Evaluation of treatment efficacy using a Bayesian mixture piecewise linear model of longitudinal biomarkers

Lili Zhao,[a*†] Dai Feng,[b] Brian Neelon[c] and Marc Buyse[d,e]

Prostate-specific antigen (PSA) is a widely used marker in clinical trials for patients with prostate cancer. We develop a mixture model to estimate longitudinal PSA trajectory in response to treatment. The model accommodates subjects responding and not responding to therapy through a mixture of two functions. A responder is described by a piecewise linear function, represented by an intercept, a PSA decline rate, a period of PSA decline, and a PSA rising rate; a nonresponder is described by an increasing linear function with an intercept and a PSA rising rate. Each trajectory is classified as a linear or a piecewise linear function with a certain probability, and the weighted average of these two functions sufficiently characterizes a variety of patterns of PSA trajectories. Furthermore, this mixture structure enables us to derive clinically useful endpoints such as a response rate and time-to-progression, as well as biologically meaningful endpoints such as a cancer cell killing fraction and tumor growth delay. We compare our model with the most commonly used dynamic model in the literature and show its advantages. Finally, we illustrate our approach using data from two multicenter prostate cancer trials. The R code used to produce the analyses reported in this paper is available on request. Copyright © 2015 John Wiley & Sons, Ltd.

Keywords: mixture model; change point; Bayesian hierarchical model; longitudinal data; PSA; tumor growth profile

1. Introduction

Prostate-specific antigen (PSA), a serine protease normally produced in the prostate, is a widely used marker in clinical trials for patients with prostate cancer. Typically, it becomes elevated when the prostatic epithelium undergoes malignant transformation. The longitudinal trajectory for PSA levels is often non-linear in clinical trials. PSA levels are conventionally categorized as complete response, partial response, no change, or progressive disease [1]. For instance, a partial response is determined if the PSA level is decreased by at least 50% from the baseline level, and remained under 50% of the baseline level for at least 28 days. However, using these categorized endpoints often leads to biased estimates [2] and underpowered comparisons [3–5].

In this paper, we propose the use of continuous longitudinal PSA data to evaluate treatment effects. Motivated by a phase III multicenter prostate cancer trial [1], we observe that, in many cases, there is a decline in PSA immediately after therapy followed by a rise in the recurrence of the cancer, and the decline in PSA represents that a subject responded to therapy. There has been extensive literature on fitting such nonlinear PSA trajectories using a piecewise linear model (PWL) with a random change point [6–9]. Similar PWL models have been applied to other studies related to HIV [10], cognitive function [11], AIDS [12], and xenograft experiments [13]. However, we also observe that not all patients have a decline in PSA following therapy. For these patients, the longitudinal PSA data increase until disease progression as a sign of not responding to therapy. Therefore, it is important for a model to accommodate

[a]Department of Biostatistics, University of Michigan, Ann Arbor 48109, MI, U.S.A.
[b]Biometrics Research, Merck Research Lab, Rahway 07033, NJ, U.S.A.
[c]Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, U.S.A.
[d]International Drug Development Institute, Louvain-la-Neuve, Belgium
[e]Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Hasselt University, Diepenbeek, Belgium
*Correspondence to: Lili Zhao, Department of Biostatistics, University of Michigan, Ann Arbor 48109, MI, U.S.A.
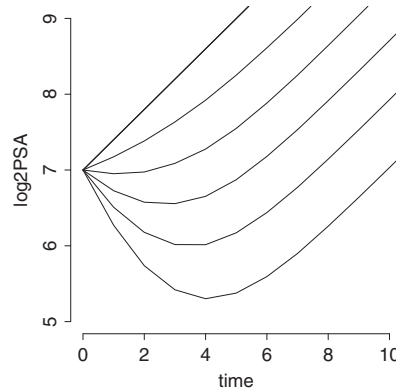†E-mail: zhaolili@umich.edu

Figure 1. Hypothetical nonlinear prostate-specific antigen (PSA) trajectories following treatment. Each curve is an average over 1000 curves simulated from a mixture model with a mixture probability, *p*, ranging from 0 to 1, at a 0.2 interval, a fixed intercept at 7, and fixed normal distributions for the random change point, slope before and after the change point.

both of these trajectories for responders and nonresponders. The most commonly seen mixture model in the literature consists of a mixture of a one-parameter intercept-only function and a three-parameter PWL function [10, 14–16]. These two functions are probably adequate to model longitudinal trajectories in their contexts. For example, a flat trajectory is expected for controls in the ovarian cancer screening study [14]. However, they will not be appropriate for modeling longitudinal efficacy biomarkers following treatment.

To accommodate responders and nonresponders, we consider a model with an additional slope parameter in both the PWL and linear functions. This additional parameter allows an initial decline in PSA among responders or a linear increase among nonresponders. For a responder, the PSA trajectory is described by a PWL function, represented by an intercept, a PSA decline rate, a period of PSA decline, and a PSA rising rate. For a nonresponder, the trajectory is described by an increasing linear function with an intercept and a PSA rising rate. Using Bayesian inference, a subject's trajectory is classified into a linear or a PWL function with a certain probability. This probability estimates how likely it is that the subject responds to therapy. Furthermore, this additional slope parameter offers a greater flexibility for modeling a variety of patterns of nonlinear PSA trajectories as indicated in Figure 1. This figure depicts hypothetical PSA trajectories simulated from our mixture model with a mixture probability ranging from 0 to 1, at 0.2 interval. When the mixture probability is 0, the trajectory is an increasing linear function of time; when the mixture probability is 1, the trajectory is a PWL function with a decline phase. Different mixture probabilities between 0 and 1 result in different patterns of PSA trajectories.

In addition to providing a sufficient fit to nonlinear trajectories, this mixture structure also enables us to derive clinically useful endpoints such as a response rate and time-to-progression, as well as biologically meaningful endpoints, such as a cancer cell killing fraction and tumor growth delay.

The remainder of this article is organized as follows. In Section 2, we develop a Bayesian mixture hierarchical longitudinal model, derive clinically and biologically meaningful endpoints, and address the nonconstant variance issue. In Section 3, we present simulation studies and compare our mixture model with a commonly used dynamic model. Finally, in Section 4, we analyze the prostate cancer trial. We conclude the article with a brief discussion.

