Hierarchical changepoint modeling of post-radiotherapy prostate-specific antigen (PSA) series in men with prostate cancer

Carine A. Bellera

Doctor of Philosophy

Department of Epidemiology, Biostatistics, and Occupational Health

McGill University

Montreal, Quebec

February 2006

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Your file Votre référence
ISBN: 978-0-494-25099-0
Our file Notre référence
ISBN: 978-0-494-25099-0

ABSTRACT

Prostate-specific antigens (PSA) help monitor the post-therapy course of prostate cancer. If radiotherapy is successful, levels reach a nadir, and remain low or possibly rise very slowly. A sustained steep increase indicates biochemical failure.

Serial PSA measurements are rarely perfectly monotonic. The American Society for Therapeutic Radiology and Oncology (ASTRO) consensus panel defines biochemical failure as three consecutive PSA increases. I examined the sensitivity and specificity of the ASTRO criterion using simulations of realistic, sophisticated data sets, that accurately reflect the systematic and random variations observed in PSA series.

In a preliminary analysis, I estimated the underlying PSA trajectories in a cohort of 470 men treated with radiotherapy for localized prostate cancer. I exploited the flexibility of Bayesian hierarchical regression models to describe the individual PSA series, each with its own changepoint, and non-constant variance.

The estimates provided by the hierarchical model allowed me to simulate a large set of *true* PSA series. From these, I generated *observed* PSA series: each underlying PSA value was distorted by adding a realistic amount of 'noise'. To evaluate the performance of rules for biochemical failure, including the ASTRO criterion, I then compared the generated observed PSA series to the underlying true PSA series. My results suggest that another rule might outperform ASTRO. This simulation-based approach can be applied to evaluate other rules that purport to rapidly and accurately detect up (down) turns in noisy series, such as in other medical data, and in data series used to monitor economic trends.

Finally, I present a practical charting paper for physicians to record post-treatment PSA values of individual patients. The plotted serial values provide rapid and accurate estimates of the PSA doubling time, without any difficult computations.

ABRÉGÉ

L'Antigène Prostatique Spécifique ou PSA (Prostate Specific Antigen en anglais), permet de suivre l'évolution du cancer de la prostate une fois qu'un traitement est initié. Si la radiothérapie est un succès, la concentration des PSA diminue, puis se stabilise à des concentrations faibles, avec éventuellement un léger taux de croissance. En revanche, une augmentation soutenue des niveaux de PSA indique généralement un échec biochimique.

Les séries de données PSA sont rarement parfaitement monotones. Le critère ASTRO définit l'échec biochimique par trois augmentations successives des concentrations de PSA. J'étudie la sensibilité et spécificité de ce critère en simulant des jeux de données réalistes et sophistiqués, qui reflètent les variations systématiques et aléatoires généralement observées dans les séries de PSA.

Dans une analyse préliminaire, j'estime les trajectoires des PSA dans une cohorte de 470 hommes atteints du cancer de la prostate, et traités par radiothérapie. J'exploite la flexibilité des modèles bayésiens hiérarchiques, afin d'estimer les 470 trajectoires, chacune ayant son propre point de jonction. De plus, je modélise la variance des observations en fonction de la concentration des PSA.

A l'aide du modèle hiérarchique, j'obtiens des estimés que j'utilise afin de simuler un nombre important de *vraies* séries de PSA. Grâce à elles, je génère des séries de PSA *observées* : chacune de ces trajectoires est générée à partir de l'une des 470 trajectoires estimées précédemment. En effet, chaque trajectoire estimée est altérée par l'ajout de variabilité, telle qu'estimée dans le groupe des 470 séries. Je compare ensuite les séries *observées* aux séries *réelles*, afin d'estimer la sensibilité et spécificité de deux critères, incluant le critère ASTRO, en fonction de la fréquence des mesures de PSA, ainsi que du temps de doublement de PSA. Mes résultats suggèrent qu'une autre règle a de meilleures performances que le critère ASTRO.

Cette technique d'estimation basée sur la simulation de trajectoires réelles peut-être facilement appliquée à d'autres règles basées sur la détection rapide de changement de direction dans des séries de données médicales, ou même économiques, sujettes à une importante variabilité.

Finalement, je propose aux médecins de noter les valeurs de PSA de leurs patients sur un graphique que j'ai conçu. Les données de PSA ainsi tracées permettent d'estimer facilement et rapidement leur temps de doublement.

PREFACE

Contributions of Authors

This thesis is the continuation of a research project undertaken by Dr. Albertsen and Dr. Hanley, who were initially interested in post-treatment outcomes in men treated for prostate cancer. After I started the literature review, Dr. Hanley and I determined statistical questions that could be answered in this research project.

I was responsible for the literature review, the statistical analysis and modeling, and the simulation study. I wrote the thesis, including each of the three manuscripts.

Dr. Hanley had responsibility for day to day supervision: he provided advice with regard to the research questions, the statistical analysis, and the interpretation of the results. Dr. Joseph offered his statistical expertise in hierarchical modeling, including the practical implementation of the models. Both Dr. Hanley and Dr. Joseph offered support during the editing of this thesis. Dr. Albertsen supplied the original data set and provided support with regard to medical issues.

Statement of Originality

The doctoral thesis consists of three manuscripts. In the first manuscript, I estimate post-radiotherapy trajectories of prostate-specific antigens (PSA) using hierarchial modeling. Similar models have been used for the last two decades; however I emphasize their flexibility. Besides accounting for the multiple characteristics of longitudinal data, the models accommodate the piecewise-linear pattern of the data, which consists of four random parameters, including a changepoint. In addition, the hierarchical models allow us to model the PSA variability as a function of the true PSA concentration. The second manuscript evaluates the performance of a rule for biochemical failure. Today, this rule is widely used by the medical community; however, only rough empirical evidence is available as to its performance. To my

knowledge, this thesis reports the first formal statistical evaluation of the performance of this rule. Finally, the third manuscript is aimed at a medical audience, and presents a simple tool that will help in the monitoring of their patients. Although the technique is simple, it had not been formally presented as such in the medical literature.

Notes to the reader

As part of the multi-disciplinary Department of Epidemiology and Biostatistics, I have included sufficient statistical material, so that most of the thesis can be understood by both statisticians and epidemiologists. The thesis includes a comprehensive literature review of both statistical and medical concepts. For that reason, this chapter may be considered longer than a traditional literature review. In addition, given the manuscript-based format of the thesis, some background material is covered in this literature review, but also in the introduction of some manuscripts. Finally, I wrote the three manuscript chapters using 'we', and chose 'I' throughout the other chapters.

ACKNOWLEDGEMENTS

First of all, I thank my supervisor, Dr. James Hanley. Dr. Hanley has provided me with an invaluable support through my doctoral studies, and gave me the chance to work on a wonderful thesis project. He also gave me the opportunity to work on other projects, and thus to improve my experience and knowledge of statistics.

I thank my parents, Salima and Jean-François, for their permanent guidance, encouragements, and trust in me. Although the Atlantic Ocean was between us for years, they have been constantly present to support, and push me always further. I also thank my young sister, Camille, for her smile and patience.

I thank my boyfriend, Grégory, for his patience, support and trust in me. His day to day encouragements made this thesis much easier to go through.

I thank my thesis committee: Dr. Lawrence Joseph and Dr. Peter Albertsen. Dr. Joseph gave me valuable statistical advices and was always available to help. Dr. Albertsen offered his inspiring data set, and provided guidance on medical issues.

I thank the Department of Epidemiology and Biostatistics for the opportunity to study in a stimulating environment. I thank Dr. Stan Shapiro and Dr. Jean-Paul Collet, who provided their valuable supports, and gave me the opportunity to work on interesting research projects. I also thank Marie-Eve Beauchamp, a graduate student, who was very helpful to me on several occasions.

Finally, I would like to thank the people and committees who provided me with financial support during these doctoral studies: Dr. James Hanley (2002/2003, 2003/2004, 2004/2005), Dr. Peter Albertsen (2003/2004, 2004/2005), the McGill Faculty of Graduate Studies and Research (2002/2003, 2003/2004), the McGill University Health Center (MUHC, 2004/2005) and the Fonds de Recherche en Santé du Québec (FRSQ, 2005/2006).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
Introduction

1.1 Rationale

Prostate cancer is the number one cancer threat to Canadian men. In 2005, an estimated 19,000 new cases of prostate cancer will be diagnosed, and 4,300 men will die from the disease [1]. The options for the management of prostate cancer include watchful waiting, potentially curative treatment (radiotherapy or surgery), or palliation (hormone therapy or chemotherapy).

Prostate-Specific Antigen (PSA) is a protease naturally produced by the prostate, and by the tumor in men with prostate cancer. Normally this antigen is at a very low level in the blood stream, but it can be persistently elevated in men with prostate cancer or recurrence of the cancer. The PSA test is used for screening and monitoring of the disease. The higher the PSA level the more likely is the presence or recurrence of prostate cancer. Following treatment, a rapid PSA doubling time is prognostic of clinical failure, usually defined as local recurrence, incidence of metastases, or death. However, independently of tumor recurrence, an observed rise in PSA may simply be due to random fluctuations, and a seemingly stable series may hide a true increase. A study conducted in healthy men, showed for example, that nearly half of the men who had one abnormal PSA level subsequently had a normal level [2].

In this research thesis, I focus on *PSA series following radiation treatment in men treated for localized prostate cancer.* If radiotherapy is successful, PSA levels decrease substantially to a nadir value, and remain at low levels with possibly a very slow increase. A subsequent steeper increase usually indicates treatment failure, and depending on the man's wishes, his physician may start hormone-withdrawal

therapy. The best timing of any secondary treatment is a compromise between wishing to delay the onset of clinically evident metastases, the side effects of the hormones, the possibility that the cancer will eventually become resistant, and that the man may die of another disease before these would become bothersome. It is therefore important to detect true treatment failure as quickly as possible, with few false positive and false negative results.

Biochemical failure is defined as a recurrence of the cancer, detected by rising PSA levels. In 1996 the American Society for Therapeutic Radiology and Oncology (ASTRO) consensus panel proposed guidelines to unify the scientific community on the use of a single definition that would standardize the reporting and comparison of treatments and end results. The ASTRO rule defines three consecutive rises in PSA as biochemical failure. This rule is widely used but has undergone limited formal evaluation. Past studies focused on its clinical validation, and its capability to predict a clinical outcome. However, a fundamental point, before one considers how well even a perfectly measured PSA trajectory correlates with clinical outcomes, is how good the ASTRO rule is at correctly (and quickly) identifying 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. To the best of my knowledge, such numerical, or statistical, evaluation has not been examined yet.

In 2003, McMullen et al. reported that only two thirds of the peer-reviewed English published articles in 1999-2000 used the ASTRO definition [3]. *Recognizing the controversy surrounding the use of the ASTRO definition*, the authors were not surprised to find that a significant minority of investigators chose not to use the ASTRO definition. This lack of enthusiasm for the use of a common criterion might be explained by the lack of a formal evaluation of the ASTRO criterion, and the possibility that other rules might outperform it. It is surprising that there has not been a statistical evaluation of the ASTRO criterion, given that several statistical

analyses have successfully described post-radiotherapy profiles, in particular, for the prediction of time to prostate cancer recurrence. I believe that there is a need for a formal evaluation of the ASTRO criterion, which is amenable using real data, appropriate statistical modeling, and simulations.

1.2 Objectives

My thesis will address three issues, whose results are of interest to both statisticians and clinicians; each issue is presented as a separate manuscript.

The first part of the thesis focuses on the precise estimation of post-radiotherapy PSA profiles. I demonstrate the considerable flexibility of a Bayesian hierarchical model to estimate longitudinal PSA profiles. The essence of hierarchical modeling is the simultaneous estimation of the individual and population PSA profiles, allowing for the borrowing of strength; the information provided by the entire set of PSA series is used to strengthen the estimation of any individual profile. The hierarchical model accurately estimates both individual and population PSA profiles, by simultaneously accounting for between and within individual variability. In addition, the model allows me to express the PSA variance as a function of the PSA concentration, a feature which is particularly appropriate given that the variability is known to be higher at lower concentrations. Finally, this approach permits the modeling of complex patterns over time, such as the typical piecewise linear pattern of the post-radiotherapy PSA series. Following radiotherapy, the PSA levels decrease to a nadir level, and then increase again. I can estimate the four random parameters of the post-radiotherapy PSA profiles at both the individual and population levels: the post-radiotherapy PSA nadir, the timing of the nadir, the number of PSA halvings preceding the PSA nadir, and the number of PSA doublings following the PSA nadir. I illustrate the practicality of such a Bayesian hierarchical model, by estimating PSA trajectories following radiotherapy in a cohort of 470 men with localized prostate

cancer. I then use these estimates to investigate possible associations with patients' pre-treatment characteristics.

In the second and most original topic of this thesis, I evaluate the properties of the ASTRO rule, as well as an alternative rule that has been suggested to outperform it. Up to date, studies investigating the performance of the ASTRO criterion examined its *clinical validation*, and thus focussed on the capabilities of the criterion to predict a distant clinical outcome. Instead, I focused on the *numerical validation* of the rule, and present the first assessment of its short-term performance in truly identifying PSA rises (sensitivity), and in reassuring those whose PSA is truly not rising, or rising so slowly as to not cause trouble within the man's life expectancy (specificity). The originality of the approach relies on the simulation of a realistic, sophisticated data set, by first fitting an appropriate model to real PSA series, and then simulating new data from these models. Data are simulated using the results of the first manuscript, specifically using the outputs provided by the hierarchical modeling of the serial PSA's in the cohort of the 470 men. These 'empirically-based simulations' are particularly flexible and allow me to evaluate any rule that relies on PSA profiles, and any schedule of measurements.

The motivation for the last manuscript arose while I was exploring the medical literature related to the monitoring of treatment outcome. It appeared to me, that often, clinicians estimate the PSA doubling time using only the last two PSA values. I present a simple and practical tool that clinicians can use to monitor PSA levels over time, to compute doubling times without requiring a calculator or formula. While the rest of the thesis focusses on PSA values following radiation treatment only, I show that this recording technique applies not only to post-radiotherapy series, but also even more naturally to post-surgery PSA series, and series under watchful waiting.

I begin this thesis with a literature review of the clinical and statistical themes of this thesis. Each objective is then addressed in a separate manuscript. Some manuscripts are followed by a technical chapter, providing additional details not included in the manuscript, but relevant for the thesis. Finally, I give a general conclusion, and discuss the clinical and statistical implications of this research.

CHAPTER 2
Literature Review

This literature review covers medical and statistical materials. In section 2.1, I present basic medical concepts. I describe prostate cancer, prostate-specific antigens (PSA), and PSA failure. The next sections cover statistical issues that will be dealt with in the thesis: Bayesian inference (section 2.2), hierarchial modeling (section 2.3) and changepoint modeling in a Bayesian framework (section 2.4). In section 2.5, I present an overview of longitudinal studies of prostate-specific antigens in the literature.

2.1 Prostate-specific antigen

2.1.1 Prostate cancer

The prostate is a roughly spherical gland in men about the size of a walnut. It is located below the bladder and in front of the rectum, and surrounds part of the urethra, the tube carrying urine from the bladder. The prostate produces some of the fluid for semen, which transports sperm during the male orgasm. Prostate cancer develops from cells of the prostate gland, and is usually a slow-growing disease. A localized cancer can take many years before the disease becomes life threatening without treatment. On the other hand, some prostate cancers can grow and spread quickly with few, if any, symptoms in the early stages.

In general, cancer cells have two characteristics: uncontrolled growth and a loss of differentiation. The loss of growth regulation can mean that cells are dividing too

Table 2–1: Incidence and mortality rates for three most common cancers

Region	Incidence ASR	Mortality ASR
North America		
Prostate	119.9	15.9
Lung	61.2	48.7
Colon and rectum	44.4	15.3
Western Europe		
Prostate	61.6	17.5
Lung	50.9	46.2
Colon and rectum	42.9	19.0

ASR: Age standardized (World) rates, per 100,000. Source: IARC [4].

quickly, or that too few of them are dying as a normal part of the cell cycle of life. The term 'loss of differentiation' indicates that the cell loses its normal structure and function. Prostate cancer begins as a few abnormal cells, and grows over the years. It becomes clinically detectable by rectal examination when it is the size of a sugar cube.

Prostate cancer is the second most deadly form of cancer in Northern America after lung cancer, and the third one in Western Europe after lung cancer, and cancers of the colon and the rectum (see table 2–1). It is the most common form of cancer by incidence in both regions before lung cancer. Although prostate cancer is more common than lung cancer, the potential years of life lost due to prostate cancer are four times lower than for lung cancer, reflecting lower mortality rates for prostate cancer and the older ages at which men develop and die from this disease. The mortality rates for prostate cancer are close across the two regions (15.9/100,000 in North America, and 17.5/100,000 in Western Europe), however, the incidence rates are very different (119.9/100,000 in North America, and 61.6/100,000 in Western Europe). These discrepancies are partly explained by the differing screening procedures across regions, which tend to be more aggressive in North America, especially in the United States. In Canada, an estimated 19,000 new cases will be diagnosed in 2005 and 4,300 men will die from the disease [1].

Prostate cancer incidence increases almost exponentially with age and most cases are diagnosed in men 65 years or older. Other risk factors include family history, high levels of testosterone, and ethnicity (highest rates are in African-American, and lowest rates in Eastern Asian). Modifiable risk factors such as smoking status, socio-economic status and diet have shown discordant findings, and are still under study, as are obesity and inactivity [5].

Screening and diagnosis

PSA is a protease naturally produced by the prostate and by the tumor in men with prostate cancer. Normally this glycoprotein is at a very low level in the blood stream. The PSA concentration will naturally rise with age and increasing prostate size, which occurs with prostate diseases, such as prostate cancer. Other common prostate problems are inflammation (prostatitis), and prostate enlargement, also called benign prostatic hyperplasia (BPH).

The PSA concentration is measured by a simple blood test called a PSA test. The PSA level is recognized as an important marker for prostate cancer [6][7], and is commonly used as a screening tool for the disease. In general, PSA levels between 0 and 4 ng/ml are considered normal; levels between 4 and 10 ng/ml may lead to an ultrasound and biopsy to determine the reason for the increase; levels greater than 10 ng/ml are usually suggestive of cancer. There is no limit to how high a PSA concentration can rise.

A digital rectal examination (DRE) is also used as a tool for screening and diagnosis. Because the prostate lies in front of the rectum, the physician can feel the prostate by inserting a finger into the rectum; it allows the physician to determine whether the prostate is enlarged or has lumps or other types of abnormal texture.

At present, there is disagreement as to the appropriateness of the PSA test for routine screening in the population, essentially because of the low classification performance of the PSA blood test (low sensitivity, and low specificity). As a result, guidelines for screening in North-America are very different [5]. For example, the Canadian Urological Society does not recommend any screening program, but rather that the PSA test should be performed only at the request of the patient. On the other hand, according to the American Cancer Society, men aged 50 and older, and those over the age of 45 who are in high-risk groups, such as African-American men and men with a family history of prostate cancer, should have a PSA blood test once every year.

Early prostate cancer usually does not cause any symptoms. As the tumor grows, the cancer may spread from the prostate to surrounding areas; in such cases, change in urination, including increased frequency, hesitancy or dribbling of urine may be experienced. These symptoms are also common to other prostate disorders, such as BPH. Determining prostate cancer involves a series of tests and exams, including a DRE, a PSA blood test and a biopsy. The DRE and PSA blood test are performed in the first steps of the diagnosis process, and used as indicators of prostatic disorders. Only a biopsy will definitely confirm prostate cancer. A biopsy is the removal of a sample of tissue, which is then examined under a microscope to check for cancerous changes.

When prostate cancer is diagnosed, the tumor is histologically graded with the Gleason score, which refers to the aggressiveness of the cancer ; that is, the likelihood of its spreading [8]. The Gleason score consists of the sum of two scores, each between 1 and 5. A score is assigned to the two areas that make up most of the cancer, and these two scores are added together; the higher the sum, the more aggressive the tumor.

Table 2–2: T-stage classification of prostate cancer.

T-Stage	Tumor
TX	Cannot be assessed
T0	No evidence
T1	Clinically unapparent, not palpable or visible by imaging
	T1a incidental histologic finding in $\leq 5\%$ of resected tissue
	T1b incidental histologic finding in $> 5\%$ of resected tissue
	T1c identified by needle biopsy (e.g., because of elevated PSA values) but not palpable or visible by imaging)
T2	Confined to the prostate
	T2a involves half a lobe or less
	T2b involves more than half a lobe, but not both
	T2c involves both lobes
T3	Extends through the prostatic tissue
	T3a extends unilaterally
	T3a extends bilaterally
	T3a invades seminal vesicle(s)
T4	Fixed or invades adjacent structures other than seminal vesicles
	T4a invades bladder neck, external sphincter or rectum
	T4b invades levator muscles or is fixed to pelvic wall (or both)

Source: NCI [9]

As any cancer, prostate cancer is also classified according to its anatomic stage. Staging is the process of gathering information about the cancer from certain tests to determine if it has already spread or how widespread it is. It is an important factor in choosing treatment options and predicting a patient's prognosis. Several tests and procedures can be used for the staging process: assessment of the PSA concentrations and Gleason score, digital rectal exams (DRE), computed tomography imaging scans, magnetic-resonance imaging scans (MRI) or bone scans. Four categories describe the primary tumor's T stage, as detailed in table 2–2.

Finally, nomograms are predictive instruments that use a set of input data (baseline PSA, Gleason, stage) to make predictions about clinical states, such as the pathological stage of the cancer and other pathological features. For example, Partin et al. have predicted specific pathological tumor stages (organ-confined diseases, isolated

capsular penetration, seminal vesicle involvement, or pelvic lymph node involvement) using the PSA concentration, the clinical stage, and the Gleason score [10].

Treatments

The options for managing prostate cancer include watchful waiting, potentially curative treatment (surgery or radiotherapy), and palliation (hormone therapy or chemotherapy) [6]. Watchful waiting is considered more appropriate for men with a life expectancy of less than ten years who have a low-stage cancer; in this case the cancer is growing so slowly that the subjects will likely die from another cause. In men with clinically localized prostate cancer (confined to the prostate), whose life expectancy is ten years or more, the goal of the treatment is to eradicate the disease. Radical prostatectomy is commonly used when the cancer is localized. Radiation therapy is used to treat cancer that is still confined within the prostate gland, or that has spread to nearby tissue only; it has been used preferentially in older, less healthy patients, i.e. not suitable candidates for surgery, or by man's preference. Radiation therapy uses high-energy rays and particles to kill cancer cells. There are two main forms of radiotherapy: brachytherapy and external beam radiation therapy, usually simply referred as radiotherapy. Brachytherapy involves implanting radioactive material (seeds) into the prostate via thin needles, while external beam radiation therapy is focused from a source outside the body on the areas affected by the cancer. High energy-radiations from X rays, γ rays, neutrons and other sources are used to kill cancer cells and thus shrink the tumor. The radiation damages one or both strands of the DNA inside the cells, thereby preventing the cells from growing and dividing further. Although normal cells are also sometimes damaged by the radiotherapy, they can repair themselves more effectively. Radiotherapy is normally given as a series of short daily treatments over a period of six to seven weeks.

When cancer has spread to other parts of the body or has returned after treatment, hormone-withdrawal therapy is available. The goal of this therapy, also called androgen-deprivation therapy, is to lower the levels of the male hormones (androgens), including testosterone, which are produced mainly in the testicles and cause prostate cancer cells to grow. This therapy can lower androgen levels, thereby making prostate cancers shrink and grow more slowly. Although not a cure, this type of treatment can increase life span and keep the disease under control for some time.

Chemotherapy is used when prostate cancer has spread outside of the prostate gland and hormone-withdrawal therapy has failed; it may slow tumor growth but it is not expected to destroy all the cancer cells.

These active treatments have serious side effects. Impotence is very frequent and urinary incontinence common with surgery and radiotherapy; erectile dysfunction is almost certain and hot flashes are common when hormonal treatments are used; gastrointestinal disorders, hair loss and immunosuppression are expected with chemotherapy.

The choice of the treatment is complex and depends on the spread of the cancer, the age and general medical condition of the patient, as well as his personal preferences (some may not be willing to accept adverse effects of some therapies). It is possible to estimate the probability of specific treatment outcomes before any therapy is initiated. Baseline characteristics, such as the initial PSA concentration, are used to predict metastasis or recurrence of the cancer under specific treatments, using nomograms and risk groups [11]. For example, using the baseline PSA level and the Gleason score, Shipley et al. identified four subgroups of men for which the risk of radiotherapy failure was similar [12]. Men were divided into four categories: PSA levels below 9.2 ng/ml, PSA levels between 9.2 ng/ml and 19.7 ng/ml, PSA level above 19.7 ng/ml and a Gleason score of 7 or above. Their respective five-year rates of survival free of biochemical failure were 81%, 69%, 47% and 29% (see later

for a definition of biochemical failure). Some of these nomograms have even become available on the internet; for example, the patients and/or physicians can specify pre-treatment information to obtain an estimated 5-year progression free probability under different treatment options [13].

In 1992, the U.S National Cancer Institute reported that among cases with localized and regional disease, the proportions of men undergoing surgery, radiotherapy, hormonal treatment and watchful waiting were respectively 37%, 32%, 8% and 23%. I could not find this information for the entire Canadian population, however, for the province of Alberta, these proportions were approximatively, 40%, 30%, 20% and 10%, for respectively surgery, radiotherapy, hormonal treatment and watchful waiting. These data are from the Alberta Cancer Registry and were personally communicated by Dr Penny Brasher, biostatistician at the Alberta Cancer Board; they concern all men diagnosed with invasive prostate cancer living in Alberta at the time of diagnosis (years 1996-2001). There may well be differences between the Canadian provinces, but I do not expect wide variations.

2.1.2 Prostate-specific antigen (PSA)

Variability of the Prostate-specific antigens

PSA concentrations provide valuable information about disease progress; unfortunately they are subject to considerable intra- and inter-individual variations due to measurement and biological factors. The PSA concentrations can be affected by different sources of variability. For example, there might have been errors in the measurement process or during data collection (errors in reporting the date or the PSA level), a change in the laboratory practice or a different laboratory might have been used. There are other known and unknown sources of variation such as recent prostate manipulations [14][15], recent sexual activity [16][17] or recent physical exercise [18]. Results of Nixon et al. suggest that an increase between two consecutive

PSA levels between 20% to 46% might be due to biological and analytical variation only [19]. Ornstein et al. obtained serum on three different occasions, two weeks apart, for 92 healthy men and reported a mean variation of 15% [20]. Eastham et al. evaluated year-to-year fluctuations in PSA, and concluded that nearly half of men who had one abnormal PSA level subsequently had a normal level, suggesting that PSA level fluctuations may result in many false elevations [2]. These studies relate to cohorts free of prostate cancer, and it is likely that the coefficients of variation (defined as the ratio of the standard deviation to the mean) obtained in these studies are lower than in men with prostate cancer, since post-treatment PSA series are also subject to the treatment effect, both on the tumor and the normal prostatic tissue.

Finally, different laboratory assays have different coefficients of variation. For example, at mean concentrations 0.7 ng/ml, 2.8 ng/ml, and 17.9 ng/ml, the assay used by Eastham et al. had coefficients of variation of 3.1%, 2.9% and 0.6% respectively [2]. On the other hand, Carter et al. reported coefficients of variation of 7.5%, 4% and 4%, at mean concentrations of 0.84 ng/ml, 2.9 ng/ml, and 40 ng/ml [21].

