ELSEVIER

Hierarchical Changepoint Models for Biochemical Markers Illustrated by Tracking Postradiotherapy Prostate-Specific Antigen Series in Men With Prostate Cancer

CARINE A. BELLERA, PHD, JAMES A. HANLEY, PHD, LAWRENCE JOSEPH, PHD, AND PETER C. ALBERTSEN, MD

PURPOSE: Biomarkers provide valuable information when detecting disease onset or monitoring disease progression; examples include bone mineral density (for osteoporosis), cholesterol (for coronary artery diseases), or prostate-specific antigens (PSA, for prostate cancer). Characteristics of markers series can then be used as prognostic factors of disease progression, such as the postradiotherapy PSA doubling time in men treated for prostate cancer. The statistical analysis of such data has to incorporate the within and between-series variabilities, the complex patterns of the series over time, the unbalanced format of the data, and the possibly nonconstant precision of the measurements.
METHODS: We base our analysis on a population-based cohort of 470 men treated with radiotherapy for prostate cancer; after treatment, the $\log_2$PSA concentrations follow a piecewise-linear pattern. We illustrate the flexibility of Bayesian hierarchical changepoint models by estimating the individual and population postradiotherapy $\log_2$PSA profiles; parameters such as the PSA nadir and the PSA doubling time were estimated, and their associations with baseline patient characteristics were investigated. The residual PSA variability was modeled as a function of the PSA concentration. For comparison purposes, two alternative models were briefly considered.
RESULTS: Precise estimates of all parameters of the PSA trajectory are provided at both the individual and population levels. Estimates suggest greater PSA variability at lower PSA concentrations, as well as an association between shorter PSAdts and greater baseline PSA levels, higher Gleason scores, and older age.
CONCLUSIONS: The use of Bayesian hierarchical changepoint models accommodates multiple complex features of longitudinal data, permits realistic modeling of the variability as a function of the marker concentration, and provides precise estimates of all clinically important parameters. This type of model should be applicable to the study of marker series in other diseases.
Ann Epidemiol 2008;18:270–282.

KEY WORDS: Bayesian Models, Hierarchical Changepoint Models, Longitudinal Data, Biomarkers, PSA, Prostate Cancer, Radiotherapy.

INTRODUCTION

A biomarker is a laboratory measure of a biological process that can act as an indicator of a current or future disease state. Examples include bone mineral density (for osteoporosis), cholesterol (for coronary artery diseases), CD4 T-cells count (for HIV/AIDS), or prostate-specific antigens (PSA, for prostate cancer). Biomarkers are extensively used to track patient status over time, and the analysis of long series can provide valuable information when estimating the time of disease onset or monitoring disease or treatment outcome.

The analysis of biomarker series requires special care because it has to accommodate multiple characteristics of longitudinal data. First, repeated observations usually are available on each individual and, thus, are likely to be correlated. Second, measurement errors and short-term biological variations (because of natural biological fluctuations, errors in laboratory procedures, or imprecise measurement tools) are important sources of within-series variability that need to be accounted for in the model. Third, marker data may not have been obtained in the context of a controlled study, so that the length of follow-up and the frequency of the measurements can vary considerably from subject to subject. Finally, the biomarker profile may exhibit

From the Department of Clinical Epidemiology and Clinical Research, Institut Bergonié, Regional Comprehensive Cancer Center, Bordeaux, France (C.B.); Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Quebec, Canada (C.B., J.H., L.J.); and Division of Urology, University of Connecticut Health Center, Farmington, Connecticut.

Address correspondence to Carine A. Bellera, Unité de Recherche et d'Epidémiologie Cliniques, Institut Bergonié, Centre de Lutte Contre le Cancer de Bordeaux et du Sud-Ouest, 229 Cours de l'Argonne, 33076 Bordeaux, France. Tel.: +33 (0)5 56 33 04 95; fax: +33 (0)5 56 33 78 01. E-mail: bellera@bergonie.org.

This work was supported by a doctoral training award from Le Fonds de la Recherche en Santé du Québec (C.B.), a studentship from the McGill University Health Center (C.B.), as well as grants from the Natural Sciences and Engineering Research Council of Canada (J.H.) and Le Fonds Québécois de la Recherche sur la Nature et les Technologies (J.H.). LJ is supported by a Senior Scientist Award from the FRSQ.

Received May 12, 2007; accepted October 5, 2007.

360 Park Avenue South, New York, NY 10010

1047-2797/08/$–see front matter
doi:10.1016/j.annepidem.2007.10.006

Selected Abbreviations and Acronyms

PSA = prostate-specific antigen
BHC = Bayesian hierarchical changepoint

a sudden change over time because of treatment, disease onset, or recurrence. For example, in patients treated for infection with the human immunodeficiency virus, a rapid decline in the CD4 T cells number is observed immediately after seroconversion, followed by a less-rapid decline after some unknown time point. Similar changes are observed, for example, for estrogen concentration before compared with after menopause, PSA concentration before compared with after prostate cancer onset or before compared with after treatment. In the presence of such abrupt changes, several parameters are of interest: the time of change, called the changepoint, as well as summaries of the overall patterns in the data, for example, the slopes of the biomarker trajectories before and after the changepoint.

Given the aforementioned characteristics, a Bayesian hierarchical changepoint (BHC) model appears particularly suited to the analysis of markers series data, including for example, postradiotherapy PSA concentration. If radiotherapy is successful, PSA levels reach a nadir, and remain low or possibly rise very slowly. A sustained steeper increase is usually an indication of treatment failure. The interest is in the estimation of both individual and group level parameters: the PSA nadir, its timing or changepoint, the PSA decline rate before the nadir, or equivalently the PSA half-life, and the subsequent PSA growth rate, or equivalently the PSA doubling time, as well as the variability of these parameters.

Bayesian methods have been developed to estimate the parameters of changepoint models (1,2); they have been successfully applied to CD4 cell counts to predict the timing of HIV viral rebound following treatment (3,4) and to longitudinal PSA series to predict cancer onset (5–7) or recurrence (8,9). One advantage of the Bayesian approach is its ability to provide a direct estimation of the probability that the changepoint and, thus, that disease onset or recurrence, has occurred.

Previous studies relying on a BHC model have considered the residual variability of the observations to be constant within the pre- and post-changepoint phases (10), but the model could easily handle another level of complexity by allowing the variance to be a function of the level of the marker. This feature is added here because, in general, laboratory assays can have a precision (i.e., the reciprocal of the variance) that is dependent on the marker's concentration. For example, this phenomenon has been reported in the context of PSA data (11,12).

Our objective is to emphasize the flexibility and accuracy of BHC models, illustrated by estimating postradiotherapy PSA series. We also compare our BHC model to other possible models. This work proceeds as follows. We first introduce our postradiotherapy PSA data set by presenting a population-based cohort of men treated for prostate cancer. We then present our BHC model and provide details with respect to previous distributions, likelihood function, and practical implementation. Next, we investigate possible associations between the individual estimates provided by the BHC model, such as the PSA doubling time and the PSA nadir, and baseline characteristics, including age at diagnosis of the cancer, pretreatment PSA concentration, and pretreatment Gleason score. Finally, to emphasize the advantages of the BHC model, we compare it with two other commonly used models, including a "naïve" piecewise model that is used to estimate each PSA profile independently, and then pools the estimates using a simple average. We also model the PSA data series over time as a polynomial function.

