by

$$\sigma_{ij}^2 = \exp[-(\theta_1 + \theta_2 \mu_{ij})] \tag{7}$$

where μ_{ij} is given by equation (2). Finally, we assigned noninformative normal priors for the regression parameters γ_1 to γ_4 and the variance parameters θ_1 and θ_2.

The longitudinal model was first fit independently to each cohort. Next, after pooling the three cohorts, we fitted the same model and replaced α_i, β_{1i}, β_{2i} and τ_i by cohort-specific parameters to account for heterogeneity.

4 Risk process

Once the longitudinal model was defined, we used a Cox proportional hazards model to asses the association of pre-treatment prognostic factors, longitudinal pattern of PSA, and HT with time to clinical failure. We fitted the following model:

$$\lambda(t) = \lambda_0(t)\exp(\delta \mathbf{Z_i} + \zeta \mathbf{W_i(t)}) \tag{8}$$

where δ and ζ denote regression parameters, Z_i are individual fixed covariates and $W_i(t)$ are time-dependent covariates. In a first simple survival model, individual fixed covariates included Gleason, PSA, T-stage, and dose of radiation received. HT was included as a time-dependent covariate set to zero before HT, and 1 after HT. In subsequent survival models, characteristics of the longitudinal PSA pattern were added as fixed covariates, namely the PSA nadir, the timing of the nadir, the slopes before and after the PSA nadir. These four quantities were estimated by the longitudinal model based on the individual predicted means.

Following a manual forward selection approach, each of the four PSA parameters was added one by one to the simple survival model, thus leading to four additional survival models. We retained the model with the PSA parameter that led to the best model improvement based on the deviance information criterion (DIC). We next tested the addition of each of the three remaining parameters, and so on. For exploratory purpose, the final model obtained through this manual forward selection approach was compared to the model obtained through a backward selection approach.

5 Implementation

Estimation was implemented in WinBUGS, a statistical software package that uses MCMC to generate random samples from the relevant posterior distributions.[23] For each model, we generated three chains with distinct sets of overdispersed initial values. Convergence was monitored using visual inspection of the trace plots and the Gelman-Rubin statistics.[24] We relied on the Raftery and Lewis criterion to confirm that three samples in the MCMC pocess provided the desired accuracy in parameter estimation.[25] Once convergence was reached, chains were pooled to estimate the posterior distributions for each parameter. Point estimates of parameters were assessed using the **median** of the posterior distribution, and 95% credible intervals (CI) were reported based on the 2.5^{th} and 97.5^{th} percentiles of the posterior distribution. We performed sensitivity analysis to investigate the impact of various prior distributions for the parameters of the longitudinal process. Additional details on the implementation process, sensitivity analysis, and strategy for model checking are described in Bellera et al.[13]

6 Results

Characteristics of the patients are summarized in Table 1. A total of 22,356 PSA measurements were available for 2,384 men, leading to a median of 9 measurements per man. Median age was 72 years. With regards to baseline prognostic factors, median PSA concentration was close across the cohorts (7.7 ng/mL). The proportion of patients with T3-T4 stage was 4.6%. The proportion of patients with Gleason 7-10 was 33%. Median radiation dose was 70.2 Grays. Median follow-up time was 5.4 years. Between 6.8% (RTOG) and 16.6% (UMRT) of the patients experienced clinical failure, leading to 315 events overall (13.2%) that were observed around four years post-EBRT. Salvage HT was prescribed in 11.2% of the population.

With regards to the longitudinal process, we first fitted a hierarchical changepoint model unadjusted for baseline characteristics, that is, assuming no covariate in equations (3) to (6). Estimated parameters are presented in Table 2. Point estimates were relatively close across cohorts. Median $\log_2$PSA nadir was -0.30 (0.7 ng/mL on the PSA natural scale) and was reached about one year post RT treatment. The estimated pre-nadir slope was -2, equivalent to a PSA half-life of 0.5 year. The population post-nadir slope was 0.29, equivalent to a PSAdt doubling-time of 3.5 years.

The estimated values for the variance parameters θ_1 and θ_2 were 1.65 and 0.19, respectively. The estimated value for θ_2 was positive, suggesting, as expected, increased PSA variability at lower PSA concentrations. For

Table 1. Description of the study cohorts.