Monitoring PSA after radiation therapy

PSA concentrations are proportional to the tumor volume, and thus help detect residual and early recurrence of tumor, and monitor responses to radiotherapy [22][23]. In the past 15 years, PSA has become an important tool in the monitoring of prostate cancer after treatment. If radiation therapy is successful, the PSA concentration decreases substantially to a nadir value, and remains at low levels with possibly a very slow increase. A subsequent steeper increase would indicate treatment failure, and depending on the man's wishes, his physician may start hormone-withdrawal therapy. The initial *log*PSA decline rate and the post-nadir *log*PSA growth rate vary across men, but are reasonably constant within-men, suggesting exponential PSA patterns before and after the nadir [23][24]. For this reason, the

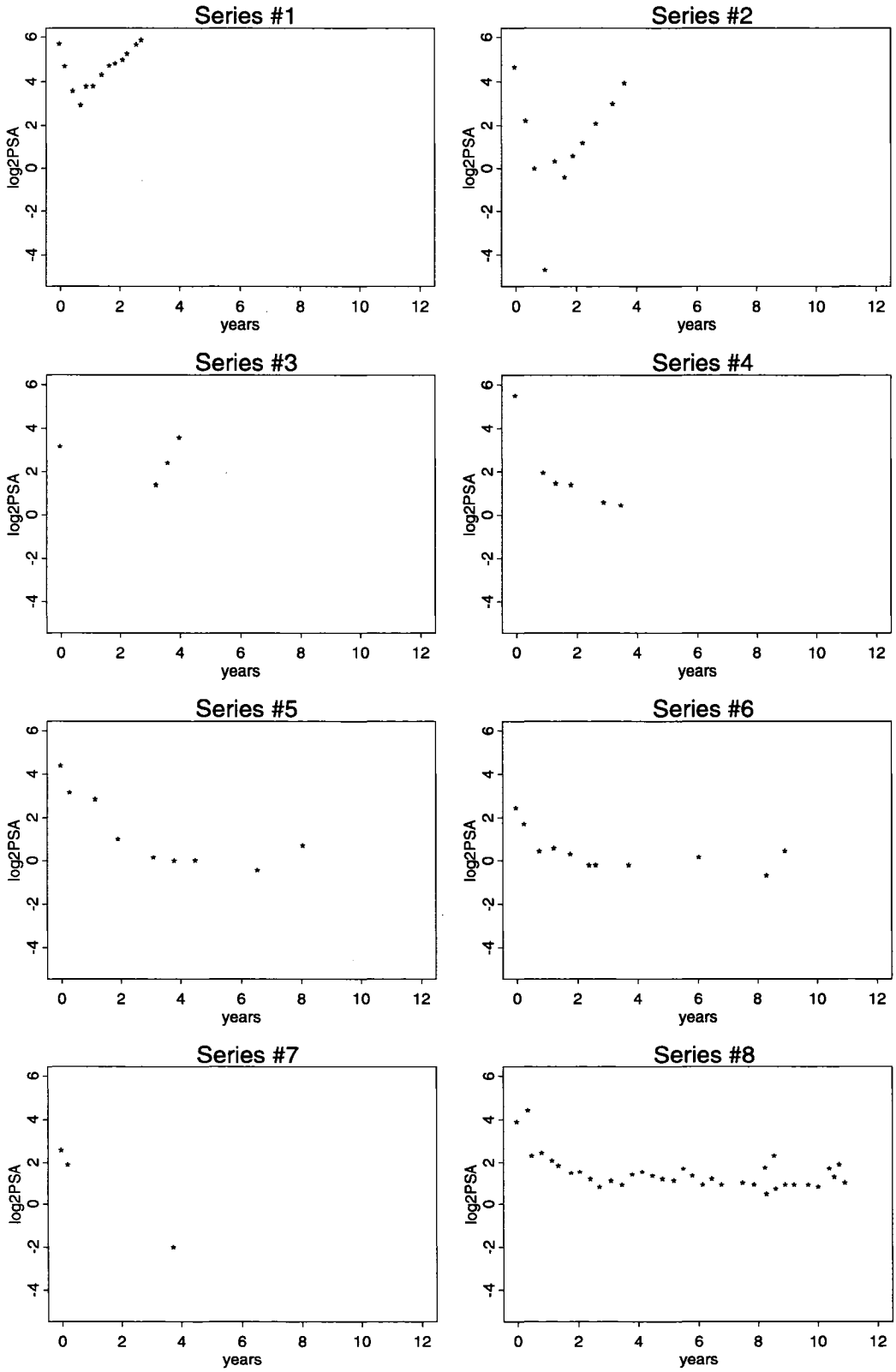
PSA series are frequently plotted on the *log* scale to obtain a piecewise linear trajectory. The base e is often used, but the base 2 is sometimes selected since the reciprocal of the post-nadir slope on this new scale corresponds to the PSA doubling time.

Figure 2–1 illustrates PSA profiles for eight men following radiotherapy (page 16). These men were selected from a cohort of 470 men, part of a larger study described by Albertsen et al. [25]; I will describe this cohort more extensively in the next chapters. Note that the data were plotted on the log_2 scale and the first measurement is taken before the treatment in initiated (time zero). This subset of data allows us to illustrate the general characteristics of PSA series. First, we observe the considerable within-subject variability. For example, the sixth and eighth series appear to stabilize after a few years, however, the PSA levels continue to depict some non-negligible variations over time. Second, between-men variability of the rates of change before and after the nadir are important. The log_2PSA concentrations decline very steeply for the first man, while the decline rate is much shallower for the sixth man. Similarly, the observed growth rates for the three first men are much steeper than for the last five men. The nadir can be reached at various points in time; for example, the lowest PSA measurement is reached within the first year for the first two men, while it still has not been reached after four years for men 4 and 5. Note finally, that unlike in a controlled clinical trial, the number of PSA measurements, the length of follow-up and the interval between measurements are naturalistic, and vary extensively between the series subjects.

Prognostic role of the post-treatment PSA concentrations

Several studies have investigated the role and distribution of the parameters of the log_2 PSA profiles: the initial decline rate, the nadir, the timing of the nadir and the subsequent growth rate. Most studies have focussed on the PSA growth rate and

Figure 2–1: log_2PSA concentrations over time for eight men



the PSA nadir. The log_2PSA growth rate or its reciprocal, the PSA doubling time, has been shown to be an important predictor of disease progression and survival [25][26][27][28], and its distribution was often reported. Pollack et al. followed 427 men treated with radiotherapy for T1-T4 cancers of the prostate [26]. The patients were scored as having a rising PSA profile when the values obtained at two or more follow-up visits were rising and when at least one value was greater than 1ng/ml. Based on this definition, 100 men had a rising PSA profile and enough data to compute the PSA doubling time. The reported mean and median doubling time were 13.5 months (SD=11.6 months) and 10.3 months respectively. Zagars et al. estimated the PSA doubling time for 154 men treated for radiotherapy [29]. Of these, 37 men had two or more consecutive rising PSA values, and were therefore considered as having an increasing post-nadir profile. The reported mean and median doubling time were 12.5 months and 8.8 months respectively.

The value of the PSA nadir has been shown to be associated with clinical outcomes such as local failure [30], distant failure [30][31], and progression free survival [32][31]. Based on 655 men, Kaminsky et al. reported a mean PSA nadir, median and range of 1.17 ng/ml, 0.50 ng/ml and 0 to 98.7 ng/ml respectively [33]. Hanlon et al. reported the distribution of the nadir from 615 men treated with radiotherapy [31]. The mean nadir was 1.8 ng/ml, its median was 0.6 ng/ml, and it ranged from 0 to 177 ng/ml. The nadir was reached on average 36 months after the initial day of radiotherapy, with a median time of 32 months and a range of 2 to 114 months. Both studies used the ASTRO criterion to define the PSA nadir (see section 2.1.3).

The rate of decline before the PSA nadir has not been evaluated as extensively and results are contradictory. Meek et al. concluded that a longer half life was associated with subsequent disease progression [24], but Zagars found no association, and the PSA rate of fall following treatment did not provide any clinical information [29]. Meek et al. reported a mean half life of 43 days (SD=11); note, however, that

the authors restricted their analysis to PSA measurements during radiation therapy only. Zagars et al. reported a mean half life of 1.9 months, and a range of 0.5 to 9.2 months [29].

The association between the parameters of the PSA profile and pre-treatment characteristics has been investigated in a few studies. In particular, a shorter PSA doubling time has been shown to be associated with a higher pretreatment Gleason score [23][26][29][34] and higher pretreatment PSA concentrations [26]. Similarly, an association has been shown between the PSA nadir and the pretreatment PSA level, but none with the Gleason score. Results do not suggest any significant association between the PSA half life and pretreatment characteristics [29].

The studies described in this section estimated the parameters of the PSA trajectories without accounting for the considerable variability of the PSA observations: the PSA data were taken at their face values. The PSA doubling time was estimated from PSA series that appeared to be rising, i.e., from series satisfying some arbitrary definition (for example, at least two consecutive PSA rises). Similarly, the lowest observed PSA concentration was used at the PSA nadir. Thus, these estimates should be handled with care, since they are likely to be biased. In order to capture, and estimate, both the within- and between- PSA series variability, appropriate statistical modeling is necessary.

2.1.3 The ASTRO criterion for biochemical failure following radiotherapy

If radiotherapy fails, a secondary treatment can be initiated, usually hormone-withdrawal therapy. The best timing of this secondary treatment is a compromise between wishing to delay the onset of clinically evident metastases, the side effects of hormone deprivation, the possibility that the cancer will eventually become resistant to the hormone manipulation, and that the man may die of another disease before

the cancer could become symptomatic. It is therefore essential to detect treatment failure as quickly as possible.

Biochemical failure is defined as a recurrence of the cancer, detected by a rising PSA. There is an ongoing debate as to a firm definition of a *rising* PSA pattern. Clearly, for validity, any measure should be strongly correlated with the probability of developing symptomatic metastases and death from prostate cancer. Before 1996, several definitions of a significant PSA increase had been used, however, the use of various definitions rendered the comparison of studies difficult [35][36]. In 1996, the American Society for Therapeutic Radiology and Oncology consensus panel proposed guidelines to unify the scientific community on the use of a single definition that would standardize the reporting and comparison of treatment outcomes; these guidelines are reported below (cited from [37]):

The panel agreed on four guidelines:

- *Biochemical failure is not justification per se to initiate additional treatment. It is not equivalent to clinical failure. It is, however, an appropriate early endpoint for clinical trials.*
- *Three consecutive increases in PSA is a reasonable definition of biochemical failure after radiation therapy. For clinical trials, the date of failure should be the midpoint between the postirradiation nadir PSA and the first of the three consecutive rises.*

  The use of three, rather than two, consecutive values reduces the risk of falsely declaring biochemical failure due to "bouncing" PSAs. This phenomenon results when sequential PSA determinations show one or two rises followed by a fall and a subsequent failure to rise again.
- *No definition of PSA failure, has yet, been shown to be a surrogate for clinical progression or survival.*

- *Nadir PSA is a strong prognostic factor, but no absolute level is a valid cut point for separating successful and unsuccessful treatments. Nadir PSA is similar in prognostic value to pretreatment prognostic variables.*

This rule, which I will refer to as the ASTRO rule, is commonly used and widely discussed in the literature. However, how well it classifies individuals (its sensitivity and specificity) has undergone only limited formal evaluation. Thus far, evaluations that have been carried out are limited to correlations with clinical outcomes, and in specific case series. Horwitz et al. reported that the ASTRO definition correlated well with clinical outcomes such as distant metastases, disease free survival and cause-specific survival [38]; they concluded further that this measure of biochemical failure may be a valid endpoint for separating successful versus unsuccessful treatment. Hanlon et al., reported that the ASTRO definition was robust to inclusion of patients with some extreme characteristics (recurrence with a very slow rise) [39].

Several authors have criticized the ASTRO criterion, in particular regarding the date of failure. In the case of a successful surgery, the PSA concentrations drop to negligible values. If there is a recurrence of the cancer, PSA concentrations start to increase, and the date of failure is the point in time when PSA become detectable. After radiotherapy, the date of failure is not as obvious. The remaining (healthy) prostatic tissue continues to produce PSA, and therefore, the post-radiotherapy PSA concentrations will eventually rise again, whether or not the subject is cured. To facilitate comparisons with surgery series, the decision was made to backdate the time of failure to the point in time when the PSA begin to rise again. This decision is controversial, and several studies have pointed out the ambiguity of this procedure. This problem is particular to recently rising PSA values, which are difficult to interpret since the patient may have recurred but does not have sufficient PSA data to be scored as a failure [40].

Although the ASTRO criterion is used in the majority of the recent publications [3], alternative definitions have been proposed. Definitions incorporating the PSA nadir and/or rises from the nadir may have a better correlation with clinical failure than the ASTRO definition [41] and may even better identify failures and rises otherwise misclassified [42]. For example, based on 727 subjects treated with radiotherapy alone, Vicini et al. reported that the ASTRO definition had 73% sensitivity and 76% specificity, both lower than other definitions including a fixed PSA level [41]. Similarly, using 688 subjects treated, Taylor et al. reported that the ASTRO definition had 44% sensitivity and 88% specificity, again lower than other definitions including fixed PSA level [43]. Finally, Thames et al. have recently reported results form a large study aimed at evaluating the ASTRO criterion ; based on data from nearly 5000 patients treated with radiotherapy, the authors reported a 61% sensitivity and a 80% specificity, again lower than some definitions that included the PSA nadir [44].

2.1.4 The ASTRO criterion, PSA series, and statistical modeling

As of the time I planned this thesis, studies investigating the performance of the ASTRO criterion focussed on its capabilities to predict distant clinical outcomes. These studies provided discordant findings, but were also very difficult to compare, since the findings were functions of specific features related to the data collection: different institutions, different inclusion criteria, different intensity of PSA surveillance, and different lengths of follow-up. Given the discordant findings of these clinical validation studies, as well as the important proportion of studies not relying on the ASTRO criterion, an investigation of the more short-term performance of the ASTRO rule appears necessary. Indeed, a fundamental point, before one considers how well even a perfectly measured PSA trajectory correlates with clinical outcomes, is how good the ASTRO rule is at correctly (and quickly) identifying 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. This *numerical validation* involves comparing observed PSA series to the underlying true PSA trajectories. It is surprising that this first-stage issue has not been evaluated, given that statistical methods have successfully described longitudinal changes in PSA to predict either the onset of prostate cancer [45][46][47], or the recurrence of the disease following treatment [48][49]. Based on these successful statistical analyses, a numerical evaluation of the ASTRO criterion using modeling of PSA post-treatment series appears feasible. This validation study is the core of this research thesis, and will be presented in the next chapters.

The analysis of longitudinal PSA data should account for the several characteristics of the PSA series, including the piecewise-linear pattern of the post-radiotherapy PSA profiles on both sides of the man's specific changepoint, and the important within and between series variability. In addition, the model should permit the modeling of the PSA variance as a function of the PSA concentration, since it is suspected that the variability of the PSA measurements is more important at lower levels [2][21]. Finally, the model should adequately handle unbalanced data, since PSA series typically have differing number and frequency of observations, and differing follow-up lengths. Given that both the individual and population PSA profiles are of interest, a random-effects model is particularly appropriate to accommodate this multi-level structure. In addition, a Bayesian approach appears computationally more flexible than the maximum-likelihood approach, given the multiple features of the longitudinal PSA data. Specifically, Bayesian hierarchical models easily handle the presence of a random changepoint, non-constant variance, and the small number of PSA observations for some men compared to the large number of parameters to estimate.

I have chosen a Bayesian hierarchical changepoint model with a random changepoint to estimate post-radiotherapy PSA series. In the next sections, I review the

characteristics of this model. I first describe the general principles of Bayesian inference in section 2.2; I do not compare the Bayesian philosophy to the frequentist approach, but rather emphasize the practicality of this approach. In section 2.3, I focus on the random-effect models, or *hierarchial models*, in the Bayesian framework. In section 2.4, I describe the hierarchical changepoint model which allows the modeling of piecewise-linear patterns.

2.2 Bayesian inference

2.2.1 Principles

In frequentist inference, the parameters of a statistical model are usually estimated using the maximum likelihood principle. The likelihood function expresses the probability of observing the data, as a function of the unknown parameters of the model, and a point estimate is obtained by maximizing the likelihood function. Confidence intervals can be constructed for each population parameter, and are based on the concept of repeated sampling. If the same population is sampled indefinitely, and 95% interval estimates are made on each occasion, then 95% of these intervals will contain the population parameter.

In contrast, the Bayesian approach views the unknown parameter as a random variable, and speaks directly about the parameter of interest. Bayesian inference makes '*explicit use of probability for quantifying uncertainty in inferences based on statistical data analysis*' [50], and inferences are in terms of posterior probability distributions, and credible intervals. Summary statistics of the parameter's distribution (or uncertainty), such as its mean or median, can be estimated. A 95% credible interval provides an estimated range of values which is likely to include the unknown parameter, i.e., it has 95% probability of containing the true parameter.

Bayesian analysis is based on Bayes rule, and relies on three components: the prior distribution, the likelihood of the data, and the posterior distribution. Assume

that we observe some data $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)$, $n \geq 1$, and we are interested in estimating a set of parameters from a model $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_k), k \geq 1$. For ease of exposition, I assume that both the data and the parameters are continuous, and that both the data and the set of parameters are multidimensional.

The prior distribution, $p(\boldsymbol{\theta})$, reflects the state of knowledge about the parameter(s) *before* one has seen the data. If one has much information available, then a sharp distribution is assigned to the prior distribution of the parameter. On the other hand, if little is known about the parameter, then one assigns a noninformative (or flat, diffuse) prior distribution: i.e., the parameter is assigned a distribution with equal probability for each possible value.

The likelihood function, $\mathcal{L}(\mathbf{y}|\boldsymbol{\theta})$, represents the probability of observing the data $\mathbf{Y}$, given the parameter $\boldsymbol{\theta}$.

The posterior distribution, $p(\boldsymbol{\theta}|\mathbf{y})$, reflects the state of our knowledge, *after* combining the information provided by the prior distribution, $p(\boldsymbol{\theta})$, and the observed data $\mathcal{L}(\mathbf{y}|\boldsymbol{\theta})$; mathematically, these two distributions are combined using Bayes rule:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\boldsymbol{\theta}, \mathbf{y})}{p(\mathbf{y})} = \frac{\mathcal{L}(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})}. \tag{2.1}$$

Note that $p(\mathbf{y}) = \int \mathcal{L}(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$ can be interpreted as an average of the likelihood with respect to the uncertainty about $\boldsymbol{\theta}$. In addition, $p(\mathbf{y})$ is independent of $\boldsymbol{\theta}$, and can be considered as constant for fixed $\mathbf{y}$. We can rewrite equation (2.1) as:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto \mathcal{L}(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}).$$

Bayes rule uses the information provided by the data to update prior beliefs about the parameters. If little is known beforehand about $\boldsymbol{\theta}$, a noninformative prior distribution $p(\boldsymbol{\theta})$ is selected, and the posterior distribution is mainly determined by the data $\mathcal{L}(\mathbf{y}|\boldsymbol{\theta})$.

Often, the integral required to obtain $p(\mathbf{y})$ can be difficult to calculate, and $p(\boldsymbol{\theta}|\mathbf{y})$ may not be available in a closed analytical form. In such case, algorithms such as Markov Chain Monte Carlo techniques allow one to draw samples from the posterior distribution, and thus provide summary statistics (mean, median, etc...) for the parameters of interest. Some of these algorithms are discussed in section 2.2.2.

2.2.2 Posterior simulations

Posterior distributions difficult to solve analytically can be estimated using computational algorithms that generate sequences of samples from the distributions of interest. Any summary statistics of the parameter can then be estimated from the generated samples. These algorithms, such as the Gibbs Sampler, are based on Markov Chain Monte Carlo (MCMC) methods, which sample from the probability distribution by constructing a Markov chain which converges to the distribution of interest. I start by reviewing the general principles of the MCMC methods, and then describe specific algorithms, including the Gibbs Sampler, the Metropolis, and the Metropolis-Hastings algorithms. These algorithms will be used in the next chapters to estimate PSA trajectories.

Markov Chain Monte Carlo methods

A first-order Markov chain is defined as a sequence of simulated random variables $X_1, ..., X_n$, where the value X_i is called the state of the process at time i. If the probability of sampling the next state X_{i+1} depends only on the current state X_i, i.e., $p(X_{i+1}|X_i) = p(X_{i+1}|X_i, X_{i-1}, ..., X_1)$, then we say that the sequence $X_1, ..., X_n$ forms a first-order Markov chain, with transition distribution $p(X_{i+1}|X_i)$. The Markov chain is aperiodic if there is no state to which the process will continually return with a fixed time period. If the Markov chain is aperiodic, and if the expected return time is finite for every state, then the Markov chain converges to some stationary

distribution $\pi()$. If in addition, the Markov chain is irreducible, i.e., every state i is accessible from every other state, then this stationary distribution is unique.

Monte Carlo algorithms are computational algorithms that simulate the behavior of complex functions. This is particularly useful in Bayesian inference, where some distributions of interest require complex multidimensional integrations. These algorithms use sequences of (approximately) random numbers, such as the ones generated by Markov chains.

Given a complex function of interest $f(.)$, and a Markov chain $X_1, ..., X_n$ with stationary distribution $\pi(.)$, Markov Chain Monte Carlo (MCMC) uses the simulations of the Markov chain to approximate the expectation μ of the function $f(.)$ with respect to π:

$$\mu = E_\pi\{f(X_i)\}$$

by a sample average of the Markov chain simulations:

$$\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i).$$

In general, there is a burn in period to allow for convergence of the Markov chain, and the first m iterations are discarded. The expectation is then estimated as, for example:

$$\widehat{E\{f(X)\}} = \frac{1}{n-m}\sum_{t=m+1}^{n} f(X_t). \tag{2.2}$$

Several algorithms use MCMC techniques; I describe two of them.

The Gibbs sampler

The objective of Bayesian inference is the estimation of the posterior distribution

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{\mathcal{L}(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int \mathcal{L}(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}, \tag{2.3}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_k)$ is the set of parameters of interest. The denominator of equation 2.3 is often intractable and thus, $p(\boldsymbol{\theta}|\mathbf{y})$ is not available analytically. The Gibbs sampler uses MCMC methods to generate a sequence of samples from the full conditional distribution $p(\boldsymbol{\theta}|\mathbf{y})$, and was formally introduced by Geman and Geman [51]. It is applicable if the conditional distribution of each variable is available explicitly, i.e., the full set of conditional distributions:

$$p(\theta_i|\boldsymbol{\theta}_{-i}, \mathbf{y}), i = 1, ..., k, \tag{2.4}$$

must be available to sample from, where $\boldsymbol{\theta}_{-i}$ represents all the components of the vector $\boldsymbol{\theta}$, except the i^{th} element, θ_i. A Markov chain is generated by successively sampling values from the conditional distributions $p(\theta_i|\boldsymbol{\theta}_{-i}, \mathbf{y})$. Each component of $\boldsymbol{\theta}$ is updated conditional on the current values of the other variables. If we let t denote the iteration step, then the algorithm proceeds as follows:

1. For $t = 0$, select a starting point $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, ..., \theta_k^{(0)})$,
2. For t=1, 2, ...:
   (a) Sample $\theta_1^{(t+1)}$ from $p(\theta_1|\theta_2^{(t)}, \theta_3^{(t)}, ..., \theta_k^{(t)}, \mathbf{y})$,
   (b) Sample $\theta_2^{(t+1)}$ from $p(\theta_2|\theta_1^{(t+1)}, \theta_3^{(t)}, ..., \theta_k^{(t)}, \mathbf{y})$,
   (c) Sample $\theta_3^{(t+1)}$ from $p(\theta_3|\theta_1^{(t+1)}, \theta_2^{(t+1)}, ..., \theta_k^{(t)}, \mathbf{y})$,
   (d) ...
   (e) Sample $\theta_k^{(t+1)}$ from $p(\theta_k|\theta_1^{(t+1)}, \theta_2^{(t+1)}, \theta_3^{(t+1)}, ..., \theta_{k-1}^{(t+1)}, \mathbf{y})$.

At each iteration t, the Gibbs sampler generates a sequence of vectors, $\boldsymbol{\theta}^{(1)}$, $\boldsymbol{\theta}^{(2)}$..., where each vector $\boldsymbol{\theta}^{(t)}$ depends only on the previous iteration $\boldsymbol{\theta}^{(t-1)}$. The resulting sequence forms a Markov chain whose stationary distribution is the requested distribution $p(\boldsymbol{\theta}|\mathbf{y})$ [50]. Once convergence is reached, usually following an initial burn-in period of m iterations, the set of vectors $\boldsymbol{\theta}^{(m+1)}, ..., \boldsymbol{\theta}^{(n)}$ is considered as a sample from the joint distribution $p(\boldsymbol{\theta}|\mathbf{y})$.

Often, the interest is in the individual components of $\boldsymbol{\theta}$, and thus in the marginal posterior distributions:

$$p(\theta_i), i = 1, ..., k.$$

Geman and Geman showed that, in the limit, the distribution of the simulated elements $\theta_i, i = 1, ..., k$ will tend to the marginal distribution $p(\theta_i)$ [51]:

$$\theta_i^{(n)} \rightarrow^d \theta_i \sim p(\theta_i) \quad \text{as} \quad n \rightarrow \infty.$$

Statistics about the marginal distribution $p(\theta_i)$ can be calculated using the sets of replicates $\theta_i^{(m+1)}, ..., \theta_i^{(n)}$. For example, following equation 2.2, the mean of $p(\theta_i)$ can be estimated using $\frac{1}{n}\sum_{j=m+1}^{n}\theta_i^{(j)}$. Credible intervals are obtained using the percentiles. In this thesis, for example, I will report 95% credible intervals which are constructed using the 2.5^{th} and 97.5^{th} percentiles.

The Metropolis and the Metropolis-Hastings algorithms

The Gibbs sampler requires some knowledge of the full conditional distributions, and uses every simulated value. An alternative, more general, updating scheme is a form of acceptance/rejection sampling, where values are first drawn from an arbitrary distribution, and corrected so that, asymptotically, they behave as random observations from the target distribution. This is the motivation behind algorithms such as the Metropolis [52], and its generalization, the Metropolis-Hastings algorithm [53]. Let $p(.)$ and t denote the joint distribution, and the iteration number respectively, then the Metropolis-Hastings algorithm proceeds as follows:

1. Select the proposal distribution $q^{(t)}(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})$ (or jumping distribution), that will be used to select candidates at each new step t. This distribution might be dependent on the current state $\boldsymbol{\theta}^{(t)}$ (for example a normal distribution centered at the current iteration).
2. For $t = 0$, select a starting point $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, ..., \theta_k^{(0)})$,

3. For t=1, 2, ...,

   (a) Sample a candidate $\boldsymbol{\theta}^*$ from the proposal distribution $q^{(t)}(\boldsymbol{\theta}^{(*)}|\boldsymbol{\theta}^{(t-1)})$.

   (b) Calculate the Hastings ratio r, where

$$r = \frac{p(\boldsymbol{\theta}^{(*)}|\mathbf{y})q^{(t)}(\boldsymbol{\theta}^{(*)}|\boldsymbol{\theta}^{(t-1)})}{p(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta})q^{(t)}(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}^{(*)})}.$$

   (c) Set

$$\boldsymbol{\theta}^{(t)} = \begin{cases} \boldsymbol{\theta}^{(*)} & \text{with probability } min(r,1) \\ \boldsymbol{\theta}^{(t-1)} & \text{with probability } 1 - min(r,1) \end{cases}$$

The Hastings ratio can be rewritten as:

$$r = \frac{p(\boldsymbol{\theta}^{(*)}|\mathbf{y})}{p(\boldsymbol{\theta}^{(t-1)}|\mathbf{y})} \times \frac{q^{(t)}(\boldsymbol{\theta}^{(*)}|\boldsymbol{\theta}^{(t-1)})}{q^{(t)}(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}^{(*)})} = \frac{p(\mathbf{y}|\boldsymbol{\theta}^{(*)})p(\boldsymbol{\theta}^*)}{p(\mathbf{y}|\boldsymbol{\theta}^{(t-1)})p(\boldsymbol{\theta}^{(t-1)})} \times \frac{q^{(t)}(\boldsymbol{\theta}^{(*)}|\boldsymbol{\theta}^{(t-1)})}{q^{(t)}(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}^{(*)})}. \quad (2.5)$$

Just like the Gibbs sampler, the Metropolis Hastings algorithm can also be applied component by component. In such case, the Hastings ratios are given by:

$$r_i = \frac{p(\theta_i^{(*)}|\boldsymbol{\theta}_{-i}, \mathbf{y})}{p(\theta^{(t-1)}|\boldsymbol{\theta}_{-i}, \mathbf{y})} \times \frac{q_i^{(t)}(\boldsymbol{\theta}^{(*)}|\boldsymbol{\theta}^{(t-1)})}{q_i^{(t)}(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}^{(*)})} = \frac{p(\mathbf{y}|\boldsymbol{\theta}^{(*)})p(\boldsymbol{\theta}^*)}{p(\mathbf{y}|\boldsymbol{\theta}^{(t-1)})p(\boldsymbol{\theta}^{(t-1)})} \times \frac{q_i^{(t)}(\boldsymbol{\theta}^{(*)}|\boldsymbol{\theta}^{(t-1)})}{q_i^{(t)}(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}^{(*)})}.$$

Given the form of the Hastings ratio (equation 2.5), the full conditional distributions are not needed, and basically any proposal distribution can be used. A good proposal distribution should be easy to sample from, and should allow to easily obtain the Hastings ratio. For faster convergence, the candidates should be selected from a reasonable distance from the last draw, and the refusal rate should not be too elevated. As mentioned in Gelman et al., the ideal proposal distribution is simply the target distribution, i.e., $q(\boldsymbol{\theta}^{(*)}|\boldsymbol{\theta}^{(t-1)}) = p(\boldsymbol{\theta}^{(*)}|\mathbf{y})$ [50]. Thus, the Hastings ratio is always 1, and candidates are always accepted. In this special setting, and if the Metropolis Hastings algorithm is applied component by component, the Metropolis-Hastings is equivalent to the Gibbs sampler.

Similarly, the Metropolis algorithm is also a special case of the Metropolis-Hastings algorithm. The proposal distribution is simply selected to be symmetric,

i.e., $q^{(t)}(\boldsymbol{\theta}^{(*)}|\boldsymbol{\theta}^{(t-1)}) = q^{(t)}(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}^{(*)})$, and in such case r is simplified to the ratio of two conditional densities:

$$r = \frac{p(\boldsymbol{\theta}^{(*)}|\mathbf{y})}{p(\boldsymbol{\theta}^{(t-1)}|\mathbf{y})}.$$

Monitoring convergence

Before drawing final inferences, it is essential to ensure that the resulting Markov chain has converged to the desired stationary distribution; if the simulation has not proceeded long enough, the sequence may not yet be representative of the target distribution. For each parameter, it is often recommended to run multiple sequences simultaneously, using several sets of overdispersed starting values. Several techniques are available to monitor convergence, using either single or multiple chains; I describe two common methods, both of which were used in this thesis.

