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

TECHNICAL REPORT NO. 1207

October 1997

Using PSA to Detect Prostate Cancer Onset: An Application of Bayesian Retrospective and Prospective Changepoint Identification

by

Elizabeth H. Slate
Larry C. Clark

The research was supported by a grant from the National Cancer Institute and NSF grant DMS 9505065.

Using PSA to detect prostate cancer onset: An application of Bayesian retrospective and prospective changepoint identification

Elizabeth H. Slate
Larry C. Clark

ABSTRACT Serum prostate-specific antigen (PSA) concentrations are now widely used to aid in the detection of prostate cancer. When prostate cancer is present, PSA levels typically increase. But a number of benign conditions will also cause elevated PSA levels and, conversely, prostate cancer has been diagnosed in the absence of raised PSA.

We analyze one of the most extensive data sets currently available for longitudinal PSA readings, obtained by an historical prospective study of frozen serum samples from the Nutritional Prevention of Cancer Trial (Clark *et al.*, 1996). These data consist of serial readings for over 1200 men taken at approximate six-month intervals over an 11 year period.

We fit a fully Bayesian hierarchical changepoint model to the longitudinal PSA readings. Our objectives include better understanding the natural history of PSA levels in patients who have and have not been diagnosed with prostate cancer and identifying subject-specific changepoints that are indicative of cancer onset. With the goal of accurate early detection, we perform a prospective sequential analysis to compare several diagnostic rules, including a rule based on the posterior distribution of individual changepoints.

1 Introduction

According the American Cancer Society Facts & Figures, prostate cancer is the most common cancer, excluding skin cancer, and the second leading cause of cancer death in American men. Prostate cancer incidence has tripled in the past decade, to an estimated 334,500 new cases in 1997, largely due to increased detection by the use of serum prostate-specific antigen (PSA) measurements. PSA is a glycoprotein produced by the prostate gland that increases with the volume of the prostate. Many studies have demonstrated the utility of PSA for detecting cancer. Catalona *et al.* (1991,

1993) found that combining PSA levels with ultrasound and digital rectal exam (DRE) significantly improved the detection of prostate cancer. They interpreted PSA levels of 4 ng/ml (on the monoclonal scale) as arousing suspicion of cancer. Oesterling *et al.* (1993) concluded from a prospective study that PSA increases gradually among healthy men and, consequently, suggested age-specific normal ranges for PSA. Recently, Carter *et al.* (1997) supported the cutoff of 4 ng/ml for maintaining the detection of curable prostate cancer. However, a single PSA measurement is an imprecise indicator of disease status. Carter *et al.* (1992a, b) and Catalona *et al.* (1991, 1993, 1997) have reported that the proportion of men with prostate cancer who have PSA levels less than 4 ng/ml is 20-35%, while 7-63% of men without prostate cancer have PSA over 4 ng/ml. Similarly Gann *et al.* (1995) and Whittemore *et al.* (1995) reported that the cutoff of 4 ng/ml has estimated sensitivities of 73-75% and specificities of 88-91% for detection of prostate cancer in the next 4-7 years.

A longitudinal series of PSA measurements taken periodically from a subject may lead to the development of diagnostic criteria with much higher sensitivities and specificities. Few published studies have investigated the behavior of longitudinal PSA measurements. Carter *et al.* (1992a, b), Pearson *et al.* (1991, 1994) and Morrell *et al.* (1995) analyzed data obtained from the Baltimore Longitudinal Study of Aging (BLSA). These data consisted of series of PSA readings obtained from frozen blood samples collected approximately bi-annually for 54 men. The series spanned the 7-25 years before determination of prostate disease status, with a median number of approximately 10 readings per person. They fit linear mixed effects models and concluded that the exponential growth rate of PSA is significantly greater for cases prior to diagnosis than for controls, indicating that the rate of change of PSA may be a more sensitive and specific marker for prostate cancer than PSA levels. These investigators also used nonlinear mixed effects models to estimate changepoints in the rate of increase of the PSA trajectories for cases. They found that the transition from slow to rapid increase typically occurred 7.3 and 9.2 years before diagnosis for local and metastatic cancers, respectively.

A second longitudinal study was described by Whittemore *et al.* (1995), based on data available from a screening program run by the Kaiser Permanente Medical Care Program (KPMC). They analyzed PSA trajectories spanning 1–5 years for 320 men, with a median number of readings of approximately 4 per person. These investigators also fit linear and nonlinear mixed effects models and found significant differences in the growth rate of PSA between cases and controls, reporting that the transition from slow to rapid growth occurred about 13–14.5 years before diagnosis. Moreover, Whittemore *et al.* evaluated diagnostic rules based on absolute levels and rates of change of PSA and concluded, surprisingly, that a single PSA measurement was a more sensitive indicator of cancer within the next 7 years than any index of change based on the entire trajectory. This result may

be partially explained by the lack of readings near diagnosis in their data.

In this paper we investigate a much larger set of longitudinal PSA readings obtained from frozen blood samples available for patients in the Nutritional Prevention of Cancer Trial (NPCT). The NPCT is a randomized double-blind cancer prevention trial using a nutritional dose of the essential trace element selenium as the intervention agent. The results of a ten-year follow up were described by Clark *et al.* (1996). The PSA data that we analyze here consist of serial readings for 1210 men spanning up to 11 years with a median number of readings of 4 per person; additional details are given in Section 2. For comparison with the BLSA and KPMC studies, we briefly describe the results of fitting linear and nonlinear mixed effects models to our data in Section 3. However, the focus of this paper is the fully Bayesian hierarchical changepoint model described in Section 4. Our model is similar to that described by Carlin *et al.* (1992) and used by Lange *et al.* (1992) for CD4 T-cell counts, but with a continuous changepoint as in Stephens (1994). As does Stephens, we use our model to retrospectively estimate changepoints, but we also consider our data as arising from an historical prospective study (see, for example, Carter *et al.*, 1997) and dynamically estimate the changepoints. This dynamic implementation enables us to evaluate and compare various diagnostic rules using receiver operating characteristic (ROC) curves adapted for the longitudinal tests; we do this in Section 5. Additional discussion appears in Section 6.