MATERIALS AND METHODS

Patients

To illustrate our approach, we used a dataset from a study aimed at linking increasing PSA profiles after treatment (surgery or radiotherapy) of prostate cancer to long-term outcomes (13). The data were assembled retrospectively, on a population-based cohort that was identified by the Connecticut Tumor Registry. The men were aged 75 years or younger and were residents of Connecticut when diagnosed with localized cancer between 1990 and 1992. Men who were known to have metastatic disease were excluded, as well as men with an initial PSA greater than 50 nanograms per millimeter (ng/mL), because this population has a very high probability of having systemic (extra prostatic or metastatic) disease. PSA values were recorded from the ambulatory records located primarily in urologists' offices but also from ambulatory records located in the offices of radiation oncologists, medical oncologists, the Connecticut Tumor Registry, and inpatient records. More details are available from Albertsen et al. (13). We based our analysis on men treated with radiotherapy and required each PSA series to have at least a baseline PSA measurement, as well as two subsequent PSA measurements. In some cases, men can receive a subsequent treatment, usually in the form of hormones. For the purposes of this study, we excluded any PSA measurements taken after hormonal therapy.

PSA Series and Notation

After radiotherapy, the PSA levels decrease and then start to increase again at variable rates across individuals, although rates are reasonably constant within-men, with close to exponential patterns before and after the nadir (14,15). For this reason, we applied a base 2 logarithmic

AEP Vol. 18, No. 4
April 2008: 270–282

transformation to the PSA observations. This transformation has several advantages. First, it allows one to obtain piecewise linear patterns within individuals. Second, the series tend to be smoother than when plotted on the original PSA scale. Finally, the post-nadir $\log_2$PSA growth rate is equivalent to the number of PSA doublings per year (i.e., the number of times the PSA level doubles during a 1-year period), and its reciprocal corresponds to the PSA doubling time, a variable of particular interest to clinicians. Similarly, the $\log_2$PSA decline rate before the nadir is equivalent to the number of PSA halvings per year (i.e., the number of times the PSA level is halved during a 1-year period), and its reciprocal corresponds to the PSA half-life.

Figure 1 illustrates a prototypic $\log_2$PSA profile plotted over time for a given man i; the time axis starts at the initiation of treatment. We denote by α_i, β_{1i}, β_{2i}, and τ_i, the $\log_2$-PSA nadir, the $\log_2$PSA decline rate before 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) respectively. We are interested in the estimation of these four parameters, their variability in the population, as well as the residual PSA variability.

On average, men with a secondary treatment reach their PSA nadir much sooner, with a steeper postnadir PSA growth rate, than men without such treatment. For this reason, we split the men into two subgroups according to whether or not they received a subsequent hormonal treatment. This division allowed us to obtain two relatively homogeneous subgroups, simplifying the fitting of our model. As not only the changepoint, but other parameters as well may differ between these groups, each subgroup was fitted independently.

A Bayesian Hierarchical Changepoint Model

We used a changepoint model with three hierarchical levels to account for the presence of a random changepoint, as well as the wide between-subjects variations in PSA trajectories

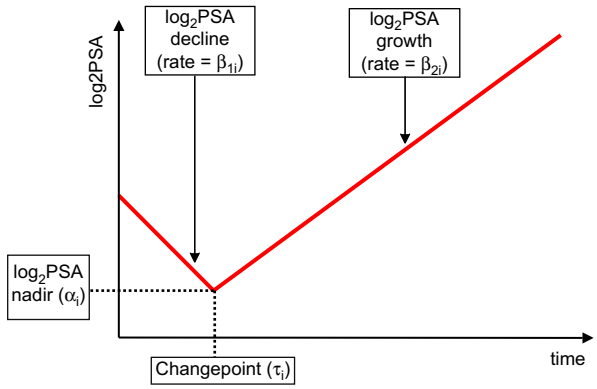

FIGURE 1. Individual piecewise linear model, with the four individual parameters.

(16). At the first level, each individual $\log_2$PSA profile was modeled as in Figure 1. Let $\log_2 PSA_{ij}$ be the PSA concentration on the $\log_2$ scale for the j^{th} measurement for the i^{th} man. We assumed that the $\log_2 PSA_{ij}$ were normally distributed, with expected value μ_{ij}, and variance σ^2_{ij}:

$$\log_2 PSA_{ij} \sim N\left(\mu_{ij}, \sigma^2_{ij}\right). \quad (1)$$

The expected $\log_2 PSA$ value, μ_{ij}, was related to the timing of the measurement, t_{ij}, through linear regression functions before and after the unknown changepoint τ_i:

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

where α_i, β_{1i}, and β_{2i}, correspond respectively to the $\log_2$-PSA nadir, the $\log_2$PSA decline rate prior to the PSA nadir, and the post-nadir $\log_2$PSA growth rate for the i^{th} man.

At the second hierarchical level, we selected noninformative prior distributions for the individual parameters α_i, β_{1i}, β_{2i}, and τ_i. We assumed that the individual parameters were a priori uncorrelated, both within and between subjects, although they are related through the likelihood function (given in the appendix). We selected normal distributions for the two slopes and the nadir, i.e., $\alpha_i \sim N(\mu_\alpha, \sigma^2_\alpha)$, $\beta_{1i} \sim N(\mu_{\beta 1}, \sigma^2_{\beta 1})$, and $\beta_{2i} \sim N(\mu_{\beta 2}, \sigma^2_{\beta 2})$. This choice was motivated by the histograms of these parameters provided the simple analysis (see section "Alternative Approaches"). The individual parameters are thus viewed as a random sample from a common distribution. For example, consider the individual slopes before the nadir. During this initial phase, the individual PSA concentrations decreased with different rates. The hierarchical model assumes that the individual slopes β_{1i} are randomly selected from a normal distribution, $\beta_{1i} \sim N(\mu_{\beta 1}, \sigma^2_{\beta 1})$, where $\mu_{\beta 1}$ and $\sigma^2_{\beta 1}$ are the common mean and between-men variance of the prenadir slopes.

The prior distribution of the changepoint was a continuous uniform distribution; the range was selected according to previous biological knowledge and depending on the subgroup of men. Secondary treatment usually is initiated when it is suspected that radiotherapy has failed, indicated by a rising PSA pattern starting within the first 2 to 3 years after radiotherapy; we thus selected a range of 5 years for this subgroup, $\tau_i \sim U(0, 5)$. Most men who do not receive a secondary treatment generally are those for whom radiotherapy is successful. In such cases, PSA are still produced by the remaining healthy prostate cells, although in very small quantities. Thus, the PSA concentrations for these men will start to rise at a later time, and at a very slow rate. For this reason, we selected a uniform distribution with a ten-year range for this subgroup: $\tau_i \sim U(0, 10)$.

In addition, our model included another feature of the PSA data not accounted for in earlier studies: we expressed the residual variance of the PSA measurements as a function of the PSA concentration. This is particularly appropriate given that PSA laboratory assays are known to have lower precision at lower PSA concentrations (11,12). We thus modeled the logarithm of the precision, that is, the reciprocal of the variance, as a linear function of the observed $\log_2$-PSA level, i.e.:

$$\log\frac{1}{\sigma_{ij}^2} = \theta_1 + \theta_2 \log_2 PSA_{ij}. \quad (3)$$

Finally, at the third hierarchical level, the parameters of the second level hierarchical distributions, that is $\mu_\alpha, \mu_{\beta 1}, \mu_{\beta 2}, \sigma_\alpha^2, \sigma_{\beta 1}^2, \sigma_{\beta 2}^2$, as well as θ_1 and θ_2, were assigned distribution functions. For the nadir, the slopes before and after the nadir, we used non-informative normal distributions on the mean parameters, and non-informative uniform prior distributions on the standard deviation parameters, i.e.,

$$\begin{aligned} \mu_\alpha &\sim N(0,100), \quad \sigma_\alpha^2 \sim U(0,4), \\ \mu_{\beta 1} &\sim N(0,100), \quad \sigma_{\beta 1}^2 \sim U(0,4), \\ \mu_{\beta 2} &\sim N(0,100), \quad \sigma_{\beta 2}^2 \sim U(0,4). \end{aligned}$$

We preferred to use a noninformative uniform prior distribution on the standard deviation parameters, because it is generally believed to work better than the commonly used inverse-gamma distribution (17). Similarly, the variance parameters, θ_1 and θ_2, were assigned noninformative normal prior distributions, i.e., $\theta_1 \sim N(0, 100)$ and $\theta_2 \sim N(0, 100)$.

