SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NY 14853-3801

TECHNICAL REPORT NO. 1189

March 1997

Changepoint Modeling of Longitudinal PSA as a Biomarker for Prostate Cancer[1]

by

Elizabeth H. Slate
and Kathleen A. Cronin

[1]This research was supported by grants PHS R01 CA61120 from the National Institute of Health and DMS-9505065 from the National Science Foundation.

Changepoint modeling of longitudinal PSA as a biomarker for prostate cancer

Elizabeth H. Slate
Kathleen A. Cronin

ABSTRACT
Prostate-specific antigen (PSA) is an important indicator of the presence of prostate disease. When the volume of the prostate increases, as when cancer is present, the levels of PSA in the blood also increase. Our work focuses on using PSA levels as a biomarker for the recurrence of prostate cancer in patients that have been previously diagnosed and treated by radiotherapy. We fit a fully Bayesian hierarchical changepoint model to the longitudinal PSA readings. Our objective is twofold; to better understand the natural history of PSA levels in patients who have completed treatment, and to use the model to identify individual changepoints that are indicative of recurrence. With the goal of accurate early detection of recurrence, we perform a prospective sequential analysis to compare several diagnostic rules, including a rule based on the posterior distribution of individual changepoints.

1 Introduction

Garnick (1994) discusses the importance of research in the detection and treatment of prostate cancer, a disease the American Cancer Society predicts will strike 244,000 men and kill 40,400 men in the United States in 1995. Prostate specific antigen (PSA) is a glycoprotein produced by the prostate gland that increases with the volume of the prostate. Many papers have examined the usefulness of PSA as a biomarker for prostate cancer. The work of Catalona *et al.* (1991, 1993) supported the usefulness of PSA levels as a diagnostic marker for prostate cancer. Gerber (1991) discussed the value of screening along with a review of current screening methods. Oesterling *et al.* (1993) performed a prospective study to understand the link between PSA and age. They concluded that PSA increases gradually with age in healthy men and suggested normal ranges of PSA for different age groups. Whittemore *et al.* (1995) compare diagnostic rules based on PSA levels in black men and white men. They conclude that a single current PSA reading outperforms changes in past PSA readings in identifying

men who will develop prostate cancer in the next 7 years.

A few papers have modeled longitudinal PSA readings to better understand the natural history of the PSA trajectories in those who do and do not develop prostate cancer. Carter *et al.* (1992), and Pearson *et al.* (1991, 1994) looked at serial PSA readings for men over a period of 7 to 25 years. They used a mixed-effects regression model to test whether the changes in PSA readings were different in men with and without prostate disease. Model parameters were estimated using a Newton-Raphson restricted maximum likelihood method. Carter *et al.* (1992) observed that PSA increases only very slowly with age before the onset of cancer and then increases more rapidly when cancer is present.

Morrell *et al.* (1995) investigated whether the natural history of PSA progression in men with locally confined prostate cancer differs from that in men with metastatic tumors. They used a mixed-effects model with a random changepoint for longitudinal PSA readings. Approximate maximum likelihood methods were used to estimate model parameters. They concluded that the difference between local/regional and advanced/metastic cancers is that advanced/metastic cases are diagnosed later in the progression.

We generalize the mixed-effects models introduced by Laird and Ware (1982) to include random changepoints to model longitudinal PSA readings. There is a great deal of literature on identifying when a process has changed and estimating the changepoint. Page (1955) used non-parametric methods to test the hypothesis that all observations are from the same distribution. Hinkley (1969, 1970) used maximum likelihood estimation to identify a shift in process mean and the intersection of a two-phase regression. Smith (1975) presented a Bayesian approach to estimating changepoints for normal and binomial distributions along with an informal sequential procedure. Carlin et al. (1992) gave a fully Bayesian hierarchical analysis of changepoint problems, including the use of the Gibbs sampler to solve for the posterior distributions of model parameters. Stephens (1994) looked at continuously distributed changepoints and multiple changepoint identification from a retrospective viewpoint.

We analyze longitudinal PSA readings recorded for patients treated by radiotherapy for prostate cancer. We start with a fully Bayesian hierarchical model with a single random changepoint representing the onset of a recurrence of cancer. The hierarchical approach permits the "borrowing of strength" from the population to estimate the subject-specific parameters while accounting for the within-subject serial correlation. This single-changepoint model was found to have several shortcomings when applied to our data set. We then expand the model to include two random changepoints; the first representing the end of the transient effects of radiotherapy, and the second representing the onset of recurrence. We use these models to perform a prospective sequential analysis and compare the timeliness of changepoint detection with other proposed diagnostic rules using receiver

operator characteristic (ROC) curves.

2 Data description

A study at the University of Michigan currently follows patients that have been diagnosed and treated for prostate cancer. The patients are treated with radiotherapy and then closely monitored for signs of recurrence. The data consist of longitudinal PSA readings recorded from May 7, 1987 through January 6, 1995 for these patients, beginning upon completion of radiotherapy. For individuals experiencing a recurrence of prostate cancer, the date of recurrence and the date when subsequent treatment (hormone therapy) began are also available. Recurrence is classified as a distant (outside the prostate), local (within the treatment field) or chemical (rising PSA levels) failure. A distant failure occurs when cancer is detected outside the prostate by bone scan, imaging or biopsy. A local failure happens when cancer is detected within the radiation treatment field by biopsy, digital rectal exam, or both. A chemical failure is determined to have occurred when either of the following happen:

- there is a sustained rise in the PSA readings of greater than 50% of the post radiotherapy nadir reaching at least 4 ng/ml,
- when the post radiotherapy nadir is not less than 4 ng/ml, there are two successive rises or two non-successive rises with no intervening decline in the PSA readings.

Those individuals experiencing a non-chemical recurrence are called cases here. Cases will be said to have encountered a "recurrence of cancer" to indicate that cancer has been detected after radiotherapy. Subjects remaining cancer-free during the time period of our data (cancer has not been detected) are termed controls. Because we are investigating the utility of using PSA to aid the detection of recurrence of cancer, subjects experiencing chemical failures are controls rather than cases since recurrence of cancer has not been confirmed.

Radiation therapy uses high-energy ionizing radiation to kill cancer cells within a carefully specified treatment field. The radiation damages one or both strands of the DNA inside the cells, thereby preventing the cells from growing and dividing. While cells in all phases of the cell cycle can be damaged by radiation, the lethal effect of radiation may not be apparent until after one or more cell divisions have occurred. This accounts for the initial decrease in PSA levels that is seen in both cases and controls (see Figures 1 and 2). After the initial decrease, the PSA levels are expected to remain relatively constant if treatment is effective. However, a rising PSA level is often experienced in the case of a recurrence of cancer. The single-changepoint model uses only those observations after the PSA has

returned to a normal level, while the two-changepoint model looks at the natural history of PSA from the completion of radiotherapy.

We have included subjects in the analysis whose PSA trajectories meet both of the following conditions:

- The post radiotherapy nadir is no more than 4 ng/ml. Because healthy men typically have PSA levels less than 4 ng/ml, this is an indication of some success of the treatment.
- There are at least two PSA observations after the first PSA reading not exceeding 4 ng/ml and before the beginning of hormone treatment, if present. Readings taken after the beginning of hormone treatment are excluded because it is believed that hormone therapy directly affects PSA levels.

