(revised 2014 for electronic submission)

Causal Mediation Analysis with failure time outcome and error-prone longitudinal covariate

Cheng Zheng

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Ross L. Prentice, Chair

Ying Huang

Peter B. Gilbert

Xiao-Hua A. Zhou

Ying Q. Chen

Program Authorized to Offer Degree:
UW Biostatistics

University of Washington

Abstract

Causal Mediation Analysis with failure time outcome and error-prone longitudinal covariate

Cheng Zheng

Chair of the Supervisory Committee:
Professor Ross L. Prentice
Biostatistics

Mediation analyses are important for understanding the biological mechanisms whereby a treatment/exposure influences an outcome of interest. For example, one may be interested in whether body fat accumulation mediates an association of certain dietary patterns with chronic diseases risk. Similarly mediation analyses may aim to achieve an understanding of which elements of a multi-faceted dietary modification intervention were most influential in affecting disease incidence. Several challenges occur in mediation analysis: (1) the longitudinal and observational nature of the dietary variables and BMI/weight; (2) the measurement error in dietary variables which are often assessed using food frequency questionnaires; (3) control of measured/unmeasured confounders. In this dissertation, we proposed a general potential outcome framework for causal mediation analysis with failure time outcome and longitudinal mediator/exposure with measurement error. We proposed a method to correct for the systematic bias in longitudinal self-reported dietary data and use the calibrated data to estimate parameters in the survival model. We also proposed a robust estimator of key survival model parameters that can accommodate the existence of certain types of unmeasured confounders. We studied the performance of regression calibration methods for multiple choices of survival models numerically. We analyzed some important epidemiologic data and provided scientific information on the interplay between dietary exposures, physical activity and BMI in relation to site-specific cancer and other chronic diseases.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to gratefully and sincerely thank my dissertation advisor Dr. Ross L. Prentice for his inspiration and guidance on this work. His mentorship and support provide me a great environment to pursue my research interest and long-term career goal. I would also like to thank my supervisory committee, Ying Q. Chen, Peter B. Gilbert, Ying Huang and Xiao-Hua A. Zhou for their insightful input on this dissertation. I would like to thank Stephen M. Schwartz to serve as my graduate school representative. I would like to thank Marian Neuhouser, Lesley Tinker and Chongzhi Di for their help in analyzing Women's Health Initiative data. Finally, I would like to thank my wife Ran and my parents, Yuan and Huailie, for their encouragement and eternal love.

DEDICATION

To my family.

Chapter 1

INTRODUCTION

In this chapter, we give the background of the motivating data from the Women's Health Initiative (WHI) and will review the current methods for mediation analyses and measurement error modeling.

1.1 Data and question of interest

Our motivating data comes from the Women's Health Initiative, a research program among 161,808 postmenopausal women, aged 50-79 at enrollment at 40 clinical center in the US. We will mainly focus on the dietary modification (DM) trial (n=48,835) and the Nutritional Biomarker Study (NBS, 2004-2005), where a 544 women biomarker subsample was randomly selected from 12 of the 40 clinical center that participated in the WHI program. To fully use the available data, we also consider use the data from WHI observation study (OS) [23] (n=93,676) to improve efficiency of our proposed estimation procedure under additional assumptions. Similar to the DM trial, a subgroup with 450 women was randomly selected from OS to enroll in a Nutrition and Physical Activity Assessment Study (NPAAS), which also included the same total energy and protein biomarkers. Below we describe these data as well as the outcome ascertainment in detail.

1.1.1 The WHI Dietary Modification Study

The DM trial component (recruited 1993-1998) of the WHI program is a randomized controlled trial, which assign of 48835 postmenopausal women to either low-fat eating pattern (40%) or usual dietary behavior (60%)[42]. The intervention was carried out in small groups (5-15 women) and includes both nutritional and behavioral components. The principal nutritional goal of the intervention was to reduce fat consumption to 20% of energy. Women who were on a low-fat diet (<32% energy from fat as estimated by a FFQ) were excluded

from the study. Approximately half of the potential participants were excluded for this reason, many of whom enrolled in the OS, leading to rather different dietary patterns in the DM trial and OS. A secondary goal with DM trial was to increase the daily vegetable and fruit intake and daily grain intake. The average follow up time was 8.1 years with disease outcomes reported semi-annually. The food frequency questionnaire (FFQ) data and weight were recorded at baseline, and 1 year, and approximately about 3 years and 6 years thereafter on a rotating basis. A random 4.3% subsample with annual FFQ and weight assessments was used to monitor the adherence to low-fat diet in the intervention group and dietary trends in the control group [34]. Another dietary assessment method, a 4-day food record (4DFR) was collected at baseline in the DM trial, and 24-hour dietary recalls were subsequently applied in certain random subsamples.

1.1.2 Nutrientional Biomarker Study

There is a Nutrientional Biomarker Study (NBS, 2004-2005), where a 544 women from DM trial subsample was randomly selected from 12 of the 40 clinical center that participated in the WHI program. For this subsample, a doubly labeled water protocal [50] was applied to provide an accurate measurement of short term total energy consumption, while 24-hour urine nitrogen [5] was used to provide an objective measurement of short term protein intake. The Nutritional Biomarker Study (NBS) also recorded FFQ measurement for all 544 women in order to build calibration equations for the self-reported data [32]. Among them, 111 (20% subsample) repeated the same protocol about 6 month later to assess repeatability.

1.1.3 The WHI Observational Study

The WHI Observational Study enrolled 93,676 postmenopausal women at 40 U.S. clinical centers during 1993-1998 [23] with the aim of studying disease risk factors broadly. Women having a predicted survival of less than three years were excluded but there was no dietary eligibility criterion. OS women completed a WHI food frequency questionnaire (FFQ) as a part of their enrollment process. The WHI FFQ collects frequency of intake and portion size information for the past three months for 122 foods or food groups, along with 19

adjustment questions having a focus on dietary fat, and four summary questions [34]. The protocol was applied at baseline and 3 years thereafter in the OS, along with standard measurements of weight and height. The average follow up time was 8.1 years in the basic WHI program with all disease outcomes self reported annually.

OS participants completed a baseline WHI Personal Habits Questionnaire. The WHI PHQ is a short, self-administered questionnaire that inquires about the usual frequency and duration of walking activity outside the home, as well as other mild, moderate, or strenuous recreational activity [19]. Standard intensity values [2], expressed as metabolic equivalent units (METs), were assigned to each activity item and multiplied by reported duration, and summed to compute activity-related energy expenditure (AREE) in MET-hours per week. Since the PHQ focuses on recreational activity only, other WHI questionnaires were used to obtain MET estimates for housework, yard work, sitting, sleeping and all other activities, and the activity sources were combined to produce a total daily AREE estimate [31]. METs were assigned to each of the activity categories using standard algorithms [2].

1.1.4 Nutrition and Physical Activity Assessment Study

The WHI Nutrition and Physical Activity Assessment Study enrolled 450 weight-stable postmenopausal women from the OS during 2007-2009 [39]. These women were recruited from OS enrollees at 9 WHI clinical centers. Black and Hispanic women were oversampled as were women in the extremes of body mass index (BMI) and relatively younger postmenopausal women. Women were excluded from NPAAS for having any medical condition precluding participation, weight instability in preceding months, or travel plans during the study period. A 20% reliability sub-sample repeated the entire biomarker study protocol at about 6 months after the original protocol application.

The NPAAS study protocol [39][31] involved two clinical center visits separated by a two-week period, along with at-home activities. The first visit included eligibility confirmation; informed consent; measured height and weight using a standardized protocol; doubly-labeled water (DLW) dosing for short term energy expenditure assessment; completion of FFQ, dietary supplement, PHQ and other questionnaires needed for AREE assessment. This visit

also included collection of a blood specimen and spot urines after DLW dosing. Participants collected 24-hour urine on the day prior to the second clinic visit. At the second clinic visit, the 24-hour urine samples were delivered to clinic staff, participants provided additional spot urine and a fasting blood, and indirect calorimetry was conducted. Dietary data were analyzed for nutrient content using the University of Minnesotas Nutrition Coordinating Center nutrient database (NDS-R).

Total energy expenditure during the two-week protocol was estimated from relative urinary elimination rates of oxygen-18 and deuterium. In weight-stable persons total energy consumption (TE) over a 2-week period is objectively estimated by this procedure [50]. Resting energy expenditure was estimated [31] by indirect calorimetry using a standard protocol [14]. Metabolic carts were calibrated each day and gasses were monitored during each test. Participants arrived after a 12-hour fast and rested in a semi-reclined position in a thermally neutral room for 30 minutes followed by a 30-minute test. Data from the first 10 minutes were excluded based on time needed to achieve steady-state metabolism, defined as 10 minutes during which the oxygen consumption, minute ventilation and the respiratory exchange ratio did not vary by more than 10% [14]. Participants who did not reach a steady state or did not have at least 10 minutes of useable data (n=16) were not included in AREE biomarker analyses. The AREE biomarker was defined as TE from doubly labeled water minus REE.

1.1.5 Outcome Ascertainment and Categories

Clinical outcomes were reported annually in the OS, by self-administered questionnaire [10]. The initial cardiovascular disease or invasive cancer event during OS follow-up were confirmed by review of medical records and pathology reports by physician-adjudicators at local clinical centers. Additionally, coronary heart disease (CHD), stroke, and all deaths were centrally reviewed by expert committees, and all cancers except non-melanoma skin cancer were centrally coded using the NCIs SEER system [10]. Centrally adjudicated data were used when available; otherwise locally adjudicated outcomes were utilized.

CVD categories included those previously considered in relation to TE [38], along with

congestive heart failure (CHF). Cancer categories included those previously considered [41], along with an obesity-related cancer category. Obesity-related cancer was defined as combined breast, colon, rectum, endometrium, and kidney cancer.

Prevalent diabetes at baseline was self-reported during eligibility screening. Incident diabetes during follow-up was documented by self-report at each annual contact. Data from a WHI diabetes confirmation study showed that prevalent and incident diabetes was consistent with medication inventories of oral agents or insulin [27].

At the end of the planned WHI program time period (April 2005), women were invited to re-enroll for an additional five years of non-intervention follow-up, and more than 80% of women chose to do so. The CVD and cancer association analyses presented here include OS follow-up through September 30, 2010.

1.1.6 Question of Interest

In this dissertation, we aim to develop methods to reliably evaluate the relationship of dietary modification intervention, total energy consumption, BMI, activity-related energy expenditure with various chronic disease risks. Previous studies showed that randomization to the DM intervention group leads to a decrease in fat intake and some modest weight reduction which tended to dissipate over time. The risk of breast cancer was somewhat reduced in the intervention group (logrank p=0.07) with stronger evidence for risk reduction among women whose baseline diet was relatively high in fat. Assuming an intervention effect, one can ask whether the breast cancer reduction was mediated by certain aspects of change in dietary patterns. Though the primary intervention goal, reduction in percentage of energy from fat, is of scientific interest, we will not focus on it as we do not have a suitable biomarker for that variable and systematic bias in the measurement error can not be handled without suitable biomarker data. Instead, we will focus on the total energy and physical activity-related energy expenditure to illustrate our mediation methods. Specifically, we are interested in the following questions: (1) does the dietary modification intervention decrease breast cancer risk through decreasing total energy consumption; (2) does low total energy consumption decrease diabetes risk through decreasing BMI; (3) does the physical activity

and total energy expenditure jointly affect the risk of cardiovascular disease, different kinds of cancer and diabetes.

To answer this kind of question, calibration equation without longitudinal data consideration have been developed [38]. Using Cox regression, it was found that the calibrated energy consumption was significantly associated with risk of breast cancer in the DM trial comparison group and OS, while FFQ energy without calibration did not show this association [41]. The association was decreased and no longer significant when adjusting for BMI. Also, using a similar calibration approach, the accuracy and measurement properties of three different dietary assessment methods have been studied and compared [39]. One limitation of the data is that the biomarker study was carried out during 2004-2005, at least 6 years after randomization for participating women, which means the calibration equation build from this data arose from those DM trial enrollees who were still alive and actively participating after more than 6 years since randomization. In this dissertation, we consider modeling longitudinal dietary profile data to develop a time-varying calibration procedure, rather than using the same calibration equation at all follow-up times. Also, we consider BMI at baseline instead of "current" BMI as a factor for the systematic bias adjustment in FFQ data. By doing so, we can separate the effect of measurement error and confounding effect of baseline BMI and mediating effect for the change in BMI from baseline. Also, we deal with the possibility of unmeasured confounders that might prevent us from developing a causal interpretation for both short-term dietary effects and time varying dietary effects. Figure 1.1 gives our hypothesized causal diagram when we don't allow time-varying effects.

1.2 Causal mediation analysis

For causal inference with a two arm randomized trial, it is common to define the average causal effect (ACE) as the mean difference of the potential outcomes [47]) between treatment and control, $Y(1) - Y(0)$. For some skewed continuous or quantitative outcomes, the 'distributional causal effect' (DCE), defined as $Pr(Y(1) > c) - Pr(Y(0) > c)$ for all possible c values has been proposed [20]. Robins (1986) [43] extends this idea to survival outcomes with the effect of exposure in a time period by defining the potential outcome in the form of $Y(\underline{a})$ where $\underline{a}$ is a pre-determined external process $a(t), t \in [0, \infty]$. Robins and

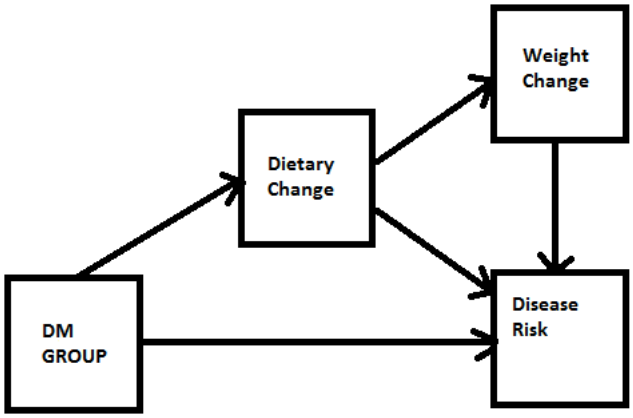


Figure 1.1: Our hypothesized directed acyclic graph (DAG).