Cohort	UM	RTOG	WBH	All patients
Period of recruitment	1988–2004	1994–2001	1987–2003	1987–2004
Number of patients	501	615	1268	2384
Total number of PSA measurements	4562	6413	11381	22356
Number of PSA per patient	8	11	8	9
	(3; 19)	(4; 17)	(3; 19)	(3; 18)
Age	69.8	69.0	73.0	72.0
(years)	(56.1; 78.7)	(54.0; 77.0)	(61.0; 83.0)	(57.0; 81.0)
Baseline PSA	8.2	7.4	7.7	7.7
	(2.3; 46.4)	(2.6; 19.2)	(2.3; 38.2)	(2.3; 35.5)
Baseline T-stage				
1	163 (32.5%)	348 (56.6%)	431 (34.0%)	942 (39.5%)
2	288 (57.5%)	253 (41.1%)	792 (62.5%)	133 (55.9%)
3	48 (9.6%)	14 (2.3%)	45 (3.5%)	107 (4.5%)
4	2 (0.4%)	0	0	2 (0.1%)
Baseline Gleason score				
2–6	275 (54.9%)	421 (68.5%)	902 (71.1%)	1598 (67.0%)
7	187 (37.3%)	156 (25.4%)	252 (19.9%)	595 (25.0%)
8–10	39 (7.8%)	38 (6.1%)	114 (9.0%)	191 (8.0%)
Total dose	70.4	78	66.6	70.2
(Gray)	(66.0; 77.8)	(70.3; 82.4)	(66; 77.4)	(66.4; 81.2)
Clinical recurrence	83 (16.6%)	42 (6.8%)	190 (15%)	315 (13.2%)
Time to clinical recurrence	2.9	4.2	4.3	3.9
(years)	(1.3; 8.0)	(1.7; 8.1)	(1.4; 9.8)	(1.4; 8.9)
Hormonotherapy	44 (8.8%)	47 (7.6%)	176 (13.9%)	267 (11.2%)
Follow-up time[a]	5.3	5.4	5.3	5.4
(years)	(1.3; 12.0)	(2.6; 9.7)	(1.5; 11.3)	(1.5; 11.0)

Note: Quantitative variables are described using the mean and the 95% reference interval. Qualitative variables are described using counts and proportions. PSA: prostate-specific antigen.
[a]Time to date of clinical recurrence or loss of follow-up.

Table 2. Estimation of the PSA features (level and timing of the nadir, and slopes before and after the nadir) based on the unadjusted hierarchical changepoint model.

	Cohort-specific parameters			Multilevel analysis
	UM	RTOG	WBH	
Log_2PSA nadir[a] ($\hat{\mu}_\alpha$)	−0.25	−0.54	−0.21	−0.30
	(−0.37; −0.13)	(−0.64; −0.43)	(−0.29; −0.14)	(−0.33; −0.28)
Timing of the nadir[b] ($\hat{\mu}_\tau$)	1.12	1.18	0.98	1.06
	(1.06; 1.19)	(1.12; 1.24)	(0.93; 1.02)	(1.03; 1.09)
Pre-nadir slope ($\hat{\mu}_{\beta_1}$)	−1.97	−1.77	−2.13	−2.00
	(−2.09; −1.85)	(−1.87; −1.67)	(−2.22; −2.05)	(−2.06; −1.95)
Post-nadir slope ($\hat{\mu}_{\beta_2}$)	0.45	0.12	0.30	0.29
	(0.37; 0.52)	(0.06; 0.19)	(0.25; 0.35)	(0.27; 0.30)
θ_1 (variance parameter)	–	–	–	1.65
				(1.62; 1.68)
θ_2 (variance parameter)	–	–	–	0.19
				(0.18; 0.21)

Note: Point estimates with 95% credible interval (mean and 95% reference range of the posterior distribution). PSA: prostate-specific antigen.
[a]log_2 scale.
[b]In years.

Table 3. Hazard ratios for prognostic factors of risk of clinical failure following radiotherapy for prostate cancer patients: parameter estimates based on a simple Cox survival model and four two-stage models.