2 The NPCT PSA data

The NPCT investigated the utility of a nutritional dose of selenium for preventing cancer. A total of 1736 patients at high risk of skin cancer were randomized into efficacy and safety arms of the trial, with 1255 of these being males with no prior history of prostate cancer. Blood samples were collected and frozen semi-annually from trial participants. Our PSA data were obtained using the Abbott IMX assay on the frozen samples from the men in the trial. After removing PSA readings from samples taken after diagnosis of prostate cancer, after the start of Proscar, which is known to affect PSA levels, and after the trial unblinding date of Feb. 1, 1996, our data consisted of 6659 readings for 1210 men. Of these men, 85 had biopsy-confirmed diagnoses of prostate cancer and will be called the cases, while the remainder will be called the controls. Note that because of the historical prospective nature of our data, the PSA readings were typically not used to aid diagnosis.

Figure 1 shows representative PSA trajectories from a sample of cases and controls from NPCT data. The horizontal scale in these plots is labeled as "Years Before Reference Date." The reference date is defined as the date of diagnosis for cases and the date of the last recorded reading for

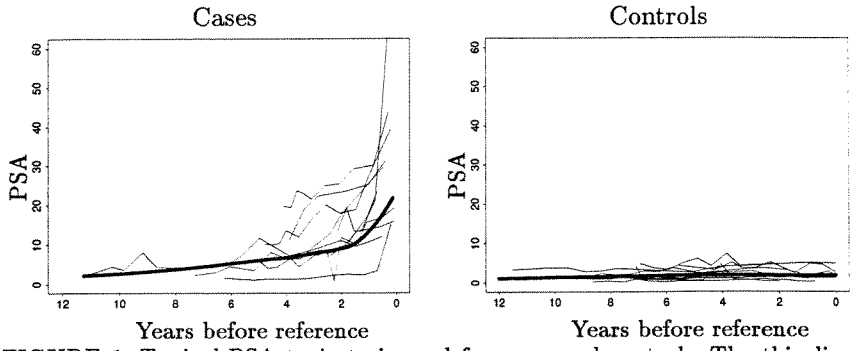

FIGURE 1. Typical PSA trajectories and for cases and controls. The thin lines are linear interpolations of selected trajectories and the thick lines, one for each graph, are the smoothed trajectories for all cases and controls.

	Cases (N = 85)			Controls (N = 1125)		
	Med.	Av.	Range	Med.	Av.	Range
No. Readings	5.0	6.0	(1–19)	4.0	4.5	(1–20)
Time Span (yrs)	2.5	3.1	(0–10.8)	4.5	4.2	(0–11.6)
Age at Ref.	72.0	72.5	(56.7–84.6)	70.5	68.9	(20.7–89.5)
Median PSA	5.3	10.0	(0.9–150.4)	1.1	1.6	(0–25.1)
YPR	3.0	3.3	(0–11.3)	2.5	3.2	(0–12.3)

TABLE 1.1. Summary of the NPCT PSA data. The median (Med.), average (Av.) and range are given for key characteristics as described in the text.

controls (as in Pearson *et al.*, 1994). Also shown are loess (Cleveland, 1988) smooth fits to all 1125 controls and 85 cases. Similar graphs can be found in Figures 1, 2 and 3 of Carter *et al.* (1992a), Figure 1 of Pearson *et al.* (1994) and in Figure 1 of Whittemore *et al.* (1995). It is important to note that there does appear to be a difference in the behavior of the PSA trajectories for the two groups: those for the controls remain relatively flat, whereas the trajectories for the cases appear to show more of an increasing trend. These (apparent) differences in the PSA trajectories according to prostate condition suggest that, indeed, longitudinal PSA readings can be informative about prostate disease status.

Table 1.1 summarizes key characteristics of the NPCT data. Separately for cases and controls, the median, average and range are given for the number of readings per person, time span of the readings per person, age at the reference date, and median PSA per person. Also summarized across all readings for cases and controls is the years prior to the reference date (YPR) at which the blood samples were taken. Although the NPCT data contain readings for many more subjects, the number of readings per subject and the time span of those readings fall between these values for the BLSA and KPMC data.

3 Mixed effects models

The analyses of both the BLSA and KPMC studies included mixed effects models for the PSA trajectories. These models incorporate both population and subject-specific effects and can capture the serial correlation expected in the marker measurements recorded within a subject. The linear mixed effects model is useful for describing the apparent differences in the PSA trajectories for the cases and controls. The nonlinear mixed effects model additionally permits the estimation of changepoints representing the time at which cancer first affects the trajectories for the cases. Here we provide an illustration of each type of model, its application to the NPCT data and a brief comparison to the BLSA and KPMC results.