2. Mixture model

Let y_{ij} be the $\log_2$ observed PSA data of subject i ($i = 1, \cdots, n$) at time $t_{ij}(j = 1, \cdots, n_i)$. Log transformation is commonly used to obtain linear profiles before and after the nadir (the lowest PSA measurement), and the base 2 logarithms are related to PSA doubling times [8]. The observed PSA data are measured with error, and can be described by $y_{ij} = \mu_{z_i, ij} + \varepsilon_{ij}$, where $\varepsilon_{ij} \sim N(0, \sigma^2)$. An unobserved indicator z_i is included in the mean function to distinguish responders and nonresponders. If $z_i = 1$, subject i is a responder, and the mean function of the $\log_2$ PSA profile is described by a PWL function, which is given by

$$\mu_{1,ij} = a_i + b_{1i}t_{ij}\mathrm{I}(t_{ij} \leqslant \tau_i) + b_{1i}\tau_i\mathrm{I}(t_{ij} > \tau_i)$$
$$+ b_{2i}(t_{ij} - \tau_i)_+ \tag{1}$$

where functions x_+ is defined as $\max(x, 0)$. Here, a_i is the $\log_2$ measurement at baseline, τ_i is the unobserved change point (period of PSA decline), b_{1i} is the rate of decline in PSA, and b_{2i} is the PSA rising rate. If $z_i = 0$, subject i is a nonresponder, and the mean function is reduced to a linear function

$$\mu_{0,ij} = a_i + b_{2i}t_{ij} \tag{2}$$

Subject i has four parameters $(a_i, b_{1i}, \tau_i, b_{2i})$ with $z_i = 1$, and only has two parameters (a_i, b_{2i}) with $z_i = 0$.

2.1. Hierarchical structure

Both mean functions in (1) and (2) include the subject-specific intercept, a_i, which represents the baseline data ($\log_2$ transformed) common to all treatment groups. Hence, we will borrow strength across all subjects (in all treatment groups) for estimation of this parameter. Specifically, for $i = 1, \cdots, n$, we consider

$$a_i | \alpha, \sigma_a^2 \sim N\left(\alpha, \sigma_a^2\right)$$

where α and σ_a^2 are the population-level parameters. Borrowing the most information across all subjects and all treatment groups will gain precision in parameter estimations.

The $\log_2$ PSA decline rate b_{1i} and period of decline τ_i are specific to $z_i = 1$, and it is very likely that they are affected by treatments. Therefore, we only allow pooling information across subjects with $z_i = 1$ in the same treatment group for the estimation of these parameters. Let g_i be the treatment applied to subject i out of $k = 1, \ldots, K$ treatments (hence, $g_i \in \{1, \ldots, K\}$); we consider

$$b_{1i} | \beta_{1g_i}, \sigma_{b_{1g_i}}^2, z_i = 1 \quad \sim \quad N\left(\beta_{1k}, \sigma_{b_{1k}}^2\right)\mathrm{I}(b_{1i} < 0) \tag{3}$$

$$\tau_i | \mu_{\tau g_i}, \sigma_{\tau g_i}^2, z_i = 1 \quad \sim \quad N\left(\mu_{\tau k}, \sigma_{\tau k}^2\right)\mathrm{I}(\tau_i > 0) \tag{4}$$

where $\beta_{1k}, \mu_{\tau k}, \sigma_{b_{1k}}^2$, and $\sigma_{\tau k}^2$ are treatment-level parameters. The restrictions of $b_{1i} < 0$ and $\tau_i > 0$ suggest that the $\log_2$ PSA profile for a responder is characterized by an initial PSA decline followed by a rebound of PSA.

The parameter b_{2i} also appears in both $z_i = 1$ and $z_i = 0$. It represents a $\log_2$ PSA growth rate when $z_i = 0$ and a regrowth rate when $z_i = 1$. To allow comparisons between treatment arms, we consider

$$b_{2i} | \beta_{2g_i}, \sigma_{b_{2g_i}}^2 \sim N\left(\beta_{2k}, \sigma_{b_{2k}}^2\right) \tag{5}$$

Additionally, indicators z_i $(i = 1, \cdots, n)$ are assumed to follow a Bernoulli distribution given by

$$z_i \sim \mathrm{Bern}\left(n_{g_i}, P_{g_i}\right) \tag{6}$$

where n_{g_i} is the number of subjects and P_{g_i} is the response rate in treatment g_i $(g_i \in \{1, \ldots, K\})$.

The hierarchical structures from (3)–(6) allow us to directly make comparisons between treatment groups using treatment-level parameters such as $P_k, \beta_{1k}, \mu_{\tau k} \ \beta_{2k}$.

2.2. Posterior computation

The Markov chain Monte Carlo (MCMC) procedure for estimating the posterior distributions is implemented by repeatedly drawing samples from the full conditional distributions of the parameters. The full conditional distributions for all parameters except τ_i and z_i are straightforward to compute because of the conjugacy, assuming posterior distributions are normal for the mean parameters and gamma for the precision parameters. The slice sampling algorithm developed by Neal [17] is used in sampling the change point τ_i. Updating the class indicator z_i requires an algorithm that can move between different parameter spaces, that is, between a space with two parameters (a_i, b_{2i}) and a space with four parameters $(a_i, b_{1i}, \tau_i, b_{2i})$. We apply the pseudo-prior method of Carlin *et al.* [18] and the reversible-jump procedure

of Green [19] that allow such moving. Details of the calculation are given in Appendix A. With the prior restrictions in (3) and (4), these algorithms can move between $z_i = 0$ and $z_i = 1$ efficiently. When no change point is present in the trajectory, τ_i is not identifiable [20], and in this situation, the algorithm will randomly pick a positive τ_i. With the constraint of $b_{1i} < 0$ and $\tau_i > 0$, the algorithm is able to correctly identify the pattern of $z_i = 0$ because of the absence of a negative pre-nadir slope.

2.3. Endpoints for drug efficacy

An important feature of our mixture model is its capability to derive clinically useful endpoints such as the response rate and time to PSA progression (TTP).

As defined in most clinical protocols with a PSA endpoint, a subject is classified as a responder if the subject experiences a decline in PSA following treatment, although the reduction threshold varies across protocols. In our mixture model, the $\log_2$ PSA profile of subject i is classified as $z_i = 0$ or $z_i = 1$ at each MCMC iteration, and the proportion of times the profile is classified as $z_i = 1$ (corresponds to a PWL function) across all MCMC iterations estimates the probability that subject is a responder. Clinically, this probability may help guide physicians in developing individualized treatments. For a treatment k, its response rate P_k can be calculated using z_i ($\forall i; g_i = k$) by the beta-binomial conjugacy as shown in (6). This P_k is the preferred endpoint of response rate in phase II trials.

Additionally, we derive the TTP endpoint as a X% increase in PSA taking the nadir (smallest value) as reference. We define TTP_k ($k = 1, \cdots, K,$) as

$$\text{TTP}_k = P_k \mu_{\tau k} + \log_2(1 + X\%)/\beta_{2k} \tag{7}$$

At the subject level, TTP is the sum of two time periods: the time from the start of treatment to nadir and the time from nadir to progression. The first time period is the period of PSA decline, τ_i, if $z_i = 1$ and zero otherwise; the second time period, by simple algebra, equals to $\log_2(1 + X\%)/b_{2i}$. At the treatment level, TTP_k is defined in a similar fashion except that subject-level parameters are replaced by treatment-level parameters as shown in (7), in which the averaged treatment responding time is estimated as $\mu_{\tau k}$ weighted by P_k, the proportion of responders in treatment k.