	Cox Model	Two-stage model 1	Two-stage model 2	Two-stage model 3	Two-stage model 4
Baseline T stage T3-T4[a]	**2.19**	**1.61**	1.26	1.23	1.30
	(1.54; 3.15)	**(1.10; 2.31)**	(0.84; 1.83)	(0.82; 1.79)	(0.54; 6.13)
Baseline PSA[b]	**1.31**	**1.19**	0.96	1.01	0.91
	(1.19; 1.43)	**(1.07; 1.31)**	(0.86; 1.07)	(0.90; 1.13)	(0.75; 2.40)
Baseline Gleason 7[c]	**2.14**	**1.69**	**1.80**	**1.79**	1.95
	(1.66; 2.74)	**(1.29; 2.22)**	**(1.38; 2.35)**	**(1.37; 2.33)**	(1.37; 16.9)
Baseline Gleason 8–10[c]	**1.66**	1.10	1.28	1.23	1.22
	(1.16; 2.33)	(0.74; 1.57)	(0.87; 1.84)	(0.84; 1.75)	(0.70; 7.87)
Radiation dose[d]	**0.93**	**0.97**	0.98	0.98	0.98
	(0.90; 0.95)	**(0.94; 1.00)**	(0.95; 1.01)	(0.95; 1.01)	(0.66; 1.02)
Hormotherapy[e]	1.04	**0.31**	**0.23**	**0.21**	0.18
	(0.74; 1.46)	**(0.20; 0.46)**	**(0.16; 0.34)**	**(0.14; 0.30)**	(0.06; 0.30)
Post-nadir slope[f]	–	**4.44**	**4.52**	**4.65**	5.40
		(3.69; 8.23)	**(3.90; 5.26)**	**(4.00; 5.45)**	(4.29; >100)
$\log_2$PSA nadir[f]	–	–	**1.50**	**1.38**	1.66
			(1.35; 1.67)	**(1.22; 1.56)**	(1.38; 7.67)
Timing of the nadir[f]	–	–	–	**0.73**	0.94
				(0.55; 0.92)	(0.65; 16.30)
Slope prior to the nadir[f]	–	–	–	–	0.59
					(0.37; >100)
DIC[g]	5597	5167	5113	5108	5088

Note: Point estimates with 95% credible interval (mean and 95% reference range of the posterior distribution). Significant hazard ratios are reported in bold. PSA: prostate-specific antigen.
[a]Reference is T1–T2 stage.
[b]Continuous and centered, $\log_2$ transformed.
[c]Reference is Gleason 2–6.
[d]Continuous and centered, total cumulative dose in Gray.
[e]Time-varying covariate.
[f]Individual parameter estimated from the longitudinal model.
[g]Deviance information criterion.

illustration, coefficient of variations (i.e. σ/μ) were 40% and 7% for mean $\log_2$PSA concentrations 1 and 4, respectively (2 and 16 ng/mL on the PSA natural scale).

In a second longitudinal changepoint model, we investigated whether expressing each of the random parameter as linear functions of baseline covariates led to model improvement. We therefore included baseline Gleason score, PSA and T-stage, as well as radiation dose as covariates in equations (3) to (6). Although we highlighted some associations, estimations for the variance parameters (σ^2_α, $\sigma^2_{\beta_1}$, $\sigma^2_{\beta_2}$ and σ^2_τ), residual variance parameters (θ_1 and θ_2), and deviance information criterion (DIC) were relatively close (data not shown). Since this modeling strategy did not lead to improved precision, we concluded that the additional complexity induced by the adjusted model compared to the unadjusted model did not result in an improved fit. The unadjusted hierarchical changepoint model appeared sufficient to describe the underlying PSA profile and was thus retained for the subsequent two-stage analysis.

A first Cox PH model included pre-EBRT factors, irradiation dose, and HT. The model revealed associations between baseline covariates and the risk of clinical failure; the multilevel analysis is presented in the first column in Table 3. A T-stage of 3–4, a greater baseline PSA level, and a more severe Gleason grade increased the risk of clinical failure. On the other hand, this risk was reduced with increased radiation dose. When looking at the per-study analysis, similar trends were observed although not systematically significant (data not shown). This can be partly explained by the reduced sample size for some classes of the prognostic factors investigated. HT was associated with an increased risk of clinical recurrence in the RTOG cohort (Hazard ratio [HR]: 4.29; 95% credible interval [CI]: [1.67; 8.61]), but no such association was observed for the other two cohorts, nor in the multilevel analysis.