After applying these inclusion criteria, our data consist of 411 patients, 47 of which are cases (*i.e.* recurrence of cancer was confirmed during our followup period).

For the single-changepoint model, we use the longitudinal observations starting with the first reading not exceeding 4 ng/ml and stopping with either the last observation available or the start of hormone treatment. These conditions yield 2402 records for the 411 subjects, with a median of 5 readings per subject (medians of 7 readings per case and 5 readings per control). For the two-changepoint model, we use the longitudinal readings starting upon completion of radiotherapy and stopping with either the last observation or the start of hormone therapy. In this case the total number of PSA readings for the 411 patients is 2755, with a median of 6 readings per patient overall and median number of readings of 9 for the 47 cases and 6 for the 364 controls.

Figures 1 and 2 illustrate typical PSA trajectories for cases and controls. The plots for control subjects 6710931 and 12316045 and case subjects 88226 and 89018 show the effect of the truncation of the initial readings for the single-changepoint model: the points marked by open circles are not included in the single-changepoint analysis.

3 Hierarchical model

3.1 Single-changepoint model

The mixed-effects model for linear growth before and after the changepoint, t_i, can be written as

$$Y_{ij} = a_{oi} + a_i x_{ij} + b_i \, (x_{ij} - t_i)^+ + \epsilon_{ij}, \tag{1.1}$$

where y_{ij} is the ln(PSA) value for subject i at observation j, x_{ij} is the time in months since completion of radiotherapy of observation j for subject

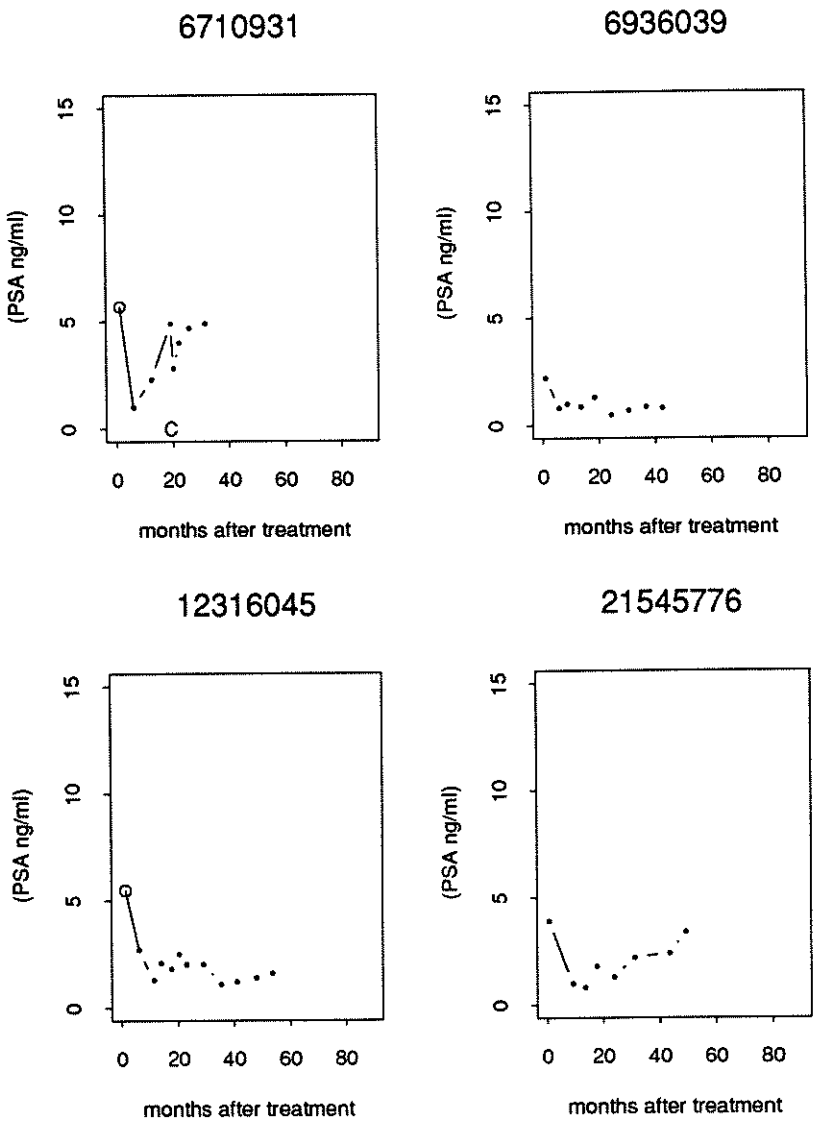

FIGURE 1. PSA trajectories for selected controls. The character C denotes the time of chemical failure for subject 6710931. For subjects 6710931 and 12316045, the open point is a reading that is not included for the single-changepoint model but is included for the two-changepoint model.

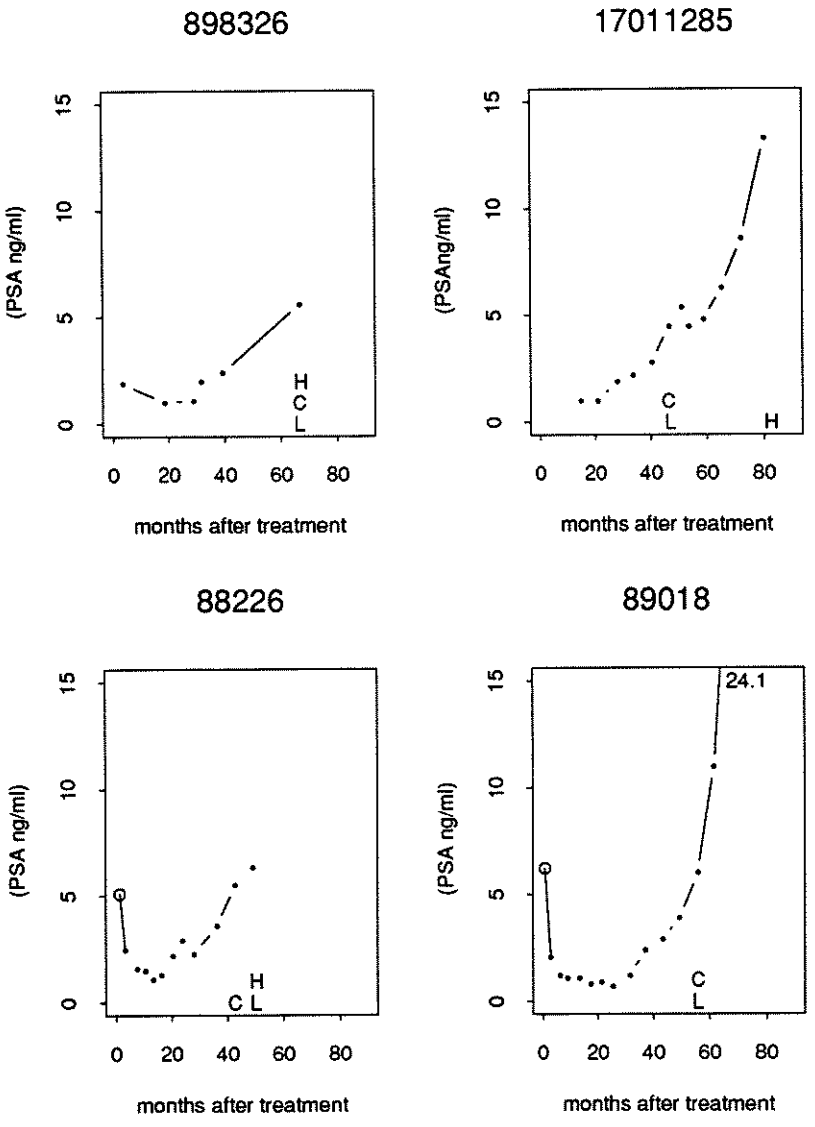

FIGURE 2. PSA trajectories for selected cases. The characters L, D, C and H denote the times of local failure, distant failure, chemical failure and the start of hormone treatment, respectively. The open points for subjects 88226 and 89018 are readings that are not included for the single-changepoint model but are included for the two-changepoint model.