Greenland (1992) [46] extend the potential outcome to the mediation context, and define the 'controlled direct effect' (CDE) and 'controlled mediator effect' (CME) as follows:

$$\begin{aligned} CDE &= EY(z_1, m) - EY(z_2, m) \\ CME &= EY(z, m_1) - EY(z, m_2), \end{aligned}$$

where the potential outcome $Y(z, m)$ means the outcome if the study subject is required to have exposure at level z and mediator at level m. This is often compared with the natural direct effect (NDE) and natural indirect effect (NIE) [18] which are defined as

$$\begin{aligned} NDE &= EY(z_1, M(z_1)) - EY(z_2, M(z_1)) \\ NIE &= EY(z_1, M(z_1)) - EY(z_1, M(z_2)), \end{aligned}$$

where $M(z)$ means the potential value for the mediator if the individual is required to have exposure at level z. The controlled direct and mediator effect can be easily extended to the longitudinal exposure setting, while the definition of natural effects is evidently limited to 'point'/time independent exposures.

Under the so-called 'sequential ignorability assumption', which requires that there is no unmeasured confounding between the mediator and outcome, many semi-parametric

and nonparametric estimators for the direct and indirect effect have been proposed [17] [59]. Sensitivity analysis methods [18] [65] [60] have been developed to partly overcome the weakness of assuming sequential ignorability. To handle the confounder issue in longitudinal settings, marginal structure model (MSM) and structural mean model (SMM) [45] have been proposed. Using G-estimation [44], Ten Have *et al.* (2007) [61] extended structural mean model (SMM) [45] and propose an estimator that allows certain departure from the sequential ignorability assumption. Small (2012) [52] proposed similar estimators from an instrumental variable point of view. Zheng and Zhou (2014) [74] extended this method to the case with multiple arm treatment and multiple mediators. Apart from the above methods, there is also some work that defines and estimates causal mediation effects via principal stratification (PS) [13] [11]. The principal stratification framework assumes that the mediator cannot be manipulated and defines principal strata by $C = (M(0), M(1))$. The strata-specific average causal effect among those with strata $M(0) = M(1)$ is then treated as direct effect while those strata specific average causal effect among those with strata $M(0) \neq M(1)$ as indirect effect. The interpretation of these effects is different from traditional regression models and is considered as associative and dissociative effects as discussed by VanderWeele [67].

When we have a time independent exposure, if the potential mediator is time-varying in a short period, for example, we are interested in mediator at year 0 and year 1 in a long-term follow-up study, then we can treat it as a multivariate mediator $\underline{M} = (M_0, M_1)$. The controlled direct and mediator effects can then be defined as

$$\begin{aligned} CDE &= EY(z_1, \underline{m}) - EY(z_2, \underline{m}), \\ CME &= EY(z, \underline{m}_1) - EY(z, \underline{m}_2). \end{aligned}$$

However, when Y is survival time, we need to assume death or censoring before last measurement of m is ignorable, which requires that survivors at the end of short period can still be viewed as being randomized. Mathematically, this condition can be written as $Z \perp Y(z, \underline{m}) | Y > t_m$ hold for all z and $\underline{m}$, where t_m is the last time that mediator is measured.

In many potential outcome applications, the treatment effect of interest is defined as the difference between the mean of two potential outcomes. The advantage of this is that the effect can be interpreted as both an average individual causal effect and a population-level causal effect. However, we can also define the effect in other scales if we focus on the population-level causal effects. In general, if we use $F^{zm}(\cdot)$ to denote the cumulative distribution function of Y^{zm}, then the comparison can be made using any functional form $G(\cdot)$ of F^{zm}. The choice of functional form may be able to be determined either by scientific interest or mathematical convenience. For example, we can define $G(F^{zm}) = logEY^{zm}$ for count outcomes and $G_t(F^{zm}) = \lambda^{zm}(t)$ for hazard rate, $\lambda(t)$ [22]. With additive hazard model, Martinussen *et al.* (2011) [28] proposed an estimator for direct effect. Vanderweele (2011) [66] also propose to use quantities like the survivor function $S^{zm}(t) = 1 - F^{zm}(t)$, the cumulative hazard function $\Lambda^{zm}(t) = -\log S^{zm}(t)$, and the event time T^{zm}. Vanderweele mentions that different quantities will be preferable for different models. Under the sequential ignorability assumption, Tchetgen [56] [58] proposed robust estimators of indirect effects with Cox and accelerated failure time (AFT) models. An important result for mediation analysis is that the total effect can be decomposed into direct and indirect effects. Under sequential ignorability, and a linear regression model, the indirect effect has a product form composed of the mediator effect and the effect of exposure on the mediator. More generally, the product form does not apply and the model conditional on the mediator may have a different form from the corresponding marginal model. However, when the product form exists and the models are in the same form, the interpretation will be straightforward. Hence, one can hope to find some convenient functional form $G(\cdot)$ to induce a product form. This suggests that we should use $\log \Lambda(t)$ for the Cox model, $\Lambda(t)$ for the additive hazard model, $\log T$ for the AFT model with failure time outcome. Here we use the transformed cumulative hazard rather than time specific hazard since the former may have a more useful interpretation. The reason is that transformed cumulative hazard can be viewed as a transformed survival probability, and the comparison is essentially between two probabilities. On the other hand, the hazard compares two conditional probability and the condition will be different for two potential outcomes. Conditioning on survival at a time t may make the two groups no longer comparable. If we are interested in the comparison of instantaneous

disease occurence rates for those who can survive for at least t regardless of their intervention assignment, then we need to make assumptions about the joint distribution of different levels of potential outcomes, rather than model them marginally.

The AFT model is a little bit difficult to use with time-varying covariates and will not be very fully discussed in this dissertation. Comparing the Cox model with the additive hazard model, the Cox model often fits real data better and there are theoretical results for the asymptotics for Cox model under model misspecification. However, the additive hazard model also has some advantages in this context: (1) the indirect effect has a product form under additive hazard model; and (2) the parameter in additive hazard model is collapsible. In this dissertation, we will focus primarily on the Cox model and we will also discuss additive hazard model, especially for the case where we do not assume sequential ignorability. For survival outcomes, we can view the hazard regression model as a process of sequential binary outcomes. As is discussed in Maxwell (2007) [29], estimation of mediation effects via cross sectional data will lead to biased results for a longitudinal mediating process. In Maxwell (2011) [30], the lag model for mediation analysis is discussed. Albert and Nelson (2011) [3] discussed how to identify suitable effects when mediators are measured at several time points. Zheng and Van der Laan (2012) [75] proposed an estimator for direct effects with the time-varying mediator model unspecified. When considering time-varying exposure, Zhang *et al.* (2011) developed models for continuous processes with discrete observed time point. We can adapt their result for time-varying covariate survival models. The mechanisms of some different types of problems with up to two time points are shown in Figure 1.2. For the top left, it is a simple mediation mechanism with time independent exposure Z and time independent mediator M while the outcome can be survival time. For the top right, it extend the model to the mediator process and allow lag-time effect. Since mediator at both time points are allowed to affect the outcome Y, here the outcome should have happened later than the measurement of M_2 and thus in general cannot be survival until some time earlier than M_2. The middle left figure shows a speical case that both mediator and exposure are time-varying but the effect lags during different time points are the same. The middle right figure shows the case for time-independent exposure, time-dependent mediator and outcome. For this case, the outcome Y_t can be considered as

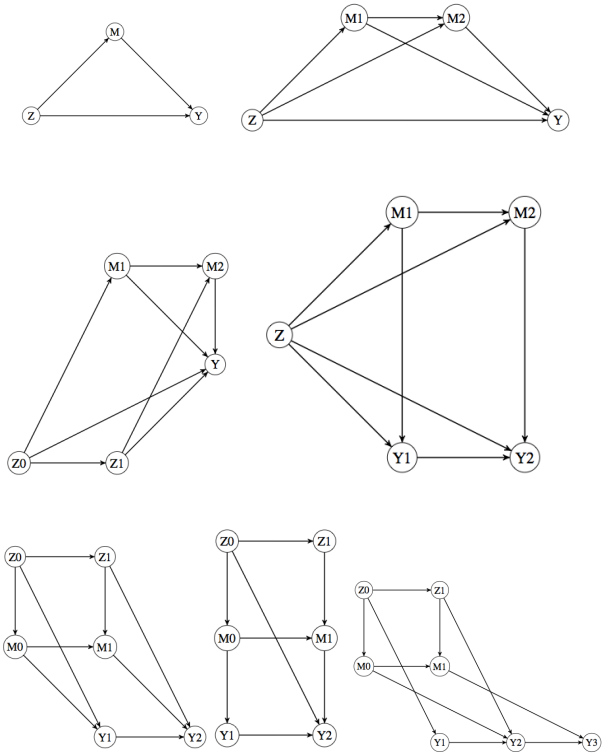
whether a subject is still alive until time t and thus the arrow can be considered as effects on hazard scale. The bottom three figures show the case that all variables are time-varying and with different possible time lags. We are also interested in quantifying the portion of effect through mediation and this will be challenging when the effect is time-varying. One approach is to define a time-varying portion of mediation $R(t) = DE(t)/TE(t)$ and then plot the cumulative weighted portion. Another is to compute the weighted average DE and TE and then compute the portion of mediation. Procedures for estimating model specific effects will be derived in later chapters, where specific formulas will be given.

1.3 Measurement error issues

The measurement error issue has been introduced in the potential outcome framework in recent years. Hernan and Cole (2009) showed that measurement error tends to bring in many difficulties in this causal inference formulation. Pearl (2010)[35] summarized two ways to incorporate measurement error in causal inference. One way is to include error-prone variable as well as the precise measured variable in the causal diagram to derive the formula for the causal parameter of interest. Another way is to derive formula for the causal parameter of interest based on the precise measured variable and then estimate the parameters in that formula via traditional techniques for handling measurement error. VanderWeele *et al.* (2012) [68] discussed issue of measurement error in mediation analysis. However, no estimation procedure is provided and the discussion is limited to single time point. Tchetgen and Lin (2012) [57] proposed an estimator for natural direct effect with measurement error in the mediator.

Outside of the potential outcome framework, measurement error issue has been intensively studied in survival analysis. The most intuitive idea is to estimate the unobserved true variable by its conditional mean in the context of induced hazard function [36]. Regression calibration is also simple way to handle measurement error. The original regression calibration method uses the same calibration equation for any time point [36][51] while the modified risk set regression calibration [70] method estimates different calibration equation at different time point using the data from those that are still at risk. Other variants [4] [25] use two-stage or monte carlo method to realize the calibration. Another view is to treat the

Figure 1.2: Different DAGs depending on point, short period or time varying mediator and exposure.



Here Z denotes exposure, M denotes mediator and Y denotes outcome. M0, M1, M2 denote mediator at different time points, Z0, Z1 denote exposure at different time points, Y1, Y2, Y3 denote indicator of survival until different time points.

precise value as missing and then use multiple imputation [49] to construct estimates [9] [24]. Other functional methods, such as corrected score[71] [54] [16], conditional score [62] and expected estimating equation [69] have been developed. Though theoretically consistent, these types of estimators tend to be less efficient compared to the calibration methods. With longitudinal data and survival outcome, many joint modeling methods [53] [63] [55] have also been developed. Zeng and Cai (2005) [72] give the asymptotics for this kind of estimator. Using joint modeling and likelihood based estimator tend to be computational complicated and is sometimes sensitive to the model assumption. Similarly, Bayesian estimators [8] also face the problem of intensive computation and sensitivity to prior distribution and model specifications.

However, most of the previous studies are based on the fact that there is some replication for the classical measurement error model or there is a validation subsample where the true underlying dietary is measured without error. However, in our example, the biomarker based measurement only plausibly provides a surrogate of the true underlying dietary value without systematic bias. Some calibration equation-based methods have been used for previous analysis without time-varying mediator and dietary exposure [15] [32]. However, this approach primarily considers models for the long-term average dietary intake and treats different time points as repeated measurement. To make the assumption that there are no time trends in different time point of self-reported variables, the analysis is limited to the dietary modification trial comparison group and OS data in the Women's Health Initiative (WHI) [41]. When we want to include the dietary intervention group in the analysis, viewing different time points as repeated measurement is inappropriate. Another issue is that although it is found that BMI is an important term in the calibration equation, we do not know whether this is due to the systematic bias in FFQ caused by BMI or due to the association between BMI and energy consumption beyond the measurement issue. Also possible is that BMI is associated with some other variables that cause systematic bias in FFQ. Such variables may be personal characteristics that are time-independent and can be captured by the baseline factors rather than the outcome after receiving intervention.

In this dissertation, we required that the measurement error in certain variable can only be affected by its previous variables. For example, if we consider the longitudinal total

energy consumption as exposure $Z(t)$ and longitudinal BMI as mediator $M(t)$. Then our assumption is that the measurement error in the year 3 self-reported total energy consumption can depend on BMI at baseline or year 1, but should not depend on BMI at year 3. Due to the large time gap between year 1 and year 3, it is possible that the true systematic bias in the self-reported data at year 3 not only depend on BMI at year 1, but also depend on later BMI, say year 2. Then in general, the method might yield biased estimators similar as what is shown in Chap 3. But for our dataset, using the biomarker subsampe, we find that there seems not a large residual systematic bias related to year 3 BMI after year 1 BMI is modeled.

In Chapter 2, we discussed the mediation analysis when the exposure is time-independent and the mediator is mis-measured. We illustrated our method by assessing how the effect of dietary modification on decreasing breast cancer risk is mediated by total energy consumption. In Chapter 3, we discussed how to perform mediation analysis when the exposure is mis-measured. We applied our method to assess how the association between total energy intake and diabetes risk is mediated by BMI. In Chapter 4, we discussed how to perform mediation analysis when both the exposure and mediator are reported with systematic bias. We applied our method to assessing how the association between physical activity and chronic disease risk is mediated by total energy consumption for various chronic diseases. In Chapter 5, we extend a simple model in Chapter 2 to allow the existence of certain unmeasured confounder. We applied our method to assess the extent to which the dietary modfication intervention decrease breast cancer risk through decreasing total energy consumption. In Chapter 6, we summarize the results and discuss future extensions.