After inclusion of the baseline covariates and HT in the survival process, each of the four parameters of the longitudinal process was introduced one by one. Introducing the post-nadir slope led to the greatest reduction in

DIC (Table 3, second column). The model suggested a close to five-fold increase in the risk of clinical failure for each unit increase in the $\log_2$ PSA slope. In contrast to the simple survival analysis, the two-stage model revealed the protective role of the implementation of HT (68% risk reduction; HR=0.32; 95% CI=[0.20; 0.46]). Of the other three PSA parameters, the $\log_2$PSA nadir was the next parameter that led to the greatest reduction in the DIC value. The model with the post-nadir slope and the nadir indicated that the $\log_2$PSA nadir was associated with an increased risk of failure (HR=1.50; 95% CI=[1.35; 1.67]) while associations between the post-nadir slope and HT with failure were attenuated (Table 3, third column). Finally, the timing of the PSA nadir also appeared as an independent prognostic factor of clinical failure (Table 3, last column) with still strong associations for the $\log_2$ post-nadir slope, the $\log_2$PSA nadir, and HT. Introducing the slope prior to the nadir led to a small reduction of the DIC; however, time to convergence was significantly increased, and the resulting confidence intervals for some parameters were much larger due to the underlying correlation between parameters of the longitudinal process. As such, the model based on three parameters, the slope after the nadir, the nadir, and its timing, was retained (two-stage model 3 in Table 3). Based on this last model, all other factors held equal, patients with a $\log_2$ PSA slope of 2 (i.e. PSA$dt = 6$ months) had a close to five-fold increased risk of failure compared to patients with a $\log_2$ PSA slope of 1 (i.e. PSA$dt = 2$ years). Similarly, increasing the $log2$ PSA nadir from 1 to 2 (e.g. PSA increase from 2 to 4 ng/mL) increased the risk of failure by 40%. Finally, each reduction of one year in the delay to reaching the nadir was associated with a 27% risk reduction. Finally, variables of the model obtained from a backward selection process were the same as those selected from the forward selection procedure. Computational time for our final model was approximately five days on an INTEL XEON X5650 2 processors computer.

7 Discussion

We have investigated the prognostic value of the post-irradiation PSA trajectory on the risk of clinical failure in patients treated with EBRT for prostate cancer, while accounting for hormonotherapy, a salvage treatment prescribed by indication. Identifying prognostic factors of clinical failure while relying on the underlying PSA pattern can provide valuable information for patients' management. We showed that the PSA nadir, its timing, and the post-nadir PSA slope are strong prognostic factors of clinical failure. These variables, especially the nadir, are expected to happen about one year after EBRT, much earlier than clinical failure, usually four to five years after EBRT. Thus, these parameters should be carefully monitored to guide the surveillance of treated patients, in addition to the usual baseline prognostic factors (Gleason, PSA, stage). Interestingly, we have shown that the prognostic value of these baseline factors was attenuated once the PSA profile was accounted for. Second, compared to a model that ignored the underlying PSA pattern, initiating HT was associated with an important risk reduction.

With regards to the underlying PSA model, we relied on large dataset which allowed us to build an appropriate model with reasonable estimations, including the PSA post-nadir slope, even for those series with fewer data points or shorter follow-up. The log-PSA profile follows a piecewise linear trajectory and thus once PSA starts to increase, the pattern is quite linear and quite deterministically driven. Finally, the PSA process was estimated based on the PSA data available before initiation of HT only. Adjusting for post-HT PSA data would yield an estimate of the HT effect beyond that due to the longitudinal process, and as such provide biased estimate of HT effect.

Previous works that focused on the PSA profile were based on observed PSA measurements and hence ignored the marker variability when estimating the individual nadir and slopes.[5–11] Our hierarchical changepoint model relies on borrowing the strength from the whole population to give more precise estimates for each subject. Our Bayesian hierarchical model was also flexible enough to express the residual PSA variability as a function of the marker level, a strategy consistent with the reported lower precision of the measurement tools at low concentrations[3,4] and confirmed by our estimates of the variance parameters. Piecewise linear modeling is commonly used to model longitudinal processes. However, the novelty of the methodology lies on the fact that (i) the variability of the PSA is modeled as a function of the PSA concentration and (ii) predictions of the longitudinal model are now linked to the survival model as these, including the random changepoint, are used as covariates. Interestingly, this approach could be relevant for additional clinical applications including, for example, the modeling of cognitive decline to predict onset of dementia or modeling tumor size to predict cancer recurrence/progression.