We performed a sensitivity analysis to ensure that the estimates provided by our hierarchical model were not driven by the choice of our prior distributions. We tried other non-informative distributions with varying shapes, including log-normal distributions for the slopes and the PSA nadir, and normal and Dirichlet distributions for the changepoint. We assessed the adequacy of the fit of the model to the data using individual predicted PSA profiles (or realizations) (16).

Implementation

Estimation was implemented in WinBUGS, a statistical software package that uses Markov Chain Monte Carlo to generate random samples from the relevant posterior distributions (18). To ensure convergence, we generated three chains with distinct sets of widely dispersed initial values; we specified initial values for the individual and population parameters. For the first chain, we selected initial values close to expected values for the slopes, the nadirs, and the changepoints; standard deviation parameters were assigned an initial value of one. For the second chain, the slopes and the nadirs were assigned extreme initial values, as well as large standard deviations; the changepoints were assigned expected initial values. For the last chain, the slopes and the changepoints were assigned expected initial values and very large standard deviations; the individual changepoints were assigned large initial values. For each chain, we ran an initial burn-in period of 2,000 iterations, and an additional set of 10,000 iterations for inference. We assessed convergence using the Gelman-Rubin convergence statistic (16); in particular, we focused on the mean and variance of the four parameters of the PSA profiles, as well as the two variance parameters. We then pooled the three chains, and used the 30,000 iterations to estimate the posterior distributions for each parameter. Point estimates of parameters were calculated using the mean of the posterior distribution; 95% credible intervals were based on the 2.5^{th} and 97.5^{th} percentiles.

Association of PSA Trajectory Features With Baseline Characteristics

To illustrate how estimates provided by the BHC model can be used in subsequent analyses, we investigated the associations between individual level parameters and baseline characteristics, including age at diagnosis of the cancer, pretreatment PSA concentration, and pretreatment Gleason score. The Gleason score refers to the aggressiveness of the cancer, and ranges from 2 to 10; the higher the score, the more aggressive the tumor. We could have built the associations with baseline factors into the model, rather than analyzing them afterwards. However, we ran separate models in our two subgroups (with and without subsequent hormonal treatment). Because we preferred to analyze the associations all at once, we first pooled the estimates across all subjects from the two subgroups, and then analyzed the baseline factors. We constructed four standard multiple linear regression models; the dependent variables were one of the four parameters of the PSA profile, i.e., the $\log_2$PSA nadir, the changepoint, or one of the two slopes, and the independent variables were the above three baseline characteristics.

Alternative Approaches

To show the advantages of our BHC model, we used two alternative models. We first present a simplistic approach that independently estimates individual PSA profiles; population estimates are next obtained by pooling individual estimates using a simple average. The second model incorporates the multilevel structure of the data, and the $\log_2$PSA values are modeled using a smooth function of time.

AEP Vol. 18, No. 4
April 2008: 270–282

Independent Estimation of Individual PSA Profiles

Several studies have taken the PSA data at their face value, including the PSA nadir (19–21). For comparison purposes, we performed an equivalent naive analysis, thus ignoring measurement errors owing to short-term biological variations. For each man, we visually identified the lowest $\log_2$-PSA observation, and considered it as the true $\log_2$PSA nadir value. Similarly, we used the timing of the lowest $\log_2$-PSA observation as the true changepoint. The $\log_2$PSA decline rate for each man was estimated by the slope of the least-squares regression (LSR) line fit to the $\log_2$PSA values for that man versus time in years, for PSA values from pretreatment up until and excluding the $\log_2$PSA nadir. Similarly, the $\log_2$PSA growth rate was estimated by the slope of the LSR line fit to the $\log_2$PSA values from and including the $\log_2$PSA nadir to the last $\log_2$PSA observation available. The population parameters were obtained using a simple average of the individual estimates; we report the 95% reference range based on the 2.5th and 97.5th percentiles.

A Smooth Model. Smooth models can be particularly appropriate to model longitudinal biological data (22). Indeed, a smooth transition between two distinct phases, such as before and after the PSA nadir, can be more credible than a sharp transition as imposed by a piecewise linear pattern. We thus also investigated fitting a polynomial model for comparison purposes. Given that the PSA nadir following radiotherapy treatment is typically the only inflexion point in the PSA trajectory, we selected a quadratic model. We assumed that the $\log_2\text{PSA}_{ij}$ values were normally distributed about our quadratic curve, with expected value μ_{ij} and variance σ^2_{ij}, where μ_{ij}, was related to the timing of the measurement, t_{ij}, through the quadratic function $\mu_{ij} = \alpha_i + \beta_i t_{ij} + \gamma_i t^2_{ij}$, and where α_i, β_i,and γ_i are the individual level parameters for the i^{th} man. We assigned normal prior distributions to the individual parameters, i.e., $\alpha_i \sim N(\mu_\alpha, \sigma^2_\alpha)$, $\beta_i \sim N(\mu_\beta, \sigma^2_\beta)$, and $\gamma_i \sim N(\mu_\gamma, \sigma^2_\gamma)$ with noninformative normal distributions for the mean parameters ($\mu_\alpha \sim N(0, 100)$, $\mu_\beta \sim N(0, 100)$, $\mu_\gamma \sim N(0, 100)$) and noninformative uniform distributions for the standard deviations ($\sigma_\alpha \sim N(0, 4)$, $\sigma_\beta \sim N(0, 4)$, $\sigma_\gamma \sim N(0, 4)$) We also modeled the logarithm of the precision as a linear function of the $\log_2$PSA level, $\log(1/\sigma^2_{ij}) = \theta_1 + \theta_2 \log_2\text{PSA}_{ij}$, and assigned noninformative normal distributions to the variance parameters, i.e., $\theta_1 \sim N(0, 100)$ and $\theta_2 \sim N(0, 100)$. The adequacy of the model was assessed using the predicted PSA profiles.

RESULTS

From the initial population-based cohort of men diagnosed with prostate cancer between 1990 and 1992, 470 men were treated with radiotherapy and satisfied our inclusion criteria. Of those, 139 men subsequently received a hormonal treatment, and 331 did not. The shortest and longest PSA series had 3 and 36 measurements, respectively, with an average of 9 PSA readings per series. The shortest and longest durations of follow-up were, respectively, 4 months and 12 years long, with a mean follow-up duration of 5.7 years after the initiation of radiotherapy. The median interval between PSA readings was 6 months during the first 2-year period (range, 2–24 months), and 7 months afterwards (range, 1–83 months), close to the recommendations of the American Society for Therapeutic Radiology and Oncology Consensus panel published in 1997 (23).

The average age at diagnosis was similar for the two subgroups (69 and 70 years old for the subgroups with and without a subsequent hormonal treatment, respectively). The baseline median PSA level was greater in the subgroup of men with a subsequent treatment compared to the subgroup of men without a subsequent treatment (17 ng/mL and 10 ng/mL, respectively). Similarly, the median Gleason score was higher in the subgroup of men with a subsequent treatment compared to the subgroup of men without a subsequent treatment (7 and 6, respectively).

