Impact of Early Detection on Treatment Effects and Cancer Mortality

by

Shih-Yuan Lee

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2010

Doctoral Committee:

Professor Alexander Tsodikov, Chair
Professor John D. Kalbfleisch
Professor Jeremy M.G. Taylor
Associate Professor David Mendez

ACKNOWLEDGEMENTS

I want to express my sincerest thanks to my advisor Dr. Alex Tsodikov for spending endless hours, even in the late evenings and on the weekends. Also for teaching me and helping me throughout the whole thesis process. I would never have accomplished this work without him. I also want to thank my family for always being there for me when I needed them. Finally, I want to thank all of the faculty members in the Biostatistics department and Dr. Franzblau and Dr. Garabrant in Environmental Health Sciences. I really enjoyed learning from them and working with them. It was one of the best experience of my life.

This work was supported by NIH/NCI grants UO1 CA097414 and T32 CA083654.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS . . . . . **ii**

LIST OF FIGURES . . . . . **v**

LIST OF TABLES . . . . . **vi**

CHAPTER

I. Introduction 1

II. Treatment Effect under Early Detection of Cancer 7

2.1 Introduction 7
2.2 The model, notation, assumptions and preliminaries 9
2.3 Lead-time 13
2.4 The true multiplicative treatment effect of a misspecified model 16
2.5 Estimating treatment effects using a misspecified PH model 19
2.5.1 Assessing the bias of point estimates 19
2.5.2 Small treatment effect approximation for point estimates 23
2.5.3 Measuring the bias of variance estimated by the PH model 24
2.5.4 Small treatment effect approximation for the variance 26
2.6 The meta-analytic correction 28
2.7 Simulation study 29
2.8 Example: Radical prostatectomy vs. watchful waiting for prostate cancer . 31
2.9 Discussion 34
2.10 Appendix 37
2.10.1 Properties of relative expectation 37

III. Predicting Cancer Progression and the Null Hypothesis of Treatment Effect under Screening 39

3.1 Introduction 39
3.2 Natural history model 42
3.2.1 Marginal incidence model 42
3.2.2 Z-specific incidence model 45
3.2.3 Lead-time 46
3.2.4 Disease progression model and the null hypothesis of treatment effect 47
3.2.5 Survival model 51
3.3 Predictions 52
3.3.1 Natural history of Prostate Cancer in US 52
3.3.2 Risk of adverse events for local-regional screen-detected patients . . 56
3.4 Data analysis and results 58
3.5 Discussions 60

IV. Mortality Model for Prostate Cancer 63

4.1 Introduction 63
4.2 Model 66
4.2.1 Basic Model 66
4.2.2 Mortality Model 66
4.2.3 Treatment Model 67
4.2.4 Disease Progression Model 68
4.2.5 Survival Model 69
4.3 Data Analysis and Results 71
4.4 Discussions 81

V. Discussion 84

BIBLIOGRAPHY **87**

LIST OF FIGURES

Figure

2.1 A comparison of the true treatment effect without screening $\theta_0 = e^{\beta_0}$ under the PH model (2.12) vs. the time-dependent true treatment effect θ_t under screening (2.13). Top left: Baseline hazard function h. Top right: The distribution (pdf with mass at zero) of the lead time W. Bottom: The true multiplier, a ratio (2.14) of the true effect under screening averaged over the lead time to the effect without screening. 20

2.2 True and large-sample characteristics in screened and unscreened populations under various shapes of the baseline hazard. Left: The baseline hazard in the unscreened population $h(t)$ and its counterpart under screening $\lambda(t)$ averaged over the lead time W as in (2.5). Right: True hazard ratio in the unscreened population θ_0, its time-dependent counterpart under screening θ_t, and $\theta^*(\tau) \approx \theta_0^{m^*(\tau)}$ representing the large-sample limit of the PH-estimated hazard ratio under screening. τ is the duration of the study; $t \leq \tau$ is a point in follow-up time. Note that the misspecified PH model produces a biased estimate $\hat{\theta}(\tau) \xrightarrow{p} \theta^*(\tau)$ of the average true hazard ratio θ_t under screening. Computation of m^* is done under the Assumption II.1 approximation. 25

2.3 Quality of the small treatment effect approximation for the misclassification bias. Left:Shapes of the baseline hazard function without screening h vs. under screening λ. Middle and right column: Variable polygons and smooth curves going through them correspond to PH model estimated and average hazard ratios at various study durations τ. The lines at $y = \theta_0$ correspond to a simulation without screening ($W = 0$ with probability 1). The dashed curve showing a departure upwards particularly with small θ_0 (large beneficial treatment effect) is the hazard ratio under screening predicted using the approximation (2.24) for the multiplier. . . . 32

3.1 Structure of the chronic disease three-state model 43

4.1 Observed and expected incidence by year. 72

4.2 Expected stage and grade distribution by age and year of diagnosis. 73

4.3 Distribution of SEER treatment by age of diagnosis and grade. 75

4.4 Distribution of SEER treatment by year of diagnosis and grade. 76

4.5 Distribution of Hormone Therapy (HT) by year of diagnosis and grade. 77

4.6 Mean lead time given stage, grade, age, and year of diagnosis. 78

4.7 Observed and expected mortality by calendar year. 82

LIST OF TABLES

Table

2.1 Empirical vs. approximate variance ($\theta_0 = 0.5$, τ=15, n=10000) 31

2.2 Estimated and predicted relative risks of radical prostatectomy vs. watchful waiting 34

3.1 Results of Estimated Progression Probability Matrix (PPM) and 95% C.I. 51

3.2 Predictions of key prostate cancer natural history events 59

3.3 Lifetime risk of adverse events for local-regional screened-detected patients 60

4.1 Estimated treatment effects by stage and grade in Generalized stage-shift model . 80

4.2 Estimated treatment effects by stage and grade in Calibration model 80

CHAPTER I

Introduction

In recent years, rapid development of modern biomarker research and continuous growth of public interest in early prevention and treatment of cancer have resulted in an increased use of early detection programs (screening). These developments have led to a dramatic increase in the incidence of early stage diagnoses, most of which would have never happened within the patient's lifespan without screening (overdiagnosis).

Screening interventions interact with the latent disease progression process. First, the diagnostic test sensitivity increases as the disease progresses and becomes easier to detect. Second, early detection advances the diagnosis by the amount of the so-called lead time (Zelen and Feinleib, 1969) which adds to the patient's survival, and the disease presents at an earlier stage with more favorable clinical characteristics. Third, slower progressing diseases are easier to catch by the test (length-bias, Zelen and Feinleib (1969)) while they are still latent, while aggressively developing disease usually results in diagnosis due to symptoms (clinical diagnosis) rather than the test (screening diagnosis). As a result of the diagnostic intervention, the population is split into less aggressive diseases diagnosed by the test (length bias) and the more aggressive ones missed by the test (anti-length bias, Zelen (2004)).

To improve cancer control policy under limited resources, a quantitative measure of the effectiveness of cancer screening and treatment interventions is needed. Two dimensions define the success of cancer control interventions: the population dimension (cancer incidence, distribution of the disease presentation at diagnosis, cancer mortality) and the subject-specific level (subject's prognosis based on clinical information available at diagnosis and knowledge of the disease heterogeneity and utilization of screening in the population from which the subject was sampled).

One way to assess the benefit of screening is via randomized screening trials. However, such trials need to recruit a lot of patients for a long period of time to be able to have sufficient power to compare outcomes. They are also conditional on specific screening patterns, and the results cannot be generalized to other populations without use of modeling.

The randomized clinical trial (RCT) is the gold standard to evaluate treatment effects. Randomization is believed to be a tool to obtain a conservative test in the presence of confounding, unbiased under the null hypothesis (Schumacher et al., 1987; Gail et al., 1984; Struthers and Kalbfleisch, 1986). However, as we show in this study, dependent on the model, the estimated treatment effects could be biased (not necessarily conservatively) if important covariates are not controlled for specifically in a model-based analysis.

These consideration lead us to pursue a statistical modeling approach to evaluate cancer screening and treatment interventions. Investigators from the Cancer Intervention and Surveillance Modeling Network (CISNET, http://cisnet.cancer.gov/), of which we are part of, use statistical modeling to study the effects of cancer control interventions on population trends in incidence (Etzioni et al., 1999, 2002; Draisma et al., 2003; Davidov and Zelen, 2004; Tsodikov et al., 2006) and mortality (Berry

et al., 2005; Lee and Zelen, 2008). Other researchers (Parker et al., 2006) developed statistical models to study disease progression and survival for screened detected patients. The challenge of modeling is that a shallow statistical model cannot explain the observed phenomena and the results cannot be generalized to populations under different screening patterns. Complex simulations on the other hand are over-parameterized and need to rely on many ad-hoc parameter specifications based on outside sources and literature. Our approach is to use mechanistic statistical modeling in conjunction with statistical inference methods to leverage the precision and confidence of statistical methodology and the explanatory and predictive power of mechanistic models.

The objective of this dissertation is to develop quantitative measures of the impact of early detection followed by treatment on clinical outcomes and to evaluate the success of the combined screening and treatment interventions. We extend our previous work on marginal and stage/grade specific incidence model (Tsodikov et al., 2006; Chefo and Tsodikov, 2009) and build a joint hierarchical family of models of prostate cancer from the point of onset to the point of death. Using the models we study the heterogeneity induced by screening interventions in the population, and its interaction with treatments applied at diagnosis of the disease as they affect the clinical and disease outcomes such as disease prognosis and cancer mortality. Our models are statistical in nature and are fit to observed population and subject-level data before predictions are made.

The rest of the dissertation is divided as follows:

- **Chapter 2. We assess how the the early detection of cancer affects the outcome from clinical trials.**

Cancer-specific survival measured from the point of diagnosis is the most common endpoint in cancer clinical trials and observational studies. Early detection of cancer leads to variability of the point of diagnosis advanced by the amount of the so-called lead time, a random variable. Estimated treatment effects by the proportional hazards (PH) model may be biased if this variability is ignored. Three distinct problems studied in this chapter are of interest with this specific model misspecification mechanism: (1) How the true multiplicative treatment effect differs in screened vs. unscreened populations; (2) How they are estimated using a misspecified model (PH); (3) How the bias and standard errors can be corrected using a meta-analytic approach that does not require the raw data. To address these questions we use a joint cancer incidence and survival modeling approach and illustrate it using simulation and real prostate cancer data. To reduce the dependence on raw data, a small treatment effect approximation to the asymptotic inference with a misspecified PH model is pursued.

- **Chapter 3. We develop an analytic joint statistical model of cancer incidence, presentation at diagnosis, and progression.**

We develop a statistical model for the natural history of the disease and its interaction with screening. The model can be decomposed into four major components: (1) The marginal incidence model with age at prostate cancer diagnosis as an endpoint. Its outcome is the prostate cancer incidence as a functional of the distributional characteristics of screening utilization process operating in the population. (2) A model of disease presentation at diagnosis (stage and grade of the disease). The disease presentation at diagnosis can be considered

a multivariate mark to the point process of diagnoses of prostate cancer in the population. Combined with the marginal incidence model this model gives a marked point process model describing the stage and grade (Z) specific incidence. The model predicts the probability of being diagnosed with a specific stage and grade at cancer incidence. (3) The disease progression model is the main contribution of this chapter. It defines the probability of disease progression for the early detected patient after the screening diagnosis. Understanding this model is critical for measuring how the early treatment intervention enabled by the early diagnosis favorably affects the patient's prognosis. For patients deciding to defer treatment, the model gives an assessment of the risk of disease progression under watchful waiting (conservative management of the disease). Treatment effects built into the progression model define the benefit by way of preventing the development of more advanced stages of the disease, the so-called stage shift.

- **Chapter 4. We develop an analytic joint statistical model of survival post-diagnosis and cancer mortality in the presence of screening and treatment interventions.**

We apply the models developed in Chapter 3 to characterize the heterogeneity of the prostate cancer patient population and describe the latent disease characteristics for the US male population. This information is used as a frailty for modeling survival post-diagnosis adjusted for the lead-time and length bias, and to adequately describe the treatment effects in the situation of variable point of diagnosis modulated by screening operating in the population. We then

synthesize all the models into the mortality model by integrating out the intermediate outcomes of cancer diagnosis, stage, grade, treatment and survival. We also estimate the risks of potential adverse events for prostate-specific antigen (PSA)-detected patients given the prognostic factors at the time of diagnosis. The model is also used to estimate the treatment efficacy of radiotherapy and radical prostatectomy against watchful waiting and study how treatment effects are influenced by the early detection programs. All analyses use data from the Surveillance, Epidemiology and End Results (SEER) database.

In our final chapter, we summarize the strengths and the limitations of our current approach and discuss the direction for future research.

CHAPTER II

Treatment Effect under Early Detection of Cancer

2.1 Introduction

Cancer is one of the top five causes of death in the US and in many developed countries. To reduce cancer mortality and cancer burden, combined use of the early detection and more effective treatments has become the major trend of cancer interventions. Rapid development of modern biomarker research in recent years makes cancer screening tests more sensitive than ever, and we are seeing cancers that were never detectable before. Early detection advances cancer diagnosis by the amount of the lead time, a random variable. This results in over-diagnosis of the disease, increasing cancer incidence and seemingly improved survival from the point of diagnosis even if treatment is not effective.

The most common approach in cancer clinical trials and observational studies is to use the proportional hazards (PH) model to estimate the multiplicative treatment effect. The model is fitted to cancer-specific survival time measured from the point of diagnosis.

Suppose the PH model is a valid one in an unscreened population. Then in the screened population, the PH model is misspecified because cancer-specific survival includes an additive random lead time effect. When the effect of the lead time is

integrated out, the model changes its form because the link function of the PH model is a nonlinear one. Model misspecification, including one induced by random effects, received considerable attention in the literature. Kempthorne (1977) and Fienberg (1980) showed that collapsing a contingency table results in a biased estimator for a binary response outcome. Gail et al. (1984) further summarized asymptotic biases for common models in the exponential family GLMs when important covariates were omitted. They showed that the asymptotic bias is not zero unless the model is linear or exponential. Similar effects were observed with the PH models for failure time data. Lagakos and Schoenfeld (1984) demonstrated how ignored covariates result in a reduced power of the log-rank test; Schumacher et al. (1987) and Gail et al. (1984) showed a conservative bias in the proportional hazards setting and Struthers and Kalbfleisch (1986) studied biased treatment effect under a misspecified PH model when the true model is accelerated failure time or when an important PH covariate is omitted.

In this study, we build on the studies of a general misspecified PH model by assessing the bias of the variance estimator, deriving a small treatment effect Taylor approximation, developing a meta-analytic correction of the biases, and studying a specific misspecification pattern associated with ignoring the heterogeneity induced by the early detection of cancer.

1. We will study the true multiplicative effect. In the correctly specified PH model, the true hazard ratio is independent of time. We assume that this model operates in an unscreened population. When the model is misspecified, in the screened population, the true hazard ratio becomes time dependent. We will study the direction and the magnitude of the difference between the two multiplicative effect measures in Section 2.4.

2. It is a common practice to fit the PH model regardless of whether patients are recruited in the screening era or not. We will study the bias measuring the difference between the true hazard ratio in the unscreened population and what the PH estimator is consistent for under screening, in Section 2.5. Note that because PH model is misspecified for the screened population, it estimates neither the unscreened hazard ratio nor the true time-dependent hazard ratio under screening correctly.

3. Correct estimation of the treatment effect can be accomplished by fitting the right model to raw survival data. However, obtaining raw data from a series of international clinical trials is a logistical challenge. Besides, the correct model reproducing treatment effects in the presence of lead time is a subject of scientific debate. Alternatively, we propose a simplified meta-analytic approach that allows us to approximately correct the bias without using raw data from the clinical trial. We assume that covariate effects (treatment) are small, and approximate the bias in point estimates and standard errors up to the first order term. Characteristics of the unscreened population and the distribution of the lead time necessary to perform the correction are estimated using large sample cancer registry data and cancer incidence models, and are assumed known in a relatively small sample analysis of a clinical trial. This leads to an approximate correction for the bias in Section 2.6

2.2 The model, notation, assumptions and preliminaries

Let g, G, λ, Λ denote the density (pdf), the survival function (sf), the hazard function (hf), and the cumulative hazard function (chf) of the true model, respectively. We assume that the true model is a departure from a PH model explained by

unobserved heterogeneity summarized by a frailty random variable (r.v.) W, possibly a vector. Given W, the baseline survival distribution is represented by the pdf f, sf F, hf h, and chf H. The effect of treatment r.v. Z is modeled via an exponential predictor $\theta(\beta, z) = \exp(\beta z)$, where β is a vector of regression coefficients. For example, Z may represent a binary treatment assignment in a clinical trial, dose of the treatment agent, or generally any set of variables characterizing the specific treatment of the disease.

At the complete data level (given W) the model is

$$\lambda(t|Z,W) = \theta(Z)h(t|W),\ \ G(t|Z,W) = F(t|W)^{\theta(Z)},\ \ g = \lambda G,\ \ \Lambda = -\log G. \tag{2.1}$$

The misspecified PH model is an average of the complete data model (2.1)

$$\lambda(t|Z) = \mathrm{E}\{\lambda(t|Z,W)|Z, T > t\},\ \ G(t|Z) = \mathrm{E}\{G(t|Z,W)|Z\}, \tag{2.2}$$

where expectations are taken over the conditional distribution of W, given Z, and T is the survival time being modeled.

In the sequel, for brevity, we will suppress the arguments t, Z, W of the functions such as $\lambda(t|Z,W)$ or $h(t|W)$, and assume that they are evaluated at the true β_0 unless noted otherwise and explicitly. Also, we will use explicit notation for conditional expectations introduced in Tsodikov (2003). Define a *relative expectation* as

$$\mathrm{E}(\mu||\nu) = \frac{\mathrm{E}(\mu\nu)}{\mathrm{E}(\nu)} \tag{2.3}$$

for any functions μ and ν of some random variables. The notation is motivated by the representation

$$\mathrm{E}\{\mu(X)|A\} = \frac{\mathrm{E}\{\mu(X)1_A(X)\}}{\mathrm{E}\{1_A(X)\}}, \tag{2.4}$$

where X is a random variable, A is a measurable event, 1_A is an indicator of the event (=1 if $X \in A$, and 0 otherwise), and $\mathrm{E}\{1_A(X)\} = \Pr\{A\}$. We arrive at (2.3)

by substituting a general function ν for $1_A(X)$ in (2.4), and observing that (2.3) is still a conditional expectation. Indeed, the right part of (2.3) is an integral of $\mu(x)$ over some (conditional) probability measure $P_{\nu,X}$ built using r.v. X and the function ν

$$\mathrm{E}\{\mu(X)||\nu\} = \int \mu(x) P_{\nu,X}(dx), \quad P_{\nu,X}(dx) = \frac{\nu(x)\Pr\{dx\}}{\int \nu(x)\Pr\{dx\}}.$$

Note that the probability measure $P_{\nu,X}(dx)$ represents a length-biased distribution of X so that the relative expectation could be written as a conditional expectation referring to the condition characterizing the length biased sampling of X. However, this interpretation is irrelevant when no length biased sampling is actually going on in the data. We use the relative expectation to avoid defining an artificial length biased sampling scheme with the only purpose of wanting to use conventional conditional expectation notation for some expressions that look like the right part of (2.3). Besides we use convenient algebra associated with (2.3) as described in the Appendix. The convenience of this notation is that it explicitly expresses a conditional expectation through unconditional ones without having to define a random variable that induces the conditional probability measure. For example, instead of conditioning on $T > t$ when averaging the hazard function among survivors at t in (2.2), we can write explicitly using (2.3),

$$\lambda(t|Z) = \mathrm{E}\{\lambda(t|Z,W)||G(t|Z,W), Z\} = \mathrm{E}\{\lambda||G, Z\} = \theta \mathrm{E}\{h||F^\theta, Z\}. \tag{2.5}$$

Here $\mathrm{E}\{\mu||\nu, Z\}$ is interpreted as (2.3) where all expectations are conditional on r.v. Z in the usual sense. Note that for any non-random ν, or when $\mu = \mu(X)$, and $\nu = \nu(Y)$, and $X \perp Y$, we have $\mathrm{E}(\mu||\nu) = \mathrm{E}(\mu)$. Under the PH model h and F in (2.5) are non-random, so $\mathrm{E}\{h||F^\theta, Z\} = h$ leading to the natural PH expression $\lambda = \theta h$.

Presented in Appendix are useful properties of the relative expectation (2.3) and its derivatives that will be used throughout the paper.

Of primary interest in the example will be a model misspecification induced by the lead time. Lead-time measures how much a clinical diagnosis due to symptoms is advanced as a result of early detection by a screening test. With W playing the role of the lead time, we have the specific model

$$h(t|W) = h(t-W)1_{\{t>W\}}. \tag{2.6}$$

Note that this simple expression implies a number of important assumptions.

1. In the absence of screening when diagnosis occurs clinically due to symptoms (CDx), the model is characterized by a non-random baseline hf $h(t)$. When patient's diagnosis is advanced by screening by the amount of the lead time W, survival during the lead time is guaranteed as patients do not die before they develop symptoms. Hence their hf is zero during the lead time as specified by the $1_{\{T>W\}}$ term in (2.6). In the sequel we will omit $1_{\{T>W\}}$ for brevity assuming that $h(t|W) = 0$, $t \leq W$.