Estimation of the two-stage model is usually addressed through a frequentist approach.[12] Yet, the Bayesian approach based on MCMC sampling is particularly natural and straightforward since it can avoid some complex

approximations required by the frequentist approach. Our approach offers an advantage with regards to the estimation process. Indeed, since the longitudinal model does not have a continuous derivative at the changepoint, estimation within a frequentist framework is limited, while the Bayesian approach is more straightforward. In addition, in the longitudinal process, there exists a non-zero probability of no change (i.e. change on the last point), which can cause the frequentist estimation problem to be non-identifiable. Bayesian methods get around this via the use of prior information which is used to separate out more likely from less likely parameter values, not possible from the likelihood alone. With non-informative priors, maximum likelihood-type estimates can be obtained. The MCMC estimation approach was particularly flexible as it easily accounted for the multiple complex features of longitudinal data including within- and between-series variabilities, complex longitudinal patterns, unbalanced format of the data, non-constant precision of the measurements, as well as inclusion of time-varying covariates in the Cox PH model. The few studies that used a two-stage model in prostate cancer data relied on a frequentist approach.[2,18] Although the longitudinal model and the modeling of the PSA variance were slightly different, hazard ratios for HT were close to our findings.

The two-stage model presents some limitations as it does not correct for event-dependent drop-out, and uncertainty in the estimated parameters is not carried forward to the survival model. As a result, our reported credible intervals might be too narrow. An alternative approach consists in estimating the longitudinal model at each failure time using only measures up to each failure time.[12] It requires fitting as many random-effects models as there are event times. It is however computationally more intensive and presents some limitations. First, the model parameters for prediction change at each time point. Moreover, if the time to event outcome is associated with the longitudinal marker, the risk sets increasingly become a selected set of subjects and the normality assumption for the random effects may not be satisfied for all the models estimated. Sequential stratification and marginal structural methods have been addressed, but comparison works have shown that the two-stage model performs reasonably well and can be used.[2,26] Finally, an alternative, more unified approach is to base estimation and inference on the likelihood from a joint model of both the longitudinal marker data and survival data.[27] Under mild and gross misspecification of the underlying marker profile, the two-stage model tends to overestimate the baseline hazard but underestimates the association parameter while the joint model performs reasonably well with the association parameter being quite robust to such model misspecificiation.[28] We can therefore expect that the estimates provided by our two-stage model, in particular the strong negative association with the risk of clinical failure of an increased nadir, an earlier changepoint, and a shorter PSAdt, as well as a benefit due to HT, could have been underestimated. A joint model, although computationally more intensive, could refine these estimations.

8 Conclusion

We used a two-stage model within a Bayesian framework to assess the role of a marker profile on a time-to-event endpoint while accounting for a treatment prescribed by indication. This approach was particularly flexible to implement and accounted for multiple complex statistical features. Our model highlighted that in patients treated with EBRT for prostate cancer, post-EBRT characteristics are important prognostic factors of clinical failure and accounting for the underlying PSA trend highlighted the protective effect of HT.

Authors' contributions

CB conceived research questions, developed study design and methods, carried out statistical analysis, interpreted results, and drafted the manuscript. CPL, JH, LJ, and JT advised on study design, methods, and statistical analysis, and commented on successive drafts. PR, HS, and SMP advised on study design and commented on successive drafts. All authors read and approved the final manuscript.

Acknowledgement

We would like to thank Ravi Nookala for his help with writing this manuscript.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by a grant from the French National Cancer Institute for research in Epidemiology and Biostatistics (INCA grant 2010-059).