The Raftery and Lewis convergence criterion applies to single chains, and uses the within-chain correlation [54]. It is intended to detect convergence to the target distribution, and to provide bounds for the accuracy of the estimated quantile of functions of variables of interest. Given a specific quantile, a degree of accuracy, and the probability of attaining this degree of accuracy, the criterion evaluates the minimum number of iterations needed to accurately estimate the quantile. This number is calculated assuming that the iterates are independent, and using the binomial variance. If enough iterates are generated, the Raftery and Lewis criterion calculates the total number of iterations needed to reach convergence, as well as the thinning interval, using the lag autocorrelations of the generated sequence. If the iterates are highly correlated, the thinning interval, and thus the length of the chain, are increased.

The Gelman-Rubin statistic can be used when multiple chains of iterates have been generated. The statistic relies on the comparison of the the within-chain and

between-chain variabilities using a classical analysis of variance [50]. At the start of the simulation process, the between-sequence variability will be greater than the within-sequence one, but as the chains approximate the target distribution, the two (or more) variabilities should become closer. As convergence is approached, the between-sequence variability approaches zero, and thus the Gelman-Rubin statistic becomes closer to 1.

WinBUGS

WinBUGS is an interactive Windows version of the BUGS program for Bayesian analysis of complex statistical models using Markov chain Monte Carlo (MCMC) techniques [55]. The user has to specify the model, namely the likelihood of the data, as well as the prior distributions, and initial values of all the parameters of the models. Then, WinBugs proceeds with MCMC sampling to estimate the unknown parameters of the model. Depending on the form of the distribution, different sampling methods can be used, including the Metropolis algorithm, Gibbs sampling, rejection sampling [56], or slice sampling [57]. The statistical analysis carried out in this thesis was performed using this software (version 1.4), which is available on the public domain.

2.3 Hierarchial models in a Bayesian framework

2.3.1 Introductory example

Hierarchical models are useful when both individual and population parameters are of interest. In the next chapters, one of the parameters I will be interested in, is the post-radiotherapy PSA nadir in men treated for prostate cancer. This nadir can be affected by the treatment and/or by specific characteristics of the patients. Thus, two questions arise:

What is the average post-radiotherapy PSA nadir?

What are the individual post-radiotherapy PSA nadirs?

The classical analysis of variance would not allow me to answer these two questions simultaneously. Indeed, I could either assess the average PSA nadir by pooling all individual nadirs, or, separately estimate each individual nadir. Hierarchical modeling permits to estimate the individual and population nadirs simultaneously by regarding the individual true PSA nadirs θ_j as being connected to each other; more specifically, they are assumed to have a common model. I review the principles of hierarchical modeling more formally in the next section.

2.3.2 Principles

Consider a set of observations $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)$, $n \geq 1$. Each observation is associated with some true effect θ_j, and thus the individual likelihoods are given by $p(y_j|\theta_j)$. The individual effects θ_j are viewed as a sample from a common population of effects, with unknown parameter(s) $\boldsymbol{\phi}$: $\theta_j \sim p(\theta_j|\boldsymbol{\phi})$, where $j = 1, ..., n$. $\boldsymbol{\phi}$ is often called the hyperparameter, and has distribution $p(\phi)$. The objective is to estimate the posterior distributions of the individual effects θ_j, as well as the population effects $\boldsymbol{\phi}$.

A major assumption is the exchangeability of the individual parameters θ_1, ..., θ_n, i.e., the joint prior distribution $p(\theta_1, ..., \theta_n)$ is invariant for any permutation of the indices:

$$p(\theta_1, \theta_2, ..., \theta_n|\boldsymbol{\phi}) = \prod_{j=1}^{n} p(\theta_j|\boldsymbol{\phi}) \tag{2.6}$$

The joint prior distribution, $p(\boldsymbol{\theta}, \boldsymbol{\phi})$, is obtained using conditional probability:

$$p(\boldsymbol{\theta}, \boldsymbol{\phi}) = p(\boldsymbol{\theta}|\boldsymbol{\phi})p(\boldsymbol{\phi}).$$

Applying Bayes rule, we obtain the joint posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{y})$:

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{y}) &\propto p(\boldsymbol{\theta}, \boldsymbol{\phi})p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\phi}) \\ &= p(\boldsymbol{\theta}, \boldsymbol{\phi})p(\mathbf{y}|\boldsymbol{\theta}), \end{aligned}$$

since the sampling distribution, $p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\phi})$, depends only on $\boldsymbol{\theta}$, and $\boldsymbol{\phi}$ affects $\mathbf{y}$ only through $\boldsymbol{\theta}$. The posterior distributions for $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ can then be estimated using simulation techniques, as described earlier in section 2.2.2.

2.3.3 A particular example: the normal hierarchical model

I illustrate the principles of hierarchical modeling using the normal hierarchical model. As an example, assume that we observe a PSA nadir y_j, for a group of n men, where $j = 1, ..., n$. In the simplest case, one could assume that this nadir is common to all men, and thus the observed nadirs y_j are normally distributed with some common unknown mean θ and known variance σ^2:

$$y_j \sim N(\theta, \sigma^2).$$

The unknown average PSA nadir θ is then assigned a normal prior distribution, with known parameters μ_0 and τ_0^2: $\theta \sim N(\mu_0, \tau_0^2)$. The objective is to estimate the posterior distribution of θ, $p(\theta|y_1, ..., y_n)$. This distribution is obtained using conditional distribution:

$$\begin{aligned} p(\theta|y_1, ..., y_n) &\propto p(\theta)p(y_1, ..., y_n|\theta) \\ &= p(\theta) \prod_{j=1}^{n} p(y_i|\theta) \\ &= N(\mu_0, \tau_0^2) \prod_{j=1}^{n} N(\theta, \sigma^2) \\ &= N(\hat{\mu}, \hat{\tau}^2), \end{aligned}$$

where, using some algebra, and given $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$:

$$\hat{\mu} = \frac{\frac{\mu_0}{\hat{\tau}^2} + \frac{n\bar{y}}{\sigma^2}}{\frac{1}{\hat{\tau}_0^2} + \frac{n}{\sigma^2}} \quad \text{and} \quad \frac{1}{\hat{\tau}^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}.$$

However, in several situations, the individual effects θ_j are known to vary across subjects. It is then common to assume that each man j, has its own PSA nadir θ_j:

$$y_j \sim N(\theta_j, \sigma_j^2),$$

where the individual variances σ_j^2 might be known. The individual true effects θ_j are then linked to each other through a common normal distribution:

$$\theta_j \sim N(\mu, \tau^2).$$

The hyperparameters μ and τ^2 correspond respectively to the mean PSA nadir, and to the between nadir variability, and have to be estimated. We denote by $p(\mu, \tau^2)$, $p(\mu|\tau^2)$, and $p(\tau^2)$, the joint distribution of the hyperparameters, the conditional distribution of μ given τ^2, and the marginal distribution of τ^2. Conditional on the hyperparameters, the distribution of the true nadirs, θ_j, is given by:

$$p(\theta_1, \theta_2, ..., \theta_n|\mu, \tau^2) = \prod_{j=1}^{n} p(\theta_j|\mu, \tau^2).$$

Thus, their joint distribution is simply:

$$\begin{aligned} p(\theta_1, \theta_2, ..., \theta_n) &= \int \prod_{j=1}^{n} p(\theta_j|\mu, \tau^2) p(\mu, \tau^2) d\mu d\tau^2 \\ &= \int \prod_{j=1}^{n} p(\theta_j|\mu, \tau^2) p(\mu|\tau^2) p(\tau^2) d\mu d\tau^2. \end{aligned}$$

The joint prior distribution is given by:

$$p(\boldsymbol{\theta}, \mu, \tau^2) = p(\boldsymbol{\theta}|\mu, \tau^2) p(\mu|\tau^2) p(\tau^2).$$

Finally, the joint posterior distribution is then provided by:

$$\begin{aligned} p(\boldsymbol{\theta}, \mu, \tau^2|y) &\propto p(\mathbf{y}|\boldsymbol{\theta}, \mu, \tau^2)p(\boldsymbol{\theta}|\mu, \tau^2)p(\mu|\tau^2)p(\tau^2) \\ &= \left(\prod_{j=1}^{n} p(\mathbf{y}|\theta_j, \mu, \tau^2)p(\theta_j|\mu, \tau^2)\right) p(\mu|\tau^2)p(\tau^2) \\ &= \left(\prod_{j=1}^{n} p(\mathbf{y}|\theta_j)p(\theta_j|\mu, \tau^2)\right) p(\mu|\tau^2)p(\tau^2). \end{aligned}$$

Then, in order to obtain the marginal distributions $p(\mu)$, $p(\tau^2)$, and $p(\theta_j)$ for $j = 1, ..., n$, one could integrate out the joint posterior distribution. If the integrals are intractable analytically, MCMC algorithms can be applied to obtain estimates of the posterior distributions of interest. Similarly, if the individual variances σ_j^2 are unknown, they can be estimated using these same algorithms.

One of my objectives is the estimation of the post-radiotherapy PSA profile over time, rather than a single parameter, such as the PSA nadir. The PSA levels are known to decrease until they reach a PSA nadir, and then increase subsequently. The problem is complex, given that the timing of the nadir, also called the changepoint, is unknown. The estimation of this changepoint, as well as the other parameters of the PSA profile require special care; I describe one of the available techniques below.

2.4 Changepoint problems in a Bayesian framework

2.4.1 Introduction

As a simple example, assume that a vaccine for a specific disease X requires some unknown time to become effective. Once inoculated, one might be interested in the changepoint, i.e., the unknown time at which the vaccine becomes effective, as well as the risk of infection, before and after the changepoint.

Several approaches are available to estimate the changepoint as well as the distribution of the parameters of the model for the data, before and after the changepoint: non-parametric methods [58], maximum likelihood estimation [59][60], and fully Bayesian methods [61][62]. Hierarchical Bayes models readily accommodate changepoint problems, as they allow the changepoint to vary across subjects; the other approaches cannot handle it so easily. In addition, they allow interpretation of the posterior distribution of the changepoint, as the probability that the changepoint has occurred in any given location. These models have successfully described various biomarker series; the changepoint corresponds to the timing of a change in the biomarker distribution, usually equivalent to a change in a clinical state, such as disease onset or recurrence. For example, these models have described longitudinal CD4 cell counts, to predict the timing of HIV viral rebound following treatment [63]. Similarly, longitudinal PSA series have been analyzed to predict prostate cancer onset [45][64], and the timing of prostate cancer recurrence [48][49].

In the next section, I review the discrete-time changepoint problem as presented, for example, by Carlin et al. [61]. I then review the continuous case, as described by Stephens [62].

2.4.2 The discrete-time changepoint problem

Suppose that a set of longitudinal data $Y_1, Y_2, ..., Y_n$ are observed at discrete times $t_1, t_2, ..., t_n$. Assume two densities f and g with respective unknown parameters θ and η. The discrete-time changepoint model is given by:

$$\begin{aligned} Y_i &\sim f(Y|\theta), i = 1, 2, ..., \delta, \\ Y_i &\sim g(Y|\eta), i = \delta + 1, ..., n, \end{aligned}$$

where δ corresponds to the *unknown* changepoint, i.e., the timing of the change of the data distribution. The interest is the estimation of the discrete changepoint δ,

as well as θ and η, the unknown parameters of the two distributions. The likelihood of the data $\mathbf{Y} = (Y_1, ..., Y_n)$ is given by:

$$L(\mathbf{Y}|\delta, \theta, \eta) = \prod_{i=1}^{\delta} f(Y_i|\theta) \prod_{i=\delta+1}^{n} g(Y_i|\eta).$$

The joint distribution of the data and the parameters is therefore:

$$p(\mathbf{Y}, \delta, \theta, \eta) = L(\mathbf{Y}|\delta, \theta, \eta)\pi(\theta, \eta, \delta), \tag{2.7}$$

where $\pi(\theta, \eta, \delta)$ is the joint prior distribution on θ, η and δ. The joint posterior distribution of the parameters given the data is proportional to the joint distribution of the data and the parameters, therefore:

$$p(\delta, \theta, \eta|\mathbf{Y}) \propto L(\mathbf{Y}|\delta, \theta, \eta)\pi(\theta, \eta, \delta). \tag{2.8}$$

The three full conditional distributions of interest $p(\delta|\mathbf{Y}, \theta, \eta)$, $p(\theta|\mathbf{Y}, \delta, \eta)$, and $p(\eta|\mathbf{Y}, \delta, \theta)$ are proportional to the joint distribution (equation 2.7). They often require complex integration, but algorithms using MCMC algorithms can provide approximations to the marginal posterior distributions of interest.

Special case: conjugacy

Carlin et al. showed that in the case of conjugate families, it is possible to obtain the full conditional distributions in closed analytical forms [61]. Suppose that θ, η, and δ are independent, and have prior distributions given by $\lambda(\theta)$, $\gamma(\eta)$, and $\tau(\delta)$ respectively. Then $\pi(\theta, \eta, \delta) = \lambda(\theta)\gamma(\eta)\tau(\delta)$. If one assumes conjugacy between $\lambda(.)$ and $f(.)$ as well as $\gamma(.)$ and $g(.)$ then $\theta|\mathbf{Y}, \delta, \eta$ is independent of η, and is the prior distribution $\lambda(.)$ updated by the data. Similarly, $\eta|\mathbf{Y}, \delta, \theta$ is independent of θ, and is $\gamma(.)$ updated by the data. $\lambda(.)$ and $\gamma(.)$ have a standard parametric form, and it is therefore easy to sample from the full posterior conditional distributions. One can extend this set up to a more general hierarchical model. One still assumes

independence of θ, η and δ, but now adds another level to the hierarchical model, by assuming a parametric family for θ and η so that: $\theta \sim \gamma(\theta|\alpha)$ and $\eta \sim \gamma(\eta|\beta)$ with respective conjugate prior distributions $\rho(\alpha)$ and $\phi(\beta)$. The joint distribution of the data is now expressed as follows:

$$p(\mathbf{Y}, \delta, \theta, \eta, \alpha, \beta) = L(\mathbf{Y}|\delta, \theta, \eta)\tau(\delta)\lambda(\theta|\alpha)\rho(\alpha)\gamma(\eta|\beta)\phi(\beta) \tag{2.9}$$

Therefore:

- $\delta|\mathbf{Y}, \theta, \eta, \alpha, \beta$ is independent of α and β and more specifically because of the discrete form of the prior density of the changepoint:

$$p(\delta|\mathbf{Y}, \theta, \eta, \alpha, \beta) = \frac{L(\mathbf{Y}|\delta, \theta, \eta)\tau(\delta)}{\sum_{\delta=1}^{n} L(\mathbf{Y}|\delta, \theta, \eta)\tau(\delta)} \tag{2.10}$$

- $\theta|\mathbf{Y}, \delta, \eta, \alpha, \beta$ is independent of η and β, and its distribution is the prior $\lambda(.)$ updated by the data Y_1 to Y_δ,
- $\eta|\mathbf{Y}, \delta, \theta, \alpha, \beta$ is independent of θ and α, and its distribution is the prior $\gamma(.)$ updated by the data $Y_{\delta+1}$ to Y_n,
- $\alpha|\mathbf{Y}, \delta, \theta, \eta, \beta$ is independent of $\mathbf{Y}, \delta, \eta$ and β, and its distribution is the hyper-prior $\rho(.)$ updated by θ,
- $\beta|\mathbf{Y}, \delta, \theta, \eta, \alpha$ is independent of $\mathbf{Y}, \delta, \theta$ and α, and its distribution is the hyper-prior $\phi(.)$ updated by η,

For conjugate families, the Gibbs sampler easily estimates the set of marginal probability distributions. Instead of marginalizing the joint posterior distribution $p(\delta, \theta, \eta, \alpha, \beta|\mathbf{Y})$, the problem is simplified into a series of easier calculations requiring simple (univariate) posterior conditional distributions.

2.4.3 The continuous changepoint problem

It may be reasonable to assume that the data have some continuous distribution over time, and thus a continuous changepoint must be incorporated into the model.

Stephens extended the work by Carlin et al. to the continuous-time changepoint problem [62]. The problem is rewritten as follows:

$$\begin{aligned} Y_t &\sim f(Y|\theta), \delta \leq t, \\ Y_t &\sim g(Y|\eta), \delta > t. \end{aligned}$$

The likelihood is given by:

$$L(Y|\delta, \theta, \eta) = \prod_{\delta \leq t} f(Y_t|\theta) \prod_{\delta > t} g(Y_i|\eta).$$

Similar to the discrete-time changepoint model, once the prior distributions are specified, the joint distribution (equation 2.8) of the data and the parameters is obtained as the product of the likelihood function and the joint prior distribution. Again, MCMC algorithms can provide estimation for the posterior distributions of the parameter of the model.

2.5 Longitudinal studies of post-radiotherapy PSA series

Post-radiotherapy PSA series provide valuable information when monitoring treatment outcome; for this reason, several studies have examined the PSA profiles, in particular the PSA nadir and the PSA doubling time. The authors usually estimated these parameters, and then correlated them with clinical outcomes. I have presented the results of some of these studies, when describing the properties of the PSA data. I focuss now on the statistical methods used by these studies.

In section 2.1, I emphasized the considerable variations of the PSA observations. However, in most studies investigating the post-radiotherapy characteristics of the PSA series, this variability was not adequately accounted for. The PSA nadir, for example, was defined as the lowest *observed* PSA concentration following radiotherapy [31][32][33][65]. By taking the observations at their face value, the analysis ignored the PSA variations, and the resulting estimates were likely to be biased.

The PSA doubling time was estimated using various approaches, including exponential models [26][29][66][67][68], and splines [69]. Again, the estimation of the parameters of the PSA profiles did not adequately account for the variability of the PSA observations. First, the PSA measurements were again taken at their face values: the authors *visually* identified PSA series they believed were rising using some arbitrary definition (based on specific number of rising PSA concentrations), and then fitted the exponential model only to these specific rising profiles [26][29][66][68]. However, as discussed in the literature review, it is common to observe rising PSA patterns, while the underlying PSA are actually stable, and vice-versa. Second, all the models, except the model by Hanlon et al. [68], were fitted separately to each series. As a result, the average doubling time was estimated by pooling the individually estimated doubling times, without regard to the varying variabilities of the individual PSA series. In addition, by estimating each PSA profile independently, the estimation process ignored the information provided by the other series. Hanlon et al. fitted a classical exponential model based to their entire cohort [68]. However, their model does not appear to have adequately accounted for the characteristics of the PSA series. Although the authors called their model a random effect model, the most relevant parameters of the model seem to have been considered as fixed, and thus not allowed to vary across subjects. In particular, the changepoint was held fixed, at 20 months. In addition, the authors applied their model to rising PSA profiles only (defined as PSA rising on two consecutive occasions to a level greater than 1 ng/ml, or three consecutive rises).

Splines provided good individual fits [69], but had the inconvenience of not accounting for the natural monotonicity of the PSA history. In addition, splines were fitted separately for each man, ignoring information provided by the rest of the cohort.

Longitudinal hierarchical models appear the most appropriate models to describe post-radiotherapy PSA series. These models were used by Slate and Cronin [49], and Pauler and Finkelstein [48], and their objective was the prediction of cancer recurrence. These two models were more sophisticated, and adequately accounted for the multiple characteristics of the PSA series.

Slate and Cronin were interested in the early detection of cancer recurrence using rules for biochemical failure, and proposed two distinct models. The first model included a single changepoint, representing the onset of the recurrence of the cancer, and the second model used two changepoints, where one changepoint represented the end of the transient effect of radiotherapy, and the second the onset of recurrence. The single changepoint model did not adequately capture the initial decline rate. However, the authors only included observations with the first measurements not exceeding 4 ng/ml. PSA concentrations tend to be much higher at the start of the treatment; by leaving out the first observations greater than 4 ng/ml, valuable information for the fitting might have been lost. The two changepoints model included three post-radiotherapy slopes: an initial slope until the end of the transient effect of radiotherapy, a second slope between the two changepoints, and a third slope after the recurrence of the cancer. Because of the high correlation between the two changepoints, convergence was much slower than for the single changepoint model. These changepoint models allowed the authors to obtain a posterior distribution reflecting the probability that there was a recurrence of the cancer. At the time of the test, the posterior probability that the changepoint had occurred was computed; if the probability exceeded some specified cutoff, then a positive result was indicated.

Pauler and Finkelstein used a two-stage model to predict time to cancer recurrence [48]. Similarly to Slate and Cronin, the authors used a hierarchical single changepoint model to describe post-radiotherapy PSA series. In addition, they used a Bayesian version of the Cox model to model the time to cancer recurrence as a

function of the parameters of the longitudinal model, such as whether or not the changepoint has occurred, and the recent PSA level. The authors compared their approach to traditional rules for biochemical failure, including the ASTRO criterion, and a time varying indicator of whether each of the rule has occurred was added to the Cox model. The authors concluded that the longitudinal model for PSA contributed significantly to the Cox model, and that no additional information was gained from any of these rules.

2.6 Summary

Slate and Cronin, and Pauler and Finkelstein have shown that the Bayesian hierarchical changepoint model appropriately accounts for the PSA variability. Given its flexibility, and ease of implementation, this model could easily incorporate additional complex features of the PSA data. For example, given the lower precision of the measurement tools, it would be interesting to express the PSA variability as a function of the PSA concentration.

Studies that examined criteria for biochemical failure, such as the ASTRO criterion, focused on their clinical validation, rather than their numerical validation. The interest was in the prediction capabilities of the rules, rather than their capacities to detect a specific signal, such as an increase in the PSA concentrations. A fundamental point, before one considers how well even a perfectly measured PSA trajectory correlates with clinical outcomes, is how good the rules are at correctly (and quickly) identifying 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. Such *numerical validation* involves comparing observed PSA series to the underlying true PSA trajectories. Given the flexibility and the sophistication of the hierarchical models, I will show that underlying PSA profiles can be appropriately estimated. Furthermore, I will show that the models easily allow ones to express the PSA variability as a function

of the PSA concentrations. Based on my estimates, I will then propose the first numerical validation study of rules for biochemical failure.

Preamble to Manuscript I

The principal objective of this thesis is the evaluation of the classification properties of the ASTRO criterion. I will create a theoretical model, from which I will be able to predict how the ASTRO system performs, as a function of the true trajectory, and other variables. The purpose of this first manuscript is to provide these realistic 'true' trajectories, and the amount of variation encountered in real life. I estimate these inputs by taking the values from a set of 470 well-documented PSA series in actual patients. If radiotherapy is successful, PSA levels reach a nadir, and remain low or possibly rise very slowly. A sustained steep increase indicates biochemical failure. I use a Bayesian hierarchical changepoint model to estimate the individual and population PSA profiles, as well as the variability of the PSA series.

The Bayesian hierarchical changepoint model is particularly flexible and accommodates the multiple complex characteristics of the PSA series: the within-series correlation, the within- and between- variability, the piecewise linear pattern on both sides of the man-specific changepoint, the unbalanced format of the data, as well as the non-constant variance.

Thanks to its flexibility and ease of implementation, the Bayesian hierarchical models easily accommodate longitudinal data, and should be applicable to the study of other marker series in other diseases.

This manuscript has been submitted to the *American Journal of Epidemiology*, and follows the submission guidelines of this journal. The references are included in the global thesis bibliography.

CHAPTER 3

Manuscript I - Detecting post-radiotherapy PSA failure in men with prostate cancer.

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

[1] Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Qc.

[2] Division of Urology, University of Connecticut Health Center, Farmington, CT.

Abstract

Biomarkers provide valuable information when monitoring disease onset or progression; examples include bone mineral density (osteoporosis), cholesterol (coronary artery diseases), and prostate-specific antigens (PSA, prostate cancer). The analysis of marker series has to incorporate the within-series correlation, the within- and between-series variability, the possibly complex pattern of the series over time, and the unbalanced format of the data.

We illustrate the practicality of the Bayesian hierarchical model, by describing individual post-radiotherapy series from 470 men with prostate cancer. We estimate the four parameters of the series at both the individual and group levels: the PSA nadir, its timing, the PSA half-life preceding the nadir, and the subsequent PSA doubling time. Given the lower detection levels at lower PSA concentration, we also estimate the variability of the PSA observations as a function of the concentration. Compared to most earlier studies, our estimates were more precise, and results indicate that the PSA nadir was reached earlier.

We emphasize the flexibility of hierarchical models. They accommodate the multiple complex features of longitudinal PSA series, and we show that they easily permit to more realistically model the variability of the PSA concentration. These models are applicable to the study of other marker series in other diseases.

3.1 Introduction

A biomarker is an objective laboratory measure of a biological parameter that can act as an indicator of a disease state. Examples include bone mineral density (osteoporosis), cholesterol (coronary artery diseases), CD4 T-cells count (HIV/AIDS), or prostate-specific antigens (prostate cancer). Biomarkers are extensively used, and the analysis of a long series can provide valuable information when monitoring disease or treatment outcome.

The analysis of biomarker series requires special care since it has to accommodate multiple characteristics of longitudinal data. First, repeated observations are usually available on each individual, and thus are likely to be correlated. Second, apart from changes in the true biomarker level, measurement errors and short term biological variations create within-series variability. Data collection, and laboratory measurement procedures are important sources of variability, such as for example cholesterol or PSA concentrations [2][70]. In addition, measurement tools can have varying measuring precisions; for example, laboratory tests are often less precise at low PSA concentrations [2][21]. Biological variations are usually less controllable than the measurement variability since not all the causes of variations are known. Third, biomarker concentrations may not have constant growth rates, and traditional linear regression techniques cannot be applied. Suitable transformations may simplify the overall pattern; the square root of the CD4 T cells number has been shown to be piecewise linear following seroconversion [63], and similarly for log_2PSA series following radiotherapy [23]. Piecewise-linear patterns require special attention, since specific techniques are needed to estimate the changepoint, the point of intersection of the two lines. Finally, marker data may not have been obtained in the context of a controlled study; the length of follow-up, the number, and the frequency of the measurements can vary considerably from subject to subject.

Various statistical models have been used to estimate individual and population marker profiles. The classical random-effect model [71] has been applied for the modeling of CD4 cell counts during HIV infection [72][73], and natural history of PSA progression in prostate cancer screening [47]. Bayesian hierarchical models are particularly appropriate in the presence of a random changepoint. They have been successfully applied to describe series of the square root of CD4 T-cell counts for the progression of HIV infection [63], and log_2PSA series to monitor radiotherapy outcomes in men with prostate cancer [49]. An advantage of the Bayesian approach is the ability to obtain a direct estimation of the probability that the changepoint, and thus disease onset or recurrence, has occurred.

The objective of this article is to take advantage of the flexibility and accuracy of hierarchical changepoint models to describe post-radiotherapy PSA series. Our analysis included an important feature not accounted for in past studies using this type of model: we have modeled the variance of the PSA observations as a function of the PSA concentrations. This is particularly appropriate, given the known lower levels of detection of the PSA measurement tools at lower PSA concentrations [2][21]. We emphasize that this model provides much more accurate estimates than a naive (and common) approach that ignores the PSA variability.

We apply our model to a set of 470 longitudinal PSA series from men treated with radiotherapy for prostate cancer. If radiotherapy is successful, PSA levels reach a nadir, and remain low or possibly rise very slowly. A sustained steep increase indicates biochemical failure. We provide estimates for the parameters of the profiles (the PSA nadir, its timing or changepoint, the PSA decline rate before the nadir, and the subsequent PSA growth rate), as well as their variability.

Our model could bring important clinical insights. First, it has been suggested that the PSA growth rates, and the PSA nadir indicate cancer recurrence. However,

most of the published results were based on statistical methods that were not optimal, as most studies ignored the random-variability of the PSA measurements. For example, linear regression analysis was applied independently to each PSA series, taking each PSA observation at their face value, and the PSA growth rate was estimated from PSA series that appeared to be rising, i.e., from series satisfying some arbitrary definition (for example, at least two consecutive PSA rises) [26][27][29]. Similarly, the lowest observed PSA concentration was used as an estimate of the PSA nadir [32][33]. Our model will provide more precise estimates of these parameters, and allow one to more accurately evaluate their associations with clinical variables. Second, a sharp PSA increase usually indicates biochemical (or PSA) failure. As of today, evaluations of the performance of rules for biochemical failure reported discordant findings, essentially because the PSA variability was ignored, or not adequately captured. A precise estimation of PSA series and their variability will provide new tools to evaluate the performance of such rules.

