Semiparametric and Joint Modeling of Cancer Screening

by

Sheng Qiu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2017

Doctoral Committee:

Professor Alex Tsodikov, Chair
Associate Professor Rafael Meza
Professor Susan Murray
Professor Douglas E. Schaubel

Sheng Qiu

shqiu@umich.edu

ORCID iD: 0000-0002-2408-4961

To my parents

ACKNOWLEDGEMENTS

I would like to express my utmost appreciation and gratitude to my advisor, Dr. Alex Tsodikov. Thank you for your invaluable guidance, mentorship, and patience throughout the years. You have made me a better student, a better researcher, and helped me learn a great many. This dissertation could never be finished without your instruction, inspiration, encouragement and motivation. It has been truly a pleasure and honor to work with you.

Great thanks to my committee members Dr. Rafael Meza, Dr. Susan Murray, and Dr. Douglas E. Schaubel, for your constructive input, help and precious time on this research. I must also thank Dr. Douglas E. Schaubel for offering me the opportunities to work with you on Survival Analysis course during summer over the last few years, which has improved a variety of my skills, statistically and behaviorally.

To my parents, Xiukang Qiu and Suying Zeng, thank you for your love encouragement, support, effort and sacrifice that have helped make this dissertation possible and make me a better person. This is dedicated to you.

Finally, gratefulness to all the knowledgeable faculty members at Michigan for their excellent lectures and guidances. To all my friends in Ann Arbor, thank you for making the past six years enjoyable and memorable. I will miss all the time with your company and forever cherish these memories and friendship. Hope one day our lives cross paths again.

This work was supported by the grants 1U01CA199338 (CISNET) and 5P50CA186786 (SPORE) of the National Cancer Institute.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Introduction of screening for prostate cancer using the prostate-specific antigen (PSA) biomarker of the disease in the late 80ies led to remarkable dynamics of the incidence of the disease and shortly after, cancer mortality showed a decline. Except for fragmentary studies, no comprehensive information exists on the PSA uptake in the European countries that would allow specification of utilization intensity by age and calendar time, which puts forward the problem of estimating PSA utilization patterns from cancer incidence and mortality data. Even in the USA the patterns have been heterogeneous and showed nontrivial dynamics. Capturing the picture by parametric methods has been very challenging. Although prostate cancer mortality rates have fallen dramatically since the widespread adoption of PSA screening in the early 1990s, conclusively establishing screening benefit requires evidence from randomized controlled trials. Former studies did not formally evaluate whether screening efficacy differed between trials when implementation details such as screening patterns are taken into account and conflicting results have been seen between trials.

In the second chapter, we formulate a joint model of cancer progression to symptomatic (clinical) diagnosis and the screening process with the associated detection mode, as both processes interact to produce the observed incidence in the population. The risks of screening and clinical diagnosis are dependent sharing the latent tumor onset and progression processes in the subject, denoted by a common shared frailty term. Intensity of screening and the hazard driving prostate cancer progression are estimated jointly and semiparametrically using the NPMLE method based on the

joint model. Asymptotic and finite sample properties of the proposed estimators are studied analytically and by simulations. An application using data from the European cancer registry EUREG is presented.

In the third chapter, we develop a semiparametric joint model of cancer progression to clinical and screening diagnosis based on screening trials data with a mixture of known PSA test schedules per protocol and random unknown schedules before and after implementation of the protocol in both control and screening arm. Ad-hoc screening patterns in both arms before recruitment and after existing the trial, and the hazard driving prostate cancer progression are estimated jointly and semiparametrically. Hypothesis tests comparing the screening risks between the arms and periods are performed to validate if the randomization was contaminated. Applications using the subject-specific incidence data for both control and screening arms from Prostate, Lung, Colorectal, and Ovarian screening trial (PLCO) and cancer incidence data from The Surveillance, Epidemiology, and End Results (SEER) Program are demonstrated.

In the fourth chapter, we derive the lead time to link cancer mortality with cancer incidence and screening efficacy. We use a two-step approach to formally test whether screening efficacy differs between trials using mean lead time as a surrogate of screening intensity. First, the mean lead time is estimated in each trial arm as a proxy for the intensity of screening. Second, the association is quantified between the mean lead time and prostate cancer mortality and tested whether it differs between trials while accounting for differences in screening and diagnosis between arms. We analyze the individual-level data from PLCO jointly with SEER US population data to prove that there is no evidence that screening efficacy differed between trials and screening can significantly reduce the risk of prostate cancer death.

CHAPTER I

Introduction

This dissertation is motivated by the problem of interpreting recent screening trials in prostate cancer, the European Randomized Study of Screening for Prostate Cancer (ERSPC) in Europe and Prostate, Lung, Colorectal, and Ovarian cancer screening trial (PLCO) in the United States. Introduction of screening for prostate cancer using the prostate-specific antigen (PSA) biomarker of the disease in the late 80ies led to remarkable dynamics of the incidence of the disease and shortly after, cancer mortality showed a decline (Schroder et al., 2009, 2012, 2014). Except for fragmentary studies, no comprehensive information exists on the PSA uptake in the European countries that would allow specification of utilization intensity by age and calendar time, which puts forward the problem of estimating PSA utilization patterns from cancer incidence and mortality data. Even in the USA the patterns have been heterogeneous and showed nontrivial dynamics. Although prostate cancer mortality rates in the United States have fallen dramatically since the widespread adoption of PSA screening in the early 1990s, conclusively establishing screening benefit requires evidence from randomized controlled trials. Previous studies did not formally evaluate whether screening efficacy differed between trials when implementation details such as screening patterns are taken into account.

The ERSPC and PLCO produced apparently conflicting results, with the ER-

SPC reporting a 21% reduction in prostate cancer mortality (Schroder et al., 2009, 2012, 2014) and the PLCO finding no mortality difference between the trial arms (Andriole, 2009, 2012). Rather than resolving questions, the trials have exacerbated long-standing uncertainty about whether screening benefits cancer survival. Indeed it has been assumed that changes in treatment, rather than PSA screening, largely explain the observed decline in mortality rates (Chou and LeFevre, 2011). It was suggested that ad-hoc screening in the control arms of the trials (contamination) was to blame for the controversial results. Control arms of the trials are thought of as being representative samples from the respective populations inheriting their population screening patterns. However, because subject-level screening schedules in the control arms as well as in the populations they come from are unobserved, adjusting for contamination is impossible without a proper model-based methodology that would allow estimation of the distributional characteristics of an ad-hoc screening utilization process in a population from observed cancer incidence. Devising such methodology and applying it to cancer registry data and trials data is one focus of this dissertation.

In the second chapter, we formulate a joint model of cancer progression to symptomatic (clinical) diagnosis and the screening process and the associated detection mode, as both processes interact to produce the observed incidence in the population. The risks of screening and clinical diagnosis are dependent sharing the latent tumor onset and progression processes in the subject. The model is formulated that treats both risks semiparametrically on two time scales, age t and calendar time (year) y. Because the model is developed under the premise of unobserved screening schedules, the screening tests are modelled as an unobserved non-homogeneous Poisson process N_{scr}, with intensity h_{scr} that depends on age and calendar time. The model for the observed data is an average over the unobserved onset time (S), the screening process (N_{scr}), and the mode of diagnosis (SDx vs. CDx). This set of unobserved

random variables and stochastic processes constitutes a complex shared "frailty" object that explains the dependence between the competing risks of diagnosis. Intensity of screening and the hazard driving prostate cancer progression are estimated jointly and semiparametrically using the NPMLE method based on the joint model. Asymptotic and finite sample properties of the proposed estimators are studied analytically and by simulations. An application using data from the European cancer registry EUREG is presented.

In the third chapter, we develop a semiparametric joint model of cancer progression to clinical and screening diagnosis based on screening trials data with a mixture of known PSA test schedules per protocol and random unknown schedules before and after implementation of the protocol as well as in the control arm. We analyze the subject-specific incidence data for both control and screening arms from PLCO. Patients in the control arm were screened following population patterns while those in the screening arm were recruited and screened for 6 years according to a specific schedule. The model is formulated under the premise of unobserved screening schedules (population patterns) for subjects in the control arm and subjects in the screening arm when they are off trials, with specific screening schedules N_{scr}^{sc} following protocols during trials. The population screening patterns are modeled as an unobserved nonhomogeneous Poisson process N_{scr}^{c}, with intensity h_{scr} that depends on age following the longitudinal study. For both arms, subjects are assumed to be screened with the same intensity as the population h_{scr} before they enter trials due to randomization. For subjects in the control arm during the trial, the intensity may change if recruitment into the trial has an effect on subjects' screening patterns (Gulati et al., 2012; Pinsky et al., 2010) and the intervention effect persists after subjects exiting the trial. Hence we set the intensity as $r_1 h_{scr}$ with the risk ratio r_1. For subjects in the screening arm during the trial, the test schedules are discrete and fixed and they return to random after exiting the trial with the intensity $r_2 h_{scr}$, where the risk

ratio r_2 models the difference v.s. the population intensity. The screening risk ratio r_3 models the difference in screening utilizations between screening arm and control arm after the trial. In addition, we incorporate incidence data from SEER with control arm in the absence of screening and intervention arm with screening intensity as $r_4 h_{scr}$. Ad-hoc screening patterns in both arms before recruitment and after existing the trial, and the hazard driving prostate cancer progression are estimated jointly and semiparametrically. Hypothesis tests comparing the screening intensities between the arms and periods using the risk ratios are performed to validate if the contamination exists. Applications using incidence data from PLCO and SEER are demonstrated.

In the fourth chapter, we derive the mean lead tim (MLT)s to link cancer mortality to cancer incidence and screening efficacy. The lead time is the amount of time to cancer diagnosis advanced due to screening, which is a counterfactual concept. We use a two-step approach to formally test whether screening efficacy differs between trials with mean lead time. First, the mean lead time is estimated in each trial arm as a proxy for the intensity of screening. We estimate the MLTs empirically, without any model assumptions about cancer progression and diagnosis. The empirical approach estimated the MLTs by calculating the differences between survival curves for time from randomization to diagnosis in each trial arm relative to an assumed baseline level. Additionally, one analytic model (UMICH) and two simulation models (FHCRC and MISCAn) estimated distributions of age at onset of latent disease and diagnosis in the absence and presence of screening. The fitted model then estimated MLTs as in the empirical approach but using projected instead of observed incidence data. Second, the association is quantified between the mean lead time and prostate cancer mortality and tested whether it differs between trials while accounting for differences in screening and diagnosis between arms. We analyze the individual-level data from PLCO jointly with SEER US population data to prove the benefits of screening in terms of reducing the risk of prostate cancer death.

CHAPTER II

A Joint Model of Cancer Incidence and Screening

2.1 Introduction

Prostate cancer is the most common cancer in men presenting a significant public health problem. Since the introduction of prostate specific antigen (PSA) screening the incidence rates of newly diagnosed prostate cancers have seen a dramatic increase in Europe (Figure 2.1) that followed country-specific PSA utilization uptake. Shortly after, cancer mortality showed a decline (Schroder et al., 2009, 2012, 2014). To study cancer incidence patterns induced by screening Tsodikov et al. (2006) developed a parametric model of prostate cancer in the US population, relying on PSA utilization patterns obtained from external studies. Except for fragmentary studies, no direct data exist on the PSA uptake in the European countries. Because PSA utilization in a population is a strong factor potentially confounding the results of survival and screening trials (Lee and Tsodikov, 2013; Gulati et al., 2012), its estimation is an important problem. In this chapter we propose a model that allows us to estimate screening utilization indirectly from cancer incidence data.

The screening utilization patterns observed in the USA using claims and survey data (Mariotto et al., 2007) have been heterogeneous and showed nontrivial dynamics, which makes their parametric specification for a different population quite challenging if not impossible. Also, parametric assumptions related to latent processes are difficult

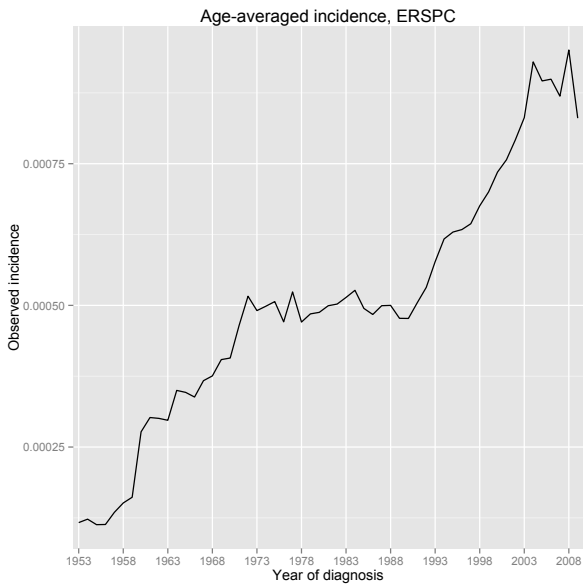

Figure 2.1: Averaged prostate cancer incidence of key countries enrolled in the European Randomized Study of Screening for Prostate Cancer (ERSPC), Finland, Italy, Spain, Belgium, Netherlands, Sweden, Switzerland. Data from the European Cancer Registry (EUREG).

to verify for lack of direct information in the observed data. Therefore, we set on developing a model where the basic dynamic patterns of disease progression and screening in the population are specified nonparametrically to avoid bias and be flexible enough to reproduce nontrivial relationships.

The classical model of cancer natural history describes irreversible transitions through three consecutive stages: disease-free stage, pre-clinical stage (asymptomatic) and clinical stage (symptomatic) (Zelen and Feinleib, 1969). The preclinical period in the absence of screening is called the sojourn time, while the clinical period is survival post-diagnosis. The end of the disease-free stage is marked by the unobserved event of tumor onset at the age of S. The end of the preclinical stage is marked by the event of diagnosis. Cancer incidence in the subject is a result of risk competition of two modes (causes) of diagnosis, one due to symptoms of the disease (clinical diagnosis, CDx), and one due to the screening process (screening diagnosis, SDx). The risks are dependent as neither of them exists before the unobserved shared tumor onset event, denoted by a shared frailty term A.

In this chapter, a joint model is formulated that treats both risks semiparametrically on two time scales, age t and calendar time (year) y. Because the model is developed under the premise of unobserved screening schedules, the screening tests are modelled as an unobserved non-homogeneous Poisson process N_{scr}, with intensity h_{scr} that depends on age and calendar time. The model for the observed data is an average over the unobserved onset time (S), the screening process (N_{scr}), the shared frailty term A and the mode of diagnosis (SDx vs. CDx). This set of unobserved random variables and stochastic processes constitutes a complex shared "frailty" object that explains the dependence between the competing risks of diagnosis.

In Section 2 we define the model and derive its essential distributional characteristics. In Section 3 we derive the likelihood function and the nonparametric maximum likelihood estimators using an iterative reweighting algorithm (Chen, 2009), and study its asymptotic properties. We apply the method to cancer registry data in Section 4, and perform a simulation study of the properties of estimators in finite samples. Finally, we discuss the results in Section 5.

2.2 Statistical Framework

2.2.1 The natural history model without screening

To specify the model in the absence of screening we follow Rice and Tsodikov (2016) who introduced a semi-parametric joint model of time to terminal event affected by a latent progression event. The idea is similar to the parametric model by Dejardin et al. (2010). We identify the latent event with tumor onset at the age of S, and the terminal event with diagnosis at the age of T. By definition, the latent event must precede the terminal one: $S \leq T$. The hazard rate associated with the r.v. T, $d\Lambda(t|x) = \lambda(t|x)dt$, where t is age at diagnosis (CDx), Λ is the cumulative hazard, λ is its instantaneous counterpart, and x is the birth year ($x = y - t$), defines

cancer incidence on both time scales (t, y). The time to the latent onset event S follows the Cox model $d\Lambda_0(t|Z_{tx})$, given covariates Z_{tx}, with predictor $\theta(Z_{tx})$, and the baseline cumulative hazard $H_0(t)$. The terminal event (CDx) follows the Cox model $d\Lambda_1(t|Z_{tx}, S)$ with predictor $\eta(Z_{tx})$ and the baseline cumulative hazard $H_1(t)$, constrained to show zero risk before the latent event by means of a multiplicative indicator process $I(t \geq S)$ that depends on the latent r.v. S.

$$d\Lambda_0(t|Z_{tx}) = \lim_{\epsilon \to 0} \frac{\Pr(S \in [t, t+\epsilon] | S \geq t, Z_{tx})}{\epsilon} = \theta(Z_{tx}) dH_0(t), \tag{2.1}$$

$$d\Lambda_1(t|Z_{tx}, S) = \lim_{\epsilon \to 0} \frac{\Pr(T \in [t, t+\epsilon] | T \geq t, S, Z_{tx})}{\epsilon} = I(t \geq S)\eta(Z_{tx}) dH_1(t). \tag{2.2}$$

Here I is an indicator function such that $I(t \geq s) = 1$ if $t \geq s$, and zero otherwise. The relative relationship between θ and η governs the duration of the sojourn time. Exponential parameterization of predictors using regression coefficients $\boldsymbol{\beta} = \left(\beta_0, \boldsymbol{\beta}'_\eta, \boldsymbol{\beta}'_\theta\right)'$ gives $\eta = \eta(\boldsymbol{\beta}; Z_{tx}) = e^{\beta_0 + Z'_{tx}\boldsymbol{\beta}_\eta}$ and $\theta = \theta(\boldsymbol{\beta}; Z_{tx}) = e^{Z'_{tx}\boldsymbol{\beta}_\theta}$. In the sequel we may suppress some of the arguments for brevity, for example writing $\eta(t)$ for $\eta(\boldsymbol{\beta}; Z_{tx})$.

Generally, when the conditional hazard function for a survival time T is a stochastic process $\lambda(t)$, then the marginal survival function $G(t) = \mathbb{E}[e^{-\int_0^t \lambda(\xi) d\xi}]$, where the expectation is taken over the random trajectory $\bar{\lambda}(t)$ of the process λ from 0 to t. Here and in the sequel the notation $\bar{F}_t$ will indicate the trajectory of a function F on $[0, t]$ as opposed to a single value $F(t)$ at the point t. Extending the formulation of Rice and Tsodikov (2016) to time-dependent predictors, we obtain the marginal diagnosis-free survival function in the form (see more derivation in Appendix A.1)

$$\begin{aligned} G_d(t|\bar{Z}_{tx}) &= \mathbb{E}\left[e^{-\int_0^t d\Lambda_1(\xi|Z_{\xi x}, S)}\right] \\ &= e^{-\int_0^t \theta(\xi) dH_0(\xi)} + \int_0^t e^{-\int_s^t \eta(\xi) dH_1(\xi) - \int_0^s \theta(\xi) dH_0(\xi)} \theta(s) dH_0(s). \end{aligned} \tag{2.3}$$

The first term in the above expression is the probability of no onset before t. The

integral of the second term averages the probability of no diagnosis during the sojourn time $[s,t]$ over the pdf of onset at time s.

2.2.2 The joint model with screening

The screening schedule in the subject is summarized by the $[0,t]$-trajectory $\bar{N}_{scr}(t|x)$ of the counting process $N_{scr}(t|x)$ that counts the number of screens performed on the subject before the age of t. The dependence on the birth cohort x is needed to account for variable utilization of screening over calendar time and age. Let $\alpha(t|x)$ be the sensitivity of a screening test, that is the probability of cancer detection by a screening test performed at age t, given that cancer is detectable ($t > S$). Given a subject's screening schedule at times τ_i, the the potential time to SDx survival function becomes

$$\begin{aligned} G_{SDx}(t|\bar{Z}_{tx}, S, \bar{N}_{scr}(t|x)) &= \prod_{i:S<\tau_i\leq t} (1-\alpha(\tau_i|x)) \\ &= \exp\left[\int_0^t I(\xi \geq S)\log(1-\alpha(\xi|x))dN_{scr}(\xi|x)\right]. \end{aligned} \tag{2.4}$$

We introduce the shared frailty term A to explain the dependence between the two modes of diagnosis in the same subject. The corresponding conditional hazard functions given tumor onset in terms of two modes of diagnosis are:

$$\begin{aligned} &\text{Time to CDx} : d\Lambda_1(t|Z_{tx}, S, A) = AI(t \geq S)\eta(Z_{tx})dH_1(t), \\ &\text{Time to SDx} : d\Lambda_2(t|Z_{tx}, S, A, N_{scr}) = -AI(t \geq S)\log(1-\alpha(t|x))dN_{scr}(t|x). \end{aligned} \tag{2.5}$$

Assuming that screen counts follow a nonhomogeneous Poisson process with intensity $h_{scr}(\xi|x)$, we have $N_{scr}(t|x) \sim \text{Poisson}\left(\int_0^t h_{scr}(\xi|x)d\xi\right)$. Using the Laplace functional

of a Poisson process with intensity h_{scr} (Serfozo (2009); Shiryaev (1960))

$$\mathcal{L}(\varphi) = \mathbb{E}\left\{e^{-\int_0^t \varphi(x)dN_{scr}(t)}\right\} = \exp\left\{-\int_0^t (1-e^{-\varphi(\xi)})dH_{scr}(\xi)\right\}, \tag{2.6}$$

for any integrable functional argument φ. We obtain the marginal survival function incorporating risks of screening and clinical diagnosis as the expectation over the distribution of latent event time S, frailty term A and screening pattern N_{scr}. Let $\Lambda_0(t) = \int_0^t \theta(Z_{\xi x})dH_0(\xi), \tilde{\Lambda}_{s,t} = \int_0^t I(\xi \geq s)\left[\eta(Z_{\xi x})dH_1(\xi) + \alpha(\xi|x)dH_{scr}(\xi|x)\right], \mathcal{L}^{(k)}(A)$ be the k_{th} derivative of the Laplace transform of A. Then we have

$$\begin{aligned} G_d(t|\bar{Z}_{tx}) &= \mathbb{E}_{S,A,N_{scr}}\left[e^{-\int_0^t \left(d\Lambda_1(\xi|Z_{\xi x},S,A)+d\Lambda_2(\xi|Z_{\xi x},S,A,N_{scr})\right)}\right] \\ &= e^{-\Lambda_0(t)} + \int_0^t \mathcal{L}^{(0)}(\tilde{\Lambda}_{s,t})e^{-\Lambda_0(s)}d\Lambda_0(s). \end{aligned} \tag{2.7}$$

Note that the above expression depends on the screening pattern only through the product $\alpha(\xi|x)dH_{scr}(\xi|x)$, so that the screening sensitivity is not identifiable jointly with the screening intensity. From now on we will therefore assume that H_{scr} absorbs α.

From (2.7) we have the marginal hazard function (see more derivation in Appendix A.1)

$$d\Lambda(t|\bar{Z}_{tx}) = -\frac{dG_d(t|\bar{Z}_{tx})}{G_d(t|\bar{Z}_{tx})} = \Psi(t|\bar{Z}_{tx}) \times \left[\eta(t)dH_1(t) + dH_{scr}(t|x)\right], \tag{2.8}$$

where

$$\Psi(t|\bar{Z}_{tx}) = \frac{-\int_0^t \mathcal{L}^{(1)}(\tilde{\Lambda}_{s,t})e^{-\Lambda_0(s)}d\Lambda_0(s)}{e^{-\Lambda_0(t)} + \int_0^t \mathcal{L}^{(0)}(\tilde{\Lambda}_{s,t})e^{-\Lambda_0(s)}d\Lambda_0(s)}, \tag{2.9}$$

and $\theta(t) = e^{Z'_{tx}\boldsymbol{\beta}_\theta}, \eta(t) = e^{\beta_0 + Z'_{tx}\boldsymbol{\beta}_\eta}$. The function Ψ summarizes the dependence between times to the two potential points of diagnosis (SDx and CDx), and a departure from additive independent risks expressed by the multiplier to the Ψ in (2.8).

2.2.3 Censoring Mechanism

In our model, there are latent (tumor onset) and terminal (cancer diagnosis) events. The time to the latent event S is never observed. By definition above, the latent event must precede the terminal one: $S \leq T$. There is a censoring time T_c that is independent of S and T, given covariates Z. We observe (T^*, Δ, Z), where $T^* = \min(T, T_c)$ and $\Delta = I(T^* = T)$. When $\Delta = 1$, we have $S \leq T \leq T_c$; otherwise when $\Delta = 0$, then either $T_c \leq S$ or $S \leq T_c \leq T$. Thus under the censoring mechanism we are unable to tell from observed data whether or not the latent event has occurred.

2.3 Estimation

2.3.1 Likelihood

Given the maximum follow up time υ, we write the log-likelihood as

$$\ell = \sum_x \int_0^\upsilon \sum_{i \in \mathcal{R}(t|x)} \log\left(d\Lambda(t|\bar{Z}_{txi})\right) dN(t|\bar{Z}_{txi}) - d\Lambda(t|\bar{Z}_{txi}), \tag{2.10}$$

where $\mathcal{R}(t|x)$ is the set of subjects at risk for diagnosis of x-birth cohort at age t, and $dN(t|\bar{Z}_{txi})$ is the indicator that subject i from the same risk set was diagnosed at t. The above likelihood recognizes the life-table-type cross-sectional data structure resulting from observations of a population at risk of the disease along with the counts of diagnosis coming from the population, over a period of time. The probability for a subject at risk to get a diagnosis over the next small period of time dt is $d\Lambda$, which makes the first term of the conditional likelihood. The second term is a log of the probability to survive without diagnosis for such subject, $\log(1 - d\Lambda) = -d\Lambda$, where the differentials are understood as first order terms with respect to dt.

With the appropriate filtration $\mathcal{F}_{t-}$, for the subject $i \in \mathcal{R}(t|x)$, we construct the

orthogonal martingale (at the true model):

$$dM(t|\bar{Z}_{txi}) = dN(t|\bar{Z}_{txi}) - d\Lambda(t|\bar{Z}_{txi}), \quad \mathbb{E}\left[dN(t|\bar{Z}_{txi})\middle|\, \mathcal{F}_{t-}\right] = d\Lambda(t|\bar{Z}_{txi}).$$

2.3.2 Nonparametric Maximum Likelihood Estimation

2.3.2.1 Functional derivative

Denote the functional derivatives of $\Psi(t|\bar{Z}_{tx})$ from (2.9) w.r.t. $dH(\tau)$ (dH_0, dH_1) and $dH_{scr}(\tau|x)$ as

$$\Psi_{dH(\tau)}(t|\bar{Z}_{tx}) = \frac{\partial \Psi(t|\bar{Z}_{tx})}{\partial dH(\tau)}, \tag{2.11}$$

$$\Psi_{dH_{scr}(\tau|x)}(t|\bar{Z}_{tx}) = \frac{\partial \Psi(t|\bar{Z}_{tx})}{\partial dH_{scr}(\tau|x)}, \tag{2.12}$$

respectively, where the functional derivatives are defined as follows. For a functional $J(f),\ f = f(u)$, the functional derivatives in (2.11) and (2.12) are defined as

$$\frac{\partial J(f)}{\partial df(s)} = \left.\frac{\partial J(f+\epsilon g)}{\partial \epsilon}\right|_{\epsilon=0, g=I(u>s)}$$

(see Hu and Tsodikov, 2014a, Section 3.2) and correspond to taking the derivative with respect to a "jump" of f at time t, where f can be discrete or continuous. For a linear functional of the form $J(f) = \int_0^t \varphi(u)\, df(u)$, the functional derivative is

$$\frac{\partial J}{\partial df(s)} = \int_0^t \varphi(u)\, d\left.\frac{\partial (f+\epsilon g)}{\partial \epsilon}\right|_{\epsilon=0, g=I(u>s)} = \int_0^t \varphi(u)\, dI(u>s) = \varphi(s) I(t \geq s).$$

Using this definition, we have

(i) $\frac{\partial H(t)}{\partial dH(s)} = \frac{\partial H_{scr}(t|x)}{\partial dH_{scr}(s|x)} = I(t \geq s)$,

(ii) $\frac{\partial \log dH(t)}{\partial dH(s)} = \frac{1}{dH(t)} \frac{\partial dH(t)}{\partial dH(s)} = \frac{1}{dH(t)} dI(t \geq s)$,

(iii) $\frac{\partial \log dH_{scr}(t|x)}{\partial dH_{scr}(s|x)} = \frac{1}{dH_{scr}(t|x)} \frac{\partial dH_{scr}(t|x)}{\partial dH_{scr}(s|x)} = \frac{1}{dH_{scr}(t|x)} dI(t \geq s)$.