i, and $z^+ = \max(0, z)$. The index i takes values $1, \ldots, N$ and j takes values $1, \ldots, n_i$ when there are N subjects in the study and the ith has n_i observations. A model for PSA that is approximately linear on the log scale has been used by many researchers (see, for example, Carter *et al.* 1992, Pearson *et al.* 1991, 1994, and Whittemore *et al.* 1995). The complete model assumes the following distributions:

$$
\begin{aligned}
\begin{pmatrix} a_{oi} \\ a_i \end{pmatrix} \Bigg| \begin{pmatrix} \alpha_o \\ \alpha \end{pmatrix}, \Sigma_a &\sim \mathrm{N}_2 \left\{ \begin{pmatrix} \alpha_o \\ \alpha \end{pmatrix}, \Sigma_a \right\} \qquad (1.2) \\
\begin{pmatrix} \alpha_o \\ \alpha \end{pmatrix} &\sim \mathrm{N}_2 \left\{ \begin{pmatrix} \mu_{\alpha_o} \\ \mu_\alpha \end{pmatrix}, \Sigma_\alpha \right\} \\
\Sigma_a^{-1} &\sim \mathrm{W}\left\{(\rho V)^{-1}, \rho\right\}, \text{ Wishart with scale } (\rho V)^{-1},\ \rho \text{ df} \\
b_i \mid \beta, \sigma_b^2 &\sim \mathrm{N}(\beta, \sigma_b^2) \\
\beta &\sim \mathrm{N}(\mu_\beta, \sigma_\beta^2)\, I(\beta > k) \\
\frac{1}{\sigma_b^2} &\sim \mathrm{Gamma}(\lambda_b, r_b) \\
t_i | \tau, \sigma_t^2 &\sim \mathrm{N}(\tau, \sigma_t^2) \\
\tau &\sim \mathrm{N}(\mu_\tau, \sigma_\tau^2) \\
\frac{1}{\sigma_t^2} &\sim \mathrm{Gamma}(\lambda_t, r_t) \\
\epsilon_{ij} | \sigma_{\epsilon_i}^2 &\sim \mathrm{N}(0, \sigma_{\epsilon_i}^2) \\
\frac{1}{\sigma_{\epsilon_i}^2} &\sim \mathrm{Gamma}(\lambda_\epsilon, r_\epsilon).
\end{aligned}
$$

The subject-specific parameters are conditionally independent, as are the within-subject errors. The prior distributions for $(\alpha_o, \alpha)^T$, Σ_a^{-1}, β, σ_b^{-2}, τ, σ_t^{-2}, and $\sigma_{\epsilon_i}^{-2}$ are assumed known.

The Gibbs sampler, as described in Gelfand and Smith (1990), is used to solve for the posterior distributions of the model parameters. The procedure is similar to that in Lange *et al.* (1992), but with a continuous changepoint as described in Stephens (1994). The complete conditional distributions for each parameter, with the exception of the $\{t_i\}$, are standard parametric distributions and can be easily sampled. Although the form of the complete conditional distribution for t_i changes at each observation time, the form of the distribution between observation times is known. Hence the $\{t_i\}$ can be generated in a two step procedure which first generates an interval and then generates a value within that interval. Thus it is straightforward to generate from all the complete conditional distributions. This procedure leads to estimates of the subject specific parameters, including the $\{t_i\}$, based on posterior distributions.

Our choice of the form for the prior distributions was motivated primarily by computational ease. Nevertheless, we sought to accurately represent known information about PSA within these classes of distributions (specif-

ically, we drew heavily on the results of Catalona *et al.*, 1991 and 1993) without being overly informative. Our prior distributions are given in Appendix 1.

Figure 3 shows the estimated posterior distributions for the population parameters after 100, 300 and 500 iterations of the Gibbs sampler. The distribution of α, the mean of the slopes before recurrence, is centered slightly below -0.02 showing the lingering effects of the radiotherapy; even once below the normal reading of 4 ng/ml the PSA readings continue to decrease for many patients. The parameter τ is the mean time to recurrence for the mix of cases and controls in the population being modeled. There is an implicit assumption that every subject will experience a recurrence of cancer eventually, although the time to recurrence may be extremely long for some. The posterior distribution of τ shown in Figure 3 shows that, for this population consisting of about 89% controls, the mean time to recurrence is a bit longer than 50 months.

Figure 4 shows the evolution of the posterior distribution of the changepoint for a selected case as readings are accumulated. The PSA readings for this case are shown in the first panel of Figure 2. The posterior distribution for the changepoint puts substantial mass to the left of the current observation time by 31.4 months after completion of radiotherapy (13.06 months after the nadir), indicating that a recurrence is likely. There can be little doubt by the final panel, which is when local failure was diagnosed for this patient. Although the posterior distribution indicates that the changepoint has occurred, it does not accurately indicate when this changepoint occurred. Rather, the location of the distribution appears to be unstable, shifting to the right as more PSA readings are obtained. There are two factors contributing to this behavior: the characteristics of the data being fit, specifically the mix of cases and controls, and the model restriction that the post-changepoint slope of the ln(PSA) trajectory be at least 0.2. The large proportion of control subjects in our data (about 89%) causes the posterior distribution of τ, the mean of the changepoints, to be shifted toward larger values, which in turn leads the model to expect larger rather than smaller changepoints. As more readings are accumulated beyond the time of cancer recurrence for a case, the slope of the ln(PSA) trajectory typically increases so that a more recent estimated changepoint permits a larger estimated post-changepoint slope.

There are two major shortcomings of the single-changepoint model. The first is that the model does not capture the initial decrease in PSA levels caused by radiotherapy. Although we tried to eliminate this problem by excluding initial readings greater than 4 ng/ml, the PSA readings continued to decrease for many cases and controls. The second shortcoming is that the incremental increase in slope b_i is poorly distinguished. The slope after the changepoint is $a_i + b_i$ where a_i is often estimated to be negative. Thus the value of k, the lower bound for β, had to remain large to separate cases and controls.

