American Journal of Epidemiology Advance Access published February 25, 2008

American Journal of Epidemiology

DOI: 10.1093/aje/kwn003

Practice of Epidemiology

Detecting Trends in Noisy Data Series: Application to Biomarker Series

Carine A. Bellera[1,2], **James A. Hanley**[2], **Lawrence Joseph**[2], **and Peter C. Albertsen**[3]

[1] Department of Clinical Epidemiology and Clinical Research, Institut Bergonié, Regional Comprehensive Cancer Center, Bordeaux, France.
[2] Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Quebec, Canada.
[3] Division of Urology, University of Connecticut Health Center, Farmington, CT.

Received for publication June 30, 2007; accepted for publication January 2, 2008.

It is common to define a change in health status or in a disease state on the basis of a sustained rise (or decline) in a biomarker over time. However, such observations are often subject to important variability unrelated to the underlying biologic process. The authors propose a method to evaluate rules that define an event on the basis of consecutive increases (or decreases) in the observations, given the presence of random variation. They examine how well these rules correctly identify a truly rising biomarker trajectory and, conversely, how often they can recognize a truly stable series or a slowly rising series. The method relies on simulation of realistic, sophisticated data sets that accurately reflect the systematic and random variations observed in marker series. These flexible, empirically based simulations enable estimation of the sensitivity and specificity of rules of consecutive rises as a function of the underlying trend, amount of random variation, and schedule of measurements (frequency and duration of follow-up). The authors illustrate the approach with postradiotherapy series of prostate-specific antigen, where three consecutive rises in prostate-specific antigen indicate treatment failure; the data are described by using a Bayesian hierarchical changepoint model. The method is particularly flexible and could be applied to evaluate other rules that purport to accurately detect upturns (downturns) in other noisy data series, including other medical data or other application areas.

Bayesian hierarchical model; changepoint; Markov chain Monte Carlo; noise; prostate-specific antigen; sensitivity and specificity; simulation; trend analysis

Abbreviations: ASTRO, American Society for Therapeutic Radiology and Oncology; MCMC, Markov chain Monte Carlo; PSA, prostate-specific antigen.

Clinical or biologic characteristics of an individual measured repeatedly are often used to assess a change in health status or in a disease state. Examples include falloffs from a growth curve (weight, height) to indicate a failure to thrive in young babies and, by extension, developmental problems; a sudden change in levels of human chorionic gonadotropin to detect pregnancy; or the depletion of CD4 T-cells as a marker of the progression of human immunodeficiency virus. Similarly in oncology, biomarkers are becoming extensively used to monitor tumor growth or regrowth and thus, by extension, disease onset or progression. Examples include the prostate-specific antigen (PSA) for prostate cancer (1–3), the cancer antigen 125 for ovarian cancer (4), or the carcinoembryonic antigen in colorectal cancer patients (5). Following cancer treatment, the primary aim is to predict clinical recurrence, usually in the form of local or distant relapse. In such cases, a secondary treatment can be initiated, which is usually more effective when provided as early as possible, that is, even before local or distant failure is observed. Biomarkers are thus particularly

Correspondence to Carine A. Bellera, Department of Clinical Epidemiology and Clinical Research, Institut Bergonié, Regional Comprehensive Cancer Center, 229 Cours de l'Argonne, 33076 Bordeaux, France (e-mail: bellera@bergonie.org).

valuable because they can indicate progression (through a simple blood test) even though no clinical progression has been observed yet (usually through more invasive tests, such as biopsies or bone scans). Given the variability of biomarkers, it is thus particularly important to provide rules that accurately detect a rise in true underlying biomarker values.

In this paper, we propose a method to evaluate rules that define events based on consecutive increases (or decreases) in the values of a marker, given the presence of random variability. We illustrate our approach by using postradiotherapy PSA series, where biochemical failure is defined as a recurrence of the cancer detected by rising PSA levels. However, debate is ongoing as to the definition of a rising PSA pattern, and, until 1996, several definitions were used. In 1996, the American Society for Therapeutic Radiology and Oncology (ASTRO) consensus panel proposed guidelines to unify the scientific community concerning the use of a single definition that would standardize the reporting and comparison of treatment outcomes. The panel considered three consecutive PSA rises as an appropriate definition of biochemical failure following radiation therapy (6). However, to date, studies investigating the performance of this rule have focused on its capabilities to predict distant clinical outcomes and have provided discordant findings (7–9); importantly, they took the observed PSA values at face value, which were thus analyzed as if they represented the true PSA concentrations. For example, sensitivity was estimated by the association between a specific distant outcome (presence of metastases, vital status) and whether or not the individual PSA series presented at least three observed PSA rises. The observed PSA level is, however, an amalgam of the unobservable true PSA concentration and random variation (measurement errors and short-term biologic variations unrelated to tumor size), which, similar to other markers, can have a large effect on observed PSA series (10, 11). Thus, PSA variability should be accounted for when looking for a specific pattern.

It is well recognized that a rising PSA concentration precedes clinical failure by several years; for this reason, it is often used as an indication for salvage therapy (12). Thus, given the possible treatment implications, a fundamental point, before one considers how well even a perfectly measured PSA trajectory correlates with clinical outcomes, is how well this rule of three rises correctly identifies a PSA trajectory that is truly rising and how often it can recognize a series that is truly stable, or rising only slowly, for what it is. It is surprising that this first-stage issue has not been evaluated to our knowledge, given that statistical methods have successfully described longitudinal changes in PSA to predict either the onset of prostate cancer (13–15) or the recurrence of the disease following treatment (16, 17).