2.3.2.2 Score equations and asymptotic properties

Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_\theta, \boldsymbol{\beta}_\eta, \boldsymbol{\beta}_\phi)$ be the parameters with onset, clinical diagnosis, and frailty parts of the model, respectively. Given the maximum follow-up time υ, differentiate the log-likelihood we arrive at the following score equations for $dH_0(\tau), dH_1(\tau), dH_{scr}(\tau|x)$ and $\boldsymbol{\beta}$ (See Appendix A.2 for details).

$$\mathcal{U}_{dH_0(\tau)} = \sum_x \int_\tau^\upsilon \sum_{i \in \mathcal{R}(t|x)} \frac{\partial \log d\Lambda(t|\bar{Z}_{txi})}{\partial dH_0(\tau)} dM(t|\bar{Z}_{txi}), \tag{2.13}$$

$$\mathcal{U}_{dH_1(\tau)} = \sum_x \sum_{i \in \mathcal{R}(\tau|x)} \frac{\eta_i(\tau) dN(\tau|\bar{Z}_{\tau xi})}{\eta_i(\tau) dH_1(\tau) + dH_{scr}(\tau|x)} - \sum_{i \in \mathcal{R}(\tau|x)} \Psi(\tau|\bar{Z}_{\tau xi}) \eta_i(\tau) w_{dH_1(\tau)}, \tag{2.14}$$

where $\tau \in (0, t]$ and

$$w_{dH_1(\tau)} = 1 - \frac{\int_\tau^\upsilon \sum_{i \in \mathcal{R}(t|x)} \frac{\Psi_{dH(\tau)}(t|\bar{Z}_{txi})}{\Psi(t|\bar{Z}_{txi})} dM(t|\bar{Z}_{txi})}{\sum_{i \in \mathcal{R}(\tau|x)} \Psi(\tau|\bar{Z}_{\tau xi}) \eta_i(\tau)}, \quad \eta_i(\tau) = \eta(\tau|\boldsymbol{\beta}; Z_{\tau xi}). \tag{2.15}$$

For the screening intensity dH_{scr} of x-birth cohort,

$$\mathcal{U}_{dH_{scr}(\tau|x)} = \sum_{i \in \mathcal{R}(\tau|x)} \frac{dN(\tau|\bar{Z}_{\tau xi})}{\eta_i(\tau) dH(\tau) + dH_{scr}(\tau|x)} - \sum_{i \in \mathcal{R}(\tau|x)} \Psi(\tau|\bar{Z}_{\tau xi}) w_{dH_{scr}(\tau|x)}, \tag{2.16}$$

where

$$w_{dH_{scr}(\tau|x)} = 1 - \frac{\int_\tau^\upsilon \sum_{i \in \mathcal{R}(t|x)} \frac{\Psi_{dH_{scr}(\tau|x)}(t|\bar{Z}_{txi})}{\Psi(t|\bar{Z}_{txi})} dM(t|\bar{Z}_{txi})}{\sum_{i \in \mathcal{R}(\tau|x)} \Psi(\tau|\bar{Z}_{\tau xi})}. \tag{2.17}$$

The score for $\boldsymbol{\beta}$ is

$$\mathcal{U}_{\boldsymbol{\beta}} = \sum_x \int_0^\upsilon \sum_{i \in \mathcal{R}(t|x)} \frac{\partial \log d\Lambda(t|\bar{Z}_{txi})}{\partial \boldsymbol{\beta}} dM(t|\bar{Z}_{txi}). \tag{2.18}$$

It can be shown that the score functions for H_0, H_1, H_{scr} are martingales at the true model (see Appendix A.3.1 for details). For the weighted Breslow-type estimators of dH_1 and dH_{scr}, we note that the weights $w = (w_{dH_1}, w_{dH_{scr}})$ have expectation of 1, given filtration $\mathcal{F}_{t\_}$. In particular, the martingale estimation equations (EE) solution can be obtained by setting all weights to 1. The EE approach provides estimators that are consistent, computationally fast, yet not fully efficient (Hu and Tsodikov, 2014). The efficiency of the NPMLE is due to the fact that optimal weights depend on martingale residuals utilizing available information in the future of the subject.

We adapt the Weighted Breslow Estimator algorithm (Chen, 2009) to maximize the likelihood by iteratively updating the weights. Following Rice and Tsodikov (2016), we impose a proportional hazards (PH) assumption in the example. Let $dH_0 = dH_1 = dH$, meaning that subject shares the same baseline hazard between tumor onset and cancer detection by symptoms (CDx). With richer population-level covariates, independent specification of these baseline hazards would perhaps be a better choice. With the weights treated as known at the inner loop of the algorithm, a set of recurrent score equations emerges similar to the computationally efficient martingale estimating equations (Chen et al., 2002) in the EE approach. An alternative computation method is the EM algorithm of Tsodikov (2003) that requires substantial theoretical development for the current model. Given initial values for $\boldsymbol{\beta}$ and initial weights $w^{(0)}$ as 1, we use Nelson-Aalen estimator as initial values for dH and dH_{scr}. For iteration count $k = 0, 1, ...$, we repeat the following steps until convergence to maximize the likelihood over the two hazards and obtain the profile likelihood $\ell_{pr}(\boldsymbol{\beta})$:

1. Fix weights $w^{(k)}$ and given $(dH^{(k)}, dH_{scr}^{(k)})$, obtain the solution $(dH^{(k+1)}, dH_{scr}^{(k+1)})$ from the score equations for dH and dH_{scr}.

2. Update the weights $w^{(k+1)}$ using (2.15) and (2.17) with $(dH^{(k+1)}, dH_{scr}^{(k+1)})$.

Maximization of the profile likelihood $\ell_{pr}(\boldsymbol{\beta})$, obtained using the above algorithm, using general maximization methods such as conjugate gradients, we arrive at the final MLEs.

Taking derivatives of the score equations w.r.t $\Omega = (\boldsymbol{\beta}, dH, dH_{scr})$ and plugging the converged estimators $\hat{\Omega}$, we obtained the observed information matrix $\mathcal{J}(\hat{\Omega})$.

$$\mathcal{J}(\hat{\Omega}) = -\begin{pmatrix} \frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} & \frac{\partial^2}{\partial\boldsymbol{\beta}\partial dH'} & \frac{\partial^2}{\partial\boldsymbol{\beta}\partial dH'_{scr}} \\ \frac{\partial^2}{\partial dH\partial\boldsymbol{\beta}'} & \frac{\partial^2}{\partial dH\partial dH'} & \frac{\partial^2}{\partial dH\partial dH'_{scr}} \\ \frac{\partial^2}{\partial dH_{scr}\partial\boldsymbol{\beta}'} & \frac{\partial^2}{\partial dH_{scr}\partial dH'} & \frac{\partial^2}{\partial dH_{scr}\partial dH'_{scr}} \end{pmatrix} \ell(\Omega)\Bigg|_{\Omega=\hat{\Omega}}. \tag{2.19}$$

Inverting $\mathcal{J}(\hat{\Omega})$ the we can obtain the estimated standard errors of $\hat{\Omega}$ and construct the confidence bands.

Consistency of the estimators is proved by empirical processes following Zeng and Lin (2007); Kosorok (2008); Hu and Tsodikov (2014b). Weak convergence is proved using the martingale structure of the score equations following Chen (2009); Hu and Tsodikov (2014b). Denote the true value of the set of model parameters Ω by Ω^0. Under regularity conditions, we have the following propositions (see Appendix A.3 for proof):

Theorem II.1. *$\hat{\boldsymbol{\beta}}$ converges to $\boldsymbol{\beta}^0$, $\hat{H}(\cdot)$ converges to $H^0(\cdot)$, $\hat{H}_{scr}(\cdot|x)$ converges to $H^0_{scr}(\cdot|x)$ uniformly in probability in the interval $[0, \upsilon]$.*

Theorem II.2. *$n^{1/2}\left[\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right)', \hat{H}(t) - H^0(t), \hat{H}_{scr}(t|x) - H^0_{scr}(t|x)\right]'$ converges weakly to a zero-mean Gaussian process whose covariance function structure is given in Appendix A.3.*

2.4 Example and Simulation Study

2.4.1 EUREG data analysis

EUREG permits the exploration of geographical patterns and temporal trends of incidence, mortality and survival observed in European population-based cancer registries for 35 major cancer entities in about 100 registration areas (EUREG, 2012). We use the incidence data with cancer cases C and population P corresponding to age interval [50,89] and calendar year interval [1953-2009] for key countries enrolled in the European Randomized Study of Screening for Prostate Cancer (ERSPC), Finland, Italy, Spain, Belgium, Netherlands, Sweden, Switzerland. Incidence of prostate cancer before the age of 50 is negligibly small.

For the shared frailty A, let $A \sim \Gamma(\frac{1}{\phi(Z_{tx})}, \phi(Z_{tx}))$, where $\phi(Z_{tx}) = e^{Z'_{tx}\boldsymbol{\beta}_\phi}$ that represents a gamma frailty model, a common choice in survival analysis with dependent data. Under this specification, from (2.9) we have $\mathcal{L}^{(0)}(s) = (1+\phi s)^{-\frac{1}{\phi}}, \mathcal{L}^{(1)}(s) = -(1+\phi s)^{-\frac{1}{\phi}-1}, \mathcal{L}^{(2)}(s) = (1+\phi)(1+\phi s)^{-\frac{1}{\phi}-2}$. In the absence of covariates affecting the tumor onset event, we set $\theta = 1$ and $\phi(Z_{tx}) = e^{\beta_\phi}$. Before the introduction of PSA screening in the late 80ies, the incidence of prostate cancer had an increasing trend in calendar time y (Figure 2.1), reportedly partially due to surgical treatment of benign prostate enlargement (the TURP treatment) (Merrill et al., 1999). To model this effect, a general linear model is specified for $\eta(\boldsymbol{\beta}; Z_{tx})$ during [1953,1989], saturating in the year 1989 when the PSA screening test was introduced and the use of TURP quickly receded,

$$\eta(\boldsymbol{\beta}; Z_{tx}) = \begin{cases} 1, & y < 1953 \\ e^{\beta_1+\beta_2(y-1953)}, & 1953 \le y < 1989 \\ e^{\beta_1+\beta_2(1989-1953)}, & y \ge 1989. \end{cases} \tag{2.20}$$

With the calendar year y as the only covariate, $\bar{Z}_{txi}$ does not depend on i, given t

and x. Let $P(t|\bar{Z}_{tx}) = \left\|\mathcal{R}(t|x)\right\|$, denote the size of the risk set for cohort x at age t (population) and let $C(t|\bar{Z}_{tx}) = \sum_{i\in\mathcal{R}(t|x)} dN_i(t|\bar{Z}_{txi})$ be the count of cancer cases. The score equations become

$$dH(\tau) = \frac{C(t|\bar{Z}_{\tau x})}{P(\tau|\bar{Z}_{\tau x})\Psi(\tau|\bar{Z}_{\tau x})\eta(\tau)w_{dH(\tau)}} - \frac{dH_{scr}(\tau|x)}{\eta(\tau)}, \tag{2.21}$$

$$dH_{scr}(\tau|x) = \frac{C(t|\bar{Z}_{\tau x})}{P(\tau|\bar{Z}_{\tau x})\Psi(\tau|\bar{Z}_{\tau x})w_{dH_{scr}(\tau|x)}} - \eta(\tau)dH(\tau). \tag{2.22}$$

Setting $\boldsymbol{\beta}' = (\beta_1, \beta_2, \beta_\phi) = (0.5, 0.5, 0.5)$ initially, and following the algorithm we obtain the MLEs $\hat{\boldsymbol{\beta}}, d\hat{H}$ and $d\hat{H}_{scr}$. Figure 2.2 and Figure 2.3 display the baseline hazard h driving prostate cancer diagnosis in the absence of screening, and the screening intensity h_{scr} with the 95% confidence interval for the age 70 group by year of diagnosis. The wider band for the screening intensity after 2007 is because of smaller available data for this period. Estimated regression coefficients are $\hat{\boldsymbol{\beta}}' = (0.30, 0.20, -0.01)$, with standard errors $\mathrm{SE}(\hat{\boldsymbol{\beta}}') = (0.009, 0.006, 0.12)$.

Figure 2.4 and Figure 2.5 display the observed cancer incidence and the predicted one with PSA screening. Before year 1989, the year trend is modeled through $\boldsymbol{\beta}$ parametrically. The prediction in this period is also driven by the nonparametrically specified H. After 1989, the screening intensity matrix parameter h_{scr} nonparametrically specified by age and calendar time comes into play, and the expected incidence depends on both age and year semi-parametrically matching the observed incidence.

2.4.2 Simulation study

To assess the finite-sample properties of the parameter obtained by the proposed methodology, we perform a simulation study. In particular, we verify that the variance of modeled parameters can be obtained using the observed information matrix $\mathcal{J}$ with reasonable accuracy. The simulation is conditional on population counts

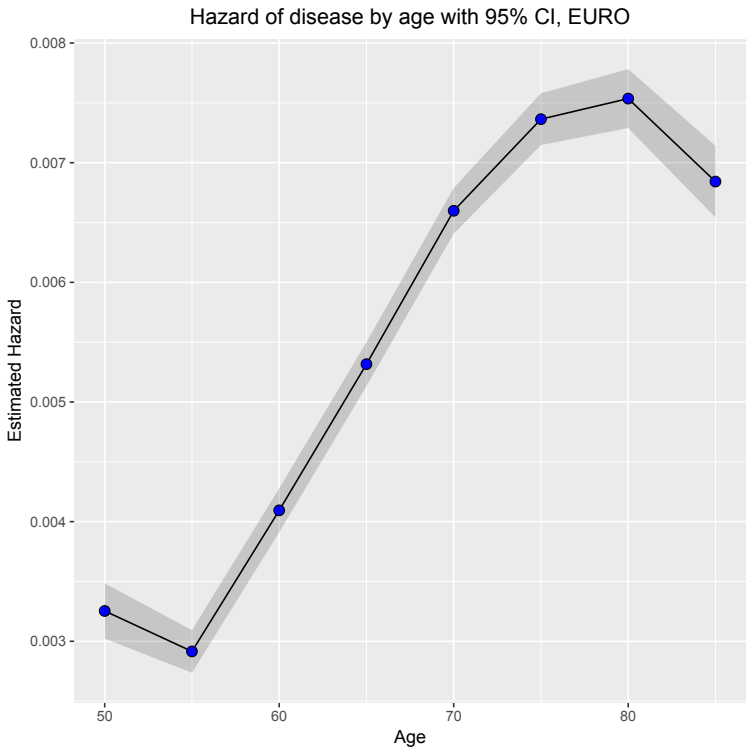

Figure 2.2: Baseline hazard of disease driving diagnosis dH with 95%CI

sampling cancer incidence from the population at risk using Bernoulli trials with probability $d\Lambda$. For each dataset k we obtain the NPMLE estimators $\hat{\Omega}^{(k)}$ and the observed information matrix $\mathcal{J}^{(k)}$, $k = 1, 2, ..., n$. For large n, $\frac{1}{n-1}\sum_{k=1}^{n}(\hat{\Omega}^{(k)} - \bar{\Omega})^2 \approx \frac{1}{n}\sum_{k=1}^{n} \text{diag}\left[(\mathcal{J}^{(k)})^{-1}\right]$. Table 1 and Table 2 present the simulation results in terms of $\boldsymbol{\beta}, dH$ and dH_{scr}, based on 200 replicates.

We note that $\hat{\beta}_\phi$ that describes the frailty term is characterized by much higher variance than the rest of β, which is generally typical of parameters characterizing latent model quantities. Same is true regarding the baseline hazard of disease driving diagnosis dH vs. the intensity of PSA screening dH_{scr}, given that PSA schedules are not observed in the dataset. The PH assumption linking dH_1 and dH_0 into a common dH allowed for the information from observed diagnoses to bear on the latent hazard of tumor onset. Regarding bias, again we see manifestation of the same effect with parameters in the latent parts of the model showing larger finite sample bias.

Table 2.1: Simulation results: Frailty and year trend before screening $\boldsymbol{\beta}$ and hazard of disease driving diagnosis dH by age.

Parameters	True	Bias	ESE	ASE
$\boldsymbol{\beta}$				
β_1	0.22	0.01	0.009	0.01
β_2	0.19	0.00	0.006	0.004
β_ϕ	-0.01	0.04	0.12	0.13
dH				
55	0.00278	0.00001	9.1×10^{-5}	9.7×10^{-5}
60	0.00390	0.00003	9.4×10^{-5}	1.3×10^{-4}
65	0.00506	0.00004	9.4×10^{-4}	1.1×10^{-4}
70	0.00631	0.00005	9.7×10^{-4}	9.7×10^{-5}
75	0.00705	0.00005	1.1×10^{-4}	1.0×10^{-4}
80	0.00724	0.00005	1.3×10^{-4}	1.2×10^{-4}
85	0.00657	0.00003	1.5×10^{-4}	1.4×10^{-4}

ESE: Empirical Standard Errors; ASE:Aysmptotic Standard Errors

Table 2.2: Simulation results: Screening intensity by calendar time dH_{scr} for age=70.

Year	True	Bias	ESE	ASE
1995	0.0032	0.0001	0.00067	0.00060
1999	0.0058	0.0007	0.00065	0.00061
2005	0.0123	0.0020	0.00068	0.00062
2009	0.0095	0.0023	0.00492	0.00451

ESE: Empirical Standard Errors; ASE:Aysmptotic Standard Errors

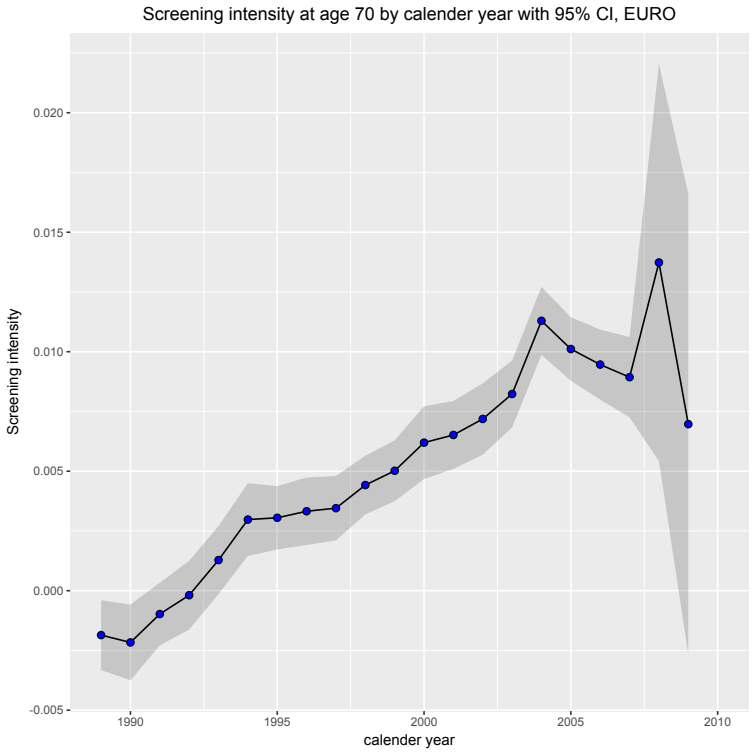

Figure 2.3: Screening intensity dH_{scr} for age 70 with 95%CI

2.5 DISCUSSION

The model we proposed provides a quantitative link between dissemination of cancer control processes with unknown schedules and their impact on population and public health measures of cancer incidence. We quantify the relationships that underlie recent trends in prostate cancer incidence in terms of model parameters $(\boldsymbol{\beta}, dH, dH_{scr})$ and perform inference on these parameters. The model can estimate screening intensity in a population jointly with the disease natural history and generate predictions for prostate cancer incidence under a variety of PSA screening patterns and for the case of no PSA screening ($H_{scr} = 0$).

The parameters in our model are estimated from population databases (cancer registries) in the most challenging situation when neither screening schedules nor the mode of diagnosis (screening vs. symptoms) are observed. We incorporate random PSA schedules into the estimation procedures. The model provides a basis for assess-

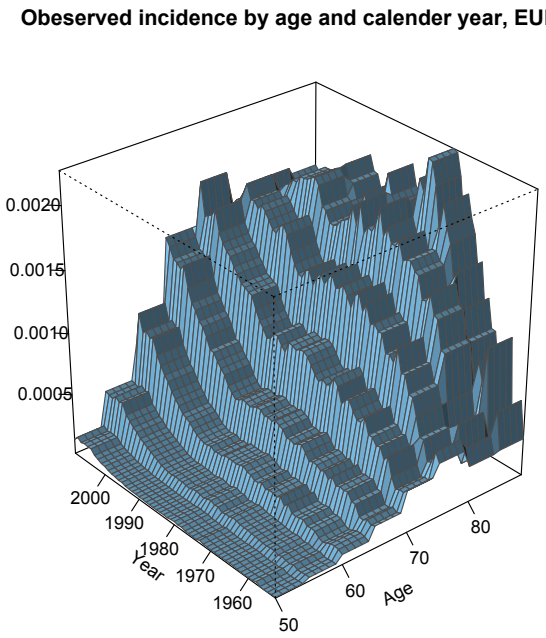

Figure 2.4: Observed prostate cancer incidence of countries enrolled in ERSPC

ing the effects of increased screening utilization on the results of survival and screening trials. In particular, when trial patients are recruited from the general population, the model provides guidance on the level of screening occurring in the control arm (contamination). The methodology of this paper can be extended to estimate contamination of the control arms of screening trials from trial data that are longitudinal and consist of a mixture of schedules performed per protocol in the screening arm and random screening occurring in the control arm, before recruitment into the trial, and after the screening protocol has ended (long term follow-up). The proposed approach lays the groundwork for model-based joint integrative analysis of population and trials data in cancer screening, confounded by ad-hoc screening outside of the control period.

Current approaches to the analysis of screening trials rely on comparisons of mortality between arms. This marginal approach fails to utilize the intermediate infor-

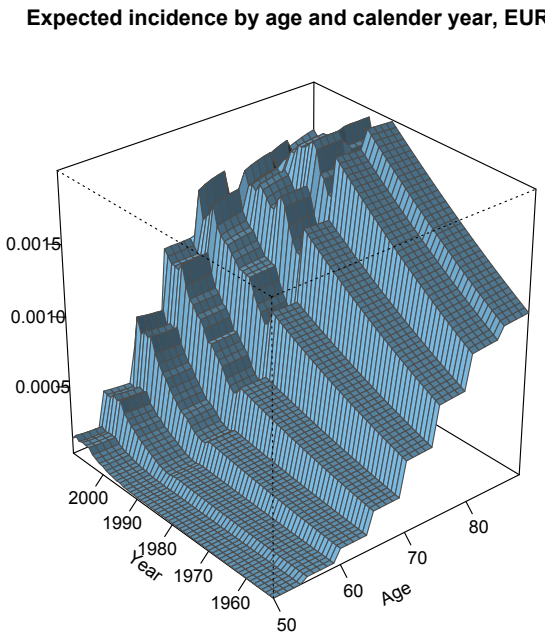

Figure 2.5: Predicted prostate cancer incidence of countries enrolled in ERSPC

mation available from observations of cancer incidence and is subject to effects of contamination of the control arm. Extending the proposed incidence model to include survival post-diagnosis and appropriate effects of early detection on survival is a promising line of future research that could lead to improved power for the effect of early detection on mortality resulting from better use of the sample information. Using the extended model, it might also be possible to resolve the conflicting results of current screening trials in the USA and Europe by careful adjustment for contamination of the control arm, and allowing for some mortality benefit in the control arm from ad-hoc screening.

The proposed joint model is quite general as it incorporates the most salient features common to all cancers. As such, it can be applied to other cancers.

CHAPTER III

Semiparametric Modeling and Analysis of Cancer Incidence with Cancer Screening Trials

3.1 Introduction

High-quality randomized controlled trials have been demonstrated essential for evaluating causal effects of medical interventions such as screening for cancer. However, trials rarely achieve perfect compliance with no contamination. Recent screening trials in prostate cancer, the ERSPC in Europe and PLCO in the United States, showed apparently different results. While ERSPC showed a benefit of screening, albeit mostly driven by the results from one center, the PLCO trial was non-conclusive at face value. Prostate cancer incidence in the US was considerably higher than in Europe before the introduction of screening, likely reflecting earlier adoption of PSA test diagnosis. The PLCO used shorter screening intervals, had a higher PSA threshold for biopsy referral and stopped regular screening after 6 rounds. The US practice setting also contributed to a lower frequency of receipt of biopsy and higher frequency of routine screening in the control arm compared with the ERSPC. Consequently, comparison of intervention and control arms in the PLCO reflects effects of an organized screening program relative to opportunistic screening rather than effects of screening versus no screening. (Berg, 2011; Andriole, 2012)

In chapter 2 we have built a model that provides a quantitative link between dissemination of cancer control processes incorporating random PSA test schedules and cancer incidence. When trial subjects are recruited from the general population, the model may provide guidance on the level of screening occurring in the control arm (contamination). In this chapter we aim to develop a model to estimate contamination of the control arm of screening trials from trial data that are longitudinal and consist of a mixture of schedules performed per protocol in the screening (intervention) arm and random screening occurring in the control arm, before recruitment into the trial, and after the screening protocol has ended (long term follow-up).

Our proposed approach performs model-based joint integrative analysis of population and trials data in cancer screening. The model is developed under the premise of unobserved screening schedules (population patterns) for subjects in the control arm and subjects in the screening arm when they are off trials, with specific screening schedules N_{scr}^{sc} following protocols during trials. The population screening patterns are modeled as an unobserved non-homogeneous Poisson process N_{scr}^{c}, with intensity h_{scr} that depends on age following the longitudinal study. For both arms, subjects are assumed to be screened with the same intensity as the population h_{scr} before they enter trials due to randomization. For subjects in the control arm during the trial, the intensity may change if recruitment into the trial has an effect on subjects' screening patterns (Gulati et al., 2012; Pinsky et al., 2010) and the intervention effect persists after subjects exiting the trial. Hence we set the intensity as $r_1 h_{scr}$ with the risk ratio r_1. For subjects in the screening arm during the trial, the test schedules are discrete and fixed and they return to random after exiting the trial with the intensity $r_2 h_{scr}$, where the risk ratio r_2 models the difference v.s. the population intensity (see Figure 3.1 for illustration). The screening risk ratio r_3 models the difference in screening

utilizations between screening arm and control arm after the trial. We have:

$$r_3 = \frac{r_2 h_{scr}}{r_1 h_{scr}} = \frac{r_2}{r_1} \tag{3.1}$$

The model for the observed data is an average over the unobserved onset time (S), the screening process (N^c_{scr}), and the mode of diagnosis (SDx vs. CDx), where the schedules and mode of diagnosis are unknown.

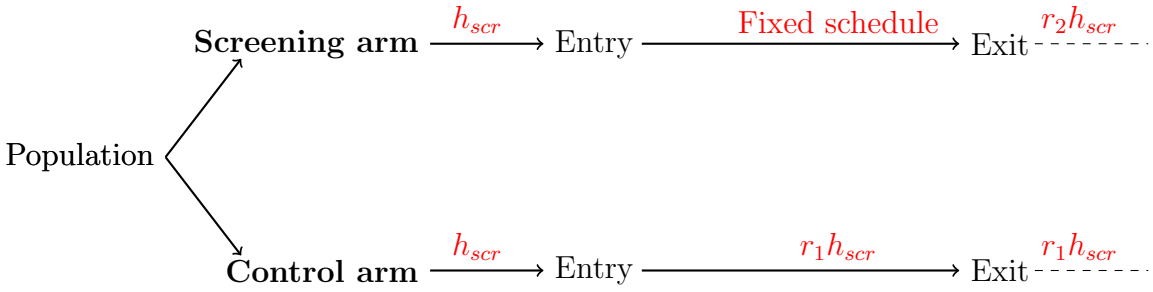

Figure 3.1: An illustration of screening risk (intensity) in different intervals for screening arm and control arm.

In Section 2 we define the joint models with screening for both control and screening arms and derive their essential distributional characteristics. In Section 3 we derive the likelihood function with marked endpoint of diagnosis (SDx or CDx) and the nonparametric maximum likelihood estimators using the iterative reweighting algorithm (Chen, 2009). We apply the method to PLCO trial data in Section 4. Finally, we discuss the results in Section 5.

3.2 Statistical Framework

3.2.1 Joint model with screening for the control arm

For control arm, subjects are screened following population pattern of random schedules with intensity h_{scr} before entering the trial. To specify the joint model during the trial we follow the semi-parametric joint model of time to terminal event affected by a latent progression event by Rice and Tsodikov (2016). Similarly, we

identify the latent event with tumor onset at the age of S, and the terminal event with diagnosis at the age of T and $S \leq T$. The hazard rate associated with the r.v. T, $d\Lambda(t) = \lambda(t)dt$, where t is age at diagnosis (CDx), Λ is the cumulative hazard, λ is its instantaneous counterpart. We assume that population screening pattern behind periods of unobserved screening schedules is stationary (no trend in calender time) because the trial falls on a relatively short and stable period in calender time. The time to the latent onset event S follows the Cox model $d\Lambda_0(t|Z_t)$, given covariates Z_t, with predictor $\theta(Z_t)$, and the baseline cumulative hazard $H(t)$. The terminal event in the absence of screening (CDx) follows the Cox model $d\Lambda_1(t|Z_t, S)$ with predictor $\eta(Z_t)$ and same baseline cumulative hazard $H(t)$, and constrained to show zero risk before the latent event by means of a multiplicative indicator process $I(t \geq S)$ that depends on the latent r.v. S.

$$d\Lambda_0(t|Z_t) = \lim_{\epsilon \to 0} \frac{\Pr(S \in [t, t+\epsilon]|S \geq t, Z_t)}{\epsilon} = \theta(Z_t)dH(t), \tag{3.2}$$

$$d\Lambda_1(t|Z_t, S) = \lim_{\epsilon \to 0} \frac{\Pr(T \in [t, t+\epsilon]|T \geq t, S, Z_t)}{\epsilon} = I(t \geq S)\eta(Z_t)dH(t). \tag{3.3}$$

The relative relationship between θ and η governs the duration of the sojourn time. Exponential parameterization of predictors using regression coefficients $\boldsymbol{\beta} = \left(\beta_0, \boldsymbol{\beta}'_\eta, \boldsymbol{\beta}'_\theta\right)'$ gives $\eta = \eta(\boldsymbol{\beta}; Z_t) = e^{\beta_0 + Z'_t\boldsymbol{\beta}_\eta}$ and $\theta = \theta(\boldsymbol{\beta}; Z_t) = e^{Z'_t\boldsymbol{\beta}_\theta}$. Still in the sequel we may suppress some of the arguments for brevity, for example writing $\eta(t)$ for $\eta(\boldsymbol{\beta}; Z_t)$.