Chapter 2

MEDIATION ANALYSIS WITH LONGITUDINAL ERROR-PRONE MEDIATOR

In this chapter, we consider how to perform mediation analysis to address the following questions: (1) Is the Women's Health Initiative DM trial intervention efficacy on reducing breast cancer risk mediated by long term average total energy intake? (2)Is the DM intervention efficacy on reducing breast cancer risk mediated by short term total energy intake? (3) Is the DM intervention efficacy on reducing breast cancer risk mediated by the long term total energy intake process?

2.1 Notation

We use Z to denote an 'exposure' which is measured at baseline and is not subject to measurement error. We use $M(t)$ to denote the error free underlying mediator value at time t. The corresponding potential outcome is denoted as $M(t, z)$. When $M(t)$ is viewed as an external process, we can define the average in a long time period $[0, \tau]$, i.e. $M = \tau^{-1} \int_0^\tau M(t)dt$, as the long term average mediator. We denote a set of baseline covariates that are included in the model to control confounding as V, and the self-reported mediator value at time t that may be subject to systematic bias is denoted as $Q(t)$. The biomarker for $M(t)$ will typically be assumed to have classical measurement error only and is denoted as $W(t)$. Here, by classical measurement error, we mean that the error is independent of all covariates and thus can be considered as the random noise. For simplicity, we first assume the longitudinal data are collected at pre-determined time points $0 = t_0 < t_1 < \cdots < t_K < \infty$ and suppose that the true processes only jump at those time points. We denote $\underline{\tilde{Q}(t)} = (Q(t_1), \cdots, Q(t_{f(t)}))$, where $f(t) = max_{i=0,\cdots,K}(t_i <= t)$. Similarly, we denote $\underline{\tilde{W}(t)} = (W(t_1), \cdots, W(t_{f(t)}))$ and $\underline{\tilde{M}(t)} = (M(t_1), \cdots, M(t_{f(t)}))$. Also, we denote $\tilde{Q}(t) = Q(t_{f(t)})$, $\tilde{W}(t) = W(t_{f(t)})$ and $\tilde{M}(t) = M(t_{f(t)})$. A discussion of extension of this notation to randomly selected time points will follow the main results. Also, we assume the

$\tilde{W}(t)$ is measured only for a randomly selected subgroup. If we consider that measurement following censoring is impossible, then we need to assume that the censoring is completely at random. Discussion will be given later for the case where censoring depends on measured covariates. To distinguish the observed variables and levels that we would like to potential assign the individual to, we use the capital letter for the former and the lower letter for the latter.

2.2 Time-independent exposure and long term error-prone mediator

When both the exposure and the mediator are time-independent, we define the potential event time and censoring time as $T_i(z, m)$ and $C_i(z, m)$. The real event time and censoring time can be denoted as T_i and C_i and the observed composite endpoint is denoted as $T_i^* = T_i \wedge C_i$, $\Delta_i = I(T_i \leq C_i)$.

2.2.1 Model

We assume the following model structures:

$$
\begin{aligned}
M_i(z) &= g_z(z, V_i; \boldsymbol{\alpha}) + \epsilon_i(z), \\
Q_i(t) &= g_{q0}(t, Z_i, V_i; \boldsymbol{\gamma}) + g_{q1}(t, Z_i, V_i; \boldsymbol{\gamma}) M_i + \varepsilon_i(t), \\
W_i(t) &= M_i + \omega_i(t).
\end{aligned}
$$

Here $\epsilon_i(z)$ is normally distributed with variance $\sigma_z^2(z, V_i; \boldsymbol{\theta})$ and is independent of other covariates, $\varepsilon_i(t)$ is a Gaussian process with variance process $K_q(t, s) = K(t, s; \boldsymbol{\theta}_q, Z_i, V_i)$ and is independent of other covariates, and $\omega_i(t)$ is a Gaussian process with variance process $K_w(t, s) = K(t, s; \boldsymbol{\theta}_w, Z_i, V_i)$ and is independent of other covariates. Here $g_z(\cdot)$, $g_{q0}(\cdot)$, $g_{q1}(\cdot)$ are known link functions and we assume $W_i(t)$ to be only available within a biomarker subsample in data analysis. Here we assume a flexible model for potential mediator and self-reported mediator. However, we do constrain the error term of potential mediator to be in an additive form and we require the self-reported mediator to be in a linear form respect to the true underlying mediator.

We have two different types of survival models: general relative risk model,

$$\lambda(t, z, m|V) = \lambda_0(t)h(t, z, m, V; \boldsymbol{\beta}),$$

and additive hazard model,

$$\lambda(t, z, m|V) = \lambda_0(t) + h(t, z, m, V; \boldsymbol{\beta}),$$

as special cases of interest.

To make causal interpretation, we make the SUTVA assumption [47] which contains two sub-assumptions: (1) no multiple version of treatment (2) no interference between subjects. The first part means that the potential outcome $Y_i(z, m)$ does not depend on how we manipulate the intervention and mediator value to level z and m. The second part means that one's intervention or mediator level will not affect another study subject's outcome. This SUTVA assumption ensures that $Y_i(z, m)$ is well-defined. For the chronic diseases and given that participants are unlikely to be in the same household, concern about interference is minimumal in our example. This assumption may not hold if for example, the outcome of interest is an infectious disease. The second assumption we need is consistency, which means that what we observed is a realization of potential outcome. When there is measurement error, this assumption does not hold, so we need to instead make this assumption on the underlying error free mediator. This assumption is required to link the potential outcome to the observed outcome.

The third assumption we would like to make is the 'sequential ignorability' assumption, which means that the intervention can be viewed as if randomized conditional on the observed covariate and the mediator can be viewed as randomized conditional on the observed covariate as well as the intervention. Mathematically, this can be written as

$$Z \perp (M(z), T(z, m), C(z, m))|V,$$
$$M \perp (T(z, m), C(z, m))|Z, V.$$

The fourth assumption is the positivity assumption. We need that the probability of

observing any levels of (z, m) for any subgroup defined by covariate V is greater than 0.

Under this set of assumptions, we can obtain the following induced model for M:

$$M_i = g_z(Z_i, V_i; \boldsymbol{\alpha}) + \epsilon_i, \tag{2.1}$$

where ϵ_i is normally distributed with variance $\sigma_z^2(Z_i, V_i; \boldsymbol{\theta})$ and relative risk model

$$\lambda(t|Z, M, V) = \lambda_0(t)h(t, Z, M, V; \boldsymbol{\beta}), \tag{2.2}$$

or additive hazard model,

$$\lambda(t|Z, M, V) = \lambda_0(t) + h(t, Z, M, V; \boldsymbol{\beta}). \tag{2.3}$$

2.2.2 Estimation of Model Parameters

We use the induced hazard method to estimate model parameter $\boldsymbol{\beta}$. Depending on whether we condition on the whole history of self-reported mediator or just the current value of the self-reported mediator, we have two kinds of induced hazards. We denote $\hat{M}^{(1)}(t) = E[M|Z, \underline{Q(t)}, V]$, $\hat{M}^{(2)}(t) = E[M|Z, \tilde{Q(t)}, V]$,
$\hat{\Sigma}_M^{(1)}(t) = Var[M|Z, \underline{Q(t)}, V]$, $\hat{\Sigma}_M^{(2)}(t) = Var[M|Z, \tilde{Q(t)}, V]$, then the induced hazard can be written as

$$\begin{aligned} E[\lambda(t|Z, \underline{Q(t)}, V)] &= E[\lambda(t|Z, M, V)|Z, \underline{Q(t)}, V, T^* > t] \\ &\approx E[\lambda(t|Z, M, V)|Z, \underline{Q(t)}, V], \end{aligned}$$

which for model (2.1)

$$\approx \lambda_0^*(t)h(t, Z, \hat{M}^{(1)}(t), V; \boldsymbol{\beta})$$

or for model (2.2)

$$\approx \lambda_0(t) + h(t, Z, \hat{M}^{(1)}(t), V; \boldsymbol{\beta})$$

The first approximation requires rare disease assumption and the second approximation is the first order approximation when h is in nonlinear form. When we have a special form for $h(t, Z, M, V; \boldsymbol{\beta})$, e.g. $h(t, Z, M, V; \boldsymbol{\beta}) = h_0(t, Z, V; \boldsymbol{\beta}) \exp\{h_1(t, Z, V; \boldsymbol{\beta})M\}$, we might assume normality for $M|Z, Q, V$ to obtain a second order approximation similar to Zhao and Prentice (2014). The induced hazards for model (2.1) and (2.2) are then in the following form respectively:

$$\lambda_0^*(t)h_0(t, Z, V; \boldsymbol{\beta}) \exp\{h_1(t, Z, V; \boldsymbol{\beta})\hat{M}^{(1)}(t) + h_1(t, Z, V; \boldsymbol{\beta})\hat{\Sigma}_M^{(1)}(t)h_1(t, Z, V; \boldsymbol{\beta})^T/2\}$$
$$\lambda_0(t) + h_0(t, Z, V; \boldsymbol{\beta}) \exp\{h_1(t, Z, V; \boldsymbol{\beta})\hat{M}^{(1)}(t) + h_1(t, Z, V; \boldsymbol{\beta})\hat{\Sigma}_M^{(1)}(t)h_1(t, Z, V; \boldsymbol{\beta})^T/2\}.$$

One can replace $\hat{M}^{(1)}(t)$ and $\hat{\Sigma}_M^{(1)}(t)$ by $\hat{M}^{(2)}(t)$ and $\hat{\Sigma}_M^{(2)}(t)$ to give another version of the induced hazards. We can show the estimating equation are in similar form using these two versions. However, when $\hat{M}^{(2)}(t)$ and $\hat{\Sigma}_M^{(2)}(t)$ is used, the condition at different time point t cannot be written as a filtration $\mathcal{F}_t = (Z, \underline{Q(t)}, \underline{Y(t)}, \underline{N(t)}, V)$ or $\mathcal{F}_t = (Z, \underline{Q(\infty)}, \underline{Y(t)}, \underline{N(t)}, V)$ and martingale results cannot be used to prove the estimator's asymptotic properties. Even for $\hat{M}^{(1)}(t)$ and $\hat{\Sigma}_M^{(1)}(t)$, we do need to account for the variation in estimation of $\hat{M}$ and $\hat{\Sigma}_M$. So a Bootstrap method is recommended for variance estimation in data analysis. Also $\hat{M}^{(2)}(t)$ and $\hat{\Sigma}_M^{(2)}(t)$ are often relatively easier to compute and are more stable, so we recommend their use in real applications.

Under a general relative risk model, we can solve the partial likelihood score equation to estimate $\boldsymbol{\beta}$. The estimating equation can be written as follows:

$$\int_0^\infty \sum_{i=1}^n [\frac{\frac{\partial h_i(t)}{\partial \boldsymbol{\beta}}}{h_i(t)} - \frac{\sum_j \frac{\partial h_j(t)}{\partial \boldsymbol{\beta}} Y_j(t)}{\sum_j h_j(t) Y_j(t)}] dN_i(t),$$

where $h(t)$ can be either the first order or second order approximation with either version of $\hat{M}$ and $\hat{\Sigma}_M$. Here $Y_i(t) = I(T_i^* \geq t)$ and $N_i(t) = I(T_i^* \leq t, \Delta_i = 1)$.

Under additive hazard model, we can solve the moment-based estimating equation

$$\int_0^\infty \sum_{i=1}^n [\frac{\partial h_i(t)}{\partial \boldsymbol{\beta}} - \frac{\sum_j \frac{\partial h_j(t)}{\partial \boldsymbol{\beta}} Y_j(t)}{\sum_j Y_j(t)}][dN_i(t) - Y_i(t)h_i(t)dt]$$

When M is observed, the baseline hazard estimator will be of the following forms:

$$\hat{\Lambda}_0(t) = \int_0^t \frac{\sum_i dN_i(s)}{\sum_i h_i(t;\hat{\boldsymbol{\beta}})Y_i(s)},$$

or

$$\hat{\Lambda}_0(t) = \int_0^t \frac{\sum_i dN_i(s) - h_i(t;\hat{\boldsymbol{\beta}})Y_i(s)ds}{\sum_i Y_i(s)}.$$

A first order approximation directly replaces M by $\hat{M}$. However, unlike the estimator for $\boldsymbol{\beta}$, such first order approximation often leads to a large bias in $\hat{\Lambda}$. So we recommend second order approximation, i.e. assume $M|Z,V,Q$ follow normal distribution and compute the expectations in the right hand side numerically using a monte carlo method.

Now the remaining question is how to estimate parameters in $\hat{M}$ and $\hat{\Sigma}_M$. We can specify a parametric model $E[M|Z,V,\underline{\tilde{Q}(t)};\boldsymbol{\delta}]$ or $E[M|Z,V,\tilde{Q}(t);\boldsymbol{\delta}]$ and use

$$E[W(t)|Z,V,\underline{\tilde{Q}(t)};\boldsymbol{\delta}] = E[M|Z,V,\underline{\tilde{Q}(t)};\boldsymbol{\delta}],$$
$$E[W(t)|Z,V,\tilde{Q}(t);\boldsymbol{\delta}] = E[M|Z,V,\tilde{Q}(t);\boldsymbol{\delta}],$$

to obtain an estimator $\hat{\boldsymbol{\delta}}$ by running a regression of $\underline{\tilde{W}(t)}$ on $Z,V,\underline{\tilde{Q}}(t)$. Then we obtain $\underline{\hat{E}_i(t)}$. Similarly we can assume a model for the variance matrix and obtain $\Sigma_M^{(1)}(t)$ and $\Sigma_M^{(2)}(t)$ and estimate them by the relationship

$$Var[W(t)|Z,V,\underline{\tilde{Q}(t)};\boldsymbol{\delta}] = Var[M|Z,V,\underline{\tilde{Q}(t)};\boldsymbol{\delta}] + Var[\omega(t)],$$
$$Var[W(t)|Z,V,\tilde{Q}(t);\boldsymbol{\delta}] = Var[M|Z,V,\tilde{Q}(t);\boldsymbol{\delta}] + Var[\omega(t)],$$

where $Var[\omega(t)]$ can be estimated from longitudinal measurements on $W(t)$. Depending on the parameter dimension in $\omega(t)$, two or more time points will be needed.

The method above supposes separate models for each time point, which is recommended when we have small number of time points and have complicated covariance structure for $Q(t)$ and $W(t)$. However, if we have large number of time points and a relatively simple covariance structure with low dimension of parameters, we might borrow information

across different time points to estimate $\hat{M}$ and $\hat{\Sigma}_M$. For this purpose, we propose either using maximum likelihood estimator or use moment-based estimator for parameters in the measurement error model.