Our paper proceeds as follows. In section 3.2, we describe post-radiotherapy PSA series from 470 men treated for prostate cancer in Connecticut. We then describe two statistical methods used for the estimation of the individual and group PSA profiles. In a preliminary analysis, we estimated each PSA profile independently, and pooled the estimates using a simple average. Although this approach is obviously far from optimal, it has often been used to estimate the post-radiotherapy PSA doubling time, and illustrates the consequences of ignoring the PSA variability. We then analyzed the data using a more complex (but also time consuming) method, and fitted a Bayesian hierarchical changepoint model. We detail the model which allows for individual changepoints, and the variance to be a function of the PSA concentrations. In section 3.3, we present results of the two statistical methods for

the Connecticut data set. We then use the estimates of the parameters of the hierarchical model to investigate associations of the PSA profile with pre-treatment patient characteristics, such as age at diagnosis, initial Gleason score of the tumor, and initial PSA concentration. We discuss our findings in section 3.4.

3.2 Materials and methods

3.2.1 The Connecticut PSA data set

The data were initially collected for an earlier study aimed at linking rising PSA profiles following treatment (surgery or radiotherapy) of prostate cancer to ten-year outcomes [25]. The data were assembled retrospectively, on a population based cohort identified by the Connecticut Tumor Registry. The men were aged 75 years or younger, and residents of Connecticut when diagnosed with localized cancer between 1990 and 1992. Men with advanced disease or initial PSA higher than 50 ng/ml were excluded. PSA values were recorded from the ambulatory records located primarily in urologists' offices, but also from ambulatory records located in the offices of radiation oncologists, medical oncologists, the Connecticut Tumor Registry, and inpatient records. More details are available in Albertsen et al. [25]. In some cases, men can receive a subsequent treatment, usually in the form of hormones. For the purposes of this study, we have excluded any PSA measurements under a hormonal therapy. Finally, we required each PSA series to have at least one pre-treatment PSA measurement as well as two post-treatment PSA measurements; 470 series satisfied our conditions.

Men with a secondary treatment reach their PSA nadir much sooner, with a steeper post-nadir PSA growth rate, than men without such treatment. For this reason, we decided to split the men into two subgroups. One subgroup consisted of the 139 men who subsequently received secondary treatment; the second subgroup included the remaining 331 men. This somewhat arbitrary division allowed us to

obtain two subgroups, each with relatively homogeneous patterns, and thus to speed up the computational phase of the analysis using the hierarchical model. Two analyses were therefore performed independently, and results are reported separately for each subgroup.

3.2.2 Statistical methods

Figure 3–1 shows post-radiotherapy PSA series over time for eight men, plotted on the log_2 scale (page 62). The time axis starts at the initiation of the treatment. Notice the typical V-shape of the log_2PSA series: following the start of the treatment, the levels drop to some nadir level, and then increase again at various rates. We can also observe the important variations between and within series.

Figure 3–2 illustrates a prototypic PSA profile plotted on the log_2 scale, for a specific man i (page 63). We denote by ϕ_{1i}, ϕ_{2i}, ϕ_{3i} and ϕ_{4i} the log_2PSA nadir, the changepoint (or location of the nadir), the log_2PSA decline rate prior to the PSA nadir (the slope of the first line), and the post-nadir log_2PSA growth rate (the slope of the second line). We will estimate these four parameters, as well as the PSA variability.

We have used the logarithm on the base 2 of the PSA level, log_2PSA, as the response variable. This transformation has several advantages. First, it allows one to obtain a piecewise linear pattern. Second, the series tend to be smoother than when plotted on the natural scale. Finally, when the unit of time is the year, the post-nadir log_2PSA 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. Similarly, the log_2PSA decline rate before the nadir is equivalent to the number of PSA halvings per year, and its reciprocal corresponds to the PSA half life. We have preferred to perform the statistical analysis based on the rates (slopes), in accord with classical least-squares regression.

Preliminary analysis

For comparison purposes, we performed a preliminary naive analysis. We took the PSA data at their face value, and thus ignored any measurement errors or short-term biological variations. We refer to this approach as the simplistic approach. For each man, we visually identified the lowest observed log_2PSA observation, and considered it as the log_2PSA nadir. Similarly, the observed timing of the PSA nadir corresponded to the timing of the lowest log_2PSA observation. The rate of log_2PSA decline for each man was estimated by the slope of the least squares regression line fit to the log_2PSA values versus time in years, for PSA values from pre-treatment up until and excluding the PSA nadir. Similarly, log_2PSA growth rate was estimated by the slope of the regression line fit to log_2PSA values from and including the log_2PSA nadir to the last log_2PSA observations available. The mean parameters were obtained using a simple average of the individual estimates, and we reported the 95% reference range using the 2.5^{th} and 97.5^{th} percentiles.

Bayesian hierarchical changepoint model

The principal analysis of this paper consists of a Bayesian hierarchical model with a changepoint. This model provided estimates for the four parameters of the log_2PSA profile at both the individual and aggregate levels. The term *hierarchical* is employed since this approach models the data as several levels, which is appropriate when there are wide between-subject variations in the parameters [50]. At the first level, the log_2PSA profile is described using the data distribution, or likelihood function. We let log_2PSA_{ij} be the PSA concentration on the log_2 scale, for the j^{th} measurement for the i^{th} man, and assume that the observations log_2PSA_{ij} are normally distributed, with mean μ_{ij} and variance σ^2_{ij}:

$$log_2PSA_{ij} \sim N(\mu_{ij}, \sigma^2_{ij}). \tag{3.1}$$

For every observation $log_2 PSA_{ij}$, the expected log_2PSA value, μ_{ij}, is related to the timing of the measurement t_{ij} through a regression function describing the expected profile before and after the changepoint ϕ_{2i}:

$$\mu_{ij} = \begin{cases} \phi_{1i} + \phi_{3i}(t_{ij} - \phi_{2i}), & t_{ij} < \phi_{2i}, \\ \phi_{1i} + \phi_{4i}(t_{ij} - \phi_{2i}), & t_{ij} \geq \phi_{2i}. \end{cases}$$

In addition, because, the interassay coefficients of variation tend to be larger at lower PSA levels [2][21], therefore, we have expressed the logarithm of the precision $\frac{1}{\sigma^2_{ij}}$ as a linear function of the PSA levels:

$$\log \frac{1}{\sigma^2}_{ij} = \theta_1 + \theta_2 \log_2 PSA_{ij}. \tag{3.2}$$

At the second level, the individual parameters are viewed as a random sample from a common population. For ease of exposition, consider the slope prior to the nadir; the same principles apply to the other three parameters. During this initial phase, the individual PSA concentrations vary with different rates, as illustrated by figure 3–3 (page 63). The hierarchical model assumes that these individual slopes ϕ_{3i} have been randomly selected from a common population of slopes, with common mean and variance. In particular, we selected this hierarchical distribution to be normal, $\phi_{3i} \sim N(\mu_{\phi_3}, \sigma^2_{\phi_3})$. We chose the normal distribution, since the estimates of the slopes as provided by the preliminary analysis looked close to normal. The parameters μ_{ϕ_3} and $\sigma^2_{\phi_3}$ correspond respectively to the population slope and the between-men variability of the slopes. The essence of hierarchical modeling is the simultaneous inference of the individual (ϕ_{3i}) and the population parameters (μ_{ϕ_3}, $\sigma^2_{\phi_3}$), allowing for the borrowing of strength. The individual slopes are estimated using what is effectively a weighed average of the observed slope and the population slope, where the weights are the corresponding precisions. As a result, the individual

slopes are pulled toward the population slope: the less the precision of the individual data compared to the precision of the population estimates, the more they will be pulled. Similarly, we assigned a normal distribution to the PSA nadir, and the post-nadir slope. The distribution of the changepoint was specified using a diffuse continuous uniform distribution; we used a range of five and ten years, for the subgroups respectively with and without a secondary treatment.

At the third level, the parameters of the second level hierarchical distributions, such as μ_{ϕ_3} and $\sigma^2_{\phi_3}$, are also assigned distribution functions. The nadir, the slope before, and the slope after the nadir were assigned diffuse normal means and diffuse uniform standard deviations, as follows:

$$\begin{aligned}
&\phi_{1,i}|\mu_{\phi_1}, \sigma^2_{\phi_1} \sim N(\mu_{\phi_1}, \sigma^2_{\phi_1}), \quad \mu_{\phi_1} \sim N(0, 100), \quad \sigma_{\phi_1} \sim U(0, 4), \\
&\phi_{3,i}|\mu_{\phi_3}, \sigma^2_{\phi_3} \sim N(\mu_{\phi_3}, \sigma^2_{\phi_3}), \quad \mu_{\phi_3} \sim N(0, 100), \quad \sigma_{\phi_3} \sim U(0, 4), \\
&\phi_{4,i}|\mu_{\phi_4}, \sigma^2_{\phi_4} \sim N(\mu_{\phi_4}, \sigma^2_{\phi_4}), \quad \mu_{\phi_4} \sim N(0, 100), \quad \sigma_{\phi_4} \sim U(0, 4).
\end{aligned}$$

We also assigned diffuse normal priors to the parameters of the PSA variation: $\theta_1 \sim N(0, 100)$, and $\theta_2 \sim N(0, 100)$. The statistical analysis was implemented in Winbugs, a statistical software package that uses Markov Chain Monte Carlo techniques to generate the posterior distributions [55]. Point estimates of parameters were estimated using the mean of the posterior distribution, and 95% credible intervals (CI) were reported based on the 2.5^{th} and 97.5^{th} percentiles of the posterior distribution. For each model, we generated three chains with distinct sets of initial values. For each chain, we ran an initial burn-in period of 2,000 iterations, and an additional set of 10,000 iterations. After assessing convergence using the Raftery and Lewis criterion [54], we pooled the three chains, and used the 30,000 iterations to estimate the posterior distributions for each parameter.

Finally, we investigated the associations of the individual level parameters of the PSA profile with baseline characteristics. We constructed four independent Bayesian

multiple linear regression models; the dependent variable was one of the four parameters of the PSA profile. The independent variables were the age at diagnosis of the cancer, the pre-treatment PSA concentration, and the pre-treatment Gleason score. The Gleason score refers to the aggressiveness of the cancer, and ranges from 2 to 10; the higher the number, the more aggressive the tumor [8]. For the purposes of this analysis, we merged the two subgroups of men, by using all the output from both analyses. We also investigated possible associations between the four estimated parameters, again using four independent Bayesian multiple regression analysis.

3.3 Results

The shortest and longest PSA series had three and 36 measurements respectively, with an average of nine PSA readings per series. The shortest and longest durations of follow-up were respectively four months and 12 years long, with a mean follow-up duration of 5.7 years following the initiation of radiotherapy.

Descriptive statistics for the three pre-treatment covariates are given in table 3–1 (page 64). Men who underwent subsequent treatment were about the same age as those who did not, but tended to have higher PSA level and Gleason score at baseline.

3.3.1 Estimation of the log_2PSA profiles

Preliminary analysis

Summary statistics for the four parameters of the PSA profile using the simplistic approach are provided in table 3–2 (page 65). Histograms are given in the first columns of figures 3–4 and 3–5 for, respectively, the subgroup of men who underwent secondary treatment and those who did not (pages 66 and 67). We excluded three series from the analysis of the number of PSA halvings in the first phase. The

first observation of these series corresponded to the lowest observed PSA measurement, and thus no regression line could be drawn in the initial phase. Based on 136 men, the average log_2PSA decline rate was -3.37/year, corresponding to an average of 3.37 PSA halvings per year. The median number of PSA halvings per year was 2.57, corresponding to a median half-life of 0.39 years. The 139 mean had an average post-nadir log_2PSA growth rate of 1.51, and thus an average of 1.51 PSA doublings per year. The estimated median number of PSA doublings per year was 1.05, corresponding to a median doubling time of 0.95 years.

The average log_2PSA decline rate for the 31 men without a secondary treatment was -2.22, corresponding to an average of 2.22 PSA halvings per year. The median number of PSA halvings per year was 1.72, corresponding to a median half-life of 0.58 years. The average post-nadir log_2PSA growth rate, or yearly number of PSA doublings, was 0.42. This average included 55 men with an estimated negative growth rate, and 16 with a null growth rate. The PSA doubling time was estimated from the 315 men with an estimated growth rate that was greater or equal to zero; in such case, estimated median number of PSA doublings per year was 0.32, corresponding to a median doubling time of 3.12 years.

Bayesesian hierarchical changepoint model

Summary statistics for the estimates of the parameters of the hierarchical model are provided in the second half of table 3–2 (page 65), and histograms are given in the second columns of figures 3–4 and 3–5 (pages 66 and 67). We estimated the four parameters, including the numbers of PSA halvings and doublings, for each series. In particular, in the subgroup with a subsequent treatment, the estimate of the median number of halvings per year was 3.50, corresponding to a median half life of 0.29 years. The estimated median number of doublings per year was 1.66, corresponding to a median doubling time of 0.60 years. For the 331 men without a secondary

treatment, the median number of halvings per year was 3.29, corresponding to a median half life of 0.30 years. The median number of doublings per year was 0.28, corresponding to a median doubling time of 3.57 years.

The variance parameters θ_1 and θ_2 were similar for the two subgroups. In addition, the estimate of θ_2 was positive, suggesting, as expected, that the variance of the PSA observations decreased at higher concentrations. Modeling the variability was particularly appropriate given the wide discrepancies between lower and higher concentrations. Indeed, for example, at mean log_2PSA concentrations of 1 and 4 (and thus PSA levels of 2 ng/ml and 16 ng/ml), the log_2PSA variability estimate was respectively 0.36 and 0.14, corresponding to coefficients of variation of 60% and 10%.

Comparisons of the two models

The PSA nadir was estimated to occur much earlier using the hierarchial model. This is particularly true for the subgroup of men without a second treatment; in this case, the nadir was reached about one year earlier using the hierarchical model, as compared with the simple model. Overall, the other parameter point estimates were relatively close when comparing the two analysis, but the credible intervals were systematically much tighter using the hierarchical model. This is also well illustrated by the spreads of the histograms.

In figure 3-6, we plotted the estimated log_2PSA profiles for eight men using both methods (page 68). In the simplistic method, the profiles are estimated independently from each other, and the random-variation is ignored. The fitted lines tend to follow the observations very closely, including the outlying ones, which are also taken as true. On the other hand, the hierarchical model allows for the borrowing of strength. By simultaneously estimating the individual and group profiles, the

model 'learnt' from the rest of the data set, and 'guessed' that the extreme observations were more likely due to random variability than true changes in the PSA levels (see, for example, the second series). When averaging the individual and group profiles, the hierarchical model put less weight to these outlying observations, while more weight was given to the population profile; extreme observations were shrunk towards the population average profile.

3.3.2 Association of PSA model features with baseline characteristics

We investigated the associations of each of the four individual level parameters of the PSA profile with baseline characteristics. We performed four independent linear regression models, where the dependent variable was one of the four parameters as estimated by the hierarchical model. Estimates for the parameters of the regression analysis are reported in table 3–3 (page 69). Each line of the table corresponds to one independent multiple regression analysis, with the three columns as the independent variables (age of the patient at diagnosis, pretreatment PSA level, and Gleason score). A higher initial PSA level was found to be associated with more PSA halvings per year (longer half-life), a higher PSA nadir, and more PSA doubings per year (shorter doubling time). A higher initial Gleason score was associated with more PSA doubings per year (shorter doubling time).

Table 3–4 shows the results of the regression analysis used for the assessment of association within the four parameters of the PSA profile (page 69).Each line of the table corresponds to one independent multiple regression analysis, with the three columns as the independent variables; foe this reason the table is not symmetrical (as a correlation table). A steeper decline rate was associated with a lower PSA nadir, and a shorter time to the nadir. Finally, there was a positive association between the PSA nadir and the $log_2 PSA$ growth rate.

3.4 Discussion

We have proposed a hierarchical changepoint model that adequately describes marker series, such as post-radiotherapy PSA series. This model easily accommodates the usual characteristics of longitudinal data series. In addition, we have shown that the model allowed us to model the variance as a function of the marker concentration. Being able to include this feature is particularly appropriate, given that measurement tools are known to have varying precisions depending on the marker level [2][21]. The estimation of these PSA series could be further improved by including additional covariates in the hierarchical model, such as the baseline PSA level and tumor stage, which have been shown to predict tumor growth.

Our cohort may not be perfectly representative of the general population or of other cohorts investigated in past studies. However, we can compare the two subgroups of men with respect to the secondary treatment within our study. In particular, men who received a secondary treatment, and thus those who tend to be the more severe cases, had a much shorter PSA doubling time, and reached the PSA nadir much sooner. In addition, the PSA nadir was higher for these men, than those who did not get any additional treatment.

The fundamental message conveyed by our analysis is the considerable improvement in precision allowed by the hierarchical model. First, unless appropriate weighting is used to adequately reflect the precision of each individual series, regression techniques such as our simplistic approach, are particularly misleading. When estimating population average profiles, long and rich series, should not be given the same weight as short and less informative series. Second, PSA series should not be analyzed independently from each other, as all the series provide valuable information. Each of the series provides some information about the other ones, and vice-versa. The hierarchical model accounts for the varying precision of the individual series, and allows us to borrow strength from all series when estimating individual profiles. The

individual profiles are estimated using what is effectively a weighed average of the observed profile and the population profile, where the weights are the corresponding precisions. As a result, the individual profiles are pulled toward the population profile: the less the precision of the individual data compared to the precision of the population estimates, the more they will be pulled. As a result, inference is strengthened, and credible intervals are much tighter than the ones obtained using a naive approach. This is particularly striking with regard to the PSA changepoint, that we estimated to occur around 18 months in men without a second treatment. The simplistic approach estimated that this nadir occured about one year later. Very few studies reported estimates of the changepoint, and estimates were highly variable. However, Hanlon et al. estimated that the PSA nadir was reached on average 36 months after the initial day of radiotherapy, with a range of 2 to 114 months, again much later than our estimate [31].

Our model provided more precise estimates of individual post-radiotherapy trajectories. As a result, the estimated features of the PSA profiles, such as the nadir, or the PSA doubling time, can be used as input data in subsequent analysis. We used our estimates to investigate associations with pre-treatment variables. Similarly, Pauler and Finkelstein modeled the time to cancer recurrence as a function of the parameters of the PSA profiles, also estimated using a hierarchical model [48].

Our results provide valuable information for the monitoring of treatment outcome, based on biochemical failure, which is defined as a recurrence of the cancer detected by rising PSA levels. The American Society for Therapeutic Radiology and Oncology (ASTRO) consensus panel considers three consecutive PSA rises as an appropriate definition of biochemical failure following radiation therapy [37]. However, this criterion has undergone limited formal evaluation. Thus far, evaluations have ignored the PSA variability, and were limited to correlations with clinical outcomes. A fundamental point, before one considers how well even a perfectly measured PSA

trajectory correlates with clinical outcomes, is how good the ASTRO rule is at correctly (and quickly) identifying 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. This numerical validation involves comparing observed PSA series to the underlying true PSA trajectories, and appears feasible using our precise estimates.

Finally, specific characteristics of the PSA profiles could be used to propose new rules for biochemical failure. For example, Slate and Cronin proposed a rule for biochemical failure based only on the estimation of the changepoint, which represents the onset of the recurrence of the cancer [49]. At the time of the test for a specific subject, they estimated the probability that the changepoint had already occurred; the performance of the rule was assessed by estimating its sensitivity and specificity in men with and without recurrence respectively. Similarly, we can easily imagine rules for biochemical failure that would be based on the probability that the post-nadir PSA growth rate is greater than some specific threshold.

Thanks to its flexibility and ease of implementation, the Bayesian hierarchical changepoint model easily accommodates features of longitudinal PSA series, and should be applicable to the study of other marker series in other diseases.

3.5 Tables and figures

Figure 3–1: Post-radiotherapy log_2PSA concentrations over time for eight men

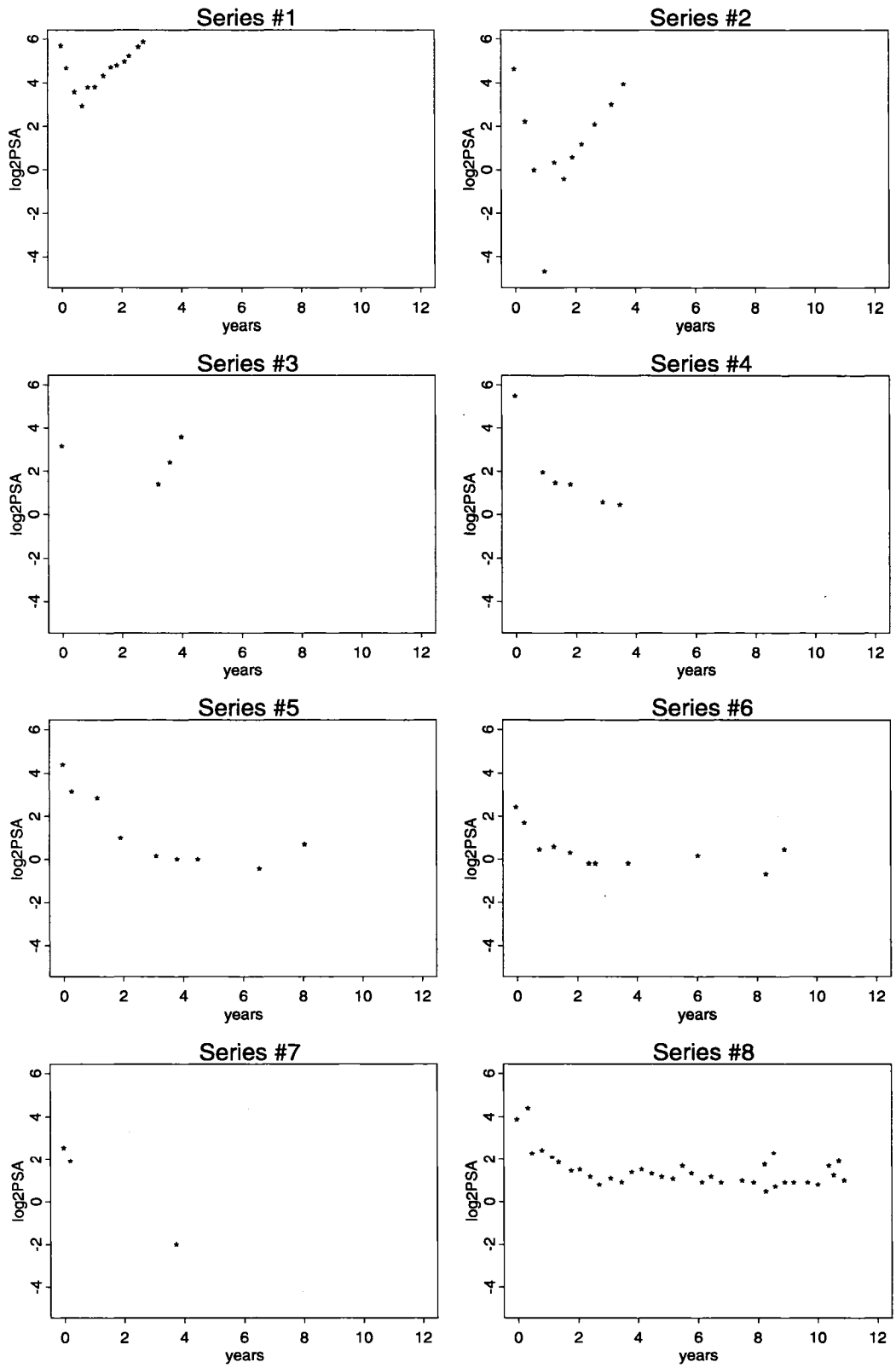


Figure 3–2: Individual piecewise linear model, with the four individual parameters.

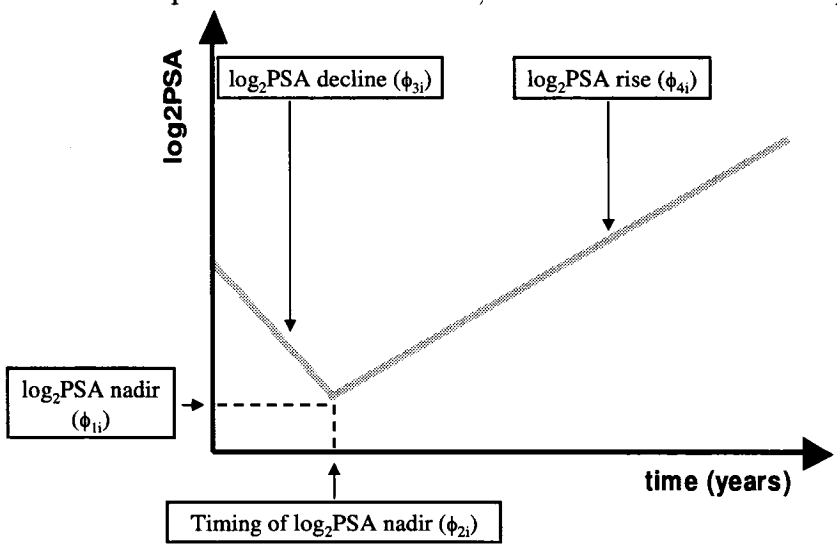


Figure 3–3: Prototypic pre-nadir PSA slopes for six men

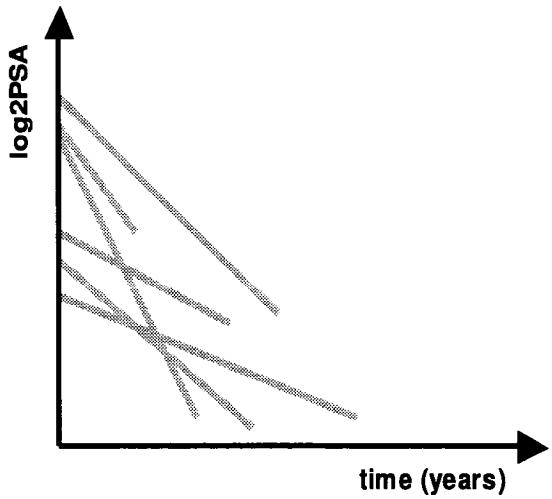


Table 3–1: Pre-treatment characteristics of the 470 men and their tumours

	139 men who subsequently underwent secondary treatment	331 men who did not undergo secondary treatment
Age at diagnosis (years)		
average	69.4	70.4
median	70.3	71.3
range	(53.2; 75.9)	(49.2; 76.0)
Pre-treatment PSA level in ng/ml[1]		
0-3.9	2% (3)	9% (30)
4-9.9	24% (34)	45% (149)
10-19.9	34% (47)	31% (102)
20-50	40% (55)	15% (50)
Pre-treatment Gleason score[1]		
2-4	2% (3)	2% (7)
5	4% (5)	6% (21)
6	30% (42)	51% (168)
7	35% (48)	24% (78)
8-10	25% (35)	17% (56)
missing	4% (6)	0% (1)

[1] Proportions and counts (in parenthesis) are provided.

Table 3–2: Estimates for parameters obtained using the simplistic and the hierarchical models.