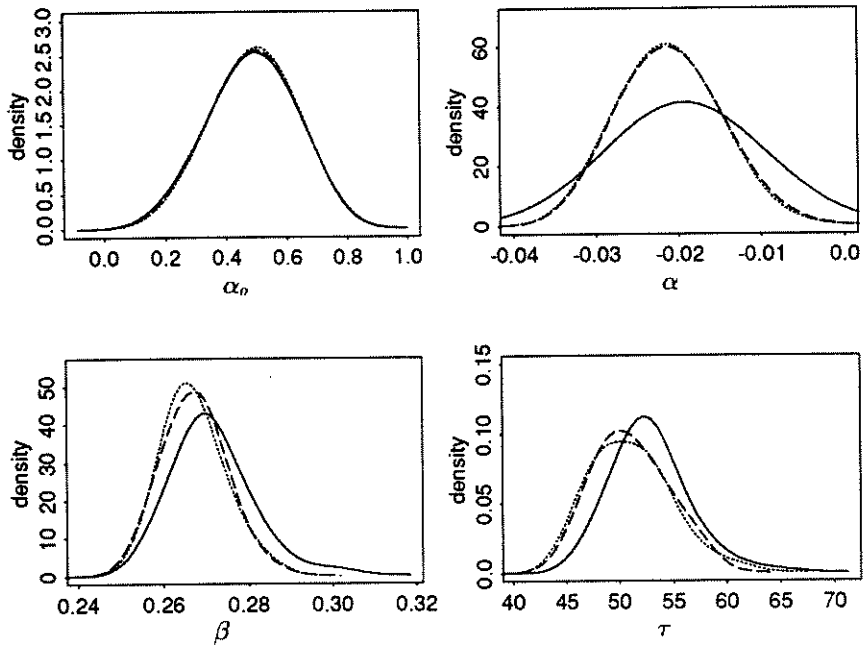

FIGURE 3. Convergence of the posterior distributions for selected population parameters in the single-changepoint model. The solid, dashed and dotted lines denote the posterior after 100, 300 and 500 iterations of the Gibbs sampler.

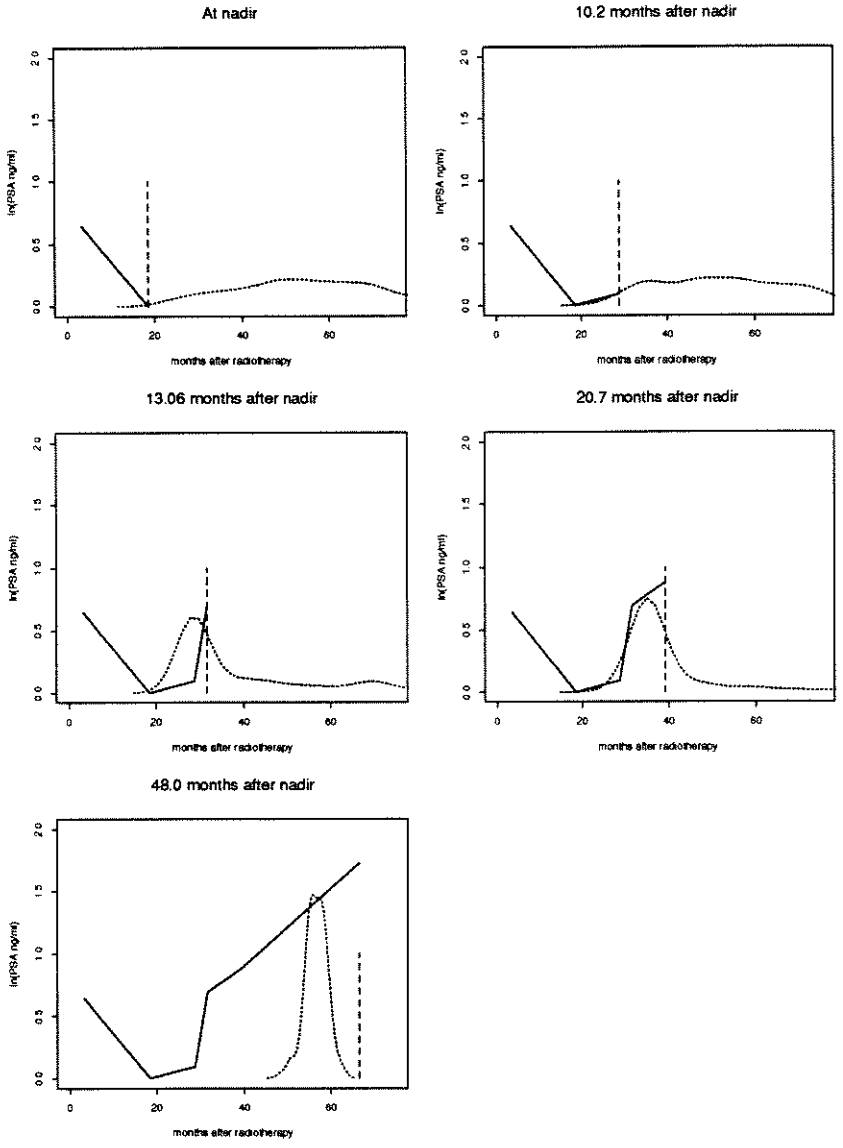

FIGURE 4. Evolution of the posterior distribution of the changepoint for a sample case (898326) for the single-changepoint model. The solid line is the ln(PSA) trajectory, the dotted line is the estimated posterior distribution of the changepoint (arbitrarily scaled) and the dashed vertical line indicates the current time. Local failure was diagnosed for this patient 66.5 months after completion of radiotherapy.

3.2 Two-changepoint model

The two-changepoint model addresses the shortcomings of the single-changepoint model. A new changepoint is introduced to enable the model to capture the initial decrease in the PSA readings after radiotherapy. Once this first changepoint has been encountered, the model is as for the single changepoint: the transformed PSA trajectories are linear with a small slope until recurrence, when the slope becomes markedly positive. Thus the first changepoint represents the end of the transient effects of radiotherapy and the second changepoint marks recurrence of cancer. To help make parameters more distinguishable, we parameterize the model in terms of the slopes of the trajectories rather than incremental slopes:

$$Y_{ij} = a_{oi} + a_i x_{ij} + (b_i - a_i)\,(x_{ij} - t_{i1})^+ + (c_i - b_i)(x_{ij} - t_{i2})^+ + \epsilon_{ij}.$$

Here a_i is the slope until the end of the transient effects of radiotherapy at t_{i1}, b_i is the slope between changepoints and c_i is the slope after recurrence of cancer at t_{i2}. An additional change for our analysis using the two-changepoint model is that here we use ln(PSA + 1) as the response rather than ln(PSA), following Whittemore (1995). This transformation diminishes the effect of very low PSA readings, for which the exact value is typically unreliable.

For this two-changepoint model, we use the longitudinal readings beginning after the completion of radiotherapy and ending with either the last observation or the start of hormone therapy.