2. In the presence of screening, survival time distribution cannot be exponential, because $h(t|W)$ cannot be a constant in t unless it is uniformly zero.

3. For the same reason, h cannot be a decreasing function of t without being uniformly zero.

4. The support of survival time in the presence of screening is affected by the missing variable W. This is a key distinction of model (2.6) from other misspecification models considered in the literature.

5. In a screened population, some patients are still detected clinically either be-

cause they were not screened, or because their cancer was missed by screening. Such patients by definition have $W = 0$ with probability 1, hence the distribution of W has a mass at 0 (see Figure 2.1, top right). Expression (2.6) presumes that survival of an early detected patient can be decomposed into the survival to the point when the patient would be detected clinically, plus the survival of the unscreened patient. In (2.6) the survival pattern post projected point of CDx (represented by the form of baseline hf h) is the same as without screening.

The bias and variance meta-analytic correction will be based on the following simplifying assumptions.

Assumption II.1.

1. *Random variables Z and W are independent $Z \perp W$;*
2. *Censoring is independent of W;*
3. *True treatment effects β_0 are small so a first order Taylor approximation with respect to β_0 can be used.*

While not essential for theoretical expressions, Assumption II.1 allows one to adjust reports based on the PH model analysis using knowledge of first three moments of Z only, without having to hypothesize multivariate distributions of Z, W and the survival time, by meta-analysis. Having to specify the multivariate distributions would defeat the purpose of meta-analysis essentially requiring a fit of the correct mixed model to raw data or a poorly justified guess.

2.3 Lead-time

Distribution of the lead-time W is a crucial piece of input for the analysis of this paper. It is estimated in Tsodikov et al. (2006) from population and cancer incidence

data amassed in the Surveillance, Epidemiology, and End Results (SEER) cancer registry (www.seer.cancer.gov). Let $\lambda_I(a \,|\, x)$ be the hazard function (incidence rate) of cancer diagnosis at age, $A = a$, for a person born in year x. Cancer development passes through the disease-free state and the pre-clinical state before being detected or censored without a diagnosis. The subject's age at tumor onset A_O represents the duration of the disease-free stage. The duration of latent cancer growth in the pre-clinical stage is given by the delay time $\xi_D = A - A_O$, a backward recurrence time, represented by the period between cancer onset at the age of A_O and its diagnosis at the age of A. Given onset time, cancer diagnosis is a result of two competing risks, the time to detection by screening, ξ_{SDx}, and the time to clinical diagnosis due to symptoms of the disease, ξ_{CDx}, so that the delay time is $\xi_D = \min(\xi_{SDx}, \xi_{CDx})$. The time ξ_{CDx} is referred to as the sojourn time. Conditional ξ_{CDx} given screen diagnosis, is stochastically larger than the unconditional one, due to the length bias. If the unconditional ξ_{CDx} is exponential then the conditional will be larger by a factor of 2 by the lack of memory property and symmetry. Suppose $\mathcal{A} = \{a_1, a_2 \dots\}$ are random ages when the subject is screened (a screening schedule). We have used a two-stage model for the screening schedule point process $\mathcal{A}$. Let $\lambda_{1S}(a, t)$ be the hazard of the first Prostate-Specific Antigen (PSA) test for a man of age a in year t. Then the probability that a man born in year x will not be tested by the age of a is

$$G_{1S}(a \,|\, x) = \exp\left\{ - \int_0^a \lambda_{1S}(\xi, x + \xi) d\xi \right\}. \tag{2.7}$$

We assume that in men who already had their first PSA tests, secondary tests $\{a_2, a_3 \dots\}$ form a non-homogeneous Poisson process with intensity $\lambda_{2S}(a, t)$. Both intensities of PSA testing λ_{1S} and λ_{2S} are treated as known bivariate functions estimated by approximating the output of a random schedule generator developed by the National Cancer Institute (NCI) Mariotto et al. (2007).

The survival function $G_{SDx}(\xi \mid x, A_O)$ representing the probability of no screening diagnosis for a subject born in year x, with tumor onset at the age of A_O, and delay time since onset ξ is derived in Tsodikov et al. (2006) as

$$G_{SDx}(\xi \mid x, A_O) = G_{1S}(A_O + \xi \mid x) + [1 - G_{1S}(A_O \mid x)]G_{2SDx}(\xi \mid x, A_O, A_O) + \int_0^\xi [1 - \alpha(\zeta)]f_{1S}(A_O + \zeta \mid x)G_{2SDx}(\xi - \zeta \mid x, A_O + \zeta, A_O)d\zeta, \tag{2.8}$$

where α is age-dependent screening sensitivity, and

(2.9)

$$G_{2SDx}(\xi \mid x, a, A_O) = \exp\left\{-\int_{\max(A_O - a, 0)}^{\xi} \lambda_{2S}(a + \zeta, x + a + \zeta)\alpha(\zeta + a - A_O)d\zeta\right\},$$

and $\int_a^b = 0$ for any $b \leq a$. The above expressions are a result of averaging over $\mathcal{A}$ and the Bernoulli outcomes of screening tests with probability of success α within the subject, given A_O. Using conditional independence of the competing risks of cancer diagnosis by screening and clinically, given age at tumor onset A_O, we have

$$\lambda_I(a \mid x) = -\frac{d}{da}\log \mathrm{E}\big\{G_{CDx}(a - A_O \mid x, A_O)G_{SDx}(a - A_O \mid x, A_O)\big\}, \tag{2.10}$$

where G_{CDx} is the survival function of the sojourn time.

This model was fitted by maximizing a parametric likelihood for incidence rates.

The joint distribution (pdf) of the lead time $W = \xi_{CDx} - \xi_{SDx}$, a forward recurrence time, and age at diagnosis A is given by

$$f_{LT}(w, a|x) = \int_0^a f_O(y|x)f_{CDx}(a - y + w|x, y)\left\{\begin{array}{ll} G_{SDx}(a - y|x, y), & w = 0 \\ f_{SDx}(a - y|x, y), & w > 0 \end{array}\right\} dy,$$

where f_O is a pdf of the age at onset, and f_{CDx}, f_{SDx} are pdfs corresponding to sf G_{CDx}, G_{SDx}, respectively. Expression under the integral is a joint pdf of age at onset y, screen diagnosis at a, and potential clinical diagnosis at $a + w$.

Finally, the conditional lead-time distribution used in this paper is obtained from the joint one

$$f_{LT}(w|a,x) = \frac{f_{LT}(w,a|x)}{f_I(a|x)}, \tag{2.11}$$

where f_I is the pdf corresponding to the cancer incidence rate λ_I. Shown in Figure 2.1 is a representative lead time distribution for a 65 year old patient diagnosed in 1995 predicted by the incidence model Tsodikov et al. (2006).

2.4 The true multiplicative treatment effect of a misspecified model

Assuming that in the reference group corresponding to no treatment $\theta(z_{\text{ref}}) = 1$, let $\theta(Z)$ represent the hazard ratio in the correctly specified PH model (the one without screening)

$$\lambda(t|Z) = \theta(Z)h(t). \tag{2.12}$$

In the the mis-specified model (2.5) the hazard ratio is time dependent

$$\theta_t = \frac{\lambda|_{\theta(Z)}}{\lambda|_{\theta=1}} = \frac{\theta \mathrm{E}\{h||F^\theta, Z\}}{\mathrm{E}\{h||F, Z\}}. \tag{2.13}$$

Define a correction factor, the true multiplier $m(t\,|\theta)$ characterizing the effect of model misspecification on the hazard ratio

$$m(t\,|\,\theta) = \frac{\theta_t}{\theta} = \frac{\mathrm{E}\{h||F^\theta, Z\}}{\mathrm{E}\{h||F\}}. \tag{2.14}$$

When the treatment is effective (i.e. $\theta < 1$), $m(t\,|\,\theta) > 1$ implies that the actual treatment effect is smaller than θ, so the effect is conservatively attenuated. On the other hand, $m(t\,|\,\theta) < 1$ means that the actual treatment effect is larger than θ_0 so the effect is anti-conservatively attenuated.

Key to the properties of the multiplier are given by the following two lemmas.

Lemma II.2. *The logarithmic derivative of the multiplier is proportional to the relative covariance between the negative baseline cumulative hazard -H = log F and the baseline hf h.*

$$\frac{\partial \log m(t \,|\, \theta)}{\partial \theta} = \frac{\mathrm{Cov}\left\{-H, h \,\middle|\middle|\, F^{\theta}, Z\right\}}{\mathrm{E}\left\{h \,\middle|\middle|\, F^{\theta}, Z\right\}}. \tag{2.15}$$

The result follows from (2.36) in the Appendix.

Lemma II.3. *A version of Chebyshev's inequality (See Shea (1979)). Let $u(x)$ and $v(x)$ be functions and W be a random variable such that $E\{u(W)\}$, $E\{v(W)\}$, and $E\{u(W)v(W)\}$ exist. Then*

1. *if u and v are both nonincreasing or both nondecreasing then*

$$\mathrm{Cov}(u(W)v(W)) \geq 0$$

2. *if one of u and v is nonincreasing and the other nondecreasing then*

$$\mathrm{Cov}(u(W)v(W)) \leq 0$$

3.

$$\mathrm{Cov}(u(W)v(W)) = 0$$

if and only if at least one of u and v is a constant.

The following observations follow immediately from (2.14). When $m(t \,|\, \theta) = 1$, there is no bias, $\theta_t = \theta$. Clearly, $m(t \,|\, \theta) = 1$ uniformly in t when there is no treatment effect ($\theta = 1$). Also, $m(t \,|\, \theta) \to 1$ as $t \to 0$, indicating that there is no bias at the start of follow up.

When the baseline hazard h, is an increasing function, $u(x) = h(t|x) = h(t-x)1_{t>x}$ is decreasing in x while $v(x) = -H(t-x)$ is increasing. By Lemmas II.2, II.3 the

logarithmic derivative is negative so $m(t\,|\,\theta)$ is a decreasing function of θ. This implies that θ_t is biased conservatively toward the null hypothesis $\theta = 1$. Indeed since $m(t\,|\,1) = 1$, in the left neighborhood of $\theta = 1$ we have $m(t\,|\,\theta) > 1$, and consequently $1 > \theta_t > \theta$ at least when θ is close to 1. Alternatively, when $\theta > 1$ (treatment is harmful) we still have attenuation towards the null as $\theta > \theta_t > 1$.

With small treatment effects when θ is close to 1 (2.15) gives the departure of the multiplier from 1. Indeed, expanding $m(t|\theta)$ around $\beta = 0$ we have

$$\begin{aligned} m(t|\theta) &= m(t|\theta)|_{\theta=1} + \left.\frac{\partial m(t|\theta)}{\partial\theta}\right|_{\theta=1}(\theta-1) + o(1) \\ \Rightarrow m(t|\theta) - 1 &\approx \left[\frac{\partial \log m(t|\theta)}{\partial\theta}\right]_{\theta=1}(\theta-1). \end{aligned}$$

It is important to note that $u(x)$ cannot be made an increasing function of x, because $u(x) = 0$ for $x \geq t$ by definition, and h is nonnegative. If, for the sake of argument, this were possible, then by Lemma II.3 we would have optimistic (anti-conservative) bias with decreasing hazard. Also, under the same fantasy, by the last statement of Lemma II.3 and the fact that H cannot be a constant, exponentially distributed survival in the absense of screening ($h = Const$) would be the only case of no bias uniformly in t. The unbiasedness under exponential survival would hold regardless of the size of the treatment effect by virtue of (2.14). None of these scenarios can take place because $h(t|x) = 0$, $x \geq t$. However, they help understand the behavior of m under non-monotonic h. Generally, when h is non-monotonic, the direction of the bias term defined by the multiplier, $m(t|\theta)$, depends on the shape of the hazard function h, and the distribution of the lead time W. Averaging over W will weigh increasing and decreasing areas of h against each other as they contribute to the opposite behavior patterns of m. The result will depend on the weights provided by the form of the pdf of W, conditional on survival up to t.

For cancers where cure is a possibility, such as prostate cancer, the hazard function typically increases initially but turns into decreasing one eventually, as with any cure model $h(t) \to 0$ as $t \to \infty$. Note that over-diagnosed cancers contribute greatly to the chance of 'cure'.

The treatment in a screened population may appear more efficient than it actually is, dependent on the time t when the effect is evaluated. Shown in Figure 2.1 is the multiplier behavior for a unimodal baseline hazard function for various values of θ of the treatment effect and under the lead time distribution shown in the upper right corner of the figure as estimated in Tsodikov et al. (2006) and outlined in Section 2.3. Note that there is one point in the follow-up time around 18 years when the true average multiplicative treatment effect is unbiased.

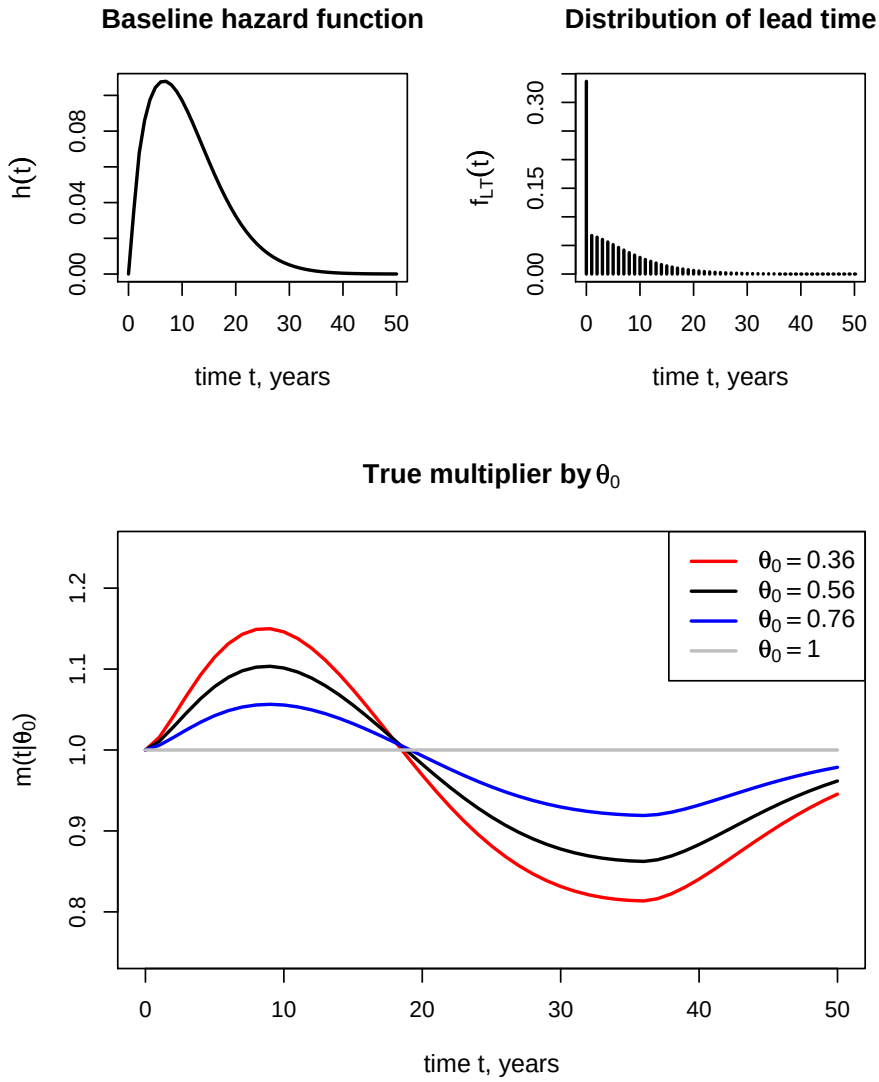
The behavior of the bias described in this section is more complicated than in the case of misspecification induced by ignored PH covariates (Lagakos and Schoenfeld, 1984; Gail et al., 1984; Struthers and Kalbfleisch, 1986; Schumacher et al., 1987) that always leads to a conservative bias. This is because in the latter case the mixed effect is not in the argument of the possibly non-monotonic h.

2.5 Estimating treatment effects using a misspecified PH model

2.5.1 Assessing the bias of point estimates

In Section 2.4, we studied how the true treatment effect is modified by the early detection. However, when the effect is estimated by the misspecified PH model, the PH estimator is generally consistent for some hazard ratio that is neither the true underlying hazard ratio θ_0, nor the average true hazard ratio θ_t. The estimators will depend on the duration of the study τ. We assume that τ is non-random and marks the right extreme of the time to censoring. Denote by $\theta^*(\tau)$, $\beta^*(\tau)$ the large sample limits of the estimators $\hat{\theta}$, $\hat{\beta}$ based on fitting the misspecified PH model.

Figure 2.1: A comparison of the true treatment effect without screening $\theta_0 = e^{\beta_0}$ under the PH model (2.12) vs. the time-dependent true treatment effect θ_t under screening (2.13). Top left: Baseline hazard function h. Top right: The distribution (pdf with mass at zero) of the lead time W. Bottom: The true multiplier, a ratio (2.14) of the true effect under screening averaged over the lead time to the effect without screening.



Let X be time to event (failure or censoring) measured from observed diagnosis (CDx or SDx whichever comes first), and $\delta = 1_{\text{failure}}$ is an indicator of failure (=1 if failure, =0 if censoring). The data is represented as a sample of independent triplets (X_i, δ_i, z_i), $i = 1 \cdots n$, where z is a vector of treatment covariates, and Z the corresponding random vector. For an individual i define the counting processes $Y_i(t) = I(X_i \geq t)$, $N_i(t) = I(X_i \leq t, \delta_i = 1)$. Define an empirical analog of the relative expectation $\mathrm{E}\{Z||\theta(\beta)GS\}$, where G, S are the true survival functions for time to failure and censoring, respectively, as

$$\hat{\mathrm{E}}\{Z||\theta(\beta), Y\} = \frac{\sum_{j=1}^{n} z_j Y_j(t) \exp(\beta z_j)}{\sum_{j=1}^{n} Y_j(t) \exp(\beta z_j)}. \tag{2.16}$$

The score function for the partial likelihood PL (Cox, 1972) can be written as

$$U_n(\tau, \beta) = \frac{\mathrm{d} \log PL(\tau, \beta)}{\mathrm{d}\beta} = \sum_{i=1}^{n} \int_0^\tau \left\{ z_i - \hat{\mathrm{E}}\{Z||\theta(\beta), Y\} \right\} \mathrm{d}N_i(t), \tag{2.17}$$

where β is the regression coefficient (an arbitrary argument), and τ is the duration of the study. By the (uniform) law of large numbers

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^{n} z_i Y_i(t) \theta(\beta, z_i) &\xrightarrow{p} \mathrm{E}\{Z\theta(\beta)GS\} \\
\frac{1}{n} \sum_{i=1}^{n} Y_i(t) \theta(\beta, z_i) &\xrightarrow{p} \mathrm{E}\{\theta(\beta)GS\} \\
\frac{1}{n} \int_0^\tau \sum_{i=1}^{n} z_i dN_i(t) &\xrightarrow{p} \int_0^\tau \mathrm{E}\{ZgS\} dt \\
\hat{\mathrm{E}}\{Z||\theta(\beta), Y\} &\xrightarrow{p} \mathrm{E}\{Z||\theta(\beta)GS\},
\end{aligned} \tag{2.18}$$

where G, S, g are parameterized by the true β_0, and β is a placeholder for the MLE of β, the solution to the score equation. Using (2.18), after a little algebra, the large sample limit of the normalized score function $\frac{1}{n} U_n(\tau, \beta)$ (2.17) can be written as

$$U^*(\tau, \beta|\beta_0) = \int_0^\tau \mathrm{E}\{gS\} \mathrm{E}\left\{ Z \Big|\Big|_{\theta(\beta)GS}^{gS} \right\} dt, \tag{2.19}$$

where $\mathrm{E}\left\{Z\middle|\middle|_A^B\right\} = \mathrm{E}\{Z||B\} - \mathrm{E}\{Z||A\}$. Notation aside, this expression is the same as in Struthers and Kalbfleisch (1986). As noticed in Section 2.2, multiplying a condition by a non-random quantity does not change the relative expectation (2.3). Therefore under the PH model conditioning on gS and on θGS are equivalent because $\lambda = \theta h$, $g = \lambda G$, and h is non-random, and the difference $\mathrm{E}\left\{Z\middle|\middle|_{\theta(\beta)GS}^{gS}\right\}$ is uniformly zero when $\beta = \beta_0$, a reflection of the fact that the PH estimation equation is consistent (Andersen and Gill, 1982). Under a misspecified PH model, the maximum partial likelihood estimator (MPLE), $\hat{\beta}(\tau)$ solving (2.17)=0 will be consistent for $\beta^*(\tau, \beta_0)$, the solution of the score equation (2.19)=0. Generally, the MPLE estimator from the misspecified PH model, $\hat{\beta}(\tau)$, is biased $\beta^*(\tau, \beta_0) \neq \beta_0$. It is easy to verify that $\hat{\beta}$ is unbiased under the null hypothesis $\beta_0 = 0$. This follows from the fact that in this case $\theta_0 = 1$, and the conditions $gS = hFS$ and $\theta GS = FS$ are both independent of Z, and the relative expectations in the difference $\mathrm{E}\left\{Z\middle|\middle|_{\theta(\beta)GS}^{gS}\right\}$ in (2.19) are equal (at $\beta = \beta_0 = 0$) to the unconditional $\mathrm{E}\{Z\}$ making the difference zero. In other words when $\theta_0 = 1$ then $\beta = 0$ is the solution to t he score equation (2.19)=0.