139 men with a secondary treatment

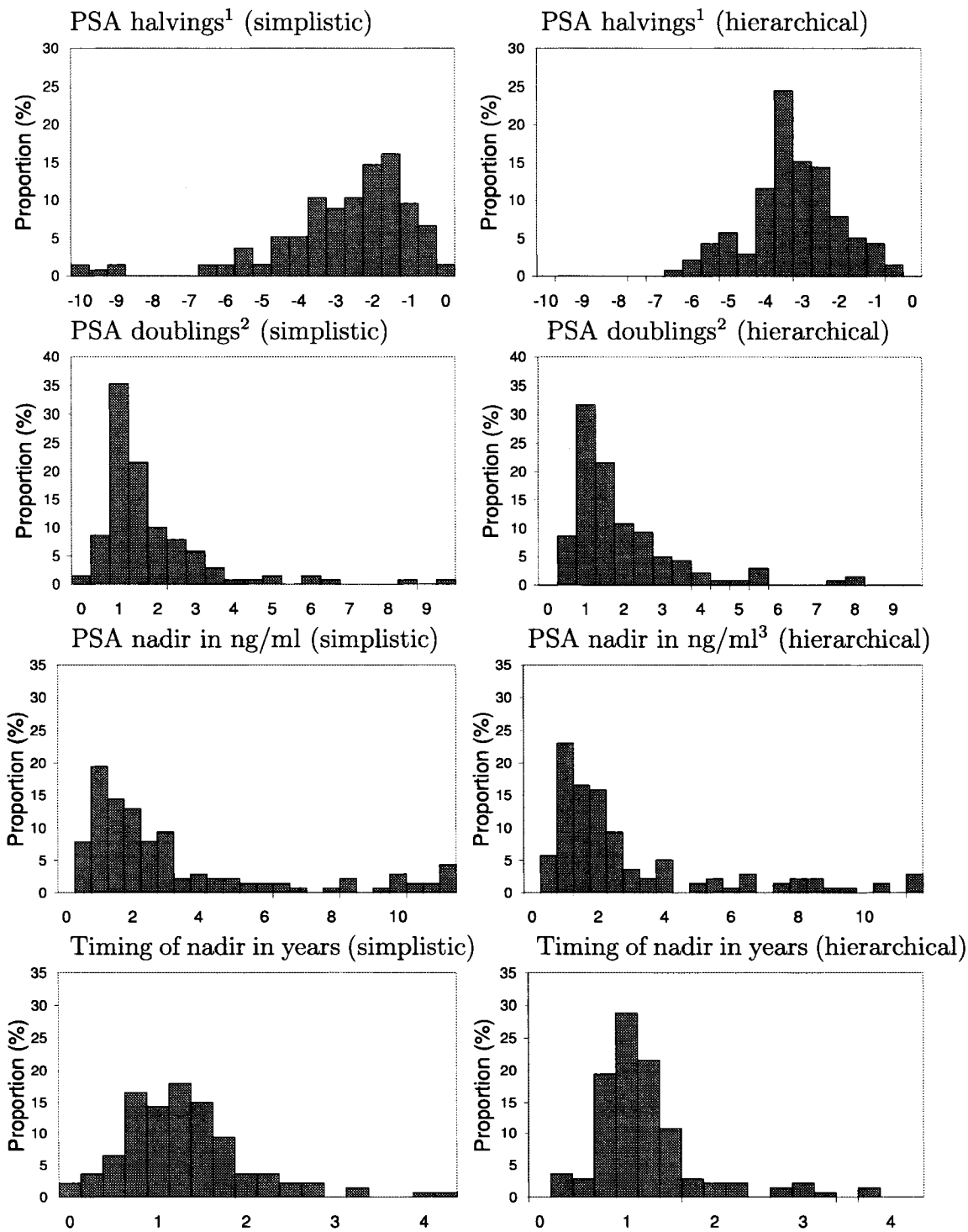
Parameter	Simplistic model[1]		Hierarchical model[2]	
log_2PSA decline rate (-PSA halvings)[3]	-3.37	(-9.18; -0.65)	-3.50	(-3.89; -3.12)
log_2PSA growth rate (PSA doublings)[4]	1.51	(0.11; 5.97)	1.66	(1.38; 1.95)
PSA nadir in ng/ml	3.48	(0.03; 16.8)	3.01	(2.75; 3.38)
Timing of the nadir in years	1.17	(0.17; 3.07)	1.08	(1.00; 1.17)
θ_1 (variance parameter[5])			0.72	(0.55; 0.87)
θ_2 (variance parameter[5])			0.31	(0.27; 0.36)

331 men without a secondary treatment

Parameter	Simplistic model[1]		Hierarchical model[2]	
log_2PSA decline rate (-PSA halvings)[3]	-2.22	(-8.26; -0.24)	-3.29	(-3.60; -2.99)
log_2PSA growth rate (PSA doublings)[4]	0.42	(-0.01; 1.85)	0.28	(0.23; 0.33)
PSA nadir in ng/ml	1.07	(0.03; 3.8)	1.25	(1.17; 1.35)
Timing of the nadir in years	2.58	(0.48; 7.17)	1.41	(1.29; 1.53)
θ_1 (variance parameter[5])			0.68	(0.62; 0.75)
θ_2 (variance parameter[5])			0.27	(0.24; 0.29)

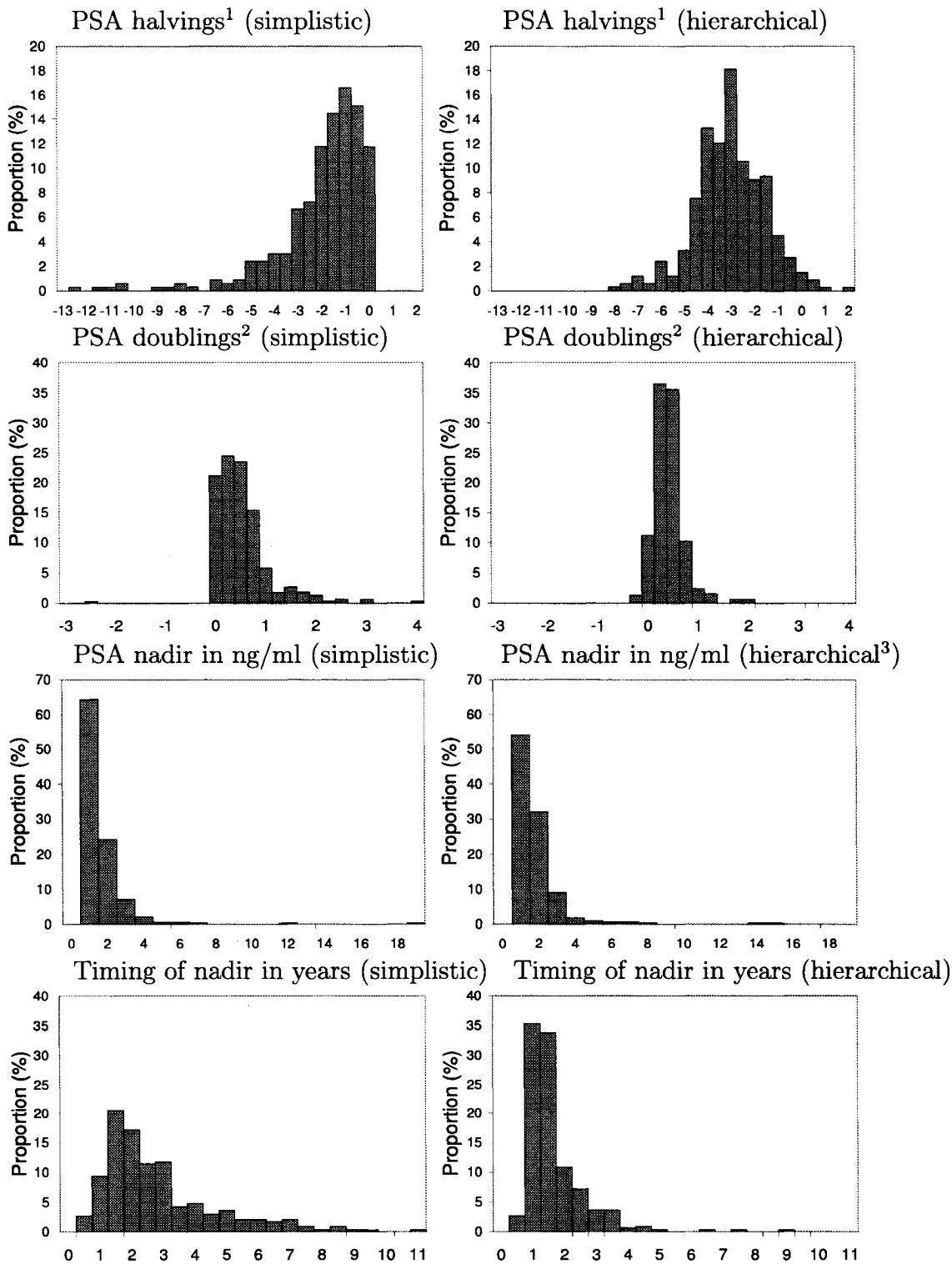
[1] The interval was estimated using the 2.5^{th} and 97.5^{th} percentiles.
[2] The 95% credible intervals are reported.
[3] Before the PSA nadir is reached.
[4] After the PSA nadir is reached.
[5] As given by equation 3.2.

Figure 3–4: Estimates for parameters obtained from the simplistic (left) and hierarchical (right) approaches, for the 139 men who underwent secondary therapy.



[1] Number of halvings before the PSA nadir is reached.
[2] Number of doublings after the PSA nadir is reached.
[3] For enhanced visualization, one estimated nadir was changed to 12 ng/ml, instead of 53 ng/ml.

Figure 3–5: Estimates for parameters obtained from the simplistic (left) and hierarchical (right) approaches, for the 331 men who did not undergo secondary therapy.



[1] Number of halvings before the PSA nadir is reached.
[2] Number of doublings after the PSA nadir is reached.

Figure 3–6: Post-radiotherapy PSA observations (dots) for eight men, with profile estimated by the simple (dashed line) and the hierarchical (straight line) models.

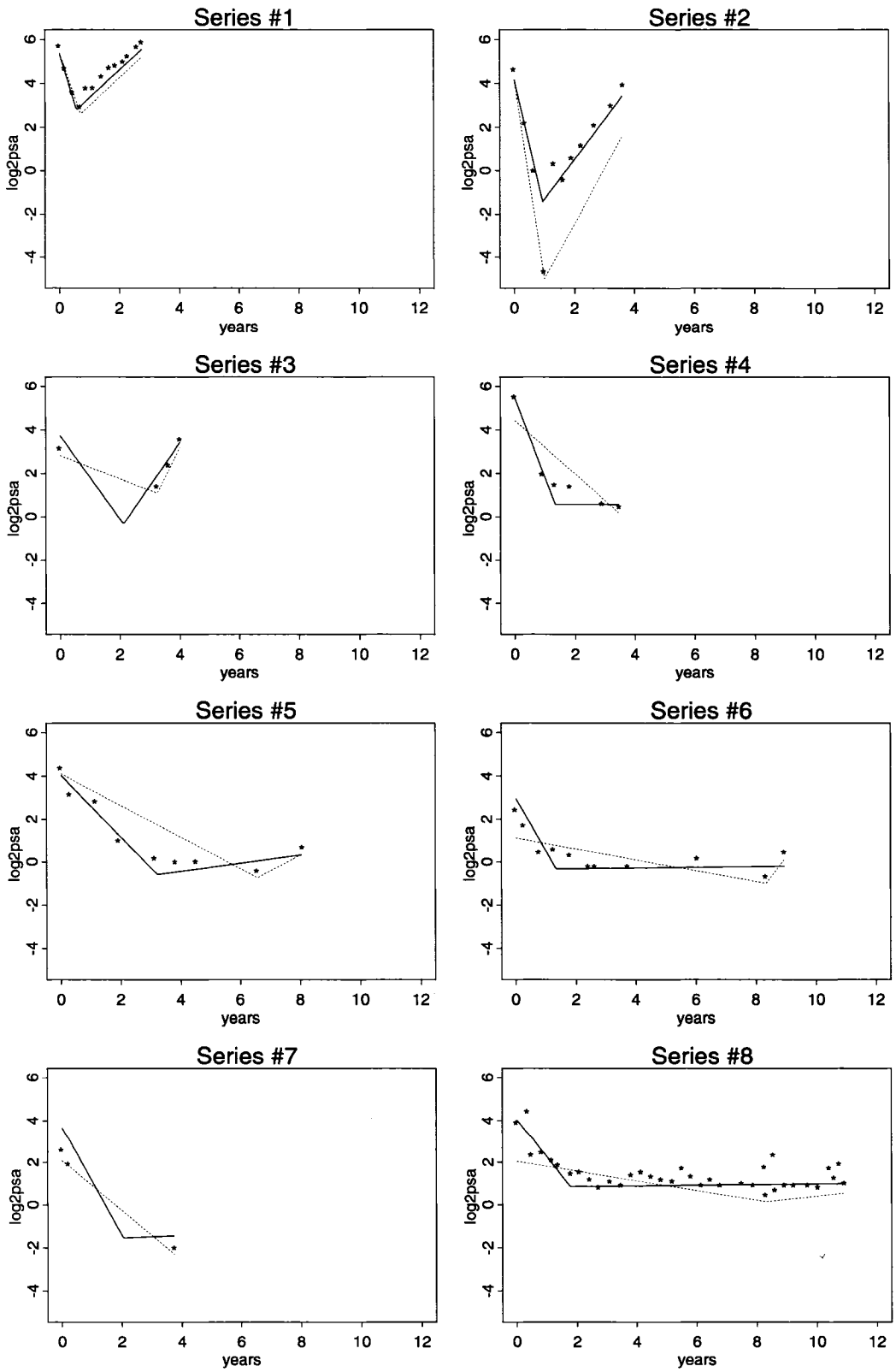


Table 3–3: Regression parameters from four independent analyses of associations between the parameters of the PSA profile and pre-treatment characteristics.

	Covariates		
Dependent variable	Initial PSA level	Gleason score	Age at diagnosis
log_2PSA decline rate[1]	-0.04 (-0.05; -0.03)	0.04 (-0.07; 0.15)	-0.03 (-0.06; 0.00)
log_2PSA growth rate[2]	0.02 (0.01; 0.03)	0.22 (0.14; 0.30)	0.00(-0.02; 0.02)
log_2PSA nadir	0.05 (0.04; 0.06)	0.09 (-0.01; 0.18)	-0.02 (-0.04; 0.01)
Timing of the nadir[3]	0.00 (-0.01; 0.00)	-0.04 (-0.10; 0.03)	0.00 (-0.02; 0.02)

[1] Before the PSA nadir is reached. It is equivalent to the negative of the number of PSA halvings.
[2] After the PSA nadir is reached. It is equivalent to the number of PSA doublings.
[3] In years
Each row corresponds to one independent regression analysis. For example, given a man with initial PSA level of 2 ng/ml (i.e. $log_2PSA = 1$), Gleason score of 2, and age 70 years, the estimated post-nadir PSA doublings per year is $0.02 * 1 + 0.22 * 2 + 0 * 70 = 0.46$. For a man with initial PSA level of 2 ng/ml (i.e. $log_2PSA = 1$), Gleason score of 10, and age 70 years, the estimated number of post-nadir PSA doublings per year is $0.02 * 1 + 0.22 * 10 + 0 * 70 = 2.2$.

Table 3–4: Regression parameters from four independent analyses of associations between the parameters of the PSA profile and pre-treatment characteristics.

	Covariates			
Dependent variable	Decline rate[1]	Growth rate[2]	PSA nadir	Timing
Decline rate[1]	-	-0.04 (-0.13; 0.05)	0.36 (0.28; 0.43)	1.07 (0.96; 1.17)
Growth rate[2]	-0.04 (-0.12; 0.05)	-	0.22 (0.15; 0.30)	-0.08 (-0.22; 0.05)
Nadir	0.46 (0.36; 0.55)	0.31 (0.20; 0.41)	-	-0.52 (-0.68; -0.36)
Timing	0.42 (0.38; 0.47)	-0.04 (-0.09; 0.02)	-0.16 (-0.21; -0.11)	-

[1] log_2PSA decline rate before the PSA nadir. It is equivalent to the negative of the number of PSA halvings.
[2] log_2PSA growth rate after the PSA nadir. It is equivalent to the number of PSA doublings.
Each row corresponds to one independent regression analysis.

CHAPTER 4
Additional material for manuscript I - Technical chapter

This chapter contains additional information for the hierarchical model used in the previous manuscript. In sections 4.1 and 4.2, I provide details for the conditional and prior distributions. I describe the practical implementation of the model in section 4.3, and provide the computer code and outputs in section 4.4.

4.1 Notation and Likelihood function

I have split the data into two subgroups depending on whether the men subsequently received a secondary treatment; two analyses were performed independently.

Let i and j be indices for the i^{th} man and the j^{th} measurement. Let N_1 and N_2 be the number of men, respectively, with and without a secondary treatment, so that $N_1 = 139$ and $N_2 = 331$. The next steps are performed independently for each subgroup and we use N generically for N_1 or N_2.

Let m_i be the total number of PSA measurements for man i and let $\eta_i(\phi_{2,i})$ be the total number of PSA measurements during the decline period.

Let $\phi_{1,i}$ be the nadir for man i, and $\phi_{2,i}$ be its index time. $\phi_{2,i}$ corresponds to a latent changepoint, and is allowed to occur any time, even after the follow-up period has ended for the given subject.

Let $\phi_{2,i}$, $\phi_{3,i}$ and $\phi_{4,i}$ be respectively the decline rate between the treatment initiation, the changepoint, and the growth rate following the changepoint.

Let Y_{ij} and t_{ij} be the log_2PSA concentration and its index time for man i on its j^{th} measurement. Let σ^2_{ij} be the variance of the log_2PSA measurement. I suspected that the variability was decreased at higher PSA levels, and specified a regression

model for σ_{ij}^2. Since a simple linear regression could lead to negative estimates of σ_{ij}^2, I have modeled the logarithm of the precision (the inverse of the variance): $\log \frac{1}{\sigma_{ij}^2} = \theta_1 + \theta_2 \log_2 PSA_{ij}$. Therefore:

$$\sigma_{ij}^2 = \exp\left(-(\theta_1 + \theta_2 Y_{ij})\right) \tag{4.1}$$

Conditional on $\phi_{1,i}, \phi_{2,i}, \phi_{3,i}, \phi_{4,i}, \theta_1$ and θ_2, I assumed:

$$\begin{aligned} Y_{ij} &\sim N(\phi_{1,i} + \phi_{3,i}(t_{ij} - \phi_{2,i}), \sigma_{ij}^2) \quad \text{for} \quad t_{ij} \leq \phi_{2,i}, \\ Y_{ij} &\sim N(\phi_{1,i} + \phi_{4,i}(t_{ij} - \phi_{2,i}), \sigma_{ij}^2) \quad \text{for} \quad t_{ij} > \phi_{2,i}. \end{aligned}$$

The conditional likelihood for man i is therefore:

$$\begin{aligned}
&\mathcal{L}(Y_i|\phi_{1,i}, \phi_{2,i}, \phi_{3,i}, \phi_{4,i}, \theta_1, \theta_2) \\
&= \prod_{t_{ij} \leq \phi_{2,i}} \left(\frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \right) \exp\left(-\frac{1}{2\sigma_{ij}^2} [Y_{ij} - \phi_{1,i} - \phi_{3,i}(t_{ij} - \phi_{2,i})]^2 \right) \\
&\quad \times \prod_{t_{ij} > \phi_{2,i}} \left(\frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \right) \exp\left(-\frac{1}{2\sigma_{ij}^2} [Y_{ij} - \phi_{1,i} - \phi_{4,i}(t_{ij} - \phi_{2,i})]^2 \right) \\
&= \prod_{t_{ij} \leq \phi_{2,i}} \left(\frac{1}{\sqrt{2\pi e^{-(\theta_1 + \theta_2 Y_{ij})}}} \right) \exp\left(-\frac{1}{2e^{-(\theta_1 + \theta_2 Y_{ij})}} [Y_{ij} - \phi_{1,i} - \phi_{3,i}(t_{ij} - \phi_{2,i})]^2 \right) \\
&\quad \times \prod_{t_{ij} > \phi_{2,i}} \left(\frac{1}{\sqrt{2\pi e^{-(\theta_1 + \theta_2 Y_{ij})}}} \right) \exp\left(-\frac{1}{2e^{-(\theta_1 + \theta_2 Y_{ij})}} [Y_{ij} - \phi_{1,i} - \phi_{4,i}(t_{ij} - \phi_{2,i})]^2 \right) \\
&\propto \left(\frac{1}{e^{-\frac{1}{2}(\theta_1 + \theta_2 Y_{ij})}} \right)^{\eta_i(\phi_{2,i})} \exp\left(-\frac{1}{2e^{-(\theta_1 + \theta_2 Y_{ij})}} \sum_{t_{ij} \leq \phi_{2,i}} [Y_{ij} - \phi_{1,i} - \phi_{3,i}(t_{ij} - \phi_{2,i})]^2 \right) \\
&\quad \times \left(\frac{1}{e^{-\frac{1}{2}(\theta_1 + \theta_2 Y_{ij})}} \right)^{m_i - (\eta_i(\phi_{2,i}))} \exp\left(-\frac{1}{2e^{-(\theta_1 + \theta_2 Y_{ij})}} \sum_{t_{ij} > \phi_{2,i}} [Y_{ij} - \phi_{1,i} - \phi_{4,i}(t_{ij} - \phi_{2,i})]^2 \right) \\
&= \left(\frac{1}{e^{-\frac{1}{2}(\theta_1 + \theta_2 Y_{ij})}} \right)^{m_i} \\
&\quad \times \exp\left(-\frac{1}{2e^{-(\theta_1 + \theta_2 Y_{ij})}} \left(\sum_{t_{ij} \leq \phi_{2,i}} [Y_{ij} - \phi_{1,i} - \phi_{3,i}(t_{ij} - \phi_{2,i})]^2 + \sum_{t_{ij} > \phi_{2,i}} [Y_{ij} - \phi_{1,i} - \phi_{4,i}(t_{ij} - \phi_{2,i}
\end{aligned}$$

Let $\mathbf{Y}$ be the vector of the N sets of individual observations. The full likelihood is obtained by multiplying the individual likelihoods over the whole subgroup:

$$\mathcal{L}(\mathbf{Y}|\phi_{2,i},\phi_{1,i},\phi_{3,i},\phi_{4,i},\theta_1,\theta_2) = \prod_{i=1}^{N} \mathcal{L}(Y_i|\phi_{2,i},\phi_{1,i},\phi_{3,i},\phi_{4,i},\theta_1,\theta_2)$$

4.2 Prior distributions

I assumed that the four key parameters were a priori independent. I selected a continuous uniform prior distribution for the changepoint, with ranges of five and ten years respectively for the subgroup treated subsequently with hormones, and for the other group. The range was selected according to biological background; secondary treatment is usually initiated when it is suspected that radiotherapy failed, which occurs within a couple of years. Eventually, PSA levels increase even if the treatment is a success, for that reason, I selected a ten year range.

Normal hierarchical models were selected for the nadir, and the two slopes. I initially visually checked histograms of slopes and nadir estimates; I plotted the lowest log_2PSA measurement for each man, as well as the observed slopes as discussed in section 3.3. The histograms looked approximately normal, although the slopes were a little skewed towards null values. I used the normal and uniform distributions for the means and variances of these three parameters. I selected normal distributions for θ_1 and θ_2. In summary the prior distributions were :

$$\begin{aligned}
\phi_{2,i} &\sim U[a,b], \\
\phi_{k,i}|\mu_k,\sigma_k^2 &\sim N(\mu_k,\sigma_k^2) && k=1,3,4, \\
\mu_k &\sim N(a_k,b_k^2) && k=1,3,4, \\
\sigma_k &\sim U(c_k,d_k) && k=1,3,4, \\
\theta_l &\sim N(v_l,w_l^2) && l=1,2.
\end{aligned}$$

4.3 Practical implementation

4.3.1 Posterior simulation

The Bayesian model was implemented in WinBUGS, the interactive Windows version of BUGS (Bayesian Analysis Using Gibbs Sampling, version 1.4) [55]. The program performs Bayesian analysis of complex statistical models using MCMC techniques, as described in details in section 2.2.2.

4.3.2 Convergence

I ran two independent analyses, and assessed convergence separately for each model using the Raftery and Lewis criterion [54]. For each analysis, I ran three chains simultaneously, with overdispersed starting values. After an initial burn-in period of 2,000 iterations, I applied the diagnostic test to the next 10,000 iterations, separately for each chain. Under the default set up (2.5% quantile with precision of $\pm$ 0.005 at 95%) the required number of iterations was between 10,000 and 20,000 for the group parameters (μ_1,σ_1^2, μ_2, σ_2^2, μ_3, σ_3^2), as well as the variance parameters (θ_1 and θ_2), but some of them required about 30,000 iterations. The default set up is highly conservative, and I thus modified the precision to 0.01. With this new set up, 10,000 iterations or less were required for the group parameters, as well as the majority of the individual parameters. Using the Raftery and Lewis outputs, the Gelman-Rubin statistic, the history plots, and my observation of the distributions of the group and individual parameters, I concluded that convergence was reached. In each subgroup, I pooled the three sets of 10,000 iterations, and estimated the posterior distributions of the parameters of interest using the 30,000 replicates.

4.4 BUGS codes

4.4.1 Hierarchical model

Below is the code for the modeling of the subgroup of men with a secondary treatment. The model was similar for the second subgroup, except for the prior distribution of the changepoint that was uniformly distributed over five and ten years for the subgroups of men with and without secondary treatment respectively. I first provide a list of the names used in the coding. See equations 3.1 and 3.2, for a complete description of the model.

Table 4–1: Codes for the variables used in the WinBugs code

Name in the code	Definition
i log2nadir[i]	log_2PSA nadir for the i^{th} man (ϕ_{1i})
change[i]	Changepoint for the i^{th} man (ϕ_{2i})
before[i]	log_2PSA decline rate before the nadir for the i^{th} man (ϕ_{3i})
after[i]	log_2PSA growth rate after the nadir for the i^{th} man (ϕ_{4i})
log2nadir.mu	Population log_2PSA nadir(μ_{ϕ_1})
change.mu	Population changepoint (μ_{ϕ_2})
before.mu	Population log_2PSA decline rate before the nadir (μ_{ϕ_3})
after.mu	Population log_2PSA growth rate after the nadir (μ_{ϕ_4})
log2nadir.var	Variance of the individual log_2PSA nadirs ($\sigma^2_{\phi_1}$)
var.change	Variance of the individual changepoints ($\sigma^2_{\phi_2}$)
before.var	Variance of the individual log_2PSA decline rates before the nadir ($\sigma^2_{\phi_3}$)
after.var	Variance of the individual log_2PSA growth rates after the nadir ($\sigma^2_{\phi_4}$)
nadir[i]	PSA nadir on the natural scale for the i^{th} man
mean.nadir	Population PSA nadir on the natural scale
var.nadir	Variance of the individual PSA nadirs on the natural scale
mean.dt	Population PSA doubling time
mean.hl	Population PSA half life
theta	Variance parameter (θ_1)
phi	Variance parameter (θ_2)

```
MODEL { for(i in 1 : N) {
log2psa[i] ~ dnorm(mu[i],prec[i] )
mu[i]<- log2nadir[id[i]] + beta[i]*newtime[i]
newtime[i]<-time[i]-change[id[i]]
indic[i]<-step(newtime[i])
beta[i]<-before[id[i]]*(1-indic[i])+after[id[i]]*indic[i]
log(prec[i]) <- theta + phi * log2psa[i] }

# PRIORS FOR FIXED EFFECTS:
theta ~ dnorm(0.0, 0.01)
phi ~dnorm(0.0, 0.01)
# PRIORS FOR RANDOM EFFECTS:
for (i in 1 : M) {
change[i] ~ dunif(0,5)
log2nadir[i] ~ dnorm(log2nadir.mu,log2nadir.prec)
before[i] ~ dnorm(before.mu,before.prec)
after[i] ~ dnorm(after.mu,after.prec)
nadir[i]<-pow(2,log2nadir[i])}

# HYPER PRIORS
log2nadir.mu ~ dnorm(0,0.01)
before.mu ~ dnorm(0.0,0.01)
after.mu ~ dnorm(0.0,0.01)
before.prec <- 1/before.var
before.var <- pow(before.sd,2)
before.sd~ dunif(0,4)
after.prec<-1/after.var
```

```
after.var <- pow(after.sd,2)
after.sd~dunif(0,4)
log2nadir.prec <- 1/log2nadir.var
log2nadir.var <- pow(log2nadir.sd,2)
log2nadir.sd ~ dunif(0,4)
sd.change<-sd(change[])
var.change<-pow(sd(change[]),2)

mean.nadir<-mean(nadir[])
var.nadir<-pow(sd(nadir[]),2)
sd.nadir<-sd(nadir[])
mean.after<-mean(after[])
mean.dt<-1/mean.after
mean.before<-mean(before[])
mean.hl<-1/mean.before }
```

4.4.2 Multiple linear regression models

Once the hierarchical model was fitted, the four estimated parameters were saved for each series. They were then used in turn as dependent variables in four separate multivariate analysis. The explanatory variables were the age at diagnosis, the initial log_2PSA level, and the initial Gleason score.

```
model { for( i in 1 : N ) {
before[i] ~ norm(mu.before[i],tau.before)
mu.before[i]<-b0.before+b2.before.age*age1[i]
    +b3.before.glea*glea1[i] +b4.before.log2psa*psa01[i]
after[i] ~ dnorm(mu.after[i],tau.after)
```

```
mu.after[i]<-b0.after+b2.after.age*age2[i]+ b3.after.glea*glea2[i]
    +b4.after.log2psa*psa02[i] log2nadir[i] ~
dnorm(mu.log2nadir[i],tau.log2nadir)
mu.log2nadir[i]<-b0.log2nadir+b2.log2nadir.age*age3[i]
    +b3.log2nadir.glea*glea3[i] +b4.log2nadir.log2psa*psa03[i]
change[i] ~ dnorm(mu.change[i],tau.change)
mu.change[i]<-b0.change+b2.change.age*age4[i]
    +b3.change.glea*glea4[i] +b4.change.log2psa*psa04[i]  }

#PRIORS
b0.before~ dnorm(0.0, 0.0001)
b2.before.age~ dnorm(0.0, 0.0001)
b3.before.glea~ dnorm(0.0, 0.0001)
b4.before.log2psa~ dnorm(0.0, 0.0001)
tau.before<-1/var.before
var.before<-pow(sd.before,2)
sd.before~dunif(0,5)
b0.after~ dnorm(0.0, 0.0001)
b2.after.age~ dnorm(0.0, 0.0001)
b3.after.glea~ dnorm(0.0, 0.0001)
b4.after.log2psa~ dnorm(0.0,0.0001)
tau.after<-1/var.after
var.after<-pow(sd.after,2)
sd.after~dunif(0,5)
b0.log2nadir~ dnorm(0.0, 0.0001)
b2.log2nadir.age~ dnorm(0.0, 0.0001)
b3.log2nadir.glea~ dnorm(0.0, 0.0001)
```

```
b4.log2nadir.log2psa~ dnorm(0.0, 0.0001)
tau.log2nadir<-1/var.log2nadir
var.log2nadir<-pow(sd.log2nadir,2)
sd.log2nadir~dunif(0,5)
b0.change~ dnorm(0.0, 0.0001)
b2.change.age~ dnorm(0.0, 0.0001)
b3.change.glea~ dnorm(0.0, 0.0001)
b4.change.log2psa~ dnorm(0.0, 0.0001)
tau.change<-1/var.change
var.change<-pow(sd.change,2)
sd.change~dunif(0,5) }
```

4.5 Inference using the outputs from the MCMC algorithms

Three chains were used in the MCMC estimation process. Once convergence was reached, I pooled the three chains; the overall estimates were shown in table 3-2, page 65. Below are the estimates for the parameters of the hierarchical model, as provided by each chain, before the pooling, and after convergence was reached.