If PSA is a surrogate for tumor size such that a decline (increase) of PSA values indicates regression (growth) of the tumor. Also, if treated tumors can be assumed to regrow at the same rate as untreated tumors (i.e., $\beta_{21} = \cdots = \beta_{2K} = \beta_2$) [21, 22], we can define cancer cell killing fraction (KF), a cell surviving fraction (LSF), and a tumor growth delay (TGD), respectively, as

$$\text{LSF}_k = -P_k \times \mu_{\tau k}(\beta_2 - \beta_{1k}),$$
$$\text{KF}_k = 1 - 2^{\text{LSF}_k},$$
$$\text{TGD}_k = -\text{LSF}_k/\beta_2$$

These clinical and biological endpoints are all defined as a function of our model parameters. Using MCMC draws of model parameters, these endpoints can be easily calculated and compared between treatment groups.

2.4. Nonconstant error variance

Because laboratory assays are known to have reduced precision at lower levels, we are concerned with the assumption that $\text{var}(y_{ij}) = \sigma^2$. To check this assumption, we modeled the $\log_2$ variance as a linear function of the observed y_{ij} [8, 23]. Additionally, to improve model performance, we normalized longitudinal data of subject i by its baseline value, y_{i1}, specifically, $\varepsilon_{ij} \sim N\left(0, \sigma_{ij}^2\right)$, where

$$\log_2\left(\sigma_{ij}^2\right) = \sigma^2 - 2r(y_{ij} - y_{i1})$$

If $r = 0$, data have a variance σ^2, which is independent of the marker value; if $r > 0$, the measurement variance decreases with the increasing marker level, and vice versa. Further details on MCMC sampling are given in Appendix B.

3. Simulation studies

3.1. Performance of our mixture model

We carried out simulations to investigate the performance of our proposed mixture model defined in Section 2 (denoted as the mixture model). We simulated two groups, each containing 40 subjects, with each subject having measurements at baseline, for 2 weeks, monthly for 6 months, and at a 3-month interval until the second year. The true parameter values were chosen to mimic the real data example, given by the following:

- $a_i \sim N(7, 2^2)$ and $\sigma^2 = 0.4^2$.
- Arm I: $P_1 = 0.6$, $b_{1i} \sim N(-0.8, 0.16^2)$, $\tau_i \sim N(3, 1.5^2)$, $b_{2i} \sim N(0.4, 0.1^2)$. Arm II: $P_2 = 0.8$, $b_{1i} \sim N(-1.2, 0.16^2)$, $\tau_i \sim N(5, 1.5^2)$, $b_{2i} \sim N(0.2, 0.1^2)$. In both arms, b_{1i} is truncated below zero and τ_i and b_{2i} are truncated above zero.

In the simulations, data were truncated at a progression point, determined on the basis of a moving average of three consecutive values of PSA. Progression was defined as an increase in PSA equal to, or larger than, 50% above the lowest prior moving average. This increase has to be either the last determination in the patient follow-up, or maintained for at least 28 days [1]. Based on this rule and the previously chosen parameters, the average median number of measurements per subject across 80 subjects is 7 (ranges from 2 to 13).

The models were implemented in **R**. We generated 5000 Gibbs samples with a burn-in of 2000 iterations. N(0, 100) priors were used for the mean parameters except that $\mu_\tau \sim N(4, 100)$. Gamma(0.01, 0.01) priors were used for the precision parameters, and a noninformative prior, Beta(0.5, 0.5), was used for P_k ($k = 1, \cdots, K$). The posterior estimates from the reversible-jump procedure and pseudo-prior approach were very similar; therefore, we only reported the estimation from the reversible-jump procedure. To improve the computation speed, a table lookup method was used in the indexing operation, and vectorized computation was used wherever possible. Table I shows the true values of the parameters (True), estimated mean (Mean), standard deviation (SD), square root of mean square error (SqrMSE), and 95% coverage probability (CP), based on 1000 simulated trials. These results indicate that the mixture model performed well for the estimation of all parameters. In Appendix D, we also present simulation results for data simulated from a different set of parameters, which resemble the real data example, and the model also performed well as shown in Table D.2.

3.2. Performance of the tumor growth inhibition model

In the literature, models to estimate longitudinal efficacy biomarkers following treatment are very limited. Dynamic models are the only models used to estimate longitudinal data (specifically, tumor size data) following treatment. These models assume that a nonlinear tumor growth profile is a result of two latent processes: a cell-growing process and a cell-killing process induced by drug exposure. Because PSA can be considered as a surrogate of tumor size, dynamic models can serve as alternative models for longitudinal PSA data. Among all dynamic models in the literature [24–28], the most commonly used

Table I. Parameter estimation and performance statistics using the mixture model based on 1000 simulated datasets.

Parameter	Arm	True	Mean	SD	SqrtMSE	CP
α		7	6.99	0.28	0.28	89
P	I	0.6	0.56	0.07	0.08	96
	II	0.8	0.79	0.07	0.07	95
β_1	I	−0.8	−0.85	0.10	0.11	94
	II	−1.2	−1.21	0.08	0.08	94
μ_τ	I	3	3.18	0.33	0.38	93
	II	5	5.03	0.28	0.28	94
β_2	I	0.4	0.39	0.03	0.03	92
	II	0.2	0.20	0.03	0.03	94

SD, standard deviation; SqrMSE, square root of mean square error; CP, coverage probability.

model was proposed by Claret *et al.* [26–28]. It is referred to as the tumor growth inhibition (TGI) model. The TGI model is described by the differential equation

$$\frac{dy_i(t)}{dt} = K_{L_i} y_i(t) - K_{D0_i} \exp(-\lambda_i t) y_i(t) \quad y_i(0) = y_{0i}$$

where $y_i(t)$ is the PSA measurement at time t for subject i and $y_i(t)$ is described as a result of two latent processes: a growing process with a rate of K_{L_i} and a killing process with a rate of $K_{D0_i} \exp(-\lambda_i t)$, which decreases exponentially with time (according to λ_i) from an initial rate of K_{D0_i} [26].

This differential equation can be shown (details in Appendix C) to be equivalent to

$$\log(y_i(t)) = \log(y_i(0)) + K_{L_i} t + \frac{K_{D0_i}}{\lambda_i}(\exp(-\lambda_i t) - 1)$$

where lognormal distributions are assumed for K_{L_i}, K_{D0_i}, and λ_i as specified in [26].

It is advisable to assume that the baseline data are also measured with error. We therefore modified the previous model by replacing $\log(y_i(0))$ with a random parameter a_i ($i = 1, \cdots, n$), which can be thought as unobserved true values at baseline. This model is denoted as TGI(a).