Note that if $\beta_0 \neq 0$ even under exponential baseline survival β_0 still does not satisfy (2.19) because of the implicit presence of $1_{\{T>W\}}$ in h making it dependent on W through the support of survival times. Explicitly, in this case

$$g = \theta h \mathrm{E}\{I_{\{W<t\}} F^\theta | Z\} \neq \theta h G = \theta h \mathrm{E}\{F^\theta | Z\},$$

and consequently $\mathrm{E}\{Z||\theta GS\} \neq \mathrm{E}\{Z||\lambda GS\} = E\{Z||gS\}$, all because of the indicator function resulting in $G = \mathrm{E}\{F^\theta I_{\{W<t\}} | Z\} + \Pr\{W \geq t |\, Z\}$.

2.5.2 Small treatment effect approximation for point estimates

By definition the function $\beta^*(\beta_0)$ is obtained by solving $U^*(\beta|\beta_0) = 0$ with respect to β (dependence on τ is suppressed for brevity). Hence, taking derivative with respect to β_0 we have

$$\frac{\partial U^*(\beta|\beta_0)}{\partial \beta}\frac{d\beta^*(\beta_0)}{d\beta_0} + \frac{\partial U^*(\beta|\beta_0)}{\partial \beta_0}\bigg|_{\beta=\beta^*(\beta_0)} \equiv 0, \tag{2.20}$$

where $\equiv$ is uniform equality with respect to β_0. Expanding $\beta^*(\beta_0)$ around $\beta_0 = 0$ and keeping in mind that $\beta^*(0) = 0$ we get

$$\beta(\beta_0) = \beta^{*'}(0)\beta_0 + o(\beta_0), \tag{2.21}$$

where the prime stands for a partial derivative with respect to β_0. The first order approximation is

$$\beta(\beta_0) \approx \beta_0 \times \beta^{*'}(0) \stackrel{def}{=} \beta_0 \times m^*(\tau). \tag{2.22}$$

The PH multiplier, $m^*(\tau) = \beta^{*'}(\tau, 0)$, describes the departure of β^* from β_0. $m^*(\tau) = 1, m^*(\tau) > 1$, and $m^*(\tau) < 1$ indicate that the effect is unbiased, overestimated, or underestimated, respectively. Combining (2.20) and (2.21) we obtain

$$m^* = -\frac{\partial U^*(\beta|\beta_0)/\partial \beta_0}{\partial U^*(\beta|\beta_0)/\partial \beta}\bigg|_{\beta=\beta_0=0}. \tag{2.23}$$

Taking derivatives in (2.23) using (2.19), Lemma II.6 in the Appendix, and under Assumption II.1 we get after a little algebra

$$m^*(\tau) = 1 - \int_0^\tau \mathrm{E}\{fS\}\mathrm{Var}\{Z||S\}\mathrm{E}\left\{H\big|\big|_{FS}^{fS}\right\} dt \left[\int_0^\tau \mathrm{E}\{fS\}\mathrm{Var}\{Z||S\}dt\right]^{-1}. \tag{2.24}$$

Here $Var(Z)$ is thought of as the covariance matrix if Z is a vector. By Lemma II.5 and (2.38) in the Appendix

$$\mathrm{E}\left\{H\big|\big|_{FS}^{fS}\right\} = \frac{\mathrm{Cov}\{h, H||FS\}}{\mathrm{E}\{h||FS\}}$$

indicating that the behavior of m^* is also governed by the covariance of instantaneous and cumulative hazards conditional on survival up to t (compare with Section 2.4). Note that if censoring does not depend on the treatment covariates Z, S will cancel from (2.24), and the bias will not depend on the censoring distribution. The PH multiplier m^* is a multiplicative modifier of β_0. In terms of hazard ratio, we expect the limit treatment effect estimated by the PH model to be $\theta^*(\tau) \approx \theta_0^{m^*(\tau)}$. For a population under screening, the estimator obtained from the PH model is biased with respect to either the underlying treatment effect θ_0 or the actual treatment effect θ_t represented by (2.13). Given θ_0, the relationship between true and large sample quantities $h(t)$, $\lambda(t)$, θ_t, and $\theta^*(\tau)$ is displayed in Figure 2.2. The large-sample limit of PH estimated hazard ratio $\theta^*(\tau)$ depends on the distributions of the underlying survival from the clinical diagnosis, and the lead time (2.24). The PH multiplier also depends on the duration of the study.

2.5.3 Measuring the bias of variance estimated by the PH model

Using the general setting of the M-estimation (Van der Vaart (2000), Chapter 5) and assuming regularity conditions hold, we have

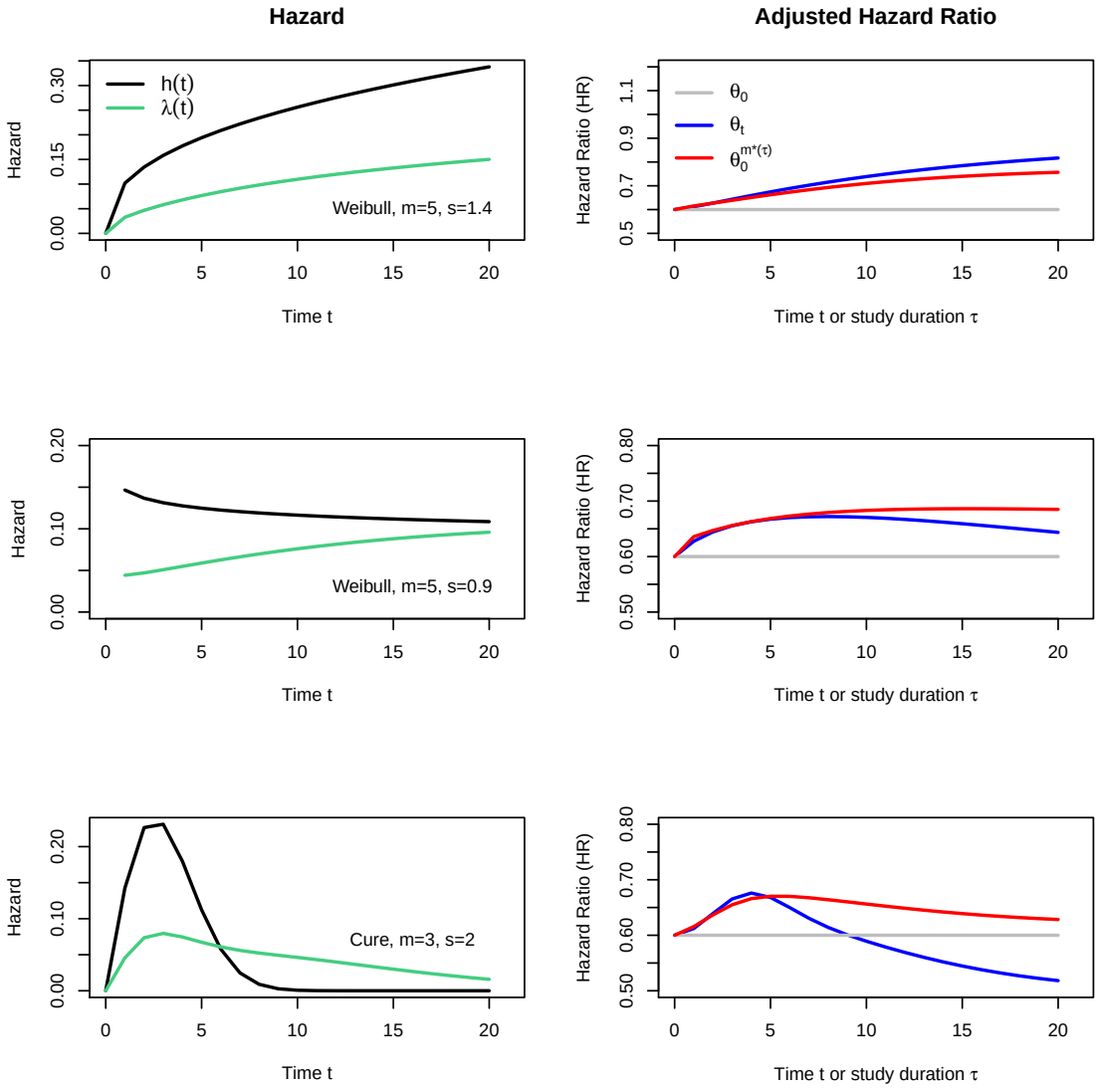
$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau, \beta_0)) \xrightarrow{d} N\left(0, \Sigma^*(\tau, \beta_0)\right),$$

where $\hat{\beta}(\tau)$ is the MLE under a misspecified PH model and

$$\Sigma^*(\tau, \beta_0) = \left[\frac{\partial^2 \mathrm{E}\{\ell\}}{\partial\beta\partial\beta^T}\right]^{-1} \mathrm{E}\left[\frac{\partial \ell_1}{\partial\beta}\frac{\partial \ell_1}{\partial\beta^T}\right]\left[\frac{\partial^2 \mathrm{E}\{\ell\}}{\partial\beta\partial\beta^T}\right]^{-1}, \tag{2.25}$$

is the covariance matrix, $\ell = \log PL$, ℓ_1 is a contribution of one observation to the likelihood, and the expression is evaluated at $\beta = \beta^*(\tau, \beta_0)$, the limit in probability of PH-model based solution. The weak convergence result stated above is valid despite the presense of 'nuisance' estimator $\hat{\mathrm{E}}\{Z|\theta(\beta), Y\}$ in the profile score function (2.17)

Figure 2.2: True and large-sample characteristics in screened and unscreened populations under various shapes of the baseline hazard. Left: The baseline hazard in the unscreened population $h(t)$ and its counterpart under screening $\lambda(t)$ averaged over the lead time W as in (2.5). Right: True hazard ratio in the unscreened population θ_0, its time-dependent counterpart under screening θ_t, and $\theta^*(\tau) \approx \theta_0^{m^*(\tau)}$ representing the large-sample limit of the PH-estimated hazard ratio under screening. τ is the duration of the study; $t \leq \tau$ is a point in follow-up time. Note that the misspecified PH model produces a biased estimate $\hat{\theta}(\tau) \xrightarrow{p} \theta^*(\tau)$ of the average true hazard ratio θ_t under screening. Computation of m^* is done under the Assumption II.1 approximation.



that can be verified directly. The information matrix is defined as

$$I(\tau) = -\frac{\partial^2 \mathrm{E}\{\ell\}}{\partial\beta\partial\beta^T}.$$

Using the corollary to Lemma II.6 in the Appendix we have

$$I = \frac{\partial U^*}{\partial\beta} = \int_0^\tau \mathrm{E}\{gS\}\mathrm{Var}\{Z||\theta(\beta)GS\}dt, \tag{2.26}$$

the information matrix estimated when the misspecified PH model is fitted to the data. Proceeding similar to Section 2.5.1, after some algebra, the variance of the individual score in the middle of the sandwich (2.25) is

$$\Sigma_1^* \stackrel{def}{=} \mathrm{E}\{U_1^* U_1^{*T}\} = \int_0^\tau \mathrm{E}\{gS\}\left[\mathrm{Var}\{Z||gS\} + \mathrm{E}^2\left\{Z\Big|\Big|_{\theta(\beta)GS}^{gS}\right\}\right], \tag{2.27}$$

where $U_1^* = \partial\ell_1/\partial\beta$. If the model is correctly specified, $\Sigma^*(\tau,\beta) = I^{-1}(\tau)$ because $I = \Sigma_1^*$. Indeed, similar to the discussion after the score equation (2.17), under the PH model or under the null hypothesis, the relative expectation difference term in the right part of (2.27) is zero, and relative variances in (2.27) and (2.26) are equal. However, generally the correct variance of the misspecified model is given by $I^{-1}\Sigma_1^* I^{-1} \neq I^{-1}$, and the variance reported as $I^{-1}(\tau)$ using the PH model is incorrect.

2.5.4 Small treatment effect approximation for the variance

Now, consider the variance (2.25) under the small treatment effect Assumption II.1. Using the lemmas presented in the Appendix we have the expansions

$$\begin{aligned}
\mathrm{E}\{gS\} &= \mathrm{E}\{fS\} + \mathrm{E}\{ZfS\}[1 - \mathrm{E}\{H||f\}]\beta_0 + o(\beta_0) \\
\mathrm{Var}\{Z||gS\} &= \mathrm{Var}\{Z||S\} + [1 - \mathrm{E}\{H||f\}]M_3\{Z||S\}\beta_0 + o(\beta_0) \\
\mathrm{E}^2\left\{Z\Big|\Big|_{\theta(\beta)GS}^{gS}\right\} &= o(\beta_0) \qquad (2.28) \\
\mathrm{Var}\{Z||\theta SG\} &= \mathrm{Var}\{Z||S\} + [m^* - \mathrm{E}\{H||f\}]M_3\{Z||S\}\beta_0 + o(\beta_0),
\end{aligned}$$

where the multiplier m^* is given by (2.24), and M_k is a central moment of order k (2.39).

Define

$$
\begin{aligned}
I_0 &= \int_0^\tau \mathrm{E}\{fS\}\mathrm{Var}\{Z||S\}dt, \\
I_1 &= \int_0^\tau M_3\{Z||S\}\mathrm{E}\{fS\}[1-\mathrm{E}\{H||f\}]dt, \\
I_2 &= \int_0^\tau \mathrm{E}\{ZfS\}\mathrm{Var}\{Z||S\}[1-\mathrm{E}\{H||f\}]dt, \\
I_3 &= \int_0^\tau M_3\{Z||S\}\mathrm{E}\{fS\}\mathrm{E}\left\{H\Big|\Big|_F^f\right\}dt, \\
I_4 &= \int_0^\tau \mathrm{E}\{fS\}\mathrm{Var}\{Z||S\}\mathrm{E}\left\{H\Big|\Big|_F^f\right\}dt. \\
I_5 &= \int_0^\tau M_3\{Z||S\}\mathrm{E}\{fS\}dt,
\end{aligned}
\tag{2.29}
$$

The after some algebra we get the expansion for the variance in the form

$$
\Sigma^* = I_0^{-1} - I_0^{-1}[I_1 + I_2 + 2I_3 - 2I_5 I_4 I_0^{-1}]I_0^{-1}\beta_0 + o(\beta_0). \tag{2.30}
$$

Note that for the PH model, a similar but simpler exercise gives the following variance approximation

$$
\Sigma^* = I_0^{-1} - I_0^{-1}[I_1 + I_2]I_0^{-1}\beta_0 + o(\beta_0). \tag{2.31}
$$

There are at least two scenarios when the PH variance is unbiased. Naturally, if the true model is PH (no misspecification), then $I_3 = I_4 = 0$, and (2.30) and (2.31) are equivalent. A non-trivial fact, however, is that when Z is not skewed, i.e. when $M_3\{Z||S\} = 0$, then $I_1 = I_3 = I_5 = 0$, (2.30) and (2.31) are also equivalent and equal to

$$
I_0^{-1} - I_0^{-1} I_2 I_0^{-1}\beta_0 + o(\beta_0). \tag{2.32}
$$

Also, when censoring is independent of the covariates, it is easy to see that (2.31)

and (2.30) are the same so that the PH model based variance is correct up to the first order of β.

2.6 The meta-analytic correction

We have shown that estimators in the PH model are generally biased under misspecification. To correct the bias, we consider a meta-analytic framework to restore the adjusted hazard ratio θ_0 from the results of reported studies without access to raw patient-level data. Let $\hat{s}_i^2$ be the sample variance of the estimated treatment effect coefficients, $\hat{\beta}_i$, in study i and ω_i be the weight assigned to that study where $\omega_i = 1/\hat{s}_i^2$, i=1 $\cdots$ k. If the distribution of the treatment covariate is not symmetric, the reported sample variance needs to be corrected as well as discussed in the sections 2.5.3, 2.5.4. The meta-analytic estimate $\hat{\beta}$ is obtained as a weighted average of study-specific $\hat{\beta}_i$ with weights inversely proportional to the sample variance of the estimates. The combined point estimator $\hat{\beta}$ and variance $\hat{\sigma}^2/n$ is

$$\begin{aligned} \hat{\beta} &= \frac{\sum \omega_i \hat{\beta}_i}{\sum \omega_i} \\ \frac{\hat{\sigma}^2}{n} &= \frac{1}{\sum \omega_i}. \end{aligned} \tag{2.33}$$

Since the MPLE estimator from the PH model, $\hat{\beta}$, is a consistent estimator of $\beta^*(\tau, \beta_0)$, approximately, by the Continuous Mapping Theorem, $\hat{\theta}_0(\tau) = \exp([\hat{\beta}/m^*(\tau)]z)$ is also a consistent estimator for θ_0 (subject to approximating Assumptions II.1). Let σ^2 denote the variance of $\hat{\beta}$. By the Delta method, we will have

$$\begin{aligned} &\sqrt{n}(\hat{\beta}_0 - \beta^*(\tau, \beta_0)) \xrightarrow{d} N\left(0, \sigma^2\right) \\ \Rightarrow &\sqrt{n}(\hat{\theta}_0(\tau) - \theta_0) \xrightarrow{d} N\left(0, \left[\frac{1}{m^*(\tau)} e^{\frac{\beta^*}{m^*(\tau)}}\right]^2 \sigma^2\right) \end{aligned}$$

The 95% confidence interval for θ_0 is

$$\hat{\theta}_0(\tau) \pm 1.96 \left[\frac{1}{m^*(\tau)} e^{\frac{\hat{\beta}}{m^*(\tau)}}\right] \frac{\hat{\sigma}}{\sqrt{n}}. \tag{2.34}$$

Alternatively, one could exercise the correction of each study first, and then provide a meta-analytic estimate based on the corrected individual study estimates. This is preferable if studies being combined have different durations.

2.7 Simulation study

We used simulations to study the accuracy of the PH multiplier under small treatment effect assumption. Sensitivity analyses were also conducted to evaluate how the shape of the baseline hazard and the magnitude of the treatment effect affect the estimated hazard ratios.