We propose to evaluate the sensitivity and specificity of the rule of three consecutive rises by simulating data that mimic what is empirically observed. We performed a simulated empirical evaluation of the ASTRO criterion by comparing observed PSA series with the underlying true PSA trajectories. Our estimation approach relied on the simulation of realistic, sophisticated data sets that accurately reflect the systematic and random variations observed in PSA series. First, using a cohort of men treated for localized prostate cancer with radiotherapy, we estimated the underlying true "error-free" PSA trajectories, as well as the variability of the PSA measurements, by fitting a hierarchical changepoint model. Next, we generated realistic PSA series, that is, those that could typically be observed. In order for our simulated series to have the most likely shapes of typical postradiotherapy PSA series, we based our simulation process on the estimates provided by our hierarchical model. We then estimated the sensitivity and specificity of the rule of three consecutive rises by comparing the simulated realistic PSA series with the estimated underlying true PSA profiles. In addition to increasing the effective sample size and providing realistic data, this simulation strategy is particularly flexible because it allows one to evaluate the performance of the decision rule under variable settings: different schedules of measurements, different underlying true marker trends, or different amounts of variation in the measurements. To our knowledge, such an evaluation has not been proposed yet. Moreover, although the rule of three rises is the most commonly used (18), we also evaluated the Houston rule, which has been suggested to outperform the ASTRO criterion (19–21); this criterion is defined as an increase of 2 ng/ml above the PSA nadir.

We also emphasize the flexibility of the Bayesian hierarchical changepoint models, and we show that these models easily account for the multiple complex features of our data including the within- and between-series variabilities, the complex patterns of the series over time, the unbalanced format of the data (different schedules of follow-up), and the nonconstant precision of the measurements.

In this paper, we first describe a population-based cohort of 470 men treated for localized prostate cancer with radiotherapy and estimate the individual true PSA profiles, as well as the PSA variability, by fitting a Bayesian hierarchical changepoint model. Next, we simulate realistic PSA series by using predictions from our hierarchical model. Finally, we estimate the sensitivity and specificity of the rule of three consecutive rises by comparing the generated realistic series with the estimated true PSA profiles.

THE PSA DATA SET

The data were assembled retrospectively from a population-based cohort 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 were men with an initial PSA level higher than 50 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 in the offices of radiation oncologists, medical oncologists, and the Connecticut Tumor Registry, as well as inpatient records. More details are available in Albertsen et al. (22). We based our analysis on men treated with radiotherapy and required each PSA series to have at least a baseline PSA measurement and two subsequent PSA

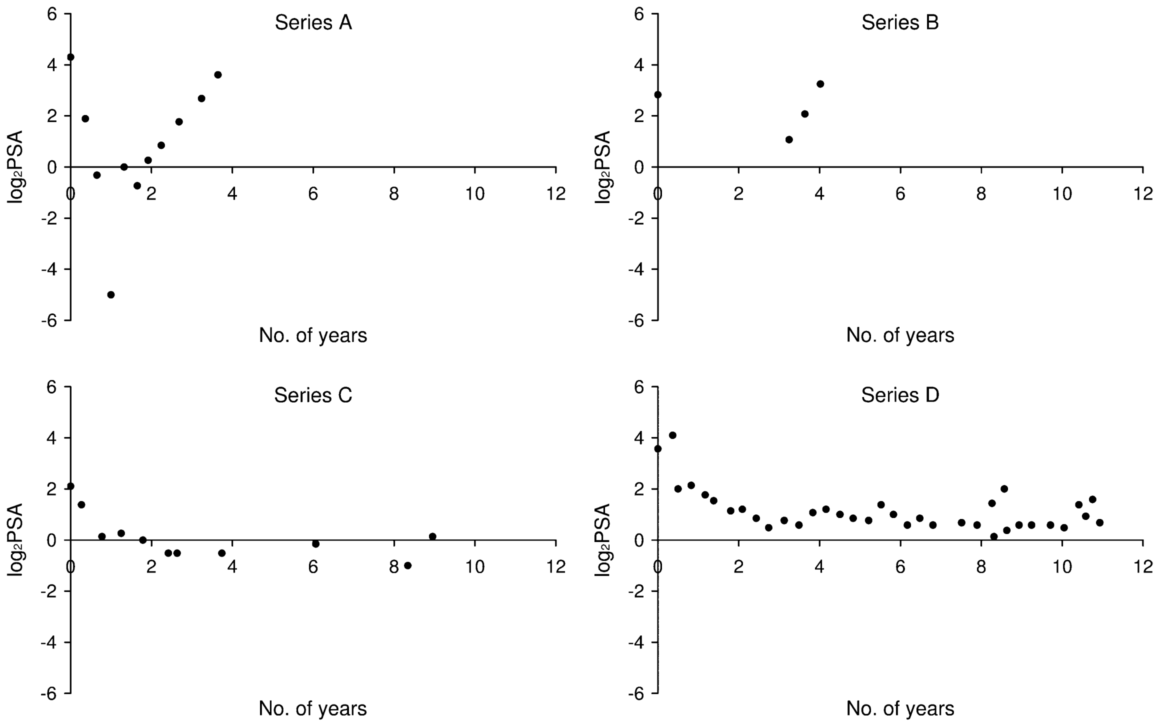

FIGURE 1. For four men (series A–D), $\log_2$PSA concentrations over time since the start of radiotherapy. PSA, prostate-specific antigen.

measurements. In some instances, men can receive a subsequent treatment, usually in the form of hormones. We excluded any PSA measurements taken following hormonal therapy.

A total of 470 series satisfied our conditions and were included in our analysis. The shortest and longest series had three and 36 measurements, respectively; there were nine PSA measurements on average, and the mean follow-up time was 5.7 years.

Following radiotherapy, PSA levels decrease and then start to rise again at variable rates across individuals, although rates are reasonably constant within men, with close to exponential patterns before and after the nadir (23, 24). For this reason, a logarithm transformation is usually applied to obtain a piecewise linear pattern. Moreover, if one uses a base 2 logarithm transformation, then the postnadir $\log_2$-PSA growth rate is equivalent to the number of PSA doublings per year, and its reciprocal corresponds to the PSA doubling time, a variable of particular interest to clinicians.

Figure 1 shows postradiotherapy PSA series over time for four men, plotted on the $\log_2$ scale. The time axis (x) starts at the initiation of treatment. We use the notation $\log_2$PSA to define the logarithm to the base 2 of the PSA concentration. Note the typical bilinear shape of the series: following the start of treatment, the levels drop to some nadir value and then increase again, with important variations in rates within and between series. In addition to this variability, analysis of postradiotherapy data has to account for the presence of a sudden change in PSA concentrations. If radiotherapy is successful, PSA levels reach a nadir and remain low or possibly rise very slowly. A sustained steeper increase usually indicates treatment failure. Given the above characteristics, a Bayesian hierarchical changepoint model appears particularly suited to the analysis of our data.