Primarily for computational simplicity we used a discrete distribution on the observation times for the two changepoints. Our model is

$$\begin{aligned}
\begin{pmatrix} a_{oi} \\ a_i \end{pmatrix} \Bigg| \begin{pmatrix} \alpha_o \\ \alpha \end{pmatrix}, \Sigma_a &\sim \mathrm{N}_2\left\{ \begin{pmatrix} \alpha_o \\ \alpha \end{pmatrix}, \Sigma_a \right\} \\
\begin{pmatrix} \alpha_o \\ \alpha \end{pmatrix} &\sim \mathrm{N}_2\left\{ \begin{pmatrix} \mu_{\alpha_o} \\ \mu_\alpha \end{pmatrix}, \Sigma_\alpha \right\} \\
\Sigma_a^{-1} &\sim W\left\{ (\rho V)^{-1}, \rho \right\}, \text{ Wishart with scale } (\rho V)^{-1},\ \rho \text{ df} \\
b_i \mid \beta, \sigma_b^2 &\sim \mathrm{N}(\beta, \sigma_b^2)\, I(|b_i| \le B) \\
(t_{1i}, t_{2i}) &\sim \text{Uniform on observation times} \\
c_i &\propto \mathrm{N}(\gamma, \sigma_c^2)\, I(c_i > k) \\
\epsilon_{ij} | \sigma_i^2 &\sim \mathrm{N}(0, \sigma_i^2) \\
\frac{1}{\sigma_i^2} &\sim \text{Gamma}(\lambda_\epsilon, r_\epsilon).
\end{aligned}$$

Because the changepoints t_{i1} and t_{i2} have the discrete uniform distribution on observation times, there is no additional level to the hierarchy for the parameters of this distribution as there was in the single-changepoint model. Furthermore, β and σ_b^2 are specified here rather than being given

distributions. The full specification of the prior distributions is given in Appendix 2.

The Gibbs sampler is again used to fit this model. The complete conditional distributions are again of standard form, except for those for the changepoints. However, by reparameterizing to t_{i1} and d_i, where $d_i = t_{i2} - t_{i1}$, a set of intervals can be defined for each subject so that the techniques used in the single-changepoint model can be applied to both t_{i1} and d_i.

Figure 5 shows the estimated posterior distributions for the population parameters α_o and α. The distribution for α, the mean of the subject-specific slopes before the first changepoint puts most of its mass below zero, as was expected. Figure 6 shows the evolution of the posterior distribution of the second changepoint for the selected case examined for the single-changepoint model. Recall that the distributions for the changepoints are discrete on the observation times. For this subject, at approximately 13 months after the nadir there is substantial mass for an observation other than the most recent. This is an indication that the (second) changepoint has occurred, since otherwise all mass will be placed on the most recent observation time. The mass again concentrates on the most recent observation time in the following panels, however, partly because of our restriction that the slope after the changepoint be larger than $k = 0.15$ and partly because of the preponderance of controls in these data. For comparison, Figure 7 shows the evolution for a second case. Here the posterior distribution of the changepoint is less concentrated on a single observation time.

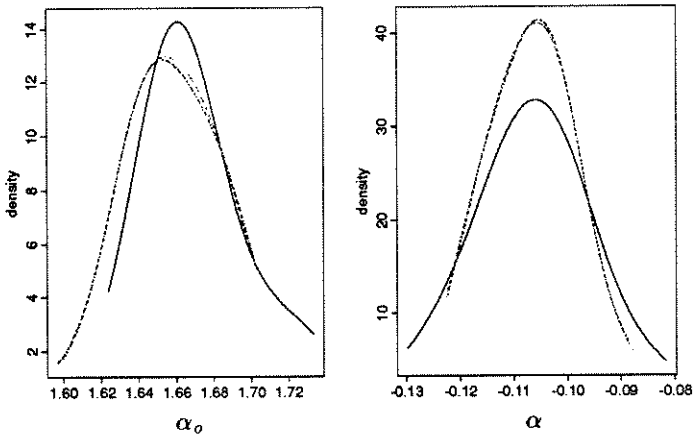

FIGURE 5. Convergence of the posterior distributions for the population parameters in the two-changepoint model. The solid, dotted and dashed lines denote the posterior after 1000, 2000 and 3000 iterations of the Gibbs sampler.

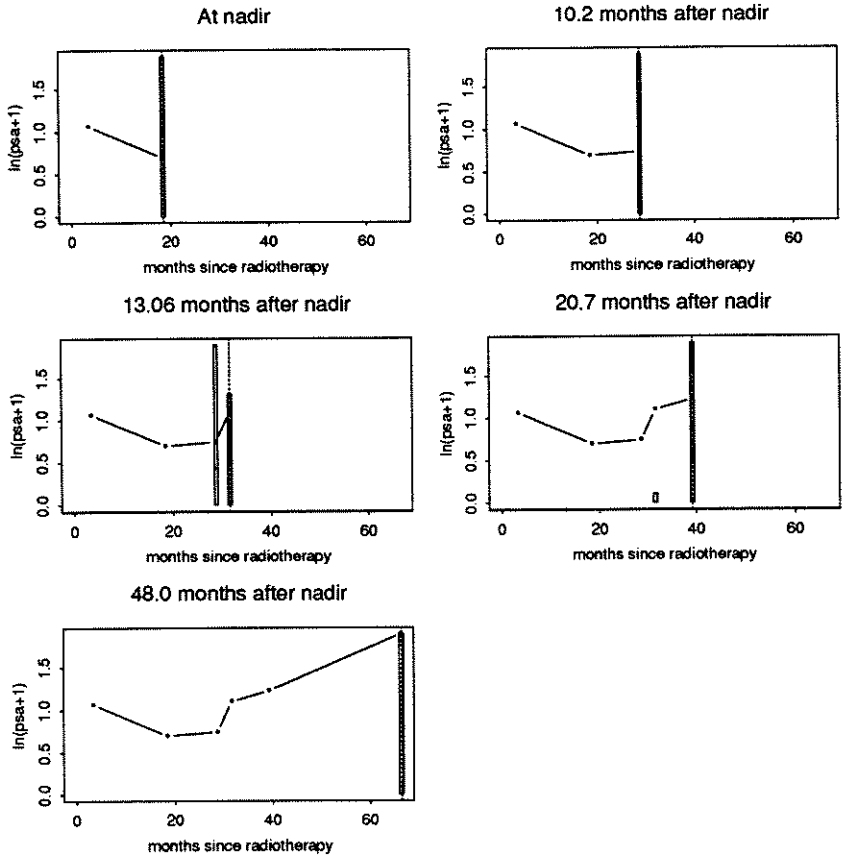

FIGURE 6. Evolution of the posterior distribution of the second changepoint for a sample case (898326) for the double-changepoint model. The solid line is the $\ln(\text{PSA} + 1)$ trajectory, the bars are the estimated posterior distribution (arbitrarily scaled) of the changepoint and the dashed vertical line (inside the last bar) indicates the current time. Local failure was diagnosed for this patient 66.5 months after completion of radiotherapy.