We adopted a 2×2 design to estimate the hazard ratio of the treatment group versus the control group using the PH model for both the unscreened and the screened populations. For the unscreened population, the survival time was calculated from the time of clinical diagnosis to time of death. For the screened population, a random lead time W was generated using the distribution displayed in Figure 2.1 and was added to the survival time in agreement with the specific convolution model (2.6). Under the PH assumption, a pre-specified treatment effect $\theta_0 = \exp(\beta_0)$ was applied to treatment group characterized by the $z = 1$ value of the dummy variable. Survival times from the clinical diagnosis were drawn from Weibull distributions for increasing and decreasing baseline hazards, or from a cure model distribution for a changing baseline hazard. Survival functions of Weibull and cure model distributions were parameterized using median, m, and shape, s, parameters, and the treatment effect θ_0, as defined in the following equation

$$\begin{aligned} G_{weibull}(t|\theta_0) &= \exp\left\{-\log(2)\theta_0\left(\frac{t}{m}\right)^s\right\} \\ G_{cure}(t|\theta_0) &= \exp\left\{-\theta_0\left(1-\exp\left\{-\log(2)\left(\frac{t}{m}\right)^s\right\}\right)\right\}. \end{aligned} \tag{2.35}$$

Hazard Ratios were estimated using the PH model. Subjects with survival times longer than τ were censored. Adjusted hazard ratios were calculated using the PH multiplier defined in (2.24) and were compared with the true result. Two hypothetical treatment groups with the sample size of 500 each were used. For one replicate survival data in the treatment and control group was generated under no screening and under screening. Hazard ratios θ_0 (no screening) and $\theta^*(\tau)$ (under screening) were estimated by the PH model applied to both sets of the data. To obtain an approximately unbiased estimate of θ_0 under a misspecified model, the log hazard ratio estimate under screening $\hat{\theta}(\tau)$ was adjusted by dividing it by the small-sample approximated multiplier $m^*(\tau)$ given by (2.24). The r.v. Z was taken to represent treatment assignment in a 1:1 ratio by simple randomization, a Bernoulli(0.5) r.v. Censoring was assumed to be independent of Z so S cancels from (2.24). Baseline distribution characteristics H and f were assumed to be known externally. The experiment was repeated for various study durations τ. Shown in Figure 2.3 are the results of a study to assess the quality of the proposed adjustment based on the small treatment effect approximation. The left part of the figure shows three scenarios of varying shape of the baseline hazard function without screening vs. under screening (averaged over W). The right part of the figure shows hazard ratios estimated or predicted under different study durations represented by the x-axis. The wiggly polygon curves in the right part of the figure give estimated hazard ratios without screening (the bottom curve), and the one under screening showing a departure upwards. The smooth curves (a line at $y = \theta_0$ in the case of no screening) going through the polygons represent an average or a large sample limit of the respective PH estimate. In the case of screening, the latter is given by the exponentiated solution of (2.19)=0 with respect to β under the true β_0, H, S. The top dashed curve

Table 2.1: Empirical vs. approximate variance ($\theta_0 = 0.5$, τ=15, n=10000)

Proportion of the treatment group	0.3	0.5
	Simulation result ($\times 10^{-4}$)	Simulation result ($\times 10^{-4}$)
Increasing $h(t)$, Weibull m=5, s=1.5		
Naïve variance	7.45	6.04
Sandwich variance	7.42	6.00
Approximated variance	6.93	5.68
Decreasing $h(t)$, Weibull m=5, s=0.9		
Naïve variance	9.62	7.49
Sandwich variance	9.61	7.48
Approximated variance	9.31	7.50
Non-monotonic $h(t)$, Cure m=3, s=2		
Naïve variance	12.52	9.65
Sandwich variance	12.55	9.68
Approximated variance	11.30	8.90

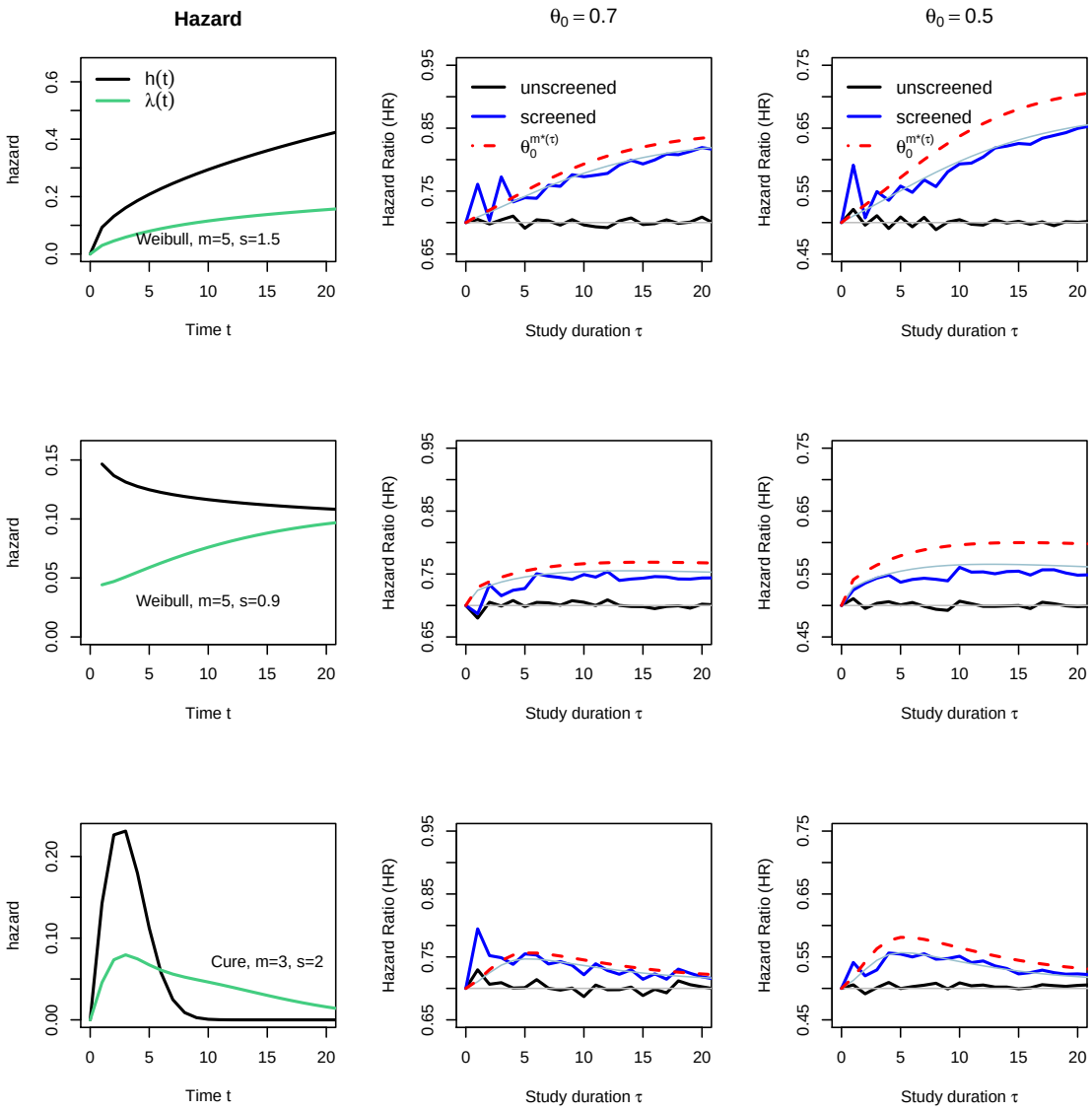
shows hazard ratio under screening predicted as $\theta_0^{m^*(\tau)}$ with the multiplier $m^*(\tau)$ obtained from the approximation (2.24). It is clear from Figure 2.3 that the bias correction using the approximation to the multiplier is reasonably accurate except perhaps in the case of large treatment effects (small θ_0).

We also conducted a simulation study to compare the difference between the corrected variance, Σ^*, and the variance reported by the PH model using the inverse of information matrix. Having generated a large sample of 10000 we found that when the treatment is effective ($i.e.\, \theta_0 < 1$), the difference between the adjusted and naive variance is quite small ($< 1\%$) regardless of the treatment allocation ratio (table 2.1).

2.8 Example: Radical prostatectomy vs. watchful waiting for prostate cancer

Radical prostatectomy is an invasive surgical procedure to remove the prostate gland. It is one of the most common treatments for patients with localized prostate

Figure 2.3: Quality of the small treatment effect approximation for the misclassification bias. Left:Shapes of the baseline hazard function without screening h vs. under screening λ. Middle and right column: Variable polygons and smooth curves going through them correspond to PH model estimated and average hazard ratios at various study durations τ. The lines at $y = \theta_0$ correspond to a simulation without screening ($W = 0$ with probability 1). The dashed curve showing a departure upwards particularly with small θ_0 (large beneficial treatment effect) is the hazard ratio under screening predicted using the approximation (2.24) for the multiplier.



cancer. Three studies, one from Europe (Bill-Axelson et al., 2005) and two from the United States (Tewari et al., 2006; Albertsen et al., 2007), have looked at large groups of localized prostate cancer patients. All patients in these three studies were newly diagnosed and many of them must have been subjected to the Prostate-Specific Antigen (PSA) tests, especially those diagnosed in the United State after 1990. Relative effects of radical prostatectomy vs. watchful waiting (conservative management) on the disease-specific survival were estimated using PH models. Assuming that the study cohorts are similar to patients in the Surveillance, Epidemiology, and End Results (SEER) registry data, and that the recruitment is following the pattern of cancer incidence in the SEER population, we recovered the underlying treatment effects θ_0 for radical prostatectomy vs. watchful waiting using the methods described in Section 2.6. Distribution of the lead time was obtained from the marginal incidence model (Tsodikov et al., 2006) fitted to SEER data, and assumed known. The baseline hazard was obtained from a PH analysis of SEER data before the year of 1988 when PSA was introduced, and assumed known for prediction of the multiplier. The lead time and survival distributions depend on age, A, and year of diagnosis, Y. Therefore we incorporated age and calendar time as covariates into the lead time and into the baseline survival characteristics essentially by tabulating their distributions for all combinations of A and Y. We then exploit the fact that we kept the development general with respect to how $h(t|W)$ depends on W. Therefore, we can redefine the vector W to include A, Y in addition to the lead time. Essentially this means that all expectations over W turn into expectations over W, A, Y, and the key expressions of the paper are valid with this understanding. These expectations were taken with respect to the lead time conditional on A, Y, and then over the empirical distribution of A, Y specified using SEER data and the calendar period of

Table 2.2: Estimated and predicted relative risks of radical prostatectomy vs. watchful waiting

Study	Reported RR(95%C.I.) from PH model	RR(95%C.I.) after correction
Bill-Axelson et al. (2005, NEJM)	0.56 (0.36-0.88)	0.51 (0.24-0.77)
Tewari et al. (2006, J. Urology)	0.37 (0.25-0.55)	0.35 (0.21-0.50)
Albertsen et al. (2007, J. Urology)	0.29 (0.16-0.52)	0.23 (0.07-0.39)
Combined result from meta-analysis	0.41 (0.31-0.53)	0.37 (0.26-0.51)

the respective study. The results of the predicted true underlying treatment effects from all the three studies as well as the combined estimate using meta-analysis are summarized in table 2.2.

This analysis indicates under-estimated treatment effects by major clinical studies. However, the correction appears to be relatively modest, 6-26% relatively to the corrected one. Note that the corrected hazard ratio refers to the survival time from clinical diagnosis (the point of diagnosis by symptoms), for which there is no direct data if the patient is screen detected while asymptomatic.

2.9 Discussion

When screening is operating in the population targeted by clinical trials, the PH model is misspecified due to the fact that the time of diagnosis is advanced by a random and unobserved amount, the lead time, as patients are detected before they develop symptoms of the disease. This type of misspecification is special because patients do not die before they develop symptoms (during the lead time) meaning that the support of the survival time is random.

While early detection might enhance the treatment effect (i.e. interact with treatment), we have conservatively assumed that screening is of no real benefit, and we asked the narrow question of what the implications of misspecification might be in this situation and how they can be reversed or at least reduced without access to

raw data.

We have shown that even with proper randomization, evaluating the treatment effect in clinical trials recruiting from a population under screening could result in biased estimates. The bias is a function of the duration of the study, the shape of baseline hazard, the lead time distribution in the population under screening, the distribution of time to censoring if it varies by treatment group, and the size of the true treatment effect.

Unlike the bias from omitting important covariates in the PH model, the bias from ignoring early detection is not always conservative. To correct the treatment effect estimated by the misspecified PH model, we propose a meta-analytic framework based on a study of bias and variance of a general misspecified PH model. To be able to provide a correction in the absence of raw data we entertained a small treatment effect approximation that simplified the formulas and reduced the complexity of information needed to launch the correction. Simulation results suggest the correction we proposed is robust and accurate in realistic scenarios of modest treatment effects.

We also found that the PH model is correct at estimating the variance if the distribution of treatment covariates is not skewed. At the same time under no circumstances other than absence of screening can the PH model point estimate of the treatment effect be unbiased.

This study also shows that a correction of the treatment effects observed in the screening era is possible using population models of cancer that provide an estimate of the distribution of the lead time. This distribution was assumed to be known throughout the paper. This is a natural assumption since population models are fitted to big populations while the survival study of this paper concerns much smaller

groups of cancer patients only.

A number of simplifying assumptions were unavoidable. We disregard stage progression between screening and clinical diagnosis in a small fraction of patients. We did not incorporate the interaction between treatment and early detection resulting in the perceived benefit of screening to the population. Last but not least, as we focused on the lead time bias, the length bias for screen-detected cancers and anti-length bias for cancers missed by screening were ignored. The latter can be recognized from (2.6) that implies, for example, that $h(t|W = 0)$ apply equally to a patient from an unscreened population or to a patient from a population under screening who is detected due to symptoms with zero lead time. Length bias exhibits itself in that screening is more likely to catch slower growing tumors that spend longer times without symptoms. As a result tumors detected by screening have better prognosis than tumors missed by screening. For this reason a patient with $W = 0$ from an unscreened population generally has better prognosis than a screened patient with $W = 0$ whose cancer was missed by screening. This effect can be incorporated by allowing more complicated forms of $h(t|W)$ such as $h(t - W, W)$, where the second W models dependence on the lead time on top of the convolution effect. The same recipe could be used to generally model baseline survival post clinical diagnosis as being correlated with the lead time. Note that we intentionally kept the development of the theoretical sections of the paper general with respect to the form of $h(t|W)$ to incorporate such scenarios.

To avoid massive over-treatment of cancer patients, screen-detected patients with good prognosis may be placed on deferred treatment regimen. Screening and treatment would then eventually represent a dynamic cancer control strategy designed to preserve the mortality benefit while reducing over-treatment. Joint dynamic models

need to be developed to analyze the combined screening and treatment interventions.

2.10 Appendix

2.10.1 Properties of relative expectation

Lemma II.4. *Let μ, ν be some functions of random variables independent of a parameter θ. Then*

$$\frac{\partial}{\partial\theta}\log \mathrm{E}\big\{\mu\big|\big|\nu^{\theta}\big\} = \frac{\mathrm{Cov}\{\mu,\log\nu||\nu^{\theta}\}}{\mathrm{E}\{\mu||\nu^{\theta}\}}. \tag{2.36}$$

Proof proceeds straightforwardly by differentiation and using the definition (2.3).

Lemma II.5. *Let μ, ν, and ξ be some functions of random variables. Define*

$$\mathrm{E}\left\{\mu\right|\big|_{\xi}^{\nu}\left.\right\} = E\{\mu||\nu\} - \mathrm{E}\{\mu||\xi\}. \tag{2.37}$$

Then

$$\mathrm{E}\left\{\mu\right|\big|_{\nu\xi}^{\nu}\left.\right\} = -\frac{\mathrm{Cov}\{\mu,\xi||\nu\}}{\mathrm{E}\{\xi||\nu\}}. \tag{2.38}$$

Proof. Using the definition (2.3) of the relative expectation we get the left part of (2.38) as a difference of two fractions. Bringing them to a common denominator and gividing numerator and denominator by $\mathrm{E}(\nu)$, and again using (2.3), we arrive at the right part of (2.38).

Lemma II.6. *Let μ be some function of random variables, Z be some random variable, and $\theta = e^{\beta Z}$. Define the central relative moment of Z of kth order as*

$$M_k(Z||\mu) = \mathrm{E}\left\{\left[Z - \mathrm{E}\{Z||\mu\}\right]^k\right|\big|\mu\big\} \tag{2.39}$$

Then if μ does not depend on β

$$\frac{\partial^k}{\partial\beta^k}\mathrm{E}\{Z||\theta\mu\} = M_{k+1}(Z||\theta\mu). \tag{2.40}$$

Proof. The proof proceeds by induction, uses straightforward differentiation and the fact that $M_1(Z||\theta\mu) = 0$. As a corollary, we have

$$\frac{\partial}{\partial\beta}\mathrm{E}\{Z||\theta\mu\} = \mathrm{Var}(Z||\theta\mu).$$

Lemma II.7. *Let $\mu = \mu(\beta_0, Z, W)$, where Z, W are r.v.s, and β_0 is a parameter. Then*

$$\frac{\partial}{\partial\beta_0}\mathrm{E}\{Z||\mu\} = \mathrm{Cov}\left\{Z, \frac{\partial\log(\mu)}{\partial\beta_0}\middle||\mu\right\}. \tag{2.41}$$

Proof by straightforward differentiation.

CHAPTER III

Predicting Cancer Progression and the Null Hypothesis of Treatment Effect under Screening

3.1 Introduction

Deciding whether, or how, to treat cancer for a newly diagnosed prostate cancer patient is difficult because a large fraction of patients are over-diagnosed and if left untreated would never die from the disease. Quantifying potential risks of the disease progression would provide valuable information to help patients and doctors make informed decisions to manage cancer. With screening in place, patients are detected earlier in a less advanced stage showing longer survival from the point of diagnosis even in the absence of any treatment benefit, which greatly complicates treatment decisions. Because the risk of cancer detection is correlated with the latent cancer growth process, patient heterogeneity at diagnosis varies dependent on the utilization of screening in the population from which the patient is sampled. Thus, the construction of a disease progression prognosis for the patient depends on joint modeling of the cancer development and heterogeneity in the population and the subject-specific risk of cancer progression within the patient given the information on his latent heterogeneity available through clinical and demographic characteristics observed at diagnosis.

Since the introduction of PSA test in 1988, the wide spread use of screening

programs resulted in profound dynamics of cancer incidence and its presentation at diagnosis. In the PSA era survival post-diagnosis has "improved" counteracting the effect of increased incidence (Nicholson and Harland, 2002). The severity of the disease at diagnosis (stage and grade) has also enjoyed a favorable shift. Due to overdiagnosis there is an increase in the probability of "cure" in prostate cancer (Draisma et al., 2003; Tsodikov et al., 2006; Draisma et al., 2009). Neither lead-time nor overdiagnosis can be directly observed so their estimation requires modeling. Parker et al. (2006) developed a competing-risks model to estimate the prostate cancer survival for screen-detected prostate cancer. While the model by Parker et al. (2006) adjusts for the lead-time, it conditions on stage and grade at diagnosis. Since stage and grade at diagnosis are themselves affected by screening, it is difficult to generalize the results to populations with different screening utilization patterns. Also, the model does not incorporate a progression mechanism and cannot predict the chance that cancer will become metastatic or high-grade if left untreated before symptoms appear, a piece of information important for treatment decisions.

The general problem that has not been addressed is that of the null hypothesis of treatment effect when patients are recruited from populations under variable patterns of screening. Consider the same patient run through two scenarios: screening and no screening under the hypothesis of no treatment benefit. Despite being under the null hypothesis we will have different survival post-diagnosis, different stage and grade of the disease at diagnosis in the two scenarios, and it is not trivial to formulate the treatment effect in this setting. However, because we are picturing the same patient running through both scenarios, and because screening is not curative without the treatment effect, the age at which symptoms appear (CDx) and survival past that point is the same in both runs. For a screen-detected patient, at the

point of observed screening diagnosis (SDx), the future cancer history resulting in the onset of symptoms, CDx, and a possibly more advanced stage and grade at that point, represents a multivariate random outcome. This outcome is the same under the null hypothesis of treatment effect or under the assumption that patient is not treated and is blinded to the fact of cancer diagnosis that we call the counterfactual ignored screening scenario (iS). Needless to say, the iS scenario is never observed in practice. Note that patients on conservative management or active surveillance are not equivalent to iS because of self-selection, subsequent monitoring, and medicine they still receive while on the regimen. It is clear that the null hypothesis of treatment effect is that of the same joint distribution of age, stage and grade at CDx, and same survival post CDx where CDx is understood as the point of observed diagnosis for the patient detected by symptoms, and a counterfactual CDx under the iS scenario for the screen-detected patient. We will give a rigorous definition later.

To address the problems mentioned above we propose a joint model for the disease presentation at SDx and real or counterfactual CDx. We build upon the "marginal" stage and grade specific incidence model describing disease presentation at the observed point of diagnosis (either CDx or SDx, but not both on the same patient).

We will use our model to describe prostate cancer related risks for the general US male population and the potential adverse events for a PSA-detected man given information at the time of diagnosis. The details of the model are discussed in section 3.2 and the expressions for key natural history events are explained in section 3.3. Finally, we provide predictions for natural history events using population data and the results are shown in section 3.4.

3.2 Natural history model

The natural history model is based on the classical three-state chronic disease model. The disease progresses through three states, disease-free state, pre-clinical state, and clinical state. The disease is chronic and the transitions are irreversible. On top of that there is a finer characterization of cancer progression through stage (Localized vs. Metastatic) and grade (Low vs. High). The model is partially specified and does not assume any specific mechanism of progression through stage and grade.

The duration of the disease-free state is represented by the age at tumor onset, described by the random variable Y. We assume there is no prostate cancer before the age of 50, so the age origin in the model resides at 50.

The pre-clinical state in the absence of screening measures the time period between tumor onset and detection via clinical symptoms appearing (clinical diagnosis, CDx). This duration is called the sojourn time. In the presence of screening, cancer may be detected while it is still asymptomatic by the screening test (screening diagnosis, SDx). Time to the two types of diagnosis T_{CDx} and T_{SDx}, respectively, represent competing risks originating at the age of Y. The clinical state describes the survival time T_s from the time of diagnosis (CDx or SDx) to the time of cancer-specific death. The structure of the chronic disease three-state model are shown in figure 3.1. Our approach is to build a series of hierarchical models to describe prostate cancer incidence, presentation at diagnosis (Z = Stage and Grade), progression, and survival.