A BAYESIAN HIERARCHICAL CHANGEPOINT MODEL

On average, men receiving a secondary treatment reach their PSA nadir much sooner, with a steeper postnadir PSA growth rate, than men not receiving such treatment. For this reason, we divided the men into two subgroups according to whether they received a subsequent hormonal treatment. This division enabled us to obtain two relatively homogeneous subgroups, simplifying the fitting of our model.

Figure 2 illustrates a prototypic PSA profile plotted on the $\log_2$ scale for a specific man i. We denote by α_i, β_{1i}, β_{2i}, and τ_i the $\log_2$PSA nadir; the $\log_2$PSA decline rate prior to the PSA nadir (the slope of the first line); the postnadir $\log_2$PSA growth rate (the slope of the second line); and the changepoint (or location of the nadir in follow-up time), respectively. The $\log_2$ scale permits a direct estimate of the individual PSA doubling time PSAdt_i, which is simply the reciprocal of the postnadir $\log_2$PSA growth rate: PSA$dt_i = \frac{1}{\beta_{2i}}$.

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 (25); this model is fully described in Bellera et al.

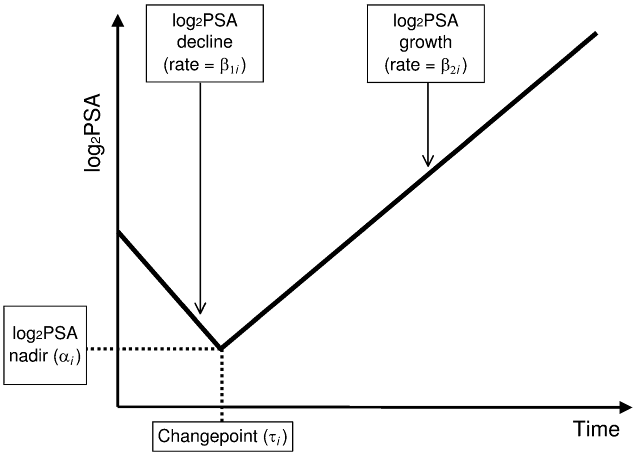
log2PSA decline (rate = β_{1i})
log2PSA growth (rate = β_{2i})
log2PSA
log2PSA nadir (α_i)
Changepoint (τ_i)
Time

FIGURE 2. Individual piecewise linear model, with four individual parameters. PSA, prostate-specific antigen.

(26). At the first level, each individual $\log_2$PSA profile was modeled as in figure 2. Let $\log_2\text{PSA}_{ij}$ be the PSA concentration on the $\log_2$ scale for the jth measurement for the ith man. We assumed that the $\log_2\text{PSA}_{ij}$ were normally distributed, with expected value μ_{ij} and variance σ^2_{ij}:

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

where

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

Our model included another feature of PSA data not accounted for in earlier postradiotherapy PSA studies (16, 17). We expressed the PSA variability σ^2_{ij} as a function of the PSA concentration because interassay coefficients of variation tend to be larger at lower PSA levels (10, 11). We modeled the logarithm of the precision as a linear function of the $\log_2$PSA level: $\log \frac{1}{\sigma^2{}_{ij}} = \theta_1 + \theta_2 \log_2\text{PSA}_{ij}$. Thus, the variance was given by

$$\sigma^2_{ij} = \exp[-(\theta_1 + \theta_2\mu_{ij})], \tag{3}$$

where μ_{ij} is given by equation 2.

We assumed that the individual parameters, α_i, β_{1i}, β_{2i}, and τ_i, were a priori uncorrelated both within and between subjects, although they are related through the likelihood function. The complete model assumes the following distributions (for details, refer to Bellera et al. (26)):

$$\alpha_i \sim N(\mu_\alpha, \sigma^2_\alpha), \quad \mu_\alpha \sim N(0,100), \quad \sigma_\alpha \sim U(0,4),$$

$$\beta_{1i} \sim N(\mu_{\beta_1}, \sigma^2_{\beta_1}), \quad \mu_{\beta_1} \sim N(0,100), \quad \sigma_{\beta_1} \sim U(0,4),$$

$$\beta_{2i} \sim N(\mu_{\beta_2}, \sigma^2_{\beta_2}), \quad \mu_{\beta_2} \sim N(0,100), \quad \sigma_{\beta_2} \sim U(0,4),$$

$$\theta_1 \sim N(0,100), \quad \theta_2 \sim N(0,100).$$

Finally, the prior distribution of the changepoint was a continuous uniform distribution; the range was selected according to prior biologic knowledge and depending on the subgroup of men. Secondary treatment is usually initiated when it is suspected that radiotherapy has failed, indicated by a rising PSA pattern starting within the first 2–3 years following 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 are generally those for whom radiotherapy is successful. In such cases, PSA is still produced by the remaining healthy prostate cells, although in very small quantities. Thus, the PSA concentrations for these men will start to rise later, and at a very slow rate. For this reason, we selected a uniform distribution with a 10-year range for this subgroup: $\tau_i \sim U(0, 10)$.

Estimation was implemented in WinBUGS, a statistical software package that uses Markov chain Monte Carlo (MCMC) to generate random samples from the relevant posterior distributions (27). Additional details on the implementation process, sensitivity analysis, and strategy for model checking are provided in Bellera et al. (26).

GENERATION OF SIMULATED REALISTIC PSA SERIES

In the previous section of this paper, we estimated the true underlying PSA trajectory including the true PSA doubling time for each man in the cohort. In this section, we simulate realistic PSA series, that is, individual PSA series that could typically be observed. In order for these simulated series to have the most-likely shapes of typical postradiotherapy PSA curves, we based our simulation process on our cohort of 470 men. We used the individual predictions provided by each MCMC iteration obtained when fitting our hierarchical model and then added a realistic amount of variability, reflecting the typical PSA variability encountered in real settings.