Following Claret *et al.* [26], the period of PSA decline (τ_i) can be calculated as

$$\tau_i = \frac{\log\left(K_{D0_i}\right) - \log\left(K_{L_i}\right)}{\lambda_i}$$

It is noteworthy that τ_i is negative when $K_{D0_i} < K_{L_i}$. Therefore, the TGI model provides an estimate of τ_i for all trajectories including the one without a decline phase. In contrast, the mixture model provides an estimate of τ_i only when $z_i = 1$.

We used the same simulation strategy to investigate the performance of the TGI and TGI(a) model. Again, true parameter values were chosen to mimic the real data example, given by the following:

- $a_i \sim N(7, 2^2)$ and $\sigma^2 = 0.4^2$
- Arm I: $\lambda_i \sim \text{LN}(-0.3, 0.5^2)$, $K_{D0_i} \sim \text{LN}(-0.4, 0.5^2)$, $K_{L_i} \sim \text{LN}(-1.2, 0.5^2)$. Arm II: $\lambda_i \sim \text{LN}(-0.5, 0.5^2)$, $K_{D0_i} \sim \text{LN}(0.4, 0.5^2)$, $K_{L_i} \sim \text{LN}(-1.5, 0.5^2)$. LN denotes the lognormal distribution.

The distribution of a_i is the same as in the mixture model. Following the calculation in [26], we estimated the posterior median (i.e., $\exp(a)$ in $\text{LN}(a, b)$) for each parameter. Based on the choices of the previous parameters, the average median number of measurements per subject across 80 subjects is 6 (ranges from 2 to 13). The models were implemented in **WinBUGS 1.4**. N(0, 100) priors were used for the mean parameters, and Gamma(0.01, 0.01) priors were used for the precision parameters. Both models provided accurate estimates for model parameters except λ, and the bias of λ was large especially in the TGI model (see Table D.1 in Appendix D). Because the TGI(a) was better from the simulation study and also coincides with the generated data (i.e, the TGI(a) model considers that baseline data are generated with errors), we compared the mixture model to the TGI(a) model in the next section.

3.3. Comparison of the mixture model and the TGI(a) model

We fitted the mixture model and TGI(a) model to the data generated in Section 3.1 and the data generated in Section 3.2 (under the TGI(a) model), which resulted in four analyses (a combination of two models and two datasets). We used three Bayesian model comparison criteria to compare the mixture model and the TGI(a) model: (1) a modified deviance information criterion (DIC$_3$) [29], (2) a Watanabe–Akaike information criterion (WAIC) [30], and (3) a log-pseudo marginal likelihood (LPML) [31]. DIC$_3$ is preferred in our setting over the standard DIC proposed by Spiegelhalter *et al.* [32] because it correctly reflects the effective number of parameters in mixture models. LPML is a cross-validated leave-one-out measure of a model's ability to predict the data. It is valid for small and large samples and does not suffer from a heuristic justification based on large sample normality. WAIC was proposed recently and can be viewed as an improvement on the standard DIC, and it also approximates Bayesian cross-validation [33]. The best model should have the smallest DIC$_3$ and WAIC, and the largest LPML.

Table II. Model comparisons.

Models	Data from mixture			Data from TGI(a)		
	DIC$_3$	WAIC	LPML	DIC$_3$	WAIC	LPML
Mixture	**710**	**778**	**−440**	754	744	−449
TGI(a)	1095	1162	−630	**690**	**711**	**−422**
% times Mixture is better	100	100	100	30	29	34

DIC$_3$, deviance information criterion; WAIC, Watanabe–Akaike information criterion; LPML, logpseudo marginal likelihood; TGI, tumor growth inhibition.

Not surprisingly, on average TGI(a) model was the best for the data that were simulated from the TGI(a) model, and the mixture model was the best for the data that were simulated from the mixture model, as shown in Table II (the best diagnostic statistics are highlighted). But it appears that the mixture model was more robust than the TGI(a) model (the differences in the diagnostic statistics were between 27 and 64 for data from the mixture model and between 190 and 385 for data from the TGI(a) model). It is also important to note that, for data that are simulated from the TGI(a), the mixture model was also chosen to be the best model 29–34% of the times.

4. Application

We applied the mixture model, and the TGI and TGI(a) models to two multicenter trials for patients with advanced (metastatic) prostate cancer [1]. An experimental retinoic acid metabolism-blocking agent, liarozole, was compared with the antiandrogenic drugs, cyproterone acetate and flutamide. As in previous analyses of these data [1], we combined data from the two trials and make no distinction between the antiandrogenic drugs. Patients in both trials were in relapse after first-line endocrine therapy. We included 485 patients in our analyses, 249 treated with liarozole and 236 treated with antiandrogenic drugs. Assessments of PSA were undertaken before the start of treatment, for 2 weeks, monthly for 6 months, and at a 3-month interval until treatment discontinuation or death. The number of PSA measurements per subject ranges from 2 to 19 (median is 6).

All priors were the same as in the simulation study. Based on the prior knowledge that PSA progression occurred within 6 months and the assumption that change point occurs about 2 months earlier than the progression time, we selected 4 as the mean for the normal prior for μ_τ (i.e., $\mu_\tau \sim N(4, 100)$). These priors were quite vague relative to the likelihood. A sensitivity analysis to investigate the effect of the prior distributions on the parameter estimates was performed. The effect on the final estimates of increasing the variance of the prior distribution was assessed. For example, increasing the variance of the prior normal distribution (e.g., σ_a^2, $\sigma_{b_{1k}}^2$, $\sigma_{b_{2k}}^2$, and σ_τ^2) from 100 to 1000 had a negligible effect on the posterior distribution of their estimates. This indicates that the priors employed were sufficiently vague relative to the likelihood, so that the prior had minimal effect on the posterior estimates.

We generated 30,000 Gibbs samples with a burn-in of 10,000 iterations, and we then used 2000 iterations obtained from every 10th iteration for computing all posterior estimates, including the posterior mean and the highest probability density (HPD) intervals. It took about 6 h for the PSA data to execute on an Intel Xeon 3.10 GHz 4 GB RAM, x64 Linux computer (Santa Clara, CA, USA). The MCMC chains mixed well, and the convergence of the MCMC sampling algorithm was further checked using several diagnostic procedures as recommended by Cowles *et al.* [34].

In the first step, we verified the assumption of a constant variance by calculating a 95% HPD interval for the parameter *r*, as defined in Section 2.4. The interval covered zero, suggesting *r* was not significantly different from zero and a constant variance was assumed in the analysis later in the text.

Table III suggests that both the Mixture and TGI(a) models are reasonable choices for the fit of PSA data. Specifically, the mixture model is preferred for the liarozole arm, and the TGI(a) is preferred for the antiandrogens arm, and the TGI model seems to be the worse choice for both arms.