3.1 Linear mixed effects model

Pearson *et al.* (1994) consider the following model:

$$\begin{aligned} \ln(\text{PSA}_{ij}+1) &= (\beta_0 + b_{0i}) + \beta_1\,\text{Case}_i + (\beta_2 + b_{2i})\,\text{PriorYears}_{ij} \\ &+ \beta_3\,\text{Case}_i \times \text{PriorYears}_{ij} + (\beta_4 + b_{4i})\,\text{PriorYears}_{ij}^2 \\ &+ \beta_5\,\text{Case}_i \times \text{PriorYears}_{ij}^2 + \beta_6\,\text{RefAge}_i + \epsilon_{ij} \qquad (3.1) \end{aligned}$$

Here, i indexes the subject and j indexes the observation within that subject. "Case_i" equals 1 if subject i has been diagnosed with prostate cancer and 0 otherwise, and "RefAge_i" is the age of subject i at the reference date (recall that the reference date is the date of diagnosis for cases and the date of the last reading for controls). Also "PriorYears" denotes years prior to the reference date with a negative sign attached. (This sign convention is needed so that the slope coefficients β_2 and β_3 will be nonnegative in situations such as those depicted in Figure 1.) The population or fixed effects are denoted by βs and are unknown constants to be estimated. The random or subject-specific effects b_{0i} and b_{2i} permit the intercept $(\beta_0 + b_{0i})$ and slope $(\beta_2 + b_{2i})$ to vary across subjects. The vectors $\{(b_{0i}, b_{2i})\}$ are assumed to be independently distributed as multivariate normal with mean $\mathbf{0}$ and arbitrary variance-covariance matrix. The errors $\{\epsilon_{ij}\}$ are modeled as independent and identically distributed normal random variables with constant (but unknown) variance σ^2. The inclusion of the predictor RefAge adjusts for the effect of age at the reference date on the PSA levels (see Oesterling *et al.*, 1993, for results concerning the effect of age on PSA levels). The transformation $\ln(\text{PSA}+1)$ is used by Whittemore *et al.* (1995) and implies linear growth on the log scale, which might be justified both biologically by the exponential growth of malignancies and statistically as an approximate variance stabilizing transformation. (The addition of one is to diminish the influence of extremely small PSA readings.)

We fit model (3.1) to the NPCT data using restricted maximum likelihood (Lindstrom and Bates, 1988, 1990). The population median fits are

shown in Figure 2. The approximately horizontal fits are for control subjects and the increasing curves are for case subjects. The effect of age at diagnosis in this model is to shift the fitted curves upwards for greater ages: for both cases and controls the three lines correspond to ages at the reference date of 50, 65 and 80 years as the PSA levels increase. All estimated coefficients are highly statistically significant with the exception of $\hat{\beta}_4$, corresponding to the quadratic term for controls. The slope of the

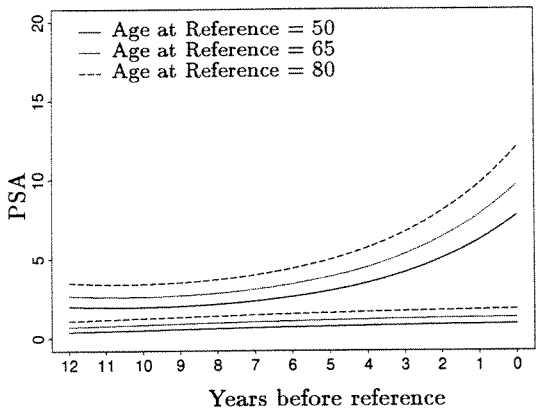

FIGURE 2. The population median PSA trajectories predicted by model (3.1) for the preliminary NPCT data. The approximately horizontal fits are for control subjects and the increasing curves are for case subjects.

$\ln(\text{PSA}+1)$ curves, here either β_2 for controls or $\beta_2+\beta_3$ for cases, is termed the exponential growth rate or "velocity" of the PSA and is of primary interest. For these data the estimated exponential growth rate is 0.01 (se = 0.005) for controls and 0.21 (0.021) for cases. Using models similar to (3.1), approximate exponential growth rates for the BLSA data are 0.037 for controls and 0.20 for cases (se not available from published reports), while the corresponding estimates are 0.03 (0.004) and 0.16 (0.012) for the KPMC data. All three studies show a significant difference between the growth rates for PSA for the cases and controls.

3.2 Nonlinear mixed effects model

Assuming that at one time the cases were cancer-free and hence controls, it is reasonable to seek a changepoint in the PSA trajectories that indicates the time when a malignancy first affects the PSA readings. Similar to Whittemore *et al.* (1995), one model that we have fit is

$$\ln(\text{PSA}_{ij}+1) = \theta_{0i} + \theta_1\,(\text{PriorYears}_{ij} - t_i) +$$

$$\theta_{2i}\,(\text{PriorYears}_{ij} - t_i)\,\text{sgn}(\text{PriorYears}_{ij} - t_i) + \epsilon_{ij}, \qquad (3.2)$$

where sgn is the sign function, θ_{0i}, θ_{2i} and t_i are subject-specific parameters, *i.e.* $\theta_{0i} = \beta_0 + b_{0i}$, $\theta_{2i} = \beta_2 + b_{2i}$ and $t_i = \alpha + a_i$, $\{(b_{0i}, b_{2i}, a_i)\}$ are independent multivariate normal vectors with mean zero, and the errors are independent and normally distributed. The interpretation of the parameters for the trajectory for subject i is that θ_{0i} is the overall level, θ_1 is the average slope, θ_{2i} is half the difference in slopes before and after the changepoint, and t_i is the changepoint. (This particular parameterization was chosen for its numerical properties for detecting changes in slope, see Seber and Wild, 1989, Sec. 9.4.1.) Thus the level, changepoint, and slopes before and after the changepoint are modeled as random. We smoothed the transition between the linear regimes by replacing $\text{sgn}(z)$ with $h(z, \gamma) = (z^2 + \gamma)^{1/2}/z$, where γ is a small positive smoothing parameter (see Seber and Wild, 1989, Sec. 9.4.1).