Following convergence, each iteration of the MCMC process generated one quartet of estimates $(\tilde{\alpha}_i, \tilde{\beta}_{1i}, \tilde{\beta}_{2i}, \tilde{\tau}_i)$ for every man i. Thus, at each MCMC iteration, the ith man's true PSA concentration, $\tilde{\mu}_{ij}$, at every time point j, was given by

$$\tilde{\mu}_{ij} = \begin{cases} \tilde{\alpha}_i + \tilde{\beta}_{1i}(t_{ij} - \tilde{\tau}_i), & t_{ij} < \tilde{\tau}_i, \\ \tilde{\alpha}_i + \tilde{\beta}_{2i}(t_{ij} - \tilde{\tau}_i), & t_{ij} \geq \tilde{\tau}_i. \end{cases} \tag{4}$$

In addition, estimates of the variance, $\tilde{\theta}_1$ and $\tilde{\theta}_2$, were also generated. Therefore, for an estimated true $\log_2$PSA concentration, $\tilde{\mu}_{ij}$, the estimated variance was given by

$$\tilde{\sigma}^2_{ij} = \exp[-(\tilde{\theta}_1 + \tilde{\theta}_2\tilde{\mu}_{ij})]. \tag{5}$$

Using these estimates, we then generated a realistic $\log_2$PSA series. We first selected the number of measurements, j, and thus their timing, t_{ij}, by specifying the duration of follow-up and the frequency of measurements. For every man i, at every time point t_{ij}, we then generated a realistic $\log_2$PSA concentration, $\log_2\text{PSA}_{ij}$, by drawing a value from a normal distribution centered at the estimated true concentration $\tilde{\mu}_{ij}$:

$$\log_2\text{PSA}_{ij} \sim N(\tilde{\mu}_{ij}, \tilde{\sigma}^2_{ij}).$$

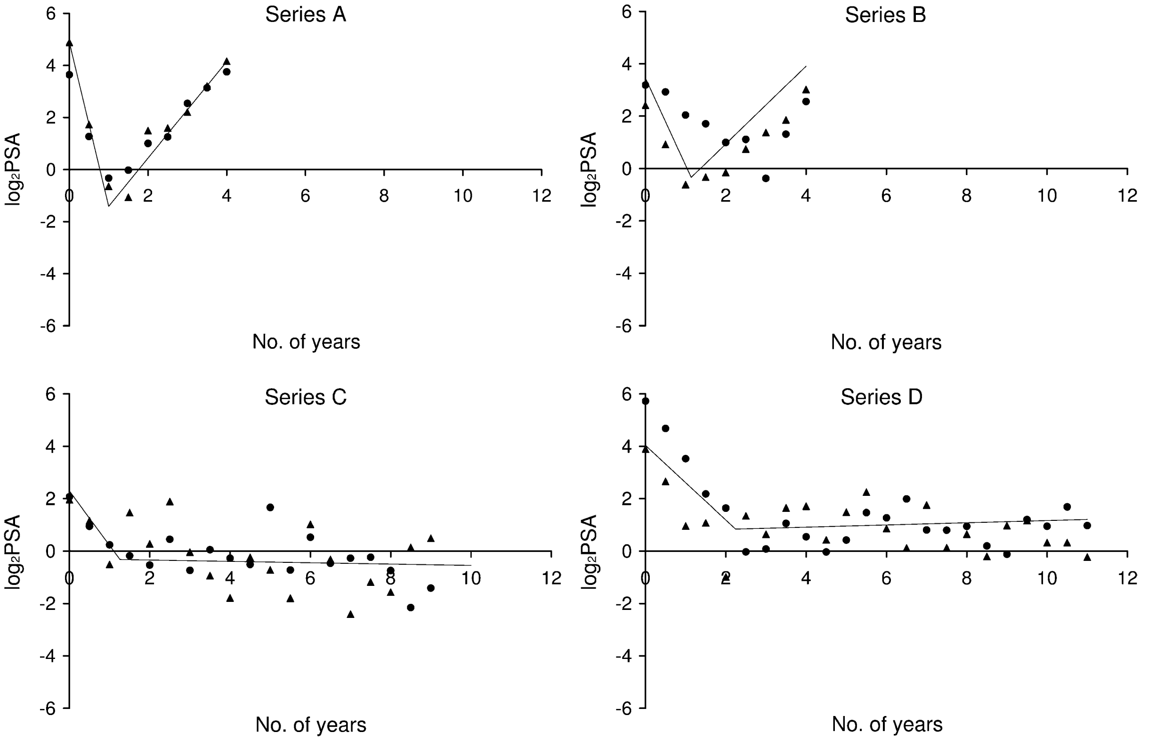

FIGURE 3. Estimated true $\log_2$PSA trajectories (solid line) with two simulated realistic $\log_2$PSA series (circles and triangles) for four men (series A–D). PSA, prostate-specific antigen. The estimated PSA doubling time was shorter than 1 year for series A and B and longer than 10 years for series C and D.

The expected true concentration $\widetilde{\mu}_{ij}$ and the variance $\widetilde{\sigma}^2_{ij}$ were given, respectively, by equations 4 and 5. Thus, one MCMC iteration provided one realistic $\log_2$PSA series for each man.

To have a set of independent realizations of a man's PSA series (i.e., several realistic PSA series), we repeated the same process by using several MCMC iterations. Recall that, following convergence, we ran 10,000 additional iterations for each of the three chains. From these, we kept the last 2,500 iterations per chain. To ensure independence, we retained only those sequences generated at every 50th iteration; this distance was more conservative than the dependence factor suggested by the Raftery and Lewis method (28). Therefore, each chain provided 50 (2,500/50) approximately independent sets of estimates per man. Because we used three chains, a total of $3 \times 50 = 150$ independent quartets $(\widetilde{\phi}_{1i}, \widetilde{\beta}_{1i}, \widetilde{\beta}_{2i}, \widetilde{\tau}_i)$, and thus 150 estimated true $\log_2$PSA profiles, were available for each man, enabling us to simulate a total of $470 \times 150 = 70{,}500$ realistic PSA series.