To obtain an estimator of $\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\theta}$, we consider the joint distribution of $(\underline{W(t)}, \underline{Q(t)}|Z, V)$. Under our normality assumptions, this is normally distributed with mean $\mu(Z, V, \boldsymbol{\alpha}, \boldsymbol{\gamma})$ and variance $\Sigma(Z, V, \boldsymbol{\gamma}, \boldsymbol{\theta})$. When Z and V are discrete, we can obtain the empirical moment $\mu_n(Z, V)$ and $\Sigma_n(Z, V)$ from the data and then we can construct generalized moment estimators $\hat{\boldsymbol{\alpha}}_{MM}, \hat{\boldsymbol{\gamma}}_{MM}, \hat{\boldsymbol{\theta}}_{MM}$, by minimizing $\|(\mu(Z, V, \boldsymbol{\alpha}, \boldsymbol{\gamma}), \Sigma(Z, V, \boldsymbol{\gamma}, \boldsymbol{\theta})), (\mu_n(Z, V), \Sigma_n(Z, V))\|$ under certain well-defined norm $\|\cdot\|$. When Z and V are continuous, a kernel smoothed estimator can be used for $\mu_n(Z, V)$ and $\Sigma_n(Z, V)$.

Another way to obtain an estimator of $\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\theta}$ is to use the maximum likelihood estimation. Since $(\underline{W(t)}, \underline{Q(t)}|Z, V)$ is normally distributed, the marginal distribution when certain $W(t)$ and $Q(t)$ is missing is easy to obtain. Thus the log likelihood can be written as

$$\begin{aligned} l_1 \quad = \quad & \sum_i \{-|\Sigma(Z_i, V_i, \boldsymbol{\gamma}, \boldsymbol{\theta})|/2 \\ & -[(\underline{W_i(t)}, \underline{Q_i(t)}) - \mu(Z_i, V_i, \boldsymbol{\alpha}, \boldsymbol{\gamma})]\Sigma^{-1}(Z_i, V_i, \boldsymbol{\gamma}, \boldsymbol{\theta})[(\underline{W_i(t)}, \underline{Q_i(t)}) - \mu(Z_i, V_i, \boldsymbol{\alpha}, \boldsymbol{\gamma})]^T\} \end{aligned}$$

for the biomarker subsample. Similarly, we can write the likelihood for those without biomarker measured as below

$$l_2 = \sum_i \{-|\Sigma(Z_i, V_i, \boldsymbol{\gamma}, \boldsymbol{\theta})|/2 - [\underline{Q_i(t)} - \mu(Z_i, V_i, \boldsymbol{\alpha}, \boldsymbol{\gamma})]\Sigma^{-1}(Z_i, V_i, \boldsymbol{\gamma}, \boldsymbol{\theta})[\underline{Q_i(t)} - \mu(Z_i, V_i, \boldsymbol{\alpha}, \boldsymbol{\gamma})]^T\}$$

We can either maximize l_1 to obtain $\hat{\boldsymbol{\alpha}}_{ML1}, \hat{\boldsymbol{\gamma}}_{ML1}, \hat{\boldsymbol{\theta}}_{ML1}$ or maximize $l_1 + l_2$ to obtain $\hat{\boldsymbol{\alpha}}_{ML}, \hat{\boldsymbol{\gamma}}_{ML}, \hat{\boldsymbol{\theta}}_{ML}$. We can compute $\hat{M}$ and $\hat{\Sigma}_M$ by pluging in any of the above versions of estimators $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\theta}}$.

Here we need to notice that both the moment estimator and likelihood approach did not handle the fact that covariate at time t is only observable when $C > t$. Since our method use the likelihood conditional on Z and V, both estimator can be biased unless the censoring is independent of T conditional on Z and V. However, under the rare disease assumption

and the assumption that $C \perp T|Z, V$, we have

$$Pr(W, Q|Z, V, T^* \geq t) = Pr(W, Q|Z, V, T \geq t) \approx Pr(W, Q|Z, V)$$

and the estimator will be approximately consistent. When the exposure is time-independent, to yield causal interpretation, we need further assume that Z process does not interact with C process.

In fact, an advantage for the maximum likelihood estimator is that we don't need to restrict the time of observing the covariate to be a sequence of pre-specified time points. We just need to assume that for each individual, the observed time is randomly selected. However, this data structure is then not amendable to moment estimators and simple regression calibration.

Ignoring the fact that the observable time for the covariate process may be related to the failure event time, we have

$$\lambda(t|M, Z, V) = \lambda(t|\underline{\tilde{Q}(\infty)}, Z, V).$$

So this gives a simple way to do the calibration, we can compute $\hat{M}^{(3)} = E[M|Z, V, \underline{\tilde{Q}(\infty)}]$ for each individual and then directly run a regression on $\hat{M}^{(3)}$, Z and V. This method is the one that is most easily implemented since $\hat{M}^{(3)}$ is no longer dependent on t. A second order approximation can also be straightforwardly extended to this settting with $\hat{\Sigma}_M^{(3)} = Var[M|Z, V, \underline{\tilde{Q}(\infty)}]$.

2.2.3 Estimation of Effects Related to Mediation Analysis

After we obtain an estimator for $\boldsymbol{\beta}$, we can estimate the controlled direct effect and controlled mediator effects among subgroup defined by V. Since the effect can be defined as either a cumulative hazard difference (log survival probability ratio), a log cumulative hazard ratio or a survival probability ratio, we can denote the effect as difference in certain transformation function, G of cumulative hazard function, where G can be identity, log, or exponential functions.

For general relative risk model, we have

$$CDE(m,t,V) \;=\; G(\int_0^t \{\lambda(u)h(u,1,m,V;\boldsymbol{\beta})\}du) - G(\int_0^t \{\lambda(u)h(u,0,m,V;\boldsymbol{\beta})\}du),$$

and

$$CME(m_1,m_2,z,t,V) \;=\; G(\int_0^t \{\lambda(u)h(u,z,m_2,V;\boldsymbol{\beta})\}du) - G(\int_0^t \{\lambda(u)h(u,z,m_1,V;\boldsymbol{\beta})\}du),$$

The form can be complicated in general. However, if we have that $h(t,z,m;V;\boldsymbol{\beta}) = h(z,m,V;\boldsymbol{\beta})$ independent of t, then the effect at subgroup V can be simplified to

$$CDE(m,t,V) \;=\; G(\Lambda(t)h(1,m,V;\boldsymbol{\beta})) - G(\Lambda(t)h(0,m,V;\boldsymbol{\beta})),$$

$$CME(m_1,m_2,z,t,V) \;=\; G(\Lambda(t)h(z,m_2,V;\boldsymbol{\beta})) - G(\Lambda(t)h(u,z,m_1,V;\boldsymbol{\beta})).$$

Choosing G to be the log function can leads to exclusion of the baseline function, $\Lambda(t)$, from the effect. This choice can lead to simple effect expression for Cox model with $h(z,m,V;\boldsymbol{\beta}) = exp\{(z,m,V)\boldsymbol{\beta}\}$.

$$\begin{aligned} CDE(m,t,V) &= \log\frac{h(1,m,V;\boldsymbol{\beta})}{h(0,m,V;\boldsymbol{\beta})}, \\ CME(m_1,m_2,z,t,V) &= \log\frac{h(z,m_2,V;\boldsymbol{\beta})}{h(z,m_1,V;\boldsymbol{\beta})}. \end{aligned}$$

Under the rare disease assumption, we have

$$\begin{aligned} \Lambda(t|z,M(z),V,\boldsymbol{\beta}) &= -\log S(t|z,M(z),V,\boldsymbol{\beta}) \\ &= -\log\int S(t|z,m,V,\boldsymbol{\beta})dF_{M(z)}(m|V;\boldsymbol{\alpha},\boldsymbol{\theta}) \\ &\approx 1-\int S(t|z,m,V,\boldsymbol{\beta})dF_{M(z)}(m|V;\boldsymbol{\alpha},\boldsymbol{\theta}) \\ &= \int[1-S(t|z,m,V,\boldsymbol{\beta})]dF_{M(z)}(m|V;\boldsymbol{\alpha},\boldsymbol{\theta}) \\ &\approx \int\Lambda(t|z,m,V,\boldsymbol{\beta})dF_{M(z)}(m|V;\boldsymbol{\alpha},\boldsymbol{\theta}). \end{aligned}$$

So, the natural direct and indirect effects have the following approximate forms:

$$\begin{aligned} NDE(z,t,V) &= \log \frac{\int h(1,m,V;\boldsymbol{\beta}) dF_{M(z)}(m|V;\boldsymbol{\alpha},\boldsymbol{\theta})}{\int h(0,m,V;\boldsymbol{\beta}) dF_{M(z)}(m|V;\boldsymbol{\alpha},\boldsymbol{\theta})}, \\ NIE(z,t,V) &= \log \frac{\int h(z,m_1,V;\boldsymbol{\beta}) dF_{M(1)}(m_1|V;\boldsymbol{\alpha},\boldsymbol{\theta})}{\int h(z,m_0,V;\boldsymbol{\beta}) dF_{M(0)}(m_0|V;\boldsymbol{\alpha},\boldsymbol{\theta})}. \end{aligned}$$

Under the simple model $g_z(z,V_i;\boldsymbol{\alpha}) = (z,V_i)\boldsymbol{\alpha}$ and assuming $\epsilon_i(z)$ does not depend on z, we have

$$\begin{aligned} NDE(z,t,V) &= \boldsymbol{\beta}_1, \\ NIE(z,t,V) &= \boldsymbol{\beta}_2\boldsymbol{\alpha}_1. \end{aligned}$$

This result is consistent with the approximation obtained from induced hazard for observed variables as shown in Huang and Prentice (2013). However, when there are interactions, the expression will be complicated. We assume $M(z)|V$ follows normal distribution with mean μ_z and variance σ_z^2 and $h(z,m,V) = \boldsymbol{\beta}_1 z + \boldsymbol{\beta}_2 m + \boldsymbol{\beta}_3 zm + \boldsymbol{\beta}_4 V$, then we have

$$\begin{aligned} NDE(z,t,V) &= \boldsymbol{\beta}_1 + \boldsymbol{\beta}_3\mu_z + \frac{1}{2}\sigma_z^2(\boldsymbol{\beta}_3^2 + 2\boldsymbol{\beta}_3\boldsymbol{\beta}_2) \\ NIE(z,t,V) &= (\boldsymbol{\beta}_2 + \boldsymbol{\beta}_3\mu_z)(\mu_1 - \mu_0) + \frac{1}{2}(\boldsymbol{\beta}_2 + \boldsymbol{\beta}_3 z)^2(\sigma_1^2 - \sigma_0^2). \end{aligned}$$

Under the assumption that $g_z(z,V_i;\boldsymbol{\alpha}) = (z,V_i)\boldsymbol{\alpha}$, and $\epsilon_i(z)$ does not depend on z, we have a simple product form for the natural indirect effect.

$$NIE(z,t,V) = (\boldsymbol{\beta}_2 + \boldsymbol{\beta}_3\mu_z)\boldsymbol{\alpha}_1.$$

Now we consider the case with $h(z,m,V;\boldsymbol{\beta}) = 1 + (z,m,V)\boldsymbol{\beta}$. We may consider the identity link G and the effect is given by the following formulas

$$\begin{aligned} CDE(m,t,V) &= \Lambda(t)[h(1,m,V;\boldsymbol{\beta}) - h(0,m,V;\boldsymbol{\beta})] \\ CME(m_1,m_2,z,t,V) &= \Lambda(t)[h(z,m_2,V;\boldsymbol{\beta}) - h(z,m_1,V;\boldsymbol{\beta})]. \end{aligned}$$

The natural direct and indirect effects have the following forms:

$$\begin{aligned} NDE(z,t,V) &= \Lambda(t)\int [h(1,m,V;\boldsymbol{\beta}) - h(0,m,V;\boldsymbol{\beta})]dF_{M(z)}(m|V;\boldsymbol{\alpha},\boldsymbol{\theta}), \\ NIE(z,t,V) &= \Lambda(t)\int [h(z,m_2,V;\boldsymbol{\beta}) - h(z,m_1,V;\boldsymbol{\beta})]dF_{(M(1),M(0))}((m_1,m_0)|V;\boldsymbol{\alpha},\boldsymbol{\theta}). \end{aligned}$$

Under the simple model $g_z(z,V_i;\boldsymbol{\alpha}) = (z,V_i)\boldsymbol{\alpha}$ and assume $\epsilon_i(z)$ does not depend on z, we have

$$\begin{aligned} NDE(z,t) &= \Lambda(t)\boldsymbol{\beta}_1, \\ NIE(z,t) &= \Lambda(t)\boldsymbol{\beta}_2\boldsymbol{\alpha}_1. \end{aligned}$$

For the additive hazard model, we have

$$\begin{aligned} CDE(m,t,V) &= G(\int_0^t \{\lambda(u) + h(u,1,m,V;\boldsymbol{\beta})\}du) \\ &\quad -G(\int_0^t \{\lambda(u)h(u,0,m,V;\boldsymbol{\beta})\}du) \\ &= G(\Lambda(t) + \int_0^t h(u,1,m,V;\boldsymbol{\beta})du) - G(\Lambda(t) + \int_0^t h(u,0,m,V;\boldsymbol{\beta})du), \end{aligned}$$

and

$$\begin{aligned} CME(m_1,m_2,z,t,V) &= G(\int_0^t \{\lambda(u) + h(u,z,m_2,V;\boldsymbol{\beta})\}du) \\ &\quad -G(\int_0^t \{\lambda(u)h(u,z,m_1,V;\boldsymbol{\beta})\}du) \\ &= G(\Lambda(t) + \int_0^t h(u,z,m_2,V;\boldsymbol{\beta})du) - G(\Lambda(t) + \int_0^t h(u,z,m_1,V;\boldsymbol{\beta})du). \end{aligned}$$

We may use also consider using an identity link G, and the effect is reduced to the following formulas:

$$\begin{aligned} CDE(m,t,V) &= \int_0^t [h(u,1,m,V;\boldsymbol{\beta}) - h(u,0,m,V;\boldsymbol{\beta})]du, \\ CME(m_1,m_2,z,t,V) &= \int_0^t [h(u,z,m_2,V;\boldsymbol{\beta}) - h(u,z,m_1,V;\boldsymbol{\beta})]du. \end{aligned}$$

Under the simple model $g_z(z, V_i; \boldsymbol{\alpha}) = (z, V_i)\boldsymbol{\alpha}$, $h(z, m, V; \boldsymbol{\beta}) = (z, m, V)\boldsymbol{\beta}$ and assume ϵ_i^z does not depend on z, we have

$$\begin{aligned} NDE(z, t, V) &= \boldsymbol{\beta}_1 t, \\ NIE(z, t, V) &= \boldsymbol{\beta}_2 \boldsymbol{\alpha}_1 t. \end{aligned}$$

If we would like to compute the whole population effect, then we need to integrate over V. The interaction between V and other variables will cause the form become complicated. However, if no interaction exist between V and other variables, then we have $NDE(z, t) = NDE(z, t, V)$, $NIE(z, t) = NIE(z, t, V)$.