Bayesian Hierarchical Changepoint Model

Summary statistics for the estimates provided by the BHC model are shown in the first column of Table 1, and histograms are given in the first columns of Figures 2 and 3. The credible intervals and the spreads of the histograms suggest that the parameters were estimated with precision. In the subgroup of 139 men with a subsequent treatment, the estimate of the median number of PSA halvings per year was 3.50, corresponding to a median PSA half life of 0.29 years. The estimated median number of PSA doublings per year was 1.66, corresponding to a median PSA doubling time of 0.60 years. For the 331 men without a secondary treatment, the median number of PSA halvings per year was 3.29, corresponding to a median PSA half life of 0.30 years. The median number of PSA doublings per year was 0.28, corresponding to a median PSA doubling time of 3.57 years. Thus, as expected, men who received a secondary treatment, and thus those who tend to be the more severe cases, had a much shorter PSA doubling time, and reached the PSA nadir much sooner. In addition, the PSA nadir was higher for these men, than those who did not receive any additional treatment.

The variance parameters θ_1 and θ_2, as given by equation 3, were similar for the two subgroups. The estimate of θ_2 was positive, suggesting, as expected, that the variance of the PSA observations decreased at higher concentrations. Modeling the variability was particularly appropriate given the wide discrepancies between lower and higher concentrations. For example, in the subgroup of men with a secondary treatment, at mean $\log_2$PSA concentrations of 1 and 4, the

TABLE 1. Estimates (with 95% credible intervals) of the parameters of the Bayesian hierarchical changepoint (BHC) model (equations 2 and 3), and the simple model based on independent estimation of individual profiles

Parameters	BHC model	Simple model
139 men with a secondary treatment		
log_2PSA decline rate[a]	−3.50 (−3.89; −3.12)	−3.37 (−9.18; −0.65)
log_2PSA growth rate[b]	1.66 (1.38; 1.95)	1.51 (0.11; 5.97)
PSA nadir in ng/ml	3.01 (2.75; 3.38)	3.48 (0.03; 16.8)
Timing of the nadir in years	1.08 (1.00; 1.17)	1.17 (0.17; 3.07)
Variance parameter[c] (θ_1)	0.72 (0.55; 0.87)	Not applicable
Variance parameter[c] (θ_2)	0.31 (0.27; 0.36)	Not applicable
331 men without a secondary treatment		
log_2PSA decline rate[a]	−3.29 (−3.60; −2.99)	−2.22 (−8.26; −0.24)
log_2PSA growth rate[b]	0.28 (0.23; 0.33)	0.42 (−0.01; 1.85)
PSA nadir in ng/ml	1.25 (1.17; 1.35)	1.07 (0.03; 3.8)
Timing of the nadir in years	1.41 (1.29; 1.53)	2.58 (0.48; 7.17)
Variance parameter[c] (θ_1)	0.68 (0.62; 0.75)	Not applicable
Variance parameter[c] (θ_2)	0.27 (0.24; 0.29)	Not applicable

[a]Before the PSA nadir is reached; equivalent to the negative of the number of PSA halvings.
[b]After the PSA nadir is reached; equivalent to the number of PSA doublings.
[c]For example, in the subgroup of men with a subsequent treatment, and from equation 3, the estimated variance at mean log_2PSA concentration of 1, is $1/\exp(0.72 + 0.31 \times 1) = 0.357$, corresponding to a coefficient of variation of 60%.

standard deviations were, respectively, 0.60 and 0.37, corresponding to coefficients of variation of 60 and 10%. The results of our sensitivity analysis were not substantially different from our main analyses. As indicated by Figure 4, the predicted profiles fit the observed data very closely.

Association of PSA Model Features With Baseline Characteristics

Estimates for the parameters of the regression analysis are reported in Table 2. A greater initial PSA level was found to be associated with a larger number of PSA halvings per year (longer half-life), a higher PSA nadir, and a larger number of PSA doubings per year (shorter doubling time). Higher initial Gleason scores were associated with shorter doubling times. Finally, there was a positive association between age at diagnosis and the number of post-nadir PSA doublings.

Alternative Approaches

Independent Estimation of Individual PSA Profiles. Summary statistics are provided in the second column of Table 1, and histograms are given in the second columns of Figures 2 and 3. We excluded 3 series from the analysis of the number of PSA halvings in the first phase; the first observation in each of these series corresponded to the lowest observed PSA measurement and, thus, no regression line could be drawn for the initial phase. Thus, although the BHC model permits to estimate each individual profile, this simpler method cannot be applied to subjects with a limited number of observations.

Although most parameter point estimates were relatively close to the ones obtained using the BHC model, the PSA nadir was estimated to occur much earlier using the hierarchical model. This is particularly true for the subgroup of men without a second treatment; in this case, the nadir was reached about 1 year earlier using the hierarchical model, compared with the simple model. One likely explanation is the poor precision of the PSA readings at lower concentrations, which affects the subgroup of men without a secondary treatment more than the others. Indeed, for these men, the PSA concentrations decrease up to the nadir, and then remain at relatively low levels, compared to men subsequently treated who have a sharp PSA increase. The simple model considers the timing of the lowest observed PSA observation as the changepoint. Suppose that a man is followed-up for 10 years, and the changepoint occurs around 2 years. Afterwards, the PSA level stabilizes at low concentration. Then, because of the important variability of the observations at low PSA level, there are chances that PSA measurements, lower than the 2-year PSA reading, will be observed. Although the BHC model accounts for these random fluctuations, the simple model only considers the lowest reading. This is particularly well illustrated by the series 5, 6, and 8 in Figure 5, which represents the estimated log_2PSA trajectories estimated using the 3 methods. The nadir appears to be reached between the third and fourth year, and the series tend to follow a close to flat pattern afterwards. On the other hand, the lowest PSA measurement is observed much later than what appears to be the true nadir.

Compared with the BHC model, the simple model also provides less precise estimates, as shown by the tighter credible intervals and histograms; this is particularly obvious for the PSA nadir. Thus, the BHC model allows one to gain precision by allowing for the simultaneous estimation of the individual and population curves. This enhanced precision is