Each man in the cohort provided one estimated true PSA doubling time and multiple realistic PSA series. As an illustration, consider figure 3; we have represented the estimated mean PSA trajectory (solid line) along with two (of the 150) simulated realistic series for the four men initially presented in figure 1. The estimated PSA doubling time was shorter than 1 year for series A and B and longer than 10 years for series C and D. It is interesting to note that, for series B, one of the generated series satisfied the ASTRO rule (triangles), whereas the other series did not (circles). Finally, note that, although the simulated series and the mean PSA profiles tend to overlap for series A, C, and D, this is not the case for series B. This difference is explained by the fact that, for series B, only four PSA observations were available; given the hierarchical structure of the model, estimation of the mean profile was largely influenced by the estimated hierarchical population parameters.

SENSITIVITY AND SPECIFICITY OF THE RULE OF CONSECUTIVE RISES

Sensitivity and specificity were estimated by comparing the generated realistic series with the associated underlying true PSA profile. Because we were interested in evaluating whether the ASTRO rule adequately detects rising PSA concentrations, we used the underlying true PSA doubling time as the "gold standard" rule.

We first categorized the simulated realistic PSA trajectories according to their associated true PSA doubling time: less than 1 year, 1–2 years, 2–5 years, 5–10 years, more than 10 years, and infinite. We estimated the sensitivity and independently for four subgroups of series depending on the underlying true PSA doubling time (less than 1 year, 1–2 years, 2–5 years, 5–10 years). (Refer to tables 1–3 for more details about the categorization process.) Within each

TABLE 1. Distribution of the 470 estimated true $\log_2$PSA* growth rates

$\log_2$PSA growth rate ($\hat{\phi}_{3i}$)	PSA doubling time in years ($\frac{1}{\hat{\phi}_{3i}}$)	Count		Cumulative	
		No.	%	No.	%
(1; ∞)	(0; 1)	92	19.6	92	19.6
(0.5; 1]	[1; 2)	86	18.3	178	37.9
(0.2; 0.5]	[2; 5)	157	33.4	335	71.3
(0.1; 0.2]	[5; 10)	42	8.9	377	80.2
(0; 0.1]	[10; ∞)	52	11.1	429	91.3
(−∞; 0]	∞	41	8.7	470	100

* PSA, prostate-specific antigen.

subgroup, we estimated the sensitivity of the ASTRO criterion as the proportion of simulated realistic series with three consecutive PSA increases.

We estimated the specificity by using the simulated series from men with a close-to-flat postnadir PSA curve, that is, for the subgroup of series with an estimated true doubling time longer than 10 years. In such instances, the postnadir PSA curves are almost flat, indicating that men can be clinically considered cured. The specificity of the ASTRO criterion was estimated as the proportion of series with two or fewer consecutive PSA rises.

We were also interested in evaluating the Houston rule, defined as any increase of 2 ng/ml above the PSA nadir (the lowest PSA measurement of the follow-up). To estimate the sensitivity, we used the same approach as for the ASTRO rule: we assessed whether the simulated realistic PSA series satisfied the Houston criterion depending on the underlying true PSA doubling time. We estimated the specificity in the subgroup of series with an associated PSA doubling time longer than 10 years as the proportion of series not satisfying this rule.

Finally, the hierarchical model provided an estimate of the PSA variability, through the estimation of θ_1 and θ_2. Thus, in an additional analysis, we also evaluated the performance of the ASTRO rule by assuming different amounts of PSA variation, that is, using different values of θ_1 and θ_2.

RESULTS

The estimated $\log_2$PSA growth rate was positive for 429 of the 470 men, providing a finite PSA doubling time (table 1). Of these men, 377 and 52 had a PSA doubling time respectively shorter and longer than 10 years. Forty-one men had a negative estimated postnadir $\log_2$PSA growth rate, varying between −0.4 and 0; in such cases, the PSA doubling time was assumed to be infinite.

We estimated the sensitivity by using the generated realistic series from the 377 men with a PSA doubling time shorter than 10 years, that is, using 377 × 150 = 56,550 series. For example, from table 1, 92 men had an estimated true PSA doubling time shorter than 1 year. Thus, 92 × 150 = 13,800 simulated series were used to estimate the sensitivity of the ASTRO rule when the PSA doubling time is shorter than 1 year. Table 2 provides estimates of the sensitivity for variable PSA doubling times and duration schedules. For example, when PSA levels are measured every 3 months over a 3-year period, 80.8 percent of the 13,800 series generated from the 92 men with a PSA doubling time shorter than 1 year had three consecutive PSA rises. Thus, under this schedule of measurements, and assuming that the PSA doubling time is shorter than 1 year, the rule of three rises has an 80.8 percent sensitivity. In the same settings, the Houston rule has an 87.1 percent sensitivity.

The specificity for the two rules is reported in table 2. It was estimated by using the generated realistic series of the 52 men with a finite doubling time longer than 10 years (52 × 150 = 7,800 series). When PSA levels are measured every 3 months over a 3-year period, 78.2 percent of the generated series with an estimated PSA doubling time longer than 10 years had at most two consecutive PSA rises, providing an estimated 78.2 percent specificity. In the same settings, the Houston rule had a specificity of 82 percent.

The ASTRO criterion requires three consecutive PSA rises and thus a minimum of four observations. Therefore, we did not evaluate the ASTRO criterion when the follow-up duration was 1 year and the frequency of measurements was 6 months only, since, in such cases, only three observations were available. In addition, the PSA nadir is reached on average the second year after radiotherapy; the PSA curve is therefore decreasing over the first 2 years following radiotherapy. This biologic process explains the early low sensitivities observed for the two rules, and similarly their high specificities.

When the first 2 years were ignored (the PSA decline period), the sensitivity was improved for both rules, longer follow-up periods, and shorter intervals between measurements; these results are intuitively reasonable. Indeed, given a fixed true doubling time and a fixed follow-up duration, a 3-month interval between measurements provides twice as many PSA readings as a 6-month interval between measurements. Therefore, the probability of observing three consecutive PSA rises is higher, which is equivalent to a larger sensitivity. Similarly, given a fixed doubling time and a fixed interval between measurements, an increased follow-up duration provides more PSA observations and thus increases the chance of observing three consecutive PSA rises. Conversely, the specificity decreases with longer follow-up and increases when intervals between measurements are extended.