3.2.1 Marginal incidence model

The marginal incidence model describes the risk of being diagnosed with prostate cancer at age a_I either clinically (CDx) or by PSA screening tests (SDx). In our

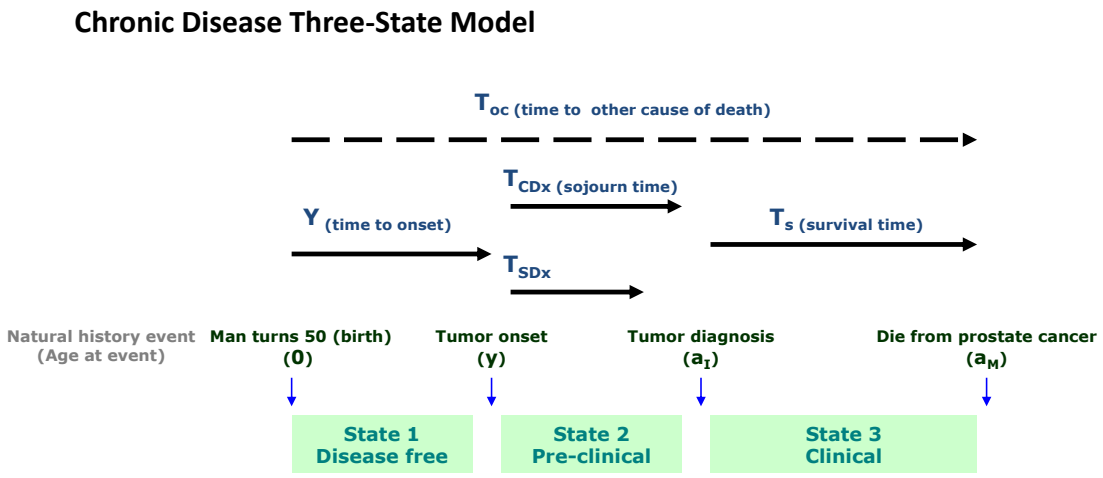


Figure 3.1: Structure of the chronic disease three-state model

previous paper, the prostate cancer incidence for x birth cohort (a cohort of men turning 50 in year x) can be written as a complex mixture model (Tsodikov et al., 2006) where missing data include the age at onset y, the screening schedule, and the detection process. The p.d.f. of cancer diagnosis given birth year x can be written as

$$f_I(a_I|x) = \int_0^{a_I} f_o(y|x) f_I(a_I - y|x, y)\, \mathrm{d}t \tag{3.1}$$

where f_o is the unconditional pdf of age at tumor onset, and $f_I(a_I - y|x, y)$ is the pdf of age at cancer diagnosis given birth year x and tumor onset time y. Under competing risks, $f_I(a_I - y|x, y)$ can be split into two crude densities,

$$f_I(a_I - y|x, y) = f^c_{CDx}(a_I - y|x, y) + f^c_{SDx}(a_I - y|x, y)$$

where

$$f^c_{CDx}(a_I - y|x, y) = f_{CDx}(a_I - y|x, y) G_{SDx}(a_I - y|x, y)$$

and

$$f^c_{SDx}(a_I - y|x, y) = f_{SDx}(a_I - y|x, y) G_{CDx}(a_I - y|x, y).$$

Here $f_{CDx}(a_I - y|x, y)$ and $G_{CDx}(a_I - y|x, y)$ are the p.d.f. and survival function (s.f.) of T_{CDx}, the sojourn time distribution, and $f_{SDx}(a_I - y|x, y)$ and $G_{SDx}(a_I - y|x, y)$ are the p.d.f and s.f. of T_{SDx}, time to PSA diagnosis, respectively computed as an average over the point process of screening schedule and the outcomes of screening tests. Distributional characteristics of the schedule process (intensity of utilization of screening in the population) was estimated in Mariotto et al. (2007). Integrating out the age at tumor onset y, we have the unconditional p.d.f. of age at diagnosis a_I

$$f_I(a_I|x) = f^c_{CDx}(a_I|x) + f^c_{SDx}(a_I|x).$$

Note that the unconditional p.d.f $f_{CDx}(a_I|x)$ is a function of the age of the subject while the conditional p.d.f $f_{CDx}(a_I - y|x, y)$ is a function of the delay time defined as $t_D = a_I - y$. Detailed expressions can be found in Tsodikov et al. (2006).

3.2.2 Z-specific incidence model

We use r.v. Z to denote four possible combinations of binary stage and grade classifications (Localized/Reginal Stage, Low Grade)=LL, (Localized/Reginal Stage, High Grade)=LH, (Distant Stage, Low Grade)=DL, (Distant Stage, High Grade)=DH. Disease stage is dichotomized into the local-regional (LR) stage and distant (D) stage. Disease grade is divided into well or moderate (WM) differentiated (low grade) and poorly differentiated or undifferentiated (PU) disease (high grade). Generally, Z can be any multivariate mark on the cancer incidence process. The Z-specific incidence model describes the probability of being diagnosed with prostate cancer at a certain age and with specific stage and grade z. Conditional on birth year x, age of tumor onset y, and age of tumor diagnosis a_I, the probability of being diagnosed with stage and grade z was modeled using mixed multinomial logit model (Chefo and Tsodikov, 2009). Missing data include the delay time between tumor onset and diagnosis (a backward recurrence time) and the mode of diagnosis. Calendar time and age are treated as fixed effects covariates. Using the Z-specific incidence model we predict the conditional (multinomial) distribution of z, $f_I(z|x, a_I)$ given the cohort x, incident age a_I. Note that $f_I(z|x, a_I)$ is a conditional average over and tumor onset age Y, and the mode of diagnosis I_{Scr} =1 if SDx and =0 if CDx. Using the model we can update the distribution of Y given the information available at diagnosis.

The conditional distribution of the age of tumor onset Y can be written as

$$Y \sim f_I(y|x, a_I, z) = \frac{f_I(a_I, z, y|x)}{f_I(a_I, z|x)} = \frac{f_I(z|x, a_I, y) f_I(a_I|x, y) f_o(y|x)}{\int_0^{a_I} f_I(z|x, a_I, y) f_I(a_I|x, y) f_o(y|x)\, \mathrm{d}y}.$$

The conditional distribution of stage and grade z at diagnosis can be written as

$$f_I(z|x, a_I) = \frac{f_I(a_I, z|x)}{f_I(a_I|x)} = \frac{\int_0^{a_I} f_I(z|x, a_I, y) f_I(a_I|x, y) f_o(y|x)\, \mathrm{d}y}{f_I(a_I|x)}.$$

The Z-specific incidence is given by

$$\lambda_I(a_I, z|x) = \lambda_I(a_I|x) f_I(z|x, a_I), \tag{3.2}$$

where $\lambda_I(a_I|x)$ is the marginal incidence, and $f_I(z|x, a_I)$ serves as a factor partitioning it into the z-specific components.

3.2.3 Lead-time

Lead-time measures the amount of time the point of diagnosis is advanced due to screening. It represents the period between SDx and real or counterfactual CDx. For a clinically detected patient, the lead-time is zero, and its distribution has a mass at zero reflecting the proportion of CDx among all diagnoses. It is an important factor in cancer survival presenting a guaranteed survival benefit (patients cannot die before symptoms appear). For an incident patient with characteristics (x, a_I, z), the updated distribution of lead-time is

$$f_{LT}(s|x, a_I, z) = \frac{f_{LT}(s, a_I, z|x)}{f_I(a_I, z|x)},$$

where

(3.3)

$$f_{LT}(s, a_I, z|x) = \int_0^{a_I} f_o(y|x) f_I(z|x, a_I, y) \begin{cases} G_{SDx}(a_I - y|x, y), & (s = 0) \\ f_{SDx}(a_I - y|x, y), & (s > 0) \end{cases} f_{CDx}(a_I - y + s|x, y)\, \mathrm{d}y,$$

$$f_I(a_I, z|x) = \int_0^{a_I} f_I(z|x, a_I, y) f_I(a_I|x, y) f_o(y|x)\, \mathrm{d}y.$$

The expressions above integrate the subject's history over the unobserved age at onset and mode of diagnosis. The history includes onset, observed diagnosis SDx or CDx, presentation z at observed diagnosis, and counterfactual CDx if observed diagnosis is SDx). Lead-time equal to 0 implies that the person had a CDx while lead-time > 0 means that the person had an SDx.

3.2.4 Disease progression model and the null hypothesis of treatment effect

The disease progression model estimates the probability of disease progression during the lead-time in the absence of treatment. Let the vector

$$V_{I_{Scr}} = (a_{I_{Scr}}, z_{I_{Scr}})$$

be the disease presentation at diagnosis indexed by the mode of diagnosis. For a screen-detected patient we have V_1 at the observed SDx and V_0 at the counterfactual CDx. Note that as a result of the Z-specific incidence model (3.2) we have a model for the marginal distribution of V when the mode of diagnosis I_{Dx} is random (unobserved).

Disease progression between SDx and CDx can be characterized by the transitional distribution $f_V(V_0|V_1, x)$ describing the p.d.f. of the disease presentation at counterfactual CDx (V_0) conditional on the observed presentation V_1 at SDx and the

birth cohort x. We can expand f_V as

$$f_V(V_0|V_1,x) = f_{LT}(t_{LT}|a_1,z_1,x,t_{LT}>0)p_b(z_0|z_1,t_{LT}), \tag{3.4}$$

where the lead-time t_{LT} for the disease detected by the test

$$f_{LT}(t_{LT}|\cdot,t_{LT}>0) = \frac{f_{LT}(t_{LT}|\cdot)}{1-f_{LT}(0|\cdot)}$$

is conditional on SDx that is equivalent to a positive lead-time, $t_{LT} > 0$. Here $p_b(z_0|z_1,t_{LT})$ are the baseline progression probabilities. While generally, p_b may depend on the lead-time, in the data analysis example we assume they are independent of t_{LT}. This gives a set of unknown parameters $p_b(z_0 = j|z_1 = i) = p_{bij}$, where $i \le j$, $i,j = 1,\ldots,4$ go over the four categories of stage and grade z, that is summarized as a progression probability matrix (PPM) in (Table 3.1). The fact that $i \le j$ reflects the assumption that cancer cannot regress.

The main difficulty in estimating the PPM is rooted in the fact that CDx and SDx are not observed on the same subject. So there is no direct subject-specific data on the disease progression.

To estimate the PPM, we first formulate the null hypothesis of treatment effect. Under the null hypothesis of treatment effect the baseline PPM probabilities p_b are not affected by treatment applied at the point of SDx. If treatment had an effect, the baseline probabilities p_bs would be transformed by a categorical regression model with treatment as a covariate and p_b corresponding to the baseline of no treatment. In the extreme, treatment applied at the point of SDx may completely prevent cancer progression in which case PPM would be an identity matrix (stage and grade are frozen at SDx and carry over unchanged to the point CDx). This introduces the so-called stage-shift resulting in distant stages being prevented by screening when cancer is detected while it is still localized and progression is arrested by treatment. It is this

stage shift assumption that is at the root of the mortality benefit of screening (and treatment) assumed in many models of cancer mortality in the presence of screening. Now consider two model predictions in the following two counterfactual scenarios:

1. z-specific Incidence $\lambda_I(a, z|\neg S)$ under no screening ($\neg S$, zero screening sensitivity); and

2. the model predicted z-specific incidence $\lambda_I(a, z|iS)$ as if screening were ignored (iS) and the patient was left undiagnosed until his lead-time expired.

While the first scenario does not involve the PPM and is expressed by the z-specific incidence model prediction (3.2) under zero screening sensitivity, the second counterfactual scenario uses PPM to predict stage and grade at the end of the lead-time.

The absence of the stage-shift is expressed as the equality

$$\lambda_I(a, z|\neg S) \equiv \lambda_I(a, z|iS), \tag{3.5}$$

where $\equiv$ denotes a uniform equality over a, z represents the first part of the null hypothesis of the treatment effect. The equality (3.5) represents the first part of the null hypothesis of treatment effect where $\lambda_I(a, z|iS)$ is computed using the baseline probabilities p_b. In the case of more general progression models, not necessarily formulated in terms of PPM, (3.5) still represents the first part of the null hypothesis expressing the general equivalence of ignored screening iS, zero screening sensitivity, and zero treatment effect. In other words there is no difference between no screening, screening with zero sensitivity, or screening combined with ineffective treatment as far as the disease presentation at real or counterfactual CDx goes.

The second part of the null hypothesis is the similar equality for cancer mortality that remains beyond the scope of the present paper.

To estimate the PPM we treat (3.5) as an equation for the unknown p_b parameters. The idea is to deduce the disease progression model from the marginal one by making the two counterfactual predictions as close as possible. The target function to be minimized for the estimate of the disease progression model, l, can be written as the Poisson likelihood "distance" (3.6) between the two predictions by treating one of them as "observed" data ($^{\neg}S$), and the other as expected (iS).

Conditional on the birth year, x:

$$l = \sum_a \sum_t \sum_z P(a,t,z) \left\{ \lambda_I(a,z|^{\neg}S) \log \lambda_I(a,z|iS) - \lambda_I(a,z|iS) \right\} \tag{3.6}$$

where $P(a,t,z)$ is the population count with age a, stage and grade z, in calendar year t, and λ_e is deduced from the corresponding joint pdf

$$\begin{aligned} f(a,z|iS) &= f(a,z,SDx|iS) + f(a,z,CDx|iS) \\ &= \int_0^{a-} \sum_{z_1 \le z} f(a_1,z_1|S) f_{LT}(a-a_1|z_1,a_1) p_b(z|z_1,a-a_1) \mathrm{d}a_1 \\ &\quad + f(a,z|S) \times f_{LT}(0|a,z). \end{aligned}$$

The latter expression represents an incident cancer under iS as either a real CDx with zero lead-time (second term) or a counterfactual one in which case possible presentations at SDx prior to the counterfactual CDx are entertained in the first term of the sum, analogous to backward Markov equations.

Standard errors of the estimates are obtained by bootstrap.

The progression model was fitted to SEER data. Only 5% to 6% patients progress in stage/grade respectively in the localized stage, low grade group at SDx. No progression from the best category (local stage and low grade) to the worst category (distant stage and high grade) is observed, perhaps due to the long time frame required for such a big transition compared to the lead-time. Local stage high grade

Table 3.1: Results of Estimated Progression Probability Matrix (PPM) and 95% C.I.

		Counterfactual Clinical Diagnosis			
	Stage/Grade	LR/L(1)	LR/H(2)	D/L(3)	D/H(4)
Screening Diagnosis	LR/L(1)	p_{b11}=0.893 (0.879,0.906)	p_{b12}=0.051 (0.039,0.062)	p_{b13}=0.056 (0.049,0.064)	p_{b14}=0 (0,0)
	LR/H(2)	0	p_{b22}=0.717 (0.676 0.757)	0	p_{b24}=0.283 (0.243 0.324)
	D/L(3)	0	0	p_{b33}=1 (1,1)	p_{b34}=0 (0,0)
	D/H(4)	0	0	0	1

Stage: LR=Local/Regional, D=Distant. Grade: L=Low(WM), H=High(PU).

patients are more likely to progress. About 28% potentially progress to distant stage. There is no grade progression in the distant stage patients likely because their lead-time is too short for the grade to change.

3.2.5 Survival model

The survival model G describes the time spent in the clinical state conditional on the age of incidence a_I, year of diagnosis t, $(t = x + a_I)$, and stage and grade z.

Two adjustments were made in survival model during the PSA era. First, lead-time adjustment was made to make sure survival times are always measured from the time of clinical diagnosis to time of death implying a guaranteed lead-time benefit. Additionally, survival is conditional on the stage and grade at the time of the clinical diagnosis, the latter being unobserved if the patient is screened-detected. The stage and grade progression during the lead-time was described by the disease progression model in section 3.2.4.

The survival function from the time of diagnosis to time of death (t_s) can be

written as

$$
\begin{aligned}
G(t_s|x, a_I, z) &= G_{LT}(t_s|x, a_I, z) \\
&\quad + f_{LT}(0|x, a_I, z) G_b(t_s|x, a_I, z) \\
&\quad + \int_{0+}^{t_s} \sum_{z_0 \geq z} p_b(z_0|z) f_{LT}(s|x, a_I, z) G_b(t_s - s|x, a_I + s, z_0)\, \mathrm{d}s
\end{aligned} \tag{3.7}
$$

where G_b is the baseline survival function measuring survival time from the clinical diagnosis to death.

All three terms on the right represent mutually exclusive possibilities. The first term is a survival function of the lead-time G_{LT} that represents guaranteed survival up to t_s if the lead-time is at least as large as t_s, i.e. the patient does not die from the disease before he develops symptoms. The second term is the survival contribution of symptomatic diagnoses with the probability of clinical diagnosis (the probability of lead-time is zero) as the weight. The last term represents cancer progression or non-progression from the presentation at SDx to projected CDx, and survival with possibly a more advanced stage thereafter.

3.3 Predictions

We use our model to predict key population and subject-specific characteristics of prostate cancer.

3.3.1 Natural history of Prostate Cancer in US

We present population predictions of major natural history events such as lifetime risk, mean age at event, and mean time between events. In this study, we look at scenarios under the null hypothesis of no treatment.

Cancer onset

Conditional on birth year x, let f_o be the p.d.f. of Y, the time from birth to tumor onset (age at tumor onset), and G_{OC} be the s.f. for T_{OC}, the time from birth to other causes of death. Let y and a_∞ denote the age at tumor onset and the maximal lifetime, respectively. The lifetime risk of tumor onset can be written as a crude cumulative probability of tumor onset Y occurring before death due to other causes (OC). The mean age of tumor onset is computed as a conditional expectation of Y given that the onset does occur.

- lifetime risk

$$\int_0^{a_\infty} f_o(y|x)G_{oc}(y|x)\mathrm{d}y$$

- mean age

$$\frac{\int_0^{a_\infty} y f_o(y|x)G_{oc}(y|x)\mathrm{d}y}{\int_0^{a_\infty} f_o(y|x)G_{oc}(y|x)\mathrm{d}y} \tag{3.8}$$

Cancer diagnosis

Using the definition described in 3.2.1, let the crude density $f^c_{CDx}(a_I|x)$ correspond to clinical diagnosis given birth year x and age of diagnosis a_I. The lifetime risk and the mean age of clinical diagnosis (CDx) under early detection can be computed using the the cumulative crude probability and the conditional expectation similar to (3.8). Conditional on birth year x, let $f^c_{CDx}(a_I, z \in D|x)$ be the joint crude density that represents incidence at age a_I with distant stage and $f_I(a_I|x) = f^c_{CDx}(a_I|x) + f^c_{SDx}(a_I|x)$ be the p.d.f. of the overall incidence at age a_I. Replacing $f^c_{CDx}(a_I|x)$ in equation (3.9) with $f^c_{CDx}(a_I, z \in D|x)$ and $f_I(a_I|x) = f^c_{CDx}(a_I|x) + f^c_{SDx}(a_I|x)$, we

have the lifetime risk and mean age for metastatic clinical diagnosis (Met CDx) and overall diagnosis (CDx or SDx), respectively.

1. Clinical diagnosis (CDx)

   - lifetime risk

   $$\int_0^{a_\infty} f^c_{CDx}(a_I|x)G_{oc}(a_I|x)\mathrm{d}a_I$$

   - mean age

   $$\frac{\int_0^{a_\infty} a_I f^c_{CDx}(a_I|x)G_{oc}(a_I|x)\mathrm{d}a_I}{\int_0^{a_\infty} f^c_{CDx}(a_I|x)G_{oc}(a_I|x)\mathrm{d}a_I} \tag{3.9}$$

2. Metastatic clinical diagnosis (Met CDx)

   - lifetime risk

   $$\int_0^{a_\infty} f^c_{CDx}(a_I, z \in D|x)G_{oc}(a_I|x)\mathrm{d}a_I$$

   - mean age

   $$\frac{\int_0^{a_\infty} a_I f^c_{CDx}(a_I, z \in D|x)G_{oc}(a_I|x)\mathrm{d}a_I}{\int_0^{a_\infty} f^c_{CDx}(a_I, z \in D|x)G_{oc}(a_I|x)\mathrm{d}a_I} \tag{3.10}$$

3. Screening diagnosis (SDx) or clinical diagnosis (CDx)

   - lifetime risk

   $$\int_0^{a_\infty} f_I(a_I|x)G_{oc}(a_I|x)\mathrm{d}a_I$$

   - mean age

   $$\frac{\int_0^{a_\infty} a_I f_I(a_I|x)G_{oc}(a_I|x)\mathrm{d}a_I}{\int_0^{a_\infty} f_I(a_I|x)G_{oc}(a_I|x)\mathrm{d}a_I} \tag{3.11}$$

4. Mean years from onset to tumor diagnosis

   The time from tumor onset to tumor diagnosis $a_I - y$ is called the delay time (tumor age). For the unscreened population, the delay time is the same as the sojourn time, T_{CDx}. For screened the population, the delay time is the minimum of T_{CDx} and T_{SDx}. Mean years from onset to tumor diagnosis can be computed as a conditional expectation of the delay time given cancer diagnosis within the lifetime.

$$\frac{\int_0^{a_\infty}(a_I - y)\int_y^{a_\infty} f_o(y|x)f(a_I - y, z|x, y)G_{oc}(a_I|x)\mathrm{d}a_I\mathrm{d}y}{\int_0^{a_\infty}\int_y^{a_\infty} f_o(y|x)f(a_I - y, z|x, y)G_{oc}(a_I|x)\mathrm{d}a_I\mathrm{d}y} \tag{3.12}$$

   where

$$f(a_I - y, z|x, y) = \begin{cases} f^c_{CDx}(a_I - y|x, y), & \text{CDx.} \\ f^c_{CDx}(a_I - y, z \in D|x, y), & \text{Met CDx.} \\ f_I(a_I - y|x, y), & \text{CDx or SDx.} \end{cases}$$

Cancer death

Conditional on birth year x, the marginal survival function $G_M(a_M|x)$ of age a_M at cancer-specific death (s.f. of mortality) can be written using the convolution of stage/grade specific incidence and survival distributions equation (3.13).

$$\begin{aligned} G_M(a_M|x) = & \int_0^{a_M} f_I(a_I|x)\sum_z f_I(z|x, a_I)G(a_M - a_I|x, a_I, z)\mathrm{d}a_I \\ & + G_I(a_M|x). \end{aligned} \tag{3.13}$$

The first part of the equation describes the probability of a man who has prostate cancer diagnosed at the age a_I (before age a_M) with stage and grade z at the time

of diagnosis, to survive at least $(a_M - a_I)$ years after the diagnosis. The second part of the equation represents the probability for a man who has never been diagnosed with prostate cancer. Similar to equations (3.8), we can compute lifetime risk and mean age of prostate caner death.