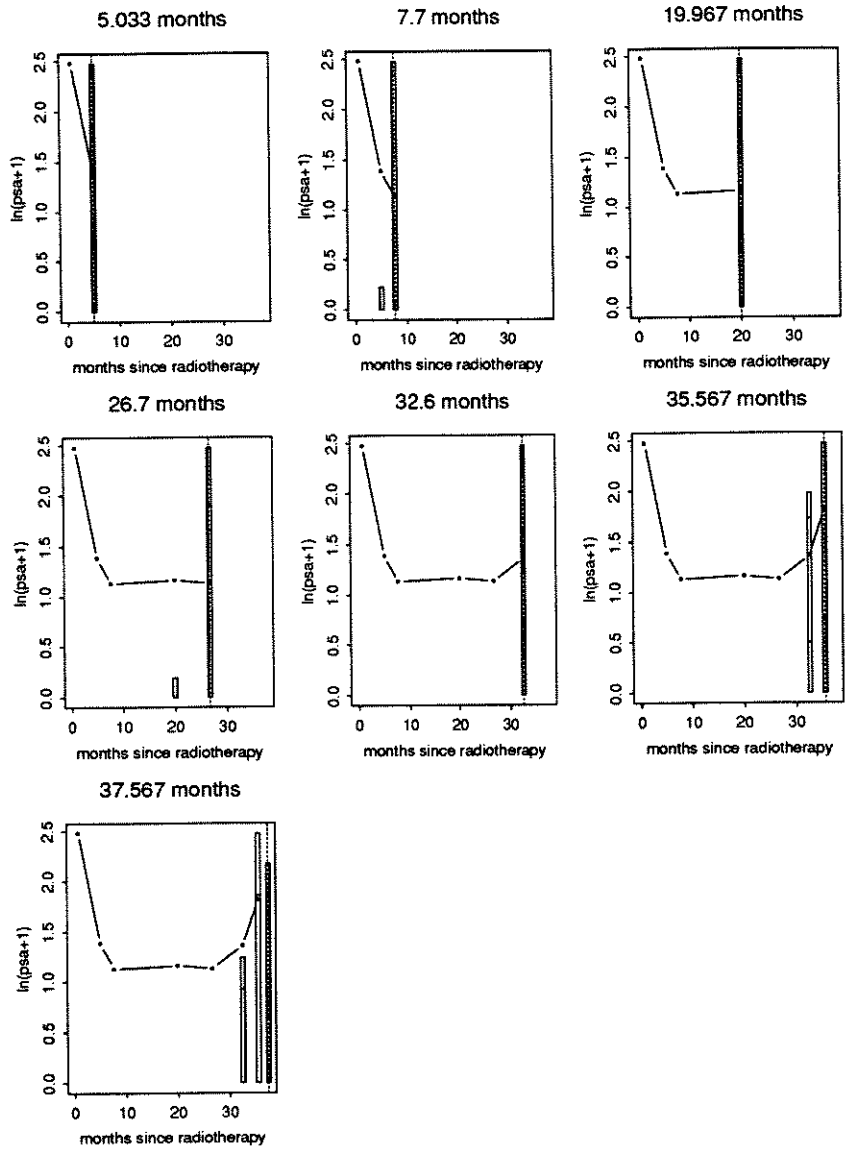

FIGURE 7. Evolution of the posterior distribution of the second changepoint for a sample case for the double-changepoint model. The solid line is the $\ln(\text{PSA}+1)$ trajectory, the bars are the estimated posterior distribution (arbitrarily scaled) of the changepoint and the dashed vertical line (inside the last bar) indicates the current time. Local failure was diagnosed for this patient 37.6 months after completion of radiotherapy.

4 Comparison of diagnostic rules

We compare three diagnostic rules for identifying a recurrence of cancer. The first rule is based on a normal range, whereby any PSA reading above a threshold value (typically 4 ng/ml) is considered to be a positive test result. The second rule is based on the definition of a chemical failure currently used when monitoring patients. In our analysis the definition of a chemical failure is a sustained increase (two or more observations) in PSA levels of more than 50% from the post-radiotherapy nadir reaching 4 ng/ml or more. The formulation we have proposed leads naturally to a third rule. At the time of the current test for a particular subject, we compute the posterior probability that the changepoint has already occurred. For the two-changepoint model, the second changepoint is used. If the probability exceeds some specified cutoff, then a positive result is indicated. We would like to compare these suggested criteria—threshold, chemical failure and posterior probability.

A standard method for comparing diagnostic rules is to use receiver operator characteristic (ROC) curves (Centor, 1991). ROC curves plot sensitivity versus (1−specificity) as the cutoff for the given criterion varies. Specificity is defined as the proportion of non-diseased subjects that test negative, and sensitivity as the proportion of diseased subjects that test positive. These definitions were developed for a single test and do not apply directly to a sequence of tests taken periodically over time. This is because, with longitudinal data, one subject may yield a false positive test result at one observation time and a true positive test result at a later observation time. Murtaugh *et al.* (1991) discussed ROC curves for repeated markers. They classified each subject as either true positive, false positive, true negative or false negative using the series of test results, thus effectively reducing the problem to the single test case.

We define a *specificity rate* for each subject as the proportion of negative test results obtained while the subject is disease-free. Thus, the specificity rate for subject i is

$$\text{spec}_i = \frac{\text{the number of negative test results before recurrence of cancer}}{\text{the number of tests before recurrence of cancer}},$$

where all tests have been performed on subject i. An estimate of the population specificity is obtained by averaging the subjects' rates, so that the information available from different subjects is weighted equally. If the times of recurrence of cancer were known for the cases, then this definition permits the use of data from both cases and controls, incorporating all information available for specificity. Because the times of recurrence of cancer are not known, however, we calculate the specificity rates for only the control subjects (those remaining cancer-free for the time frame of our data) and average only these specificity rates to obtain our estimate of population specificity.

We use a different approach to define sensitivity than we do specificity for two reasons. The first is that sensitivity is time-dependent; a negative result ten years after recurrence of cancer cannot be compared with a negative result within two years of recurrence. Second, a true positive result ends the series of observations. This leads us to define a sensitivity indexed by time, *K-period sensitivity*, where a period is the time between tests. Here, for convenience, we assume the same period for all subjects. K-period sensitivity is the proportion of diseased subjects that test positive *at any time* within K periods after recurrence of cancer. Thus, for K-period sensitivity, a subject is classified as a true positive if he has any positive test results within K periods after recurrence of cancer, and he is classified as a false negative if he has no positive test results within K periods after recurrence of cancer. We use only cases in our calculation of K-period sensitivity and, because the times of recurrence of cancer are unknown, we begin the K-period sensitivity calculation at the post radiotherapy nadir. This substitution of the nadir for the unknown time of cancer recurrence may introduce bias in the sensitivity estimates, but it does not affect the comparability of the detection rules.

Figure 8 shows the ROC curves for the threshold, chemical failure, and posterior probability diagnostic rules based on these definitions of sensitivity and specificity. The heading for each plot gives the value of K that was used for the calculation of K-period sensitivity. The ideal ROC curve would hug the vertical axis, jumping to a sensitivity of one as soon as the specificity became less than one. The posterior probability rule based on the single-changepoint model (dotted line) outperforms the threshold rule (sold line) consistently. The posterior probability rule based on the second changepoint in the two-changepoint model performs better than the single-changepoint rule for short periods (6 and 12 months), but otherwise is comparable to or worse than the threshold rule. This is most likely due to too large a lower bound specified for the slopes of the $\ln(\text{PSA}+1)$ trajectories after the second changepoint. Our model set this lower bound to $k = 0.15$. We are investigating the behavior of the posterior probability rule when this bound is lowered. The chemical failure rule performs consistently better than the threshold rule (at the corresponding specificity), but not as well as the single-changepoint rule.