Overall, the Houston criterion had better classification properties than the ASTRO criterion. For illustration, we constructed a "receiver operating characteristic–like" plot to compare both rules when the true PSA doubling time lies between 1 and 2 years (figure 4). Using a 3-month interval between the measurements, we observed that the Houston rule performs systematically better than the ASTRO criterion.

Finally, and as expected, performance of the rule of three consecutive rises decreased as PSA variability increased (table 3).

DISCUSSION

In this paper, we have proposed a method to evaluate rules that define an event on the basis of consecutive rises (or

TABLE 2. Sensitivity and specificity (%) of the ASTRO* and Houston criteria

	Duration of follow-up (years)						
	1	2	3	4	5	6	7
Sensitivity, given an interval of 3 months between PSA measurements*							
PSA doubling time in years and rule							
(0; 1)							
ASTRO	7.9	50.8	80.8	93.2	97.5	98.7	99.1
Houston	24.7	66.1	87.1†	94.0†	98.8†	99.6†	99.6†
[1; 2)							
ASTRO	1.1	19.0	45.2	66.1	80.3	89.6	94.4
Houston	11.3	37.2	61.8†	78.3†	88.1†	92.8†	96.5†
[2; 5)							
ASTRO	0.8	11.2	27.1	42.9	56.0	66.7	75.3
Houston	4.0	15.4	29.6†	43.6	56.6†	67.4†	75.3†
[5; 10)							
ASTRO	0.7	8.0	20.6	32.8†	43.7†	52.6†	61.2†
Houston	5.0	13.3	21.1†	28.2	35.4	41.5	47.2
Sensitivity, given an interval of 6 months between PSA measurements							
PSA doubling time in years and rule							
(0; 1)							
ASTRO		24.0	75.2	90.5	96.7	98.2	98.8
Houston	18.9	61.1	85.5†	93.3†	98.6†	99.5†	99.6†
[1; 2)							
ASTRO		6.7	34.0	58.9	76.0	86.9	92.9
Houston	5.7	27.2	53.2†	73.6†	85.9†	91.6†	95.8†
[2; 5)							
ASTRO		2.2	14.2	28.1	41.4	53.6	63.4
Houston	1.6	8.5	19.2†	32.6†	46.4†	59.3†	69.0†
[5; 10)							
ASTRO		1.1	8.0	15.6	23.9	32.1†	39.3†
Houston	2.2	6.9	12.3†	18.2†	25.0†	32.0	38.3
Specificity, given 3- and 6-month intervals between PSA measurements							
Interval and rule							
Every 3 months							
ASTRO	99.1	90.6	78.2	67.1	57.2	48.3	41.2
Houston	93.5	87.2	82.0†	77.1†	71.9†	68.1†	63.6†
Every 6 months							
ASTRO		98.0	91.4	84.1	76.3	69.3	63.2
Houston	96.4	91.9	88.3	84.9†	80.5†	76.8†	72.8†

* ASTRO, American Society for Therapeutic Radiology and Oncology; PSA, prostate-specific antigen.

† The criterion with the best sensitivity or specificity.

decreases), given the presence of random variation. We used the example of postradiotherapy PSA series, where three consecutive PSA rises indicate treatment failure; we also evaluated a rule relying on both the nadir and the subsequent trend. We estimated the sensitivity and specificity of both rules as a function of the true marker doubling time, the PSA random variability, and the schedule of measurements.

This study emphasizes the role of simulation models in decision analysis. These models are commonly used to evaluate the benefit of screening and surveillance strategies, such as in ovarian cancer (29, 30). The natural progression of the disease is simulated, and various screening programs are superimposed by using either Monte Carlo methods or Markov modeling. In such cases, the working data sets are

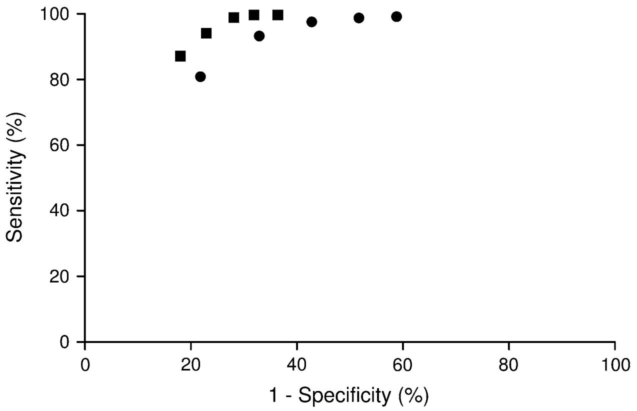

FIGURE 4. Sensitivity and specificity for the American Society for Therapeutic Radiology and Oncology (ASTRO) and Houston criteria when the prostate-specific antigen (PSA) doubling time lies between 1 and 2 years and PSA is measured every 3 months. ●, ASTRO rule; ■, Houston rule. The five circles represent the performance of the ASTRO rule when the duration of follow-up varies from 3 years (leftmost circle) to 7 years (rightmost circle). Similarly, the performance of the Houston rule is shown by the squares.

TABLE 3. Performance (%) of the ASTRO* rule assuming small and large amounts of PSA* variability†

	Duration of follow-up (years)				
	3	4	5	6	7
	Sensitivity				
PSA doubling time in years and amount of variation					
(0; 1)					
Small	86.8	94.8	98.0	99.0	99.2
Large	68.4	87.1	95.2	97.6	98.4
[1; 2)					
Small	82.0	92.8	96.7	98.8	99.1
Large	26.0	48.4	66.7	80.3	88.8
[2; 5)					
Small	60.8	77.8	85.1	88.5	90.0
Large	12.2	24.4	36.5	47.9	57.9
[5; 10)					
Small	27.2	44.0	54.6	62.1	67.1
Large	7.8	15.2	23.2	30.8	37.9
	Specificity				
Amount of variation					
Small	79.3	66.5	57.2	50.6	46.4
Large	91.7	84.5	77.0	70.3	64.4

* ASTRO, American Society for Therapeutic Radiology and Oncology; PSA, prostate-specific antigen.