Posterior estimates from the Mixture and TGI(a) models are displayed in Table IV. This Table shows that both models provided very close estimates for α and σ^2. A 95% HPD interval was calculated for the difference in each parameter between liarozole and antiandrogens (denoted by difference in the table); if the interval does not cover zero for a parameter, the two arms are considered to be significantly different for that parameter (denoted by *). Both models show that Liarozole was more effective. In the mixture model, while the response rate in liarozole was similar to antiandrogens ($\approx 40\%$), patients treated with

Table III. Model comparisons.

Models	Antiandrogens			Liarozole		
	DIC$_3$	LPML	WAIC	DIC$_3$	LPML	WAIC
Mixture	1758	−1125	1998	**1771**	**−1086**	**1950**
TGI(a)	**1697**	**−1070**	**1884**	1842	−1124	2021
TGI	2020	−1192	2179	1965	−1140	2104

DIC$_3$, deviance information criterion; WAIC, Watanabe–Akaike information criterion; LPML, logpseudo marginal likelihood; TGI, tumor growth inhibition.

Table IV. Posterior mean and 95% HPD intervals in prostate cancer trials using the mixture model and TGI(a) model.

Common Parameters			
Mixture		0.46	(0.45,0.48)
TGI(a)	σ	0.46	(0.45,0.48)
Mixture		7.10	(6.91,7.28)
TGI(a)	α	7.04	(6.84,7.21)

Different Parameters				
		Antiandrogens	Liarozole	Difference
Mixture	P	0.40(0.33,0.48)	0.43(0.35,0.50)	(−0.08, 0.13)
	μ_τ	2.02(1.57,2.46)	4.09(3.28,4.96)	(1.12,3.07)*
	β_1	−0.76(−0.97, −0.57)	−0.74(−0.90, −0.59)	(−0.25, 0.27)
	β_2	0.37(0.33,0.42)	0.35(0.32,0.39)	(−0.07, 0.04)
	TTP	2.39(2.15,2.67)	3.41(2.94,3.87)	(0.46,1.53)*
	KF	0.46(0.38,0.55)	0.73(0.63,0.82))	(0.13,0.39)*
	TGD	2.50(1.90,3.18)	5.30(3.95,6.81)	(0.28,4.48)*
TGI(a)	λ	0.97(0.42,1.73)	0.82(0.39,1.41)	(−1.24, 0.79)
	K_{D0}	0.26(0.09,0.42)	0.62(0.40,0.97)	(0.08,0.79)*
	K_L	0.27(0.22,0.32)	0.34(0.29,0.40)	(0.00,0.15)*

HPD, highest probability density; TGI, tumor growth inhibition; TTP, time to progression; TGD, tumor growth delay; KF, killing fraction.

liarozole responded to therapy significantly longer than those treated with antiandrogenic drugs, which led to a significantly prolonged TTP and greatly increased cancer cell killing fraction, as well as a longer tumor growth delay. In the TGI(a) model, the initial drug-killing rate was higher in liarozole, and the tumor growth rate was also marginally higher in liarozole. The estimates using the TGI model were similar to that of the TGI(a) model except that K_L is not significant in the TGI model.

Figure 2 displays the estimated PSA trajectories on the observed data for nine selected subjects. All three models fit the data reasonably well. The mixture model was noticeably better for subject G and slightly worse for subject I. Based on the mixture model, each fitted curve was estimated by a weighted average of a linear function and a PWL function. The weight (mixture probability) is shown on top of each panel, which indicates how likely it is that the subject responds to therapy. For instance, patient A is a responder, patient D is a nonresponder, and patient F is a responder with a 48% probability. It is also important to note that the fitted curve is not limited to a linear or PWL function; for instance, patients A, B, and C had a quadratic fitted curve, and patients F and I had a fitted curve with a slower increasing linear function followed by a rapidly increasing linear function. These nine subjects represent different patterns of PSA trajectories, and they were selected for illustrative purposes. We want to emphasize that the model fit all 485 subjects fairly well (plots are available upon request). It is worth mentioning that profiles with two data points (such as patient H) should be interpreted with caution. If the two data points are far apart in time, it is unknown if there was a PSA decline between the two time points. In this study, there are only a few such subjects, and the two time points were reasonably close; otherwise, these subjects should be excluded before using the mixture model.

Lastly, we assessed the goodness of fit of each model using a quantitative measure based on posterior predictive checks. The predicted values were obtained from the posterior predictive distribution for each
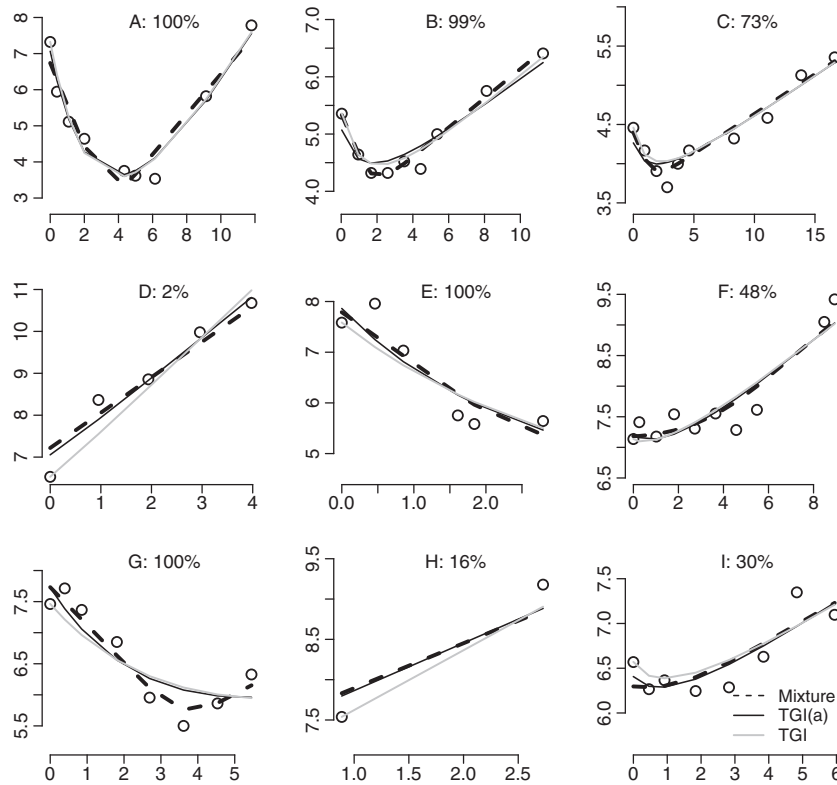
Figure 2. $\log_2$ prostate-specific antigen profiles for nine selected patients in the prostate cancer trial. The horizontal axis is the months at which PSA are measured. Each plot represents one patient denoted by $A, B, C, \cdots, I$. The circles are the observed data. The percentage on the top of each panel represents a mixture probability, that is, the proportion of times that the trajectory is classified as a piecewise linear function. TGI, tumor growth inhibition.
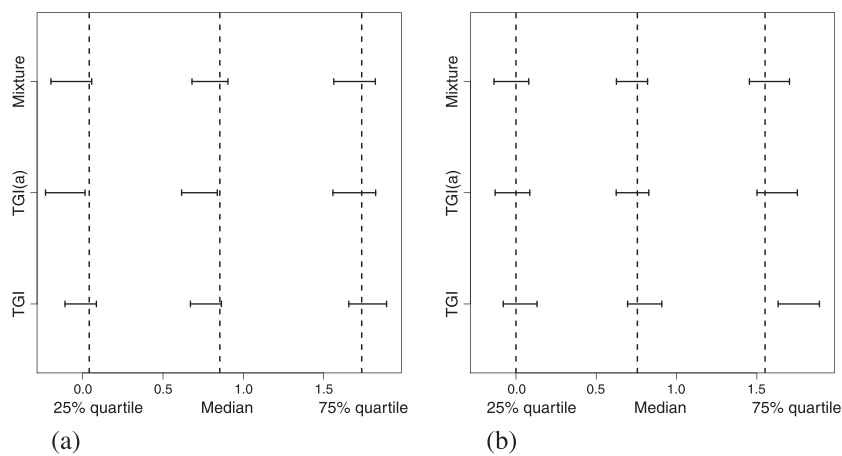