Figure 3 shows the median maximum likelihood fit of this model to the PSA trajectories for 84 cases from the NPCT data set ($\gamma = 1$). (The two PSA readings of 167.4 and 133.3 for one case were omitted.) The population

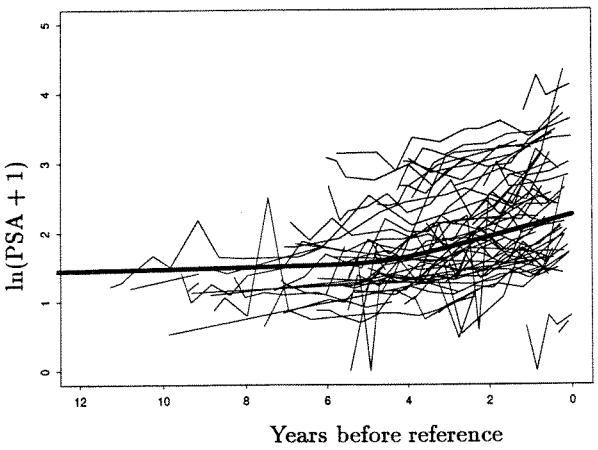

FIGURE 3. The ln(PSA + 1) trajectories for 84 cases from the NPCT data set and the median fit from the model (3.2).

parameters are highly significant in this model. The median changepoint is 4.45 years before diagnosis, and all subjects' estimated changepoints (*i.e.* $\hat{t}_i = \hat{\alpha} + \hat{a}_i$) are before diagnosis. The transition time of 4.45 years before diagnosis is considerably less than the values of 7–9 and 13–14 years obtained by Pearson *et al.* (1994) and Whittemore *et al.* (1995). This difference might be explained by the fact that these data involved a selected population actively participating in a clinical trial with frequent physician visits. Residual plots (not shown) confirm that this model provides an ad-

equate fit to these data.

4 Bayesian hierarchical changepoint model

Although the nonlinear mixed effects model accomodates serial correlation and subject-specific changepoints in the PSA trajectories, there are difficulties associated with these models. For most standard fitting routines, model (3.2) can be fit to the cases only and cannot be used to predict when in the future the changepoint may be encountered by controls. Also, these fitting routines tend to rely on smoothness of the model function and hence require smoothing at the transition point. Furthermore, linear approximations in fitting may create biases in the estimates of the random effects (Breslow and Lin, 1995). A Bayesian model is appealing because it has the benefits of the mixed effects model and enables prediction of changepoints for controls, is easily fit using Monte Carlo Markov chain (MCMC) techniques, and, most importantly, provides an immediate answer to the question "What is the probability that this subject has encountered his changepoint?" For example, if PSA levels react immediately to the underlying prostate condition, then this question is equivalent to "What is the probability that prostate cancer has initiated in this subject?" The answer that the Bayesian model provides is the probability assigned prior to the current time by the posterior distribution of the particular subject's changepoint.

We use a segmented linear regression model for the transformed PSA:

$$\ln(\mathrm{PSA}_{ij}+1) = a_{0i} + a_{1i}\, x_{ij} + (b_i - a_{1i})(x_{ij} - t_i)^+ + \epsilon_{ij}$$

where x_{ij} is the age of subject i at j-th reading, and $z^+ = z$ if $z > 0$ and 0 otherwise. Note that our predictor here is age, rather than years prior to the reference date. This allows incorporation of the dependence on age and obviates the need for the concept of a "reference date" as used in Section 3. The intercept a_{0i}, slope before the changepoint a_{1i}, slope after the changepoint b_i and the changepoint t_i are all random effects.

The full model, including the specification of the prior distributions, is shown below.

$$\begin{aligned}
\begin{pmatrix} a_{0i} \\ a_{1i} \end{pmatrix} \Bigg| \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix},\ \Omega_a \quad &\sim \quad \mathrm{MVN}\left\{ \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix},\ \Omega_a \right\} \\
\begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} \quad &\sim \quad \mathrm{MVN}\left\{ \begin{pmatrix} 1 \\ .02 \end{pmatrix},\ \begin{pmatrix} 100 & 0 \\ 0 & 10000 \end{pmatrix} \right\} \\
\Omega_a \quad &\sim \quad W\left\{ \left[5 \begin{pmatrix} .1 & 0 \\ 0 & .0001 \end{pmatrix} \right]^{-1},\ 5 \right\} \\
b_i \mid \beta,\ \tau_b \quad &\propto \quad \mathrm{N}(\beta,\ \tau_b)\ I(b_i > .08)
\end{aligned} \tag{4.3}$$

$$
\begin{aligned}
\beta &\sim \text{N}(.15,\ 3600) \\
\tau_b &\sim \text{Gamma}(48.0,\ .0133) \\
t_i|\mu_t,\ \tau_t &\sim \text{N}(\mu_t,\ \tau_t) \\
\mu_t &\sim \text{N}(80,\ 0.10) \\
\tau_t &\sim \text{Gamma}(47.0,\ 4700) \\
\epsilon_{ij}|\tau_i &\sim \text{N}(0,\ \tau_i) \\
\tau_i &\sim \text{Gamma}(5.0,\ 0.25).
\end{aligned}
$$

The subject-specific parameters are conditionally independent, as are the within-subject errors. All normal distributions are parameterized in terms of a mean and a precision: thus Ω_a is a 2×2 precision matrix, and τ_b, τ_t and τ_i are precisions. The precision matrix Ω_a follows a Wishart distribution, as parameterized in Press (1982, Chapter 5), for example. The slope after the changepoint, b_i, is constrained to be larger than 0.08, a conservative value consistent with previous estimates of the exponential growth rate for cases. This restriction facilitates the model's distinction between a_{1i} and b_i. Because of the conjugate structure, it is straightforward to fit this model using the Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990), as described in Slate and Cronin (1997), Cronin *et al.* (1994) and Cronin (1995).