Let the counting process $N^c_{scr}(t)$ count the number of screen tests performed on the subject by the age of t, with the screening schedule expressed by the $[0, t]$-trajectory $\bar{N}^c_{scr}(t)$. With the risks of CDx and SDx competing, we assume that CDx and SDx are independent given covariates $\bar{Z}_t$, tumor onset S and $\bar{N}^c_{scr}(t)$. Then the conditional diagnosis-free survival function is

$$G_d(t|\bar{Z}_t, S, \bar{N}^c_{scr}(t)) = G_{CDx}(t|\bar{Z}_t, S)G_{SDx}(t|\bar{Z}_t, S, \bar{N}^c_{scr}(t)),$$

where $G_{CDx}(t|\bar{Z}_t, S) = e^{-\int_0^t I(\xi \geq S)\eta(\xi)dH(\xi)}$ is the potential conditional survival function for time to CDx in the absence of screening from (3.3). Given a subject's screening schedule at times τ_i, the the potential time to SDx survival function becomes

$$\begin{aligned} G_{SDx}(t|\bar{Z}_t, S, \bar{N}^c_{scr}(t)) &= \prod_{i:S<\tau_i\leq t} (1-\alpha(\tau_i)) \\ &= \exp\left[\int_0^t I(\xi \geq S)\log(1-\alpha(\xi))dN^c_{scr}(\xi)\right], \end{aligned} \tag{3.4}$$

where α is the sensitivity of screening. Let A_e and A_x be the ages of entry into and exit from the trial protocol respectively. Note that the follow-up period starts at A_e and generally extends beyond A_x. When subjects enter the control arm of the trial from the population, the screen counts (assumed to follow a non-homogeneous Poisson process) change their intensity from $h_{scr}(\xi)$ to $r_1 h_{scr}(\xi)$. We have $N^c_{scr}(t) \sim$ Poisson $\left(\int_0^t r_1^{I(\xi\geq A_e)} h_{scr}(\xi)d\xi\right)$. Here r_1 is the risk ratio of screening for subjects of control arm in the trial compared to the population screening intensity. $I(\xi \geq A_e) = 1$ if $\xi \geq A_e$ and $I(\xi \geq A_e) = 0$ if $\xi < A_e$. Using the Laplace functional of a Poisson process with intensity (Serfozo (2009); Shiryaev (1960)), we obtain the marginal survival function incorporating risks of screening and clinical diagnosis as the expectation over the distribution of latent event time S and screening pattern N^c_{scr} for subjects in the control arm. Let $\pi_c(t|\bar{Z}_t)$ be the averaged probability of no diagnosis at time t. For subjects in the population before recruitment into the trial,

$$\pi_c(t|\bar{Z}_t) = \int_0^t e^{-\int_s^t \left[\eta(\xi)dH(\xi)+\alpha(\xi)dH_{scr}(\xi)\right] - \int_0^s \theta(\xi)dH(\xi)} \theta(s)dH(s).$$

During the trial when $A_e \leq t \leq A_x$,

$$\begin{aligned} \pi_c(t|\bar{Z}_t) = &\int_0^{A_e} e^{-\int_s^t \left[\eta(\xi)dH(\xi)+\alpha(\xi)dH_{scr}(\xi)\right] - \int_0^s \theta(\xi)dH(\xi)} \theta(s)dH(s) \\ &+ \int_{A_e}^t e^{-\int_s^t \left[\eta(\xi)dH(\xi)+r_1\alpha(\xi)dH_{scr}(\xi)\right] - \int_0^s \theta(\xi)dH(\xi)} \theta(s)dH(s). \end{aligned}$$

The first term in the above expression averages the probability of no diagnosis before the trial entry, while the second term averages the probability of no diagnosis after the trial entry. After exiting the trial with the same screening risk, $t > A_x$,

$$\begin{aligned}\pi_c(t|\bar{Z}_t) = &\int_0^{A_e} e^{-\int_s^t \left[\eta(\xi)dH(\xi)+\alpha(\xi)dH_{scr}(\xi)\right]-\int_0^s \theta(\xi)dH(\xi)}\theta(s)dH(s)\\ &+\int_{A_e}^{A_x} e^{-\int_s^t \left[\eta(\xi)dH(\xi)+r_1\alpha(\xi)dH_{scr}(\xi)\right]-\int_0^s \theta(\xi)dH(\xi)}\theta(s)dH(s)\\ &+\int_{A_x}^{t} e^{-\int_s^t \left[\eta(\xi)dH(\xi)+r_1\alpha(\xi)dH_{scr}(\xi)\right]-\int_0^s \theta(\xi)dH(\xi)}\theta(s)dH(s).\end{aligned}$$

The sum of the first and second terms in the above expression averages the probability of no diagnosis before exiting the trial protocol at the age of A_x, while the third term averages the probability of no diagnosis during the follow-up after A_x. Let $G_d^c(t)$ be the marginal survival function of the time to diagnosis T for the control arm. Elaborating further on the model quantities we have

$$G_d^c(t|\bar{Z}_t) = \mathbb{E}\left[G_{CDx}(t|\bar{Z}_t, S)G_{SDx}(t|\bar{Z}_t, S, \bar{N}_{scr}^c(t))\right] = e^{-\int_0^t \theta(\xi)dH(\xi)} + \pi_c(t|\bar{Z}_t), \quad (3.5)$$

and the corresponding hazard function is :

$$\begin{aligned}d\Lambda^c(t|\bar{Z}_t) &= -\frac{dG_d^c(t|\bar{Z}_t)}{G_d^c(t|\bar{Z}_t)} = \Psi^c(t|\bar{Z}_t) \times \left[\eta(t)dH(t) + r_1^{I(t\geq A_e)}\alpha(t)dH_{scr}(t)\right],\\ \Psi^c(t|\bar{Z}_t) &= \left\{\pi_c^*(t|\bar{Z}_t)^{-1} + 1\right\}^{-1},\end{aligned} \quad (3.6)$$

where

$$\pi_c^*(t|\bar{Z}_t) = \begin{cases} \int_0^t e^{-\int_s^t \left[(\eta(\xi)-\theta(\xi))dH(\xi)+\alpha(\xi)dH_{scr}(\xi)\right]}\theta(s)dH(s), & t < A_e, \\ \int_0^{A_e} e^{-\int_s^t \left[(\eta(\xi)-\theta(\xi))dH(\xi)+\alpha(\xi)dH_{scr}(\xi)\right]}\theta(s)dH(s) & \\ + \int_{A_e}^t e^{-\int_s^t \left[(\eta(\xi)-\theta(\xi))dH(\xi)+r_1\alpha(\xi)dH_{scr}(\xi)\right]}\theta(s)dH(s), & A_e \le t \le A_x, \\ \int_0^{A_e} e^{-\int_s^t \left[(\eta(\xi)-\theta(\xi))dH(\xi)+\alpha(\xi)dH_{scr}(\xi)\right]}\theta(s)dH(s) & \\ + \int_{A_e}^{A_x} e^{-\int_s^t \left[(\eta(\xi)-\theta(\xi))dH(\xi)+r_1\alpha(\xi)dH_{scr}(\xi)\right]}\theta(s)dH(s) & \\ + \int_{A_x}^t e^{-\int_s^t \left[(\eta(\xi)-\theta(\xi)dH(\xi)+r_1\alpha(\xi)dH_{scr}(\xi)\right]}\theta(s)dH(s), & t > A_x. \end{cases} \tag{3.7}$$

The function Ψ^c summarizes the dependence between times to the two potential points of diagnosis (SDx and CDx), and a departure from additive independent risks for subjects in the control arm.

3.2.2 Joint model with screening for the screening arm

In terms of the screening arm, subjects were recruited from the population into a specific screening schedule N_{scr}^{sc}. While the detection mode (CDx or SDx) in the control arm is unknown, in the screening arm the detection mode is available during the trial protocol period $[A_e, A_x]$ until loss of follow-up. Let I_{scr} be the indicator of diagnosis by screening,

$$I_{scr} = \begin{cases} 1, & \text{screening diagnois}, \\ 0, & \text{clinical diagnosis}. \end{cases}$$

It is natural to assume that subjects in all arms share the same disease natural history model (in particular, the same hazard function (3.3)). For the screening arm before A_e and after A_x (period off trial), the intensities of screening are h_{scr} and $r_2 h_{scr}$, respectively. When $r_1 = 1$ and $r_2 = 1$, subjects in all arms, outside of the period of screening per protocol in the screening arm, have the same intensity as the

population they were recruited from.

Let $\pi_{sc}(t|\bar{Z}_t)$ be the averaged probability of no diagnosis at time t for subjects in the screening arm. Before recruitment, $t < A_e$,

$$\pi_{sc}(t|\bar{Z}_t) = \int_0^t e^{-\int_s^t \left[\eta(\xi)dH(\xi)+\alpha(\xi)dH_{scr}(\xi)\right]-\int_0^s \theta(\xi)dH(\xi)}\theta(s)dH(s).$$

During the period of screening per protocol, when $A_e \leq t \leq A_x$,

$$\begin{aligned}\pi_{sc}(t|\bar{Z}_t) = &\int_0^{A_e} e^{-\int_s^t \left[\eta(\xi)dH(\xi)+\alpha(\xi)dH_{scr}(\xi)\right]-\int_0^s \theta(\xi)dH(\xi)}\theta(s)dH(s)\\ &+\int_{A_e}^t e^{-\int_s^t \left[\eta(\xi)dH(\xi)-\log(\bar{\alpha}(\xi))dN^{sc}_{scr}(\xi)\right]-\int_0^s \theta(\xi)dH(\xi)}\theta(s)dH(s).\end{aligned}$$

In subsequent follow-up when $t > A_x$,

$$\begin{aligned}\pi_{sc}(t|\bar{Z}_t) = &\int_0^{A_e} e^{-\int_s^t \left[\eta(\xi)dH(\xi)+\alpha(\xi)dH_{scr}(\xi)\right]-\int_0^s \theta(\xi)dH(\xi)}\theta(s)dH(s)\\ &+\int_{A_e}^{A_x} e^{-\int_s^t \left[\eta(\xi)dH(\xi)-\log(\bar{\alpha}(\xi))dN^{sc}_{scr}(\xi)\right]-\int_0^s \theta(\xi)dH(\xi)}\theta(s)dH(s)\\ &+\int_{A_x}^t e^{-\int_s^t \left[\eta(\xi)dH(\xi)+r_2\alpha(\xi)dH_{scr}(\xi)\right]-\int_0^s \theta(\xi)dH(\xi)}\theta(s)dH(s).\end{aligned}$$

Let $G_d^{sc}(t)$ be the marginal survival function of the time to diagnosis T for the screening arm, and $f(t, I_{scr} = k), k = 0, 1$ be the two sub-distribution (crude) probability density functions (pdf) by mode of diagnosis. Elaborating further on the model quantities (see Appendix B), we have

$$\begin{aligned}&G_d^{sc}(t|\bar{Z}_t) = \mathbb{E}\left[G_d^{sc}(t|\bar{Z}_t, S, \bar{N}^{sc}_{scr}(t), \bar{N}^{c}_{scr}(t))\right] = e^{-\int_0^t \theta(\xi)dH(\xi)} + \pi_{sc}(t|\bar{Z}_t),\\ &f(t, I_{scr} = 0|Z_t) = \pi_{sc}(t|\bar{Z}_t)\eta(t)dH(t),\\ &f(t, I_{scr} = 1|Z_t) = \pi_{sc}(t|\bar{Z}_t)d\Lambda_{SDx}(t),\end{aligned} \tag{3.8}$$

where

$$d\Lambda_{SDx}(t) = \begin{cases} \alpha(t)dH_{scr}(t), & t < A_e \\ -\log(\bar{\alpha}(t))dN^{sc}_{scr}(t), & A_e \le t \le A_x \\ r_2\alpha(t)dH_{scr}(t), & t > A_x. \end{cases}$$

The corresponding mode-specific hazards are

$$\begin{aligned} d\Lambda^0(t|Z_t) &= \frac{f(t, I_{scr} = 0|Z_t)}{G^{sc}_d(t|Z_t)} = \Psi^{sc}(t|\bar{Z}_t)\eta(t)dH(t), \\ d\Lambda^1(t|Z_t) &= \frac{f(t, I_{scr} = 1|Z_t)}{G^{sc}_d(t|Z_t)} = \Psi^{sc}(t|\bar{Z}_t)d\Lambda_{SDx}(t), \end{aligned} \tag{3.9}$$

where

$$\Psi^{sc}(t|\bar{Z}_t) = \left\{ \left[\pi^*_{sc}(t|\bar{Z}_t)\right]^{-1} + 1 \right\}^{-1}, \tag{3.10}$$

$$\pi^*_{sc}(t|\bar{Z}_t) = \begin{cases} \int_0^t e^{-\int_s^t \left[(\eta(\xi)-\theta(\xi))dH(\xi)+\alpha(\xi)dH_{scr}(\xi)\right]}\theta(s)dH(s), & t < A_e, \\ \int_0^{A_e} e^{-\int_s^t \left[(\eta(\xi)-\theta(\xi))dH(\xi)+\alpha(\xi)dH_{scr}(\xi)\right]}\theta(s)dH(s) & \\ + \int_{A_e}^t e^{-\int_s^t \left[(\eta(\xi)-\theta(\xi))dH(\xi)-\log(\bar{\alpha}(\xi))dN^{sc}_{scr}(\xi)\right]}\theta(s)dH(s), & A_e \le t \le A_x, \\ \int_0^{A_e} e^{-\int_s^t \left[(\eta(\xi)-\theta(\xi))dH(\xi)+\alpha(\xi)dH_{scr}(\xi)\right]}\theta(s)dH(s) & \\ + \int_{A_e}^{A_x} e^{-\int_s^t \left[(\eta(\xi)-\theta(\xi))dH(\xi)-\log(\bar{\alpha}(\xi))dN^{sc}_{scr}(\xi)\right]}\theta(s)dH(s) & \\ + \int_{A_x}^t e^{-\int_s^t \left[(\eta(\xi)-\theta(\xi)dH(\xi)+r_2\alpha(\xi)dH_{scr}(\xi)\right]}\theta(s)dH(s), & t > A_x. \end{cases} \tag{3.11}$$

3.3 Estimation and Hypothesis Testing

3.3.1 Likelihood

The joint log-likelihood function for both control and screening arm is

$$\begin{aligned}\ell &= \ell_c + \ell_{sc} \\ &= \int_0^{\upsilon} \sum_{j\in\mathcal{R}_c(t)} \log(d\Lambda^c(t|Z_{tj}))dN^c(t|Z_{tj}) - d\Lambda^c(t|Z_{tj}) \\ &+ \int_0^{\upsilon} \sum_{i\in\mathcal{R}_{sc}(t)} \sum_{I_{scr}=0}^{1} \left\{\log d\Lambda^{I_{scr}}(t|Z_{ti})dN^{I_{scr}}(t|Z_{ti}) - d\Lambda^{I_{scr}}(t|Z_{ti})\right\}\end{aligned} \tag{3.12}$$

where υ is the maximum follow up time, $\mathcal{R}(t)$ is the risk set at time t specific to the arm (c for control, sc for screening), $dN(t|Z_{tj})$ is the counting process of cancer diagnosis for subject j. For the screening arm, we adapt the likelihood with marked endpoint (T, I_{scr}) from Hu and Tsodikov (2014b). In screening arm, for subject i in the risk set $\mathcal{R}_{sc}(t)$ at time t, $d\Lambda^0(t|Z_{ti}) = \frac{f(t,I_{scr}=0|Z_{ti})}{G_d^{sc}(t|Z_{ti})}$ is the crude hazard of clinical diagnosis; and $dN^0(t|Z_{ti})$ is the counting process of clinical diagnosis in subject i. Similarly, $d\Lambda^1(t|Z_{ti}) = \frac{f(t,I_{scr}=1|Z_{ti})}{G_d^{sc}(t|Z_{ti})}$ is the crude hazard of screening diagnosis and $dN^1(t|Z_{ti})$ is the counting process of screening diagnosis in subject i. The parameter set $\Omega = \{\boldsymbol{\beta}, r_1, r_2, \alpha, H, H_{scr}\}$ enters the likelihood through $\theta, \eta, \Psi^c, \Psi^{sc}$, with observed data $\{\mathcal{R}_c, \mathcal{R}_{sc}, I_{scr}, A_e, A_x, N, N_{scr}^{sc}\}$.

With the appropriate filtration $\mathcal{F}_{t-}$ for subject $i \in \mathcal{R}_c(t)$ or $\mathcal{R}_{sc}(t)$, we construct the orthogonal martingale at the true model:

$$\begin{aligned}&dM^c(t|Z_{ti}) = dN^c(t|Z_{ti}) - d\Lambda^c(t|Z_{ti}), \quad \mathbb{E}\left[dN^c(t|Z_{ti})\middle|\mathcal{F}_{t-}\right] = d\Lambda^c(t|Z_{ti}), \\ &dM^{I_{scr}}(t|Z_{ti}) = dN^{I_{scr}}(t|Z_{ti}) - d\Lambda^{I_{scr}}(t|Z_{ti}), \\ &\mathbb{E}\left[dN^{I_{scr}}(t|Z_{ti})\middle|\mathcal{F}_{t-}\right] = d\Lambda^{I_{scr}}(t|Z_{ti}), I_{scr} = 0, 1.\end{aligned}$$

3.3.2 Score equations

Define the partial derivatives of $\Psi^c(t|\bar{Z}_t), \Psi^{sc}(t|\bar{Z}_t)$ w.r.t $dH(\tau), dH_{scr}(\tau)$as

$$\Psi^c_{dH(\tau)}(t|\bar{Z}_t) = \frac{\partial \Psi^c(t|\bar{Z}_t)}{\partial dH(\tau)}, \Psi^{sc}_{dH(\tau)}(t|\bar{Z}_t) = \frac{\partial \Psi^{sc}(t|\bar{Z}_t)}{\partial dH(\tau)} \tag{3.13}$$

$$\Psi^c_{dH_{scr}(\tau)}(t|\bar{Z}_t) = \frac{\partial \Psi^c(t|\bar{Z}_t)}{\partial dH_{scr}(\tau)}, \Psi^{sc}_{dH_{scr}(\tau)}(t|\bar{Z}_t) = \frac{\partial \Psi^{sc}(t|\bar{Z}_t)}{\partial dH_{scr}(\tau)}. \tag{3.14}$$

respectively. Given the maximum follow-up time υ, differentiating the log-likelihood we arrive at the following score equations for $dH(\tau), dH_{scr}(\tau)$ and $\boldsymbol{\beta}, \alpha(\tau), r_1, r_2$, where $\tau \in (0, t]$ (See Appendix B.2 for details).

$$\begin{aligned}\mathcal{U}_{dH(\tau)} = &\sum_{j\in\mathcal{R}_c(\tau)} \frac{\eta_j(\tau) dN^c(\tau|Z_{\tau j})}{\eta_j(\tau)dH(\tau) + r_1^{I(\tau\geq A_{ej})}\alpha(\tau)dH_{scr}(\tau)} - \sum_{j\in\mathcal{R}_c(\tau)} \Psi^c(\tau|\bar{Z}_{\tau j})\eta_j(\tau) w^c_{dH(\tau)} \\ &+ \sum_{i\in\mathcal{R}_{sc}(\tau)} \sum_{I_{scr}=0} \frac{dN^{I_{scr}}(\tau|Z_{\tau i})}{dH(\tau)} - \sum_{i\in\mathcal{R}_{sc}(\tau)} \sum_{I_{scr}=0} \Psi^{sc}(\tau|\bar{Z}_{\tau i})\eta_i(\tau) w^{sc}_{dH(\tau)},\end{aligned} \tag{3.15}$$

where

$$\begin{aligned} w^c_{dH(\tau)} &= 1 - \frac{\int_\tau^\upsilon \sum_{j\in\mathcal{R}_c(t)} \frac{\Psi^c_{dH(\tau)}(t|\bar{Z}_{tj})}{\Psi^c(t|\bar{Z}_{tj})} dM^c(t|\bar{Z}_{tj})}{\sum_{j\in\mathcal{R}_c(\tau)} \Psi^c(\tau|\bar{Z}_{\tau j})\eta_j(\tau)}, \\ w^{sc}_{dH(\tau)} &= 1 - \frac{\int_\tau^\upsilon \sum_{i\in\mathcal{R}_{sc}(t)} \sum_{I_{scr}=0}^1 \frac{\Psi^{sc}_{dH(\tau)}(t|\bar{Z}_{ti})}{\Psi^{sc}(t|\bar{Z}_{ti})} dM^{I_{scr}}(t|\bar{Z}_{ti})}{\sum_{i\in\mathcal{R}_{sc}(\tau)} \sum_{I_{scr}=0} \Psi^{sc}(\tau|\bar{Z}_{\tau i})\eta_i(\tau)}. \end{aligned} \tag{3.16}$$

$$
\begin{aligned}
\mathcal{U}_{dH_{scr}(\tau)} = & \sum_{j\in\mathcal{R}_c(\tau)} \frac{r_1^{I(\tau\geq A_{ej})}\alpha(\tau)dN^c(\tau|Z_{\tau j})}{\eta_j(\tau)dH(\tau)+r_1^{I(\tau\geq A_{ej})}\alpha(\tau)dH_{scr}(\tau)} \\
& - \sum_{j\in\mathcal{R}_c(\tau)} \Psi^c(\tau|\bar{Z}_{\tau j}) r_1^{I(\tau\geq A_{ej})}\alpha(\tau) w^c_{dH_{scr}(\tau)} \\
& + \sum_{i\in\mathcal{R}_{sc}(\tau)} \sum_{I_{scr}=1} \frac{I(\tau<A_{ei},\tau>A_{xi})dN^{I_{scr}}(\tau|Z_{\tau i})}{dH_{scr}(\tau)} \\
& - \sum_{i\in\mathcal{R}_{sc}(\tau)} \sum_{I_{scr}=1} \Psi^{sc}(\tau|\bar{Z}_{\tau i}) \left[I(\tau<A_{ei})\alpha(\tau)+I(\tau>A_{xi})r_2\alpha(\tau)\right] w^{sc}_{dH_{scr}(\tau)},
\end{aligned}
\tag{3.17}
$$

where $I(\tau < A_{ei}, \tau > A_{xi}) = 1$ if $\tau < A_{ei}$ or $\tau > A_{xi}$; 0 otherwise.

$$
\begin{aligned}
w^c_{dH_{scr}(\tau)} &= 1 - \frac{\int_\tau^v \sum_{i\in\mathcal{R}_c(t)} \frac{\Psi^c_{dH_{scr}(\tau)}(t|\bar{Z}_{tj})}{\Psi^c(t|\bar{Z}_{tj})} dM^c(t|\bar{Z}_{tj})}{\sum_{j\in\mathcal{R}_c(\tau)} \Psi^c(\tau|\bar{Z}_{\tau j}) r_1^{I(\tau\geq A_{ej})}\alpha(\tau)}, \\
w^{sc}_{dH_{scr}(\tau)} &= 1 - \frac{\int_\tau^v \sum_{i\in\mathcal{R}(t)} \sum_{I_{scr}=0}^{1} \frac{\Psi^{sc}_{dH_{scr}(\tau)}(t|\bar{Z}_{ti})}{\Psi^{sc}(t|Z_{ti})} dM^{I_{scr}}(t|Z_{ti})}{\sum_{i\in\mathcal{R}_{sc}(\tau)} \sum_{I_{scr}=1} \Psi^{sc}(\tau|\bar{Z}_{\tau i}) \left[I(\tau<A_{ei})\alpha(\tau)+I(\tau>A_{xi})r_2\alpha(\tau)\right]}.
\end{aligned}
\tag{3.18}
$$

The scores for $\boldsymbol{\beta}, r_1, r_2, \alpha(\tau)$ are

$$
\mathcal{U}_{\boldsymbol{\beta}} = \int_0^v \sum_{j\in\mathcal{R}_c(t)} \frac{\partial \log d\Lambda^c(t|\bar{Z}_{tj})}{\partial\boldsymbol{\beta}} dM^c(t|Z_{tj}) + \int_0^v \sum_{i\in\mathcal{R}_{sc}(t)} \sum_{I_{scr}=0}^{1} \frac{\partial \log d\Lambda^{I_{scr}}(t|\bar{Z}_{ti})}{\partial\boldsymbol{\beta}} dM^{I_{scr}}(t|Z_{ti}),
\tag{3.19}
$$

$$
\mathcal{U}_{r_1} = \int_0^v \sum_{j\in\mathcal{R}_c(t)} \frac{\partial \log d\Lambda^c(t|\bar{Z}_{tj})}{\partial r_1} dM^c(t|Z_{tj}),
\tag{3.20}
$$

$$\mathcal{U}_{r_2} = \int_0^{\upsilon} \sum_{i \in \mathcal{R}_{sc}(t)} \sum_{I_{scr}=0}^{1} \frac{\partial \log d\Lambda^{I_{scr}}(t|\bar{Z}_{ti})}{\partial r_2} dM^{I_{scr}}(t|Z_{ti}), \tag{3.21}$$

$$\mathcal{U}_{\alpha(\tau)} = \int_{\tau}^{\upsilon} \sum_{j \in \mathcal{R}_c(t)} \frac{\partial \log d\Lambda^{c}(t|\bar{Z}_{tj})}{\partial \alpha(\tau)} dM^{c}(t|Z_{tj}) + \int_{\tau}^{\upsilon} \sum_{i \in \mathcal{R}_{sc}(t)} \sum_{I_{scr}=0}^{1} \frac{\partial \log d\Lambda^{I_{scr}}(t|\bar{Z}_{ti})}{\partial \alpha(\tau)} dM^{I_{scr}}(t|Z_{ti}). \tag{3.22}$$

The scores for $\boldsymbol{\beta}, r_1, r_2, \alpha(\tau)$ are martingales at the true model, same as the score functions for H, H_{scr}(see Appendix B for details). We have the weighted Breslow-type estimators dH and dH_{scr} from the NPMLE, with weights $w = (w^c_{dH}, w^{sc}_{dH}, w^c_{dH_{scr}}, w^{sc}_{dH_{scr}})$ that have expectation of 1, given filtration $\mathcal{F}_{t_-}$. The optimal weights depend on martingale residuals evaluated over the future of the subject ensure the efficiency of the NPMLE. To enhance the robustness of the parameters in our model, we also include cancer incidence data from the Surveillance, Epidemiology, and End Results (SEER, 2017) into the estimating procedure. Data before PSA screening began in the US can be used to estimate the baseline hazard dH in the absence of screening, treated as another control arm without screening contamination. While data after screening test was introduced can be used to estimate the screening risk $r_4 h_{scr}$ in the population from SEER, treated as another intervention arm. Here r_4 denotes the risk ratio of screening in the population from SEER.

3.3.3 Estimating algorithm and Hypothesis tests

Similar to Chapter 2, we keep adapting the Weighted Breslow Estimator algorithm (Chen, 2009) to maximize the likelihood by iteratively updating the weights. While a set of recurrent score equations emerges similar to the computationally efficient martingale estimating equations (Chen et al., 2002). The weights are treated as known at the inner loop of the algorithm. For simplicity we assume constant screening

sensitivity α. Set initial values for $\Omega = (\boldsymbol{\beta}, r_1, r_2, r_4, \alpha)$ and initial weights $w^{(0)}$ as 1, use Nelson-Aalen estimator as initial values for dH and dH_{scr}. For iteration count $k = 0, 1, ...$, we repeat the following steps until convergence to maximize the likelihood over the two hazards and obtain the profile likelihood $\ell_{pr}(\Omega)$:

1. Fix weights $w^{(k)}$ and given $(dH^{(k)}, dH_{scr}^{(k)})$, obtain the solution $(dH^{(k+1)}, dH_{scr}^{(k+1)})$ from the score equations for dH and dH_{scr}.

2. Update the weights $w^{(k+1)}$ using (3.16) and (3.18) with $(dH^{(k+1)}, dH_{scr}^{(k+1)})$.

Maximization of the profile likelihood $\ell_{pr}(\Omega)$ is obtained using the above algorithm. We arrive at the final MLEs with general maximization methods such as conjugate gradients applied to ℓ_{pr}. Given the MLEs of our proposed estimators, we construct the likelihood ratio test (LRT) for our hypotheses.