References

1. Robins JM. The control of confounding by intermediate variables. *Stat Med* 1989; **8**: 679–701.
2. Kennedy EH, Taylor JMG, Schaubel D, et al. The effect of salvage therapy on survival in a longitudinal study with confounding by indication. *Stat Med* 2010; **29**: 2569–2580.
3. Eastham J, Riedel E, Scardino P, et al. Variation of serum prostate-specific antigen levels. *J Am Med Assoc* 2003; **289**: 2695–2700.
4. Ballentine Carter H, Pearson J, Metter J, et al. Longitudinal evaluation of prostate-specific antigen levels in men with and without prostate disease. *J Am Med Assoc* 1992; **267**: 2215–2220.
5. Crook JM, Bahadur YA, Robertson SJ, et al. Evaluation of radiation effect, tumor differentiation, and prostate specific antigen staining in sequential prostate biopsies after external beam radiotherapy for patients with prostate carcinoma. *Cancer* 1997; **79**: 81–89.
6. Crook JM, Choan E, Perry GA, et al. Serum prostate-specific antigen profile following radiotherapy for prostate cancer: implications for patterns of failure and definition of cure. *Urology* 1998; **51**: 566–572.
7. Albertsen PC, Hanley JA, Penson DF, et al. Validation of increasing prostate specific antigen as a predictor of prostate cancer death after treatment of localized prostate cancer with surgery or radiation. *J Urol* 2004; **171**: 2221–2225.
8. Valicenti RK, DeSilvio M, Hanks GE, et al. Posttreatment prostatic-specific antigen doubling time as a surrogate endpoint for prostate cancer-specific survival: an analysis of Radiation Therapy Oncology Group Protocol 92-02. *Int J Radiat Oncol Biol Phys* 2006; **66**: 1064–1071.
9. Ray ME, Thames HD, Levy LB, et al. PSA nadir predicts biochemical and distant failures after external beam radiotherapy for prostate cancer: a multi-institutional analysis. *Int J Radiat Oncol Biol Phys* 2006; **64**: 1140–1150.
10. Cheung R, Tucker SL and Kuban DA. First-year PSA kinetics and minima after prostate cancer radiotherapy are predictive of overall survival. *Int J Radiat Oncol Biol Phys* 2006; **66**: 20–24.
11. Hanlon AL, Diratzouian H and Hanks GE. Posttreatment prostate-specific antigen nadir highly predictive of distant failure and death from prostate cancer. *Int J Radiat Oncol Biol Phys* 2002; **53**: 297–303.
12. Tsiatis AA, DeGruttola V and Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error: applications to survival and CD4 counts in patients with AIDS. *J Am Stat Assoc* 1995; **90**: 27–37.
13. Bellera C, Hanley J, Joseph L, et al. Detecting trends in noisy data series: application to biomarker series. *Am J Epidemiol* 2008; **167**: 1130–1139.
14. Taylor JMG, Yu M and Sandler HM. Individualized predictions of disease progression following radiotherapy for prostate cancer. *J Clin Oncol* 2005; **23**: 816–825.
15. Michalski JM, Winter K, Purdy JA, et al. Toxicity after three-dimensional radiotherapy for prostate cancer on RTOG 9406 dose Level V. *Int J Radiat Oncol Biol Phys* 2005; **62**: 706–713.
16. Roach M, Winter K, Michalski JM, et al. Penile bulb dose and impotence after three-dimensional conformal radiotherapy for prostate cancer on RTOG 9406: findings from a prospective, multi-institutional, Phase I/II dose-escalation study. *Int J Radiat Oncol Biol Phys* 2004; **60**: 1351–1356.
17. Kestin LL, Vicini FA, Ziaja EL, et al. Defining biochemical cure for prostate carcinoma patients treated with external beam radiation therapy. *Cancer* 1999; **86**: 1557–1566.
18. Proust-Lima C, Taylor JMG, Williams SG, et al. Determinants of change in prostate-specific antigen over time and its association with recurrence after external beam radiation therapy for prostate cancer in five large cohorts. *Int J Radiat Oncol Biol Phys* 2008; **72**: 782–791.
19. Schmid H, McNeal J and Stamey T. Observations on the doubling time of prostate cancer. The use of serial prostate-specific antigen in patients with untreated disease as a measure of increasing volume. *Cancer* 1993; **71**: 2031–2040.
20. Meek A, Park T, Oberman E, et al. A prospective study of prostate specific antigen levels in patients receiving radiotherapy for localized carcinoma of the prostate. *Int J Radiat Oncol Biol Phys* 1990; **19**: 733–741.
21. Slate E and Cronin K. Changepoint modeling of longitudinal PSA as biomarker for prostate cancer. In: Gatsonis C, Hodges J, Kass R, et al. (eds) *Case studies in Bayesian statistics*. Vol 3. New York: Springer-Verlag New York, Inc, 1999, pp.435–456.
22. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Anal* 2006; **1**: 515–533.
23. Spiegelhalter D, Thomas A, Best N, et al. *WinBUGS: Bayesian inference using Gibbs sampling, version 1.3*. Cambridge, UK: MRC Biostatistics Unit, 2000.
24. Gelman A, Carlin J, Stern H, et al. *Bayesian data analysis*, 2nd ed. Boca Raton, FL: CRC Press, 2004.
25. Raftery A and Lewis S. How many iterations in the Gibbs sampler? In: Bernardo B and Dawid S (eds) *Bayesian statistics* Vol 4. Oxford: Oxford University Press, 1992, pp.763–773.

26. Taylor JMG, Shen J, Kennedy EH, et al. Comparison of methods for estimating the effect of slavage therapy in prostate cancer when treatment is given by indication. *Stat Med* 2014; **33**: 257–274.
27. Wulfsohn MS and Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; **53**: 330–339.
28. Sweeting MJ and Thompson SG. Joint modelling of longitudinal and time-to-event data with application to predicting abnominal aortic aneurysm growth and rupture. *Biometric J* 2011; **53**: 750–763.