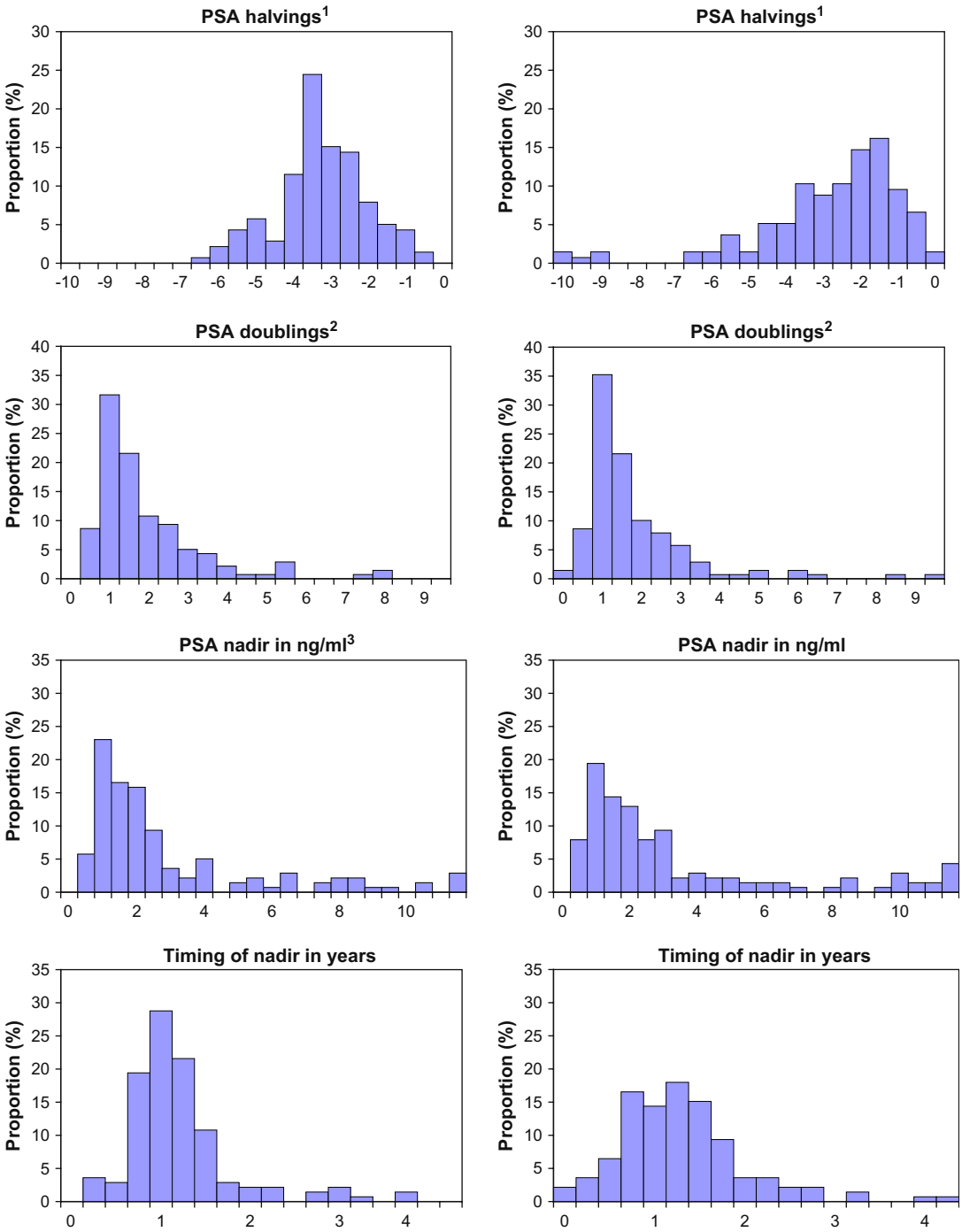

FIGURE 2. Estimates of the parameters of the BHC model (left) and the simple model using LSR (right), for the 139 men who underwent secondary therapy. [1]Number of PSA halvings before the PSA nadir is reached. [2]Number of PSA doublings after the PSA nadir is reached. [3]For enhanced visualization, one estimated nadir with value 53 ng/ml is not shown on the LSR plot.

also particularly well illustrated by Figure 5. The PSA trajectories estimated by the simple model follow all observations very closely, whereas the BHC model provides smoother pattern, putting less weight on observations that appear as outliers, again, a consequence of the simultaneous inference. Finally, note that this simple model does not allow one to provide an estimate of the residual PSA variability, nor to express it as a function of the PSA concentration.

A Smooth Model. Estimates of the quadratic model are presented in Table 3. Because of the hierarchical structure of

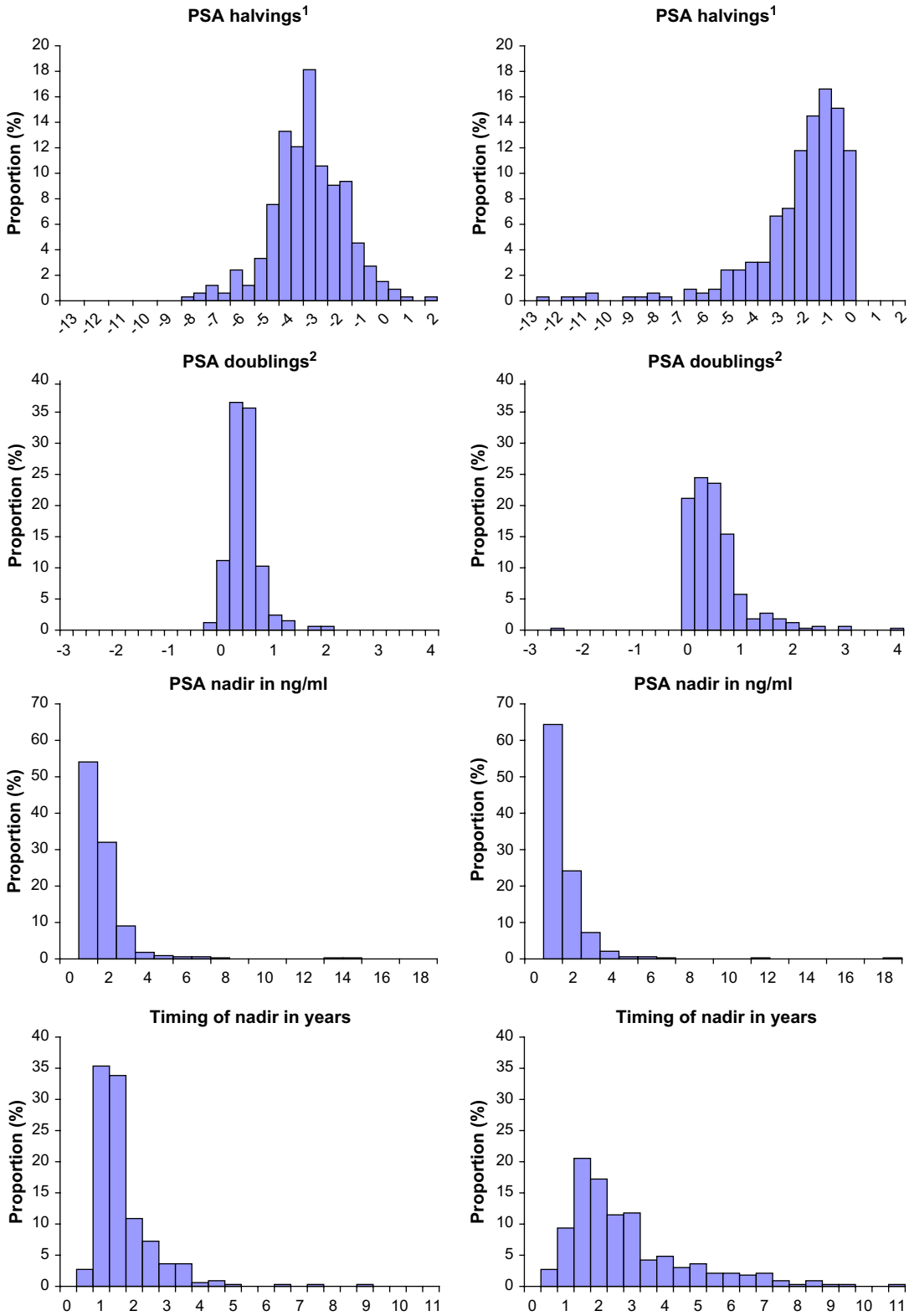

FIGURE 3. Estimates of the parameters of the BHC model (left) and the simple model using LSR (right), for the 331 men who did not undergo secondary therapy. [1]Number of PSA halvings before the PSA nadir is reached. [2]Number of PSA doublings after the PSA nadir is reached.