2.2.4 Asymptotic Results

Here we list some regularity conditions required for the Asymptotic Results. For simplicity, we denote $\boldsymbol{\Theta} = (\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\theta})$ and denote its true value as $\boldsymbol{\Theta}_0$. The parameter $\boldsymbol{\Theta}$ has finite dimension.

- (A) The true parameter $\boldsymbol{\Theta}_0$ is an interior point of the possible space of $\boldsymbol{\Theta}$ and the space is compact.

- (B) The model parameter is identifiable, i.e. if the joint distribution of $(\underline{\tilde{Q}(t)}, \underline{\tilde{W}(t)}, Z, V)$ is the same for two parameters $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$, then we have $\boldsymbol{\Theta}_1 = \boldsymbol{\Theta}_2$. This assumption need to be verified case by case. For many commonly used models, the assumption holds.

- (C) There exist measurable functions μ_0, μ_1, μ_2, μ_3, μ_4 such that

$$\begin{aligned} &\lim_n \sup_{\boldsymbol{\beta}} n^{-1} \sum_i (h(t, Z_i, \hat{M}_i(\boldsymbol{\Theta}_0), V_i; \boldsymbol{\beta}) - \mu_0)^2 \to 0 \\ &\lim_n \sup_{\boldsymbol{\beta}} n^{-1} \sum_i (\frac{\partial h(t, Z_i, \hat{M}_i(\boldsymbol{\Theta}_0), V_i; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} - \mu_1)^2 \to 0 \\ &\lim_n \sup_{\boldsymbol{\beta}} n^{-1} \sum_i (\frac{\partial^2 h(t, Z_i, \hat{M}_i(\boldsymbol{\Theta}_0), V_i; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} - \mu_2)^2 \to 0 \end{aligned}$$

$$\lim_n \sup_{\boldsymbol{\beta}} n^{-1} \sum_i (\frac{\partial \log h(t, Z_i, \hat{M}_i(\boldsymbol{\Theta}_0), V_i; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} - \mu_3)^2 \to 0$$

$$\lim_n \sup_{\boldsymbol{\beta}} n^{-1} \sum_i (\frac{\partial^2 \log h(t, Z_i, \hat{M}_i(\boldsymbol{\Theta}_0), V_i; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} - \mu_4)^2 \to 0.$$

Theorem 2.3.1: When Z and V are discrete, under assumptions (A) and (B), we have

$$\sqrt{n}\left(\hat{\boldsymbol{\Theta}}_{MM} - \boldsymbol{\Theta}_0\right) \to_d N(\mathbf{0}, \boldsymbol{V}_{MM})$$

Proof: Since $(\underline{\tilde{Q}(t)}, \underline{\tilde{W}(t)}|Z, V)$ is jointly normal distributed, its fourth order moment exist. Thus we have $\mu_n(z, v)$ is regular asymptotic linear (RAL) estimator for $\mu(z, v)$. Similar, $\Sigma_n(z, v)$ is RAL estimator for $\Sigma(z, v)$. The estimating equation for $\hat{\boldsymbol{\Theta}}_{MM}$ are functional forms of $\mu_n(z, v)$, $\Sigma_n(z, v)$. Under the identifiable assumption, the expected derivatives shall be either positive definite or negative definite. So $\hat{\boldsymbol{\Theta}}$ is RAL estimator for $\boldsymbol{\Theta}$.

Theorem 2.3.2: Under Assumption (A) and (B),

$$\sqrt{n}\left(\hat{\boldsymbol{\Theta}}_{ML1} - \boldsymbol{\Theta}_0\right) \to_d N(\mathbf{0}, \boldsymbol{V}_{ML1})$$

$$\sqrt{n}\left(\hat{\boldsymbol{\Theta}}_{ML} - \boldsymbol{\Theta}_0\right) \to_d N(\mathbf{0}, \boldsymbol{V}_{ML})$$

Proof: Assumption (B) ensure that the maximum likelihood is unique and the fisher information matrix is positive definite. Since the data follow a joint normal distribution, the likelihood has continuous second partial derivatives and the envelope function exists. So with Assumption (A), we know that the maximum likelihood is RAL for the first one and $\boldsymbol{V}_{ML1}$ is the inverse of fisher information matrix. The second one is similar due to missing at random assumption.

Theorem 2.3.3a: For the general relative risk model, under regularity conditions (A)-(C), $\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*) \to N(0,\Sigma_{\boldsymbol{\beta}})$. Here $\boldsymbol{\beta}^*$ is the solution for the following estimating equation

$$0 = E\int_0^\tau \left[\frac{\frac{\partial h(t,Z,\hat{M}(\boldsymbol{\Theta}_0),V;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}}{h(t,Z,\hat{M}(\boldsymbol{\Theta}_0),V;\boldsymbol{\beta})} - \frac{E\frac{\partial h(t,Z,\hat{M}(\boldsymbol{\Theta}_0),V;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}Y(t)}{Eh(t,Z,\hat{M}(\boldsymbol{\Theta}_0),V;\boldsymbol{\beta})Y(t)}\right] dN(t).$$

Proof: Under the assumption of locally concavity of the right hand side, we just need to show the original estimating equation is an RAL estimator for the right hand side in above equation. As Theorem 3.2 and 3.1 showed, we have

$$\sqrt{n}(\hat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}_0) = \sum_i \phi(X_i) + o_p(1),$$

where the detail form of ϕ depend on the estimator we use. Then for a function $G(X_i)$, we will have

$$\begin{aligned}
& n^{-1/2}\sum_i [G(X_i,\hat{\boldsymbol{\Theta}}) - EG(X_i,\boldsymbol{\Theta}_0)] \\
= \;& n^{-1/2}\sum_i [G(X_i,\hat{\boldsymbol{\Theta}}) - G(X_i,\boldsymbol{\Theta}_0)] + n^{-1/2}\sum_i [G(X_i,\boldsymbol{\Theta}_0) - EG(X,\boldsymbol{\Theta}_0)] \\
= \;& n^{-1/2}\sum_i \frac{\partial G(X_i,\boldsymbol{\Theta})}{\partial\boldsymbol{\Theta}}|_{\boldsymbol{\Theta}_0}(\hat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}_0) + o_p(1) + n^{-1/2}\sum_i [G(X_i,\boldsymbol{\Theta}_0) - EG(X,\boldsymbol{\Theta}_0)] \\
= \;& n^{-1}\sum_i \frac{\partial G(X_i,\boldsymbol{\Theta})}{\partial\boldsymbol{\Theta}}|_{\boldsymbol{\Theta}_0}(\sum_j \phi(X_j)) + o_p(1) + n^{-1/2}\sum_i [G(X_i,\boldsymbol{\Theta}_0) - EG(X,\boldsymbol{\Theta}_0)] \\
= \;& E\frac{\partial G(X_i,\boldsymbol{\Theta})}{\partial\boldsymbol{\Theta}}|_{\boldsymbol{\Theta}_0}(\sum_j \phi(X_j)) + o_p(1) + n^{-1/2}\sum_i [G(X_i,\boldsymbol{\Theta}_0) - EG(X,\boldsymbol{\Theta}_0)] \\
= \;& \sum_i [\phi(X_i)E\frac{\partial G(X_i,\boldsymbol{\Theta})}{\partial\boldsymbol{\Theta}}|_{\boldsymbol{\Theta}_0} + n^{1/2}[G(X_i,\boldsymbol{\Theta}_0) - EG(X,\boldsymbol{\Theta}_0)]] + o_p(1)
\end{aligned}$$

When $Var\frac{\partial G(X_i,\boldsymbol{\Theta})}{\partial\boldsymbol{\Theta}} < \infty$ and $VarG(X_i,\boldsymbol{\Theta}_0) < \infty$, we have $n^{-1}\sum_i G(X_i,\hat{\boldsymbol{\Theta}})$ is RAL for $EG(X_i,\boldsymbol{\Theta})$.

Now let

$$G_0(X,\boldsymbol{\Theta}) \;=\; \int_0^\tau \frac{\frac{\partial h(t,Z,\hat{M}(\boldsymbol{\Theta}),V;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}}{h(t,Z,\hat{M}(\boldsymbol{\Theta}),V;\boldsymbol{\beta})} dN(t)$$

$$\begin{aligned} G_1(X,\boldsymbol{\Theta}) &= \frac{\partial h(t,Z,\hat{M}(\boldsymbol{\Theta}_0),V;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} Y(t) \\ G_2(X,\boldsymbol{\Theta}) &= h(t,Z,\hat{M}(\boldsymbol{\Theta}_0),V;\boldsymbol{\beta})Y(t), \end{aligned}$$

then we know that for $k=0,1,2$, $n^{-1}\sum_i G_k(X_i,\hat{\boldsymbol{\Theta}})$ is RAL for $EG_k(X_i,\boldsymbol{\Theta})$. We would like to point out that the detailed form of G also depends on the choice of estimator $\hat{M}$. But now for simplicity, we just denote the influence functions by ϕ_k. Then we know that the score can be expressed as

$$\begin{aligned} & \sum_i \int_0^\tau \left[\frac{\frac{\partial h(t,Z_i,\hat{M}_i(\boldsymbol{\Theta}_0),V_i;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}}{h(t,Z_i,\hat{M}_i(\boldsymbol{\Theta}_0),V;\boldsymbol{\beta})} - \frac{\sum_j \frac{\partial h(t,Z_j,\hat{M}_j(\boldsymbol{\Theta}_0),V_j;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} Y_j(t)}{\sum_j h(t,Z_j,\hat{M}_j(\boldsymbol{\Theta}_0),V_j;\boldsymbol{\beta})Y_j(t)} \right] dN_i(t). \\ - \quad & E\int_0^\tau \left[\frac{\frac{\partial h(t,Z,\hat{M}(\boldsymbol{\Theta}_0),V;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}}{h(t,Z,\hat{M}(\boldsymbol{\Theta}_0),V;\boldsymbol{\beta})} - \frac{E\frac{\partial h(t,Z,\hat{M}(\boldsymbol{\Theta}_0),V;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} Y(t)}{Eh(t,Z,\hat{M}(\boldsymbol{\Theta}_0),V;\boldsymbol{\beta})Y(t)} \right] dN(t) \\ = \quad & \sum_i \{\phi_{0i} + \int_0^\tau \frac{\phi_{1i}}{EG_2} - \frac{EG_1\phi_{2i}}{EG_2^2} dN_i(t) + \int_0^\tau \frac{EG_1}{EG_2}[dN_i(t) - EdN_i(t)]\} + o_p(1) \\ = \quad & \sum_i \psi_i \end{aligned}$$

Theorem 2.3.3b: For the additive hazard model, under regularity conditions (A)-(C), $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \to N(0, \Sigma_{\boldsymbol{\beta}})$. Here $\boldsymbol{\beta}$ is the solution for the following estimating equation

$$0 = E\int_0^\tau [\frac{\partial h(t,Z,\hat{M}(\boldsymbol{\Theta}_0),V;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} - \frac{E\frac{\partial h(t,Z,\hat{M}(\boldsymbol{\Theta}_0),V;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} Y(t)}{EY(t)}][dN(t) - Y(t)h(t,Z,\hat{M}(\boldsymbol{\Theta}_0),V;\boldsymbol{\beta})dt]$$

Proof: The proof can be done follow the same logic as theorem 2.3.3a. The only difference is now we first have the RAL related to

$$\begin{aligned} G_0(X) &= \frac{\partial h(t,Z,\hat{M}(\boldsymbol{\Theta}_0),V;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \\ G_1(X) &= \partial h(t,Z,\hat{M}(\boldsymbol{\Theta}_0),V;\boldsymbol{\beta})Y(t) \\ G_2(X) &= Y(t) \\ G_3(X) &= Y(t)h(t,Z,\hat{M}(\boldsymbol{\Theta}_0),V;\boldsymbol{\beta}) \\ G_4(X) &= \frac{\partial h(t,Z,\hat{M}(\boldsymbol{\Theta}_0),V;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} Y(t)h(t,Z,\hat{M}(\boldsymbol{\Theta}_0),V;\boldsymbol{\beta}). \end{aligned}$$

Comment: Although $\boldsymbol{\beta}^*$ are the same for different choice of estimator $\hat{\boldsymbol{\Theta}}$, the asymptotic variance $\Sigma_{\boldsymbol{\beta}}$ can be different since the influence function does depend on the choice. Here we would like to comment that we can estimate the variance matrix by sandwich estimator $\hat{\Sigma}_{\boldsymbol{\beta}}$, however, it converges to $\Sigma^*_{\boldsymbol{\beta}}$ which is just approximately equal to $\Sigma_{\boldsymbol{\beta}}$. So in practice, we recommend use bootstrap method to estimate the variance $\Sigma_{\boldsymbol{\beta}}$.

Theorem 2.3.4a: For the general relative risk model, under regularity conditions, $\sqrt{n}(\hat{\Lambda}(t) - \Lambda^*(t)) \to GP(0, K)$, where GP means Gaussian process and K is the covariance process. Here

$$\Lambda^*(t) \quad = \quad \int_0^t \frac{EdN(s)}{Eh(s, Z, \hat{M}(\boldsymbol{\Theta}_0), V; \boldsymbol{\beta}^*))Y(s)},$$

or

$$\tilde{\Lambda}^*(t) \quad = \quad \int_0^t \frac{EdN(s)}{Eh(s, Z, M, V; \boldsymbol{\beta}^*))Y(s)},$$

We would like to comment that $\hat{M}(\boldsymbol{\Theta}_0)$ is just a functional form of Z, V, Q and does not involve estimated quantities. The second one is needed for higher order approximation.