5 Discussion

It is now becoming standard clinical practice to monitor PSA levels in men after radiation treatment for prostate cancer. Through appropriate modeling of the PSA trajectories in a dynamic fashion, it may be possible to speed the detection of the recurrence of cancer. We analyzed the PSA readings after the transient effects of radiotherapy dissipated using a Bayesian

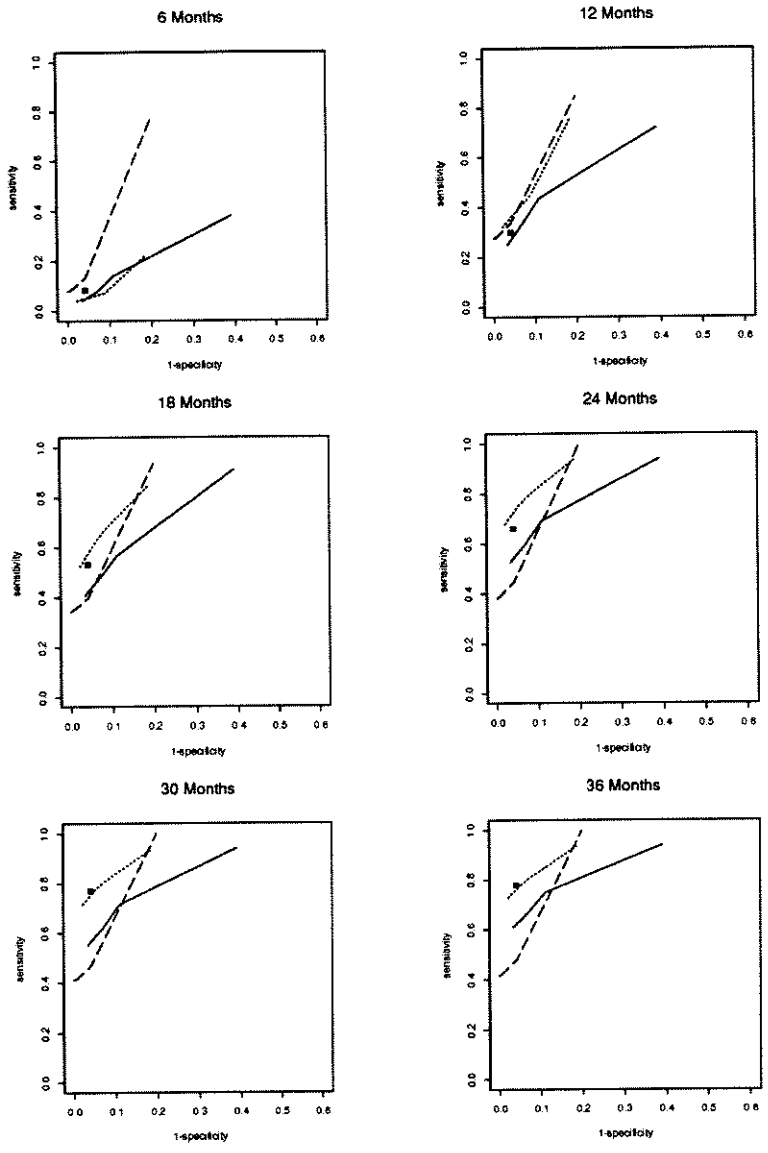

FIGURE 8. ROC curves for the diagnostic rules. The heading for each plot gives the value of K that was used for the calculation of K-period sensitivity. The solid, dotted and dashed lines are for the threshold, single-changepoint posterior probability and second-changepoint posterior probability rules. The square denotes the chemical failure rule.

hierarchical model with a single changepoint. We have also used a two-changepoint model for the post-radiotherapy PSA readings, in which the first changepoint marks the end of the transient effects of radiotherapy and the second indicates cancer recurrence. We defined notions of sensitivity and specificity for longitudinal tests and used ROC curves to compare diagnostics rules for detecting cancer recurrence. The detection rules based on the posterior probability of the changepoint signifying recurrence perform better than a threshold rule, which is common in practice.

Our emphasis is on the posterior distribution of the subject-specific changepoint as it evolves in time. The Bayesian framework is ideal for our purposes because this posterior distribution enables us to answer the most natural and arguably most important question (at least to the subjects): What is the probability that cancer is present now? We have not emphasized the distributions of the population parameters (*e.g.* the mean of the changepoints or the mean slopes of the trajectories). Although they may be extremely important for health policy issues, care must be taken in the interpretation of the population parameters here because their distributions are heavily influenced by the proportion of cases in the data set, which may not be representative of the proportion of recurrences in the general population.

Another benefit of the Bayesian framework is the incorporation of expert opinion or prior information. Our prior distributions contain information drawn from many studies of PSA (especially Carter *et al.* 1992, Pearson *et al.* 1991, 1994, Oesterling *et al.* 1993 and Whittemore *et al.* 1995). Our posterior distributions incorporate this prior information plus that contained in the data we analyzed, and now serve as a database for the analysis of new data. For example, a physician wishing to use our techniques to enhance the monitoring of patients after radiotherapy may use our posterior distributions as the prior distributions for his analyses, updating these as he collects data from his patient population. At some time his posterior distributions may be used as the prior information for the analysis of another physician's patients. In this way the posterior distributions can evolve into a comprehensive database for the behavior of PSA following radiotherapy for prostate cancer.

The two-changepoint model can be generalized easily in a number of ways. We are currently fitting continuous distributions for the two changepoints that permit more flexibility in the modeling. Our choice of discrete distributions here was motivated by the desire to speed convergence of the Gibbs sampler. A major hindrance to convergence is the tendency for the generated changepoints to change very slowly due to high correlations among the changepoints. The speed of convergence can be enhanced by sampling the changepoints simultaneously. The models extend easily to multiple changepoints, also, as discussed from a retrospective viewpoint by Stephens (1994).

The prospective use of these hierarchical models for changepoint detec-

tion may be applied in screening for prostate cancer in healthy men. We are applying these techniques to PSA readings obtained from (*a priori*) healthy men in the Nutritional Prevention of Cancer Trial (Abu-Libdeh *et al.* 1990, Clark *et al.* 1991).

Acknowledgments: We are grateful to Myla Strawderman for suggesting these data to us, to Dr. Howard Sandler for making the data available to us, and to Larry C. Clark for first introducing us to the importance of PSA as a biomarker for prostate cancer. We also thank Bruce W. Turnbull for many helpful discussions and suggestions pertinent to longitudinal modeling, changepoint detection and modeling PSA. We are grateful to the editor and a referee for comments and suggestions leading to an improved paper. This research was supported by grants PHS R01 CA61120 from the National Institute of Health and DMS-9505065 from the National Science Foundation.

6 References

Abu-Libdeh, H., Turnbull, B.W. and Clark, L.C. (1990). "Analysis of multitype recurrent events in longitudinal studies: Application to a skin cancer prevention trial." *Biometrics*, **46**, 1017-1023.

Buhle, E. L. (1995). "Introduction to radiation oncology," The University of Pennsylvania.