- lifetime risk

$$\int_0^{a_\infty} f_M(a_M|x)G_{oc}(a_M|x)\mathrm{d}a_M$$

- mean age

$$\frac{\int_0^{a_\infty} a_M f_M(a_M|x)G_{oc}(a_M|x)\mathrm{d}a_M}{\int_0^{a_\infty} f_M(a_M|x)G_{oc}(a_M|x)\mathrm{d}a_M} \tag{3.14}$$

3.3.2 Risk of adverse events for local-regional screen-detected patients

In this section we provide subject-specific predictions for a PSA-detected man diagnosed with localized-regional stage.

Let a_i and z_i, $i = 0, 1$ denote age and stage/grade at CDx and SDx, respectively. For a man diagnosed by the PSA test, the disease presentation at CDx is unobserved. Conditional on the disease presentation at SDx (a_1, z_1) and the fact that the patient was detected by screening, the p.d.f. of the lead-time distribution is

$$f_{LT}(s|x, a_1, z_1 \in LR, T_{LT} > 0) = \frac{f_{LT}(s|x, a_1, z_1 \in LR)}{1 - f_{LT}(0|x, a_1, z_1 \in LR)}, \tag{3.15}$$

where $s = a_0 - a_1, a_0 \geq a_1$ is the lead-time argument (section 3.2.3). Also the probability of stage and grade progression during the lead-time can be estimated using the PPM (Table 3.1).

1. Clinical diagnosis (CDx)

   Conditional on birth year x and age at diagnosis a_1, the lifetime risk of coun-

terfactual CDx can be presented by the crude probability

$$\int_{a_1^+}^{a_\infty} f_{LT}(a_0 - a_1|x, a_1, z_1 \in LR, T_{LT} > 0) G_{oc}(a_0|x) \mathrm{d}a_0. \tag{3.16}$$

The event of counterfactual CDx defines the cancer as a relevant one (as opposed to overdiagnosed) that would eventually present symptoms without screening.

2. Metastatic clinical diagnosis (Met CDx)

   For the same man, the lifetime risk of metastatic clinical diagnosis is calculated as follows

$$\int_{a_1^+}^{a_\infty} f_{LT}(a_0 - a_1|x, a_1, z_1 \in LR, T_{LT} > 0) p_b(z_0 \in D|z_1 \in LR) G_{oc}(a_0|x) \mathrm{d}a_0. \tag{3.17}$$

3. Prostate cancer death

   In addition to the potential clinical diagnosis, the chance that the man dies from prostate cancer in his lifetime can be expressed using the survival time t_s measured from the age of cancer diagnosis a_1 to age of cancer death a_M ($t_s = a_M - a_1$) as discussed in section 3.2.5. Conditional on SDx ($T_{LT} > 0$), the survival function for the PSA-detected man at the age of a_1 and with stage/grade z_1 is

$$\begin{aligned} G(t_s|x, a_1, z_1) &= G_{LT}(t_s|x, a_1, z_1 \in LR) \\ &\quad + \int_{0+}^{t_s} \sum_{z_0 \geq z_1} p_b(z_0|z_1 \in LR) f_{LT}(s|x, a_1, z_1 \in LR) G_b(t_s - s|x, a_1 + s, z_0)\, \mathrm{d}s. \end{aligned}$$

   Similarly, the lifetime risk of prostate cancer specific death is given by the crude probability

$$\int_{a_1}^{a_\infty} f(a_M - a_1|x, a_1, z_1 \in LR) G_{oc}(a_M|x) \mathrm{d}a_M \tag{3.18}$$

3.4 Data analysis and results

Our analysis was performed using the SEER9 database from 1973 to 2000 which contains more than 350,000 cases of prostate cancer from 9 registries: Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, and Utah. Parameters (except the PPM) in our model were estimated using maximum likelihood methods. Distributions entering the z-specific incidence model were estimated in our previous study (Tsodikov et al., 2006; Chefo and Tsodikov, 2009).

Predictions of key prostate cancer population characteristics in US are summarized in Table 3.2. Numbers listed on the table are averages over the 50-84 age window (a typical age interval for SEER-based statistics) and 1975 to 2000 calendar year window.

On average, the lifetime risk of a man developing prostate cancer is 20%. Among men who develop prostate cancer, their average age at tumor onset is 72 years. If there were no screening, the average lifetime risk of prostate cancer clinical diagnosis with any stage and with distant stage would be 10% and 1%, respectively. Both groups have same average age at diagnosis of 81. Under screening, the average lifetime risk of CDx, Met CDx, and any diagnosis are 6%, 1%, and 15%, respectively. The average age at diagnosis is 80 for a clinically detected case and 75 for any mode of diagnosis. The mean time from onset to diagnosis is 6-7 years for CDx, 3-4 years

Table 3.2: Predictions of key prostate cancer natural history events

	Lifetime risk(%)	Mean age(years)	Mean years from onset
Cancer onset	20	72	
Cancer diagnosis			
- CDx	$10^a/6^b$	$81^a/80^b$	$7^a/6^b$
- Metastatic CDx	$1^a/\ 1^b$	$81^a/80^b$	$4^a/3^b$
- SDx or CDx	15	75	6
Cancer death	$4^a/\ 4^b$	$81^a/81^b$	$10^a/10^b$

$^{a,\ b}$ indicates events under scenarios without and with screening, respectively.

for Met CDx, and 6 years for SDx or CDx cases. The average delay time (time from the tumor onset to diagnosis) is shorter for Met CDx patients compared with the average time from all cases (i.e. 7 vs. 4 years under no screening scenario). This observation indicates that our partially specified model favors heterogeneity in tumor aggressiveness, and more aggressive tumors are detected earlier due to their shorter latency times. Also under the screening scenario, the lifetime risk of CDx is smaller because some of them would be diagnosed earlier by screening. The lifetime risk of death from prostate cancer is 4%. Among those men who die from prostate cancer, the average age of death is 81 years old and it takes about 10 years from the tumor onset. Without any treatments, the risks of cancer death are about the same for screened and unscreened population.

In Table 3.3, we provide projected risks of adverse events for men diagnosed by PSA with local-regional stage prostate cancer in 2000 using SEER data. The predictions are stratified by disease grade (low vs. high) and age groups.

For a PSA-detected local-regional stage prostate cancer case, the probability of CDx within his lifetime is about 59% to 93% depending on the age at diagnosis.

Table 3.3: Lifetime risk of adverse events for local-regional screened-detected patients

Stage/Grade	Age group	Clinical diagnosis	Metastatic clinical diagnosis	Prostate cancer death
	50-54	93	6	28
	55-59	90	5	29
	60-64	86	5	30
LR/L	65-69	80	5	30
	70-74	74	4	28
	75-79	66	4	24
	80-84	59	3	17
	50-54	90	30	60
	55-59	90	26	63
	60-64	86	24	62
LR/H	65-69	80	23	59
	70-74	74	21	53
	75-79	66	19	45
	80-84	59	16	33

Regardless of the grade at disease detection, patients diagnosed at a younger age are more likely to develop symptoms and be detected clinically within their lifetime. Risk of being diagnosed with metastatic disease clinically or the risk dying from prostate cancer within lifetime are also decreasing by age of diagnosis. For high grade PSA-detected patients, the chance of being dignossed later with metastatic disease if left untreated is almost 5 times higher when compared with low grade patients (16-30% in high grade vs. 3-6% in low grade). They have about twice the chance of dying from the prostate cancer in their lifetime (33-60% in high grade vs. 17-28% in low grade).

3.5 Discussions

We presented an analytical statistical model for the joint disease presentation at potentially two diagnoses per subject, SDx and CDx, the latter being counterfactual. The model combines explicit mechanistic assumptions with a partially specified disease progression mechanism.

We formulated the null hypothesis of the treatment effect on the stage-shift as the equality of stage/grade-specific cancer incidence predictions under two counterfactual scenarios of zero screening sensitivity and ignored screening applied to the PSA era. This understanding allowed us to devise an estimation procedure for the progression probabilities for the screen-detected cancer patient despite the absence of longitudinal observations of cancer progression within the subject.

The model is explicit enough as it shows the points where treatment may have an effect. The model can be used to relax the traditional stage-shift assumption by incorporating the treatment effect into a nominal or ordinal model for the progression probability matrix. Another point of application of treatment effects is the treatment by lead-time interaction and the treatment main effect acting on the survival function of time post counterfactual CDx resulting in a complex frailty model. Estimation of such treatment effects would require fitting the model jointly to survival and incidence data.

The minimization of the distance between the two counterfactual incidence predictions to estimate the PPM will not exactly satisfy (3.5). In our data analysis, the quality of the approximation is good for all practical purposes. The fact that only an approximate model is available is a consequence of an independent model formulation for the progression probabilities and the marginal V that could be inconsistent with each other. The alternative would be to define the latent cancer growth and progression process explicitly and fit it to the observed data. However, the mechanism of prostate cancer progression (i.e. whether cancer progresses in grades or whether the grade is fixed at onset) is very much under debate by cancer biologists. In addition strong unjustified assumptions would have to be made to assure identifiability of the complex latent process model from the aggregate observed population

data. Pursuing a robust partially specified model and leaving the exact mechanism of cancer progression open seems to be a better approach.

Nevertheless, some elements of mechanistic modeling of the mechanism of tumor growth and cancer detection may be worth pursuing. For example, in table 3.2, the met CDx patients have a shorter delay time compared with all CDx patients (i.e. 4 vs. 7 years under no screening scenario). This suggests that more aggressive tumors tend to be detected earlier. The model could include the heterogeneity of tumor growth rates in some form. This indicates that the so-called early onset prostate cancers (cancer diagnosed before 55) may represent a subset enriched with aggressive tumors in line with some recent genetics research targeting such patients in search for markers of the aggressive disease.

CHAPTER IV

Mortality Model for Prostate Cancer

4.1 Introduction

There is growing interest in cancer screening programs. The goal of any screening program is to diagnose patients early so they would be detected in a more favorable stage and have better prognosis for treatment. The effect however cannot be evaluated directly using patient-specific data because of the favorable shift in stage of the disease and survival post-diagnosis that would be occurring with screening even if treatment were of no benefit. Prostate-specific antigen (PSA) test was approved by the U.S. Food and Drug Administration (FDA) in year 1988 and it is commonly used by physicians in routine physical exams to screen for prostate cancer in men at risk. However, survival benefit is a matter of debate. Although USA prostate cancer mortality has dropped more than 30 percent since early 1990, coincidently after the introduction of PSA, there has long been a controversy and speculations on the survival benefit of prostate cancer screening. Several ecological studies conducted in United State and Europe found no conclusive evidence to prove the association between the intensity of PSA screening and mortality reduction (Shaw et al., 2004; Collin et al., 2008). One way to assess whether PSA screening contribution to recent mortality decline is by conducting randomized screening trials. Such trials require

large amount of patients and long term follow-up to be able to have sufficient power to compare outcomes. Three large-scale randomized screening trials, the Prostate, Lung, Colorectal and Ovary cancer trial (PLCO) conducted by National Cancer Institute (NCI) in United State, the European Randomized Screening for Prostate Cancer (ERSPC), and the Comparison Arm for ProtecT study in UK (CAP, 2009), are on-going to test whether PSA screening tests reduce the mortality of prostate cancer but it is still too early to make final conclusions. Interim results from PLCO found no effect of prostate cancer screening (Andriole et al., 2009) while ERSPC showed some survival benefit from screenings (Schroder et al., 2009). Besides a statistical convenience approach of testing for benefit in a screening trial will make the effect estimate specific to the trial populations and its generalizability to populations with a different pattern of screening would be a challenge.

The questions can be addressed by statistical modeling. Mathematical and simulation models have been developed in the past to evaluate screening programs. Oftentimes cancer progression is modeled using a semi-Markov stochastic process. Lee and Zelen (2008) developed a general probability model and used it to assess the role of screening programs in breast cancer mortality. The model derives the benefit of screening from more favorable stage distribution of screen-detected cases vs. the clinically diagnosed cases. The assumption that stage at screening diagnosis (SDx) determines the prognosis has been referred to as the stage-shift. In prostate cancer, watchful waiting is a legitimate "treatment" option for patients whose perceived chance of dying from prostate cancer is low. The stage-shift is essentially an interaction effect between early detection and curative treatment based on the assumption that the treatment applied at SDx prevents stage progression, and a patient who would be in an advanced stage at clinical diagnosis (CDx) is prognosti-

cated using early stage survival having been detected at an early stage by screening. This assumption may be problematic particularly if the patient receives no curative treatment on watchful waitings.

This paper was motivated by the following refinements of the general stage-shift mortality model:

1. We wanted to use cancer registry data, where screening schedules are unobserved, to specify the inputs for the prediction of mortality. Therefore, we treat screening schedules as an unobserved point process with known distribution. This makes the mortality prediction a functional of screening policy as conditioning on intermediate outcomes is avoided.

2. The prediction of survival for a patient detected with cancer is based on updating the distribution of this patient's natural history of the disease conditional on age, stage and grade observed at diagnosis, recognizing that the mode of diagnosis (SDx, or CDx) is not observed in SEER data. Treatment effects interact with the latent natural history.

3. We used an categorical model for treatment allocation based on patient's information available at diagnosis that modeled treatment patterns over time.

4. We provide a flexible stage-shift approach to modeling and estimating the chance of stage progression between SDx and the counterfactual CDx (one that would occur if the results of screening were ignored). Patients who progress in stage or grade by the end of the lead time will have advanced stage/grade survival.

One dimension of the screening and treatment interaction is how the treatment affects the stage-shift and the chance of cancer progression.

Section 4.2 will describe the main elements of the mortality model. In Section 4.3 we apply the model to analyze US prostate cancer mortality. Finally, Section 4.4 will discuss the significance of our model and potential problems encountered using current approach.

4.2 Model

4.2.1 Basic Model

Our basic model used the classical three-state chronic disease model. Three states, disease-free state, pre-clinical state, and clinical state, represent the progression of the natural history of the disease, and transitions are irreversible. The disease progression with respect to stage and grade is only partially specified as will be described in Section 4.2.4. The disease-free state is measured from the time a man turns fifty to the time of the tumor onset. For population under screening, the screening diagnosis and the clinical diagnosis are two competing risks operating at the pre-clinical state. The pre-clinical state is calculated from time of tumor onset to time of clinical diagnosis or screening diagnosis whichever comes first. The clinical state describes the survival time from the time of diagnosis to the time of death or last follow up.

4.2.2 Mortality Model

Conditional on birth year x, cancer mortality is a hazard function, λ_M, of the age at prostate cancer death, a_M. By definition,

$$\lambda_M(a_M|x) = \frac{f_M(a_M|x)}{G_M(a_M|x)}$$

where $f_M(a_M|x)$ is the probability density function (p.d.f.) and $G_M(a_M|x)$ is the survival function (s.f.). For a man from the birth cohort x, the probability of surviving

from prostate cancer at age a_M can be written as

$$G_M(a_M|x) = \int_0^{a_M} f_I(a_I|x) \sum_z f_I(z|x, a_I) \sum_{Tx} f_{Tx}(Tx|x, a_I, z) G(a_M - a_I|x, a_I, z, Tx) \mathrm{d}a_I$$

$$+ G_I(a_M|x) \tag{4.1}$$

The first part of the equation describes the probability for the man, who had prostate cancer diagnosed at the age a_I (before age a_M) with stage and grade z, received treatment Tx at the time of diagnosis, and survived at least $(a_M - a_I)$ years after the diagnosis. The second part of the equation represents the probability for the man who has never been diagnosed with prostate cancer. Mortality is a convolution of five models addressing the development of the disease under screening and treatment interactions: Marginal incidence(section 3.2.1), Z-specific incidence (section 3.2.2), treatment(section 4.2.3), disease progression between SDx and counterfactual CDx (section 4.2.4), and survival models (section 4.2.5). The latter two models are used to specify the s.f. G in equation 4.1 describing survival time post real or counterfactual CDx conditional on stage and grade at that point .

4.2.3 Treatment Model

The treatment model describes the probability of receiving a certain treatment combination at the time of cancer diagnosis. Using SEER data, we classified treatments into three major categories: Conservative Management (CM), Radiation Therapy (RT), and Radical Prostatectomy (RP). Hormone Therapy (HT) is commonly used as an adjuvant therapy following one of those three primary treatments, RT most commonly. Note that HT information is not available in cancer registries, hence we used a two-stage model to predict treatment allocations. In the first stage, we modeled the probability of receiving one of the three treatments defined in SEER data conditional on birth year x, age of diagnosis a_I, and grade using

multinomial logit model. In the second stage, we conditioned on the treatment obtained from the first stage, birth year x, age of diagnosis a_I, and grade and used logistic regression to model the probability of receiving the additional adjuvant hormone therapy. The second stage model was fitted to CaPSURE data (`http://urology.ucsf.edu/capsure/overview.htm`)

By combining first and second stage, we can obtain the probability of actual treatment received given birth year x, age of diagnosis a_I, and grade.

4.2.4 Disease Progression Model

The disease progression model estimates the probability of disease progression during the lead time in the absence of treatment. Assuming no stage and grade regression, the disease presentation at the end of the lead time is given by a set of baseline transition probabilities, $p_b s$, that can be summarized as a progression probability matrix (PPM) (Table 3.1).

The impact of treatment can be introduced into the transition probabilities p_b using a cumulative logit ($k > 2$) or a logistic regression ($k = 2$) model working with a row of the PPM matrix. In the logistic regression model, let p_b be the baseline transition probability, and η_{Tx} be the treatment effect (odds) on the baseline transition probability. The transition probability given the treatment effect is

$$p_{ij} = \Pr(z_0 = j | z_1 = i, Tx) = \frac{p_{bij}\eta_{Tx}}{p_{bij}(\eta_{Tx} - 1) + 1}, j > i, \tag{4.2}$$

where $z_{0,1}$ is stage-grade response at CDx, and SDx, respectively.

In the cumulative logit model, let cp_b be the cumulative baseline transition proba-

bility. The cumulative transition probability given the treatment effect is

$$cp_{ij} = p(z_0 \leq j | z_1 = i, Tx) = \frac{cp_{bij}/\eta_{Tx}}{cp_{bij}(1/\eta_{Tx} - 1) + 1}, j \geq i \tag{4.3}$$

$$\text{and } p_{ij} = cp_{ij} - cp_{i(j-1)}.$$

When there is no treatment effect on the probabilities (i.e. $\eta_{Tx} = 1$), $p_{ij} = p_{bij}$. When the treatment effect is really large (i.e. $\eta_{Tx} \to 0$), then PPM will become an identity matrix corresponding to the full stage-shift when the stage and grade carry over from the screening diagnosis to the counterfactual clinical diagnosis.

4.2.5 Survival Model

The survival model G describes the time spent in the clinical state conditional on the age of incidence a_I, year of diagnosis t, $(t = x + a_I)$, stage and grade Z, and treatment Tx. As discussed in our earlier paper (Lee and Tsodikov, 2010), the estimated treatment effect using the Cox PH model is biased under early detection (with random lead time added to the survival time). In this paper, we propose two approaches to assess the treatment efficacy using the full likelihood while accounting for the heterogeneity and survival biases induced by the early detection.

Generalized Stage-shift Model

The Generalized stage-shift model is motivated by the traditional stage-shift model by allowing disease progression during the lead time. Two adjustments were made in the survival model during the PSA era. First, lead time adjustment was made to make sure survival times are always measured from the time of clinical diagnosis to time of death implying a guaranteed lead time benefit. Additionally, survival is conditional on the stage and grade at the time of the clinical diagnosis, the latter being unobserved if the patient is screened-detected. Three treatment effects were investigated in the model. The treatment main effect θ_{Tx} describes how

treatment affects post-lead time survival. The treatment effect applied to the disease progression (η_{Tx}) measures how treatment prevents stage and grade progression during the lead-time. Finally, even if the patient progresses during the lead time, the one treated earlier is still better off than the one treated close to the point of symptoms, and this is represented by the treatment by lead-time interaction effect, $\theta_{Tx\times LT}$ in the s.f. G post real or counterfactual CDx. We did not model treatment by early detection interaction effects in distant stage at CDx because the predicted conditional lead time was too short to have an effect.

$$\begin{aligned} G(t_s|a_I,t,z,Tx) &= G_{LT}(t_s|a_I,t,z) \\ &\quad + f_{LT}(0|a_I,t,z)G_b(t_s|a_I,t,z,Tx) \\ &\quad + \int_{0+}^{t_s} \sum_{z_0\geq z} \Pr(z_0|z,Tx) f_{LT}(s|a_I,t,z) G_b(t_s-s|a_I+s,t,z_0,Tx)\,\mathrm{d}s \end{aligned} \tag{4.4}$$

and

$$G_b(t_s-s|a_I+s,t,z_0,Tx,s) = G_0(t_s|a_I,t,z_0)^{\theta_{Age}\theta_{Tx}\theta_{Tx\times LT}}$$

where G_0 is the model under a reference group under conservative management.