Proof: It is obvious that $n^{-1}\sum_i N_i(s)$ is a RAL estimator for $EN(s)$ and is a tight process respect to s. So we just need to show that $n^{-1}\sum h(s, Z_i, \hat{M}_i(\hat{\boldsymbol{\Theta}}), V_i; \hat{\boldsymbol{\beta}})Y_i(s)$ is a RAL estimator for $Eh(s, Z, \hat{M}(\boldsymbol{\Theta}_0), V; \boldsymbol{\beta}^*))Y(s)$ and the process is tight. Notice

$$\begin{aligned} & n^{-1/2}\sum_i (h(s, Z_i, \hat{M}_i(\hat{\boldsymbol{\Theta}}), V_i; \hat{\boldsymbol{\beta}}) - h(s, Z_i, \hat{M}_i(\boldsymbol{\Theta}_0), V_i; \boldsymbol{\beta}^*))Y(s)) \\ = \quad & [n^{-1}\sum_i \frac{\partial h_i(s)}{\partial \boldsymbol{\beta}}]\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) + [n^{-1}\sum_i \frac{\partial h_i(s)}{\partial \boldsymbol{\Theta}}]\sqrt{n}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0) + o_p(1) \end{aligned}$$

So we just need the two partial derivatives to be bounded and tight in s, which holds for some commonly used h, such as exponential or linear. Replacing $h(s, z, \hat{M}(\boldsymbol{\Theta}), V; \boldsymbol{\beta})$ with $h(s, Z, M, V; \boldsymbol{\beta}^*)$ gives the requirement for second order approximation.

Theorem 2.3.4b: For the additive hazard model, under regularity conditions, $\sqrt{n}(\hat{\Lambda}(t) - \Lambda^*(t)) \to GP(0, K)$, where GP again means Gaussian process and K is the covariance pro-

cess. Here

$$\Lambda^*(t) \quad = \quad E\int_0^t \frac{dN(s) - h(t, Z, \hat{M}(\boldsymbol{\Theta}_0), V; \boldsymbol{\beta}^*))Y(s)ds}{EY(s)}$$

or

$$\tilde{\Lambda}^*(t) \quad = \quad E\int_0^t \frac{dN(s) - h(t, Z, M, V; \boldsymbol{\beta}^*))Y(s)ds}{EY(s)}$$

Proof: There is no problem with the expansion of $N(s)$ and $Y(s)$ and the expansion of $h(t, Z, \hat{M}(\boldsymbol{\Theta}_0), V; \boldsymbol{\beta}^*)$ and $h(t, Z, M, V; \boldsymbol{\beta}^*)$ follows the same proof as in theorem 2.3.4a.

Corollary 2.3.5: All controlled and natural effects consistent and asymptotically normal distributed to certain value. This results follows by the fact that all effects can be written as function of $\boldsymbol{\Theta}$. If the derivative of the functional form is invertable, then using the delta method, we obtain the asymptotically normality.

2.2.5 Simulation

To explore the finite sample property of our proposed estimator for model parameters and to examine whether rare disease is a suitable approximation in different situations, we performed several simulation studies. For simplicity, in this simulation, we assume there is only one baseline covariate V. We also assume the time points at which variables are measured are 1, 3, 6.

The data is generated from the model below:

$$M_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 V_i + \alpha_3 Z_i \times V_i + \epsilon_i,$$

where ϵ_i is normally distributed with variance $\sigma^2_{m0}(1 - Z_i) + \sigma^2_{m1} Z_i$ and is independent of other covariates. Here $\boldsymbol{\alpha} = (1.5, 1, -0.7, -0.5)$, $\sigma_m = (0.5, 1)$.

$$Q_i(t) = \gamma_0 + \gamma_1 Z_i + \gamma_2 V_i + \gamma_3 Z_i \times V_i + \gamma_4 M_i + \gamma_5 Z_i \times M_i + \varepsilon_i(t),$$

where $\varepsilon_i(t)$ is Gaussian process with variance process $K_q(t, s) = \sigma^2_q \rho_q^{|t-s|}$ and is independent

of other covariates. Here $\boldsymbol{\gamma} = (1, 0.8, 0.7, 0.5, 0.3, 0.2)$ and $\sigma_q = 1$, $\rho_q = 0$.

$$W_i(t) = M_i + \omega_i(t),$$

where $\omega_i(t)$ is Gaussian process with variance process $K_q(t,s) = \sigma_w^2 \rho_w^{|t-s|}$ and is independent to other covariates. Here $\sigma_w = 0.1$, $\rho_w = 0$. We assume the models for Cox model, linear intensity function model and additive hazard model as follows:

$$\begin{aligned}
\lambda(t|Z,M,V) &= \lambda_0(t)\exp\{\beta_1 Z + \beta_2 M + \beta_3 V\}, \\
\lambda(t|Z,M,V) &= \lambda_0(t)[1 + \beta_1 Z + \beta_2 M + \beta_3 V], \\
\lambda(t|Z,M,V) &= \lambda_0(t) + \beta_1 Z + \beta_2 M + \beta_3 V.
\end{aligned}$$

We consider two kinds of censoring mechanisms. The first uses a common censoring time $C = 7$ for all individuals. The other lets C be exponentially distributed with terminal censoring at $C = 9$. We choose the parameter to make the event rate be around 5%. For three models, we use $\boldsymbol{\beta} = (0.3, 0.1, 0.2)$, $\boldsymbol{\beta} = (0.3, 0.1, 0.2)$ and $\boldsymbol{\beta} = (0.003, 0.001, 0.002)$ respectively. For each parameter setting, we study the relative performance of the following methods.

- Naive: Ignore measurement error issue and treat $\bar{Q}_t$ as Z.
- True I: Use true parameter to estimate $\hat{M}^{(1)}$.
- True II: Use true parameter to estimate $\hat{M}^{(2)}$.
- True III: Use true parameter to estimate $\hat{M}^{(3)}$.
- True2 I: Use true parameter to estimate $\hat{M}^{(1)}$, $\hat{\Sigma}_M^{(1)}$.
- True2 II: Use true parameter to estimate $\hat{M}^{(2)}$, $\hat{\Sigma}_M^{(2)}$.
- True2 III: Use true parameter to estimate $\hat{M}^{(3)}$, $\hat{\Sigma}_M^{(3)}$.

- RC I: Use regression method to estimate $\hat{M}^{(1)}$.
- RC II: Use regression method to estimate $\hat{M}^{(2)}$.
- RC III: Use regression method to estimate $\hat{M}^{(3)}$.
- RC2 I: Use regression method to estimate $\hat{M}^{(1)}$, $\hat{\Sigma}_M^{(1)}$.
- RC2 II: Use regression method to estimate $\hat{M}^{(2)}$, $\hat{\Sigma}_M^{(2)}$.
- RC2 III: Use regression method to estimate $\hat{M}^{(3)}$, $\hat{\Sigma}_M^{(3)}$.
- MME I: Use generalized method of moment to estimate $\hat{M}^{(1)}$.
- MME II: Use generalized method of moment to estimate $\hat{M}^{(2)}$.
- MME III: Use generalized method of moment to estimate $\hat{M}^{(3)}$.
- MME2 I: Use generalized method of moment to estimate $\hat{M}^{(1)}$, $\hat{\Sigma}_M^{(1)}$.
- MME2 II: Use generalized method of moment to estimate $\hat{M}^{(2)}$, $\hat{\Sigma}_M^{(2)}$.
- MME2 III: Use generalized method of moment to estimate $\hat{M}^{(3)}$, $\hat{\Sigma}_M^{(3)}$.
- MLE I: Use maximum likelihood from biomarker subsample to estimate $\hat{M}^{(1)}$.
- MLE II: Use maximum likelihood from biomarker subsampleto estimate $\hat{M}^{(2)}$.
- MLE III: Use maximum likelihood from biomarker subsample to estimate $\hat{M}^{(3)}$.
- MLE2 I: Use maximum likelihood from biomarker subsample to estimate $\hat{M}^{(1)}$, $\hat{\Sigma}_M^{(1)}$.
- MLE2 II: Use maximum likelihood from biomarker subsample to estimate $\hat{M}^{(2)}$, $\hat{\Sigma}_M^{(2)}$.
- MLE2 III: Use maximum likelihood from biomarker subsample to estimate $\hat{M}^{(3)}$, $\hat{\Sigma}_M^{(3)}$.

	Censoring I				Censoring II			
	$\beta_1 = 0.3$		$\beta_2 = 0.1$		$\beta_1 = 0.3$		$\beta_2 = 0.1$	
Method	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$	Bias $\times 10^3$	SD$\times 10^3$
Naive	-46.9	85.7	-34.2	33.8	-39.9	90.6	-37.2	33.5
True I	2.1	75.6	-2.6	57.6	6.3	77.7	-6.5	56.9
True II	1.0	76.1	-0.8	63.0	4.1	79.6	-3.3	64.3
True III	2.9	70.4	-4.5	47.4	6.8	74.8	-7.9	47.1
RC I	2.3	75.3	-3.2	56.8	6.3	77.2	-6.7	56.2
RC II	1.9	75.9	-2.1	62.5	4.9	78.8	-4.3	63.2
RC III	2.8	70.1	-4.8	46.6	6.5	74.3	-7.8	46.3
MME I	1.0	76.1	-1.6	58.6	5.3	77.	-5.4	57.8
MME II	0.5	76.8	-0.4	63.6	3.5	80.1	-2.9	64.7
MME III	1.9	70.8	-3.6	48.0	5.7	74.7	-6.8	47.7
MLE I	1.6	75.8	-1.9	58.1	5.6	77.9	-5.8	57.4
MLE II	0.9	76.4	-0.7	63.3	3.8	80.0	-3.1	64.6
MLE III	2.4	70.5	-3.9	47.5	6.1	74.8	-7.1	47.2
FMLE I	2.3	75.8	-2.8	57.5	6.3	77.7	-6.6	56.8
FMLE II	2.0	76.3	-2.0	62.7	4.9	79.7	-4.4	63.8
FMLE III	2.8	70.5	-4.5	47.2	6.5	74.7	-7.7	46.9

Table 2.1: Simulation result comparing different simple methods for Cox model. Total sample size is 20000 with 1000 biomarker subsample. The event rate is about 5% and $\boldsymbol{\beta} = (0.3, 0.1, 0.2)$.

- FMLE I: Use maximum likelihood from the entire dataset to estimate $\hat{M}^{(1)}$.
- FMLE II: Use maximum likelihood from the entire dataset to estimate $\hat{M}^{(2)}$.
- FMLE III: Use maximum likelihood from the entire dataset to estimate $\hat{M}^{(3)}$.
- FMLE2 I: Use maximum likelihood from the entire dataset to estimate $\hat{M}^{(1)}$, $\hat{\Sigma}_M^{(1)}$.
- FMLE2 II: Use maximum likelihood from the entire dataset to estimate $\hat{M}^{(2)}$, $\hat{\Sigma}_M^{(2)}$.
- FMLE2 III: Use maximum likelihood from the entire dataset to estimate $\hat{M}^{(3)}$, $\hat{\Sigma}_M^{(3)}$.

The simulation results for Cox model without/with second order approximation is summarized in Table 2.1 and Table 2.2. The simulation results for linear intensity function model and additive hazard model are summarized in Table 2.3 and Table 2.4.

From Table 2.1-2.4, we notice that all calibration estimators have little bias compared to the naive method ignoring measurement error. Among different calibration methods, we

	Censoring I				Censoring II			
	$\beta_1 = 0.3$		$\beta_2 = 0.1$		$\beta_1 = 0.3$		$\beta_2 = 0.1$	
Method	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$
True2 I	3.8	75.5	-4.4	58.3	4.6	78.7	-6.5	57.1
True2 II	3.7	75.5	-4.8	57.6	1.6	81.4	-3.5	64.5
RC2 I	2.8	75.7	-3.5	59.1	4.6	78.2	-6.7	56.4
RC2 II	3.1	75.7	-3.7	58.7	2.3	80.5	-4.3	63.3
MME2 I	3.8	75.7	-4.6	58.1	3.5	78.9	-5.4	58.0
MME2 II	2.1	75.0	-2.8	62.5	0.9	81.9	-2.9	64.8
MLE2 I	2.5	75.3	-3.7	62.3	3.9	78.9	-5.7	57.6
MLE2 II	1.6	75.2	-2.5	62.8	1.3	81.8	-3.1	64.7
FMLE2 I	1.9	75.2	-2.7	62.8	4.6	78.8	-6.6	57.0
FMLE2 II	2.9	75.2	-3.8	62.3	2.4	81.4	-4.4	64.0

Table 2.2: Simulation result comparing different second order calibration methods for Cox model. Total sample size is 20000 with 1000 biomarker subsample. The event rate is about 5% and $\boldsymbol{\beta} = (0.3, 0.1, 0.2)$.

	Censoring I				Censoring II			
	$\beta_1 = 0.3$		$\beta_2 = 0.1$		$\beta_1 = 0.3$		$\beta_2 = 0.1$	
Method	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$
Naive	-67.5	109.2	-19.1	56.6	-66.4	114.7	-19.3	58.1
True I	2.4	97.4	-0.6	81.4	3.6	97.4	-1.2	80.5
True II	1.7	98.3	2.7	103.7	2.8	98.4	6.1	106.8
True III	3.8	94.1	-2.6	67.6	4.8	95.9	-3.1	48.9
RC I	2.3	97.2	-0.6	81.0	3.5	97.7	-0.7	80.6
RC II	2.1	98.3	1.3	101.7	2.9	99.0	6.1	105.8
RC III	3.7	94.1	-2.8	66.7	4.6	96.2	-2.9	68.4
MME I	1.2	98.0	1.6	83.7	2.4	97.8	1.1	82.9
MME II	1.1	99.1	3.6	104.6	1.9	99.4	8.0	108.9
MME III	3.0	94.7	-1.3	68.9	3.7	96.3	-1.6	70.1
MLE I	1.7	97.4	0.5	82.0	3.1	97.5	-0.1	80.9
MLE II	1.6	98.6	2.3	102.7	2.5	98.9	6.6	106.2
MLE III	3.7	94.3	-2.2	67.4	4.7	96.1	-2.6	68.6
FMLE I	1.9	97.3	-0.2	81.4	3.3	97.5	-0.8	80.4
FMLE II	1.9	98.2	1.0	101.3	2.9	98.8	5.4	105.3
FMLE III	3.7	94.3	-2.6	67.0	4.8	96.1	-3.1	68.3

Table 2.3: Simulation result comparing different simple methods for linear intensity function model. Total sample size is 20000 with 1000 biomarker subsample. The event rate is about 5% and $\boldsymbol{\beta} = (0.3, 0.1, 0.2)$.