Hypothesis 1. The recruitment into the trial did not change the ad-hoc screening patterns, control arm.

$$H_0 : r_1 = 1 \text{ vs. } H_1 : r_1 \neq 1$$

Hypothesis 2. Being on trial's protocol did not change ad-hoc screening patterns, screening arm.

$$H_0 : r_2 = 1 \text{ vs. } H_1 : r_2 \neq 1$$

Hypothesis 3. No difference in ad-hoc screening patterns between arms.

$$H_0 : r_3 = 1(r_1 = r_2) \text{ vs. } H_1 : r_3 \neq 1$$

Hypothesis 4. All ad-hoc screening patterns in the trial are the same as in the population.

$$H_0 : r_1 = r_2 = r_3 = 1 \text{ vs. } H_1 : \text{at least one } \quad r_i \neq 1$$

Hypothesis 5. Population in SEER have the same screening pattern as the general one.

$$H_0 : r_4 = 1 \text{ vs. } H_1 : r_4 \neq 1$$

3.4 Data Analysis Example

3.4.1 PLCO and SEER data analysis

The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial is a large population-based randomized trial evaluating screening programs for these cancers. The primary goal of this long-term trial of the National Cancer Institute's (NCI) Division of Cancer Prevention (DCP) is to determine the effects of screening on cancer-related mortality and on secondary endpoints. Ten screening centers located across the United States enrolled 76,685 men and 78,216 women, ages 55 to 74, and randomized them to an intervention arm, which received trial screening, or control arm, which received standard care (can still receive screening test following the population pattern). Participants included in the intervention arm of the trial received yearly screening per protocol cancer during their first 6 years of participation in the trial, and follow-up continued for at least 7 additional years: 39,105 women and 38,340 men were in this part of the trial. In the first 6 years, men received PSA blood tests. Participants in the control arm were followed for 13 years after enrollment, but did not receive any planned screening examinations: 38,111 women and 38,345 men

were in the control arm. Eligibility requirements included age (between 55 and 74 at enrollment), and no previous history of any PLCO cancer (PLCO, 2016). To fit the model we use the incidence data of both control and screening arms from PLCO trial, including age at randomization, age at diagnosis, screening schedule and mode of diagnosis during the protocol phase and follow-up time for both arms.

The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute provides information on cancer statistics in an effort to reduce the cancer burden among the U.S. population. SEER is supported by the Surveillance Research Program, which provides national leadership in the science of cancer surveillance as well as analytical tools and methodological expertise in collecting, analyzing, interpreting, and disseminating reliable population-based statistics (SEER, 2017). We use cancer incidence data from 9 main cancer registries. The SEER 9 registries are Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, and Utah. Data are available for cases diagnosed from 1973 and later for these registries with the exception of Seattle-Puget Sound and Atlanta. The Seattle-Puget Sound and Atlanta registries joined the SEER program in 1974 and 1975, respectively. For the control arm, we use cancer incidence data with population by age from 1982-1986, just before PSA screening began in the US. And for the intervention arm data is from 1991-2001.

Set $\theta = e^{Z'\beta_\theta} = 1, \eta = e^{\beta_0 + Z'\beta_\eta} = e^{\beta_0}$, and initial values $\beta_0 = 0.5, \alpha = 0.9, r_1 = 1, r_2 = 1, r_4 = 1$. Following the algorithm we obtain the MLEs $\hat{\eta}, (\hat{\alpha}), \hat{r}_1, \hat{r}_2, \hat{r}_3, \hat{r}_4$, $d\hat{H}, d\hat{H}_{scr}$. Table 3.1 summarizes the MLEs with 95%CI from the likelihood ratio tests. The estimated screening sensitivity $\hat{\alpha} = 0.87$, that is when a patient is receiving a screening test, the chance of having positive result is 87%, given tumor onset. This is consistent with Gulati et al. (2010). $\hat{r}_1 = 1.04$, 95%CI=(0.94,1.14). Within the trial control arm, the screening intensity is slightly yet not significantly higher during the trial compared with the population screening pattern, reflecting the oppor-

tunistic screening pattern (Berg, 2011; Andriole, 2012). $\hat{r}_2 = 3.04$, 95%CI=(2.89,3.19) indicates that the screening intensity in the screening arm after 6 years is significantly higher than that in the population, approximately three times higher screening risk than the population pattern, which reflects the non-compliance and contamination that were previously described in Gulati et al. (2012). $\hat{r}_4 = 0.94$, 95%CI=(0.92,0.95) shows that the screening intensity in the population from SEER is slightly lower than the general population. Table 3.2 summarizes the Hypotheses testing results. The recruitment into the trial did not change the ad-hoc screening patterns for control arm (Hypothesis 1); while being on trial's protocol did change ad-hoc screening patterns for screening arm (Hypothesis 2); there exists significant difference in ad-hoc screening patterns between arms (Hypothesis 3); ad-hoc screening patterns in the trial are different from the population (Hypothesis 4); screening risk in the population from SEER is slightly lower from the general population pattern (Hypothesis 5).

Table 3.1: Maximized likelihood estimators with 95%CI

Parameter	$\hat{\mu}$	95%lower	95%upper
η	1.73	1.71	1.75
α	0.87	0.86	0.88
r_1	1.04	0.94	1.14
r_2	3.04	2.89	3.19
r_4	0.94	0.92	0.95

η: predictor for terminal event; α: screening test sensitivity

Table 3.2: Hypotheses testing results of screening risk ratios

Hypothesis	χ^2	DF	P-value
$r_1 = 1$	0.61	1	0.43
$r_2 = 1$	710.54	1	< 0.0001
$r_3 = 1(r_1 = r_2)$	245.86	1	< 0.0001
$r_1 = r_2 = r_3$	554.25	2	< 0.0001
$r_4 = 1$	61.47	1	< 0.0001

DF: degree of freedom; χ^2 statistics from Likelihood ratio test

Figure 3.2 displays the incidence comparison between arms and within arms for PLCO trial. Observed incidence in the screening arm is higher than that in the

control arm and merge at age 80, showing continued elevated incidence for older ages with the introduction of PSA testing. If subjects in the screening arm resumed screening pattern as the population after trial's 6-year intervention period ($r_2 = 1$), the incidence would be still higher than the control arm, but lower than observed incidence in screening arm (null predicted incidence, screening arm, PLCO). While subjects in the control arm resumed screening pattern as the population after trial's follow up period, the incidence is approximately the same as the observed one (null predicted incidence, control arm, PLCO). In terms of model fit, the observed and predicted incidence of both arms match well in Figure 3.2. Figure B.1 in Appendix B is the corresponding Kaplan-Meier curves associated with subjects' diagnosis-free survival by age.

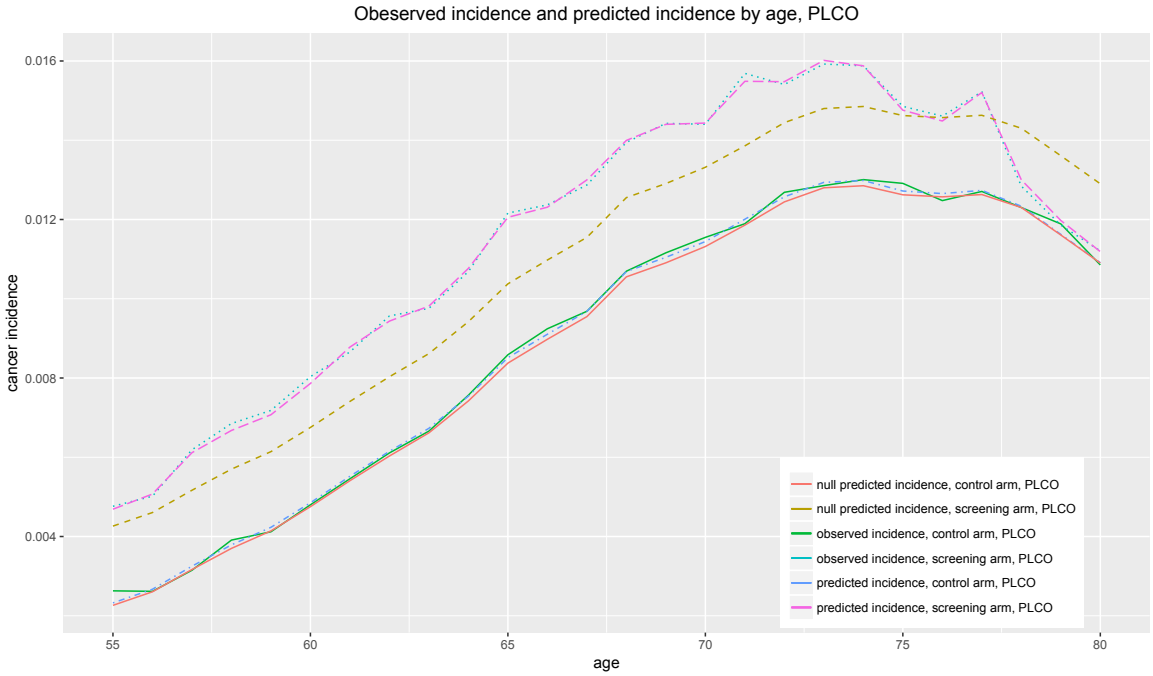

Figure 3.2: Observed incidence and predicted incidence by age for PLCO trial; null incidence with risk ratio as 1.

Figure 3.3 displays the incidence comparison between arms and within arms of SEER. Observed incidence in the screening arm is higher than that in the control arm and converge at age 80, showing continued elevated incidence for older ages with

the introduction of PSA testing. When subjects in the screening arm follow the same screening pattern as the population ($r_4 = 1$), the incidence is approximately the same as the observed one (null predicted incidence, screening arm, SEER). The observed and predicted incidence of both arms match well in Figure 3.3. Figure B.2 in Appendix B is the corresponding Kaplan-Meier curves associated with subjects' diagnosis-free survival by age.

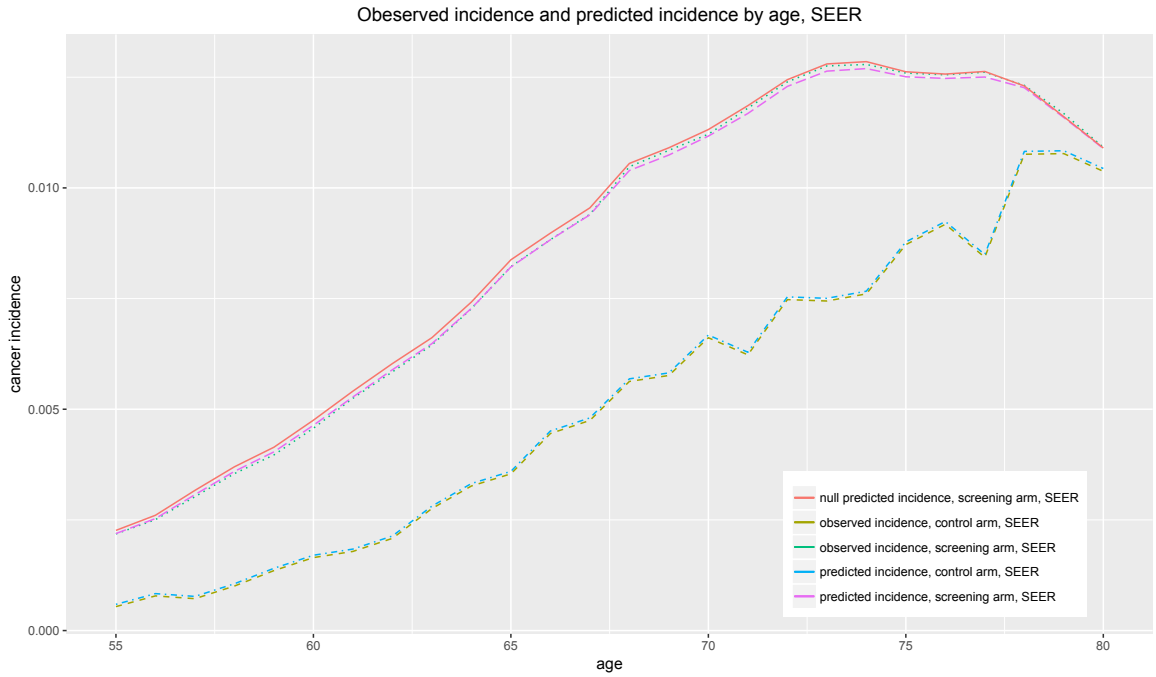

Figure 3.3: Observed incidence and predicted incidence by age for SEER; null incidence with risk ratio as 1.

3.5 DISCUSSION

The model we formulated provides a quantitative link between screening with a combination of unknown and random population pattern and specific trial known schedules and their impact on cancer incidence. We quantify the relationships and differences in prostate cancer incidence between arms in both PLCO trial and SEER in terms of model parameters $(\boldsymbol{\beta}, \alpha, r_1, r_2, r_3, r_4, dH, dH_{scr})$ and perform inference on these parameters. The model can estimate ad-hoc screening intensity (contamination)

jointly with the disease natural history, given specific screening schedules for cancer screening trials during the protocol phase, and perform hypotheses tests of screening intensities between and within intervention and control arms.

The model can assess the effects of screening intended utilization and unintended contamination on the results of survival and screening trials. In particular, when trial patients are recruited from the general population, the model evaluates the level of screening occurring in the control arm and after the protocol phase in the screening arm (contamination) using trial data that are longitudinal follow-up and consist of a mixture of periods of known and unknown schedules. The proposed approach can easily be extend to perform model-based joint integrative analysis of population and trials data in cancer screening settings.

Current marginal approaches to the analysis of screening trials rely on comparisons of mortality between arms but fail to utilize the information available from observations of cancer incidence. Mostly importantly, they fail to adjust for unintended screening contamination. The proposed joint model gives the foundation for the analysis of mortality and the effects of early detection on survival, adjusted for contamination. With the extended models, it might be possible to resolve the conflicting results of current screening trials in the USA and Europe by careful adjustment for contamination of the control arm, and allowing for some mortality benefit in the control arm from ad-hoc screening.

CHAPTER IV

Marginal Analysis of Cancer Screening Effect on Mortality Adjusting for Screening Contamination

4.1 Introduction

More than two decades after prostate-specific antigen (PSA) screening for prostate cancer entered clinical practice, the US Preventive Services Task Force (USPSTF) determined there was very low probability of preventing a death from prostate cancer in the long term and recommended against routine use of the test (Moyer and Force, 2012). Since then, PSA screening rates and prostate cancer incidence rates in the United States have declined significantly (Jemal et al., 2015, 2016).

The USPSTF recommendation relied heavily on results from the European Randomized Study of Screening for Prostate Cancer (ERSPC; ISRCTN49127736) and the Prostate, Lung, Colorectal, and Ovarian cancer screening trial (PLCO; NCT00002540). However, the trials produced apparently conflicting results, with the ERSPC reporting a 21% reduction in prostate cancer mortality (Schroder et al., 2009, 2012, 2014) and the PLCO finding no mortality difference between the trial arms (Andriole, 2009, 2012). Interpreting results of these trials is complicated by differences in their implementations, including design and adherence, and practice settings. The PLCO used shorter screening intervals (annual screening versus every 2-4 years), had a higher

PSA threshold for biopsy referral (4.0 g/L versus 3.0 g/L in most ERSPC centers and rounds), and stopped regular screening after 6 rounds. Prostate cancer incidence in the US was higher than in Europe before the trials started, reflecting different populations and clinical diagnosis patterns. The US practice setting also contributed to a higher frequency of screening in the control arm and a lower frequency of biopsy compared with the ERSPC. Consequently, the PLCO compared effects of an organized screening program relative to opportunistic screening rather than effects of screening versus no screening (Berg, 2011; Andriole, 2012; Pinsky et al., 2016). Nonetheless, the PLCO results have been viewed as more relevant to the US setting.

The perceived inconclusiveness of the trial results about screening efficacy persists despite studies indicating that differences in their implementation might explain at least some of the variability in their results. A prior investigation by Gulati et al. (2012) indicates that the trial results may be more consistent than they appear. However, these studies did not formally evaluate whether screening efficacy differed between trials when implementation details such as screening patterns are taken into account. In this chapter we use a two-step approach to formally test whether screening efficacy differs between the ERSPC and PLCO using mean lead time as a surrogate of screening intensity. Lead time is the time by with diagnosis of cancer is advanced due to screening in patients who would be detected anyway in the absence of screening tests, can be expressed by $T_{\mathrm{CDx}} - T_{\mathrm{SDx}}$ where T is the r.v of time to diagnosis (Tsodikov et al., 2006). First, the mean lead time is estimated in each trial arm as a proxy for the intensity of screening. Second, the association is quantified between the mean lead time and prostate cancer mortality and tested whether it differs between trials.

The objectives of this chapter are to (1) formally test whether the effects of screening on prostate cancer mortality differed between the ERSPC and PLCO after accounting for differences in implementation and practice settings and (2) to estimate the effects of screening in both trials relative to no screening.In Section 2 we derive

the mean lead time based on the semiparametic transformation model for screening. In Section 3 we fit the regression model of cancer mortality with prediction of mean lead time given trials and SEER data. Finally, we discuss the results in Section 4.

4.2 Methods

This study used individual records from both trials in a collaboration between trial investigators and the Cancer Intervention and Surveillance Modeling Network (CISNET) prostate cancer working group. In the intervention arms, these records included age and year of randomization, enrollment center, dates and results of PSA tests and rectal exams, whether biopsy was performed, date of cancer diagnosis, and date and cause of death. In the control arms, the records included age and year of randomization, enrollment center, date of cancer diagnosis, and date and cause of death. For consistency with prior publications, ERSPC data included men aged 55-69 years at randomization, while PLCO data included men aged 55-74 years at randomization.

We first examined a traditional statistical analysis that combined data from both trials and compared hazards of prostate cancer death in the intervention versus control arms adjusting for participant age and trial setting. However, this analysis is questionable due to remaining differences in implementation between trials. To overcome this limitation, we also examined extended analyses that accounted for variable screening and diagnostic workup (hereafter screening intensity) in each trial arm, which we operationalized in terms of mean lead times (MLTs). The MLTs quantify the magnitude of increased prostate cancer incidence relative to a baseline level expected in the absence of screening, thus capturing differences in both design and adherence (see below). We estimated the MLTs both empirically and using analytic or microsimulation models; using multiple approaches allowed us to assess robustness of results to this uncertain quantity.

4.2.1 Estimating mean lead times

Lead time is usually defined as the amount of time by which clinical diagnosis (i.e., diagnosis without screening) is advanced by screening. In this chapter we define:

$$LT = \begin{cases} 0, & T_{\text{CDx}} \leq T_{\text{SDx}} \\ T_{\text{CDx}} - T_{\text{SDx}}, & T_{\text{CDx}} > T_{\text{SDx}}. \end{cases} \tag{4.1}$$

Here T_{CDx} is the time from randomization to clinical diagnosis and T_{SDx} is the time from randomization to screen detection. Under complete follow-up (i.e., where all pre-clinical cases are eventually diagnosed in the no-screening setting), the MLT corresponds to the difference in areas under two survival curves for time from randomization to diagnosis (mean survival time to diagnosis): one in the absence of screening minus one in the presence of screening . From (4.1)

$$MLT = \mathbb{E}\left[\max(0, T_{\text{CDx}} - T_{\text{SDx}})\right] = \mathbb{E}(T_{\text{CDx}}) - \mathbb{E}\left[\min(T_{\text{CDx}}, T_{\text{SDx}})\right]. \tag{4.2}$$

Under limited follow-up time $T_{\max}$, we can define a restricted version of the MLT as an analogous difference in areas under survival curves up to a specified time point (Uno et al., 2014). Restricting to the duration of the trial recognizes that events after the trial period cannot affect the mortality during the trial. Specifically, the time of screen detection becomes $T'_{\text{SDx}} = \min(T_{\max}, T_{\text{SDx}})$,and the time of clinical diagnosis becomes $T'_{\text{CDx}} = \min(T_{\max}, T_{\text{CDx}})$. To make estimates between trials comparable, follow-up was restricted to 11 years. We have the restricted lead time

$$LT = \begin{cases} 0, & T'_{\text{CDx}} \leq T'_{\text{SDx}} \\ T'_{\text{CDx}} - T'_{\text{SDx}}, & T'_{\text{CDx}} > T'_{\text{SDx}}. \end{cases} \tag{4.3}$$

Correspondingly, the mean restricted lead time

$$\begin{aligned} MLT &= \mathbb{E}\left[\max(0, T'_{\text{CDx}} - T'_{\text{SDx}})\right] = \mathbb{E}(T'_{\text{CDx}}) - \mathbb{E}\left[\min(T'_{\text{CDx}}, T'_{\text{SDx}})\right] \\ &= \int_0^{T_{\max}} G_{\text{CDx}}(\xi) - G_d(\xi) dt, \end{aligned} \tag{4.4}$$

where G_{CDx} is the survival function for time to clinical diagnosis in the absence of screening and G_d is the survival function for time to diagnosis with screening.

Finally, to standardize the measure across trial arms in this study, we scaled each estimated MLT by a common baseline probability of (screen or clinical) diagnosis during follow-up $T_{\max}$, making the MLT a conditional average given any mode of prostate cancer diagnosis during the trial,

$$MLT = \frac{\int_0^{T_{\max}} G_{\text{CDx}}(\xi) - G_d(\xi) dt}{1 - G_d(T_{\max})}. \tag{4.5}$$

Please not that this version of lead time is only defined in patients who are detected by screening and who, in the absence of screening, would be clinically diagnosed. This definition excludes overdiagnosed patients (i.e., patients detected by screening who would not be clinically diagnosed in the absence of screening), patients clinically diagnosed, and patients without any diagnosis. Our goal here, however, is to derive a generic surrogate of the intensity of screening and diagnosis in a given population that is amenable to empirical (model-free, robust) estimation and prediction. Nevertheless, due to its close theoretical relationship with the mean lead time (described below), we adopt this terminology.

In a screening trial, the time from randomization to screen detection and the time from randomization to clinical diagnosis are competing risks that are never both observed for the same patient. Consequently, their full joint distribution is non-identifiable without specific modeling assumptions (Tsiatis, 1975). However, the mean lead time (i.e., the mean time difference $T_{\text{CDx}} - T_{\text{SDx}}$, which is non-zero only

if $T_{\mathrm{SDx}} < T_{\mathrm{CDx}}$) is identifiable and can be calculated empirically.

Note that our estimates of the MLTs are different from other estimates in the literature that can be interpreted as the average time by which screening advances diagnosis among cases that would have been clinically diagnosed (Draisma et al., 2009). Our MLTs are designed to represent proxies for the intensity of screening and diagnosis, with higher values reflecting higher attendance rates at screening exams, more frequent screening exams, less conservative criteria for biopsy referral, and/or higher frequencies of biopsy. Thus, accounting for variable MLTs across trial arms captures in a single measure important differences in the trial screening protocols, participant adherence to those protocols in the intervention arms, and control arm screening. We estimated the MLTs empirically, without any model assumptions about cancer progression and diagnosis, and also using three models of cancer natural history and diagnosis. The empirical approach estimated the MLTs by calculating the difference between survival curves for time from randomization to diagnosis in each trial arm relative to an assumed baseline level. The assumed baseline probability of diagnosis in the absence of screening was derived using incidence rates from the Surveillance, Epidemiology, and End Results (SEER) program in 1986, just before PSA screening began in the US, adjusted to reflect distributions of age at randomization in each trial. Additionally, one analytic model (UMICH) and two simulation models (FHCRC and MISCAN) estimated distributions of ages at onset of latent disease and diagnosis in the absence and presence of screening using individual patient attendance, screening, and incidence data. The fitted models then estimated MLTs as in the empirical approach but using projected instead of observed incidence rates. Additional details are described in the Supplementary Materials C.

4.2.2 Effects of screening on prostate cancer mortality

We used Cox regression to model survival from randomization to prostate cancer death, censoring individuals who died of other causes or were alive at last follow-up. We examined both a traditional statistical analysis and extended analyses that incorporated the measure of screening intensity captured by the estimated MLTs. Both types of analysis included participant age at randomization and a trial setting indicator (PLCO versus ERSPC), which allowed for a different baseline risk of prostate cancer death in the absence of screening between the two trial settings.

4.2.2.1 Traditional statistical analysis

We first conducted a traditional analysis to test whether the effect of screening differed between trials. Specifically, we tested the effect of being randomized to the intervention arm (relative to the control arm) on the risk of prostate cancer death and all cause of death. For the prostate specific survival model, the exponential of the coefficient for the trial arm indicator is the hazard ratio for prostate cancer death in the intervention arm relative to the control arm; in other words, it reflects the effect of screening on prostate cancer mortality in an intent-to-screen analysis. We fitted this model with and without allowing separate effects of screening in each trial (i.e., with and without interaction between the trial arm and trial indicator), then used a likelihood ratio test to evaluate evidence of differential effects of screening between trials.

4.2.2.2 Extended statistical analysis

Next we replaced the trial arm indicator with the corresponding MLT estimated empirically or using a model-based approach and fit both prostate specific survival model and overall survival model with all cause of death. For the prostate specific survival model, the exponential of the coefficient for the MLT represents the hazard

ratio for prostate cancer death per additional year of MLT; in other words, it reflects screening efficacy standardized by screening intensity. As in the traditional analysis, we fitted this model with and without allowing separate effects of screening on prostate cancer mortality in each trial (i.e., with and without interaction between the MLT and trial indicator), then used a likelihood ratio test to evaluate evidence of differential effects of screening between trials. Our extended analyses are consistent with the analyses in the trial publications (Schroder et al., 2009; Andriole, 2009) with two important differences. First, rather than relying on an intent-to-treat effect of screening determined by the assigned arm in a single trial, we explicitly included a covariate (MLT) to capture the intensity of screening in each arm. This represents a transition from thinking about screening as all or nothing, corresponding to an intervention and control arm, to a continuous metric of screening intensity, with resulting coefficient estimates interpreted relative to a no-screening setting (i.e., a setting where MLT=0). Second, we used combined data from both trials in a single analysis, adding an indicator for trial to capture differences between trials in baseline cancer-specific survival without screening and an interaction term to test whether screening efficacy (per year of MLT) differed between trials.

4.3 Data Analysis

Table 4.1 summarizes participants, follow-up, and prostate cancer cases and deaths in the two trials using all available follow-up and restricted to 11 years of follow-up. The data under all available follow-up differ modestly from published results (Schroder et al., 2009; Andriole, 2012) due to additional cleaning and updating. Nonetheless, the cleaned and updated data restricted to 11 years of follow-up yielded values similar to published prostate cancer incidence rate ratios (PLCO: 1.12 vs 1.12; ERSPC: 1.68 vs 1.63) and mortality rate ratios (PLCO: 1.02 vs 1.09; ERSPC: 0.79 vs 0.79) and preserve the greater effects of screening on prostate cancer incidence and

mortality rates in the ERSPC relative to the PLCO.

Table 4.1: Summary of participants, follow-up, and prostate cancer cases and deaths in the ERSPC and PLCO under all available follow-up or restricted to 11 years of follow-up.

	ERPSC		PLCO	
	Control	Screening	Control	Screening
No. of participants	88,921	72,473	38,343	38,340
Age at randomization (years)				
median	59	60	62	62
range	55-69	55-69	55-74	55-74
All available follow-up				
Follow-up from randomization (years)				
median	11.0	11.1	12.5	12.5
range	0.4-17.5	0.4-17.3	0.0-13.0	0.0-13.0
No. of prostate cancer cases	5,398	6,967	4,040	4,430
Person-years of follow-up for incidence	933,854	740,775	403,955	400,008
No. of deaths	17,019	13,652	7,149	6,940
other causes	16,557	13,353	7,003	6,788
prostate cancer	462	299	146	152
Person-years of follow-up for mortality	990,678	827,148	426,720	427,824
Restricted to 11 years of follow-up				
Follow-up from randomization (years)				
median	11.0	11.0	11.0	11.0
range	0.4-11.0	0.4-11.0	0.0-11.0	0.0-11.0
No. of prostate cancer cases	4,961	6,586	3,641	4,038
Person-years of follow-up for incidence	868,834	686,766	368,844	365,129
No. of deaths	13,207	10,397	5,880	5,798
other causes	12,822	10,150	5,771	5,687
prostate cancer	385	247	109	111
Person-years of follow-up for mortality	890,581	725,997	387,027	387,861

ERSPC=European Randomized Study of Screening for Prostate Cancer; PLCO=Prostate, Lung, Colorectal, and Ovarian cancer screening trial

To compare the screening intensity in the intervention and control arms of the two trials, Figure 4.1 illustrates MLTs estimated empirically or using a model-based approach. All estimation approaches found similar ordering and relative magnitudes of MLTs across trial arms. The ERSPC and PLCO intervention arms had similar MLTs, while the PLCO control arm had substantially longer MLTs than the ERSPC control arm, consistent with more intensive screening (i.e., greater contamination) in the PLCO control arm. Table 4.2 reports results of the traditional analysis. A likelihood ratio test associated with this analysis modestly suggested different effects of

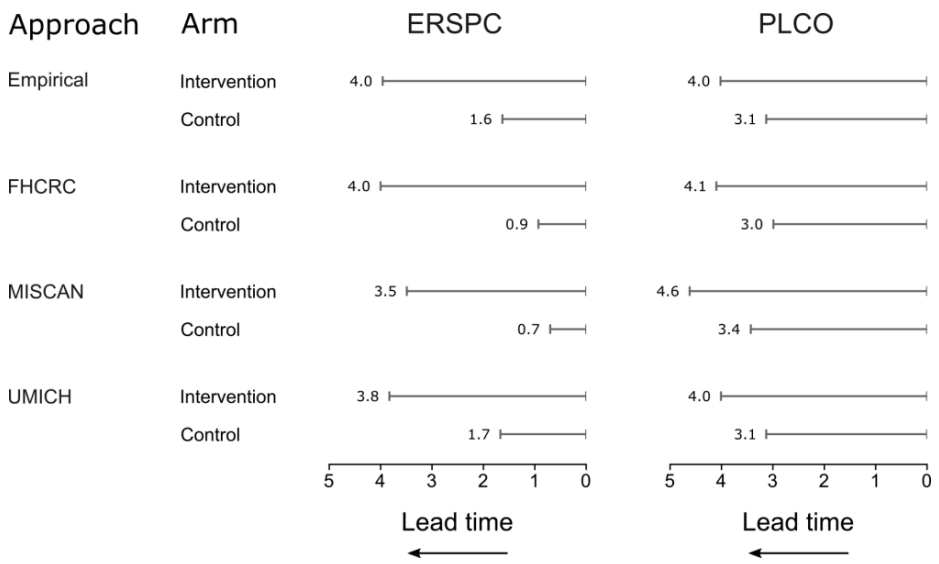

Figure 4.1: Estimated mean lead times (years) in the intervention and control arm of the ERSPC and PLCO relative to a hypothetical no-screening setting (where lead time is always zero). Estimated MLTs are visualized as increasing to the left to suggest the extent to which prostate cancer diagnosis is advanced by more intensive screening and diagnostic workup.

screening on mortality between trials (P=0.09). Under a common effect of screening, screening was estimated to reduce the risk of prostate cancer death by 16% (95% CI 4-27%; P=0.01) after accounting for different baseline risks of prostate cancer death in the PLCO setting relative to the ERSPC setting and participant age at randomization. This result essentially corresponds to a weighted average of the effect in each trial with the relative sizes of the trials as weights.