AEP Vol. 18, No. 4
April 2008: 270–282

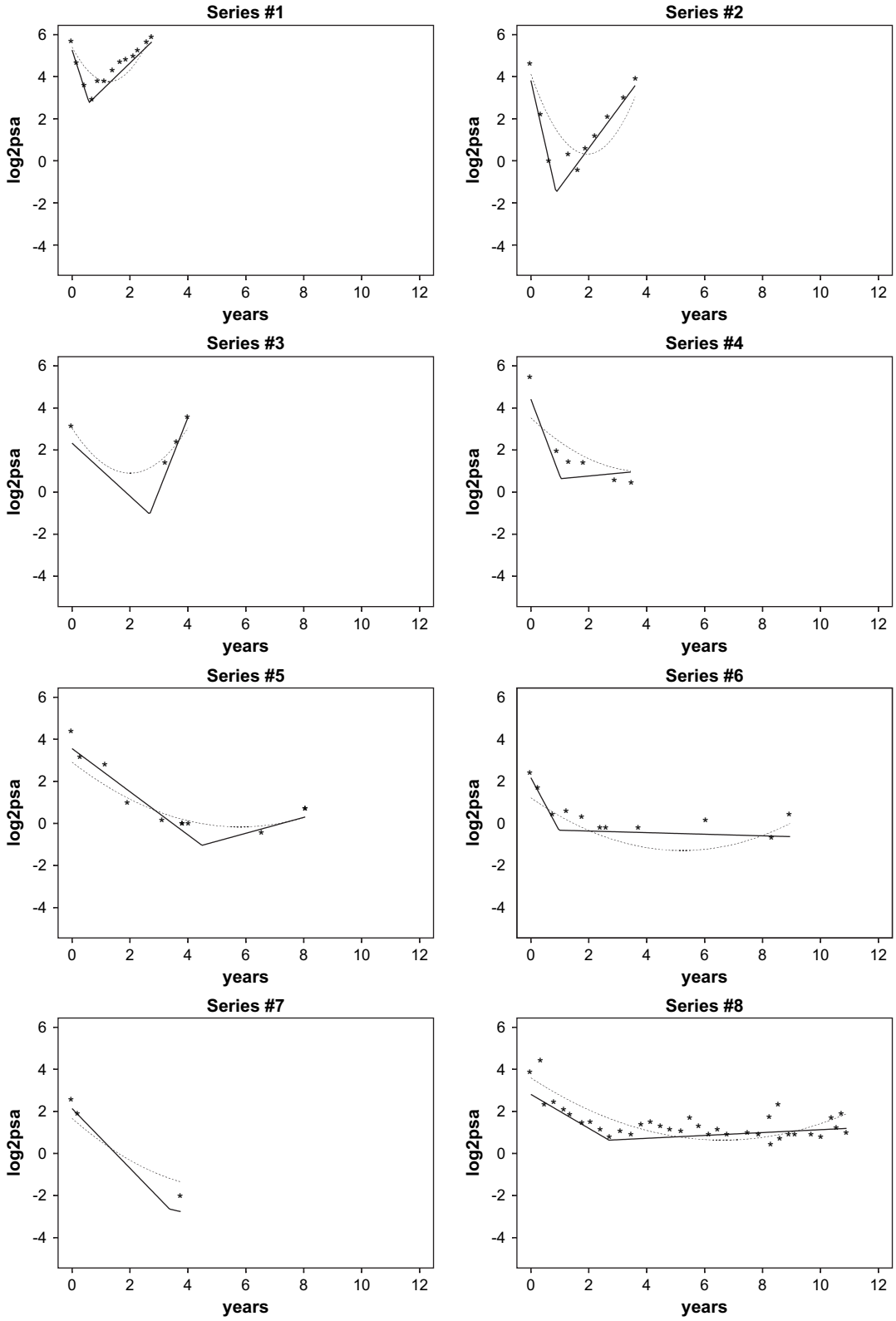

FIGURE 4. Postradiotherapy PSA observations (stars) for eight men, with predicted PSA profiles using the BHC model (solid line) and the quadratic model (dotted line).

TABLE 2. Regression parameters from four analyses of associations between the parameters of the PSA profile and pretreatment characteristics

	Covariates		
Dependent variable	Initial PSA level	Gleason score	Age at diagnosis
log_2PSA decline rate[a]	−0.04 (−0.05; −0.03)	0.04 (−0.07; 0.15)	−0.03 (−0.06; 0.00)
log_2PSA growth rate[b]	0.02 (0.01; 0.03)	0.22 (0.14; 0.30)	0.00 (−0.02; 0.02)
log_2PSA nadir	0.05 (0.04; 0.06)	0.09 (−0.01; 0.18)	−0.02 (−0.04; 0.01)
Timing of the nadir[c]	0.00 (−0.01; 0.00)	−0.04 (−0.10; 0.03)	0.00 (−0.02; 0.02)

[a]Before the PSA nadir is reached.
[b]After the PSA nadir is reached.
[c]In years.
Each row corresponds to one regression analysis. For example, using the second line, and given a man with initial PSA level of 2 ng/mL (i.e., log_2PSA = 1), Gleason score of 2, and age 70 years, the estimated number of postnadir PSA doublings per year is $0.02 \times 1 + 0.22 \times 2 + 0 \times 70 = 0.46$. Similarly, with a Gleason score of 10, the estimated number of post-nadir PSA doublings per year is 2.2.

the model, inference is strengthened, and as a result, parameters are estimated with precision, as suggested by the relatively tight credible intervals.

The smooth model, however, has some major drawbacks compared with the BHC model. First, although the parameters of the BHC model have all an important clinical interpretation, the parameters provided by a polynomial model are not as clinically useful. Importantly, no estimate of the PSA doubling time is directly available. Moreover, the smooth model does not appear in accordance with the underlying biological process. The quadratic model allows for a smooth transition, however, it assumes that the PSA pattern is symmetric about the nadir, which is in contradiction with the typical postradiotherapy PSA pattern. This results in a poor fit to the data, as confirmed by the predicted PSA profiles (Fig. 4). One could then use a greater degree polynomial; however, this process would also be biologically inappropriate, because the PSA nadir is typically the only inflexion point of the postradiotherapy PSA profiles. The mean log_2PSA profiles as estimated by the 3 models are shown in Figure 5.

DISCUSSION

We have shown that the BHC model is particularly appropriate for modeling longitudinal data with abrupt changes, because it easily allows the user to account for the between and within-subjects variabilities, as well as other complex features, such as the presence of a random changepoint, and nonconstant variance. As a result, estimates are more precise compared with a simpler model relying on independent individual estimations, all clinically important parameters are estimated, and estimation of the PSA variability as a function of the PSA concentration is available. In addition, the flexible hierarchical structure of the model could be easily extended to account for an additional level, or source of variability. For example, one can pool PSA series from multiple studies, and use the study as another hierarchical level, assuming then that subjects from the same study share common mean parameters (7).

One limitation of hierarchical modeling is the number and timing of measurements which can be potentially informative about the shape of the subsequent curve, and thus introduce some bias. Because of the hierarchical structure, subjects with larger number of observations will contribute more information to the hierarchical means and variances compared to lesser-tracked subjects; if these lesser-tracked subjects are different in some way, this can introduce some bias. In the case of post-treatment PSA data, healthier subjects are usually followed up less closely than men with poorer prognosis. However, healthy subjects tend to have lower pretreatment PSA levels and, thus, potentially different subsequent slopes. The possibility of such a bias should be recognized when interpreting hierarchical models.