† The hierarchical model provided estimates of the variance parameters that led to coefficients of variation of 60% and 10% at $\log_2$PSA concentrations of 1 and 4 (26). Additional simulations were performed by assuming first little PSA variation (using variance estimates that led to coefficients of variations of 7% and 1% at $\log_2$PSA concentrations of 1 and 4) and then large PSA variability (using variance estimates that led to coefficients of variations of 90% and 20% at $\log_2$PSA concentrations of 1 and 4).

simulated on the basis of previously reported population parameters (e.g., the incidence of cancer).

Similarly, we relied on a simulation model to evaluate a decision rule aimed at detecting rising marker series; however, our simulation process was more complex because we relied on real data. For each man in our original cohort, these simulations provided 150 prototypic patients from which to simulate data. The advantages were threefold. First, we could create a much more realistic data set than if fewer patients had been used for the original modeling and data simulation. Second, allowing repeats, that is, 150 simulations per man instead of a single one, increased the effective sample size, and thus the precision of the estimates, while still retaining realism. Note that even while using the same prototypic patients' parameters, we drew different values from the posterior distribution of their parameters, as if different patients with slightly different underlying data were used. Third, the simulation enabled us to estimate the performance of the decision rules under variable study settings that may affect the sensitivity and specificity (different schedules of measurement, different underlying true PSA trends, and different amounts of PSA variability). Another issue in implementing the ASTRO rule is unequally spaced observations, which similarly could be generated by using our simulation process. We have shown poor results for this rule with equally spaced observations and would suppose that unevenly observed results through time would decrease sensitivity and specificity because observations closer in time will have fewer real differences between them, elevating the importance of the measurement error.

The Bayesian hierarchical changepoint model was particularly appropriate for describing longitudinal data because it easily accounted for the between- and within-men variabilities, as well as other complex features, such as the presence of a random changepoint, and nonconstant variance. Predicted profiles suggested that the model fit the data well, and the sensitivity analysis confirmed that the estimates provided by the model were not driven by the choice of the prior distribution (26).

Although our model does appear to fit the data well, other choices can be made as in any complex modeling situation. For example, PSA readings have been modeled in different ways. Some authors have modeled the logarithm of the PSA observation (16, 31); others have advocated fitting log(PSA + 1) to diminish the influence of extremely small PSA readings (32). Even though the latter transformation does reduce the variability, it does not remove it entirely. We preferred to fit the logarithm because the PSA variability was a parameter that we wanted to describe, not eliminate, so that we could evaluate the decision rules accordingly. Another modeling strategy involves modeling the raw PSA data by using 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 posttreatment PSA level (33, 34). However, information on subject status is necessary because, in the case of cured patients, the parameter b_2 is set to zero.

Similarly, several strategies are available for modeling the PSA variability. To our knowledge, residual variability has been modeled generally as constant over time in hierarchical changepoint models or, in some cases, as constant within the pre- and postchangepoint phases but different between the two phases (35). We have shown that the model can easily handle another level of complexity by allowing the variance to be a function of the level of the marker. This modeling strategy is sensible given the reported lower precision of the measurement tools at low concentrations. Moreover, it enabled us to evaluate the performance of the ASTRO rule according to different amounts of variation. Our choice to model the logarithm of the precision enabled us to ensure that the resulting estimate is positive, but, alternatively, one can impose range constraints. Finally, we assumed correlated measurement errors, because independence appeared a strong assumption. The use of measurement tools poorly scaled might, for example, result in correlated measurements; such known or unknown factors of variability should not be ignored.

In addition to determining whether a marker is indeed rising, it is of equal interest to detect when the increase takes place. In the case of postradiotherapy PSA data, the ASTRO panel defined the timing of PSA failure as the time midway between the posttreatment PSA nadir and the first of the three rises. Although this rule has been widely criticized because of the backdating that it imposes, Bayesian modeling could provide an alternative definition because it gives posterior distribution for all parameters of the model. Thus, some have proposed using the posterior distribution of the changepoint to estimate the timing of biochemical failure (17).

We estimated the specificity of the rule of consecutive rises from men with a PSA doubling time longer than 10 years, that is, with a close-to-flat PSA curve. Our reported estimates can be compared with rates that would be obtained by assuming a completely flat underlying profile. In such cases, it is possible to use exact estimation methods developed in the context of nonparametric runs tests. These procedures enable one to test whether a sequence of observations comes from a random process and to estimate exactly the probability of observing a specific number of consecutive rises when the underlying pattern is flat, as described by Olmstead (36) and Levene and Wolfowitz (37). Adapting Olmstead's results to our problem, we find that if 13 PSA measurements are taken (corresponding, for example, to measurements every 3 months over a 3-year period), and assuming that the underlying PSA pattern is flat, then by chance alone, the probability of observing at least three consecutive rises is 27 percent (i.e., 73 percent specificity), close to our estimate (78 percent). These results emphasize the impact of random variations on the specificity estimates. If, on the other hand, the observations follow some rising (decreasing) trend, these nonparametric results do not apply and exact methods are particularly difficult to derive analytically, since they imply complex multiple integrations. Our empirically based simulation approach enables estimation of such probabilities.

In conclusion, both our model and our simulation process were particularly flexible to evaluate the measurement properties of a decision rule based on consecutive rises. Our approach can be applied to evaluate decision rules that purport to rapidly and accurately detect upturns (downturns) in noisy series, such as in other medical data, and even other application areas, including, for example, economics, where expansion (or recession) phases are defined as periods during which economic activity tends to trend up (or down) (38).

ACKNOWLEDGMENTS

This work was supported by a doctoral training award from Le Fonds de la Recherche en Santé du Québec (FRSQ; 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.). L. J. is supported by a Senior Scientist Award from the FRSQ.

Conflict of interest: none declared.