Figure 3. Ninety-five per cent confidence intervals of predicted $\log_2$ prostate-specific antigen reduction at the last measurement relative to baseline (a), or relative to the second measurement (b), compared with observed values (vertical lines).

posterior sample from the MCMC algorithm. Figure 3 suggests that the mixture model fits the PSA data better than the TGI and TGI(a) models. In the TGI(a) model, the observed 25% quartile and median were above the 95% confidence intervals of predictive distribution of the model; in the TGI model, the observed 75% quartile is below the 95% confidence interval. In contrast, the observed median and quartiles (25% and 75% quartiles) for the mixture model were all well within the 95% confidence intervals.

5. Discussion

In this article, we extended existing mixture models to estimate nonlinear PSA trajectories following treatment by introducing a mixture structure consisting of a linear function and a PWL function. These two functions have an important clinical implication: a responder is described by a PWL function and a nonresponder is described by a linear function. The model is able to classify each subject as a responder or nonresponder with a certain probability. Using this probability as a weight, the weighted average of the two trajectories sufficiently characterizes a variety of patterns of PSA trajectories. Furthermore, this mixture structure enables us to derive clinically useful endpoints such as a response rate and TTP, as well as biologically meaningful endpoints such as a cancer cell killing fraction and tumor growth delay.

A major limitation of dynamic models is its inability to classify the PSA trajectories; therefore, it provides an estimate of the PSA decline period even when this period does not exist. We illustrated the advantage of our model through simulation studies and an actual clinical trial data, and conclude that it is an attractive alternative to dynamic models for longitudinal PSA data following treatment. The proposed mixture model is generalizable to many other diseases when longitudinal efficacy biomarker data are available (for example, CD125 in ovarian cancer and circulating tumor cells in various solid tumors). Additionally, this model allows the inclusion of covariates for the estimation of the change point, $\log_2$ PSA value at the change point, and $\log_2$ PSA rates of change before and after the change point. For instance, the mean of τ_i, μ_τ can be modeled as a linear function of the treatment groups and patient-level covariates.

The model can also incorporate data that fall below the limit of detection. Such data are not observed in the prostate cancer trials, but they are likely to occur in most clinical trials. For example, subjects with complete response have data that are below the limit of detection. Such data can be considered left-censored, and their values are less than a threshold. In the posterior sampling, these data can be sampled from a truncated normal distribution as illustrated in [13].

In this article, we used partial follow-up data on PSA for patients who dropped out of the study before PSA progression by assuming uninformative dropout. Appropriate adjustment is needed if the dropout is informative of the PSA trajectory, and the inclusion of some time-dependent covariates may potentially be useful for this adjustment.

Appendix A: Posterior computation

The MCMC procedure for estimating the posterior distributions was implemented by repeatedly drawing samples from the full conditional distributions of the parameters. The full conditional distributions for all parameters except τ_i and z_i are straightforward to compute due to the conjugacy, assuming posterior distributions are normal for the mean parameters and gamma for the precision parameters.

A.1. Sampling τ_i

The likelihood function is only continuous but not differentiable in τ_i. Skates *et al.* [14] and Pauler *et al.* [7] propose the use of an approximation of the likelihood to obtain a proposal density in the Metropolis–Hastings algorithm. We found that the slice sampling algorithm developed by Neal [17] is more efficient in sampling the change point τ_i. In the slice sampling algorithm, the lower limit is zero and the slice width is 0.2.

A.2. Sampling z_i

Updating the class indicator z_i of whether or not there is a treatment responding period is not straightforward, because different values of z_i imply different parameter spaces. If $z_i = 0$, two parameters (a_i and b_{2i}) will need to be updated; if $z_i = 1$, four parameters (a_i, b_{1i}, τ_i, and b_{2i}) will need to be updated. Hence, sampling from the full conditional distribution of z_i requires an algorithm that can move between $z_i = 0$ and $z_i = 1$. We applied two procedures that allow such moving between models, the reversible-jump procedure of Green [19] and the pseudo-prior method of Carlin *et al.* [18] . In both procedures, we took $p(z_i = 1) = p(z_i = 0) = \frac{1}{2}$ as a priori. In the reversible-jump procedure, prior densities of τ_i and b_{1i} were chosen as their proposal densities to propose a move to $z_i = 1$. Specifically, we generated a candidate

$b_{1i}^* \sim N\left(\beta_{1k}, \sigma_{b_{1k}}^2\right) \mathrm{I}(b_{1i} < 0)$ and a candidate $\tau_i^* \sim N\left(\mu_{\tau k}, \sigma_{\tau k}^2\right) \mathrm{I}(\mu_{\tau k} > 0)$, and the move was accepted with probability

$$\min\{\text{likelihood ratio} \times \text{prior ratio} \times \text{proposal ratio} \times \text{Jacobian}, 1\}$$

Because the proposal densities are their prior distributions, prior ratio $\times$ proposal ratio $= 1$, and the Jacobian was also equal to 1 because b_{1i}^* and τ_i^* were generated directly from proposal distributions. Therefore, the acceptance probability is reduced to the minimum of likelihood ratio and 1,

$$\min\left\{ \frac{l\left(\mathbf{y_i}|a_i, b_{1i}^*, \tau_i^*, b_{2i}, \sigma^2\right)}{l\left(\mathbf{y_i}|a_i, b_{2i}, \sigma^2\right)}, 1 \right\}$$

where $l(.)$ is the likelihood of longitudinal data of subject i (i.e., $\mathbf{y_i} = (y_{i1}, \cdots, y_{in_i})$). The acceptance probability for a proposed move from $z_i = 1$ to $z_i = 0$ is the inverse of the previous probability.