Table 4.2 also presents our extended analyses, which account for the MLT in each trial arm estimated empirically or using a model-based approach. The analyses are highly consistent and indicate no evidence of different effects of screening on mortality between trials (P=0.37-0.47 for interaction, range across estimation approaches). Under a common effect of screening, all approaches indicated strong evidence that a longer MLT was associated with a lower risk of prostate cancer death after accounting for differential baseline risks of prostate cancer death between trial settings and participant age at randomization (P=0.0027-0.0032). These analyses showed that screening was estimated to confer a 7-9% lower risk of prostate cancer death per year

Table 4.2: Results of traditional and extended Cox regression analyses of prostate cancer death and estimated mortality reductions in the settings of the ERSPC and PLCO intervention arms relative to no screening.

		Cox regression analysis			Estimated mortality reduction relative to no screening					
					Setting of ERSPC intervention arm			Setting of PLCO intervention arm		
	Covariate	HR	95% CI	P	MLT	Reduction	95% CI	MLT	Reduction	95% CI
Traditional analysis										
	PLCO setting	0.53	(0.45-0.62)	<0.0001						
	Age	1.13	(1.11-1.14)	<0.0001						
	Intervention arm	0.84	(0.73-0.96)	0.0099	n/a	16%	(4-27%)	n/a	16%	(4-27%)
Extended analyses										
Empirical	PLCO setting	0.57	(0.48-0.67)	<0.0001						
	Age	1.13	(1.11-1.14)	<0.0001						
	MLT	0.92	(0.87-0.97)	0.0027	3.96	29%	(11-43%)	4.02	29%	(11-44%)
FHCRC	PLCO setting	0.58	(0.49-0.69)	<0.0001						
	Age	1.13	(1.11-1.14)	<0.0001						
	MLT	0.93	(0.88-0.97)	0.0029	4.00	27%	(10-40%)	4.10	27%	(10-41%)
MISCAN	PLCO setting	0.63	(0.51-0.77)	<0.0001						
	Age	1.13	(1.11-1.14)	<0.0001						
	MLT	0.92	(0.87-0.97)	0.0032	3.49	25%	(9-38%)	4.62	32%	(12-47%)
UMICH	PLCO setting	0.57	(0.48-0.68)	<0.0001						
	Age	1.13	(1.11-1.14)	<0.0001						
	MLT	0.91	(0.85-0.97)	0.0029	3.83	31%	(12-45%)	4.01	32%	(12-47%)

HR=hazard ratio; CI=confidence interval; PLCO setting=indicator of PLCO setting relative to the ERSPC setting to account for differential baseline risk of prostate cancer death; Age=participant age at randomization (continuous); Intervention arm=indicator of randomization to intervention arm; MLT=mean lead time (continuous) estimated in each trial arm by the specified estimation approach; FHCRC=Fred Hutchinson Cancer Research Center; MISCAN=Erasmus University Medical Center Microsimulation Screening Analysis; UMICH=University of Michigan

of MLT. Let R_{risk} be the risk reduction in the expected risk of prostate cancer death by lead time, using the formula

$$R_{risk} = 1 - HR^{MLT}, \tag{4.6}$$

this would translate into an estimated 25-31% and 27-32% reduction in the expected risk of prostate cancer death in the setting of screening as performed in the ERSPC and PLCO intervention arms, respectively, over 11 years of follow-up relative to no screening.

Table 4.3 presents both traditional and extended analyses of all cause of death survival models, accounting for the MLT in each trial arm estimated empirically or using a model-based approach. The analyses are highly consistent and indicate no evidence of different effects of screening on mortality of all death between trials (P=0.36-0.91 for interaction, range across estimation approaches). Under a common effect of screening, all approaches indicated that a longer MLT was associated with a

slightly lower risk of prostate cancer death (hazard ratio 0.96-0.98) after accounting for differential baseline risks of cancer death between trial settings and participant age at randomization. These analyses showed that screening test reduced the risk of prostate cancer death specifically, instead of all cause of cancer death.

Table 4.3: Results of traditional and extended Cox regression analyses of all cause of cancer death.

	Cox regression analysis			
	Covariate	HR	95% CI	P-value
Traditional analysis				
	PLCO setting	0.53	(0.52-0.54)	<0.0001
	Age	1.10	(1.09-1.11)	<0.0001
	Intervention arm	0.96	(0.94-0.98)	<0.0001
Extended analyses				
Empirical	PLCO setting	0.54	(0.53-0.55)	<0.0001
	Age	1.10	(1.09-1.11)	<0.0001
	MLT	0.98	(0.97-0.99)	<0.0001
FHCRC	PLCO setting	0.54	(0.53-0.55)	<0.0001
	Age	1.10	(1.09-1.11)	<0.0001
	MLT	0.98	(0.97-0.99)	<0.0001
MISCAN	PLCO setting	0.55	(0.53-0.56)	<0.0001
	Age	1.10	(1.09-1.11)	<0.0001
	MLT	0.98	(0.97-0.99)	<0.0001
UMICH	PLCO setting	0.53	(0.52-0.55)	<0.0001
	Age	1.10	(1.09-1.10)	<0.0001
	MLT	0.97	(0.96-0.98)	0.0029

HR=hazard ratio; CI=confidence interval; PLCO setting=indicator of PLCO setting relative to the ERSPC setting to account for differential baseline risk of cancer death; Age=participant age at randomization (continuous); Intervention arm=indicator of randomization to intervention arm; MLT=mean lead time (continuous) estimated in each trial arm by the specified estimation approach; FHCRC=Fred Hutchinson Cancer Research Center; MISCAN=Erasmus University Medical Center Microsimulation Screening Analysis; UMICH=University of Michigan

Figure 4.2 illustrates prostate cancer survival from randomization in each trial arm obtained by Kaplan-Meier estimation and predicted under a common effect of screening given MLTs estimated by the empirical approach. Predictions obtained using MLTs estimated by the model-based approaches (not shown) are similar. The predicted curves closely reproduce observed differences in prostate cancer survival between the intervention and control arms in both trials, showing that screening intensity as captured by the MLT is highly informative about between-arm differences

in risks of prostate cancer death in both trials.

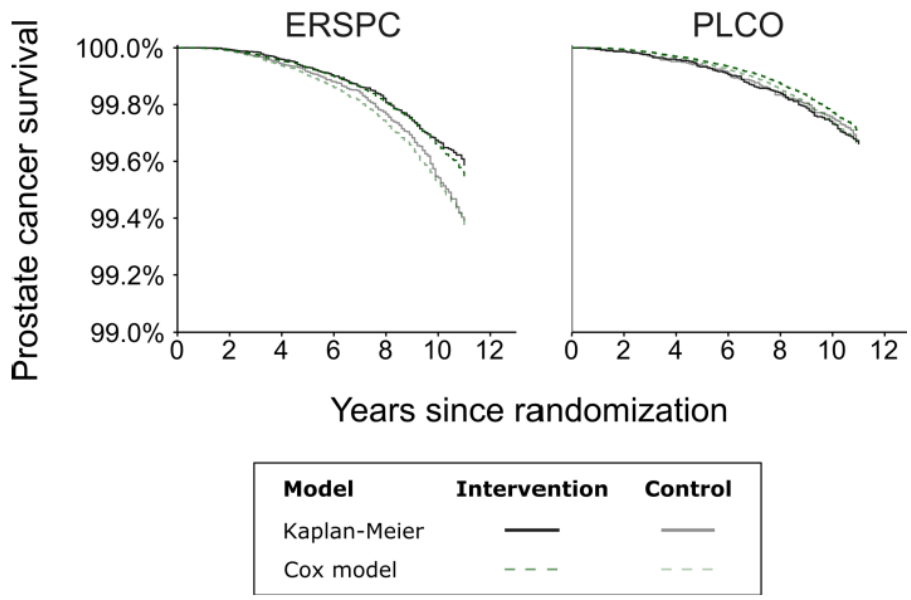

Figure 4.2: Prostate cancer survival from randomization in the ERSPC and PLCO estimated by Kaplan-Meier or Cox regression model using mean lead time estimated by the empirical approach.

4.4 Discussion

The USPSTF is currently updating its recommendations about PSA screening and it has looked at the ERSPC and PLCO at the main sources of evidence about screening benefit in the past. Primary publications from these high-quality randomized controlled trials are irreplaceable for evaluating causal effects of screening for prostate cancer. Yet analyses like the one in this article that attempt to overcome limitations of traditional statistical analyses critically complement the empirical trial findings, including informing about whether the evidence from the trials is compatible and about the expected reduction in prostate cancer mortality relative to no screening.

Rather than comparing the trial arms as if they represent screened and non-screened populations, this study estimated the intensity of screening in each arm relative to no screening. This framework allowed us to formally assess whether screening

effects differed between the trials when accounting for differential screening intensity between arms in each trial. By decoupling screening intensity from trial arm labels and investigating how benefit depends on screening intensity, we concluded that differences between the ERSPC and PLCO results are largely attributable to differences in screening intensities between arms within each trial. Finding no evidence of different effects of screening on prostate cancer mortality between trials given the screening intensities, we estimated a common effect of screening on mortality using pooled data on 19,226 prostate cancer cases. The pooled estimate demonstrates a highly significant benefit of screening. This is the first time that data from both trials have been harnessed to estimate screening benefit.

It is possible that this analysis had insufficient power to detect a significant difference in screening efficacy between trials. Thus, while there is no evidence of different screening efficacies, we cannot unequivocally conclude they were identical. Nevertheless, our combined analysis of both trials permits the most powerful examination of this question to date.

Our analysis indicates that the baseline risk of prostate cancer death differed between trials. This could be due to different incidence, stage distributions, and treatment patterns in the trial populations in the absence of screening. A lower-than-expected mortality (relative to pre-PSA-era survival) was observed in the PLCO, possibly due to participants being healthier or reflecting an era with improved disease-specific survival (Pinsky et al., 2007). By quantifying screening efficacy as a function of screening intensity, we projected that screening lowered the expected risk of prostate cancer mortality in both PLCO arms.

We used multiple approaches to estimate screening intensity. The empirical approach reflects catch-all differences in the risk of prostate cancer diagnosis between arms and calculates the MLT most consistent with incidence in each arm relative to a common baseline level. In contrast, the model-based approaches explicitly account

for trial protocols and practice settings details that are known or can be quantified, e.g., age distributions, enrollment and attendance patterns, and screening and biopsy frequencies within each ERSPC center. As expected, the estimates are shorter than in other studies (Draisma et al., 2009) due to the different estimation approach and because we restricted to 11 years of follow-up. In general, results are highly consistent across estimation approaches and suggest robustness of our conclusions to these ways of estimating screening intensity.

The finding that effects of screening on mortality appear to be consistent between trials after accounting differences in implementation and practice setting corroborate other analyses. For example, a prior investigation of the PLCO found that control arm screening substantially limited the power of that trial to detect a clinically important reduction in prostate cancer mortality (Gulati et al., 2012). However, that study did not formally evaluate whether effects of screening on prostate cancer mortality differed between the ERSPC and PLCO when implementation and setting details are taken into account.

A limitation of this study is that we do not explicitly account for differences between trials in characteristics of cancer cases (e.g., clinical stage or Gleason score) or primary treatments. Any differences in these factors between trials will be absorbed into the trial-specific baseline risks of prostate cancer death. Also, the model-based approaches to estimate lead times assume that cancers are progressive, although they allow heterogeneity in progression risk across individuals. Ultimately, it is impossible to know whether some cancers could remain indolent indefinitely or regress spontaneously and permanently. However, all estimation approaches closely match incidence trends in each trial arm. We also assume that incidence in the absence of screening was constant across calendar years before and after the trials began. This too is a simplification. Finally, we consider only the mean lead time as a surrogate for screening intensity. It is possible that other metrics could have different associations with

risk of prostate cancer death than we found.

In conclusion, taken together, the data from the two screening trials do not provide evidence that screening efficacy differed between the ERSPC and PLCO after accounting for differences in implementation between arms in the trials. Our estimation results of the common effect of screening suggest that screening can significantly reduce the risk of prostate cancer death. However, as for all interventions, the benefit of screening must be weighed against its potential harms for informed clinical and shared decision making.

CHAPTER V

Conclusion and Future Work

In this dissertation, we have formulated cancer incidence and mortality models by incorporating cancer screening patterns with disease progression to the events of cancer diagnosis and death. With these model-based methodologies, we are able to estimate cancer screening effect on both cancer incidence and mortality using cancer registries and screening trial data, especially when the screening patterns are random and unobserved. Instead of intent-to-screen analysis, which is less powerful to evaluate causal effect of cancer screening on reducing risk of death and prolonging post-diagnosis survival with screening contamination in the control arm, we proposed the term mean lead time to quantify the screening intensities and estimate the screening efficacy after adjusting for the contamination. The mean lead time is predicted by the incidence models with diagnosis-free survival functions with and in the absence of screening, incorporated with both disease progression features and screening patterns.

In Chapter II, we have proposed a general model to quantitatively link cancer control processes in the population with unknown schedules to the impact on cancer incidence, jointly with the disease natural progression stages. Three baseline risks are nonparametrically specified to the latent tumor onset (S), cancer detection by symptoms (CDx) and screening (SDx) respectively. Given tumor onset, the two

modes of diagnoses are dependent with a shared frailty term A. We demonstrated an example with a proportional hazards assumption between S and CDx and a gamma frailty model between CDx and SDx. As the model is generally specified, independent specification of these baseline hazards and more complex frailty model with larger frailty effect can be adapted into the data analysis, given sufficient identifiability of model parameters.

Given the basis work in Chapter II, we extended the joint model with a mixture of cancer screening schedules to the analysis of cancer screening trials data. The extended model enables us to estimate ad-hoc screening intensity (contamination) jointly with the disease natural history an perform hypotheses testing for the screening patterns across trial until loss of follow up. In addition, we can obtain the diagnosis-free survival functions by adapting the trial settings and mixture of screening patterns, as well as the natural disease progression features. For model simplification, in this chapter we have assumed the common baseline hazard between tumor onset S and clinical diagnosis CDx, and conditional dependence between CDx and SDx given S. The two assumptions may be released by specifying different baseline hazard and introducing frailty term into the model.

To link the formulated incidence models to mortality models for evaluating screening effect on reducing the risk of cancer death, we used the term mean lead time as the proxy of screening intensity. By fitting the prostate cancer specific survival models, we found that screening can significantly reduce the risk of prostate cancer death. As the lead time was predicted with the incidence models, its standard error can also be estimated for statistical inference purpose. In addition, because of the dependency between time to CDx and time to SDx, more distributional characteristics can be studied on the mean lead time with copula models.

APPENDICES

APPENDIX A

Supplementary Materials for Chapter II

A.1 Derivation of Model Quantities

The p.d.f of tumor onset S is:

$$f(s)ds = G(s)d\Lambda_0(s) = e^{-\int_0^s \theta(\xi)dH_0(\xi)}\theta(s)dH_0(s). \tag{A.1}$$

Generally, when the conditional hazard function for a survival time T is a stochastic process $\lambda(t)$, then the marginal survival function $G(t) = \mathbb{E}[e^{-\int_0^t \lambda(\xi)d\xi}]$. Extending the formulation of Rice and Tsodikov (2016) to time-dependent predictors, we obtain the marginal clinical diagnosis-free survival function in the form

$$\begin{aligned} G_d(t|\bar{Z}_{tx}) &= \mathbb{E}\left[e^{-\int_0^t d\Lambda_1(\xi|Z_{\xi x},S)}\right] \\ &= \int_0^t e^{-\int_s^t \eta(\xi)dH_1(\xi)}f(s)ds + \int_t^\infty e^{-\int_s^t \eta(\xi)dH_1(\xi)}f(s)ds \\ &= \int_0^t e^{-\int_s^t \eta(\xi)dH_1(\xi)-\int_0^s \theta(\xi)dH_0(\xi)}\theta(s)dH_0(s) + \int_t^\infty e^{-\int_0^s \theta(\xi)dH_0(\xi)}\theta(s)dH_0(s) \\ &= e^{-\int_0^t \theta(\xi)dH_0(\xi)} + \int_0^t e^{-\int_s^t \eta(\xi)dH_1(\xi)-\int_0^s \theta(\xi)dH_0(\xi)}\theta(s)dH_0(s). \end{aligned} \tag{A.2}$$

We obtain the marginal survival function incorporating risks of screening and clinical diagnosis as the expectation over the distribution of latent event time S, frailty term A and screening pattern N_{scr}. Let $\tilde{\Lambda}_{s,t} = \int_0^t I(\xi \geq s) \left[\eta(Z_{\xi x})dH_1(\xi) + \alpha(\xi|x)dH_{scr}(\xi|x)\right]$.

$$\begin{aligned}
G_d(t|\bar{Z}_{tx}) &= E_{S,A,N_{scr}}\left[e^{-\int_0^t \left(d\Lambda_1(\xi|Z_{\xi x},S,A)+d\Lambda_2(\xi|Z_{\xi x},S,A,N_{scr})\right)}\right] \\
&= E_{S,A,N_{scr}}\left[e^{-\int_0^t AI(\xi\geq S)\left[\eta(\xi)dH_1(\xi)-\log(1-\alpha(\xi|x))dN_{scr}(\xi|x)\right]}\right] \\
&= E_S\left\{E_A\left[e^{-\int_0^t AI(\xi\geq S)\left[\eta(\xi)dH_1(\xi)+\alpha(\xi|x)dH_{scr}(\xi|x)\right]}\right]\right\} \\
&= E_S\left\{\mathcal{L}^{(0)}(\tilde{\Lambda}_{s,t})\right\} \\
&= e^{-\Lambda_0(t)} + \int_0^t \mathcal{L}^{(0)}(\tilde{\Lambda}_{s,t})e^{-\Lambda_0(s)}d\Lambda_0(s).
\end{aligned} \tag{A.3}$$

The the corresponding p.d.f is:

$$\begin{aligned}
g_d(t|\bar{Z}_{tx}) &= -\frac{dG_d(t|\bar{Z}_t)}{dt} \\
&= -\left\{-e^{-\Lambda_0(t)}d\Lambda_0(t) + \int_0^t \mathcal{L}^{(0)'}(\tilde{\Lambda}_{s,t})e^{-\Lambda_0(s)}d\Lambda_0(s) + \mathcal{L}^{(0)}(0)e^{-\Lambda_0(t)}d\Lambda_0(t)\right\} \\
&= -\left\{\int_0^t \mathcal{L}^{(0)'}(\tilde{\Lambda}_{s,t})e^{-\int_0^s \theta(Z_{sx})dH_0(s)}\theta(Z_{sx})dH_0(s)\right\} \\
&= -\int_0^t \mathcal{L}^{(1)}(\tilde{\Lambda}_{s,t})e^{-\Lambda_0(s)}d\Lambda_0(s)\left[\eta(Z_{tx})dH_1(t) + \alpha(t|x)dH_{scr}(t|x)\right],
\end{aligned} \tag{A.4}$$

and

$$\lambda(t|\bar{Z}_{tx}) = \frac{g_d(t|\bar{Z}_{tx})}{G_d(t|\bar{Z}x_t)} = \Psi(t|\bar{Z}_{tx})\left[\eta(Z_{tx})dH_1(t) + \alpha(t|x)dH_{scr}(t|x)\right], \tag{A.5}$$

where

$$\Psi(t|\bar{Z}_{tx}) = \frac{-\int_0^t \mathcal{L}^{(1)}(\tilde{\Lambda}_{s,t})e^{-\Lambda_0(s)}d\Lambda_0(s)}{e^{-\Lambda_0(t)} + \int_0^t \mathcal{L}^{(0)}(\tilde{\Lambda}_{s,t})e^{-\Lambda_0(s)}d\Lambda_0(s)}.$$

A.2 Derivation of Score Equations

$$\ell = \sum_x \int_0^{\upsilon} \sum_{i \in \mathcal{R}(t|x)} \log\left(d\Lambda(t|\bar{Z}_{txi})\right) dN(t|\bar{Z}_{txi}) - d\Lambda(t|\bar{Z}_{txi}),$$

Follow (Rice and Tsodikov, 2016), we impose a proportional hazards (PH) assumption linking dH_1 and dH_0 into a common dH. Given the maximum follow-up time υ, differentiating the log-likelihood (2.10) we arrive at the following score equations for $dH(\tau)$, $dH_{scr}(\tau|x)$, where $\tau \in (0, t]$.

$$\begin{aligned}
\mathcal{U}_{dH(\tau)} =& \frac{\partial \ell}{\partial dH(\tau)} \\
=& \sum_x \sum_{i \in \mathcal{R}(\tau|x)} \frac{\partial \log d\Lambda(\tau|\bar{Z}_{\tau xi})}{\partial dH(\tau)} dN(\tau|\bar{Z}_{\tau xi}) - \frac{\partial d\Lambda(\tau|\bar{Z}_{\tau xi})}{\partial dH(\tau)} \\
&+ \int_{\tau+}^{\upsilon} \sum_{j \in \mathcal{R}(t|x)} \frac{\partial \log d\Lambda(t|\bar{Z}_{txj})}{\partial dH(\tau)} dN(t|\bar{Z}_{txj}) - \frac{\partial d\Lambda(t|\bar{Z}_{txj})}{\partial dH(\tau)} \\
=& \sum_x \sum_{i \in \mathcal{R}(\tau|x)} \left(\frac{\Psi_{dH(\tau)}(\tau|\bar{Z}_{\tau xi})}{\Psi(\tau|\bar{Z}_{\tau xi})} + \frac{\eta_i(\tau)}{\eta_i(\tau) dH(\tau) + dH_{scr}(\tau|x)} \right) dN(\tau|\bar{Z}_{\tau xi}) \\
&- \frac{\Psi_{dH(\tau)}(\tau|\bar{Z}_{\tau xi})}{\Psi(\tau|\bar{Z}_{\tau xi})} d\Lambda(\tau|\bar{Z}_{\tau xi}) - \Psi(\tau|\bar{Z}_{\tau xi})\eta_i(\tau) \\
&+ \int_{\tau+}^{\upsilon} \sum_{j \in \mathcal{R}(t|x)} \frac{\Psi_{dH(\tau)}(t|\bar{Z}_{txj})}{\Psi(t|\bar{Z}_{txj})} dN(t|\bar{Z}_{txj}) - \frac{\Psi_{dH(\tau)}(t|\bar{Z}_{txj})}{\Psi(t|\bar{Z}_{txj})} d\Lambda(t|\bar{Z}_{txj}) \\
=& \sum_x \sum_{i \in \mathcal{R}(\tau|x)} \frac{\eta_i(\tau) dN(\tau|\bar{Z}_{\tau xi})}{\eta_i(\tau) dH(\tau) + dH_{scr}(\tau|x)} \\
&- \sum_{i \in \mathcal{R}(\tau|x)} \Psi(\tau|\bar{Z}_{\tau xi})\eta_i(\tau) \left(1 - \frac{\int_{\tau}^{\upsilon} \sum_{j \in \mathcal{R}(t|x)} \frac{\Psi_{dH(\tau)}(t|\bar{Z}_{txj})}{\Psi(t|\bar{Z}_{txj})} dM(t|\bar{Z}_{txj})}{\sum_{i \in \mathcal{R}(\tau|x)} \Psi(\tau|\bar{Z}_{\tau xi})\eta_i(\tau)} \right) \\
=& \sum_x \sum_{i \in \mathcal{R}(\tau|x)} \frac{\eta_i(\tau) dN(\tau|\bar{Z}_{\tau xi})}{\eta_i(\tau) dH(\tau) + dH_{scr}(\tau|x)} - \sum_{i \in \mathcal{R}(\tau|x)} \Psi(\tau|\bar{Z}_{\tau xi})\eta_i(\tau) w_{dH(\tau)},
\end{aligned}$$

and

$$\Psi_{dH(\tau)}(t|\bar{Z}_{txi}) = \frac{a_{dH}B - Ab_{dH}}{\left\{e^{-\Lambda_0(t)} + \int_0^t \mathcal{L}^{(0)}(\tilde{\Lambda}_{s,t})e^{-\Lambda_0(s)}d\Lambda_0(s)\right\}^2} = \frac{a_{dH}b - ab_{dH}}{b^2}, \qquad (A.6)$$

where

$$\begin{aligned}
a_{dH}b - ab_{dH} = &\left\{-\int_0^\tau (1+\phi)(1+\phi\tilde{\Lambda}_{s,t})^{-\frac{1}{\phi}-2}\eta(\tau)e^{-\Lambda_0(s)}d\Lambda_0(s) - (1+\phi\tilde{\Lambda}_{\tau,t})^{-\frac{1}{\phi}-1}e^{-\Lambda_0(\tau)}d\Lambda_0(\tau)\right.\\
&\left.+(1+\phi\tilde{\Lambda}_{\tau,t})^{-\frac{1}{\phi}-1}e^{-\Lambda_0(\tau)} - \int_{\tau^+}^t (1+\phi\tilde{\Lambda}_{s,t})^{-\frac{1}{\phi}-1}e^{-\Lambda_0(s)}d\Lambda_0(s)\right\}\\
&\cdot\left\{e^{-\Lambda_0(t)} + \int_0^t (1+\phi\tilde{\Lambda}_{s,t})^{-\frac{1}{\phi}}e^{-\Lambda_0(s)}d\Lambda_0(s)\right\} - \left\{\int_0^t (1+\phi\tilde{\Lambda}_{s,t})^{-\frac{1}{\phi}-1}e^{-\Lambda_0(s)}d\Lambda_0(s)\right\}\\
&\cdot\left\{-e^{-\Lambda_0(t)} - \int_0^\tau (1+\phi\tilde{\Lambda}_{s,t})^{-\frac{1}{\phi}-1}\eta(\tau)e^{-\Lambda_0(s)}d\Lambda_0(s) - (1+\phi\tilde{\Lambda}_{\tau,t})^{-\frac{1}{\phi}}e^{-\Lambda_0(\tau)}d\Lambda_0(\tau)\right.\\
&\left.+(1+\phi\tilde{\Lambda}_{\tau,t})^{-\frac{1}{\phi}}e^{-\Lambda_0(\tau)} - \int_{\tau^+}^t (1+\phi\tilde{\Lambda}_{s,t})^{-\frac{1}{\phi}}e^{-\Lambda_0(s)}d\Lambda_0(s)\right\}.
\end{aligned}$$