4.5.1 Summary statistics for the subgroup who received secondary treatment

CHAIN 1	Mean	SD	0.025	0.975
after.mu	1.6586116	0.14764291	1.3759750	1.9560000
after.sd	1.5463882	0.13005325	1.2990000	1.8160000
before.mu	-3.5115960	0.19880467	-3.9150250	-3.1320000
before.sd	1.6610438	0.16875512	1.3469750	2.0170000
log2nadir.mu	0.7561532	0.13695806	0.4864975	1.0270000
log2nadir.sd	1.4894563	0.10421223	1.3020000	1.7100000
theta1	0.7164148	0.08368749	0.5458975	0.8754025
theta2	0.3141198	0.02454568	0.2670975	0.3624025
dt	0.6034155	0.02232287	0.5604000	0.6478025
hl	-0.2852755	0.01143283	-0.3086000	-0.2633000
mean.change	1.0812962	0.04251550	1.0010000	1.1700000
mean.nadir	3.0130848	0.15197405	2.7500000	3.3660000
var.change	0.5586769	0.09336875	0.3930975	0.7536125
var.nadir	27.690782	14.72249995	16.7100000	59.0110000

CHAIN 2	Mean	SD	0.025	0.975
after.mu	1.6556686	0.14394975	1.3779750	1.9380000
after.sd	1.5470824	0.13317460	1.3030000	1.8240000
before.mu	-3.4958449	0.19489616	-3.8760000	-3.1159750
before.sd	1.6500339	0.16502979	1.3500000	1.9910250
log2nadir.mu	0.7526957	0.14013010	0.4798900	1.0260000
log2nadir.sd	1.4967171	0.10693893	1.3070000	1.7310000
theta1	0.7172615	0.08059577	0.5546000	0.8724025
theta2	0.3120555	0.02331956	0.2673975	0.3583000
dt	0.6046310	0.02237225	0.56320	0.6512025

hl	-0.2862984	0.01123988	-0.30960	-0.2659000
mean.change	1.0845825	0.04248087	1.00500	1.1680000
mean.nadir	3.0224475	0.16421914	2.76600	3.4060000
var.change	0.5628162	0.09643483	0.40120	0.7641050
var.nadir	28.403154	15.27464061	16.93975	65.1500000

CHAIN 3	Mean	SD	0.025	0.975
after.mu	1.6603680	0.14492363	1.3820000	1.9530
after.sd	1.5479436	0.13328310	1.3050000	1.8280
before.mu	-3.4869220	0.19413482	-3.8780000	-3.1170
before.sd	1.6467105	0.16066348	1.3570000	1.9870
log2nadir.mu	0.7490921	0.13719704	0.4802000	1.0180
log2nadir.sd	1.4896081	0.10519854	1.2960000	1.7110
theta1	0.7170486	0.07818849	0.5608950	0.8685
theta2	0.3119343	0.02289185	0.2667975	0.3573
dt	0.6028889	0.02196408	0.5602000	0.6461
hl	-0.2872009	0.01127074	-0.3103000	-0.2868
mean.change	1.0895908	0.04114699	1.0100000	1.0890
mean.nadir	2.9996391	0.17763244	2.7370000	2.9820
var.change	0.5655563	0.08504158	0.4106875	0.5631
var.nadir	29.270223	28.03964340	16.6197500	24.7300

4.5.2 Summary statistics for the subgroup who did not receive secondary treatment

CHAIN 1	Mean	SD	0.025	0.975
after.mu	0.2813794	0.02546580	0.2314975	0.33080250
after.sd	0.3743906	0.02256343	0.3325000	0.42080250
before.mu	-3.2908883	0.15535209	-3.6060000	-2.99797500
before.sd	2.0021615	0.12531991	1.7600000	2.24600000
log2nadir.mu	-0.2567505	0.08231129	-0.4202050	-0.09335875
log2nadir.sd	1.2927791	0.06465787	1.1720000	1.42700000
theta1	0.6846694	0.03312022	0.6197000	0.75030250
theta2	0.2679680	0.01234011	0.2431000	0.29170000
dt	3.5701327	0.19333029	3.222000	3.9860000
hl	-0.3042417	0.01015532	-0.324100	-0.2848975
mean.change	1.4118572	0.05950837	1.295000	1.5280000
mean.nadir	1.2629858	0.04749968	1.176975	1.3630000
var.change	1.9244957	0.30296348	1.376975	2.5590000
var.nadir	2.6222261	0.71744754	1.596975	4.3270750

CHAIN 2	Mean	SD	0.025	0.975
after.mu	0.2803973	0.02561997	0.2301950	0.3304050
after.sd	0.3749433	0.02227293	0.3333975	0.4210000
before.mu	-3.2834199	0.15554784	-3.5970000	-2.9899750
before.sd	2.0096231	0.13137688	1.7670000	2.2770250
log2nadir.mu	-0.2595304	0.08245708	-0.4264025	-0.1012975
log2nadir.sd	1.2919511	0.0006660908	0.0018732	1.1670000
theta1	0.6847597	0.03270984	0.6195975	0.7475000
theta2	0.2674351	0.01285369	0.2422000	0.2923000
dt	3.5755157	0.19529808	3.219000	3.984025

hl	-0.3046514	0.01017459	-0.325300	-0.285300
mean.change	1.4111155	0.05950166	1.297000	1.527000
mean.nadir	1.2549980	0.04519230	1.171000	1.347000
var.change	1.9044374	0.29093421	1.371975	2.510000
var.nadir	2.4673449	0.60648859	1.568975	3.919075

CHAIN 3	Mean	SD	0.025	0.975
after.mu	0.2821356	0.02533348	0.23210	0.3309000
after.sd	0.3720399	0.02237768	0.33010	0.4187000
before.mu	-3.2815673	0.15163218	-3.58600	-2.9960000
before.sd	1.9932755	0.12974811	1.74500	2.2550250
log2nadir.mu	-0.2651456	0.08328143	-0.42911	-0.0992755
log2nadir.sd	1.2850562	0.06635509	1.16100	1.4200000
theta1	0.6806468	0.03298656	0.61600	0.7455025
theta2	0.2673895	0.01264005	0.24210	0.2921025
dt	3.5581844	0.19151490	3.211000	3.955000
hl	-0.3050761	0.01001095	-0.325100	-0.285700
mean.change	1.4138736	0.06533431	1.281000	1.544000
mean.nadir	1.2464734	0.04709868	1.161000	1.347000
var.change	1.9434053	0.31893973	1.358975	2.603025
var.nadir	2.4621183	0.66070100	1.544950	4.063025

Preamble to Manuscript II

In this second manuscript, I address the main focus of this research project: the evaluation of the ASTRO criterion for biochemical failure.

Biochemical failure is defined as a proxy for the recurrence of the cancer, as indicated by rising PSA levels. The ASTRO definition considers three consecutive rises in PSA levels as an appropriate definition of biochemical failure following radiation therapy [37]. This rule is widely used but only rough empirical evidence is available as to its short-term classification performance in truly identifying PSA rises, and in reassuring those whose PSA is truly not rising, or rising so slowly as to not cause trouble within the man's life expectancy.

In the next article, I present the first formal evaluation of the ASTRO criterion. I evaluate the sensitivity and specificity of the criterion, as well as an alternative rule, as a function of the underlying true PSA trajectory, the number, and timing of the PSA measurements, and the amount of measurement and biological fluctuations. I generate a large number of data series: the features of the underlying trajectories are taken as known, having been estimated using the hierarchical model, applied to the 470 data series in the first manuscript. Each underlying PSA value is then distorted by adding the amounts of 'noise' seen in the actual data. Rules for biochemical failure are applied to each simulated data series, and their sensitivity and specificity described as a function of the schedule and the true doubling time.

This 'rule of three', to detect an up (down) turn is also informally used in other fields, again albeit without a formal evaluation. The approach presented here could

be used to evaluate its performance in these other settings, such as trends in economic data.

This article has been submitted to the journal *Statistics in Medicine*, and follows the submission guidelines of this journal. The references are included in the global thesis bibliography.

CHAPTER 5

Manuscript II - Detecting trends in noisy data series: application to the detection of PSA failure, defined as three consecutive PSA rises in men treated for prostate cancer.

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

[1] Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Qc.

[2] Division of Urology, University of Connecticut Health Center, Farmington, CT.

Abstract

When studying longitudinal data, it is common to define an event based on a sustained rise (or decline) of the observations. Often, these data are subject to variability independent of the event of interest. Appropriate statistical methods are necessary to account for these extraneous variations, in order to correctly determine whether a specific outcome occurred. We propose a method to evaluate rules that define events based on consecutive increases (or decreases) of observations, given the presence of extra-variability. We illustrate our approach using post-radiotherapy series of prostate-specific antigen (PSA).

PSA is used as a monitoring tool for prostate cancer recurrence following radiotherapy. If radiotherapy is successful, PSA levels decrease substantially to a nadir value, and remain at low levels. A permanent subsequent steeper increase indicates treatment failure. The ASTRO criterion defines biochemical failure as three consecutive PSA increases. Although this 'rule of three' is widely used, it has been criticized for its low levels of sensitivity and specificity.

We carried out a numerical validation study of the ASTRO criterion, and examined its short-term performance in correctly identifying a PSA trajectory that is truly rising (sensitivity), and how often it can recognize a series that is truly stable for what it is (specificity). Our method relies on the simulation of realistic, sophisticated data sets, that accurately reflect the systematic and random variations that can be observed in such series. To do so, we first fitted an appropriate model to real PSA series, and then simulated new data from this model. These 'empirically based simulations' were particularly flexible, and allowed us to evaluate any rule that relied on the marker concentration, as well as any schedule of measurements. The approach can also be applied to evaluate other rules that purport to rapidly and accurately detect up (down) turns in noisy series, such as in other medical data, and in data series used to monitor economic trends.

5.1 Introduction

Prostate-specific antigen (PSA) is a glycoprotein naturally produced by the prostate and by cancer cells. The observed PSA concentration is an amalgam of the unobservable true PSA concentration and random variation (measurement errors and short-term biological variations unrelated to tumor size). In men with prostate cancer, the cancer's contribution to the PSA concentration is approximately proportional to the tumor volume, and thus the PSA level help detect residual and early recurrence of tumors when monitoring response to radiotherapy [22][23]. If radiotherapy is successful, the PSA level decreases substantially to a nadir value during the first two years, and remains at low levels with possibly a very slow increase. A subsequent ongoing steeper increase would indicate treatment failure, and depending on the man's wishes, his physician may start hormone-withdrawal therapy. Treatment success does not necessarily imply a flat PSA curve or equivalently an infinite PSA doubling time. Indeed, PSA are not only produced by the tumor, but also by the remaining healthy prostate cells; thus, the PSA concentrations will rise eventually, even if the patient is cured.

Biochemical failure is defined as a recurrence of the cancer, detected by rising PSA levels. The American Society for Therapeutic Radiology and Oncology (ASTRO) consensus panel considers *three consecutive PSA rises* as an appropriate definition of biochemical failure following radiation therapy, and the date of failure should be the midpoint between the post-irradiation nadir PSA, and the first of the three consecutive rises [37]. We will refer to this criterion as the ASTRO rule. Several studies have shown that this criterion is a robust measure that correlates well with various clinical endpoints [38][39][74]; however, criticisms of the ASTRO definition have emerged since it was proposed in 1996. First, the ASTRO rule has undergone limited formal evaluation. Studies investigating the performance of the ASTRO

criterion focussed on its *clinical validation*, that is, on its ability to predict distant clinical outcomes. The sensitivity was estimated by assessing whether observed PSA trajectories satisfied the ASTRO rule, and results were reported according to various clinical outcomes, such as evidence of distant metastases or death from prostate cancer. Similarly, the specificity was estimated as the proportion of men with PSA series not satisfying the criterion, given that they did not undergo a recurrence of the cancer (defined as no metastases, or still alive at a certain point in time). Given the different endpoints, inclusion criteria, and schedules of measurements, these studies are difficult to compare to each other. In addition, some studies also suggested that alternative definitions of biochemical failures incorporating the PSA nadir and/or the post-nadir profile, may outperform the ASTRO definition [41][42][43][44].

Second, the ASTRO criterion was examined without accounting for the PSA variability. The observed PSA values were taken at their face values, and thus analyzed as if they represented the true PSA concentrations.

Third, the ASTRO definition considers the date of PSA failure as the midpoint between the last non-rising and first rising PSA. This process, referred to as back-dating, poses problem with recently rising PSA values, and makes the value difficult to interpret.

In 2003, McMullen et al. reported that only two thirds of the peer-reviewed English published articles in 1999-2000 used the ASTRO definition [3]. *Recognizing the controversy surrounding the use of the ASTRO definition*, the authors were not surprised to find that a significant minority (35%) of investigators chose not to use the ASTRO definition or used some modification of it. Given the discordant findings of the published clinical validation studies, as well as the important proportion of studies relying on the ASTRO criterion, an investigation of the more short-term performance of this criterion appears necessary. Indeed, a fundamental point, before one considers how well even a perfectly measured PSA trajectory correlates with

clinical outcomes, is how good the ASTRO rule is at correctly (and quickly) identifying 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. This *numerical validation* involves comparing observed PSA series with the underlying true PSA trajectories. It is surprising that this first-stage issue has not been evaluated, given that statistical methods have successfully described longitudinal changes in PSA to predict either the onset of prostate cancer [45][46][47], or the recurrence of the disease following treatment [48][49].

We carried out a numerical evaluation of the ASTRO rule; we assessed both its short-term performance in truly identifying PSA rises (sensitivity), and in reassuring those whose PSA is truly not rising, or rising so slowly as to not cause trouble within the man's life expectancy (specificity). Unlike the clinical validation where a clinical outcome was used as the final endpoint, we used the underlying true PSA curve as the gold-standard. More specifically, we evaluated the sensitivity as the proportion of *observed* PSA series satisfying the ASTRO rule, given that the underlying post-nadir *true* curve is indeed rising, and thus as a function of the underlying PSA doubling time. Similarly, the specificity was the probability of not satisfying the criterion, given that the underlying PSA trend was flat, or rising very slowly. We evaluated this rule using a set of *simulated* post-radiotherapy PSA trajectories. In order for our simulated series to have the most-likely shapes of typical post-radiotherapy PSA curves, we based our simulation process on a cohort of 470 men treated for localized prostate cancer with radiotherapy only. This rich data set allowed us to estimate underlying realistic error-free trajectories, and the variability of the PSA measurements. We then simulated realistic post-radiotherapy PSA curves using the estimates provided by the Markov Chain Monte Carlo outputs. The flexibility of the simulated

data set allowed us to examine the ASTRO rule for different lengths of follow-up and frequencies of measurements.

The use of consecutive rises (declines) of a specific marker is not specific to PSA data. Because of the flexibility of the statistical model and the simulation process, our approach can be applied to other rules that purport to accurately detect up (down) turns in other noisy series.

In addition, we evaluated the Houston rule, which has been suggested to outperform the ASTRO criterion [44][41]. This criterion defines biochemical failure as any increase of 2 ng/ml above the PSA nadir (defined as the lowest PSA measurement of the follow-up).

Our paper proceeds as follows. In section 5.2, we present the Connecticut cohort of 470 men treated for prostate cancer with radiotherapy. We then fitted a hierarchical model to this data set, and obtained estimates for each individual PSA error-free PSA profiles, as well as for the variability of the PSA measurements. In section 5.3, we describe the simulation process, and in section 5.4, we detail the estimation of the sensitivity and specificity. In section 5.5, we present results for the ASTRO and Houston criteria. We discuss our findings in section 5.6.

5.2 Preliminary analysis: Estimation of real PSA trajectories

We describe the data set, and the Bayesian hierarchical changepoint model that was fitted to these PSA series.

5.2.1 The Connecticut data

The data were initially collected for a previous study aimed at linking rising PSA trajectories following treatment of prostate cancer (surgery or radiotherapy) to ten-year outcomes [25]. The data were assembled retrospectively, on a population based cohort identified by the Connecticut Tumor Registry. The men were aged 75 years

or less and residents of Connecticut when diagnosed with localized cancer between 1990 and 1992. More details are available in Albertsen et al. [25]. We base our study on the 647 men diagnosed with a localized cancer of the prostate and treated with radiotherapy without any hormonal pre-treatment. Men with advanced disease or an initial PSA higher than 50 ng/ml were excluded. In addition, we required each PSA series to have at least a baseline PSA measurement as well as two subsequent PSA measurements; thus 470 men 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.

5.2.2 The Bayesian hierarchical changepoint model

Figure 5–1 provides post-radiotherapy log_2PSA series for four men of the Connecticut data set, with the time axis starting at the initiation of the treatment (page 98). Notice the typical V-shape of these series: following the start of the treatment, the PSA levels drop to some nadir level, and then increase again at various rates. The logPSA decline and growth rates vary across men but are reasonably constant within-men, suggesting exponential patterns before and after the nadir [23][24]. As a result, the log_2PSA series follow approximately a piecewise linear pattern, as summarized by the prototypic PSA trajectory provided in figure 5–2 (page 99). Note that the log_2 scale allows one to obtain directly the PSA doubling time, which is simply the reciprocal of the post-nadir log_2PSA growth rate.

We used a Bayesian hierarchical model with a random changepoint to estimate the log_2PSA trajectories, and their variability. Let log_2PSA_{ij} be the PSA concentration on the log_2 scale, for the j^{th} measurement for the i^{th} man. We assumed that the observations log_2PSA_{ij} were normally distributed, with mean μ_{ij} and variance σ^2_{ij}:

$$log_2PSA_{ij} \sim N(\mu_{ij}, \sigma^2_{ij}).$$

For every observation log_2PSA_{ij}, the expected log_2PSA value, μ_{ij}, was related to the timing of the measurement t_{ij} through a regression function describing the expected profile:

$$\mu_{ij} = \begin{cases} \phi_{1i} + \phi_{3i}(t_{ij} - \phi_{2i}), & t_{ij} < \phi_{2i}, \\ \phi_{1i} + \phi_{4i}(t_{ij} - \phi_{2i}), & t_{ij} \geq \phi_{2i}, \end{cases} \tag{5.1}$$

where ϕ_{1i}, ϕ_{2i}, ϕ_{3i} and ϕ_{4i} correspond respectively to the log_2 PSA nadir, the change-point (i.e., the timing of the nadir), the log_2PSA decline rate prior to the PSA nadir, and the post-nadir log_2PSA growth rate for man i. The reported interassay coefficients of variation tend to be larger at lower PSA levels [2][21], therefore, we expressed the precision as a linear function of the PSA levels, i.e. $\log \frac{1}{\sigma_{ij}^2} = \theta_1 + \theta_2 \mu_{ij}$, and thus:

$$\sigma_{ij}^2 = \exp[-(\theta_1 + \theta_2 \mu_{ij})], \tag{5.2}$$

where μ_{ij} is given by equation 5.1.

We split the data into two subgroups. One subgroup consisted of 139 men who subsequently received a secondary treatment; the second subgroup included the remaining 331 men. Note that we are not in the context of a clinical trial, and the assignment of a secondary treatment is based on the decision of the physicians, as well as the patients. This somewhat arbitrary division was performed for two reasons related to the estimation process of the hierarchical parameters. First, men who undergo secondary treatment are those for whom radiotherapy appears to have failed. These men tend to reach a PSA nadir much sooner, with a steeper post-nadir PSA growth rate. By splitting the data, we therefore obtained two subgroups with similar patterns that would satisfy the distributional assumptions of the hierarchical model. Second, homogeneous groups allow one to speed up the computational phase of the estimation process. Two analysis were therefore performed independently. The two

hierarchical models were implemented in WinBUGS, a statistical software that uses MCMC techniques to generate the distributions of the parameters of interest [55].

5.3 Generation of simulated realistic PSA series

To generate several independent realizations of a man's PSA series, one needs to first estimate the man's true PSA profile, as well as the variability of the PSA measurements. These estimates are obtained from the MCMC outputs provided after fitting our hierarchical model, and following convergence.

In section 5.2.2, we described how we fitted our hierarchical model to the 470 log_2PSA series, and obtained estimates $\widetilde{\theta}_1$ and $\widetilde{\theta}_2$, for the variance parameters θ_1 and θ_2, as defined by equation 5.2. Thus, for a true log_2PSA concentration μ_{ij}, the estimated variance was given by:

$$\widetilde{\sigma}^2_{ij} = \exp[-(\widetilde{\theta}_1 + \widetilde{\theta}_2 \mu_{ij})], \tag{5.3}$$

In addition, each iteration of the MCMC process generated one quartet of estimates $(\widetilde{\phi}_{1i}, \widetilde{\phi}_{2i}, \widetilde{\phi}_{3i}, \widetilde{\phi}_{4i})$ for every man i, and thus the i^{th} man's error-free PSA concentration, $\widetilde{\mu}_{ij}$, at every time point j, as given by the model 5.1. To generate a simulated realistic PSA concentration, we added the corresponding amount of variability. Thus, at each iteration, for every man i, at every time point j, a simulated realistic log_2PSA concentration, $log_2 PSA_{ij}$, was generated by drawing a value from a normal distribution centered at the estimated true concentration:

$$log_2 PSA_{ij} \sim N(\widetilde{\mu}_{ij}, \widetilde{\sigma}^2_{ij}),$$

where $\widetilde{\sigma}^2_{ij}$ is given by equation 5.3. Thus, each iteration of the MCMC process was used to generate a separate realistic simulated log_2PSA curve for each of the 470 men. We used several iterations from each man to obtain a set of profiles representative of each individuals predictive distribution.

In order to obtain these parameters, one hierarchical model was fitted to each of the two subgroups of PSA series. For each model, we ran three sequences of iterates, each beginning with a different set of initial values. We ran an initial burn-in period of 2,000 iterates, and performed 10,000 additional iterations. We then selected a sample of independent simulated realizations. For each man, we kept the last 2,500 randomly generated quartets of each chain. To ensure independence, we retained only sequences generated at every fiftieth iteration; this distance was more conservative than the dependence factor suggested by the Raftery and Lewis method [54]. Therefore, each chain provided a set of 50 approximately independent quartets for each man. As a result, a total 150 independent quartets $(\widetilde{\phi}_{1i}, \widetilde{\phi}_{2i}, \widetilde{\phi}_{3i}, \widetilde{\phi}_{4i})$ were available from each man, where $i = 1, ..., 470$.

5.4 Estimation of sensitivity and specificity

For each man i of the Connecticut data set, the hierarchical model 5.1 provided an estimate of the PSA doubling time, the reciprocal of the post nadir log_2PSA growth rate, ϕ_{4i}. A fast PSA doubling time is usually a sign of radiotherapy failure, therefore we estimated the sensitivity as a function of this variable.

The sensitivity of each criterion was estimated in men with estimated short PSA doubling times (less than ten years). We classified the men according to their estimated doubling time (less than one year, one to two years, two to five years, and five to ten years). In each of this subgroup, we estimated the sensitivity of the ASTRO criterion as the proportion of simulated series with three consecutive PSA increases. Similarly for the Houston rule, we first identified the lowest observation of the generated profile, and the sensitivity was then expressed as the proportion of generated series with at least one PSA measurement 2 ng/ml above (and after) this nadir.

The specificity was estimated from men with an estimated doubling time greater than ten years. In such cases, the post-nadir PSA curves are almost flat, that men

can be clinically considered as cured. We estimated the specificity of each criterion as the proportion of series not satisfying the rule.

Notice that we did not evaluate the ASTRO criterion when the follow-up duration was one year, and when measurements were done every six months. In such cases, only three observations were available, and thus the ASTRO criterion which uses three consecutive PSA increases could not be assessed. Finally, the PSA nadir is reached on average at the second year after radiotherapy, and the PSA curve is therefore decreasing over this period. For this reason, we will ignore the first two years of follow-up.

5.5 Results

We sorted the $470 \times 150 = 70,500$ simulated PSA trajectories according to the man's estimated doubling time (table 5–1, page 99). Note that 41 men had a negative estimated post-nadir growth rate, varying between -0.4 and 0. Thus, the PSA doubling time was assumed to be infinite for the corresponding $41 \times 150 = 6,150$ series. The mean PSA doubling time was finite for 429 men, and shorter than ten years for 377 men. We estimated the sensitivity from these 377 men, and thus using the $377 \times 150 = 56,050$ corresponding series. The specificity was estimated from the other 52 men with a finite doubling time greater than ten years, and thus using the $52 \times 150 = 7,800$ series.

Tables 5–2 and 5–3 report sensitivity and specificity for the ASTRO and Houston criteria, for various follow-up durations and frequencies of measurements (pages 100 and 101). For both rules, the sensitivity was improved for longer follow-up periods, and shorter intervals of follow-up. Conversely, the specificity decreased with longer follow-up, and increased when intervals between measurements were extended.

In the table for the Houston criterion, we show in bold when the Houston rule had a better sensitivity, or specificity than the ASTRO criterion (table 5–3, page 101).

In most situations, the Houston criterion appeared to outperform the ASTRO criterion. Indeed, the specificity was systematically better for the Houston rule, except at three year follow-up, for an interval between measurements of six months. In this case only, the specificity of the ASTRO criterion appeared slightly better (91.4% for the ASTRO criterion, and 88.3% for the Houston criterion). We also estimated that the Houston criterion had a better sensitivity than the ASTRO criterion, however, results differed slightly depending on the PSA doubling time. When the PSA doubling time was shorter than five years, the sensitivity of the Houston criterion was systematically better, whatever the follow-up durations and intervals between measurements. For series with a longer time to PSA failure, the sensitivities provided by both rules were very close, without a systematically better rule. For illustration, we constructed a type of receiver operating (ROC) curve to compare both rules (page 102). We selected a three-month interval measurements, and follow-up durations between three and seven years. As illustrated by the curve in figure 5–3, the Houston rule performed better in these specific settings.

5.6 Discussion

Rules based on marker progression have been used to assess biochemical failure, and by extension treatment failure. Because of the important within-subject variability of the concentration of biomarkers, the evaluation of these rules is complex. We have proposed a method that utilizes the underlying biomarker trajectory as the gold standard, and expresses the sensitivity and specificity of the biochemical failure rules as a function of the marker rate of growth, and the schedule of measurements. Such a short-term numerical validation is essential, before one considers how well a perfectly measured marker series correlates with a more distant clinical outcome, and given that clinical evaluations have suggested that the rule might not perform as well as hoped for, possibly because of censoring due to deaths from other causes.

We applied our method to evaluate the classification performances of biochemical failure rules in men treated with radiotherapy for prostate cancer, and expressed our results as a function of the PSA doubling time. By using an appropriate statistical model, we could generate simulated PSA series representative of typical observed PSA series. The estimation of these PSA series could be further improved by including additional covariates in the hierarchical model, such as the baseline PSA level and tumor stage, factors which have been shown to predict tumor growth.

We focussed on two specific rules, the popular and controversial ASTRO criterion, and the Houston rule, which had been suggested to outperform ASTRO. The PSA nadir is reached on average at the second year after radiotherapy, and the PSA curve is therefore decreasing over this period. This biological process explains the low sensitivities observed for the two rules in this initial period, and similarly their high specificities. When focussing on longer follow-up durations, we observed that although the superiority of the Houston rule was not systematic for the various schedules of measurements investigated, its estimated sensitivity and specificity were higher than the ASTRO criterion in most situations. However, it is not possible to compare our findings to earlier results, since earlier studies have focussed on the clinical validation of the ASTRO rule, and thus used clinical outcomes as gold standards. The hierarchical model and the simulation process were particularly flexible, and it is obvious that our method could easily be extended to the evaluation of other rules, and schedules of measurements.

Finally, the use of consecutive rises, or decreases, for a specific marker is not specific to medical data, and our approach can be applied to evaluate other rules that purport to accurately detect up (down) turns in noisy series, such as in other medical data, and in data series used to monitor economic situations. For example, in economics, expansion (or recession) phases are defined as periods when economic activity tends to trend up (down); these periods are identified when consecutive rises

(or declines) of some specific indicators, such as employment rates are observed [75]. In addition, if the underlying pattern of the series is known to be stable, it is possible to calculate exactly the probability of observing consecutive rises (or decreases) using tests for randomness, as described by Olmstead [76], and Levene and Wolfowitz [77].