The prior densities of τ_i and b_{1i} were also chosen to be their pseudo-prior densities. As the name suggests, a pseudo-prior is not really a prior but only a conveniently chosen linking density, required to define completely the joint model specification. In other words, we augmented the trajectory under $z_i = 0$ by defining a probability distribution for a hypothetical b_{1_i} and τ_i. These variables have no meaningful interpretation under $z_i = 0$, and they are only introduced to match the parameter dimensions. The full conditional distributions of b_{1i} and τ_i are

$$f(b_{1i}, \tau_i|z_i, \mathbf{y_i} \cdots) = \begin{cases} l\left(\mathbf{y_i}|b_{1i}, \tau_i, z_i = 1\right) g(b_{1i}, \tau_i|z_i = 1) & \text{if } z_i = 1; \\ g\left(b_{1i}^*, \tau_i^*|z_i = 0\right) & \text{if } z_i = 0 \end{cases}$$

where $g\left(b_{1i}^*, \tau_i^*|z_i = 0\right) = N\left(\beta_{1k}, \sigma_{b_{1k}}^2\right) \mathrm{I}(b_{1i} \leqslant 0) \times N\left(\tau_0, \sigma_{\tau_0}^2\right) \mathrm{I}(\mu_{\tau k} > 0)$. When $z_i = 1$, we generate b_{1i}^* and τ_i^* from the usual full conditional given data; when $z_i = 0$, we generate b_{1i}^* and τ_i^* from their pseudo-prior densities $g(b_{1i}^*, \tau_i^*|z_i = 0)$. In this case, when the pseudo-priors are their prior densities, we have

$$p(z_i = 1) = \frac{l\left(\mathbf{y_i}|a_i, b_{1i}^*, \tau_i^*, b_{2i}, \sigma^2\right)}{l\left(\mathbf{y_i}|a_i, b_{1i}^*, \tau_i^*, b_{2i}, \sigma^2\right) + l\left(\mathbf{y_i}|a_i, b_{2i}, \sigma^2\right)}$$

Hence, the pseudo-prior method generates z_i as a discrete random variable of 0 or 1, in contrast to the Metropolis step in the reversible-jump procedure.

Appendix B: Sampling the nonconstant variance

In Section 2.4, we modeled the variance as a function of the observed y_{ij}, that is,

$$\varepsilon_{ij} \sim N\left(0, \sigma_{ij}^2\right)$$

The full conditional distributions of σ^2 and r are

$$[1/\sigma^2|\cdots] = \text{Gamma}\left(0.01 + \frac{\sum_{i=1}^{N} n_i}{2}, 0.01 + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{n_i}(v_{ij}/v_{i1})^{2r}(y_{ij} - \mu_{z_i,ij})^2\right)$$

$$f(r|\cdots) \propto \exp\left\{ r\sum_{i=1}^{N}\sum_{j=1}^{n_i}\log(v_{ij}/v_{i1}) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{n_i}(v_{ij}/v_{i1})^{2r}(y_{ij} - \mu_{z_i,ij})^2\right\}$$

We sampled r by means of the adaptive Metropolis–Hastings algorithm [35]. The normal proposal density was centered at the previous value, and the variance in the proposal was 'refined's by using the empirical covariance from an extended burn-in period.

Appendix C: Implementation of the TGI model

$$\frac{dy_i(t)}{dt} = K_{L_i} y_i(t) - K_{D0_i} \exp(-\lambda_i t) y_i(t) \quad \text{and} \quad y_i(0) = y_{0i}$$

$$\frac{dy_i(t)}{dt} \frac{1}{y_i(t)} = K_{L_i} - K_{D0_i} \exp(-\lambda_i t)$$

$$\frac{dlog(y_i(t))}{dt} = K_{L_i} - K_{D0_i} \exp(-\lambda_i t)$$

$$log(y_i(t)) = K_{L_i} t + \frac{K_{D0_i}}{\lambda_i} \exp(-\lambda_i t) + c$$

$$log(y_i(0)) = \frac{K_{D0_i}}{\lambda_i} + c \quad \Rightarrow \quad c = log(y_i(0)) - \frac{K_{D0_i}}{\lambda_i}$$

Therefore,

$$log(y_i(t)) = log(y_i(0)) + K_{L_i} t + \frac{K_{D0_i}}{\lambda_i} (\exp(-\lambda_i t) - 1)$$

In the TGI model, K_{L_i}, K_{D0_i}, and λ_i were assumed to be lognormally distributed.

Appendix D: Additional simulation results

Simulation results using the TGI and TGI(a) models are presented in Table D.1.

Data were simulated from true parameters in the succeeding text, and simulation results from the mixture model are shown in Table D.2.

- Arm I: $P_1 = 0.4$, $b_{1i} \sim N(-0.5, 0.1^2)$, $\tau_i \sim N(4, 1.5^2)$, $b_{2i} \sim N(0.35, 0.1^2)$. Arm II: $P_2 = 0.6$, $b_{1i} \sim N(-0.75, 0.16^2)$, $\tau_i \sim N(2, 1^2)$, $b_{2i} \sim N(0.35, 0.1^2)$.

Table D.1. Parameter estimation and performance statistics using TGI and TGI(a) models based on 1000 simulated datasets.

Parameter	Arm	True	baseline	Mean	SD	SqrMSE	CP
α		7	TGI(a)	6.99	0.22	0.22	95
λ	I	0.74	TGI(a)	1.96	10.99	11.06	93
			TGI	6.43	86.86	87.04	85
	II	0.61	TGI(a)	0.67	0.15	0.16	93
			TGI	0.67	0.15	0.17	91
K_{D0}	I	0.67	TGI(a)	0.99	3.33	3.34	96
			TGI	2.79	32.18	32.25	92
	II	1.49	TGI(a)	1.54	0.22	0.23	94
			TGI	1.54	0.22	0.22	93
K_L	I	0.30	TGI(a)	0.29	0.07	0.07	90
			TGI	0.29	0.06	0.00	90
	II	0.22	TGI(a)	0.21	0.04	0.03	92
			TGI	0.22	0.03	0.00	92

TGI, tumor growth inhibition; SD, standard deviation; SqrMSE, square root of mean square error; CP, coverage probability.

Table D.2. Parameter estimation and performance statistics using the mixture model based on 1000 simulated datasets.