Carlin, B.P., Gelfand, A.E. and Smith, A.F.M. (1992). "Hierarchical Bayesian analysis of changepoint problems." *Appl. Statist*, **41**, No. 2, 389-405.

Carter, H.B., Pearson, J.D., Metter, E.J., Brant, L.J., Chan, D.W., Andres, R., Fozard, J.L. and Walsh, P.C. (1992). "Longitudinal evaluation of prostate-specific antigen levels in men with and without prostate disease." *J. Amer. Med. Assoc.*, **267**, No. 16, 2215-2220.

Catalona, W.J., Smith, D.S., Ratliff, T.L., Dodds, K.M., Coplen, D.E., Yuan, J.J., Petros, J.A. and Andriole, G.L. (1991). "Measurement of prostate-specific antigen in serum as a screening test for prostate cancer." *New Engl. J. Med.*, **324**, No. 17, 1156-1161.

Catalona, W.J., Smith, D.S., Ratliff, T.L. and Basler, J.W. (1993). "Detection of organ-confined prostate cancer is increased through prostate-specific antigen-based screening." *J. Amer. Med. Assoc.*, **270**, No. 8, 948-954.

Centor, R.M. (1991). "Signal detection: the use of ROC curves and their analyses." *Medical Decision Making*, **11**, No. 2, 102-106.

Clark, L.C., Patterson, B.H., Weed, D.L. and Turnbull, B.W. (1991). "Design issues in cancer chemoprevention trials using micronutrients: application to skin cancer." *Cancer Bulletin*, **43** No. 6, 519-524.

Cronin, K.A. (1995). *Detection of Changepoints in Longitudinal Data*, Ph.D. Thesis, School of Operations Research, Cornell University.

Garnick, M.B. (1994). "The dilemmas of prostate cancer." *Scientific American*, **270**, No. 4, 72-81.

Gelfand, A.E. and Smith, A.F.M. (1990). "Sampling-based approaches to calculating marginal densities." *J. Amer. Statist. Assoc.*, **85**, No. 410, 398-409.

Gerber, G.S. and Chodak, G.W. (1991). "Routine screening for cancer of the prostate." *J. Nat. Cancer Inst.*, **83**, No. 5, 329-335.

Hinkley, D.V. (1969). "Inference about the intersection in two-phase regression." *Biometrika*, **56**, No. 3, 495-504.

Hinkley, D.V. (1970). "Inference about the change-point in a sequence of random variables." *Biometrika*, **57**, No. 1, 1-17.

Laird, N.M. and Ware, J.H. (1982). "Random-effects models for longitudinal data." *Biometrics*, **38**, 963-974.

Lange, N., Carlin, B.P. and Gelfand, A.E. (1992). "Hierarchical Bayes models for the progression of HIV infection using longitudinal CD4 T-cell numbers." *J. Amer. Statist. Assoc.*, **87**, No. 419, 615-632.

Morrell, C.H., Pearson, J.D., Carter, H.B., and Brant, L.J. (1995). "Estimating unknown transition times using a piecewise nonlinear mixed-effects model in men with prostate cancer.' *J. Amer. Statist. Assoc.*, **87**, No. 429, 45-53.

Murtaugh, P.A., Wieand, H.S. and Schaid, D. (1991). "Application to ROC curve methodology when markers are repeated measures." Paper presented at Spring Meeting of The Biometric Society (ENAR), Houston Texas, March 24-27.

Oesterling, J.E., Jacobsen, S.J., Chute, C.G., Guess, H.A., Girman, C.J., Panser, L.A. and Lieber, M.M. (1993). "Serum prostate-specific antigen in a community-based population of healthy men." *J. Amer. Med. Assoc.*, **270**, No. 5, 660-664.

Page, E.S. (1955). "A test for a change in a parameter occurring at an unknown point." *Biometrika*, **42**, 523-527.

Pearson, J.D., Kaminski, P., Metter, E.J., Fozard, J.L., Brant, L.J., Morrell, C.H. and Carter, H.B. (1991). "Modeling longitudinal rates of change in prostate specific antigen during aging." *1991 Proceedings of the Social Statistics Section of the American Statistical Association*, Washington, D.C.

Pearson, J.D., Morrell, C.H., Landis, P.K., Carter, H.B. and Brant, L.J. (1994). "Mixed-effects regression models for studying the natural history of prostate disease." *Statistics in Medicine*, **13**, 587-601.

Smith, A.F.M. (1975). "A Bayesian approach to inference about a change-point in a sequence of random variables." *Biometrika*, **62**, No 2, 407-416.

Stephens, D.A. (1994). "Bayesian retrospective multiple-changepoint identification." *Appl. Statist*, **43**, No. 1, 159-178.

Whittemore, A.S., Lele, C., Friedman, G.C., Stamey, T. Vogelman, J.J., ans Orentreich, N. (1995). "Prostate-specific antigen as predictor of prostate cancer in black man and white men,' *J. Nat. Cancer Inst.*, **87**, No. 5, 354-360.

1 Prior distributions for the single changepoint model

These are the prior distributions used for the single changepoint model described in Section 3.1.

$$\begin{aligned}
\begin{pmatrix}\alpha_o\\ \alpha\end{pmatrix} &\sim \mathrm{N}_2\left\{\begin{pmatrix}1\\ 0.04\end{pmatrix}, \begin{pmatrix}4 & 0\\ 0 & 1\end{pmatrix}\right\}\\
\Sigma_a^{-1} &\sim W\left\{\begin{pmatrix}4 & 0\\ 0 & 0.1\end{pmatrix}^{-1}, 2\right\} = W((\rho V)^{-1}, \rho)\\
\beta &\sim \mathrm{N}(0.4, 0.01)\, I(\beta \geq 0.2)\\
\frac{1}{\sigma_b^2} &\sim \mathrm{Gamma}(3, 0.06)\\
\tau &\sim \mathrm{N}(70, 25)\\
\frac{1}{\sigma_t^2} &\sim \mathrm{Gamma}(3, 1200)\\
\frac{1}{\sigma_{\epsilon_i}^2} &\sim \mathrm{Gamma}(3, 0.27)
\end{aligned}$$

2 Prior distributions for the two changepoint model

These are the prior distributions used for the two changepoint model described in Section 3.2.

$$\begin{pmatrix}\alpha_o\\ \alpha\end{pmatrix} \sim \mathrm{N}_2\left\{\begin{pmatrix}1\\ 0.04\end{pmatrix}, \begin{pmatrix}5 & 0\\ 0 & -0.2\end{pmatrix}\right\}$$

$$
\begin{aligned}
\Sigma_a^{-1} &\sim W\left\{\begin{pmatrix} 4 & 0 \\ 0 & 0.1 \end{pmatrix},\ 2\right\} = W((\rho V)^{-1}, \rho) \\
b_i \mid \beta, \sigma_b^2 &\sim \mathrm{N}(0, 0.001)\, I(|b_i| \leq 0.05) \\
c_i &\propto N(0.30, 0.008)\, I(c_i > 0.15) \\
\frac{1}{\sigma_{\epsilon_i}^2} &\sim \mathrm{Gamma}(3, 0.27)
\end{aligned}
$$