5.7 Tables and figures

Figure 5–1: log_2 PSA concentrations over time for four men

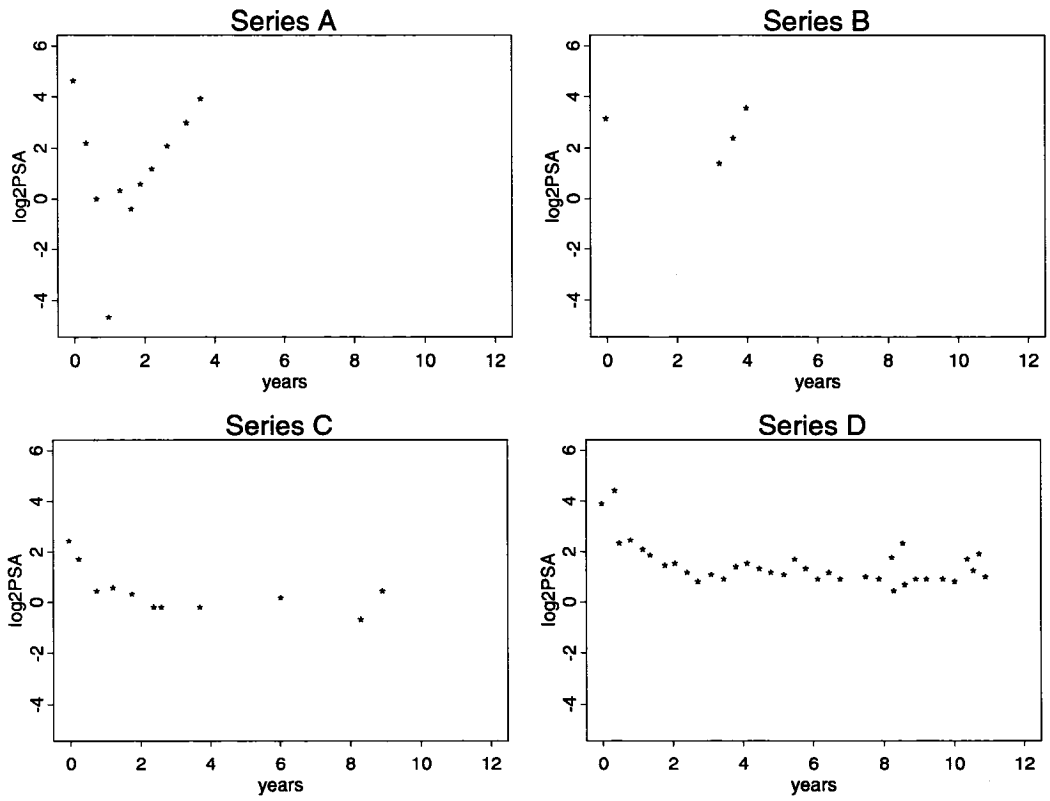


Figure 5–2: Individual piecewise linear model, with the four individual parameters.

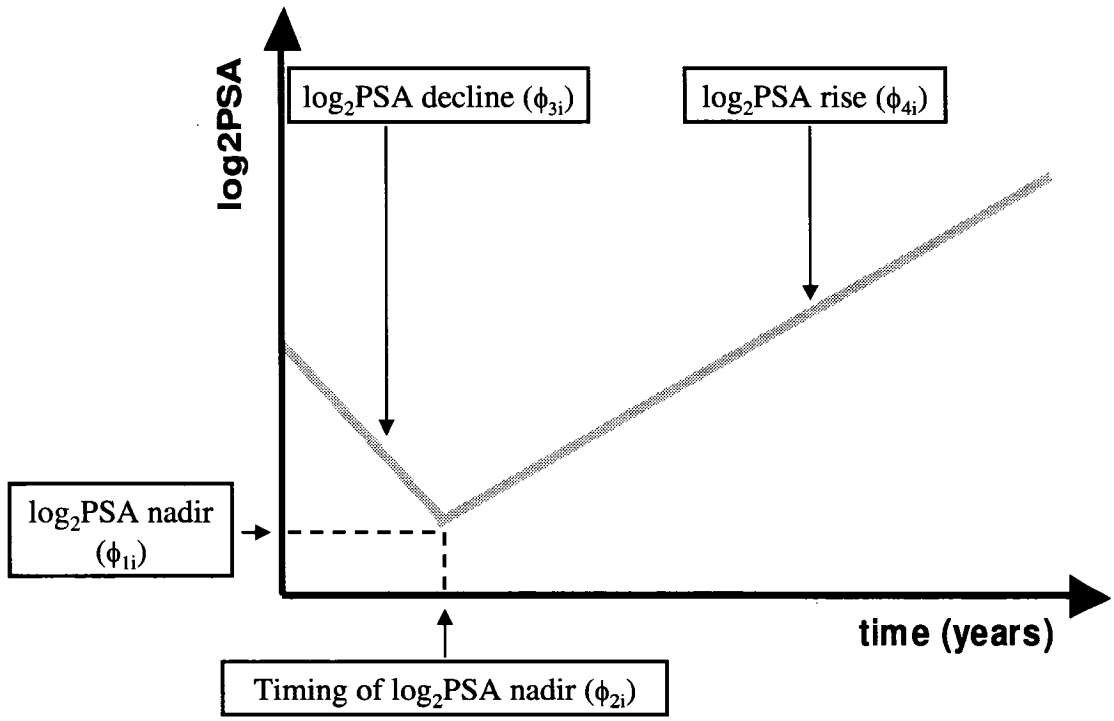


Table 5–1: Distribution of the 470 estimated doubling times and the corresponding log_2PSA growth rates ($\widehat{\phi}_{4i}$), using the hierarchical Bayes model.

Doubling time	($\widehat{\phi}_{4i}$)	count (%)
0 to 1 year	(≥ 1)	92 (19.6%)
1 to 2 years	(0.5 to 1)	86 (18.3%)
2 to 5 years	(0.2 to 0.5)	157 (33.4%)
5 to 10 years	(0.1 to 0.2)	42 (8.9%)
more than 10 years	(0 to 0.1)	52 (11.1%)
-	(≤ 0)	41 (8.7%)

Table 5–2: Sensitivity and specificity of the ASTRO criterion.

Sensitivity, given an interval of three months between PSA measurements.

	Duration of follow-up (years)						
Doubling time	1	2	3	4	5	6	7
0 to 1 year	7.9%	50.8%	80.8%	93.2%	97.5%	98.7%	99.1%
1 to 2 years	1.1%	19.0%	45.2%	66.1%	80.3%	89.6%	94.4%
2 to 5 years	0.8%	11.2%	27.1%	42.9%	56.0%	66.7%	75.3%
5 to 10 years	0.7%	8.0%	20.6%	32.8%	43.7%	52.6%	61.2%

Sensitivity, given an interval of six months between PSA measurements.

	Duration of follow-up (years)						
Doubling time	1	2	3	4	5	6	7
0 to 1 year	-	24.0%	75.2%	90.5%	96.7%	98.2%	98.8%
1 to 2 years	-	6.7%	34.0%	58.9%	76.0%	86.9%	92.9%
2 to 5 years	-	2.2%	14.2%	28.1%	41.4%	53.6%	63.4%
5 to 10 years	-	1.1%	8.0%	15.6%	23.9%	32.1%	39.3%

Specificity, given three and six month interval between PSA measurements.

	Duration of follow-up (years)						
Interval	1	2	3	4	5	6	7
every 3 months	99.1%	90.6%	78.2%	67.1%	57.2%	48.3%	41.2%
every 6 months	-	98.0%	91.4%	84.1%	76.3%	69.3%	63.2%

Table 5–3: Sensitivity and specificity of the Houston criterion.

Sensitivity, given an interval of three months between PSA measurements.

	Duration of follow-up (years)						
Doubling time	1	2	3	4	5	6	7
0 to 1 year	24.7%	66.1%	**87.1%**	**94.0%**	**98.8%**	**99.6%**	**99.6%**
1 to 2 years	11.3%	37.2%	**61.8%**	**78.3%**	**88.1%**	**92.8%**	**96.5%**
2 to 5 years	4.0%	15.4%	**29.6%**	**43.6%**	**56.6%**	**67.4%**	**75.3%**
5 to 10 years	5.0%	13.3%	**21.1%**	28.2%	35.4%	41.5%	47.2%

Sensitivity, given an interval of six months between PSA measurements.

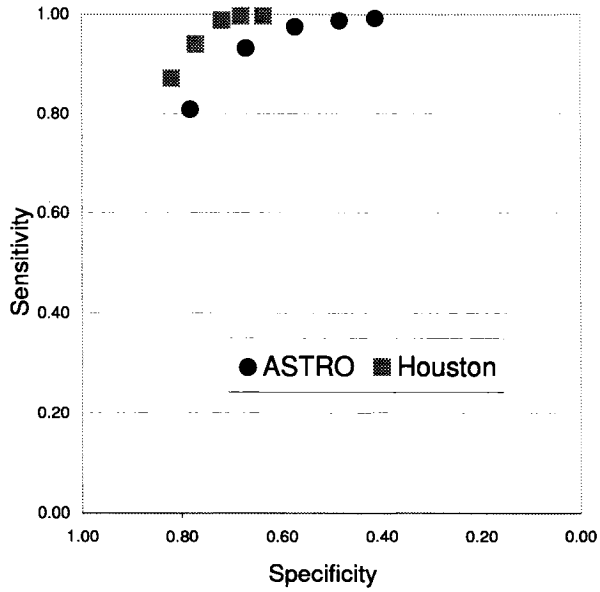
	Duration of follow-up (years)						
Doubling time	1	2	3	4	5	6	7
0 to 1 year	18.9%	61.1%	**85.5%**	**93.3%**	**98.6%**	**99.5%**	**99.6%**
1 to 2 years	5.7%	27.2%	**53.2%**	**73.6%**	**85.9%**	**91.6%**	**95.8%**
2 to 5 years	1.6%	8.5%	**19.2%**	**32.6%**	**46.4%**	**59.3%**	**69.0%**
5 to 10 years	2.2%	6.9%	**12.3%**	**18.2%**	**25.0%**	32.0%	38.3%

Specificity, given three and six month interval between PSA measurements.

	Duration of follow-up (years)						
Interval	1	2	3	4	5	6	7
every 3 months	93.5%	87.2%	**82.0%**	**77.1%**	**71.9%**	**68.1%**	**63.6%**
every 6 months	96.4%	91.9%	88.3%	**84.9%**	**80.5%**	**76.8%**	**72.8%**

The values in bold indicate when the Houston rule has a higher sensitivity or specificity than the ASTRO criterion, for the given setting.

Figure 5–3: ROC curve to illustrate the performances of the ASTRO and Houston criteria



The five black circles represent the performance of the ASTRO rule when the duration of the follow-up varies from three (leftmost circle) to seven years (rightmost circle). The interval between PSA measurements was three months. Similarly, the performance of the Houston rule is shown using the grey squares.

Preamble to Manuscript III

The options for the management of prostate cancer include watchful waiting, or potentially curative treatment (radiotherapy or surgery). The success of these treatments is carefully monitored through a frequent measurement of a patient's PSA concentrations, and more specifically, through the PSA doubling time.

Following treatment, the PSA levels tend to follow a close to exponential pattern, as a result, their logarithms are approximately linear. Furthermore, if the base 2 is selected, the PSA doubling time is simply the reciprocal of the slope of the log_2 transformed data. This property is well known in the scientific community. When reviewing the clinical literature, I found however, that most methods used for the estimation of the PSA doubling time were either not accurate, or accurate, but eventually too time consuming. In general, the authors plot the rough PSA data on their natural scale, and then visually evaluate the doubling time. Although this approach is very simple, it is far from appropriate: first, the post-treatment PSA measurements are highly variable, and second, they do not follow a linear pattern. Given the underlying exponential pattern of post-treatment PSA series, other authors transformed the data using a logarithmic scale, and then estimated the slope of the resulting line. This second approach is more appropriate, but when a physician has to meet several patients per day, he might not find the time to perform a regression analysis to fit a slope for each of them!

Instead of plotting transformed data, and thus first calculating the logarithm of each data point, I suggest simply plotting the actual PSA data on a graph that already accounts for the logarithm scale. Since the resulting plot is approximately linear, it is then much easier to fit a straight line by eye, and thus estimate its

slope, the PSA doubling time. Although this approach sounds straightforward, to my knowledge, no publication has ever mentioned it. Therefore, I wrote this last article for a clinical audience, and specifically for those medical specialists who use PSA doubling times, but cannot afford to spend much time on this calculation. I show how one can easily record serial PSA values over time on a graph sheet in order to quickly and easily obtain accurate estimates of the doubling time, without any difficult computations.

This manuscript has been submitted to the *International Journal of Radiation Oncology*Biology*Physics*, and follows the submission guidelines of this journal. The references are included in the global thesis bibliography.

CHAPTER 6

Manuscript III - A tool for monitoring PSA patterns after treatment for prostate cancer

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

[1] Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Qc.

[2] Division of Urology, University of Connecticut Health Center, Farmington, CT.

Abstract

Purpose Prostate-specific antigen (PSA) is used to monitor the treatment outcome in men with prostate cancer. The PSA doubling time appears to be an important predictor of treatment outcome, and thus essential to monitor over time.

Methods and Materials PSA observations are recorded over time using a logarithm (*log*) graph. The *log* graph avoids the calculation of the logarithm of each PSA observation by changing the scale in the graph.

Results Recording post-treatment PSA series on a *log* sheet provides approximately linear trajectories, with less variation than when using the natural scale (ng/ml). Using these properties, we show how one can record a series of PSA values over time on a graph sheet in order to quickly and easily obtain accurate estimates of the doubling time, without any difficult computations. In addition, given the considerable PSA variability, we emphasize that all available PSA measurements should be used to obtain precise estimates, rather than only the last two observations.

Conclusion Our recording technique particularly applies to post-surgery PSA measurements, but also to post-radiotherapy series following the PSA nadir, and series under watchful waiting. For convenience, we have provided a blank *log* sheet that can be used to record individual PSA series, and to estimate the PSA doubling time.

6.1 Introduction

Prostate-specific antigen (PSA) is produced by both normal prostate tissue and tumor cells in men with prostate cancer; it is used in monitoring of treatment outcome for men with prostate cancer. The clinical usefulness of the post-treatment PSA doubling time was made clear in the early 1990s [23][66]. A rapid PSA doubling time is highly prognostic of clinical failure (defined as incidence of metastatic disease, local recurrence, or prostate cancer death) in men treated with radiotherapy [25][27][28][78], radical prostatectomy [25][28][79][80], or watchful waiting [81][82].

The left-hand column of figure 6–1 represents PSA series for five men treated for prostate cancer for the first time (page 114). The first two men were treated with surgery, the next two men with radiotherapy, and the last man was under watchful waiting. These data are part of a large study, described in more details in Albertsen et al. [25]. In short, the data were assembled on a retrospective population-based cohort of men, diagnosed with localized cancer; all post-treatment PSA values and the one immediately before therapy were recorded. In addition, men who received hormones concurrently are excluded from this analysis. The time axis starts at the first day of the treatment. In the case of surgery, the first PSA measure is taken right before the procedure; it is therefore excluded when computing the PSA doubling time. For radiotherapy, any PSA reading preceding the PSA nadir is excluded. Dots in light gray represent the PSA measurements excluded from the computation of the doubling time.

The plots allow us to make several observations. First, we recognize the typical post-treatment PSA patterns. PSA levels drop to negligible values in the case of surgery and then rise slowly again with an exponential pattern. After radiotherapy, the PSA levels decrease first, and start to rise again. The *log*PSA decline and growth rates vary across men, but are reasonably constant within-men, suggesting

exponential PSA patterns before and after the nadir. In the case of watchful waiting there is variability in PSA levels over time, but a generally increasing pattern.

Second, although we observe underlying exponential trends, patient-specific PSA measurements do not fall exactly on the same curve because of the within-patient variability in the PSA measurements. For example, patient C depicts an unusual jump in PSA concentration at 4.5 years, and this is even more obvious for patient E who had a high PSA level measured at one year and a half. These outlying values may be due to factors unrelated to the disease and the treatment. For example, there might have been errors in the measurement process or during data collection (errors in reporting the date or the PSA level), or a change in the laboratory practice or a different laboratory might have been used. There may also be other (known or unknown) sources of variation such as recent prostate manipulations [14][15], recent sexual activity [16][17] or recent physical exercise [18]. This occasional large variability in PSA concentrations has been frequently reported [2][19][20]; it should not be ignored when analyzing an individual PSA series, and in particular, when estimating its doubling time. Given such noise, it is inaccurate to estimate the PSA doubling time using only the current and immediately previous observations. However, the considerable variability and the non-linearity of the PSA measurements do not allow us to easily visualize (i.e., fit by eye) the underlying profile to entire set of relevant data points, and thus to accurately estimate the doubling time.

Finally, if we apply a logarithmic transformation to an exponential trajectory, such as an individual PSA series, then apart from the random fluctuations, the resulting profile is linear (or piecewise linear, in the case of post-radiotherapy PSA series). Furthermore, if we transform using the logarithm with base 2, then the slope of this resulting line has an immediate interpretation as the reciprocal of the PSA doubling time. Thus, the PSA doubling time is a valid summary of the complete PSA profile (or post-nadir profile in the case of radiotherapy).

Our primary objective is to show how one can use a simple data recording technique to quickly and easily obtain an accurate estimate of the PSA doubling time. To our knowledge, there is no study reviewing how PSA levels are presented to clinicians and by the laboratories reporting them. From an informal survey of a small number of urologists, it appears that there is no standard practice. Laboratory reports contain different amounts of data (the last PSA measurement or the PSA history can be provided) and have different formats (data can be listed or plotted); it is up to the physician to analyze the results and estimate the PSA doubling time. In addition, we also emphasize that simply comparing the current PSA measurement with the one immediately prior to it, will provide a less precise estimate than one based on all available relevant measurements.

6.2 Methods and Materials

The individual PSA profiles in the left-hand column of figure 6–1 are plotted on the natural scale (ng/ml, page 114). Because of the usual variations in the PSA concentrations, the individual data points do not all fall exactly on the underlying approximately exponential curve. Observe now the second column of the same figure. These are the same data, but now plotted on a logarithm (*log*) scale. If one applies a logarithm transformation to exponential data, then the resulting profile is linear. Plotting the data on the supplied graph (see figure 6–2, page 115) is very simple, and is equivalent to the process of applying a *log* transformation to each PSA measurement, a step which can be time consuming. Basically, we replace the need for calculation with a change in scale to the y-axis of the graph, as shown by the y-axis of the graphs. As a result of this new presentation of the data, the PSA profiles have now two interesting properties. First, the data provide approximately linear trajectories (or piecewise linear in the case of radiotherapy). Second, because taking logarithms tends to shrink extreme and overly influential values, the trends

have been 'smoothed', depicting less variation. The outliers observed earlier are of course still present, but appear as less extreme on this scale.

6.3 Results

The linearity and the reduction of random fluctuations allow us to visually fit a straight line to the post-surgery data, and to the post-nadir data in the specific case of radiotherapy. The base 2 selected for the scale leads to another valuable advantage: the PSA doubling time is directly available from the graph sheet. For each patient, we can keep the same recording sheet, and add PSA information as more data become available, i.e. we only need to count how many horizontal (time) units on average are required to jump by one vertical (log_2) unit. In the case of radiotherapy, the doubling time is estimated using the same process, but focusing only on the post-nadir data.

We illustrate our procedure using the data for patient B. These data suggest nine doublings (i.e. nine vertical units) over the 4.5 years of follow-up (i.e. 4.5 horizontal units), therefore the doubling time is $4.5/9 = 0.5$ year, or approximately two doublings per year. Note finally, that if we base our computations only on the last two measurements, the estimated doubling time is about 0.2 years ((1/3)/1.5), much faster than the doubling time obtained from the complete series. This discrepancy is explained by the fact that, if we fit a line by eye through the data, the penultimate observation appears below the fitted line, lower than expected when compared to the overall trend. Therefore, the gap between the last two observations is unusually high, leading to a faster doubling time. Estimates based on longer series will provide more accurate estimates, since the impact of unusually high (or low) PSA measurements will tend to be attenuated as more data are used in the estimation process.

6.4 Discussion

We have shown how one can easily record serial PSA values over time on a graph sheet in order to quickly and easily obtain accurate estimates of the doubling time, without any difficult computations. The *log* scale (base 2 or e) has been used for more than ten years for radiotherapy [34][66], surgery [80], or watchful waiting [23].

Our discussion particularly applies to post-surgery PSA measurements; once the prostate is removed, the only source of PSA production is the tumor cells, which are commonly assumed to follow an exponential growth. Because of their V-shape, post-radiotherapy PSA data are more complex. PSA are produced by both tumor cells and remaining healthy cells; while tumor cells increase exponentially, the PSA production rate in healthy cells is constant and much smaller [83]. Because of the negligible production rate of benign cells as compared to tumor cells, and from our experience with handling data transformation after radiotherapy, the transformation approach still appears valuable, especially well beyond the nadir; the PSA profiles on the *log* scale look more linear, allowing us to estimate the PSA doubling time. Although the exponential pattern of the PSA series under watchful waiting is not so pronounced, plotting data on a *log* sheet allows us to attenuate the extra variation. Finally, estimates of the doubling time are more accurate when more data are used; relying only on the last two measurements should be avoided.

It has been suggested that the post-treatment PSA doubling time might not be constant over time, such as in [80]. However, in this study the PSA doubling time was estimated using only two measurements, without accounting for the PSA variations. Given the considerable variability of the PSA observations, any estimate of the slope or doubling time based on only two measurements is not as reliable. Thus, we feel that these findings do not contradict our results.

PC and Palm-pilot based tools that calculate the PSA doubling time have recently become available to physicians. We propose a tool that does not require any

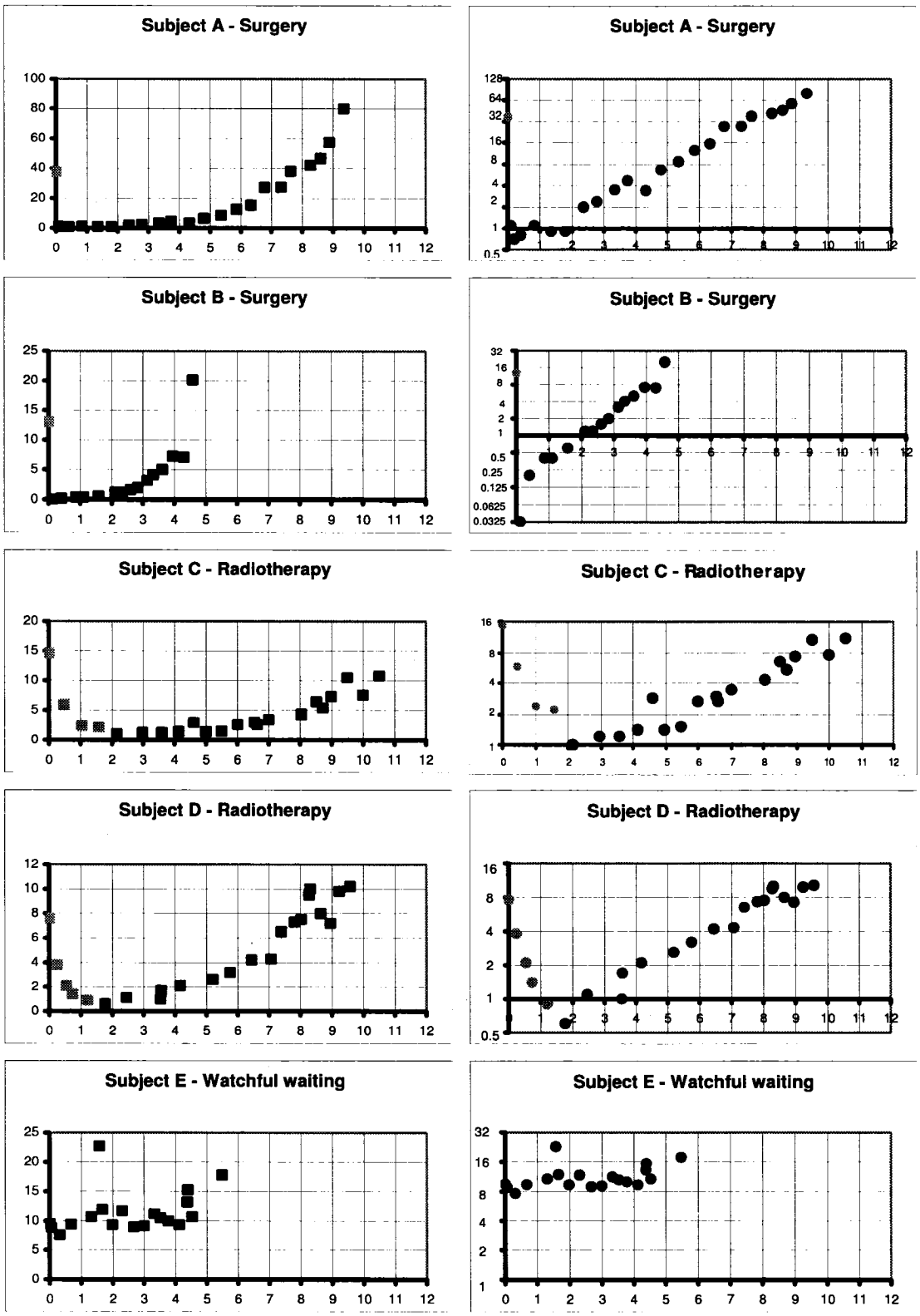
such calculators, and the graph sheet can be kept in the patient file for quick references. For convenience, we have provided a blank *log* sheet that can be used to record individual PSA series (figure 6–2, page 115). Clinicians could keep a copy of this sheet with their patient's chart to estimate the PSA doubling time; PSA measurements can be recorded as new data accumulate, and estimates of the doubling time can be revised.

6.5 Figures

Figure 6–1: PSA history for five men on the natural (ng/ml, left) and log_2 (right) scales. The first two patients were treated by surgery, the next two by radiotherapy, and the last one received no treatment. Dots in light gray represent PSA measurements excluded from the computation of the doubling time. In the case of surgery, the PSA measurement preceding the treatment is excluded; PSA measurements in the decline phase of radiotherapy are also excluded from computation of the doubling time.

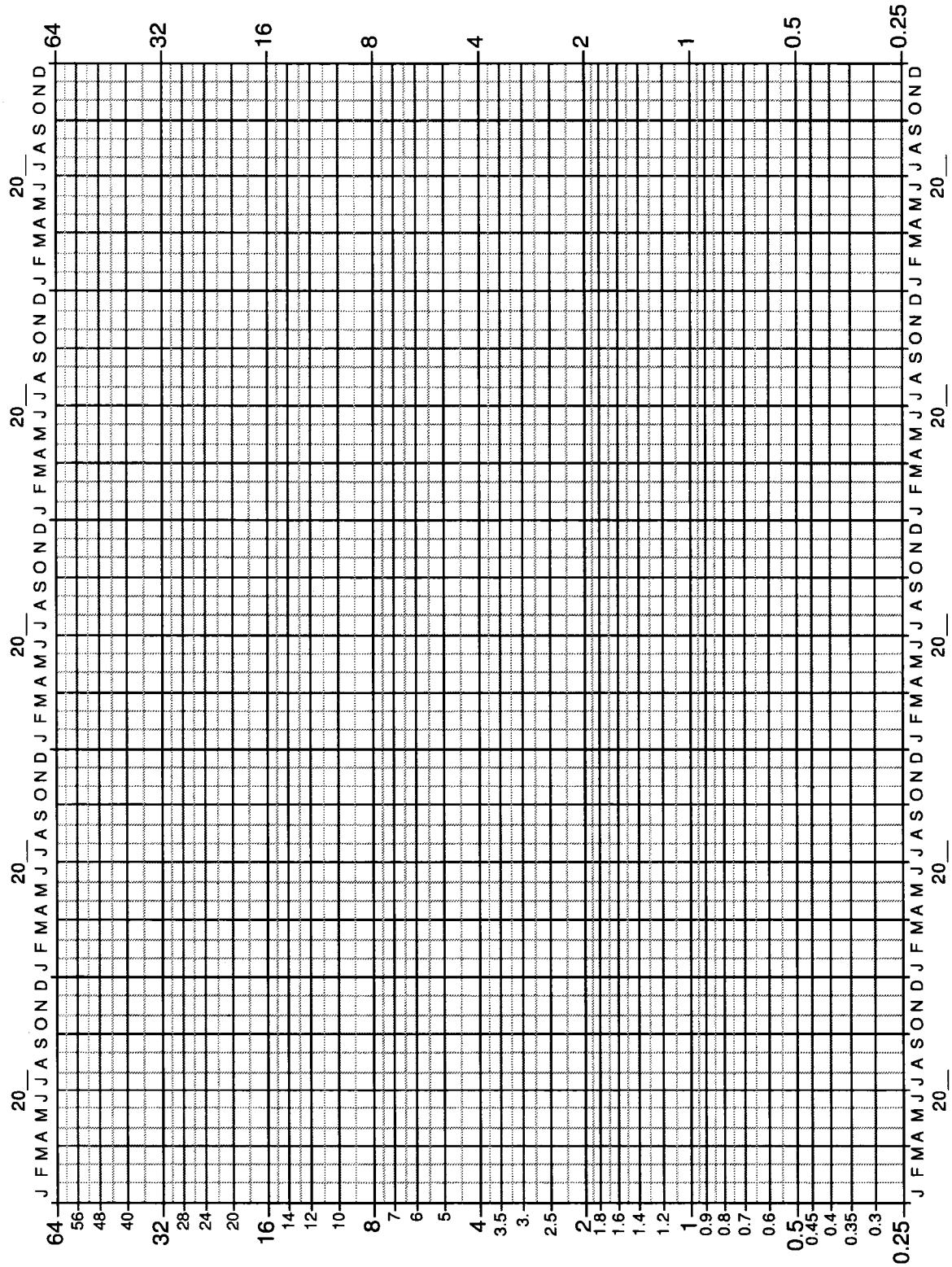
Figure 6–2: Proposed PSA recording sheet for estimating the doubling time. One can use this graph to record PSA levels over time and visually compute the doubling time. One major horizontal unit corresponds to one year with minor units for the months, and one major vertical unit represents one PSA doubling in a year. It is recommended not to join the PSA measurements directly, but rather to fit a line by eye to the observations. The PSA doubling time is the reciprocal of the slope of the fitted line: i.e., the doubling time is the number of horizontal units (years) required to jump by one vertical unit (doubling).