Alternative approaches to the BHC model are available. Instead of a single master model, one could use a simpler model such as the first alternative analysis presented in this paper. Because the method is relatively simple, it is commonly applied in the clinical literature. The PSA nadir, for example, is often defined as the lowest observed PSA concentration; the population nadir is then estimated using the average of the observed individual nadirs (24–27). To estimate the PSA doubling time, some authors identified PSA series they believed were rising (based on some arbitrary definition), and fitted a least-squares regression line only to these specific rising profiles; the mean of the resulting estimated slopes was used as an estimate of the mean PSA doubling time (20,21,28). Although these methods, similar to our first alternative analysis are straightforward, and not time consuming, they do not adequately account for the PSA variability. By using the observed nadir as the true nadir, the variability is completely ignored.

With respect to the slopes, the simple average gives equal weight to each series regardless of the number of observed data points; the hierarchical structure of the data is ignored,

AEP Vol. 18, No. 4
April 2008: 270–282

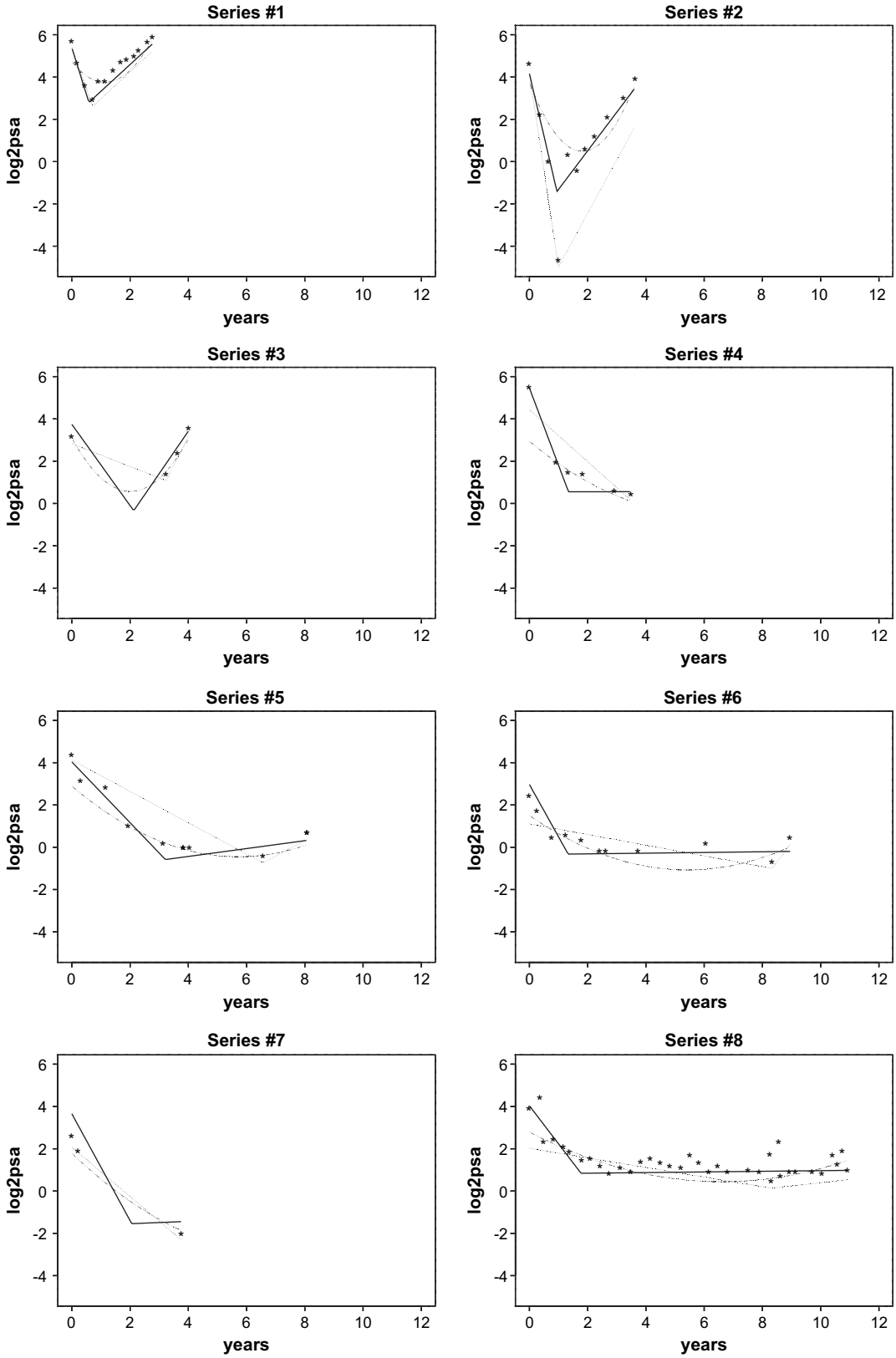

FIGURE 5. Postradiotherapy PSA observations (stars) for eight men, with mean PSA profiles as estimated by the BHC model (solid line), the simple model using LSR (dotted line), and the quadratic model (dashed line).

TABLE 3. Estimates (with 95% credible intervals) of the parameters of the smooth model

Parameters	Estimates
139 men with a secondary treatment	
Constant term (μ_α)	3.70 (3.53; 3.87)
Linear term (μ_β)	−1.90 (−2.18; −1.62)
Quadratic term (μ_γ)	0.68 (0.58; 0.80)
Variance parameter (θ_1)	0.48 (0.41; 0.56)
Variance parameter (θ_2)	−0.51 (−0.70; −0.34)
331 men without a secondary treatment	
Constant term (μ_α)	1.88 (1.54; 2.14)
Linear term (μ_β)	−0.87 (−0.96; −0.78)
Quadratic term (μ_γ)	0.10 (0.09; 0.11)
Variance parameter (θ_1)	0.06 (−0.12; 0.16)
Variance parameter (θ_2)	−0.23 (−0.35; −0.14)

and thus so are the multiple sources of variability. On the other hand, the BHC model relies on the borrowing of strength. Individual and population parameters are estimated simultaneously, and information from other series is used when estimating parameters from a specific trajectory. The individual profiles are estimated using what is effectively a weighted average of the observed profiles and the population profile, where the weights are the corresponding precisions. Thus, the individual profiles are pulled (or shrunk) toward the population profile: the less precision of the individual series compared to the precision of the population estimates, the more pulling, and vice-versa. As a result, overall inference is strengthened at the expense of possible bias at the individual level.

Finally, the hierarchical model allowed us to estimate the PSA variability as a function of the underlying PSA levels. Incorporating this feature into the simpler model would be theoretically possible. However, in practice, with relatively few data points per subject, without the borrowing of strength inherent to hierarchical models, it would be very difficult to accurately estimate parameters from a model that simultaneously estimate the trajectory and variance within each subject when the variance is a function of the observed levels within each man. We therefore did not attempt to estimate such an unstable model.

Smooth models, such as splines (29) or polynomials, can be appropriate for longitudinal data. Just as the BHC model, the smooth model presented in this paper adequately accounted for the multilevel structure of the data. However, from a biological perspective, a piecewise linear pattern seems more appropriate, and provides clinically useful estimates.

One could also model the raw data without relying on a logarithm transformation, since some have used a nonlinear model such as the exponential decay -exponential growth model in the form $PSA = a_1 \exp(-b_1 t) + a_2 \exp(b_2 t)$, where a_1, a_2, b_1, and $b_2 > 0$ are the parameters of interest, and $\ln 2/b_1$, $\ln 2/b_2$ and $a_1 + a_2$ provide, respectively, the PSA half life, the subsequent PSA regrowth, and the post-treatment PSA level (30,31). However, information on the subject status is necessary, because in the case of cured patients, the parameter b_2 is set to 0.