Parameter	Arm	True	Mean	SD	SqrtMSE	CP
α		7	6.97	0.31	0.31	85
P	I	0.4	0.42	0.06	0.07	97
	II	0.6	0.55	0.07	0.08	97
β_1	I	−0.5	−0.49	0.05	0.05	98
	II	−0.75	−0.75	0.09	0.09	97
μ_τ	I	4	3.96	0.45	0.46	95
	II	2	2.21	0.24	0.32	92
β_2	I	0.35	0.35	0.03	0.03	90
	II	0.35	0.34	0.03	0.03	90

Acknowledgements

References

1. Buyse M, Vangeneugden T, Bijnens L, Geys H, Renard D, Burzykowski T, Molenberghs G. Validation of biomarkers and surrogates for clinical endpoints, chap. In *Biomarkers in Clinical Drug Development*, Bloom JC, Dean RA (eds). Marcel Dekker: New York, 2003; 149–168.
2. Panageas KS, Ben-Porat L, Dickler MN, Chapman PB, Schrag D. When you look matters: the effect of assessment schedule on progression-free survival. *Journal of the National Cancer Institute* 2007; **99**:428–432.
3. Karrison T, Maitland M, Stadler W, Ratain M. Design of phase II cancer trials using a continuous endpoint of change in tumor size: application to a study of sorafenib and erlotinib in non small-cell lung cancer. *Journal of the National Cancer Institute* 2007; **99**:1455–1461.
4. Adjei AA, Christian M, Ivy P. Novel designs and end points for phase II clinical trials. *Clinical Cancer Research* 2009; **15**:1866–1872.
5. Wason J, Mander A, Eisen T. Reducing sample sizes in two-stage phase II cancer trials by using continuous tumour shrinkage end-points. *European Journal of Cancer* 2011; **47**:983–999.
6. Slate EH, Clark LC. Using PSA to detect prostate cancer onset: an application of Bayesian retrospective and prospective changepoint identification. *Journal of Educational and Behavioral Statistics* 1999; **26**:443–468.
7. Pauler DK, Finkelstein DM. Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes. *Statistics in Medicine* 2002; **21**:3897–3911.
8. Bellera CA, Hanley JA, Joseph L, Albertsen PC. Hierarchical changepoint models for biochemical markers illustrated by tracking postradiotherapy prostate-specific antigen series in men with prostate cancer. *Annals of Epidemiology* 2008; **18**:270–282.
9. Bellera CA, Hanley JA, Joseph L, Albertsen PC. A statistical evaluation of rules for biomedical failure after radiotherapy in men treated for prostate cancer. *International Journal of Radiation Oncology Biology Physics* 2009; **75**:1357–1363.
10. Pauler DK, Laird NM. A mixture model for longitudinal data with application to assessment of noncompliance. *Journal of the Royal Statistical Society: Series A* 2000; **56**:464–472.
11. Hall CB, Ying J, Kuo L, Lipton RB. Bayesian and profile likelihood change point methods for modeling cognitive function over time. *Computational Statistics and Data Analysis* 2003; **42**:91–109.
12. Kiuchi A, Hartigan J, Holford T, Rubinstein P, Stevens C. Change points in the series of T4 counts prior to AIDS. *Biometrics* 1995; **51**:236–248.
13. Zhao L, Morgan MA, Parsels LA, Maybaum J, Lawrence TS, Normolle D. Bayesian hierarchical changepoint methods in modeling the tumor growth profiles in xenograft experiments. *Clinical Cancer Research* 2010; **17**:1–7.
14. Skates SJ, Pauler DK, Jacobs IJ. Screening based on the risk of cancer calculation from Bayesian hierarchical changepoint and mixture models of longitudinal markers. *Journal of the American Statistical Association* 2001; **96**:429–439.
15. Garre FG, Zwinderman AH, Geskus RB, Sijpkens YWJ. A joint latent class changepoint model to improve the prediction of time to graft failure. *Journal of the Royal Statistical Society: Series A* 2008; **171**:299–308.
16. Zhao L, Banerjee M. Bayesian piecewise mixture model for racial disparity in prostate cancer. *Computational Statistics and Data Analysis* 2012; **56**:362–369.
17. Neal RM. Slice sampling. *The Annals of Statistics* 2003; **31**:705–767.
18. Carlin BP, Gelfand AE, Smith AFM. Hierarchical Bayesian analysis of changepoint problems. *Journal of the Royal Statistical Society. Series C-Applied Statistics* 1992; **41**:389–405.
19. Green PJ. Reversible-jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 1995; **82**:711–732.
20. Andrews DWK. Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica* 2001; **69**: 683–773.

21. Demidenko E. Three endpoints of in vivo tumour radiobiology and their statistical estimation. *International Journal of Radiation Biology* 2010; **86**:164–173.

22. Wu J. Assessment of antitumor activity for tumor xenograft studies using exponential growth models. *Journal of Biopharmaceutical Statistics* 2011; **21**:472–483.

23. Zangeneh SZ. Model-based methods for robust finite population inference in the presence of external information. *Ph.D. Thesis*, University of Michigan Ann Arbor, 2012.

24. Wang Y, Sung C, Dartois C, Ramchandani R, Booth BP, Rock E, Gobburu J. Elucidation of relationship between tumor size and survival in non-small-cell lung cancer patients can aid early decision making in clinical drug development. *Clinical Pharmacology & Therapeutics* 2009; **86**:167–174.

25. Stein W, Gulley J, Schlom J, Madan R, Dahut W, Figg W, Ning YM, Arlen P, Price D, Bates S, Fojo T. Tumor regression and growth rates determined in five intramural NCI prostate cancer trials: the growth rate constant as an indicator of therapeutic efficacy. *Clinical Cancer Research* 2011; **17**:907–917.

26. Claret L, Girard P, Hoff PM, Cutsem EV, Zuideveld KP, Van Cutsem E, Jorga K, Fagerberg J, Bruno R. Model-based prediction of phase III overall survival in colorectal cancer on the basis of phase II tumor dynamics. *Journal of Clinical Oncology* 2009; **66**:4103–4108.

27. Claret L, Lu JF, Sun YN, Bruno R. Development of a modeling framework to simulate efficacy endpoints for motesanib in patients with thyroid cancer. *Clinical Pharmacology & Therapeutics* 2010; **66**:1141–1149.

28. Claret L, Gupta M, Han K, Joshi A, Sarapa N, He J, Powell B, Bruno R. Evaluation of tumor-size response metrics to predict overall survival in Western and Chinese patients with first-line metastatic colorectal cancer. *Journal of Clinical Oncology* 2013; **31**:2110–2114.

29. Celeux G, Forbes F, Robert C, Titterington D. Deviance information criteria for missing data models. *Bayesian Analysis* 2006; **1**:651–674.

30. Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* 2010; **11**:3571–3594.

31. Ibrahim JG, Chen MH, Sinha D. *Bayesian Survival Analysis*. Springer: New York, 2001.

32. Spiegelhalter DJ, Best NG, Carlin B, van de Linde A. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society* 2002; **64**:583–639.

33. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis,* 3rd edition. Chapman & Hall/CRC: Boca Raton, 2013.

34. Cowles MK, Carlin BP. Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association* 1996; **91**:883–904.

35. Haario H, Saksman S, Tamminen J. An adaptive metropolis algorithm. *Bernoulli* 2001; **7**:223–242.