Similarly, for the screening intensity dH_{scr} of x-birth cohort,

$$\begin{aligned}
\mathcal{U}_{dH_{scr}(\tau|x)} =&\frac{\partial \ell}{\partial dH_{scr}(\tau|x)}\\
=&\sum_{i\in\mathcal{R}(\tau|x)} \frac{\partial \log d\Lambda(\tau|\bar{Z}_{\tau xi})}{\partial dH_{scr}(\tau|x)}dN(\tau|\bar{Z}_{\tau xi}) - \frac{\partial d\Lambda(\tau|\bar{Z}_{\tau xi})}{\partial dH_{scr}(\tau|x)}\\
&+\int_{\tau+}^{\upsilon}\sum_{j\in\mathcal{R}(t|x)} \frac{\partial \log d\Lambda(t|\bar{Z}_{txj})}{\partial dH_{scr}(\tau|x)}dN(t|\bar{Z}_{txj}) - \frac{\partial d\Lambda(t|\bar{Z}_{txj})}{\partial dH_{scr}(\tau|x)}\\
=&\sum_{i\in\mathcal{R}(\tau|x)}\left(\frac{\Psi_{dH_{scr}(\tau|x)}(\tau|\bar{Z}_{\tau xi})}{\Psi(\tau|\bar{Z}_{\tau xi})} + \frac{1}{\eta_i(\tau)dH(\tau) + dH_{scr}(\tau|x)}\right)dN(\tau|\bar{Z}_{\tau xi})\\
&-\frac{\Psi_{dH_{scr}(\tau|x)}(\tau|\bar{Z}_{\tau xi})}{\Psi(\tau|\bar{Z}_{\tau xi})}d\Lambda(\tau|\bar{Z}_{\tau xi}) - \Psi(\tau|\bar{Z}_{\tau xi})\\
&+\int_{\tau+}^{\upsilon}\sum_{j\in\mathcal{R}(t|x)} \frac{\Psi_{dH_{scr}(\tau|x)}(t|\bar{Z}_{txj})}{\Psi(t|\bar{Z}_{txj})}dN(t|\bar{Z}_{txj}) - \frac{\Psi_{dH_{scr}(\tau|x)}(t|\bar{Z}_{txj})}{\Psi(t|\bar{Z}_{txj})}d\Lambda(t|\bar{Z}_{txj})\\
=&\sum_{i\in\mathcal{R}(\tau|x)} \frac{dN(\tau|\bar{Z}_{\tau xi})}{\eta_i(\tau)dH(\tau) + dH_{scr}(\tau|x)}
\end{aligned}$$

$$
\begin{aligned}
&- \sum_{i\in\mathcal{R}(\tau|x)} \Psi(\tau|\bar{Z}_{\tau xi}) \left(1 - \frac{\int_\tau^\upsilon \sum_{j\in\mathcal{R}(t|x)} \frac{\Psi_{dH_{scr}(\tau|x)}(t|\bar{Z}_{txj})}{\Psi(t|\bar{Z}_{txj})} dM(t|\bar{Z}_{txj})}{\sum_{i\in\mathcal{R}(\tau|x)} \Psi(\tau|\bar{Z}_{\tau xi})} \right) \\
&= \sum_{i\in\mathcal{R}(\tau|x)} \frac{dN(\tau|\bar{Z}_{\tau xi})}{\eta_i(\tau)dH(\tau) + dH_{scr}(\tau|x)} - \sum_{i\in\mathcal{R}(\tau|x)} \Psi(\tau|\bar{Z}_{\tau xi}) w_{dH_{scr}(\tau|x)},
\end{aligned}
$$

and

$$
\Psi_{dH_{scr}(\tau|x)}(t|\bar{Z}_{txi}) = \frac{a_{dH_{scr}} b - a b_{dH_{scr}}}{\left\{ e^{-\Lambda_0(t)} + \int_0^t \mathcal{L}^{(0)}(\tilde{\Lambda}_{s,t}) e^{-\Lambda_0(s)} d\Lambda_0(s) \right\}^2} = \frac{a_{dH_{scr}} b - a b_{dH_{scr}}}{b^2}, \tag{A.7}
$$

where

$$
\begin{aligned}
a_{dH_{scr}} b - a b_{dH_{scr}} = &\left\{ - \int_0^\tau (1+\phi)(1+\phi\tilde{\Lambda}_{s,t})^{-\frac{1}{\phi}-2} e^{-\Lambda_0(s)} d\Lambda_0(s) \right\} \\
&\cdot \left\{ e^{-\Lambda_0(t)} + \int_0^t (1+\phi\tilde{\Lambda}_{s,t})^{-\frac{1}{\phi}} e^{-\Lambda_0(s)} d\Lambda_0(s) \right\} - \left\{ \int_0^t (1+\phi\tilde{\Lambda}_{s,t})^{-\frac{1}{\phi}-1} e^{-\Lambda_0(s)} d\Lambda_0(s) \right\} \\
&\cdot \left\{ - \int_0^\tau (1+\phi\tilde{\Lambda}_{s,t})^{-\frac{1}{\phi}-1} e^{-\Lambda_0(s)} d\Lambda_0(s) \right\}.
\end{aligned}
$$

A.3 Asymptotics Properties of NPMLE Estimators

A.3.1 Martingale properties

We write the score functions for $\Omega = (\boldsymbol{\beta}, \{dH\}, \{dH_{scr}\})$ as martingales (at the true Ω^0). Let $\gamma^\tau_{iH(t)} = \frac{\Psi_{dH(\tau)}(t|\bar{Z}_{txi})}{\Psi(t|\bar{Z}_{txi})}, \gamma^\tau_{iH_{scr(t|x)}} = \frac{\Psi_{dH_{scr}(\tau|x)}(t|\bar{Z}_{txi})}{\Psi(t|\bar{Z}_{txi})}$. For simplicity, let $d\Lambda_i(t) = d\Lambda(t|\bar{Z}_{txi}), dN_i(t) = dN(t|\bar{Z}_{txi}), dM_i(t) = dM(t|\bar{Z}_{txi}), \Psi_i(t) = \Psi(t|\bar{Z}_{txi})$. Here x is dropped from arguments except when it indexes a parameter, i.e. $dH_{scr}(t|x)$.

$$
\mathcal{U}_{\boldsymbol{\beta}} = \sum_x \int_0^\upsilon \sum_{i\in\mathcal{R}(s|x)} \frac{\partial \log d\Lambda_i(s)}{\partial \boldsymbol{\beta}} dM_i(s),
$$

and

$$\mathcal{U}_{H(t)} = \int_0^t \sum_x \sum_{i \in \mathcal{R}(s|x)} \left\{ dM_i(s) + \int_{s^+}^{\upsilon} \sum_{j \in \mathcal{R}(u|x)} \gamma^s_{iH(u)} dM_j(u) \left[\eta_i(s) dH(s) + dH_{scr}(s|x) \right] \right\},$$

$$\mathcal{U}_{H_{scr}(t|x)} = \int_0^t \sum_{i \in \mathcal{R}(s|x)} \left\{ dM_i(s) + \int_{s^+}^{\upsilon} \sum_{j \in \mathcal{R}(u|x)} \gamma^s_{iH_{scr(u|x)}} dM_j(u) \left[\eta_i(s) dH(s) + dH_{scr}(s|x) \right] \right\},$$

Exchanging the integrals, we have:

$$\mathcal{U}_{H(t)} = \sum_x \int_0^{\upsilon} \sum_{i \in \mathcal{R}(u|x)} \varepsilon_i(u, t, H; \boldsymbol{\beta}) dM_i(u), \tag{A.8}$$

$$\mathcal{U}_{H_{scr}(t|x)} = \int_0^{\upsilon} \sum_{i \in \mathcal{R}(u|x)} \varepsilon_i(u, t, H_{scr}; \boldsymbol{\beta}) dM_i(u), \tag{A.9}$$

where $\varepsilon_i(u, t, H; \boldsymbol{\beta}) = I(u \leq t) + \left[\int_0^{u \wedge t} \eta_i(s) dH(s) + dH_{scr}(s|x) \right] \gamma^s_{iH(u)}$;
$\varepsilon(u, t, H_{scr}; \boldsymbol{\beta}) = I(u \leq t) + \left[\int_0^{u \wedge t} \eta_i(s) dH(s) + dH_{scr}(s|x) \right] \gamma^s_{iH_{scr(u|x)}}$.
Properties of martingale transform. (A.8) and (A.9) are martingales as $\varepsilon_i(u, t, H; \boldsymbol{\beta}) dM_i(u)$ and $\varepsilon_i(u, t, H_{scr}; \boldsymbol{\beta}) dM_i(u)$ do not depend on t for $u \leq t$. Let $V_i(t) = \int_0^{\upsilon} \varepsilon_i(u, t, H; \boldsymbol{\beta}) dM_i(u)$. Consider an increment of the martingale transform

$$dV_i(t) = \int_0^{\upsilon} \varepsilon_i(u, t + dt) dM_i(u) - \varepsilon(u, t) dM_i(u) = \int_0^{\upsilon} d_t \varepsilon_i(u, t) dM_i(u),$$

where d_t is the partial differential of $\varepsilon(u, t)$ with respect to t. Taking an expectation given filtration $\mathcal{F}_{t-}$,

$$\mathbb{E}\left[dV_i(t) | \mathcal{F}_{t-} \right] = \int_0^{\upsilon} \mathbb{E}\left[d_t \varepsilon(u, t) dM_i(u) | \mathcal{F}_{t-} \right] = \int_0^{\upsilon} d_t \varepsilon(u, t) \, \mathbb{E}\left[dM_i(u) | \mathcal{F}_{t-} \right] = \int_0^{\upsilon} d_t \varepsilon(u, t) dM_i(u).$$

By definition, when $u < t$, $d_t\varepsilon(u,t) = 0$ then we have $\mathbb{E}\left[dV_i(t)|\mathcal{F}_{t-}\right]$=0. Therefore $V_i(t)$ is a martingale. Same proof can be derived for H_{scr} with $V_i(t) = \int_0^v \varepsilon_i(u,t,H_{scr};\boldsymbol{\beta})dM_i(u)$. Therefore, the score functions $\mathcal{U}_{\boldsymbol{\beta}}$, $\mathcal{U}_{H(t)}$ and $\mathcal{U}_{H_{scr}(t|x)}$ are martingales at the true model.

A.3.2 Consistency

We now present the consistency and weak convergence results for the proposed NPMLE estimators $\hat{\Omega} = (\hat{\boldsymbol{\beta}}, \hat{H}(t), \hat{H}_{scr}(t|x))$, mainly adapted from Hu and Tsodikov (2014b). Let $\|\cdot\|_{l^\infty[0,v]}$ denote the supremum norm in $[0,v]$ and $\|\omega\|_{V[0,v]}$ the total variation of $\omega(t)$ in $[0,v]$. Define $\mathcal{Q} = \{\omega(t) : \|\omega\|_{V[0,v]} < \infty\}$ such that $\hat{H}(t)$ and $\hat{H}_{scr}(t|x)$ may be regarded as a bounded linear functional in $l^\infty[\mathcal{Q}]$, and $\Delta\hat{\Omega}^0 = \{\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0, d\hat{H}(t) - dH^0(t), d\hat{H}_{scr}(t|x) - dH^0_{scr}(t|x)\}$ a random element in the metric space $\mathcal{S}^p \times l^\infty[\mathcal{Q}]$, where p is the dimension of $\boldsymbol{\beta}^0$. We denote $\mathcal{H}$ as the compact convex set in the metric space $\mathcal{S}^p \times l^\infty[\mathcal{Q}]$ in which Ω^0 is contained.

Proof. To establish the consistency result from Proposition II.1: $\|\hat{H}(\cdot) - H^0(\cdot)\|_{l^\infty[0,v]} \xrightarrow{p} 0$, $\|\hat{H}_{scr}(\cdot|x) - H^0_{scr}(\cdot|x)\|_{l^\infty[0,v]} \xrightarrow{p} 0$ and $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\| \xrightarrow{p} 0$, the following conditions are verified:

1. *Identifiability condition*: The model is identifiable so that $\Lambda = \Lambda^0$ uniformly over Ω implies $\Omega = \Omega^0$, which will ensure that for any sequence $\Omega_n \in \mathcal{H}$, the compact convex set in the metric space $\mathcal{S}^p \times l^\infty[\mathcal{Q}]$, $\liminf_{n\to\infty} \ell(\Omega_n) \geq \ell(\Omega^0)$ implies $\|\Omega_n - \Omega^0\| \xrightarrow{p} 0$.

2. *Uniform convergence condition*: for any sequence $\Omega \in \mathcal{H}$ we have uniform convergence,

$$\sup_{\Omega\in\mathcal{H}} |\ell_n(\Omega) - \ell(\Omega)| \xrightarrow{p} 0.$$

With these two conditions, since $\ell_n(\hat{\Omega}) = sup_{\Omega\in\mathcal{H}}\ell_n(\Omega)+o_p(1)$, then based on Theorem 2.12 in Kosorok (2008), we have $\|\hat{\Omega} - \Omega^0\| \xrightarrow{p} 0$. We use three steps to verify these

two conditions.

Step 1 Convexity and unique maximum of the likelihood function ℓ.

Let $d\Lambda(t) = d\Lambda(t|\bar{Z}_{tx})$, the model can be characterized through corresponding hazard function as

$$d\Lambda(t) = \Psi(\Omega)\left[\eta(t) \cdot dH(t) + dH_{scr}(t|x)\right],$$

which are all functionals that depend on the processes $H(\cdot)$ and $H(\cdot|x)$ on $[0, t]$. Let $F(t)$ be the cumulative incidence function , $R(t)$ be the survival function respectively. Thus $dF(t) = R(t)d\Lambda(t)$, we can rewrite the true likelihood as

$$\ell(\Omega, \Omega^0) = \mathbb{E}\int_0^{\upsilon} \log d\Lambda(t)dF^0(t) - R^0(t)d\Lambda(t),$$

where R^0 and F^0 denote the corresponding true quantities respectively and the expectation is taken with respect to the covariate process Z_{tx}.

Consider the negative true Kullback-Leibler distance,

$$D = \ell(\Omega, \Omega^0) - \ell(\Omega^0, \Omega^0).$$

We have

$$D = \mathbb{E}\int_0^{\upsilon} \left[\log \frac{d\Lambda(t)}{d\Lambda^0(t)} - \frac{d\Lambda(t)}{d\Lambda^0(t)} + 1\right] dF^0(t).$$

consider a non-positive concave function $f(t) = \log t - t + 1$, $t = 1$ is the unique maximizer and $f(1) = 0$. Therefore, D has a uniformly unique maximum when $d\Lambda(t|x) = d\Lambda^0(t|x)$.

Step 2 Identifiability conditions.

Since Λ is assumed to be a differentiable functional of H and H_{scr}, so is the likelihood function $\ell(\Omega)$. Step 1 suggests that $\Omega^0 = \text{argmax}_{\Omega\in\mathcal{H}}\ell(\Omega)$ is unique, and $\ell(\Omega, \Omega^0)$ is identifiable so that $\Lambda = \Lambda^0$ uniformly over Ω implies $\Omega = \Omega^0$. Therefore, based on

Lemma 14.3 of Kosorok (2008), we have $\liminf_{n\to\infty} \ell(\Omega_n) \geq \ell(\Omega^0)$, thus the identifiability condition is satisfied.

Step 3 Uniform convergence condition.

Assume that Ω is in the class of functions of bounded variation with integrable envelope, which implies that H and H_{scr} are bounded. Therefore, $\mathcal{H}$ belongs to a Glivenko-Cantelli class, whose ϵ-entropy with bracketing number is bounded by A/ϵ, where A is some constant. By the assumption of continuity of the functionals Λ and $\ell(\Omega)$, and the integrability of the envelope of Ω, the integrand in $\ell(\Omega)$ is also Glivenko-Cantelli. Therefore, we apply the uniform law of large numbers for the empirical process counterpart of D and ℓ as

$$D_n = \ell_n(\Omega, \Omega^0) - \ell_n(\Omega^0, \Omega^0)$$

where

$$\ell_n(\Omega, \Omega^0) = n^{-1} \sum_x \left\{ \int_0^v \sum_{i \in \mathcal{R}(t|x)} \left[\log(\eta(t) dH(t) + dH_{scr}(t|x)) + \log \Psi(H(t), H_{scr}(t|x); \boldsymbol{\beta}) \right] dN_i(t) \right.$$
$$\left. - \Psi(H(t), H_{scr}(t|x); \boldsymbol{\beta}) [\eta_i(t) dH(t) + dH_{scr}(t|x)] \right\},$$

and $n = \sum_x \left\| \mathcal{R}(t|x) \right\|$,

Such that

$$\sup_{\Omega \in \mathcal{H}} |D_n(\Omega) - D(\Omega)| \xrightarrow{p} 0, \quad \sup_{\Omega \in \mathcal{H}} |\ell_n(\Omega) - \ell(\Omega)| \xrightarrow{p} 0.$$

and this completes the verification of uniform convergence condition in Proposition II.1. □

A.3.3 Weak convergence

Consider a linear functional

$$n^{1/2}\left[\mathbf{a}'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0)+\int_0^{\upsilon} b(t)d\left(\hat{H}(t)-H^0(t)\right)+\int_0^{\upsilon} e(t|x)d\left(\hat{H}_{scr}(t|x)-H^0_{scr}(t|x)\right)\right],$$

where $\mathbf{a}$ is a real vector and $b(t), e(t|x)$ are functions with bounded total variation. Let $\mathbf{B}$ denote the vector consisting of the values of $b(t)$ evaluated at the observed failure times corresponding to the set $\{dH\}$, let $\mathbf{E}$ denote the vector consisting of the values of $e(t|x)$ evaluated at the observed failure times corresponding to the set $\{dH_{scr}\}$ for cohort x; let $\boldsymbol{\mathcal{E}}' = (\mathbf{a}', \mathbf{B}', \mathbf{E}')$. We have Proposition II.2.

Proof. Our proof closely follows that of Hu and Tsodikov (2014a, Supplementary Materials C) and Rice and Tsodikov (2016, Supplementary Materials E). Denote the scores as $\mathcal{U}(\Omega) = \left(\mathcal{U}'_{\boldsymbol{\beta}}, \mathcal{U}_{H(t)}, \mathcal{U}_{H_{scr}(t|x)}\right)'$, and the proposed NPMLE $\hat{\Omega}$ be the solution to the equation $\mathcal{U}(\Omega) = \mathbf{0}$. Note that in our case this solution involves the profile likelihood for $\boldsymbol{\beta}$:

$$\ell_{\text{pr}}(\boldsymbol{\beta}) = \sup_{H, H_{scr}} \ell(H(\cdot), H_{scr}(\cdot|\cdot); \boldsymbol{\beta}),$$

where ℓ is defined in equation (2.10) and the estimate of H, H_{scr} are obtained using the iterative reweighted algorithm of Chen (2009). Asymptotically, this is equivalent to solving the marginal score directly.

Based on the martingale representation of $\mathcal{U}(\Omega^0)$ where Ω^0 is the set of true parameters, and the fact that $N_i(t), i = 1, ..., n$ are orthogonal, it follows by the martingale central limit theorem that $n^{-1/2}\mathcal{U}(\Omega^0)$ converges weakly to $U(t) = \left(\mathbf{U}'_{\boldsymbol{\beta}}, U_{H(t)}, U_{H_{scr}(t|x)}\right)'$, where $\mathbf{U}_{\boldsymbol{\beta}}$ is a mean-zero p-variate normal random variable and $U_{H(t)}, U_{H_{scr}(t|x)}$ are mean-zero Gaussian processes. The variance-covariance function of $U(t)$ can be characterized by covariance functions $\sigma^2_{H_{scr}}(s, t; \boldsymbol{\beta}^0, H^0, H^0_{scr})$, $\sigma^2_H(s, t; \boldsymbol{\beta}^0, H^0, H^0_{scr})$, $\sigma^2_{\boldsymbol{\beta}}(\boldsymbol{\beta}^0)$, $\sigma^2_{H_{scr},H}(s, t; \boldsymbol{\beta}^0, H^0, H^0_{scr})$, $\sigma^2_{H,\boldsymbol{\beta}}(t; \boldsymbol{\beta}^0, H^0, H^0_{scr})$, and $\sigma^2_{H_{scr},\boldsymbol{\beta}}(t; \boldsymbol{\beta}^0, H^0, H^0_{scr})$.

The predictable variation process for the score process $\mathcal{U}_{H(t)}$ in (A.8) (scaled by $n^{-1/2}$) is

$$\begin{aligned}\mathrm{Var}\left(n^{-1/2}\,\mathcal{U}_{H(t)}\middle|\,\mathcal{F}_{u^-}\right) &= \frac{1}{n}\,\mathrm{Var}\left[\sum_x \int_0^{\upsilon} \sum_{i\in\mathcal{R}(t|x)} \varepsilon_i(u,t,H;\boldsymbol{\beta})dM_i(u)\middle|\,\mathcal{F}_{u^-}\right] \\ &= \frac{1}{n}\sum_x \int_0^{\upsilon} \sum_{i\in\mathcal{R}(t|x)} \varepsilon_i^2(u,t,H;\boldsymbol{\beta})\,\mathrm{Var}\left[dM_i(u)|\mathcal{F}_{t^-}\right] \\ &= \frac{1}{n}\sum_x \int_0^{\upsilon} \sum_{i\in\mathcal{R}(t|x)} \varepsilon_i^2(u,t,H;\boldsymbol{\beta})\Psi_i(u)\,\left[\eta_i(u)dH(u)+dH_{scr}(u|x)\right],\end{aligned}$$

which converges weakly as $n\to\infty$ to a mean-zero Gaussian process with covariance function

$$\sigma_H^2(s,t;\boldsymbol{\beta}^0,H^0,H^0_{scr}) = \sum_x \int_0^{\upsilon} \varepsilon(u,s,H;\boldsymbol{\beta})\varepsilon(u,t,H;\boldsymbol{\beta})P(T\geq u)\Psi\,(u)\,[\eta(u)dH(u)+dH_{scr}(u|x)],$$

for $s,t\in[0,\upsilon]$. Similarly, $n^{-1/2}\mathcal{U}_{H_{scr}(t|x)}$, $n^{-1/2}\mathcal{U}_{\boldsymbol{\beta}}$ are martingales converging to the mean-zero Gaussian processes with covariance functions

$$\sigma^2_{H_{scr}}(s,t;\boldsymbol{\beta}^0,H^0,H^0_{scr}) = \int_0^{\upsilon} \varepsilon(u,s,H_{scr};\boldsymbol{\beta})\varepsilon(u,t,H_{scr};\boldsymbol{\beta})P(T\geq u)\Psi\,(u)\,[\eta(u)dH(u)+dH_{scr}(u|x)],$$

$$\sigma^2_{\boldsymbol{\beta}}(\boldsymbol{\beta}^0) = \sum_x \int_0^{\upsilon} \frac{\Psi^2_{\boldsymbol{\beta}}\,(u)}{\Psi\,(u)}P(X\geq u)\,[\eta(u)dH(u)+dH_{scr}(u|x)].$$

In addition,$n^{-1/2}\mathcal{U}_{H_{scr}(t|x),H(s)}$, $n^{-1/2}\mathcal{U}_{H_{scr}(t|x),\boldsymbol{\beta}}$ and $n^{-1/2}\mathcal{U}_{H(t),\boldsymbol{\beta}}$ are martingales that converge to mean-zero Gaussian processes with covariance functions

$$\sigma^2_{H_{scr},H}(s,t;\boldsymbol{\beta}^0,H^0,H^0_{scr}) = \int_0^{\upsilon} \varepsilon(u,s,H;\boldsymbol{\beta})\varepsilon(u,t,H_{scr};\boldsymbol{\beta})P(T\geq u)\Psi\,(u)\,[\eta(u)dH(u)+dH_{scr}(u|x)],$$

$$\sigma^2_{H_{scr},\boldsymbol{\beta}}(t;\boldsymbol{\beta}^0,H^0,H^0_{scr}) = \int_0^{\upsilon} \varepsilon(u,t,H_{scr};\boldsymbol{\beta})\Psi_{\boldsymbol{\beta}}(u)P(T\geq u)\Psi\,(u)\,[\eta(u)dH(u)+dH_{scr}(u|x)],$$

$$\sigma^2_{H,\boldsymbol{\beta}}(t;\boldsymbol{\beta}^0,H^0)=\sum_x\int_0^v \varepsilon(u,t,H;\boldsymbol{\beta})\Psi_{\boldsymbol{\beta}}(u)P(T\geq u)\left[\eta(u)dH(u)+dH_{scr}(u|x)\right].$$

Let the limit in probability of the likelihood function (2.10) normalized as ℓ/n, be ℓ_∞. Define a linear information operator as

$$\mathcal{I}_\infty(t,s|x)=\frac{\partial\mathcal{U}^0}{\partial\Omega}=-\left[\begin{array}{ccc}\frac{\partial^2\ell_\infty}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} & \frac{\partial^2\ell_\infty}{\partial\boldsymbol{\beta}\partial dH(s)} & \frac{\partial^2\ell_\infty}{\partial\boldsymbol{\beta}\partial dH_{scr}(s|x)}\\ \frac{\partial^2\ell_\infty}{\partial dH(u)\partial\boldsymbol{\beta}'} & \frac{\partial^2\ell_\infty}{\partial dH(t)\partial dH(s)} & \frac{\partial^2\ell_\infty}{\partial dH(t)\partial dH_{scr}(s|x)}\\ \frac{\partial^2\ell_\infty}{\partial dH_{scr}(t|x)\partial\boldsymbol{\beta}'} & \frac{\partial^2\ell_\infty}{\partial dH_{scr}(t|x)\partial dH(s)} & \frac{\partial^2\ell_\infty}{\partial dH_{scr}(t|x)\partial dH_{scr}(s|x)}\end{array}\right]_{\Omega=\Omega^0},\tag{A.10}$$

where $\mathcal{U}^0=\left(\frac{\partial\ell_\infty}{\partial\boldsymbol{\beta}'},\frac{\partial\ell_\infty}{\partial dH(t)},\frac{\partial\ell_\infty}{\partial dH_{scr}(t|x)}\right)'$. The operator $\mathcal{I}_\infty$ acts on an arbitrary vector function element $\Omega_t=\left(\boldsymbol{\beta}',dH(t),dH_{scr}(t|x)\right)'$ as

$$\mathcal{I}_\infty(t,s)\Omega_s=$$
$$-\left[\begin{array}{c}\frac{\partial^2\ell_\infty}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\boldsymbol{\beta}+\int_0^v\frac{\partial^2\ell_\infty}{\partial\boldsymbol{\beta}\partial dH(s)}\,dH(t)+\int_0^v\frac{\partial^2\ell_\infty}{\partial\boldsymbol{\beta}\partial dH_{scr(s|x)}}\,dH_{scr}(s|x)\\ \frac{\partial^2\ell_\infty}{\partial dH(t)\partial\boldsymbol{\beta}'}\boldsymbol{\beta}+\int_0^v\frac{\partial^2\ell_\infty}{\partial dH(t)\partial dH(s)}\,dH(s)+\int_0^v\frac{\partial^2\ell_\infty}{\partial dH(t)\partial dH_{scr}(s|x)}\,dH_{scr}(s|x)\\ \frac{\partial^2\ell_\infty}{\partial dH_{scr}(t|x)\partial\boldsymbol{\beta}'}\boldsymbol{\beta}+\int_0^v\frac{\partial^2\ell_\infty}{\partial dH_{scr}(t|x)\partial dH(s)}\,dH(s)+\int_0^v\frac{\partial^2\ell_\infty}{\partial dH_{scr}(t|x)\partial dH_{scr}(s|x)}\,dH_{scr}(s|x)\end{array}\right].\tag{A.11}$$

With this notation, expanding the score $\mathcal{U}(\hat{\Omega})$ about the true parameter Ω^0, we have

$$\mathcal{I}_\infty(t,s)n^{1/2}\left(\hat{\Omega}_s-\Omega^0_s\right)=U(t)+o_p(1).\tag{A.12}$$

Assuming that the Fredholm operator expressed by the kernel $\mathcal{I}_\infty$ of the Fredholm integral equation (A.12) of the first kind is square integrable, and that the equation $\mathcal{I}_\infty\Omega=0$ has only the trivial solution $\Omega=0$, then equation (A.12) has a unique solution. By Theorem 3.3.1 of van der Vaart and Wellner (1996), there exists an inverse information operator $\mathcal{I}^{-1}_\infty(s,t)$ such that

$$n^{1/2}\left(\hat{\Omega}_t-\Omega^0_t\right)=\mathcal{I}^{-1}_0(u,t)U(t)+o_p(1).$$

Upon differentiation of the equation $\mathbb{E}\left[\mathcal{U}(\Omega^0)\right] = 0$ with respect to Ω at the truth Ω^0, we obtain the usual equivalence between $\mathcal{I}_\infty$ represented by second derivatives and

$$\mathcal{I}_\infty(t,s) = \left[\begin{array}{ccc} \frac{\partial \ell_\infty}{\partial \boldsymbol{\beta}}\frac{\partial \ell_\infty}{\partial \boldsymbol{\beta}'} & \frac{\partial \ell_\infty}{\partial \boldsymbol{\beta}}\frac{\partial \ell_\infty}{\partial dH(s)} & \frac{\partial \ell_\infty}{\partial \boldsymbol{\beta}}\frac{\partial \ell_\infty}{\partial dH_{scr}(s|x)} \\ \frac{\partial \ell_\infty}{\partial dH(t)}\frac{\partial \ell_\infty}{\partial \boldsymbol{\beta}'} & \frac{\partial \ell_\infty}{\partial dH(t)}\frac{\partial \ell_\infty}{\partial dH(s)} & \frac{\partial \ell_\infty}{\partial dH(t)}\frac{\partial \ell_\infty}{\partial dH_{scr}(s|x)} \\ \frac{\partial \ell_\infty}{\partial dH_{scr}(t|x)}\frac{\partial \ell_\infty}{\partial \boldsymbol{\beta}'} & \frac{\partial \ell_\infty}{\partial dH_{scr}(t|x)}\frac{\partial \ell_\infty}{\partial dH(s)} & \frac{\partial \ell_\infty}{\partial dH_{scr}(t|x)}\frac{\partial \ell_\infty}{\partial dH_{scr}(s|x)} \end{array}\right]_{\Omega=\Omega^0},$$

which represents the variance of the normalized score Gaussian process $U(t)$. By the functional delta method (Kosorok, 2008, Section 2.2.4), for a differentiable functional $F(\Omega)$, $n^{1/2}\left[F(\hat{\Omega}) - F(\Omega^0)\right]$ converges weakly to a mean-zero Gaussian process with variance-covariance function $\dot{F}(\Omega^0)'\mathcal{I}_0^{-1}\dot{F}(\Omega^0)$, where $\dot{F}(\Omega) = \frac{\partial F}{\partial \Omega}$ and the operator products are defined similarly to (A.11). Applying this to (A.3.3) and replacing operator products by matrix algebra, and $\mathcal{I}_\infty$ by its consistent (matrix) estimator $n^{-1}\hat{\boldsymbol{\mathcal{I}}}_n$, we obtain the weak convergence results. □