The prior information used for this analysis is drawn from the PSA literature, particularly Carter *et al.* (1992a, b), Pearson *et al.* (1994), Oesterling *et al.* (1993) and Whittemore *et al.* (1995). This prior information about the PSA trajectories is best conveyed via the plots in Figure 4, which show the .1, .5, and .9 quantiles of the $\ln(\text{PSA}+1)$ and PSA trajectories implied by the chosen prior distributions. Note in particular, the changepoint is typically at about age 80.

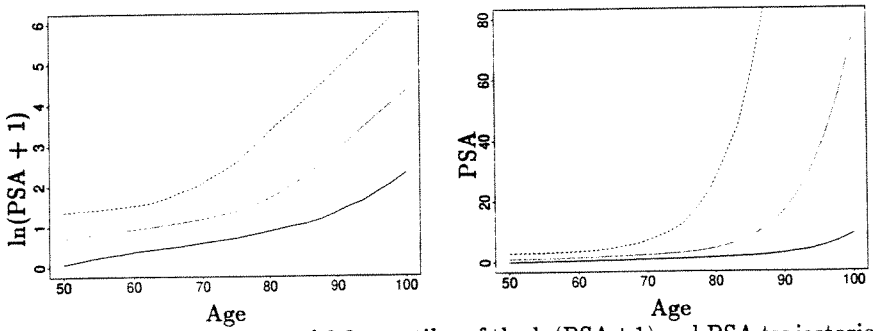

FIGURE 4. The 0.1, 0.5 and 0.9 quantiles of the $\ln(\text{PSA}+1)$ and PSA trajectories implied by the prior distributions given in equations (4.3).

Figure 5 shows the prior and posterior distributions based on the data for the 1210 men for selected population parameters after 100,000, 900,000 and

one million iterations of the Gibbs sampler. These kernel density estimates are based on 1000 sampled points obtained after lagging successive iterates by up to 75. For all population parameters, the posterior distributions are much less diffuse than the prior distributions. The characteristics of these distributions depend on the mix of cases and controls in the data set. The NPCT data is nearly 93% controls. The posterior mode of the mean of the slopes before the changepoint, α_1, is approximately 0.015, whereas the mean of the slopes after the changepoint, β, has mode about 0.14. The parameter μ_t is the mean age at the changepoint (cancer onset) and has mode approximately 87, substantially higher than the prior mean of 80 years. This shift in the distribution of μ_t is not surprising given that controls dominate the data set. Because there are few PSA readings taken at ages above 80 (less than 5%), the posterior for μ_t indicates that many subjects have yet to experience their changepoint.

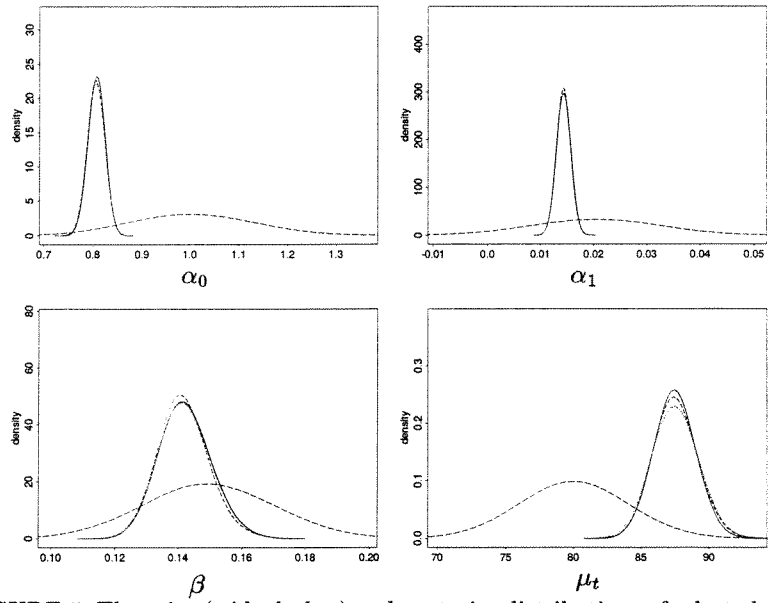

FIGURE 5. The prior (wide dashes) and posterior distributions of selected population parameters after 100,000 (dotted), 900,000 (dashed) and one million (solid) iterations of the Gibbs sampler.

4.1 Retrospective changepoint identification

The posterior distribution of the changepoint for each case subject can be used to estimate the age of cancer onset. Similarly, the posterior distribution of the changepoint for a control subject summarizes all current information about when onset is likely to occur for this subject (and this

distribution may indicate that onset has already occurred despite the lack of diagnosis). Figure 6 shows two sample PSA trajectories, one for a case subject and one for a control, and the corresponding posterior distributions for the changepoints. The variability in the posterior distribution for the changepoint for the control subject is nearly the same as the variability in the prior distribution. The posterior exhibits some right skewness and has mode at about age 90, which is well beyond the range of the PSA data available for this subject. The posterior distribution of the changepoint for the case subject, however, is quite peaked and centered at about age 64. Thus the model fit indicates that this case experienced his changepoint before entering the trial. Note that his first PSA is above 4 ng/ml.