Calibration Model

An alternative approach to the Generalized stage-shift model is the Calibration model. Similar to the stage-specific incidence model we use the delay time as a frailty. Additionally, we use the mean lead time predicted given information observed at diagnosis as a covariate for s.f. G. The latter aspect has motivated the name "calibration" by analogy with measurement error models. With this model integrating over future latent development is avoided which makes the computation a little faster. The modeled treatment effects are conceptually similar to the Generalized Stage-Shift Model. For instance, this model includes the treatment by *expected* lead time interaction effect to assess the benefit of treatment under early detection.

Replacing the latent lead time with the expected lead time allows us to avoid making any assumptions on disease progression. The main treatment effect operates on the survival time regardless of the mode of diagnosis.

$$ELT = \int_0^\infty s\, f_{LT}(s|a_I, t, z)\mathrm{d}s$$

$$G(t_s|a_I, t, z, Tx) = G_0(t_s|a_I, t, z_0, Tx_0)^{\theta_{Age}\theta_{Surrogate}\theta_{Tx}\theta_{Tx\times ELT}}. \tag{4.5}$$

4.3 Data Analysis and Results

Our analysis was performed using SEER9 database from 1973 to 2000 and parameters were estimated using maximum likelihood methods. It contains more than 350,000 cases of prostate cancer from 9 registries: Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, and Utah. All observed or predicted rates were adjusted for the US 2000 population age distribution obtained from the Human Mortality Database(HMD), Max Plank Institute for Demographic Research (2009). Figure 4.1 shows the observed and expected US prostate cancer incidence $\lambda_I(a_I|x)$ by calendar year for men over 50. Prostate cancer incidence showed an increasing trend over time before the PSA era followed by the surge resulting from increased use of PSA tests after year 1988.

Figure 4.2 shows the predicted stage and grade distribution $f_I(z|a_I, x)$ at diagnosis. The majority of prostate cancer patients are likely to be diagnosed in local/regional stage and with low grade (WM). Getting older increases the chance of being diagnosed in advanced stage. For patients diagnosed at the same age, those who are diagnosed more recently have more favorable stage and grade distributions.

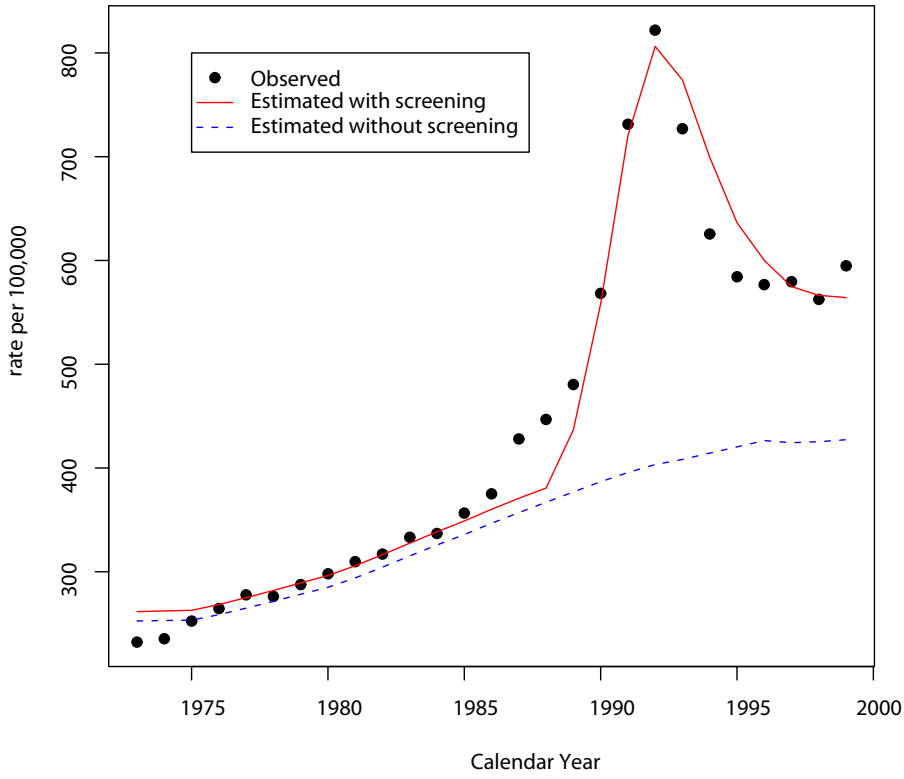


Figure 4.1: Observed and expected incidence by year.

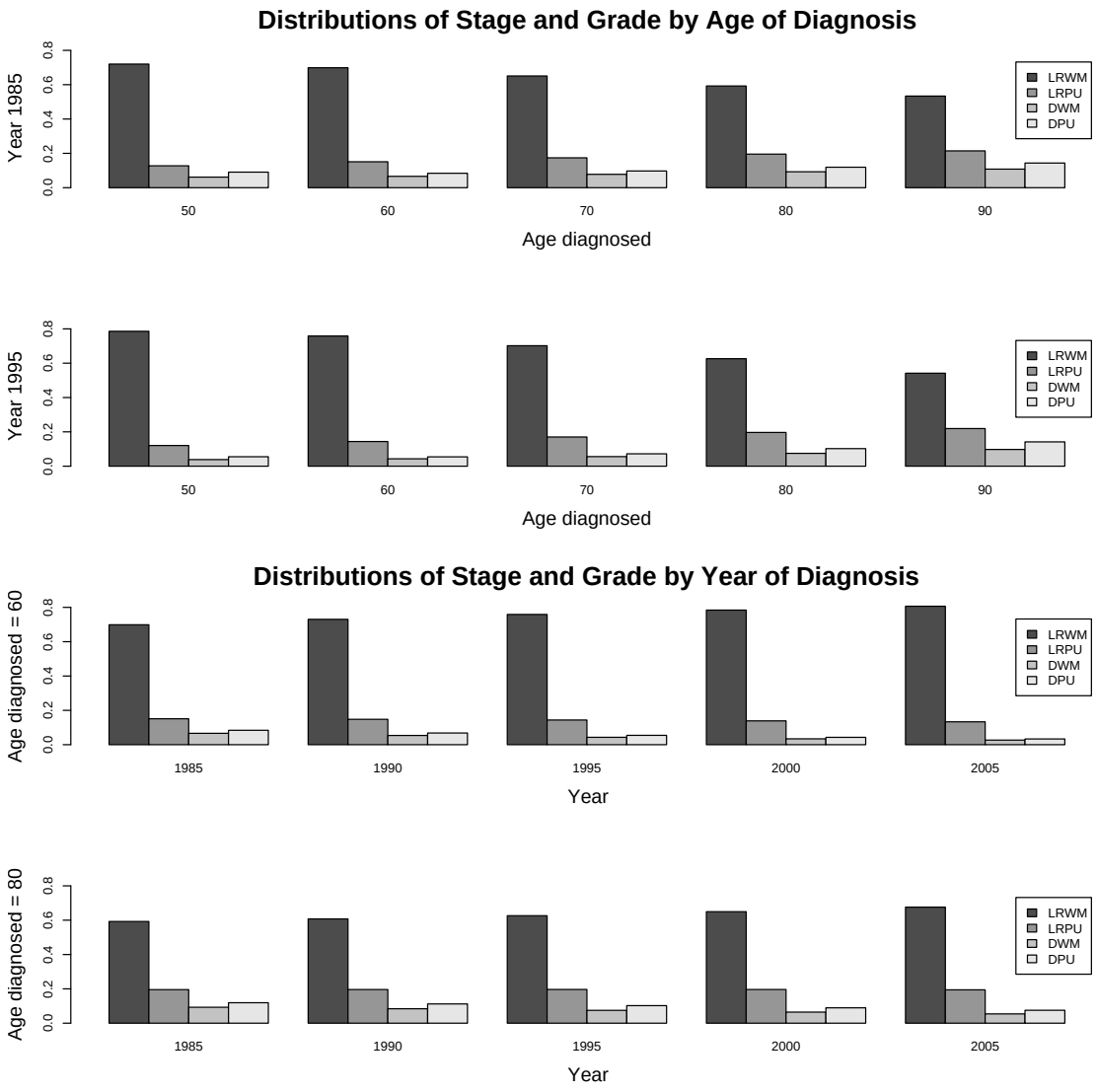


Figure 4.2: Expected stage and grade distribution by age and year of diagnosis.

Figure 4.3 and 4.4 show the probabilities of receiving particular treatments as used by the model, by age of diagnosis, year of diagnosis, and grade using the multinomial logit model. As seen in Figure 4.3, older patients are more likely to get conservative management while younger patients tend to receive more aggressive treatments such as radical prostatectomy (RP) or radio therapy (RT). The trends are similar for both low and high grade patients. Figure 4.4 shows that the use of conservative management (CM) was gradually decreasing over the years while radio therapy (RT) became more popular in recent years. More radical prostatectomy procedures were performed around year 95. This trend is more dramatic in low grade patients than it is in high grade patients. Figure 4.5 shows use of hormone therapy (HT) stratified by the primary treatment received, year of diagnosis, and grade (logistic regression model). Use of hormone therapy increased in recent years particularly after RT.

Figure 4.6 shows the estimated mean lead time conditional on age, year, stage, and grade at the time of diagnosis. The highest mean lead time occurs in men diagnosed in localized stage at around the age of 70, in year 1992. Local/regional disease patients have much longer mean lead time than distant stage patients. Regressing lead time on stage is therefore essential. Within each stage, there are only small differences between lower and higher grade patients.

Table 4.1 and 4.2 showed the estimated treatment effect under early detection stratified by stage and grade. Assuming other covariates being equal, Radical Prostatectomy (RP) is associated with a substantial reduction of the hazard of prostate-specific death (65% - 86% in the Generalized Stage-shift Model, and 56% - 89% in the Calibration model). Applying the cumulative logit formula in equation (4.3), the Generalized stage-shift model showed that RP reduced the probability of disease progression especially in local-regional high grade patients (LR/L to LR/H from 5.1%

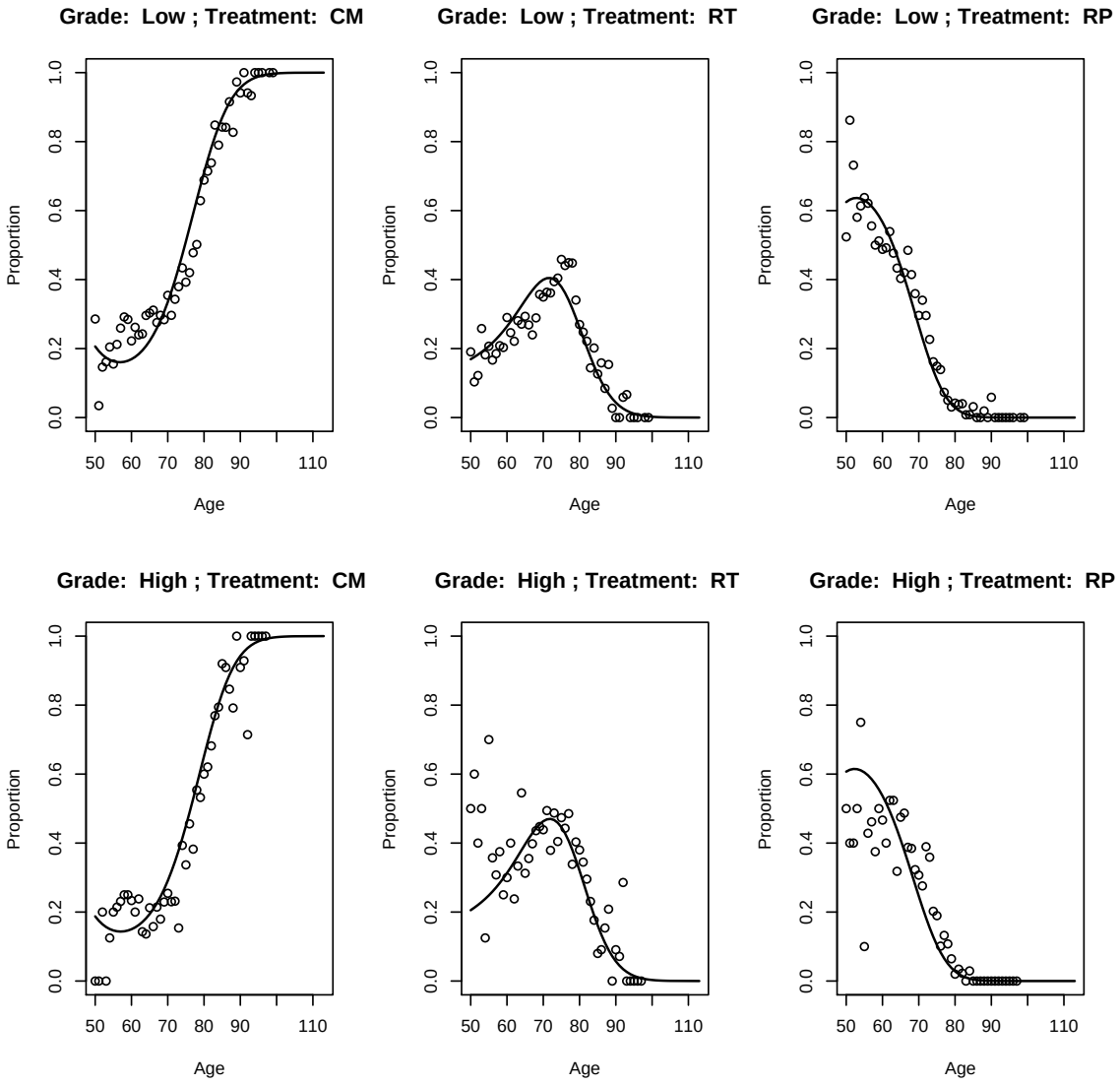


Figure 4.3: Distribution of SEER treatment by age of diagnosis and grade.

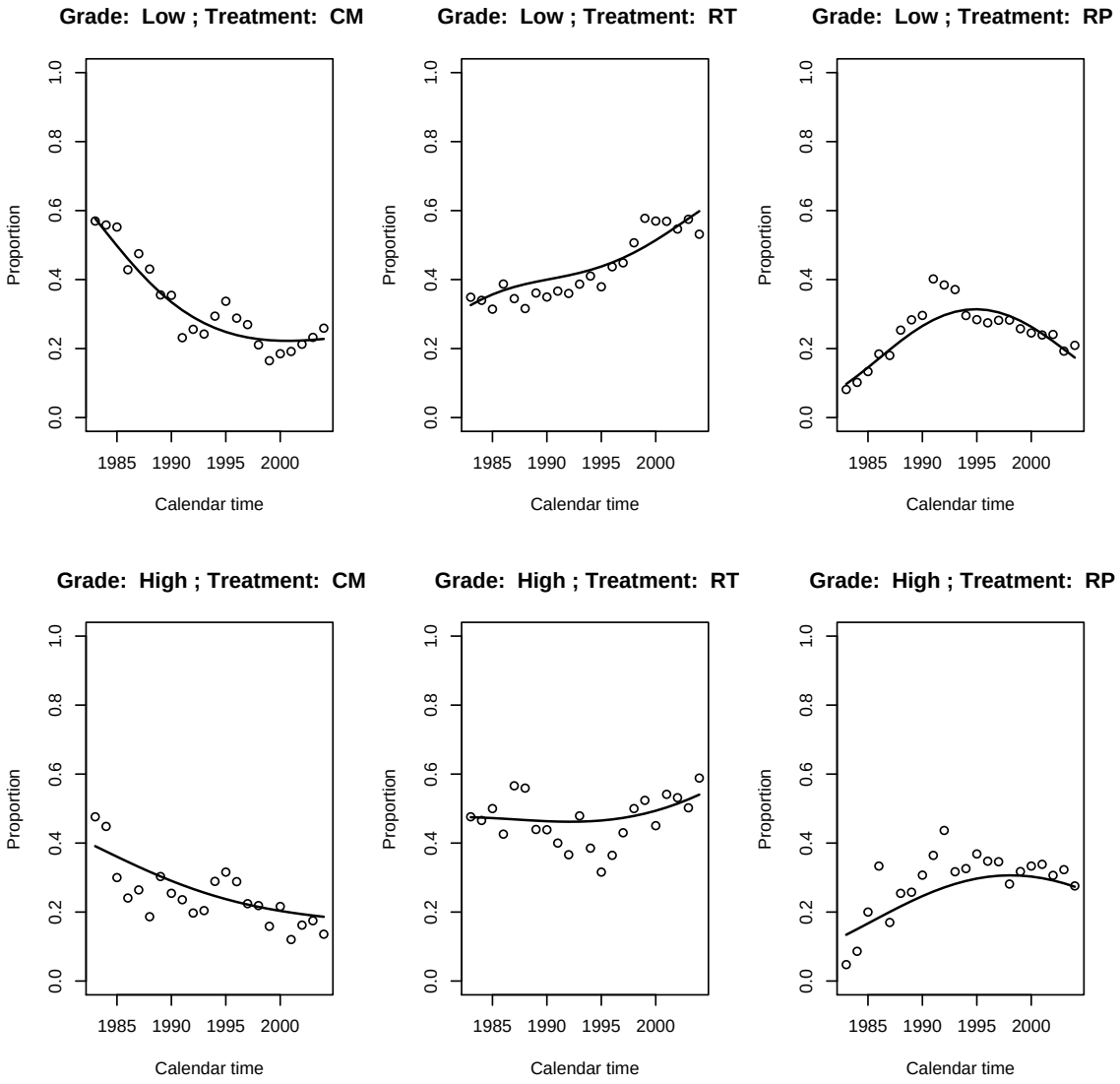


Figure 4.4: Distribution of SEER treatment by year of diagnosis and grade.

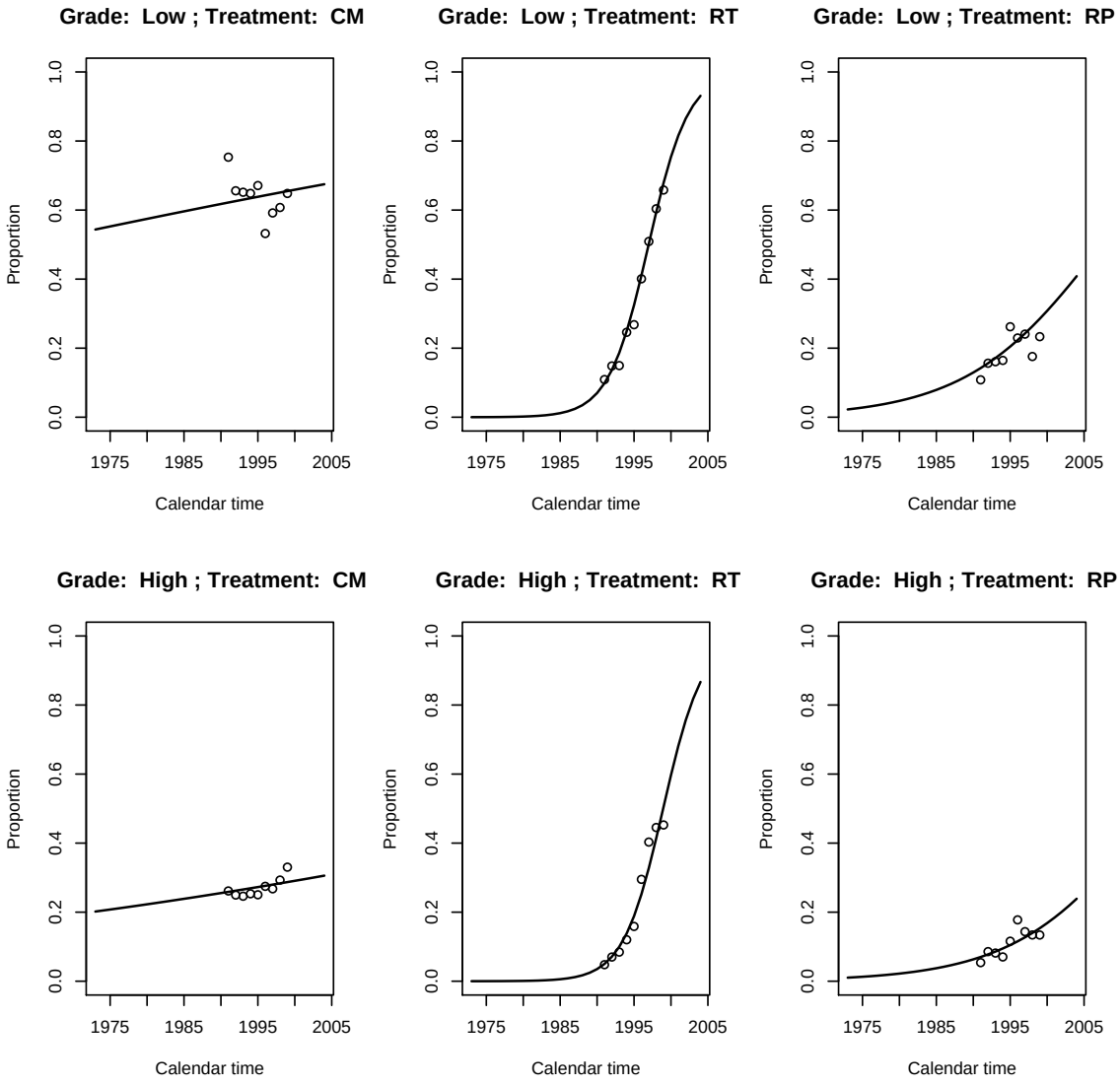


Figure 4.5: Distribution of Hormone Therapy (HT) by year of diagnosis and grade.