Figure 6–1: PSA history for five men on the natural (ng/ml, left) and log_2 (right) scales



See legend, page 113.

Figure 6–2: Proposed PSA recording sheet for estimating the doubling time



6.6 Appendix

The following discussion applies to post-surgery PSA series, post-nadir PSA series in the case of radiotherapy, and PSA data under watchful waiting. We briefly show that, when the PSA levels are reported using the log_2 scale, then the PSA doubling time is equal to the inverse of the slope of the line fitted to these series.

log_2PSA is the number such that $2^{log_2 PSA} = PSA$. It is a one-to-one increasing function that applies only to numbers strictly positive. Given its mathematical properties, the log_2 function is suitable for PSA values and the transformed data cover a tighter range based on this new *scale*.

Assume that a treatment is initiated and we observe PSA levels in a given man (PSA_1, PSA_2, ..., PSA_n) at several time intervals ($time_1$, $time_2$, ..., $time_n$). Because of the exponential shape of the PSA trajectory, we apply the log_2 transformation and obtain a set of transformed data ($log_2 PSA_1$, $log_2 PSA_2$, ..., $log_2 PSA_n$) that follows a linear pattern. We can now compute the slope of this line. The log_2PSA trajectory as a function of time can be expressed as follows:

$$log_2 PSA(time_i) = \alpha + \beta * time_i, \tag{6.1}$$

where α and β are the intercept and slope of the line (constant over time). To estimate the doubling time PSA_{DT}, we select time points $time_1$ and $time_2$ such that $PSA_2 = 2 * PSA_1$ and the doubling time is thus $time_2 - time_1 = PSA_{DT}$. After applying the log_2 transformation, equation (6.1) provides two equations:

$$\begin{cases} log_2 PSA(time_1) = \alpha + \beta * time_1 \\ log_2 PSA(time_2) = \alpha + \beta * time_2 \end{cases}$$

Subtracting one equation from the other, we obtain:

$$\begin{aligned}
\beta &= \frac{log_2 PSA_2 - log_2 PSA_1}{time_2 - time_1} \\
&= \frac{log_2(2 * PSA_1) - log_2 PSA_1}{PSA_{DT}} \\
&= \frac{log_2 \frac{2*PSA_1}{PSA_1}}{PSA_{DT}} \\
&= \frac{log_2 2}{PSA_{DT}} \\
&= 1/PSA_{DT}
\end{aligned}$$

Therefore, the PSA doubling time is equal to the inverse of the slope when the PSA levels are reported using the log_2 scale.

CHAPTER 7
Conclusion

Although each manuscript contains its own discussion, I end this thesis by providing some overall conclusions. I recall the findings of each manuscript, propose some possible statistical refinements, and discuss the clinical implications of this work.

7.1 Results

In the first manuscript, I demonstrated and took advantage of the flexibility of the Bayesian hierarchical model to estimate post-radiotherapy PSA trajectories. Each individual PSA trajectory was estimated using not only the information provided by the relevant series, but from all the PSA series; this borrowing of strength process has permitted me to strengthen the inference. In addition, the model incorporated the multiple complex characteristics of the longitudinal PSA data: the within and between-series variability, the piecewise-linear pattern over time, the presence of a random changepoint, and the non-constant variance. Because of the several advantages of this statistical model, including the borrowing of strength, flexibility and ease of implementation, it should be applicable to the study of other markers series in other diseases.

The second manuscript reported the results of the first numerical validation of the ASTRO criterion. The originality of the statistical approach relied on the simulation of a realistic, sophisticated data set, by fitting an appropriate statistical model, to a rich set of real PSA data. These empirically based simulations were particularly flexible, and allowed me to evaluate the ASTRO criterion, as well as the Houston rule. My results suggested that in most of the situations investigated, the ASTRO

criterion was outperformed by the Houston rule. Finally, the use of consecutive rises, or decreases, is not specific to PSA data. This approach can be applied to evaluate other rules that purport to rapidly and accurately detect up (down) turns in noisy series, such as in other medical data, and in data series used to monitor economic situations.

The last manuscript was directed to physicians, whose task involves the follow-up of patients treated for prostate cancer. I proposed that physicians chart PSA values using a practical charting paper, to obtain a rapid and relatively accurate estimate of the PSA doubling time, without any difficult computations.

7.2 Possible statistical refinements

The Bayesian hierarchical model of the first manuscript could be improved by incorporating additional features of the PSA series. I expressed the PSA trajectory as a function of four parameters: the nadir, the changepoint, the number of PSA halvings before the nadir, and the number of PSA doublings after the nadir. Several variables, such as pre-treatment PSA concentration, or pre-treatment tumor grade have been shown to correlate well with features of the PSA profile, in particular, the PSA doubling time and the PSA nadir. Thus, these covariates could be added to the model to improve estimation.

The assays used to measure the PSA concentrations are less precise at lower PSA levels. In the PSA data set, some concentrations were recorded as 0 ng/ml. Unlike surgery, where the prostate is removed, the prostate keeps producing PSA after radiotherapy, and thus the concentrations should not be null. For this reason, I deliberately assigned a concentration of 0.03125 ng/ml to these observations (corresponding to a log_2PSA value of -5). This procedure affected 81 of the 4667 PSA observations (1.7%). Given this small proportion, my priority was to ensure that the other complex characteristics of the PSA data were adequately accounted

for. I therefore performed the analysis without treating these values as left-censored. Although handling these data as left-censored is unlikely to change the estimates, additional manipulations of the model could be performed in order to properly describe these censored data. Just as the other complex features of the data, left-censoring can be easily handled in WinBugs.

Post-radiotherapy trajectories are more complicated to fit than post-surgery series. In the case of surgery, the prostate is removed, and if any PSA is produced, it is due to cancer tumors only. In the case of radiotherapy, the prostate is not removed, and PSA are produced by both the remaining prostatic tissue, and possibly by the tumor. In further research, one could attempt to account for these remaining healthy cells. This step might be complex however, given that the PSA produced by these cells tend to follow an exponential pattern. Thus, one could model the sum of two exponential PSA curves. This problem has already been addressed by Fitzgerald et al. to model HIV viral rebound following antiretroviral therapy [73]. In their cases, the viral load was assumed to be the sum of two distinct components, one of which declines in response to therapy, and the other may either decline or increase. The viral model was fitted using a non-linear mixed effect model, and expressed as the logarithm of the sum of two exponential components. The same approach could be investigated using a Bayesian model, but my suspicion, based on a very limited investigation, is that the precision of the measurements is not sufficient to clearly estimate the two separate components.

In the second manuscript, I evaluated the probability of observing three consecutive PSA rises, for different underlying PSA growth rates. When the underlying pattern is flat, this probability can be estimated exactly using non-parametric runs tests, or tests of randomness. These tests were initially proposed to assess whether the fluctuations exhibited by a series of observations are random. In the context of

PSA series, a run up (down) is defined as a succession of PSA increases (decreases), preceded and possibly followed by a PSA decrease (increase). The randomness of such series can be statistically tested using the theory of runs; both the number and the lengths of the runs, which are interrelated, should reflect the existence of some sort of pattern. Tables of exact probabilities of at least r runs of length p or more were given for up to 14 observations by Olmstead [76]. For example, Olmstead's results tell us that if a man has an underlying flat PSA curve, and if ten serial PSA measurements are taken, then by chance alone, there is 21% probability of observing at least three consecutive PSA rises, and thus a 21% probability that at some stage the ASTRO criterion will be satisfied, corresponding to a specificity of 79%. On the other hand, if the alternative hypothesis assumes a rising (or declining) pattern, exact methods are particularly difficult to derive analytically, since they imply complex multiple integrations. My approach, based on simulations, allows one to estimate the probability of observing a certain number of consecutive increases (or decline), for various underlying slopes.

Finally, in the last manuscript, I proposed that physicians estimate the PSA doubling time by fitting a line to the log_2 PSA observations. In an empirical study, Mosteller et al. showed that lines fitted by eyes tended to be consistently steeper or shallower than the corresponding least-square regression lines [84]. One possible explanation is that some people tend to minimize the sum of the squares of perpendicular rather than vertical distances. It would be interesting to compare the lines fitted by eye to the log_2PSA data, with the corresponding least squares regression lines, as well as with the lines obtained by minimizing the sum of the perpendicular distances of the points from the line.

7.3 Research implications

My evaluation of the ASTRO criterion for PSA failure was numerical, rather than clinical; I assessed its short-term, rather than its long-term classification properties. Following PSA failure, there may be several years before one observes some specific clinical outcomes, such as death from prostate cancer. Given that both numerical and clinical evaluations are now available, it would be particularly interesting to analyze them simultaneously. Pauler and Finkelstein, for example, evaluated whether some rules for biochemical failure, including the ASTRO criterion, were good predictors of time to clinical progression [48]. They fitted a Cox model using the post-treatment parameters of the PSA profile (PSA nadir, growth rate ...), as well as time varying indicators of whether the rules were satisfied. The authors concluded that the longitudinal model for PSA contributed significantly to the Cox model, and no additional information was gained from any of the rules tested. Further such analysis are needed to evaluate the effect of biochemical failure on clinical failure.

The Connecticut data set may not be representative of the general population of men treated for prostate cancer with radiotherapy. Therefore, the estimates provided in this thesis should not be used directly for health research. However, these estimates bring new insights regarding the parameters of the PSA profile, in particular the changepoint. I estimated that the changepoint occurred about one year earlier than in other published studies, and in addition, my results were far more precise, essentially because of the borrowing of strength allowed by the model. Given the clinical implications of the PSA changepoint, it would be interesting to obtain additional data sets, with more data available around the estimated timing of the changepoint, i.e. measured at shorter intervals.

The method used for the evaluation of rules for biochemical failure was very flexible, and may bring new information regarding the ongoing debate on the validity of the ASTRO rule. I have shown that although not systematic, the Houston criterion

often outperformed the ASTRO criterion. Although my analysis was restricted to specific measurement schedules, other lengths of follow-up or frequencies of measurements could be investigated easily. In addition, many other rules for biochemical failure have been proposed in the literature; again, these could be evaluated easily, given the flexibility of the estimation process.

One advantage of the Bayesian approach is the ability to obtain direct probability statements about the parameters of the PSA trajectory. As a result, specific characteristics, such as the timing of the PSA nadir, or the post-nadir PSA doubling time can be used to propose new definitions for biochemical failure. For example, Slate and Cronin proposed a rule for biochemical failure based on the probability that the PSA changepoint is reached after a specific period of time [49]. Similarly, a rule could use the probability that the post-nadir PSA growth rate, and thus the PSA doubling time, is greater than some specific threshold.

Finally, the estimates of the sensitivity and specificity of the ASTRO rule are 'after the fact' in theoretical scenarios, and pertain to the behavior in a group of men with known error-free trajectories. Given that estimates of the true (error-free) series are now available, it would also be of interest to address the more 'real-time' predictive value question. Specifically, given that the PSA have been measured at the times they have *up to now* for a particular man, and given these observed values, what is the probability that the man's PSA series is truly rising, i.e., what is the probability that the slope of the post-nadir PSA profile is rising? Credible intervals could be obtained for the parameters of the individual PSA profile, such as the doubling time for this man's PSA. One could approach this problem, by dynamically refitting the model described in this thesis. This might be computationally intensive, given that the model would have to be constantly refitted; however, it appears feasible if one uses the same approach based on an appropriate model and fitting.

Appendix 2 : Waivers

Enclosed are the signed waivers from the co-authors of the manuscripts : Dr James Hanley, Dr Lawrence Joseph and Dr Peter Albertsen.

References

[1] National Cancer Institute of Canada. Canadian cancer statistics 2004. Published Online at http://www.cancer.ca/ccs. Accessed April 2005.

[2] Eastham J., Riedel E., Scardino P., Shike M., Fleisher M., Schatzkin A., Lanza E., Latkany L., , and Begg C. Variation of serum prostate-specific antigen levels. *Journal of the American Medical Association*, 289(20):2695–2700, 2003.

[3] McMullen K. and Robert Lee W. A structured literature review to determine the use of the ASTRO definition of biochemical failure. *Urology*, 61(2):391–396, 2003.

[4] International Agency for Research on Cancer. GLOBOCAN 2002. Published Online at http://www-dep.iarc.fr. Accessed March 2005.

[5] Ellison L., Stokes J., Gibbons L., Linday J., and Morrison H. Monograph series on aging-related diseases: X. Prostate cancer. *HEALTH CANADA: Chronic diseases in Canada*, 19(1), 2000.

[6] Catalona W., Smith D., Ratliff T., Dodds K., Coplen D., Yuan J., Petros J., and Andriole G. Measurement of prostate-specific antigen in serum as screening test for prostate cancer. *New England Journal of Medicine*, 324(17):1156–1161, 1991.

[7] Catalona W., Smith D., Ratliff T., and Basler J. Detection of organ-confined prostate cancer is increased through prostate-specific antigen-based screening. *Journal of the American Medical Association*, 270(8):948–954, 1993.

[8] Gleason D.F., Mellinger G.T., and the Veterans Administration Cooperative Urological Research Group. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. *Journal of Urology*, 111:58–64, 1974.

[9] US National Cancer Institutes Of Health. Prostate cancer : Treatments. Health professional version. Published Online at http://cancernet.nci.nih.gov/cancertopics/pdq/treatment/prostate/HealthProfessional/. Accessed March 2005.

[10] Partin A., Kattan M., Subong E., Walsh P., Wojno, Oesterling J., Scardino P., and Pearson J. Combination of prostate-specific antigen, clinical stage and Gleason score to predict pathological stage of localized prostate cancer. *Journal of the American Medical Association*, 277(18):1445–1451, 1997.

[11] Ross P., Scardino P., and Kattan M. A catalog of prostate cancer nomograms. *Journal of Urology*, 165:1562–1568, 2001.

[12] Shipley W., Thames H., Sandler H., Hanks G., Zietman A., Perez C., Kuban D., Hancok S., and Smith C. Radiation therapy for clinically localized prostate cancer. *Journal of the American Medical Association*, 281(17):1598–1604, 1999.

[13] Memorial Sloan-Kettering Cancer Center. www.nomograms.org.

[14] Oesterling J. Prostate-specific antigen: a critical assessment of the most useful tumor marker for adenocarcinoma of the prostate. *Journal of Urology*, 145(5):907–923, 1991.

[15] Stamey T., Yang N., Hay A., McNeal J., Freiha F., and Redwine E. Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. *New England Journal of Medicine*, 317(15):909–916, 1987.

[16] Tchetgen M., Song J., Strawderman M., Jacobsen S., and Oesterling J. Ejaculation increases the serum prostate-specific antigen concentration. *Urology*, 47(4):511–516, 1996.

[17] Herschman J., Smith D., and Catalona W. Effect of ejaculation on serum prostate-specific antigen concentrations. *Urology*, 50(2):239–243, 1997.

[18] Oremek G. and Seiffert U. Physical activity releases prostate-specific antigen (PSA) from the prostate gland into blood and increase serum PSA concentrations. *Clinical Chemistry*, 42(5):691–695, 1996.

[19] Nixon R., Wener M., Smith K., Parson R., Strobel S., and Brawer M. Biological variation of PSA levels in serum: An evaluation of day-to-day physiological fluctuations in a well-defined cohort of 24 patients. *Journal of Urology*, 157(6):2183–2190, 1997.

[20] Ornstein D., Smith D., Rao G., Basler J., Ratliff T., and Catalona W. Biological variation of total, free and percent free serum prostate specific antigen levels in screening volunteers. *Journal of Urology*, 157, 1997.

[21] Ballentine Carter H., Pearson J., Metter J., Brant L., Chan D., Andres R., Fozard J., and Walsh P. Longitudinal evaluation of prostate-specific antigen levels in men with and without prostate disease. *Journal of the American Medical Association*, 267(16):2215–2220, 1992.

[22] Stamey T., Kabalin J., McNeal J., Johnstone J., Freiha F., Redwine E., and Yang N. Prostate-specific antigen in the diagnosis of adenocarcinoma of the prostate. II. Radical prostatectomy treated patients. *Journal of Urology*, 141(5):1076–1083, 1989.

[23] Schmid H., McNeal J., and 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*, 71(6):2031–2040, 1993.

[24] Meek A., Park T., Oberman E., and Wielopolski L. A prospective study of prostate specific antigen levels in patients receiving radiotherapy for localized carcinoma of the prostate. *International Journal of Radiation Oncology Biology Physics*, 19:733–741, 1990.

[25] Albertsen P., Hanley J., Penson D., and Fine J. Validation of increasing prostate specific antigen as a predictor of prostate cancer death after treatment of localized cancer with surgery or radiation. *Journal of Urology*, 171:2221–2225, 2004.

[26] Pollack A., Zagars G., and Kavadi V. Prostate specific antigen doubling time and disease relapse after radiotherapy for prostate cancer. *Cancer*, 74(2):670–678, 1994.

[27] D'Amico A., Cote K., Loffredo M., and Renshaw A. Schultz D. Determinants of prostate cancer specific survival following therapy during the prostate specific antigen era. *Journal of Urology*, 170:S42–S47, 2003.

[28] D'Amico A., Moul J., Carroll P., Sun L., Lubeck D., and Chen M. Surrogate end point for prostate cancer-specific mortality after radical prostatectomy or radiation therapy. *Journal of the National Cancer Institute*, 95:1376–1383, 2003.

[29] Zagars G. and Pollack A. The fall and rise of prostate-specific antigen: kinetics of serum prostate specific antigen levels after radiation therapy for prostate cancer. *Cancer*, 72:832–842, 1993.

[30] Crook J., Bahadur Y., Bociek R., Perry G., Robertson S., and Esche B. Radiotherapy for localized prostate carcinoma. The correlation of pretreatment prostate-specific antigen and nadir prostate-specific antigen with outcome as assessed by systematic biopsy and serum prostate specific antigen. *Cancer*, 79:328–336, 1997.

[31] Hanlon A., Diratzouian H., and Hanks G. Posttreatment prostate-specific antigen nadir highly predictive of distant failure and death from prostate cancer. *International Journal of Radiation Oncology Biology Physics*, 53(2):297–303, 2002.

[32] DeWitt K., Sandler H., Weinberg V., McLaughlin P., and Roach M. What does post-radiotherapy PSA nadir tell us about freedom from PSA failure and progression-free survival in patients with low and intermediate-risk localized prostate cancer? *Urology*, 62(3):492–496, 2003.

[33] Kaminski J., Hanlon A., Horwitz E., Pinover W., Mitra R., and Hanks G. Relationship between prostate volume, prostate-specific antigen nadir, and biochemical control. *International Journal of Radiation Oncology Biology Physics*, 52(4):888–892, 2002.

[34] Hanks G., D'Amico A., Epstein B., and Schultheiss T. Prostate-specific antigen doubling times in patients with prostate cancer: a potentially useful reflection of tumor doubling time. *International Journal of Radiation Oncology Biology Physics*, 27(1):125–127, 1993.

[35] Horwitz E., Vicini F., Ziaja E., Gonzalez J., Dmuchowski C., Stromberg J., Brabbins D., Hollander J., Chen P., and Martinez A. Assessing the variability of outcome for patients treated with localized prostate irradiation using different definitions of biochemical control. *International Journal of Radiation Oncology Biology Physics*, 36(3):565–571, 1996.

[36] Critz F. A standard definition of disease freedom is needed for prostate cancer: Undetectable prostate specific antigen compared with ASTRO consensus definition. *The Journal of Urology*, 167, 2002.

[37] American Society for Therapeutic Radiology and Oncology Consensus Panel. Consensus statement: Guidelines for PSA following radiation therapy. *International Journal of Radiation Oncology Biology Physics*, 37(5):1035–1041, 1997.

[38] Horwitz E., Vicini F., Ziaja E., Dmuchowski C., Stromberg J., and Martinez A. The correlation between the ASTRO consensus panel definition of biochemical failure and clinical outcome for patients with prostate cancer treated with external beam radiation therapy. *International Journal of Radiation Oncology Biology Physics*, 41(2):267–272, 1998.

[39] Hanlon A. and Hanks G. Scrutiny of the ASTRO consensus definition of biochemical failure in irradiated prostate cancer patients demonstrates its usefulness and robustness. *International Journal of Radiation Oncology Biology Physics*, 46(3):559–566, 2000.

[40] Vicini F., Kestin L., and Martinez A. The importance of adequate follow-up in defining treatment success after external beam irradiation for prostate cancer. *International Journal of Radiation Oncology Biology Physics*, 45(3):553–561, 1999.

[41] Vicini F., Kestin L., and Martinez A. The correlation of serial prostate specific antigen measurements with clinical outcome after external beam radiation therapy of patients for prostate carcinoma. *Cancer*, 88(10):2305–2318, 2000.

[42] Hodgson D., Catton C., Warde P., Gospodarowicz M., Milosevic M., McLean M., and Catton P. The impact of irregularly rising prostate-specific antigen and impending failure on the apparent outcome of localized prostate cancer following radiotherapy. *International Journal of Radiation Oncology Biology Physics*, 49(4):957–963, 2001.

[43] Taylor J., Griffith K., and Sandler H. Definitions of biochemical failure in prostate cancer following radiation therapy. *International Journal of Radiation Oncology Biology Physics*, 50(5):1212–1219, 2001.

[44] Thames H., Kuban D., Levy L., Horwitz E., Kupelian P., Martinez A., Michalski J., Pisansky T., Sandler H., Shipley W., Zelefsky M., and Zietman A. Comparison of alternative biochemical failure definitions based on clinical outcome in 4839 prostate cancer patients treated by external beam radiotherapy between 1986 and 1995. *International Journal of Radiation Oncology Biology Physics*, 57(4):929–943, 2003.

[45] Slate E. and Clark L. *Case Studies in Bayesian Statistics*, volume 4, pages 395–412. Springer, 1999.

[46] Pearson J., Morrell C., Landis P., Ballentine Carter H., and Brant L. Mixed-effects regression models for studying the natural history of prostate disease. *Journal of the American Statistical Association*, 13:587–601, 1994.

[47] Morrell C., Pearson J., Ballentine Carter H., and Brant L. Estimating unknown transition times using a piecewise nonlinear mixed-effects model in men with prostate cancer. *Journal of the American Statistical Association*, 90:45–53, 1995.

[48] Pauler D. and Finkelstein D. Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcome. *Statistics in Medicine*, 21:3897–3911, 2002.

[49] Slate E. and Cronin K. *Case Studies in Bayesian Statistics*, volume 3, pages 435–456. Springer, 1999.

[50] Gelman A., Carlin J., Stern H., and Rubin D. *Bayesian data analysis.* Chapman and Hall, 2nd edition, 2004.

[51] Geman S. and Geman D. Stochastic relaxation, Gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[52] Metropolis N., Rosenbluth A., Rosenbluth M., Teller A., and Teller E. Equations of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21:1087–1092, 1953.

[53] Hastings W. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[54] Raftery A. and Lewis S. *Bayesian Statistics 4*, chapter 4. How many iterations in the Gibbs sampler?, pages 763–773. Oxford University Press, 1992.

[55] Spiegelhalter D., Thomas A., Best N., and Gilks W. WinBUGS: Bayesian inference using Gibbs sampling, version 1.3. Available from MRC Biostatistics Unit, Cambridge, 2000.

[56] Ripley B. *Stochastic simulations.* New York : J. Wiley, 1987.

[57] Neal R. Markov Chain Monte Carlo methods based on slicing the density function. Technical Report 9722, Department of Statistics, University of Toronto, 1997.

[58] Page E. A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3/4):523, 1955.

[59] Hinkley D. Inference about the change-point in a sequence of random variables. *Biometrika*, 57(1):1–17, 1970.

[60] Hinkley D. Inference about the intersection in two-phase regression. *Biometrika*, 56(3):495, 1969.

[61] Carlin B., Gelfand A., and Smith A. Hierarchical Bayesian analysis of change-point problems. *Applied statistics-Journal of the Royal Statistical Society, Series C*, 41(2):389–405, 1992.

[62] Stephens D. Bayesian retrospective multiple-changepoint identification. *Applied statistics*, 43(1):159–178, 1994.

[63] Lange N., Carlin B., and Gelfand A. Hierarchical Bayes models for the progression of HIV infection using longitudinal CD4 T-cell numbers. *Journal of the American Statistical Association*, 87(419):615–632, 1992.

[64] Slate E. and Turnbull B. Statistical models for longitudinal biomarkers of disease onset. *Statistics in medicine*, 19:617–637, 2000.

[65] Critz F., Levinson K., Williams W., Holladay D., Griffin V., and Holladay C. Prostate specific antigen nadir achieved by men apparently cured of prostate cancer by radiotherapy. *Journal of Urology*, 161:1199–1205, 1999.

[66] D'Amico A. and Hanks G. Linear regression analysis using prostate-specific antigen doubling time for predicting tumor biology and clinical outcome in prostate cancer. *Cancer*, 72(9):2638–2643, 1993.

[67] Vollmer R. and Montana G. The dynamics of prostate-specific antigen after definitive radiation therapy for prostate cancer. *Clinical Cancer Research*, 5:4119–4125, 1999.

[68] Hanlon A., Moore D., and Hanks G. Modeling postradiation prostate specific antigen level kinetics. *Cancer*, 83:130–134, 1998.

[69] Guess B., Jennrich R., Johnson H., Refheffer R., and Scholz M. Using splines to detect changes in PSA doubling times. *The Prostate*, 54:88–94, 2003.

[70] Irwig L., Glasziou P., Wilson A., and Macaskill P. Estimating an individual's true cholesterol level and response to intervention. *Journal of the American Medical Association*, 266(12):1678–1685, 1991.

[71] Laird N. and Ware J. Random-effects models for longitudinal data. *Biometrics*, 38:963–974, 1982.

[72] Hughes M., Stein D., Gundacker H., Valentine F., Phair J., and Volberding P. Within-subject variation in CD4 lymphocyte count in asymptomatic human immunodeficiency virus infection: Implications for patient monitoring. *Journal of Infectious Diseases*, 169:28–36, 1994.

[73] Fitzgerald A., DeGruttola V., and Vaida F. Modeling HIV viral rebound using non-linear mixed-effect models. *Statistics in Medicine*, 21:2093–2108, 2002.

[74] Pollack A., Hanlon A., and Movsas B. Biochemical failure as a determinant of distant metastasis and death in prostate cancer treated with radiotherapy. *International Journal of Radiation Oncology Biology Physics*, 57:19–23, 2003.

[75] Chauvet M. and Piger J. Identifying business cycle turning points in real time. *The Federal Reserve Bank of St Louis review*, 85(2):47–62, 2003.

[76] Olmstead P. Distribution of sample arrangements for runs up and down. *Annals of Mathematical Statistics*, 17:24–33, 1946.

[77] Levene H. and Wolfowitz J. The covariance matrix of runs up and down. *Annals of Mathematical Statistics*, 15:58–69, 1944.

[78] Sartor C., Strawderman M., Lin X., Kish K., McLaughlin P., and Sandler H. Rate of PSA rise predicts metastatic versus local recurrence after definitive radiotherapy. *International Journal of Radiation Oncology Biology Physics*, 38(5):941–947, 1997.

[79] Pound C., Partin A., Eisenberger M., Chan D., Pearson J., and Walsh J. Natural history of progression of PSA elevation following radical prostatectomy. *Journal of the American Medical Association*, 281(17):1591–1597, 1999.

[80] Patel A., Dorey F., Franklin J., and DeKernion J. Recurrence patterns after radical retropubic prostatectomy: clinical usefulness of prostate specific antigen doubling times and log slope prostate specific antigen. *Journal of Urology*, 158(4):1441–1445, 1997.

[81] Stephenson A., Aprikian A., Souhami L., Behlouli H., Jacobson A., Begin L., and Tanguay S. Utility of PSA doubling time in follow-up of untreated patients with localized prostate cancer. *Urology*, 59(5):652–656, 2002.

[82] McLaren D., McKensie M., Duncan G., and Pickles T. Watchful waiting or watchful progression? Prostate specific antigen doubling times and clinical behavior in patients with early untreated prostate carcinoma. *Cancer*, 82(2):342–348, 1998.

[83] Swanson K., True L., Lin D., Buhler K., Vessella R., and Murray J. A quantitative model for the dynamics of serum prostate-specific antigens as a marker for

cancerous growth. An explanation for a medical anomaly. *American Journal of Pathology*, 158(6):2195–2199, 2001.

[84] Mosteller F., Siegel A., Trapido E., and Youtz C. Eye fitting straight lines. *The American Statistician*, 35(3):150–152, 1981.