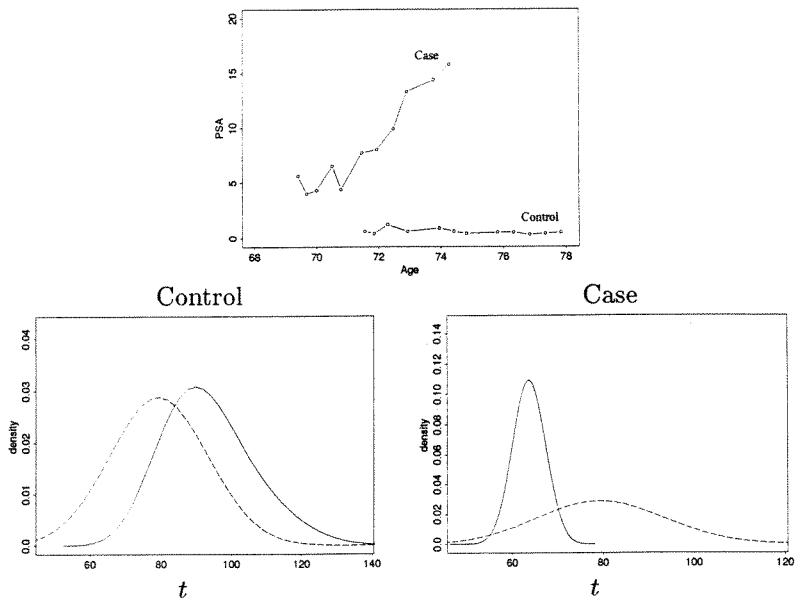

FIGURE 6. The PSA trajectories (top panel) and prior (dashed) and posterior (solid) distributions of the changepoint for selected case and control subjects.

4.2 Prospective changepoint identification

In practice it is of interest to assess the probability that the changepoint has occurred for a subject with each new PSA reading. Figure 7 illustrates the evolution of the posterior distribution of the changepoint for the case subject whose PSA trajectory is given in Figure 6. Model (4.3) was first fit to all 1210 subjects, and then the PSA readings for this case were added to the data set one-by-one and the posterior distribution of this changepoint was computed each time. The probability that onset has occurred by the

age of the current reading is estimated by the proportion of 1000 sampled changepoints (at lags of 75) less than this age and is indicated by "Pr =" in the figure. The "rugs" in the graphs depict the sampled points. The early estimated onset probabilities follow the fluctuation of the initial PSA readings, but climb rapidly once the PSA begins a consistent increase. By the eighth PSA reading there is a probability of approximately 90% that onset has occurred, yet diagnosis was not made for at least 1.5 years later. The modal estimate of the age of onset is about 64 years.

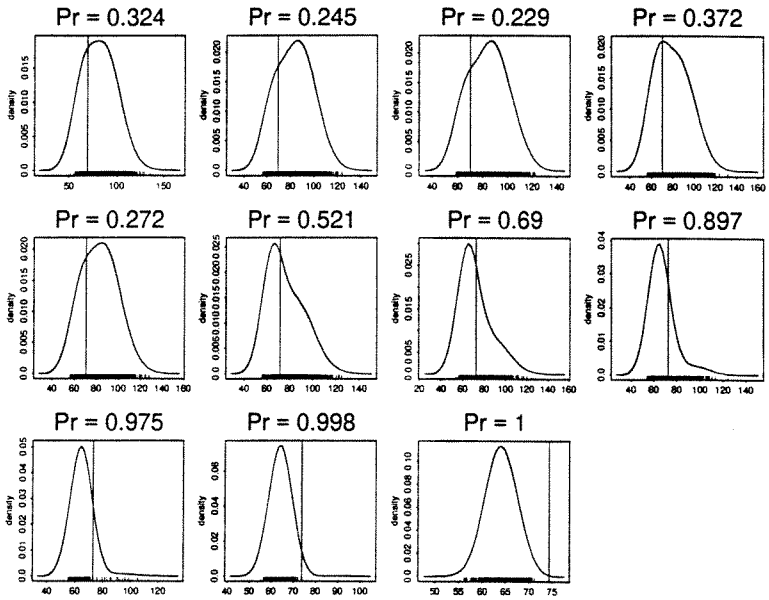

FIGURE 7. The evolution of the posterior distribution of the changepoint for the case subject illustrated in Figure 6. The vertical line is the age of the current PSA reading.

5 ROC Methodology

Figure 7 suggests a diagnostic rule that declares that the changepoint has occurred once the cumulative posterior probability of onset exceeds a specified threshold. The performance of this rule can be compared to that of other diagnostic rules by examining the ROC curves. To do so, the notions of sensitivity and specificity must be extended from the usual context of a one-time test to a series of longitudinal tests. Murtagh and Schaid (1991), Murtagh (1995) and Emir *et al.* (1995) have discussed longitudinal ROC curves. We describe and apply a method similar to that of Slate and Cronin (1997).

Specificity is the probability of a negative test given the subject is disease-free. In the longitudinal setting, we define specificity as an average of individual specificity rates, where the specificity rate for a subject is the probability of a negative test while the subject is free of disease. To estimate specificity, we restrict attention to the control subjects and, for each of these, estimate the specificity rate as the proportion of negative tests. Then the estimated specificity is the average of the estimated rates:

$$\begin{aligned} \widehat{\text{spec}}_i &= \text{proportion of negative tests for control subject } i \\ \widehat{\text{spec}} &= \text{average of the } \widehat{\text{spec}}_i. \end{aligned}$$

The estimated specificity rates are equally weighted when computing the estimate of the overall specificity.

The sensitivity of a test is the probability that the test is positive (indicates that disease is present) given that disease is indeed present. In the longitudinal context, sensitivity depends on the proximity of the test to disease onset. For example, a negative test one month after onset is not comparable to a negative test five years after onset. Moreover, a positive test result ends the series of tests. To account for this time dependence, we use *K-period sensitivity*, which is the probability that a test based on data available K time periods from an origin is positive given that disease is (ultimately) present. The appropriate choice of origin depends on the context and will often vary across individuals. In our PSA setting, we use the time of diagnosis as the origin with K extending backward in time. Thus our estimate of K-period sensitivity is the proportion of case subjects who test positive according to the most recent test that can be formed using (potentially) all readings taken prior to K years before diagnosis.