REFERENCES

1. Catalona W, Smith D, Ratliff T, et al. Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. N Engl J Med 1991;324:1156–61.
2. D'Amico A, Cote K, Loffredo M, et al. Determinants of prostate cancer specific survival following therapy during the prostate specific antigen era. J Urol 2003;170(pt 2):S42–7.
3. Sartor C, Strawderman M, Lin X, et al. Rate of PSA rise predicts metastatic versus local recurrence after definitive radiotherapy. Int J Radiat Oncol Biol Phys 1997;38:941–7.
4. Rustin GJ, Nelstrop AE, Tuxen MK, et al. Defining progression of ovarian carcinoma during follow-up according to CA 125: a North Thames Ovary Group Study. Ann Oncol 1996;7:361–4.
5. Bast RC Jr, Ravdin P, Hayes DF, et al. 2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology. J Clin Oncol 2001;19:1865–78.
6. American Society for Therapeutic Radiology and Oncology Consensus Panel. Consensus statement: guidelines for PSA following radiation therapy. Int J Radiat Oncol Biol Phys 1997;37:1035–41.
7. Thames H, Kuban D, Levy L, et al. Comparison of alternative biochemical failure definitions based on clinical outcome in 4839 prostate cancer patients treated by external beam radiotherapy between 1986 and 1995. Int J Radiat Oncol Biol Phys 2003;57:929–43.
8. Buyyounouski M, Hanlon A, Eisenberg D, et al. Defining biochemical failure after radiotherapy with and without androgen deprivation for prostate cancer. Int J Radiat Oncol Biol Phys 2005;63:1455–62.
9. Taylor J, Griffith K, Sandler H. Definitions of biochemical failure in prostate cancer following radiation therapy. Int J Radiat Oncol Biol Phys 2001;50:1212–19.
10. Eastham J, Riedel E, Scardino P, et al. Variation of serum prostate-specific antigen levels. JAMA 2003;289:2695–700.

11. 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;267:2215–20.
12. Shipley W, Thames H, Sandler H, et al. Radiation therapy for clinically localized prostate cancer: a multi-institutional pooled analysis. JAMA 1999;281:1598–604.
13. Slate E, Clark L. Using PSA to detect prostate cancer onset: an application of Bayesian retrospective and prospective changepoint identification. In: Gatsonis C, Kass RE, Carlin B, et al, eds. Case studies in Bayesian statistics. Vol 4. New York, NY: Springer-Verlag, 1999:395–412.
14. Pearson J, Morrell C, Landis P, et al. Mixed-effects regression models for studying the natural history of prostate disease. J Am Stat Assoc 1994;13:587–601.
15. 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.
16. Pauler D, Finkelstein D. Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes. Stat Med 2002;21: 3897–911.
17. Slate E, Cronin K. Changepoint modeling of longitudinal PSA as biomarker for prostate cancer. In: Gatsonis C, Hodges J, Kass RE, et al, eds. Case studies in Bayesian statistics. Vol 3. New York, NY: Springer-Verlag, 1999:435–56.
18. McMullen KR, Lee WR. A structured literature review to determine the use of the American Society for Therapeutic Radiology and Oncology consensus definition of biochemical failure. Urology 2003;61:391–6.
19. Vicini F, Kestin L, Martinez A. The correlation of serial prostate specific antigen measurements with clinical outcome after external beam radiation therapy of patients for prostate carcinoma. Cancer 2000;88:2305–18.
20. Horwitz E, Thames H, Kuban D, et al. Definitions of biochemical failure that best predict clinical failure in prostate cancer patients treated with external beam radiation alone: a multi-institutional pooled analysis. J Urol 2005;173:797–802.
21. Kuban D, Thames H, Shipley W. Defining recurrence after radiation for prostate cancer. J Urol 2005;173:1871–8.
22. 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;171:2221–5.
23. 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;71:2031–40.
24. Meek A, Park T, Oberman E, et al. A prospective study of prostate specific antigen levels in patients receiving radiotherapy for localized carcinoma of the prostate. Int J Radiat Oncol Biol Phys 1990;19:733–41.
25. Gelman A, Carlin J, Stern H, et al. Bayesian data analysis. 2nd ed. (Texts in statistical science series). London, United Kingdom: Chapman & Hall/CRC, 2004.
26. Bellera C, Hanley J, Joseph L, et al. Hierarchical change point models for biochemical markers illustrated by tracking post-radiotherapy PSA series in men with prostate cancer. Ann Epidemiol (in press).
27. Spiegelhalter D, Thomas A, Best N, et al. WinBUGS: Bayesian inference using Gibbs sampling, version 1.3. Cambridge, MA: MRC Biostatistics Unit, 2000.
28. Raftery A, Lewis S. How many iterations in the Gibbs sampler? In: Bernardo JM, Berger JO, Dawid AP, et al, eds. Bayesian statistics. Vol 4. Oxford, United Kingdom: Oxford University Press, 1992:763–73.
29. Urban N, Drescher C, Etzioni R, et al. Use of a stochastic simulation model to identify an efficient protocol for ovarian cancer screening. Control Clin Trials 1997;18: 251–70.
30. Hopkins M, Coyle D, Le T, et al. Cancer antigen 125 in ovarian cancer surveillance: a decision analysis model. Curr Oncol 2007;14:167–72.
31. Ankerst DP, Finkelstein D. Clinical monitoring based on joint models for longitudinal biomarkers and event times. In: Crowley J, Ankerst DP, eds. Handbook of statistics in clinical oncology. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC Taylor & Francis Group, 2006.
32. Slate E, Turnbull B. Statistical models for longitudinal biomarkers of disease onset. Stat Med 2000;19:617–37.
33. 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;3:547–63.
34. Taylor J, Yu M, Sandler H. Individualized predictions of disease progression following radiotherapy for prostate cancer. J Clin Oncol 2005;23:816–25.
35. Joseph L, Wolfson DB, Belisle P, et al. Taking account of between-patient variability when modeling decline in Alzheimer's disease. Am J Epidemiol 1999;149:963–73.
36. Olmstead P. Distribution of sample arrangements for runs up and down. Ann Math Stat 1946;17:24–33.
37. Levene H, Wolfowitz J. The covariance matrix of runs up and down. Ann Math Stat 1944;15:58–69.
38. Chauvet M, Piger J. Identifying business cycle turning points in real time. The Federal Reserve Bank of St Louis Review 2003;85:47–62.