One could prefer a method based on maximum likelihood instead of a Bayesian estimation approach (32,33); these models have been applied to describe the natural history of prostate disease in healthy men where the changepoint represents cancer onset (34,35). Beyond the fundamental conceptual differences between the Bayesian and classical procedures, there are however, some complexities in the frequentist approach that are not (or less) encountered when relying on a Bayesian estimation process. Specifically, there may be a problem of non-identifiability if there is a non-zero probability that there is no changepoint, that is, the changepoint corresponds to the timing of the last observation. This is not an issue when using a Bayesian approach, as the prior distribution on the changepoint allows the user to separate out more likely from less likely parameter values, not possible from the likelihood alone.

Finally, from a clinical perspective, it is difficult to compare our estimates to published results. First, as suggested by our analysis, pretreatment characteristics, such as the PSA level or the Gleason score are associated with characteristics of the PSA profile. Thus, it is likely that studies with different selection criteria will produce different estimates. Also, as discussed earlier, the statistical methods used by other studies are not always appropriate, and a measure of precision is not systematically reported. Overall, however, results tend to be in accordance. Reported estimates of the PSA nadir range between 1 and 3 ng/mL (24,25); the PSA doubling time and half-life were estimated to be approximately 1 year and 2 months, respectively (20,21). The association between the parameters of the PSA profile and pretreatment characteristics has been investigated in a few studies, and results are in accordance with our findings. In particular, a shorter PSA doubling time has been shown to be associated with a higher pretreatment Gleason score (14,20,21), and higher pretreatment PSA concentrations (21). Similarly, an association has been reported between the PSA nadir and the pretreatment PSA level, but none with the Gleason score.

The BHC model has multiple advantages. The model is in accordance with the underlying biological principles and provides clinically useful estimates, accounts for the multiple sources of variability, allows for the borrowing of strength, and finally the estimation process permits complex modeling of the variance function. The resulting precise estimates of the biomarker profile can then provide valuable information when monitoring disease and treatment outcomes; they can be used, for example, as prognostic factors in analysis of long-term outcomes such as survival. This

type of model should be applicable to the study of marker series in other diseases.

REFERENCES

1. Carlin B, Gelfand A, Smith A. Hierarchical Bayesian analysis of changepoint problems. Appl Stat. 1992;41:389–405.
2. Stephens D. Bayesian retrospective multiple-changepoint identification. Appl Stat. 1994;43:159–178.
3. Lange N, Carlin B, Gelfand A. Hierarchical Bayes models for the progression of HIV infection using longitudinal CD4 T-cell numbers. J Am Stat Assoc. 1992;87:615–632.
4. Ghosh P, Vaida F. Random changepoint modelling of HIV immunologic responses. Stat Med. 2007;26:2074–2087.
5. Slate E, Lark L. Using PSA to detect prostate cancer onset: an application of Bayesian retrospective and prospective changepoint identification. In: Gatsonis C, Kass R, Carlin B, et al., eds. Case Studies in Bayesian Statistics, volume 4. New York: Springer; 1999:395–412.
6. Slate E, Turnbull B. Statistical models for longitudinal biomarkers of disease onset. Stat Med. 2000;19:617–637.
7. Inoue L, Etzioni R, Slate E, Morrell C, Penson D. Combining longitudinal studies of PSA. Biostatistics. 2004;5:483–500.
8. Slate E, 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, volume 3. New York: Springer; 1999:435–456.
9. Pauler D, Finkelstein D. Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcome. Stat Med. 2002;21:3897–3911.
10. Joseph L, Wolfson D, Belisle P, et al. Taking account of between-patient variability when modeling decline in Alzheimer's disease. Am J Epidemiol. 1999;149:10963–10973.
11. Eastham J, Riedel E, Scardino P, et al. Variation of serum prostate-specific antigen levels. JAMA. 2003;289:2695–2700.
12. Ballentine Carter H, Pearson J, Metter J, et al. Longitudinal evaluation of prostate-specific antigen levels in men with and without prostate disease. JAMA. 1992;267:2215–2220.
13. Albertsen P, Hanley J, Penson D, et al. Validation of increasing prostate specific antigen as a predictor of prostate cancer death after treatment of localized cancer with surgery or radiation. J. Urol. 2004;171:2221–2225.
14. Schmid H, McNeal J, 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.
15. Meek A, Park T, Oberman E, Wielopolski L. 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.
16. Gelman A, Carlin J, Stern H, Rubin D. Bayesian data analysis. 2nd ed. New York: Chapman and Hall; 2004.
17. Gelman A. Prior distributions for variance parameters in hierarchical models. Bayesian Anal. 2006;1:515–533.
18. Spiegelhalter D, Thomas A, Best N, Gilks W. WinBUGS: Bayesian inference using Gibbs sampling, version 1.3. Cambridge, MA: MRC Biostatistics Unit; 2000.
19. D'Amico A, Cote K, Loffredo, et al. Determinants of prostate cancer specific survival following therapy during the prostate specific antigen era. J. Urol. 2003;170:S42–47.
20. Zagars G, Pollack A. The fall and rise of prostate-specific antigen: kinetics of serum prostate specific antigen levels after radiation therapy for prostate cancer. Cancer. 1993;72:832–842.
21. Pollack A, Zagars G, Kavadi V. Prostate specific antigen doubling time and disease relapse after radiotherapy for prostate cancer. Cancer. 1994; 74:670–678.
22. Ulm K. A statistical method for assessing a threshold in epidemiological studies. Stat Med. 1991;10:341–349.
23. American Society for Therapeutic Radiology and Oncology Consensus Panel. Consensus statement: Guidelines for PSA following radiation therapy. Int J Radiat Oncol Biol Phys. 1997;37:1035–1041.
24. Hanlon A, Diratzouian H, Hanks G. 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.
25. DeWitt K, Sandler H, Weinberg V, McLaughlin P, Roach M. What does post-radiotherapy PSA nadir tell us about freedom from PSA failure and progression-free survival in patients with low and intermediate-risk localized prostate cancer? Urology. 2003;62:492–496.
26. Kaminski J, Hanlon A, Horwitz E, Pinover W, Mitra R, Hanks G. Relationship between prostate volume, prostate-specific antigen nadir, and biochemical control. Int J Radiat Oncol Biol Phys. 2002;52:888–892.
27. Critz F, Levinson K, Williams W, Holladay D, Griffin V, Holladay C. Prostate specific antigen nadir achieved by men apparently cured of prostate cancer by radiotherapy. J Urol. 1999;161:1199–1205.
28. Hanlon A, Moore D, Hanks G. Modeling postradiation prostate specific antigen level kinetics. Cancer. 1998;83:130–134.
29. Guess B, Jennrich R, Johnson H, Refheffer R, Scholz M. Using splines to detect changes in PSA doubling times. Prostate. 2003;54:88–94.
30. Law N, Taylor J, Sandler H. The joint modeling of longitudinal disease progression marker and the failure time process in the presence of cure. Biostatistics. 2002;3:547–563.
31. Taylor J, Yu M, Sandler H. Individualized predictions of disease progression following radiotherapy for prostate cancer. J Clin Oncol. 2005;23:816–825.
32. Hinkley D. Inference about the change-point in a sequence of random variables. Biometrika. 1970;57:1–17.
33. Hinkley D. Inference about the intersection in two-phase regression. Biometrika. 1969;56:495–504.
34. Morrell C, Pearson J, Ballentine Carter H, et al. Estimating unknown transition times using a piecewise nonlinear mixed-effects model in men with prostate cancer. J Am Stat Assoc. 1995;90:45–53.
35. Pearson J, Morrell C, Landis P, Ballentine Carter H, Brant L. Mixed-effects regression models for studying the natural history of prostate disease. J Am Stat Assoc. 1994;13:587–601.