A.4 Likelihood Hessian and the information matrix

Using direct algebraic manipulation, we have the following expression for the normalized observed information matrix

$$\mathcal{I}_{n,\Omega\Omega'} = \frac{1}{n}\sum_x \int_0^v \sum_{i\in\mathcal{R}(t|x)} \left\{ \frac{\partial^2 \log d\Lambda_i(t)}{\partial\Omega\partial\Omega'} dM_i(t) - \frac{\partial \log d\Lambda_i(t)}{\partial\Omega}\frac{\partial \log d\Lambda_i(t)}{\partial\Omega'} d\Lambda_i(t) \right\},$$

where $\Lambda_(t)$ is the subject-specific cumulative hazard, $\mathcal{R}(t|x)$ is the cohort x at risk process, $N_i(t)$ is the subject's failure counting process, and $dM_i = dN_i - d\Lambda_i$ is a martingale increment under the true model for $i \in \mathcal{R}(t|x)$. We note that given $t > s$, $\partial\Lambda(s)/\partial dH(t) = 0, \partial\Lambda(s)/\partial dH_{scr}(t|x) = 0$ so terms corresponding to $dH(t)$ and $dH_{scr}(t|x)$ functional component of Ω under the integral are zero until t. Therefore as the martingale term turns into an $o_p(1)$, and the first term into a consistent estimate

of the covariance of the normalized score, we have

$$\mathrm{Cov}\left(\sqrt{n}\mathcal{U}^0\right) = \mathcal{I}_\infty + o_p(1).$$

Now, since

$$\sqrt{n}\left(\hat{\Omega} - \Omega^0\right) = \mathcal{I}_\infty^{-1}\sqrt{n}\mathcal{U}^0 + o_p(1),$$

we have the variance operator

$$\begin{aligned}
\mathrm{Var}\left[\sqrt{n}\left(\hat{\Omega} - \Omega^0\right)\right] &= \mathcal{I}_\infty^{-1}\,\mathrm{Cov}\left(\sqrt{n}\mathcal{U}^0\right)\mathcal{I}_\infty^{-1} \\
&= \mathcal{I}_\infty^{-1}\mathcal{I}_\infty\mathcal{I}_\infty^{-1} \\
&= \mathcal{I}_\infty^{-1}.
\end{aligned}$$

APPENDIX B

Supplementary Materials for Chapter III

B.1 Derivations of Model Quantities

The conditional diagnosis-free survival functions for each mode (CDx and SDx):

$$G_{CDx}(t|\bar{Z}_t, S) = e^{-\int_0^t I(\xi \geq S)\eta(\xi)dH(\xi)},$$
$$G_{SDx}(t|\bar{Z}_t, S, \bar{N}_{scr}(t)) = \exp\left[\int_0^t I(\xi \geq S)\log(1-\alpha(\xi))dN_{scr}(\xi)\right].$$

The marginal diagnosis-free survival function for screening arm is:

$$G_d^{sc}(t|\bar{Z}_t) = \mathbb{E}\left[G_{CDx}(t|\bar{Z}_t, S)G_{SDx}(t|\bar{Z}_t, S, \bar{N}_{scr}(t))\right] = e^{-\int_0^t \theta(\xi)dH(\xi)} + \pi_{sc}(t|\bar{Z}_t)$$

Marginal density:

$$f(t|Z_t) = -\frac{\partial G_d^{sc}(t|\bar{Z}_t)}{\partial t} = f(t, I_{scr}=0|\bar{Z}_t) + f(t, I_{scr}=1|\bar{Z}_t)$$

$$f(t, I_{scr}=0|\bar{Z}_t) = \mathbb{E}_{S,N_{scr}}\left\{f_{CDx}(t|Z_t, S)G_{SDx}(t|\bar{Z}_t, S, \bar{N}_{scr}(t))\right\}$$

$$
\begin{aligned}
&= \mathbb{E}_{N_{scr}} \left\{ \int_0^t \eta(t) dH(t) e^{-\int_s^t [\eta(\xi) dH(\xi) - \log(\bar{\alpha}(\xi)) dN_{scr}(\xi)] - \int_0^s \theta(\xi) dH(\xi)} \theta(s) dH(s) \right\} \\
&= \eta(t) dH(t) \pi_{sc}(t|\bar{Z}_t), \\
f(t, I_{scr} = 1|\bar{Z}_t) &= \mathbb{E}_{S, N_{scr}} \left\{ f_{SDx}(t|Z_t, S, N_{scr}(t)) G_{CDx}(t|\bar{Z}_t, S) \right\} \\
&= \mathbb{E}_{N_{scr}} \left\{ \int_0^t d\Lambda_{SDx}(t|Z_t) e^{-\int_s^t [\eta(Z_\xi) dH(\xi) - \log(\bar{\alpha}(\xi)) dN_{scr}(\xi)] - \int_0^s \theta(\xi) dH(\xi)} \theta(s) dH(s) \right\} \\
&= d\Lambda_{SDx}(t|\bar{Z}_t) \pi_{sc}(t|\bar{Z}_t).
\end{aligned}
$$

B.2 Derivation of Score Equations

The joint log-likelihood function for both control and screening arm is:

$$
\begin{aligned}
\ell &= \ell_c + \ell_{sc} \\
&= \int_0^\upsilon \sum_{j \in \mathcal{R}_c(t)} \log(d\Lambda_i^c(t|Z_{tj})) dN_j^c(t|Z_{tj}) - d\Lambda_j^c(t|Z_{ti}) \\
&+ \int_0^\upsilon \sum_{i \in \mathcal{R}_{sc}(t)} \sum_{I_{scr}=0}^{1} \left\{ \log d\Lambda_i^{I_{scr}}(t|Z_{ti}) dN_i^{I_{scr}}(t|Z_{ti}) - d\Lambda_i^{I_{scr}}(t|Z_{ti}) \right\}.
\end{aligned}
$$

Given the maximum follow-up time υ, differentiating the log-likelihood (3.12) we arrive at the following score equations for $dH(\tau)$, $dH_{scr}(\tau)$, where $\tau \in (0, t]$.

$$
\begin{aligned}
\mathcal{U}_{dH(\tau)} =& \frac{\partial \ell}{\partial dH(\tau)} \\
=& \sum_{j \in \mathcal{R}_c(\tau)} \frac{\partial \log d\Lambda^c(\tau|\bar{Z}_{\tau j})}{\partial dH(\tau)} dN^c(\tau|\bar{Z}_{\tau j}) - \frac{\partial d\Lambda^c(\tau|\bar{Z}_{\tau j})}{\partial dH(\tau)} \\
&+ \int_{\tau+}^\upsilon \sum_{k \in \mathcal{R}_c(t)} \frac{\partial \log d\Lambda^c(t|\bar{Z}_{tk})}{\partial dH(\tau)} dN^c(t|\bar{Z}_{tk}) - \frac{\partial d\Lambda^c(t|\bar{Z}_{tk})}{\partial dH(\tau)} \\
&+ \sum_{i \in \mathcal{R}_{sc}(\tau)} \frac{\partial \log d\Lambda^{I_{scr}}(\tau|\bar{Z}_{\tau i})}{\partial dH(\tau)} dN^{I_{scr}}(\tau|\bar{Z}_{\tau i}) - \frac{\partial d\Lambda^{I_{scr}}(\tau|\bar{Z}_{\tau i})}{\partial dH(\tau)} \\
&+ \int_{\tau+}^\upsilon \sum_{l \in \mathcal{R}_{sc}(t)} \frac{\partial \log d\Lambda^{I_{scr}}(t|\bar{Z}_{tl})}{\partial dH(\tau)} dN^{I_{scr}}(t|\bar{Z}_{tl}) - \frac{\partial d\Lambda^{I_{scr}}(t|\bar{Z}_{tl})}{\partial dH(\tau)}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{j\in\mathcal{R}_c(\tau)} \left(\frac{\Psi^c_{dH(\tau)}(\tau|\bar{Z}_{\tau j})}{\Psi^c(\tau|\bar{Z}_{\tau j})} + \frac{\eta_j(\tau)}{\eta_j(\tau)dH(\tau) + r_1^{I(\tau\geq A_{ej})}\alpha dH_{scr}(\tau)} \right) dN^c(\tau|\bar{Z}_{\tau j}) \\
&\quad - \frac{\Psi^c_{dH(\tau)}(\tau|\bar{Z}_{\tau j})}{\Psi^c(\tau|\bar{Z}_{\tau j})} d\Lambda^c(\tau|\bar{Z}_{\tau j}) - \Psi^c(\tau|\bar{Z}_{\tau j})\eta_j(\tau) \\
&\quad + \int_{\tau+}^{\upsilon} \sum_{k\in\mathcal{R}_c(t)} \frac{\Psi^c_{dH(\tau)}(t|\bar{Z}_{tk})}{\Psi^c(t|\bar{Z}_{tk})} dN^c(t|\bar{Z}_{tk}) - \frac{\Psi^c_{dH(\tau)}(t|\bar{Z}_{tk})}{\Psi^c(t|\bar{Z}_{tk})} d\Lambda^c(t|\bar{Z}_{tk}) \\
&\quad + \sum_{i\in\mathcal{R}_{sc}(\tau)} \sum_{I_{scr}=0} \left(\frac{\Psi^{sc}_{dH(\tau)}(\tau|\bar{Z}_{\tau i})}{\Psi^{sc}(\tau|\bar{Z}_{\tau i})} + \frac{1}{dH(\tau)} \right) dN^{I_{scr}}(\tau|\bar{Z}_{\tau i}) \\
&\quad - \frac{\Psi^{sc}_{dH(\tau)}(\tau|\bar{Z}_{\tau i})}{\Psi^{sc}(\tau|\bar{Z}_{\tau i})} d\Lambda^{I_{scr}}(\tau|\bar{Z}_{\tau i}) - \Psi^{sc}(\tau|\bar{Z}_{\tau i})\eta_i(\tau) \\
&\quad + \int_{\tau+}^{\upsilon} \sum_{l\in\mathcal{R}_{sc}(t)} \sum_{I_{scr}=0} \frac{\Psi^{sc}_{dH(\tau)}(t|\bar{Z}_{tl})}{\Psi^{sc}(t|\bar{Z}_{tl})} dN^{I_{scr}}(t|\bar{Z}_{tl}) - \frac{\Psi^{sc}_{dH(\tau)}(t|\bar{Z}_{tl})}{\Psi^{sc}(t|\bar{Z}_{tl})} d\Lambda^{I_{scr}}(t|\bar{Z}_{tl}) \\
&= \sum_{j\in\mathcal{R}^c(\tau)} \frac{\eta_j(\tau)dN^c(\tau|\bar{Z}_{\tau j})}{\eta_j(\tau)dH(\tau) + r_1^{I(\tau\geq A_{ej})}\alpha dH_{scr}(\tau)} \\
&\quad - \sum_{j\in\mathcal{R}_c(\tau)} \Psi^c(\tau|\bar{Z}_{\tau j})\eta_j(\tau) \left(1 - \frac{\int_\tau^\upsilon \sum_{k\in\mathcal{R}_c(t)} \frac{\Psi^c_{dH(\tau)}(t|\bar{Z}_{tk})}{\Psi^c(t|\bar{Z}_{tk})} dM^c(t|\bar{Z}_{tk})}{\sum_{j\in\mathcal{R}_c(\tau)} \Psi^c(\tau|\bar{Z}_{\tau j})\eta_j(\tau)} \right) \\
&\quad + \sum_{i\in\mathcal{R}_{sc}(\tau)} \sum_{I_{scr}=0} \frac{dN^{I_{scr}}(\tau|\bar{Z}_{\tau i})}{dH(\tau)} \\
&\quad - \sum_{i\in\mathcal{R}_{sc}(\tau)} \sum_{I_{scr}=0} \Psi^{sc}(\tau|\bar{Z}_{\tau i})\eta_i(\tau) \left(1 - \frac{\int_\tau^\upsilon \sum_{l\in\mathcal{R}_{sc}(t)} \frac{\Psi^{sc}_{dH(\tau)}(t|\bar{Z}_{tl})}{\Psi^{sc}(t|\bar{Z}_{tl})} dM^{sc}(t|\bar{Z}_{tl})}{\sum_{i\in\mathcal{R}_{sc}(\tau)} \sum_{I_{scr}=0} \Psi^{sc}(\tau|\bar{Z}_{\tau i})\eta_i(\tau)} \right) \\
&= \sum_{j\in\mathcal{R}_c(\tau)} \frac{\eta_j(\tau)dN^c(\tau|\bar{Z}_{\tau j})}{\eta_j(\tau)dH(\tau) + r_1^{I(\tau\geq A_{ej})}\alpha dH_{scr}(\tau)} - \sum_{j\in\mathcal{R}_c(\tau)} \Psi^c(\tau|\bar{Z}_{\tau j})\eta_j(\tau)w^c_{dH(\tau)} \\
&\quad + \sum_{i\in\mathcal{R}_{sc}(\tau)} \sum_{I_{scr}=0} \frac{dN^{I_{scr}}(\tau|Z_{\tau i})}{dH(\tau)} - \sum_{i\in\mathcal{R}_{sc}(\tau)} \sum_{I_{scr}=0} \Psi^{sc}(\tau|\bar{Z}_{\tau i})\eta_i(\tau)w^{sc}_{dH(\tau)}.
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{U}_{dH_{scr}(\tau)} =& \frac{\partial \ell}{\partial dH_{scr}(\tau)} \\
&= \sum_{j\in\mathcal{R}_c(\tau)} \frac{\partial \log d\Lambda^c(\tau|\bar{Z}_{\tau j})}{\partial dH_{scr}(\tau)} dN^c(\tau|\bar{Z}_{\tau j}) - \frac{\partial d\Lambda^c(\tau|\bar{Z}_{\tau j})}{\partial dH_{scr}(\tau)} \\
&\quad + \int_{\tau+}^{\upsilon} \sum_{k\in\mathcal{R}_c(t)} \frac{\partial \log d\Lambda^c(t|\bar{Z}_{tk})}{\partial dH_{scr}(\tau)} dN^c(t|\bar{Z}_{tk}) - \frac{\partial d\Lambda^c(t|\bar{Z}_{tk})}{\partial dH_{scr}(\tau)}
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{i\in\mathcal{R}_{sc}(\tau)}\frac{\partial\log d\Lambda^{I_{scr}}(\tau|\bar{Z}_{\tau i})}{\partial dH_{scr}(\tau)}dN^{I_{scr}}(\tau|\bar{Z}_{\tau i})-\frac{\partial d\Lambda^{I_{scr}}(\tau|\bar{Z}_{\tau i})}{\partial dH_{scr}(\tau)}\\
&+\int_{\tau+}^{\upsilon}\sum_{l\in\mathcal{R}_{sc}(t)}\frac{\partial\log d\Lambda^{I_{scr}}(t|\bar{Z}_{tl})}{\partial dH_{scr}(\tau)}dN^{I_{scr}}(t|\bar{Z}_{tl})-\frac{\partial d\Lambda^{I_{scr}}(t|\bar{Z}_{tl})}{\partial dH_{scr}(\tau)}\\
=&\sum_{j\in\mathcal{R}_c(\tau)}\left(\frac{\Psi^c_{dH_{scr}(\tau)}(\tau|\bar{Z}_{\tau j})}{\Psi^c(\tau|\bar{Z}_{\tau j})}+\frac{r_1^{I(\tau\geq A_{ej})}}{\eta_j(\tau)dH(\tau)+r_1^{I(\tau\geq A_{ej})}\alpha dH_{scr}(\tau)}\right)dN^c(\tau|\bar{Z}_{\tau j})\\
&-\frac{\Psi^c_{dH(\tau)}(\tau|\bar{Z}_{\tau j})}{\Psi^c(\tau|\bar{Z}_{\tau j})}d\Lambda^c(\tau|\bar{Z}_{\tau j})-\Psi^c(\tau|\bar{Z}_{\tau j})r_1^{I(\tau\geq A_{ej})}\\
&+\int_{\tau+}^{\upsilon}\sum_{k\in\mathcal{R}_c(t)}\frac{\Psi^c_{dH_{scr}(\tau)}(t|\bar{Z}_{tk})}{\Psi^c(t|\bar{Z}_{tk})}dN^c(t|\bar{Z}_{tk})-\frac{\Psi^c_{dH_{scr}(\tau)}(t|\bar{Z}_{tk})}{\Psi^c(t|\bar{Z}_{tk})}d\Lambda^c(t|\bar{Z}_{tk})\\
&+\sum_{i\in\mathcal{R}_{sc}(\tau)}\sum_{I_{scr}=1}\left(\frac{\Psi^{sc}_{dH(\tau)}(\tau|\bar{Z}_{\tau i})}{\Psi^{sc}(\tau|\bar{Z}_{\tau i})}+\frac{I(\tau<A_{ei},\tau>A_{xi})}{dH_{scr}(\tau)}\right)dN^{I_{scr}}(\tau|\bar{Z}_{\tau i})\\
&-\frac{\Psi^{sc}_{dH_{scr}(\tau)}(\tau|\bar{Z}_{\tau i})}{\Psi^{sc}(\tau|\bar{Z}_{\tau i})}d\Lambda^{I_{scr}}(\tau|\bar{Z}_{\tau i})-\Psi^{sc}(\tau|\bar{Z}_{\tau i})\left[I(\tau<A_{ei})\alpha+I(\tau>A_{xi})r_2\alpha\right]\\
&+\int_{\tau+}^{\upsilon}\sum_{l\in\mathcal{R}_{sc}(t)}\sum_{I_{scr}=1}\frac{\Psi^{sc}_{dH_{scr}(\tau)}(t|\bar{Z}_{tl})}{\Psi^{sc}(t|\bar{Z}_{tl})}dN^{I_{scr}}(t|\bar{Z}_{tl})-\frac{\Psi^{sc}_{dH_{scr}(\tau)}(t|\bar{Z}_{tl})}{\Psi^{sc}(t|\bar{Z}_{tl})}d\Lambda^{I_{scr}}(t|\bar{Z}_{tl})\\
=&\sum_{j\in\mathcal{R}^c(\tau)}\frac{r_1^{I(\tau\geq A_{ej})}\alpha dN^c(\tau|\bar{Z}_{\tau j})}{\eta_j(\tau)dH(\tau)+r_1^{I(\tau\geq A_{ej})}\alpha dH_{scr}(\tau)}\\
&-\sum_{j\in\mathcal{R}_c(\tau)}\Psi^c(\tau|\bar{Z}_{\tau j})r_1^{I(\tau\geq A_{ej})}\alpha\left(1-\frac{\int_\tau^\upsilon\sum_{k\in\mathcal{R}_c(t)}\frac{\Psi^c_{dH_{scr}(\tau)}(t|\bar{Z}_{tk})}{\Psi^c(t|\bar{Z}_{tk})}dM^c(t|\bar{Z}_{tk})}{\sum_{j\in\mathcal{R}_c(\tau)}\Psi^c(\tau|\bar{Z}_{\tau j})r_1^{I(\tau\geq A_{ej})}\alpha}\right)\\
&+\sum_{i\in\mathcal{R}_{sc}(\tau)}\sum_{I_{scr}=1}\frac{I(\tau<A_{ei},\tau>A_{xi})dN^{I_{scr}}(\tau|\bar{Z}_{\tau i})}{dH_{scr}(\tau)}\\
&-\sum_{i\in\mathcal{R}_{sc}(\tau)}\sum_{I_{scr}=1}\Psi^{sc}(\tau|\bar{Z}_{\tau i})\left[I(\tau<A_{ei})\alpha+I(\tau>A_{xi})r_2\alpha\right]\\
&*\left(1-\frac{\int_\tau^\upsilon\sum_{l\in\mathcal{R}_{sc}(t)}\frac{\Psi^{sc}_{dH_{scr}(\tau)}(t|\bar{Z}_{tl})}{\Psi^{sc}(t|\bar{Z}_{tl})}dM^{sc}(t|\bar{Z}_{tl})}{\sum_{i\in\mathcal{R}_{sc}(\tau)}\sum_{I_{scr}=1}\Psi^{sc}(\tau|\bar{Z}_{\tau i})\left[I(\tau<A_{ei})\alpha+I(\tau>A_{xi})r_2\alpha\right]}\right)\\
=&\sum_{j\in\mathcal{R}_c(\tau)}\frac{r_1^{I(\tau\geq A_{ej})}\alpha dN^c(\tau|\bar{Z}_{\tau j})}{\eta_j(\tau)dH(\tau)+r_1^{I(\tau\geq A_{ej})}\alpha dH_{scr}(\tau)}-\sum_{j\in\mathcal{R}_c(\tau)}\Psi^c(\tau|\bar{Z}_{\tau j})r_1^{I(\tau\geq A_{ej})}\alpha w^c_{dH_{scr}(\tau)}\\
&+\sum_{i\in\mathcal{R}_{sc}(\tau)}\sum_{I_{scr}=1}\frac{I(\tau<A_{ei},\tau>A_{xi})dN^{I_{scr}}(\tau|Z_{\tau i})}{dH_{scr}(\tau)}
\end{aligned}
$$

$$- \sum_{i \in \mathcal{R}_{sc}(\tau)} \sum_{I_{scr}=1} \Psi^{sc}(\tau|\bar{Z}_{\tau i}) \left[I(\tau < A_{ei})\alpha + I(\tau > A_{xi}) r_2 \alpha \right] w^{sc}_{dH_{scr}(\tau)}.$$

The score equations $\mathcal{U}_{dH(\tau)}$ and $\mathcal{U}_{dH_{scr}(\tau)}$ are martingales at the true model:

$$\begin{aligned}
\mathcal{U}_{dH(\tau)} =& \frac{\partial \ell}{\partial dH(\tau)} \\
=& \sum_{j \in \mathcal{R}_c(\tau)} \frac{\eta_j(\tau) dM^c(\tau|\bar{Z}_{\tau j})}{\eta_j(\tau) dH(\tau) + r_1^{I(\tau \geq A_{ej})} \alpha dH_{scr}(\tau)} + \int_\tau^v \sum_{j \in \mathcal{R}_c(t)} \frac{\Psi^c_{dH(\tau)}(t|\bar{Z}_{tj})}{\Psi^c(t|Z_{tj})} dM^c(t|\bar{Z}_{tj}) \\
&+ \sum_{i \in \mathcal{R}_{sc}(\tau)} \sum_{I_{scr}=0} \frac{dM^0_i(\tau|Z_{\tau i})}{dH(\tau)} + \int_\tau^v \sum_{i \in \mathcal{R}_{sc}(t)} \sum_{I_{scr}=0}^{1} \frac{\Psi^{sc}_{dH(\tau)}(t|\bar{Z}_{ti})}{\Psi^{sc}(t|Z_{ti})} dM^{I_{scr}}(t|\bar{Z}_{ti})
\end{aligned}$$

$$\begin{aligned}
\mathcal{U}_{dH_{scr}(\tau)} =& \frac{\partial \ell}{\partial dH_{scr}(\tau)} \\
=& \sum_{j \in \mathcal{R}_c(\tau)} \frac{r_1^{I(\tau \geq A_{ej})} \alpha dM^c(\tau|\bar{Z}_{\tau j})}{\eta_j(\tau) dH(\tau) + r_1^{I(\tau \geq A_{ej})} \alpha dH_{scr}(\tau)} + \int_\tau^v \sum_{j \in \mathcal{R}_c(t)} \frac{\Psi^c_{dH_{scr}(\tau)}(t|\bar{Z}_{tj})}{\Psi^c(t|Z_{tj})} dM^c(t|\bar{Z}_{tj}) \\
&+ \sum_{i \in \mathcal{R}_{sc}(\tau)} \sum_{I_{scr}=1} \frac{I(\tau < A_{ei}, \tau > A_{xi}) dM^1_i(\tau|Z_{\tau i})}{dH(\tau)} \\
&+ \int_\tau^v \sum_{i \in \mathcal{R}_{sc}(t)} \sum_{I_{scr}=1}^{1} \frac{\Psi^{sc}_{dH_{scr}(\tau)}(t|\bar{Z}_{ti})}{\Psi^{sc}(t|Z_{ti})} dM^{I_{scr}}(t|\bar{Z}_{ti})
\end{aligned}$$

B.3 Data analysis results

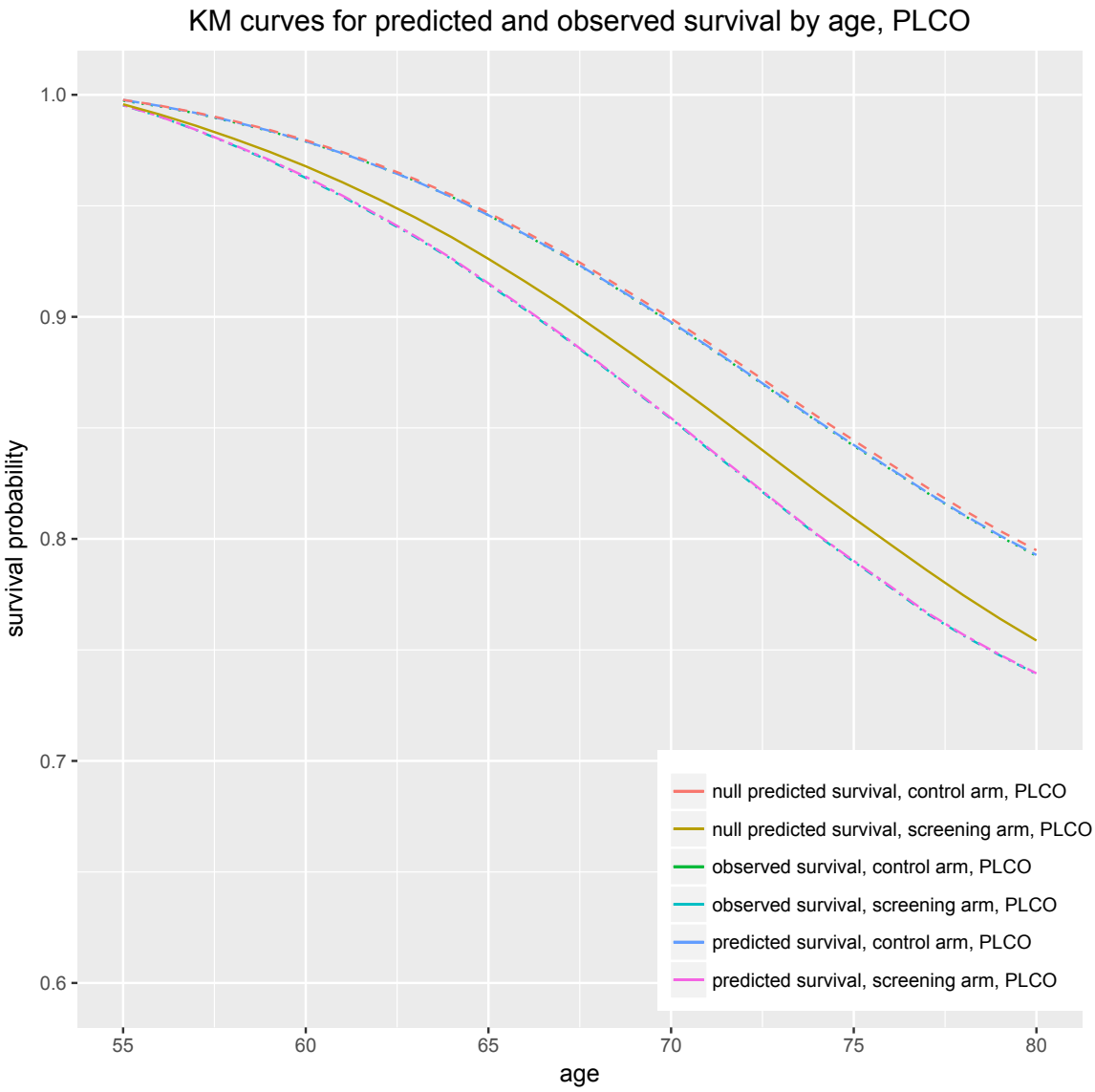

Figure B.1: Kaplan-Meier curves for predicted and observed survival from diagnosis by age for control arm and screening arm, PLCO.