We compare three diagnostic rules here: the threshold rule, a one-year slope rule, and the posterior probability. The threshold rule gives a positive result if the most recent PSA reading exceeds a cutoff. The one-year slope rule gives a positive result if the increase per year in PSA, as determined from the two most recent readings (typically 6-12 months apart for the NPCT data) exceeds a cutoff. The posterior probability rule gives a positive result for subject i at time t if the posterior probability that t_i is less than t exceeds a cutoff.

Using these definitions of sensitivity and specificity for the longitudinal setting, the ROC curves shown in Figure 8 result for a subset of 54 cases and 54 controls matched on various characteristics for the NPCT data. The curves for the posterior probability rule were obtained by first fitting the model to all 1210 subjects, and then separately adding each of the 108 subjects in the matched data, one reading at a time, and computing the requisite posterior probability. In this data set, the slope rule emerges as markedly inferior. The posterior probability rule performs at least as well as the threshold rule for all specificities of serious interest ($> 80\%$), and the superiority appears greater for the large values of K.

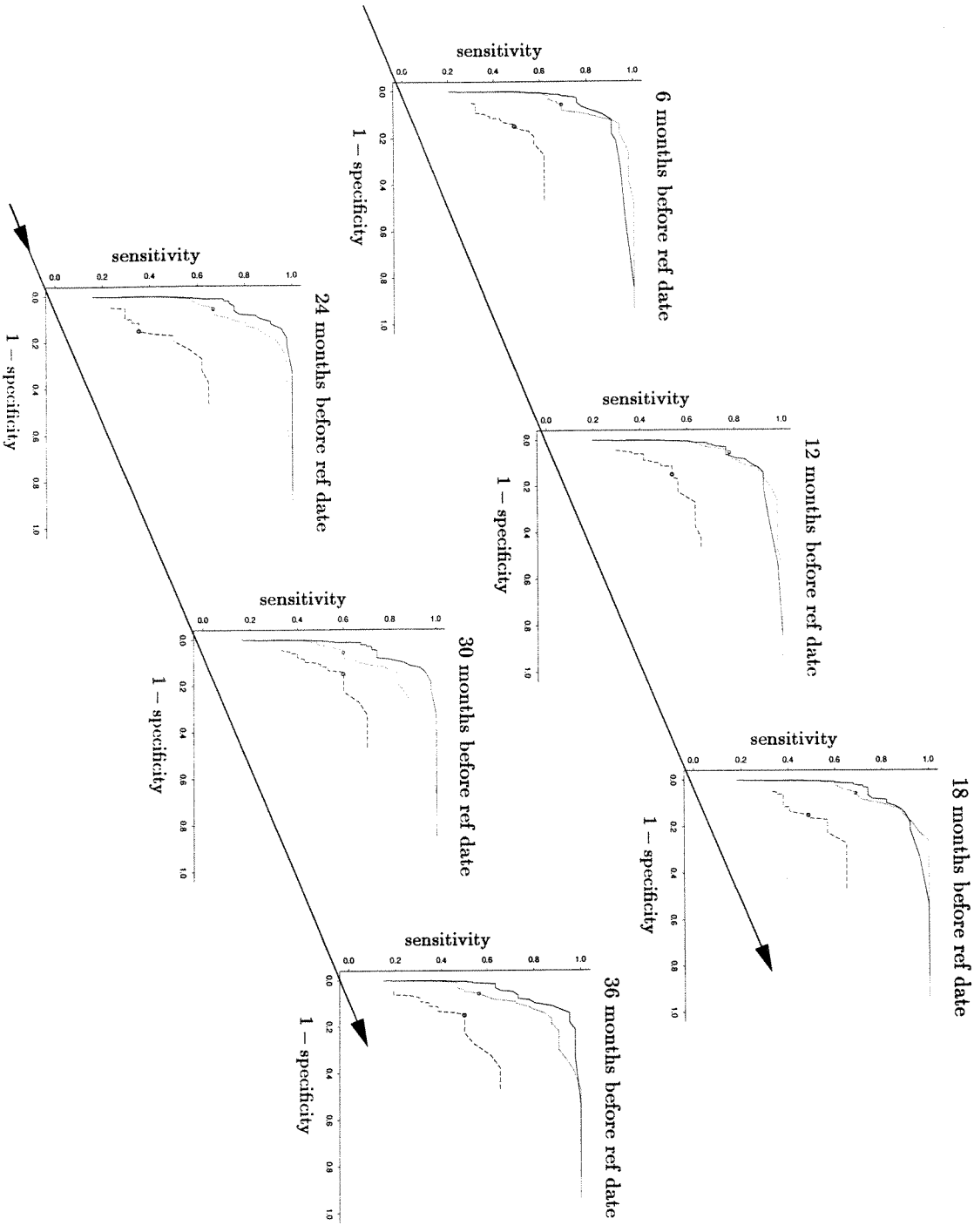

FIGURE 8. The ROC curves for the threshold (dotted line), slope (dashed line) and posterior probability (solid line) rules computed using a sample of 108 subjects from the NPCT data set. The circle on the threshold line corresponds to the usual cutoff of 4 ng/ml. The circle on the slope line corresponds to the cutoff of 1 ng/ml/year.