	Censoring I				Censoring II			
	$\beta_1 = 0.003$		$\beta_2 = 0.001$		$\beta_1 = 0.003$		$\beta_2 = 0.001$	
Method	Bias$\times 10^4$	SD$\times 10^4$	Bias$\times 10^4$	SD$\times 10^4$	Bias$\times 10^4$	SD$\times 10^4$	Bias$\times 10^4$	SD$\times 10^4$
Naive	-4.6	7.0	-3.6	2.9	-4.5	7.3	-3.7	2.8
True I	-0.1	6.1	-0.1	5.2	0.0	6.5	-0.4	4.8
True II	-0.3	6.2	0.1	5.8	-0.2	6.9	-0.1	5.7
True III	0.1	5.6	-0.3	4.3	0.2	6.3	-0.6	4.2
RC I	-0.1	6.1	-0.2	5.2	0.0	6.5	-0.4	4.7
RC II	-0.3	6.2	0.1	5.8	-0.2	6.9	-0.2	5.7
RC III	0.1	5.6	-0.4	4.3	0.1	6.3	-0.6	4.1
MME I	-0.2	6.2	0.0	5.4	-0.1	6.5	-0.3	4.8
MME II	-0.4	6.3	0.3	5.9	-0.3	6.9	0.0	5.7
MME III	0.0	5.6	-0.2	4.4	0.1	6.3	-0.5	4.2
MLE I	-0.1	6.1	-0.1	5.2	0.0	6.5	-0.3	4.8
MLE II	-0.4	6.3	0.3	5.9	-0.2	6.9	-0.1	5.7
MLE III	0.1	5.6	-0.3	4.3	0.1	6.3	-0.5	4.2
FMLE I	-0.3	6.3	0.2	5.8	0.0	6.5	-0.4	4.8
FMLE II	-0.3	6.3	0.1	5.8	-0.1	6.9	-0.2	5.7
FMLE III	0.1	5.6	-0.3	4.3	0.2	6.3	-0.6	4.2

Table 2.4: Simulation result comparing different simple methods for additive hazard model. Total sample size is 20000 with 1000 biomarker subsample. The event rate is about 5% and $\boldsymbol{\beta} = (0.003, 0.001, 0.002)$.

notice that those conditional on current status only is the least efficient while conditional on the whole process is the most efficient. However, the efficiency loss is less than 10%. The conclusions are consistent for the two different censoring mechanisms. For the Cox regression model, comparing to the estimator using first order approximation, the estimator using second order approximation has slightly smaller bias and similar efficiency. The estimators for Cox model have smaller variance than corresponding linear intensity model estimators.

Then we explore the case when the observed time is randomly selected. For each individual, we randomly select three observe time with the length of interval uniformly distribution from 1 to 2. We compared the following simple estimators

- Naive: Ignore measurement error issue and treat $\bar{Q}(t)$ as Z.
- Regression Calibration(RC): Run regression of $E[W(t)|R, V, \bar{Q}_t]$ and assume the coefficients are the same over t and different measurement times.
- True Estimator (True):

	Censoring I				Censoring II			
Cox Model								
	$\beta_1 = 0.3$		$\beta_2 = 0.1$		$\beta_1 = 0.3$		$\beta_2 = 0.1$	
Method	Bias $\times 10^3$	SD $\times 10^3$	Bias $\times 10^3$	SD $\times 10^3$	Bias $\times 10^3$	SD $\times 10^3$	Bias $\times 10^3$	SD$\times 10^3$
Naive	-58.9	96.1	-28.2	38.1	-56.1	103.3	-30.2	38.9
RC	-3.5	76.8	2.0	52.9	-2.2	82.6	-0.8	54.4
True	-3.4	75.6	2.1	53.3	-2.1	81.5	-0.7	54.4
MLE	-3.9	76.7	2.6	53.1	-2.6	82.5	-0.2	54.5
Linear Intensity Function Model								
	$\beta_1 = 0.3$		$\beta_2 = 0.1$		$\beta_1 = 0.3$		$\beta_2 = 0.1$	
Method	Bias $\times 10^3$	SD $\times 10^3$	Bias $\times 10^3$	SD $\times 10^3$	Bias $\times 10^3$	SD $\times 10^3$	Bias $\times 10^3$	SD$\times 10^3$
Naive	-79.2	133.8	-8.9	63.5	-78.2	148.8	-12.7	69.6
RC	0.8	114.0	8.2	74.1	-0.9	116.3	3.3	79.2
True	1.2	112.1	8.5	74.6	-0.6	115.0	3.4	79.8
MLE	0.6	114.0	8.3	74.1	-1.1	116.4	3.6	79.2
Additive Hazard Model								
	$\beta_1 = 0.003$		$\beta_2 = 0.001$		$\beta_1 = 0.003$		$\beta_2 = 0.001$	
Method	Bias $\times 10^4$	SD $\times 10^4$	Bias $\times 10^4$	SD $\times 10^4$	Bias $\times 10^4$	SD $\times 10^4$	Bias $\times 10^4$	SD$\times 10^4$
Naive	-5.1	7.6	-3.2	3.3	-4.5	7.4	-3.7	2.6
RC	-0.1	5.6	-0.2	4.6	0.5	5.9	-0.6	4.0
True	0.0	6.0	-0.3	5.0	0.7	6.2	-0.9	4.2
MLE	0.1	0.6	-0.2	5.0	0.8	6.0	-0.9	4.4

Table 2.5: Simulation result comparing different simple methods for random measurement time. Total sample size is 20000 with 1000 biomarker subsample. The event rate is about 5%.

Using $\underline{\tilde{E}_j(t)} = E[Z_j|R_j, V_j, \underline{\tilde{Q}_j(\infty)}; \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\theta}]$

- Subsample Likelihood Estimator(MLE):
  Using $\underline{\tilde{E}_j(t)} = E[Z_j|R_j, V_j, \underline{\tilde{Q}_j(\infty)}; \hat{\boldsymbol{\alpha}}_{ML1}, \hat{\boldsymbol{\gamma}}_{ML1}, \hat{\boldsymbol{\theta}}_{ML1}]$

We summarize the results in Table 2.5. From the Table, we notice that all calibration method works well compared with the naive method. Different calibration regression method performs similarly, this is due to the fact that the model $(Q(t), W(t))|Z, M, V$ is time-independent.

2.2.6 Data Analysis: DM Intervention and Long Term Energy Consumption Effects

Now we apply our methods to the WHI DM trial data to evaluate whether the DM intervention decrease breast cancer risk through decreasing long term average energy intake. We consider the measurement at time 1, 3 and 6. Here we group the measurement from year

	CDE			CME			NIE		
Method	HR	95%CI		HR	95%CI		HR	95%CI	
Naive	0.95	0.83	1.07	0.99	0.96	1.03	1.00	1.00	1.01
RC I	0.96	0.84	1.10	0.96	0.86	1.07	1.01	0.98	1.03
RC II	0.96	0.84	1.10	0.97	0.86	1.09	1.01	0.98	1.03
RC III	1.09	0.84	1.10	0.89	0.28	2.82	1.03	0.87	1.22
MLE I	0.95	0.84	1.08	0.98	0.93	1.04	1.00	0.99	1.01
MLE II	0.95	0.84	1.08	0.98	0.93	1.04	1.00	0.99	1.01
MLE III	0.89	0.78	1.01	0.80	0.71	0.91	1.04	1.02	1.06

Table 2.6: Controlled direct effect (CDE), controlled mediator effect (CME) and natural indirect effect (NIE) of mediation analysis of DM intervention on breast cancer risk through long term average total energy consumption.

2-4 as year 3 and year 5-7 as year 6. We only use subjects without missing values and we use inverse probability weighting to handle the missingness. The model is the same as that for simulation except that the baseline covariate V now is multi-dimensional. Due to the limited sample size of biomarker subsample, we did not allow general interaction between V and Z or M. Since V includes continuous variables, we did not use moment based estimator. We included age, race, BMI, income as covariate in our model. We use the Cox model for analysis. We use 100 Bootstrap to compute the standard error for our estimates in order to build 95% confidence interval (CI). The results are shown in Table 2.6.

From the results, we find that the type III regression calibration yield strange results, which might due to the fact that the true dietary intake shall not be considered as constant across the three time points and including later time point dietary data will thus yield a biased result. For results from other methods, we did not find any significant signal. We found that long-term average energy do not appear to explain the DM intervention effect on breast cancer risk and we did not find evidence that long-term average energy is causally related to breast cancer risk. This might be due to the limited sample size after removing all missing data (n=17948).

2.3 Time-independent exposure and short period error-prone mediator

For short period mediator, we consider time points $t_1, \cdots, t_L$. We use the method of multiple mediators to handle short period mediators. $\underline{\tilde{M}_S(t)} = (M(t_1), \cdots, M(t_{f_S(t)}))$, where $f_S(t) =$

$max_{i=0,\cdot,L}(t_i <= t)$. $\tilde{M}_S(t) = M(t_{f_S(t)})$. We denote $\boldsymbol{m} = (m(t_1), \cdots, m(t_{f_S(t)}))$, $\underline{\tilde{m}(t)} = (m(t_1), \cdots, m(t_{f_S(t)}))$.

2.3.1 Model

We assume the following model structures:

$$
\begin{aligned}
M_i(z,t) &= g_z(z, V_i; \boldsymbol{\alpha}) + \epsilon_i(z,t), \\
Q_i(t) &= g_{q0}(t, Z_i, V_i; \boldsymbol{\gamma}) + g_{q1}(t, Z_i, V_i; \boldsymbol{\gamma}) M_i(t) + \varepsilon_i(t), \\
W_i(t) &= M_i(t) + \omega_i(t).
\end{aligned}
$$

Here $\epsilon_i(z,t)$ is Gaussian process with variance $K_z(t,s) = K(t,s;\boldsymbol{\theta}_z, z, V_i)$ and is independent of other covariates, $\varepsilon_i(t)$ is Gaussian process with variance process $K_q(t,s) = K(t,s;\boldsymbol{\theta}_q, Z_i, V_i)$ and is independent of other covariates, and $\omega_i(t)$ is Gaussian process with variance process $K_w(t,s) = K(t,s;\boldsymbol{\theta}_w, Z_i, V_i)$ and is independent of other covariates. All gs are known link functions.

Same as previous section, we have two different types of survival models: general relative risk model,

$$\lambda(z, \boldsymbol{m}, t|V) = \lambda_0(t) h(t, z, \underline{m(t)}, V; \boldsymbol{\beta}),$$

and additive hazard model,

$$\lambda(z, \boldsymbol{m}, t|V) = \lambda_0(t) + h(t, z, \underline{m(t)}, V; \boldsymbol{\beta}),$$

as special cases of interest.

Under SUTVA, consistency, positivity and sequential ignorability assumption, i.e

$$
\begin{aligned}
Z &\perp (\boldsymbol{M}^z, T^{z\boldsymbol{m}}, C^{z\boldsymbol{m}})|V, \\
\boldsymbol{M} &\perp (T^{z\boldsymbol{m}}, C^{z\boldsymbol{m}})|Z, V,
\end{aligned}
$$

we have the observed model

$$M_i(t) = g_z(Z_i, V_i; \boldsymbol{\alpha}) + \epsilon_i(t), \tag{2.4}$$

where $\epsilon_i(t)$ is Gaussian process with variance $K_z(t,s) = K(t,s;\boldsymbol{\theta}_z, z, V_i)$ and

$$\lambda(t|Z, M\underline{\tilde{M}_S(t)}, V) = \lambda_0(t)h(t, Z, \underline{\tilde{M}_S(t)}, V; \boldsymbol{\beta}), \tag{2.5}$$

and additive hazard model,

$$\lambda(t|Z, \underline{\tilde{M}_S(t)}, V) = \lambda_0(t) + h(t, Z, \underline{\tilde{M}_S(t)}, V; \boldsymbol{\beta}). \tag{2.6}$$

2.3.2 Estimation of Model Parameters

We use the induced hazard method to estimate model parameter $\boldsymbol{\beta}$. Depending on whether we conditional on the whole history of self-reported mediator or just current self-reported mediator, we have two kinds of induced hazards. We denote $\hat{M}^{(1)}(t) = E[M(t)|Z, \underline{Q(t)}, V]$, $\hat{M}^{(2)}(t) = E[M(t)|Z, \tilde{Q}(t), V]$, $\hat{\Sigma}_M^{(1)}(t) = Cov[\underline{M(t)}|Z, \underline{Q(t)}, V]$, $\hat{\Sigma}_M^{(2)}(t) = Cov[\underline{M(t)}|Z, \tilde{Q}(t), V]$. Note the difference here compared to the previous section is the variance is replaced by the covariance matrix. The induced hazard can be written as

$$\begin{aligned} E[\lambda(t|Z, \underline{Q(t)}, V)] &= E[\lambda(t|Z, \underline{M(t)}, V)|Z, \underline{Q(t)}, V, T^* > t] \\ &\approx E[\lambda(t|Z, \underline{M(t)}, V)|Z, \underline{Q(t)}, V], \end{aligned}$$

which for model (2.5)

$$\approx \lambda_0(t)h(t, Z, \hat{M}^{(1)}(t), V; \boldsymbol{\beta})$$

or for model (2.6)

$$\approx \lambda_0(t) + h(t, Z, \hat{M}^{(1)}(t), V; \boldsymbol{\beta})$$

The first approximation require rare disease assumption and the second approximation is the first order approximation when h is in nonlinear form. When we have special form for $h(t, Z, \underline{M(t)}, V; \boldsymbol{\beta})$, e.g. $h(t, Z, \underline{M(t)}, V; \boldsymbol{\beta}) = h_0(t, Z, V; \boldsymbol{\beta}) \exp\{h_1(t, Z, V; \boldsymbol{\beta})\underline{M(t)}\}$, we might assume normality for $\underline{M(t)}|Z, Q, V$ to obtain second order approximation. The induced hazard for model (2.5) and (2.6) are in the following form respectively:

$$\lambda_0(t)h_0(t, Z, V; \boldsymbol{\beta}) \exp\{h_1(t, Z, V; \boldsymbol{\beta})\underline{\hat{M}^{(1)}(t)} + h_1(t, Z, V; \boldsymbol{\beta})\hat{\Sigma}_M^{(1)}(t)h_1(t, Z, V; \boldsymbol{\beta})^T/2\}$$

$$\lambda_0(t) + h_0(t, Z, V; \boldsymbol{\beta}) \exp\{h_1(t, Z, V; \boldsymbol{\beta})\underline{\hat{M}^{(1)}(t)} + h_1(t, Z, V; \boldsymbol{\beta})\hat{\Sigma}_M^{(1)}(t)h_1(t, Z, V; \boldsymbol{\beta})^T/2\}.$$

Replace $\hat{M}^{(1)}(t)$ and $\hat{\Sigma}_M^{(1)}(t)$ by $\hat{M}^{(2)}(t)$ and $\hat{\Sigma}_M^{(2)}(t)$ give another version of induced hazard. We can show the estimating equation are in similar form using these two versions though the martingale property is lost when $\hat{M}^{(2)}(t)$ and $\hat{\Sigma}_M^{(2)}(t)$ is used. However, we do need to account for the variation in estimation of $\hat{M}$ and $\hat{\Sigma}_M$, bootstrap method is recommended in practical and $\hat{M}^{(2)}(t)$, $\hat{\Sigma}_M^{(2)}(t)$ often more easier to compute and more stable, so we recommend use $\hat{M}^{(2)}(t)$ and $\hat{\Sigma}_M^{(2)}(t)$ in real applications.