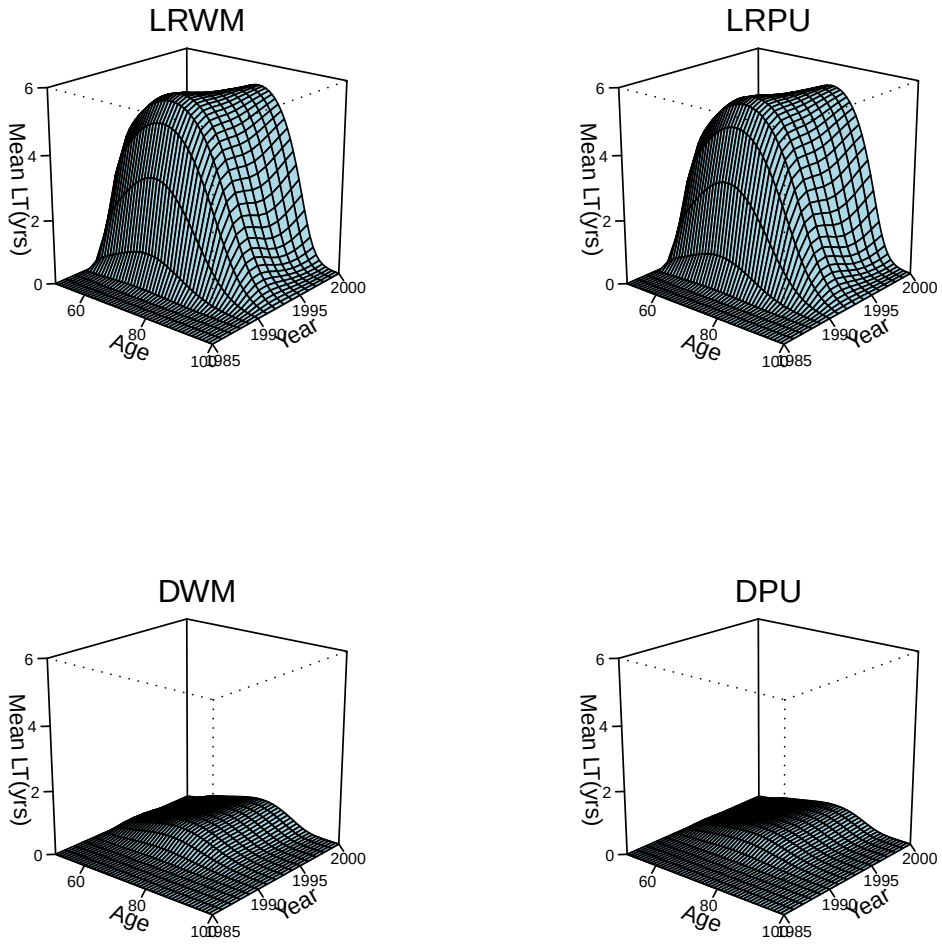


Figure 4.6: Mean lead time given stage, grade, age, and year of diagnosis.

to 0.7%, LR/L to D/L from 5.6% to 0.6%, and L/H to D/H from 28% to 0%) but no additional survival post lead-time benefits from treating patients early were found as a result of variable selection. The result from the Calibration model also supports the evideince of the RP by early detection benefit. Each additional year of mean lead time predicted at the time of diagnosis will lower the hazard by additional 9% to 24% for patients treated with RP.

Similar to RP, Radiotherapy (RT) worked better for local-regional stage high grade patients (HR=0.53 and 0.56 in Generalized stage-shift model and Calibration model, respectively), and it had an additional survival benefit by treating patient early. Advancing treatment by each year will reduce the hazard of dying from the prostate cancer by additional 3%. However, it has no advantage on preventing disease progression according to table 4.1. In Calibration model, a moderate RT early detection interaction effect was found in local-regional stage patients (HR=0.82 and 0.89 for LR/L and LR/H, respectively). For Local regional stage low grade patients, the combination of negative main treatment effect and the postive treatment expected lead time interaction effect might be interpretated as the trend of Radiotherapy effect improving with calendar time into the PSA era.

Figure 4.7 shows the observed and expected age-adjusted US prostate cancer mortality for men over 50 using our models. The US prostate caner mortality increased slightly from year 1980 to year 1990 and decreased more than 30% since. Both mortalities using survival function estimated from the Generalized stage-shift model and the Calibration model predicted the US mortality well. As shown in table 4.1 and 4.2, a decrease in mortality resulted from increased uses of screening and advanced treatments in recent years. Other conditions held equal, treating patients earlier may further improve survival by preventing disease progression or increasing survival post

Table 4.1: Estimated treatment effects by stage and grade in Generalized stage-shift model

	Hazard ratio (HR)			
	LR/L	LR/H	D/L	D/H
Treatment main effect, θ_{Tx}				
RT vs. CM	1.07	0.53	0.96	1.10
RP vs. CM	0.35	0.19	0.14	0.28
Treatment effect on disease progression, η_{Tx}				
RT vs. CM	1.00	1.00	N/A	N/A
RP vs. CM	0.11	0.00	N/A	N/A
Treatment lead time interaction, $\theta_{Tx\times LT}$				
RT vs. CM	1.00	0.97	N/A	N/A
RP vs. CM	1.00	1.00	N/A	N/A

Table 4.2: Estimated treatment effects by stage and grade in Calibration model

	Hazard ratio (HR)			
	LR/L	LR/H	D/L	D/H
Treatment main effect, θ_{Tx}				
RT vs. CM	1.40	0.56	0.96	1.36
RP vs. CM	0.46	0.19	0.11	0.12
Treatment expected lead time interaction, $\theta_{Tx\times ELT}$				
RT vs. CM	0.82	0.89	N/A	N/A
RP vs. CM	0.76	0.91	N/A	N/A
Surrogate				
Expected lead time	0.99	0.96	N/A	N/A
Delay time	1.12	1.17	1.18	0.51

the lead time in selected groups of patients.

4.4 Discussions

We have presented a hierarchical family of models synthesized into a causal model for cancer mortality linking the mortality endpoint with the population trends of treatment and screening utilization. The model is an analytical model based on a mechanistic description of the history of disease development, partially specified disease progression mechanism, and its interaction with screening and treatment. It provides an assessment of the benefit of screening and treatment for US prostate cancer mortality and explains the mortality decline in the PSA era.

We developed an estimate of survival adjusted for the lead time and the disease progression between the point of screening diagnosis and the counterfactual clinical diagnosis. Length-biased sampling is reflected in our conditional lead time distribution given the disease presentation at diagnosis. We were able to vary the stage-shift assumptions of the early detection program by addressing the disease progression after the screening diagnosis. This allowed us to question the traditional stage-shift assumption that stage and grade at the time of screening diagnosis carries over to the time of projected clinical diagnosis.

Both survival models we proposed were able to explain the mortality decline resulting from the combined effect of screening and treatment for prostate cancer. A more mechanistic model, the Generalized stage-shift model, is more specific about how the treatment effect is helped by early detection. Because this model operates on the counterfactual CDx scenario, it can be used to assess the effects of deferred treatment of prostate cancer in early stage patients. There might be other selection

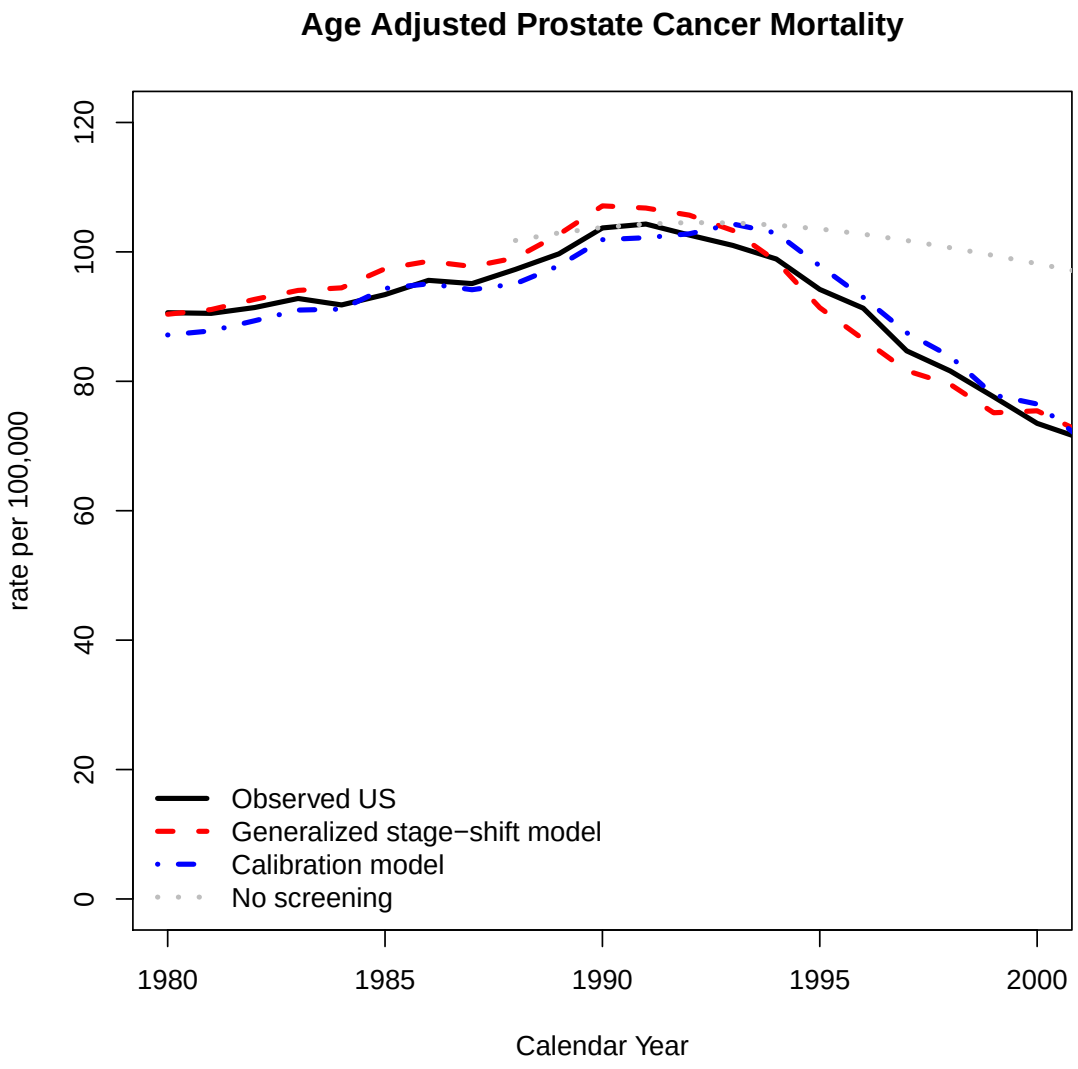


Figure 4.7: Observed and expected mortality by calendar year.

effects that can be incorporated into the more mechanistic Stage-shift model. One example is a possible length biased effect on s.f. G post real or counterfactual CDx. For the purpose of reproducing the population mortality and survival trends, the Calibration model does an equally good job and is slightly less complex computationally.

We have shown that treatment and early detection effects were varied by treatment types and the clinical characteristics of the disease. Answering how and when to treat patients will be the key to help patients, physicians and policy makers make the best decision to improve treatment outcomes while managing the quality of life and to achieve overall cost-effectiveness. This avenue can be explored by incorporating dynamic deferred treatment regimen into the mechanistic models such as the Generalized Stage-shift model.

CHAPTER V

Discussion

The goal of any early detection program is to diagnose and treat patients early to improve the prognosis of disease and improve treatment. While screening intervention might enhance treatment and improve the survival outcome and quality of life in some patients, it also has a profound impact on heterogeneity and the meaning of clinical variables for newly diagnosed patients. Analyzing data using such a dynamic population could be challenging.

Our study is motivated by the controversy of the benefit of PSA screening. We propose an analytic joint statistical model based on the classical three-state chronic disease model to assess the benefit of screening and treatment on US prostate cancer mortality. We are able to relax the traditional stage-shift assumption and present a new approach to estimate disease progression probabilities without direct within-subject observations. We also develop an estimate of survival adjusted for the lead time and the disease progression between the point of screening diagnosis and the counterfactual clinical diagnosis. This allows us to quantify the stage-shift and the treatment effect explicitly and model interaction between treatment and early detection. The model can be applied to describe the latent natural history disease characteristics and help newly diagnosed patients make decisions on cancer manage-

ment based on information given at the time of the diagnosis.

In the study, we also demonstrate how the early detection affects the clinical survival outcome and show how evaluating treatment effect using patients recruited from screened population could lead to bias. In addition, we study the direction of the bias and propose a meta-analytic approach to correct the bias.

While our study provides innovative methods and interesting findings, there are limitations in the current approach. For example, the disease progression probability in our model does not depend on the lead-time. This means that the probability of disease progression after the screening diagnosis is the same no matter how far in advance the disease is detected. Further study is needed to validate this assumption and to develop new methods to include lead-time into the current setting.

We also make simplifying assumptions to study the lead-time bias on treatment effect. Length bias, disease progression, and associated treatment effects are ignored and should be incorporated in our future setting.

To fight cancer and reduce the burden of cancer, it is important to continue searching for more effective treatments and utilizing early detection programs. Currently, treatments are mainly evaluated using randomized clinical trials (RCT) and other follow-up studies, while the early detection of the disease and its benefits are evaluated using population data, observational studies, and screening trials. However, estimating benefits from a single data source and study design are vulnerable to population heterogeneity due to unmeasured or unknown factors and selection effects. For example, cancer registries (i.e. SEER) have less detailed representation of disease-specific clinical characteristics while clinical databases often miss on population based processes, including utilization of diagnostic tests in the specific population from which the study group was recruited. Besides, limited availability

of raw data from RCT often provide only a summary measure of the observed effect and its variability obtained by using a statistical convenience model.

We believe that it is not possible to generate valid predictions of treatment effectiveness while restricting the analysis to a particular study design or dataset. We need to continue to develop new methodology based on the joint analysis of multiple sources of data under various study designs providing information on heterogeneity, unmeasured factors, and treatment outcomes.

Last but not least, the mechanistic models presented in this thesis are amenable to the introduction of dynamic deferred treatment regimen. The paradigm of detecting and treating cancer as early as possible is definitely not adequate in prostate cancer because of the massive overdiagnosis of the disease and diminished quality of life due to treatment. A sensible strategy of balancing the risks, quality of life, and costs lies in making treatments dynamic and deferring treatment until some indication of disease progression. Providing the mechanistic basis for disentangling the causality of such studies is an exciting future development for the proposed models.

Bibliography

Albertsen, P., Hanley, J., Penson, D., Barrows, G., and Fine, J. (2007). 13-year outcomes following treatment for clinically localized prostate cancer in a population based cohort. *The Journal of Urology* **177,** 932–936.

Andersen, P. and Gill, R. (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics* pages 1100–1120.

Andriole, G., Crawford, E., Grubb III, R., Buys, S., Chia, D., Church, T., Fouad, M., Gelmann, E., Kvale, P., Reding, D., et al. (2009). Mortality results from a randomized prostate-cancer screening trial. *The New England journal of medicine* **360,** 1310.

Berry, D., Cronin, K., Plevritis, S., Fryback, D., Clarke, L., Zelen, M., Mandelblatt, J., Yakovlev, A., Habbema, J., Feuer, E., et al. (2005). Effect of screening and adjuvant therapy on mortality from breast cancer. *The New England journal of medicine* **353,** 1784.

Bill-Axelson, A., Holmberg, L., Ruutu, M., Haggman, M., Andersson, S., Bratell, S., Spangberg, A., Busch, C., Nordling, S., Garmo, H., et al. (2005). Radical Prostatectomy versus Watchful Waiting in Early Prostate Cancer. *New England Journal of Medicine* **352,** 1977–1984.

CAP (last access Sept. 2009). The comparison arm for protect (cap) study. `http://www.controlled-trials.com/ISRCTN92187251`.

Chefo, S. and Tsodikov, A. (2009). Stage-specific cancer incidence: an artificially mixed multinomial logit model. *Statistics in Medicine* **28,** 2054–2076.

Collin, S., Martin, R., Metcalfe, C., Gunnell, D., Albertsen, P., Neal, D., Hamdy, F., Stephens, P., Lane, J., Moore, R., et al. (2008). Prostate-cancer mortality in the USA and UK in 1975–2004: an ecological study. *Lancet Oncology* **9,** 445–452.

Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)* pages 187–220.

Davidov, O. and Zelen, M. (2004). Overdiagnosis in early detection programs. *Biostatistics* **5,** 603.

Draisma, G., Boer, R., Otto, S., Van Der Cruijsen, I., Damhuis, R., Schroder, F., and De Koning, H. (2003). Lead times and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. *JNCI Journal of the National Cancer Institute* **95,** 868.

Draisma, G., Etzioni, R., Tsodikov, A., Mariotto, A., Wever, E., Gulati, R., Feuer, E., and De Koning, H. (2009). Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. *JNCI Journal of the National Cancer Institute* .

Etzioni, R., Legler, J., Feuer, E., Merrill, R., Cronin, K., and Hankey, B. (1999). Cancer surveillance series: interpreting trends in prostate cancerpart III: quantifying the link between population prostate-specific antigen testing and recent declines

in prostate cancer mortality. *JNCI Journal of the National Cancer Institute* **91,** 1033.

Etzioni, R., Penson, D., Legler, J., Di Tommaso, D., Boer, R., Gann, P., and Feuer, E. (2002). Overdiagnosis due to prostate-specific antigen screening: lessons from US prostate cancer incidence trends. *JNCI Journal of the National Cancer Institute* **94,** 981.

Fienberg, S. (1980). The Analysis of Cross-Classified Categorical Data (ed.).

Gail, M., Wieand, S., and Piantadosi, S. (1984). Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* **71,** 431–444.

Kempthorne, O. (1977). Why randomize. *J. Statist. Plann. Inference* **1,** 1–25.

Lagakos, S. and Schoenfeld, D. (1984). Properties of proportional-hazards score tests under misspecified regression models. *Biometrics* **40,** 1037–48.

Lee, S. and Tsodikov, A. (2010). Treatment effects under early detection of cancer. *Submitted to Biometrics. Manuscript under revision* .

Lee, S. and Zelen, M. (2008). Modelling the early detection of breast cancer. *Biometrics* **64,** 386–95.

Mariotto, A., Etzioni, R., Krapcho, M., and Feuer, E. (2007). Reconstructing PSA testing patterns between black and white men in the US from Medicare claims and the National Health Interview Survey. *Cancer* **109,** 1877–1886.

Max Plank Institute for Demographic Research (last access Sept. 2009). Human mortality database. `http://www.mortality.org`.

Nicholson, P. and Harland, S. (2002). Survival prospects after screen-detection of prostate cancer. *BJU international* **90,** 686–693.

Parker, C., Muston, D., Melia, J., Moss, S., and Dearnaley, D. (2006). A model of the natural history of screen-detected prostate cancer, and the effect of radical treatment on overall survival. *British journal of cancer* **94,** 1361–1368.

Schroder, F., Hugosson, J., Roobol, M., Tammela, T., Ciatto, S., Nelen, V., Kwiatkowski, M., Lujan, M., Lilja, H., Zappa, M., et al. (2009). Screening and prostate-cancer mortality in a randomized European study. *New England Journal of Medicine* **360,** 1320.

Schumacher, M., Olschewski, M., and Schmoor, C. (1987). The impact of heterogeneity on the comparison of survival times. *Statistics in Medicine* **6,** 773–784.

Shaw, P., Etzioni, R., Zeliadt, S., Mariotto, A., Karnofski, K., Penson, D., Weiss, N., and Feuer, E. (2004). An Ecologic Study of Prostate-specific Antigen Screening and Prostate Cancer Mortality in Nine Geographic Areas of the United States. *American Journal of Epidemiology* **160,** 1059–1069.

Shea, G. (1979). Monotone regression and covariance structure. *The Annals of Statistics* **7,** 1121–1126.

Struthers, C. and Kalbfleisch, J. (1986). Misspecified proportional hazard models. *Biometrika* **73,** 363–369.

Tewari, A., Raman, J., Chang, P., Rao, S., Divine, G., and Menon, M. (2006). Long-term survival probability in men with clinically localized prostate cancer treated either conservatively or with definitive treatment (radiotherapy or radical prostatectomy). *Urology* **68,** 1268–1274.

Tsodikov, A. (2003). Semiparametric models: a generalized self-consistency approach. *Journal of the Royal Statistical Society. Series B, Statistical Methodology* pages 759–774.

Tsodikov, A., Szabo, A., and Wegelin, J. (2006). A population model of prostate cancer incidence. *Statistics in Medicine* **25,** 2846–2866.

Van der Vaart, A. (2000). *Asymptotic statistics*. Cambridge Univ Pr.

Zelen, M. (2004). Forward and backward recurrence times and length biased sampling: age specific models. *Lifetime Data Analysis* **10,** 325–334.

Zelen, M. and Feinleib, M. (1969). On the theory of screening for chronic diseases. *Biometrika* **56,** 601–614.