6 Discussion

For detecting cancer onset, we have emphasized the posterior distribution of the subject-specific changepoint as it evolves in time. The Bayesian framework is ideal here because this posterior distribution allows us to answer, for each subject, the question "What is the probability that cancer is present now?" Furthermore, upon diagnosis, the posterior distribution of the slope after the changepoint, b_i, may be valuable for the selection of appropriate therapy. The size of this slope may distinguish aggressive from slow-growing tumors and hence aid the decision of whether to remove the cancer or to pursue watchful waiting. We are currently compiling tumor grade and stage information for the prostate cancers that have been detected in the NPCT participants. By including these variables as covariates in the Bayesian model, we will be able to assess differences in the PSA trajectories for aggressive and slow-growing tumors.

One potentially confounding factor in our data is the presence of benign prostatic hyperplasia (BPH), an enlargement of the prostate that will also cause elevated PSA levels. Carter *et al.* (1992a, b) showed that PSA levels increase at a faster rate among BPH cases than among controls, but that this rate of increase is nonetheless significantly less than that among cancer cases. Diagnoses of BPH are available for some subjects in the NPCT data, but it is believed much of the incidence has not been reported and that among men in the age group of our cohort, BPH is the rule rather than the exception. In our analysis, we categorized the men diagnosed with BPH as controls.

It is also of interest to use PSA to monitor for recurrence among men who have undergone radiotherapy for prostate cancer. Slate and Cronin (1997) investigated one- and two-changepoint Bayesian models in this context.

Acknowledgments

We thank Beth Jacobs and Dawn Houghton for performing the PSA assays. We also thank Bruce Turnbull for many valuable discussions. This work was supported by a grant from the National Cancer Institute and NSF grant DMS 9505065.

References

Breslow, N.E. and Lin, X. (1995). Bias correction in generalised linear mixed models with a single component of dispersion, *Biometrika*, **82**, 81–91.

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. (1992a). Longitudinal evaluation of prostate-specific antigen levels in men with and without prostate disease. *J. Amer. Med. Assoc.*, **267**, No. 16, 2215–2220.

Carter, H.B., Morrell, C.H., Pearson, J.D., Brant, L.J., Plato, C.C., Metter, E.J., Chan, E.W., Fozard, J.L. and Walsh, P.C. (1992b). Estimation of prostatic growth using serial prostate-specific antigen measurements in men with and without prostate disease, *Cancer Research*, **52**, 2232–3328.

Carter, H.B., Epstein, J.I., Chan, D.W., Fozard, J.L. and Pearson, J.D. (1997). Recommended prostat-specific antigen testing intervals for the detection of curable prostate cancer, *J. Amer. Med. Assoc.*, **277**, 1456–1460.

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.

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. and Ornstein, D.K. (1997). Prostate Cancer detection in men with serum PSA concentrations of 2.6 to 4.0 ng/mL and benign prostate examination: enhancement of specificity with free PSA measurements, *J. Amer. Med. Assoc.*, **277**, 1452–1455.

Clark, L.C., Combs, G. F., Turnbull, B.W., Slate, E.H. . Chalker, D.K., Chow, J., Davis, L.S., Glover, R.A., Graham, G.F., Gross, E.G., Krongrad, A., Lesher, J.L., Park, H.K., Sanders, B.B., Smith, C.L., Taylor, J.R. and the Nutritional Prevention of Cancer Study Group (1996). Effects of selenium supplementation for cancer prevention in patients with carcinoma of the skin: A randomized clinical trial, *J. Amer. Med. Assoc.*, **276**, No. 24, 1957–1963.

Cleveland, W.S., and Devlin, S.J. (1988). Locally-weighted Regression: An Approach to Regression Analysis by Local Fitting, *J. Amer. Statist. Assoc.*, **83**, 596–610.

Cronin, K.A., Slate, E.H., Turnbull, B.W. and Wells, M.T. (1994). Using the Gibbs sampler to detect changepoints: application to PSA as a longitudinal marker for prostate cancer, *Computing Science and Statistics*. **26**, eds. J. Sall and A. Lehman, Interface Foundation of North America, pp. 314–318.

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

Emir, B., Wieand, S., Su, J.Q. and Cha, S. (1995). The analysis of repeated markers used to predict progression of cancer. *Proceedings of the International Biometric Society (ENAR).*

Gann, P.H., Hennekens, C.H. and Stampfer, J.J. (1995). A prospective evaluation of plasma prostate-specific antigen for detection of prostatic cancer, *J. Amer. Med. Assoc.*, **273**, 289–294.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities, *J. Amer. Statist. Assoc.*, **85**, 398–409.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.

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.

Lindstrom, M.J. and Bates, D.M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data, *J. Amer. Statist. Assoc.*, **83**, 1014–1022.

Lindstrom, M.J. and Bates, D.M. (1990). Nonlinear Mixed Effects Models for Repeated Measures Data, *Biometrics*, **46**, 673–687.

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.*, **90**, 45–53.

Murtagh, P.A. (1995). ROC curves with multiple marker measurements. *Biometrics*, **51**, No. 4, 1514–1522.

Murtagh, P.A. and Schaid, D.J. (1991). Application of R.O.C. curve methodology when markers are repeated measures. *Proc. of the International Biometric Society (ENAR)*, Houston, TX, 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.

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, *Statist. Med.*, **13**, 587–601.

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, *Proc. Social Statistics Section of the Amer. Statist. Assoc.*

Press, S.J. (1982). *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*, Robert E. Krieger Publishing, New York.

Seber, G.A.F. and Wild, C.J. (1989). *Nonlinear Regression,* Wiley, New York.

Slate, E.H. and Cronin, K.A. (1997). Changepoint modeling of longitudinal PSA as a biomarker for prostate cancer, in C. Gatsonis, J.S. Hodges, R.E. Kass, R. McCulloch, P. Rossi and N.D. Singpurwall, (eds), *Case Studies in Bayesian Statistics III*, Springer-Verlag, New York, pp. 444–456.

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.