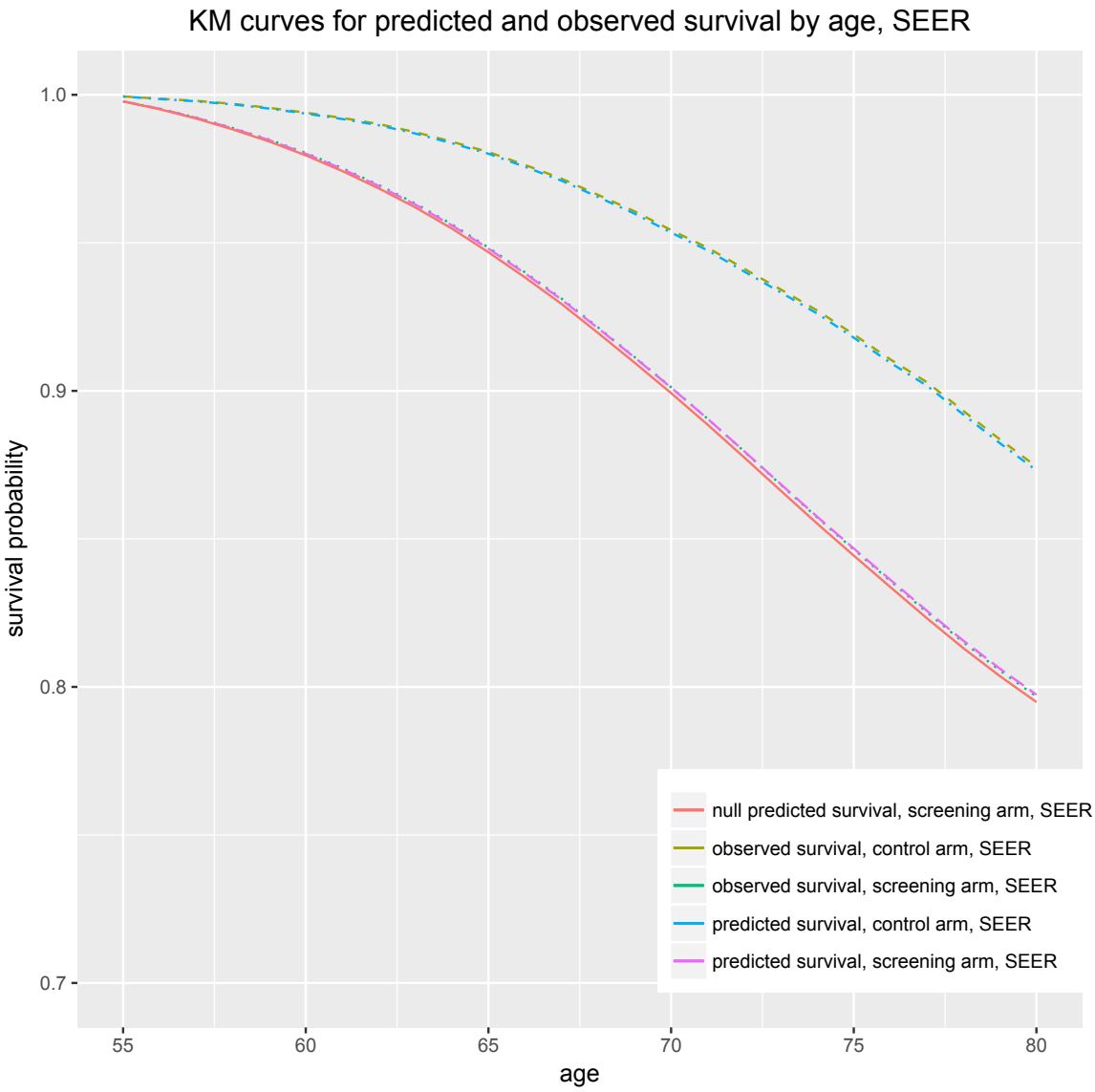

Figure B.2: Kaplan-Meier curves for predicted and observed survival from diagnosis by age for control arm and screening arm, SEER.

APPENDIX C

Supplementary Materials for Chapter IV

C.1 The mean lead time and empirical estimation

Lead time is usually defined as the amount of time by which clinical diagnosis (i.e., diagnosis without screening) is advanced by screening. This version of lead time is only defined in patients who are detected by screening and who, in the absence of screening, would be clinically diagnosed. This definition excludes overdiagnosed patients (i.e., patients detected by screening who would not be clinically diagnosed in the absence of screening), patients clinically diagnosed, and patients without any diagnosis. Our goal here, however, is to derive a generic surrogate of the intensity of screening and diagnosis in a given population that is amenable to empirical (model-free, robust) estimation and prediction. Nevertheless, due to its close theoretical relationship with the mean lead time (described below), we adopt this terminology.

In a screening trial, the time from randomization to screen detection (T_{SDx}) and the time from randomization to clinical diagnosis (T_{CDx}) are competing risks that are never both observed for the same patient. Consequently, their full joint distribution is non-identifiable without specific modeling assumptions (Tsiatis, 1975). However,

the mean lead time (i.e., the mean time difference $T_{\mathrm{CDx}} - T_{\mathrm{SDx}}$, which is non-zero only if $T_{\mathrm{SDx}} < T_{\mathrm{CDx}}$) is identifiable and can be calculated empirically. Formally, the mean lead time is equal to the difference between the mean time to diagnosis without screening minus the mean time to diagnosis with screening:

$$\mathbb{E}\left(\min(0, T_{\mathrm{CDx}} - T_{\mathrm{SDx}})\right) = \mathbb{E}(T_{\mathrm{CDx}}) - \mathbb{E}\left(\min(T_{\mathrm{CDx}}, T_{\mathrm{SDx}})\right)$$

The two terms on right-hand side can be estimated empirically as the areas under the survival curves for diagnosis in the absence and in the presence of screening. Further, this empirical approach for estimating the mean lead time extends to a trial with finite follow-up ($T_{\max}$). Specifically, the time of screen detection becomes $T'_{\mathrm{SDx}} = \min(T_{\max}, T_{\mathrm{SDx}})$, the time of clinical diagnosis becomes $T'_{\mathrm{CDx}} = \min(T_{\max}, T_{\mathrm{CDx}})$, and the mean times from randomization to diagnosis in the absence and in the presence of screening can be estimated as the areas under the survival curves for diagnosis up to the landmark $T_{\max}$ (Appendix Figure C.1). Restricted survival times have been used in medical statistics and this time-restricted version of the mean lead time corresponds exactly to the so-called restricted mean survival time (Uno et al., 2014) when the survival event of interest is cancer diagnosis.

Note that the mean lead time estimated in this way is not interpretable as the mean lead time for non-overdiagnosed screen-detected cases who would have been clinically diagnosed in the absence of screening (within the follow-up of the trial). Rather, this version of the mean lead time is a population-level measure for all participants in the trial arm, and greater magnitudes can indicate both greater numbers of non-overdiagnosed cases detected early as well as greater numbers of overdiagnosed cases. Because this version of the mean lead time is positively related to the number of non-overdiagnosed cases, it is an acceptable surrogate for the intensity of screening and diagnostic workup for our analyses. After all, if greater magnitudes were only

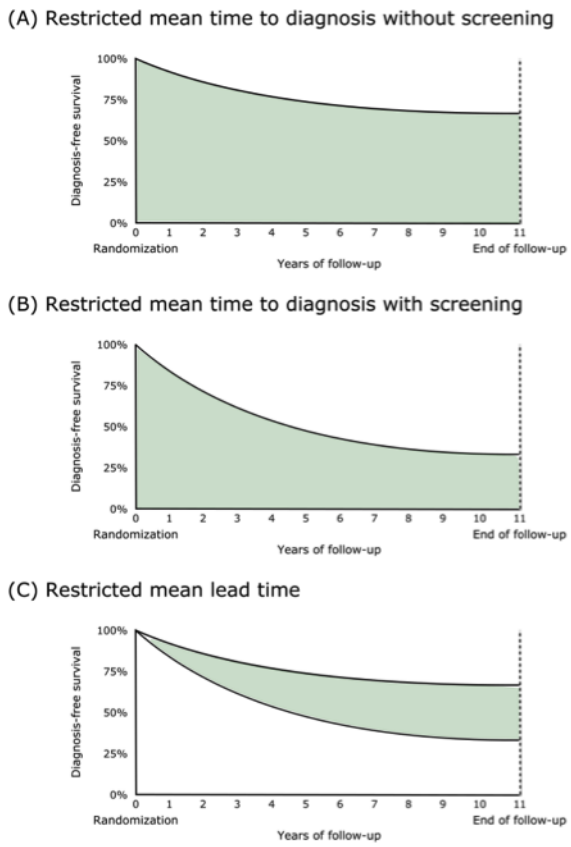

Figure C.1: (A) Mean time from randomization to diagnosis without screening, (B) mean time from randomization to diagnosis with screening, and (C) mean lead time, each restricted to 11 years of follow-up, correspond to areas under the associated diagnosis-free survival curves or to the difference between these areas.

attributable to greater numbers of overdiagnosed cases, there would be no observed prostate cancer mortality reduction that could be associated with this measure.

Finally, to standardize the measure across trial arms in this study, we scaled each estimated MLT by a common baseline probability of (screen or clinical) diagnosis during follow-up, $1/Pr(\text{diagnosis in } [0, T_{\max}])$, making the MLT a conditional average given any mode of prostate cancer diagnosis during the trial.

C.2 Descriptions of three prostate cancer natural history models, adaptations to trial settings, and estimation methods

Three models of prostate cancer development, progression, and detection were used to estimate MLTs associated with prostate-specific antigen (PSA) screening in each arm of the European Randomized Study of Screening for Prostate Cancer (ERSPC) and the Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial. Two of the models are microsimulation models that generate individual life histories comprised of cancer natural history and diagnosis; the third is an analytic mathematical model that represents corresponding events in an integrated set of analytic probability models. The models previously estimated age-specific risks of onset of latent cancer, transitions through histologic stages and grades, and clinical diagnosis in the absence of screening (Appendix Figure C.2) in the general US population (https://resources.cisnet.cancer.gov/registry/site-summary/prostate). For this study, the models were adapted using data on screening and biopsy behavior before and after the start of each trial and prostate cancer incidence data during the trials, including age, stage, and grade at diagnosis and mode of detection. Details for each model are given below. The adapted models closely reproduced incidence patterns observed in intervention and control arms of both trials. These incidence patterns obtained from the models were then used in place of observed incidence data to estimate the MLT in each trial arm using the empirical approach described above.

In the intervention arms, the models implemented center-specific screening and biopsy-referral protocols subject to observed adherence, which implied an average of 2.1-3.3 (ERSPC) or 6.9-7.5 (PLCO) tests per person during 11 years of follow-up (ranges across models). In the PLCO control arm, the models assumed screening test frequencies by age and birth year as previously reconstructed for the US population

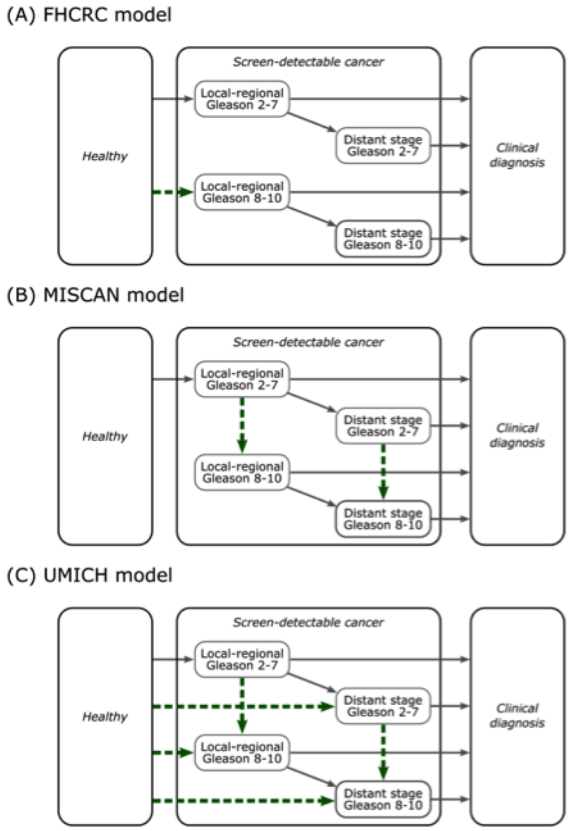

Figure C.2: Health states and modeled transitions between states in the FHCRC, MISCAN, and UMICH models.

with increased tests during the intervention period as previously noted (Pinsky et al., 2010; Gulati et al., 2012), which implied an average of 3.8-5.1 tests per person. In the ERSPC control arm, the models assumed screening occurred at 5% of the frequency of PSA screening in the US population over the same calendar period as the trial, which implied an average of 0.2-0.4 tests per person.

C.2.1 The Fred Hutchinson Cancer Research Center (FHCRC) model

The FHCRC model consists of two connected pieces: PSA growth and disease progression. PSA growth is linear on the log scale, with a larger slope after disease onset. PSA growth rates were estimated using serial screening results from the placebo arm of the Prostate Cancer Prevention Trial (Thompson et al., 2003). Disease progression encompasses tumor onset, metastatic spread, and clinical diagnosis

that would occur in the absence of screening. The risk of onset increases with age and risks of progression to subsequent events increase with PSA levels. All cancers are T-stage $\leq$ T2a at onset and can progress to T-stage $>$ T2a and to T4 or M1. Cancer grade (Gleason score 2-6, 7, or 8-10) is fixed at onset. To estimate progression risk parameters for the US population, we superimposed PSA screening and biopsy frequencies using published patterns and determined the parameter values (using simulated maximum likelihood) so that model incidence closely matched incidence in the Surveillance, Epidemiology, and End Results (SEER) program by age, year, stage, and grade (Gulati et al., 2010).

To adapt the model to the PLCO, we assumed eligible men underwent PSA screening based on our reconstruction for general US population, ensured all men were undiagnosed at entry into the trial, and specified enrollment, attendance, and follow-up patterns to match observed data for all men ages 55-74 years at enrollment. Men randomized to the intervention arm underwent 6 years of annual DRE and PSA screening, while men randomized to the control arm underwent more intensive screening than in the general population (Pinsky et al., 2010). Men in either arm with suspicious DRE or PSA>4.0 ng/mL received biopsies based on observed frequencies in the intervention arm by age and PSA level, and we assumed the sensitivity of a biopsy to detect latent cancer improved from 80% in 1993 to 93% in 2000 and later (Gulati et al., 2010). After the trials 6-year intervention period, men were randomized to either arm resumed screening according to population screening patterns. Model assumptions of non-compliance and contamination were previously described (Gulati et al., 2012). While the model for the general US population projected incidence that matched observed incidence reasonably well without adjustment, we re-estimated risks of onset and of progression to produce modestly improved fit. The modest changes are consistent with a cohort enriched for healthy behavior and higher socioeconomic status (Pinsky et al., 2010).

Adaptation to the ERSPC core age group (ages 55-69 years at enrollment) was similar except that enrollment, attendance, follow-up, DRE and PSA screening, PSA threshold for biopsy, and receipt of biopsy were modeled for each center individually to reflect heterogeneity in recruitment and protocols. We assumed the control arm underwent less intensive screening than the US population (receiving 5% of screens), consistent with contamination reported for several centers (Ciatto et al., 2003; Bokhorst et al., 2014). As for the PLCO, while the model projected incidence that matched observed incidence reasonably well without adjustment, we re-estimated risks of onset and of progression to produce modestly improved fit. In particular, allowing for differential risks of clinical presentation for each stage and grade, the model achieved a noticeably improved fit for metastatic incidence.

C.2.2 The Erasmus University Medical Center MIcrosimulation SCreening ANalysis (MISCAN) model

The MISCAN model is also a microsimulation model of individual life histories. The risk of onset increases with age, and risks of cancer progression are modeled as a semi-Markov process over a sequence of tumor states. There are 18 preclinical detectable states determined by combinations of histologic grade (SEER categories well, moderately, and poorly differentiated), clinical T-stages (American Joint Committee on Cancer stages T1, T2, and T3), and clinical M-stages (M0 and M1). Risks of stage and grade progression are both modeled and allowed to vary across combinations and from a preclinical to clinical state. The chance of a screen protocol detecting a preclinical tumor depends on screening frequency, attendance rates, the PSA threshold for a positive test, and, after a positive PSA test, biopsy compliance and sensitivity.

Baseline model parameters were originally estimated using data from the Rotterdam section of the ERSPC (Draisma et al., 2003). For adaptation to the US population, we re-estimated the PSA test sensitivity parameters and a modified stage-specific

risk of clinical diagnosis to capture different pre-PSA disease diagnosis patterns. US-specific estimates for the parameters were obtained by calibrating the model to the observed age-specific incidence and age-specific SEER stage distribution using maximum likelihood (Wenver et al., 2010).

Adaptations of the US model to the PLCO and of the Rotterdam model to the ERSPC were similar to those for the FHCRC model. Specifically, after accounting for differences in settings and protocols, we re-estimated model parameters in each trial arm to match observed incidence by age and year, and the associated stage distribution, by minimizing a sum of the chi-square errors. The minimization was achieved using an adapted version of the Nelder-Mead simplex algorithm. The algorithm was initially run using small sample sizes (i.e., 20,000 men) then repeated with larger sample sizes (i.e., 2,000,000 men) when the initial optimization step could not be further improved. For this analysis, we first re-estimated the disease progression rates and the PSA test sensitivity relative to stage- and grade-specific incidence of clinical cancers in the control arms and then relative to screen-detected and interval cancers in the intervention arms.

C.2.3 The University of Michigan (UMICH) model

The model is similar with the models in Chapter 2 and Chapter 3, which is an analytic model comprised of three components. A marginal incidence model (Tsodikov et al., 2006) is a two-stage model for a given individual screening schedule (a point process A). The first stage is defined by the hazard of the first PSA test for a man at a given age and calendar time. A second hazard is defined for men who already had their first PSA test. Both hazards of PSA testing rely on a retrospective analysis of PSA testing. Cancer diagnosis is defined as a result of two competing risks, clinical diagnosis (CDx) and diagnosis due to screening (SDx), whichever comes first. The risks are dependent based on a common natural history of the disease, with both risks

equal to zero until the onset of a detectable tumor. Estimation is based on parametric maximum likelihood, which averages over the unobserved screening schedule and natural history processes. Once the stochastic process mixed model is fit, predictions for lead time, overdiagnosis, age of tumor onset, and other characteristics in the patient and the population are predicted using Bayesian conditional probabilities.

Cancer stage and grade are represented as a categorical mark (Z) on the incident cancer. We use a mixed multinomial model to specify the distribution of stage and grade at diagnosis, where the mixing variables represent key unobserved features of latent disease progression (e.g., age at onset), predicted as conditional distributions, given results of the marginal incidence model. Stage and grade are modeled using a mixed multinomial model. The model is estimated by maximum likelihood. We developed a special method of artificial mixtures and the quasi-EM algorithm to deal with the curse of dimensionality in complex models (Tsodikov et al., 2014). Applications of the multinomial model and the stage- and grade-specific incidence model have been previously described (Tsodikov and Chefo, 2008; Chefo and Tsodikov, 2009; Wang and Tsodikov, 2010). Given age, year, stage, and grade at diagnosis, the model projects unobserved characteristics of disease natural history and clinical diagnosis in the absence of screening.

Finally, disease progression is modeled as a stochastic process $Z(\xi)$ as a function of time ξ measured from the point of onset. Given the two potential competing risks of clinical and screening diagnosis, we can define the corresponding potential values of the cancer development process $Z(\xi_{SDx})$ and $Z(\xi_{CDx})$ for the same individual. Let I_{SDx} be 1 for screening and 0 for clinical diagnosis and define the vector $V = (a, z)$ combine age, stage, and grade at diagnosis. The disease progression model defines the probability of disease progression between these two potential points of diagnosis by the transition model $[V_0|V_1]$. For a screen-detected patient, let $f_V(V_0|V_1, x)$ be the joint pdf of disease presentation at CDx (with characteristics V_0) given observed

characteristics V_1 at SDx and birth cohort x. The transition probabilities between the two points of diagnosis (SDx, CDx) are modeled as functions of the lead time $\xi_L, p_b(z_0|z_1, \xi_L)$, summarized as a progression probability matrix (PPM). Under the null hypothesis of no screening benefit, the baseline PPM probabilities p_b are not affected by intervention applied at the point of SDx. Thus, we consider two model predictions for cancer incidence: (1) $\lambda_I(a, z|\bar{S})$ under no screening and (2)$\lambda_I(a, z|IS)$ under ignored screening (no effect of early detection). The first scenario does not involve the PPM, while the second scenario uses PPM. Making the two counterfactual incidence predictions as close as possible using a Poisson-type distance measure provides a robust estimating procedure.

To estimate mean lead times in this study, only the marginal incidence model needed to be estimated. Because screening intensities for the trial populations are only partially known, the competing risk of screen diagnosis during the follow-up period was estimated nonparametrically and jointly with the parametric natural history model as a matrix by age and follow-up time using maximum likelihood. Thus, in contrast to the MISCAN and FHCRC microsimulation models, which combined empirical data with trial protocols, the UMICH model directly incorporated empirical patterns of enrollment, attendance, and follow-up; PSA screening and receipt of biopsy; and age, year, stage, grade, and mode of detection for incident cancers into the aggregate data-driven nonparametric risk of screening diagnosis and parametric risk of clinical diagnosis. The resulting version of the marginal incidence model was used to generate the expected incidence in each trial arm. Lead times in each trial arm were predicted using the method similar to the empirical approach, except that the model-based incidence predictions in each trial arm were used instead of the empirical time to diagnosis survival curves. Comparable assumptions were made concerning prior screening for both trials.

BIBLIOGRAPHY

BIBLIOGRAPHY

Andriole, G. (2009). Mortality results from a randomized prostat-cancer screening trial. *New England Journal of Medicine* **360,** 1310–9.

Andriole, G. (2012). Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. *Journal of the National Cancer Institute* **104,** 125–32.

Berg, C. (2011). The Prostate, Lung, Colorectal, and Ovarian cancer screening trial: The prostate cancer screening trial context. *Acta Oncologica* **50 Suppl 1,** 12–7.

Bokhorst, L., Bangma, C., van Leenders, G., Lous, J., Moss, S., Schroder, F., and et al (2014). Prostate-specfic antigen-based prostate cancer screening: reduction of prostate cancer mortality after correction for nonattendance and contamination in the Rotterdam section of the European Randomized Study of Screening for Prostate Cancer. *Eur Urol* **65,** 329–36.

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

Chen, K., Jin, Z., and Ying, Z. (2002). Semiparametric analysis of transformation models with censored data. *Biometrika* **89,** 659–668.

Chen, Y. H. (2009). Weighted Breslow-type and maximum likelihood estimation in semiparametric transformation models. *Biometrika* **96,** 591–600.

Chou, R. and LeFevre, M. (2011). Prostate cancer screening–the evidence, the recommendations, and the clinical implications. *Journal of the American Medical Association* **306,** 2721–2.

Ciatto, S., Zappa, M., Villers, A., Otto, S., and Auvinen, A. (2003). Contamination by opportunistic screening in the European Randomized Study for Prostate Cancer Screening. *BJU Int* **92,** Suppl 2:97–100.

Dejardin, D., Lesaffre, E., and Verbeke, G. (2010). Joint modeling of progression-free survival and death in advanced cancer clinical trials. *Statistics in Medicine* **29,** 1724–1734.

Draisma, G., Boer, R., SJ., O., van der Cruijsen, I., Damhuis, R., Schroder, F., and et al (2003). Lead times and overdetection due to prostate-specific antigen

screening: Estimations from the European Randomized Study of Screening for Prostate Cancer. *Journal of the National Cancer Institute* **95,** 868–78.

Draisma, G., Etzioni, R., Tsodikov, A., Mariotto, A., Wever, E., Gulati, R., and et al (2009). Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. *Journal fo the National Cancer Institute* **101,** 374–83.

EUREG (2012). European Cancer Observatory. `http://eco.iarc.fr/eureg/`.

Gulati, R., Inoue, L., Katcher, J., Hazelton, W., and Etzioni, R. (2010). Calibrating disease progression models using population data: a critical precursor to policy development in cancer control. *Biostatistics* **11,** 707–19.

Gulati, R., Tsodikov, A., Wever, E., A.Mariotto, Heijnsdijk, E., Katcher, J., de Koning, H., and Etzioni, R. (2012). The impact of PLCO control arm contamination on perceived prostate-specific antigen screening efficacy. *Cancer Causes and Control* **23,** 827–835.

Hu, C. and Tsodikov, A. (2014a). Joint modeling approach for semicompeting risks data with missing nonterminal event status. *Lifetime Data Analysis* **20,** 563–583.

Hu, C. and Tsodikov, A. (2014b). Semiparametric regression analysis for time-to-event marked endpoints in cancer studies. *Biostatistics* **15,** 513–525.

Jemal, A., Fedewa, S., Ma, J., Siegel, R., Lin, C., Brawley, O., and et al (2015). Prostate cancer incidence and PSA testing patterns in relation to USPSTF screening recommendations. *JAMA* **314,** 2054–61.

Jemal, A., Ma, J., Siegel, R., Fedewa, S., Brawley, O., and Ward, E. (2016). Prostate cancer incidence rates 2 years after the US Preventive Services Task Force recommendations against screening. *JAMA Oncol* **2,** 1657–60.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer.

Lee, S.-Y. and Tsodikov, A. (2013). Correcting the results of the wrong model: Treatment effects under early detection of cancer. *Journal of Statistical Theory and Practice* **7,** 421–441.

Mariotto, A., Etzioni, R., Krapcho, M., and Feuer, E. (2007). Reconstructing prostate-specific antigen (PSA) testing patterns among black and white men in the US from Medicare claims and the National Health Interview Survey. *Cancer* **109,** 18771886.

Merrill, R., Feuer, E., Warren, J., Schussler, N., and Stephenson, R. (1999). Role of transurethral resection of the prostate in population-based prostate cancer incidence rates. *American Journal of Epidemiology* **150,** 848–860.

Moyer, V. and Force, U. (2012). Screening for prostate cancer: U.S.Preventive Services Task Force recommendation statement. *Ann Intern Med* **157,** 120–34.

Pinsky, P., Black, A., Kramer, B., Miller, A., Prorok, P., and Berg, C. (2010). Assessing contamination and compliance in the prostate component of the Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial. *Clinical Trials* **7,** 303–11.

Pinsky, P., Miller, A., Kramer, B., Church, T., Reding, D., Prorok, P., and et al (2007). Evidence of a healthy volunteer effect in the prostate, lung, colorectal, and ovarian cancer screening trial. *Am J Epidemiol* **165,** 874–81.

Pinsky, P., Prorok, P., Yu, K., Kramer, B., Black, A., Gohagan, J., and et al (2016). Extended mortality results for prostate cancer screening in the PLCO trial with median 15 years of follow-up. *Cancer* .

PLCO (2016). Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (plco). `https://prevention.cancer.gov/major-programs/prostate-lung-colorectal`.

Rice, J. D. and Tsodikov, A. (2016). Semiparametric time-to-event modeling in the presence of a latent progression event. *Biometrics* **doi:10.1111/biom.12580,**.

Schroder, F., Hugosson, J., Roobol, M., and et al (2009). Screening and prostate-cancer mortality in a randomized European study. *New England Journal of Medicine* **360,** 1320–1328.

Schroder, F., Hugosson, J., Roobol, M., and et al (2012). Prostate-cancer mortality at 11 years of follow-up. *New England Journal of Medicine* **366,** 981–990.

Schroder, F., Hugosson, J., Roobol, M., and et al (2014). Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. *Lancet* **384,** 2027–2035.

SEER (2017). The surveillance, epidemiology, and end results (seer) program. `https://seer.cancer.gov/about/`.

Serfozo, R. (2009). *Basics of Applied Stochastic Processes*. Springer-Verlag, Berlin.

Shiryaev, A. N. (1960). Some problems in the spectral theory of higher order moments I. *The Probability Application* **5,** 265–284.

Thompson, I., Goodman, P., Tangen, C., Lucia, M., Miller, G., Ford, L., and et al (2003). The influence of finasteride on the development of prostate cancer. *New England Journal of Medicine* **349,** 215–24.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risk. *Proc Natl Acad Sci USA* **72,** 20–2.

Tsodikov, A. (2003). Semiparametric models: a generalized self-consistency approach. *Journal of the Royal Statistical Society, Series B (Methodological)* **65,** 759–774.

Tsodikov, A. and Chefo, S. (2008). Generalized Self-Consistency: Multinomial logit model and Poisson likelihood. *J Stat Plan Inference* **138,** 2380–97.

Tsodikov, A., Liu, L., and Tseng, C. (2014). Likelihood transformations and artificial mixtures. *IMS Collections* .

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

Uno, H., Claggett, B., Tian, L., Inoue, E., Gallo, P., Miyata, T., and et al (2014). Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *J Clin Oncol* **32,** 2380–5.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes*. Springer.

Wang, S. and Tsodikov, A. (2010). A Self-consistency Approach to Multinomial Logit Model with Random Effects. *J Stat Plan Inference* **140,** 1939–47.

Wenver, E., Draisma, G., Heijnsdijk, E., Roobol, M., Boer, R., Otto, S., and et al (2010). Prostate-specific antigen screening in the United States vs in the European Randomized Study of Screening for Prostate Cancer-Rotterdam. *Journal of the National Cancer Institute* **102,** 352–5.

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

Zeng, D. and Lin, D. Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society, Series B (Methodological)* **69,** 507–564.