Under general relative risk model, we can solve the partial likelihood score equation to estimate $\boldsymbol{\beta}$. The estimating equation can be written as follows:

$$\int_0^\infty \sum_{i=1}^n [\frac{\frac{\partial h_i(t)}{\partial \boldsymbol{\beta}}}{h_i(t)} - \frac{\sum_j \frac{\partial h_j(t)}{\partial \boldsymbol{\beta}} Y_j(t)}{\sum_j h_j(t) Y_j(t)}] dN_i(t),$$

where $h(t)$ can be either the first order or second order approximation with either version of $\underline{\hat{M}}(t)$ and $\hat{\Sigma}_M(t)$.

Under an additive hazard model, we can solve the moment based estimating equation

$$\int_0^\infty \sum_{i=1}^n [\frac{\partial h_i(t)}{\partial \boldsymbol{\beta}} - \frac{\sum_j \frac{\partial h_j(t)}{\partial \boldsymbol{\beta}} Y_j(t)}{\sum_j Y_j(t)}][dN_i(t) - Y_i(t)h_i(t)dt]$$

When $M(t)$ is observed, the baseline hazard estimator shall be in the forms as follows:

$$\hat{\Lambda}_0(t) = \int_0^t \frac{\sum_i dN_i(s)}{\sum_i h_i(t; \hat{\boldsymbol{\beta}}) Y_i(s)},$$

or

$$\hat{\Lambda}_0(t) \quad = \quad \int_0^t \frac{\sum_i dN_i(s) - h_i(t;\hat{\boldsymbol{\beta}})Y_i(s)ds}{\sum_i Y_i(s)}.$$

First order approximation is directly replace $M(t)$ by $\underline{\hat{M}}(t)$, however, unlike the estimator for $\boldsymbol{\beta}$, such first order approximation often lead to large bias in $\hat{\lambda}$. So we recommend second order approximation, i.e. assume $\underline{M}(t), \underline{Q}(t)|Z, V$ follow normal distribution and compute the expectations in the right hand side numerically by monte carlo method.

Now the remaining question is how to estimate parameters $\underline{\hat{M}}(t)$ and $\hat{\Sigma}_M$. We can specify a parametric model $E[M(t)|Z, V, \underline{\tilde{Q}(t)}; \boldsymbol{\delta}]$ or $E[M(t)|Z, V, \tilde{Q}(t); \boldsymbol{\delta}]$ and using

$$E[W(t)|Z, V, \underline{\tilde{Q}(t)}; \boldsymbol{\delta}] = E[M(t)|Z, V, \underline{\tilde{Q}(t)}; \boldsymbol{\delta}],$$
$$E[W(t)|Z, V, \tilde{Q}(t); \boldsymbol{\delta}] = E[M(t)|Z, V, \tilde{Q}(t); \boldsymbol{\delta}],$$

to obtain an estimator $\hat{\boldsymbol{\delta}}$ by running regression of $\underline{\tilde{W}(t)}$ on $Z, V, \underline{\tilde{Q}(t)}$. Then we obtain $\underline{\hat{E}_i(t)}$. Similarly we can assume a model for the covariance matrix and obtain $\Sigma_M^{(1)}(t)$ and $\Sigma_M^{(2)}(t)$ and estimate them by the relationship

$$Var[W(t)|Z, V, \underline{\tilde{Q}(t)}; \boldsymbol{\delta}] = Var[\underline{M(t)}|Z, V, \underline{\tilde{Q}(t)}; \boldsymbol{\delta}] + Var[\omega(t)],$$
$$Var[W(t)|Z, V, \tilde{Q}(t); \boldsymbol{\delta}] = Var[\underline{M(t)}|Z, V, \tilde{Q}(t); \boldsymbol{\delta}] + Var[\omega(t)],$$

where $Var[\omega(t)]$ can be estimated from the longitudinal measuring of $W(t)$. Depending on the parameter dimension in $\omega(t)$, two or more time points are needed.

As with methods for long-term average M, the method above supposes separate model for each time points, which is recommended when we have small number of time points and have complicated covariance structure for the $Q(t)$ and $W(t)$. However, if we have large number of time points and relative simple covariance structure with low dimension of parameters, we might borrow information across different time points to estimate $\underline{\hat{M(t)}}$ and $\hat{\Sigma}_M$. For this purpose, we propose either use approximated maximum likelihood estimator or use moment based estimator for parameters in the measurement error model.

Similar to the section with long-term average M, we can estimate $\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\theta}$ from either moment based estimator that minimizing $\|(\mu(Z,V,\boldsymbol{\alpha},\boldsymbol{\gamma}),\Sigma(Z,V,\boldsymbol{\gamma},\boldsymbol{\theta})),(\mu_n(Z,V),\Sigma_n(Z,V))\|$ under certain well-defined norm $\|\cdot\|$ or maximum likelihood based estiamtor by maximizing full likelihood or likelihood among biomarker subsample. However, now we should not condition on the whole process $Q(t)$ since it might contain future information on the mediator process when the calibration time point is smaller than t.

2.3.3 Estimation of Effects Related to Mediation Analysis

Unlike the observed mediator, the 'manipulated' mediator process is always external and is not affected by any other covariates due to the manipulation. So we can manipulate the future mediator but it should not affect the current cumulative hazard.

So for general relative risk model, we have

$$\begin{aligned} CDE(\boldsymbol{m},t) &= G(\int_0^t \{\lambda(u)h(u,1,\boldsymbol{m},V;\boldsymbol{\beta})\}du) \\ &\quad -G(\int_0^t \{\lambda(u)h(u,0,\boldsymbol{m},V;\boldsymbol{\beta})\}du), \end{aligned}$$

and

$$\begin{aligned} CME(\boldsymbol{m}_1,\boldsymbol{m}_2,z,t) &= G(\int_0^t \{\lambda(u)h(u,z,\boldsymbol{m}_2,V;\boldsymbol{\beta})\}du) \\ &\quad -G(\int_0^t \{\lambda(u)h(u,z,\boldsymbol{m}_1,V;\boldsymbol{\beta})\}du), \end{aligned}$$

The natural direct and indirect effects have the following forms:

$$\begin{aligned} NDE(z,t) &= \log \frac{\int h(1,\boldsymbol{m},V;\boldsymbol{\beta})dF_{M^z}(\boldsymbol{m}|V;\boldsymbol{\alpha},\boldsymbol{\theta})}{\int h(0,\boldsymbol{m},V;\boldsymbol{\beta})dF_{M^z}(\boldsymbol{m}|V;\boldsymbol{\alpha},\boldsymbol{\theta})}, \\ NIE(z,t) &= \log \frac{\int h(z,\boldsymbol{m}_1,V;\boldsymbol{\beta})dF_{(M^1,M^0)}((\boldsymbol{m}_1,\boldsymbol{m}_0)|V;\boldsymbol{\alpha},\boldsymbol{\theta})}{\int h(z,\boldsymbol{m}_0,V;\boldsymbol{\beta})dF_{(M^1,M^0)}((\boldsymbol{m}_1,\boldsymbol{m}_0)|V;\boldsymbol{\alpha},\boldsymbol{\theta})}. \end{aligned}$$

When $h(z,m,V;\boldsymbol{\beta}) = 1+(z,m,V)\boldsymbol{\beta}$, we may use identity link G and the effect is in the

following formula

$$\begin{aligned} CDE(\boldsymbol{m},t) &= \Lambda(t)[h(1,\boldsymbol{m},V;\boldsymbol{\beta})-h(0,\boldsymbol{m},V;\boldsymbol{\beta})] \\ CME(\boldsymbol{m}_1,\boldsymbol{m}_2,z,t) &= \Lambda(t)[h(z,\boldsymbol{m}_2,V;\boldsymbol{\beta})-h(z,\boldsymbol{m}_1,V;\boldsymbol{\beta})]. \end{aligned}$$

The natural direct and indirect effects have the following forms:

$$\begin{aligned} NDE(z,t) &= \Lambda(t)\int[h(1,\boldsymbol{m},V;\boldsymbol{\beta})-h(0,\boldsymbol{m},V;\boldsymbol{\beta})]dF_{M^z}(\boldsymbol{m}|V;\boldsymbol{\alpha},\boldsymbol{\theta}), \\ NIE(z,t) &= \Lambda(t)\int[h(z,\boldsymbol{m}_2,V;\boldsymbol{\beta})-h(z,\boldsymbol{m}_1,V;\boldsymbol{\beta})]dF_{(M^1,M^0)}((\boldsymbol{m}_1,\boldsymbol{m}_0)|V;\boldsymbol{\alpha},\boldsymbol{\theta}). \end{aligned}$$

For additive hazard model, we have

$$\begin{aligned} CDE(\boldsymbol{m},t) &= G(\int_0^t\{\lambda(u)+h(u,1,\boldsymbol{m},V;\boldsymbol{\beta})\}du) \\ &\quad -G(\int_0^t\{\lambda(u)h(u,0,\boldsymbol{m},V;\boldsymbol{\beta})\}du) \\ &= G(\Lambda(t)+\int_0^t h(u,1,\boldsymbol{m},V;\boldsymbol{\beta})du)-G(\Lambda(t)+\int_0^t h(u,0,\boldsymbol{m},V;\boldsymbol{\beta})du), \end{aligned}$$

and

$$\begin{aligned} CME(\boldsymbol{m}_1,\boldsymbol{m}_2,z,t) &= G(\int_0^t\{\lambda(u)+h(u,z,\boldsymbol{m}_2,V;\boldsymbol{\beta})\}du) \\ &\quad -G(\int_0^t\{\lambda(u)h(u,z,\boldsymbol{m}_1,V;\boldsymbol{\beta})\}du) \\ &= G(\Lambda(t)+\int_0^t h(u,z,\boldsymbol{m}_2,V;\boldsymbol{\beta})du)-G(\Lambda(t)+\int_0^t h(u,z,\boldsymbol{m}_1,V;\boldsymbol{\beta})du). \end{aligned}$$

We may use identity link G and the effect is in the following formula

$$\begin{aligned} CDE(\boldsymbol{m},t) &= \int_0^t[h(u,1,\boldsymbol{m},V;\boldsymbol{\beta})-h(u,0,\boldsymbol{m},V;\boldsymbol{\beta})]du, \\ CME(\boldsymbol{m}_1,\boldsymbol{m}_2,z,t) &= \int_0^t[h(u,z,\boldsymbol{m}_2,V;\boldsymbol{\beta})-h(u,z,\boldsymbol{m}_1,V;\boldsymbol{\beta})]du. \end{aligned}$$

Under the simple model $g_z(z,V_i;\boldsymbol{\alpha})=(z,V_i)\boldsymbol{\alpha}$, $h(z,\boldsymbol{m},V;\boldsymbol{\beta})=(z,\boldsymbol{m},V)\boldsymbol{\beta}(t)$ and assume

ϵ_i^z does not depend on z, we have

$$\begin{aligned} NDE(z,t) &= \int_0^t \boldsymbol{\beta}_1(s)ds, \\ NIE(z,t) &= \int_0^t \boldsymbol{\beta}_2(s)ds\boldsymbol{\alpha}_1. \end{aligned}$$

2.3.4 Asymptotic Results

The asymptotic results are in the similar form as in the previous section. So we just list them here without showing proof details.

Theorem 2.4.1: When Z and V is discrete, under the assumption that $\boldsymbol{\alpha}_0$, $\boldsymbol{\gamma}_0$ and $\boldsymbol{\theta}_0$ is interior point of possible range and the model is identifiable, we have

$$\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\alpha}}_{MM} - \boldsymbol{\alpha}_0 \\ \hat{\boldsymbol{\gamma}}_{MM} - \boldsymbol{\gamma}_0 \\ \hat{\boldsymbol{\theta}}_{MM} - \boldsymbol{\theta}_0 \end{pmatrix} \rightarrow_d N(\mathbf{0}, \boldsymbol{V}_{MM})$$

Theorem 2.4.2: Under regularity conditions,

$$\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\alpha}}_{ML1} - \boldsymbol{\alpha}_0 \\ \hat{\boldsymbol{\gamma}}_{ML1} - \boldsymbol{\gamma}_0 \\ \hat{\boldsymbol{\theta}}_{ML1} - \boldsymbol{\theta}_0 \end{pmatrix} \rightarrow_d N(\mathbf{0}, \boldsymbol{V}_{ML1})$$

$$\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\alpha}}_{ML} - \boldsymbol{\alpha}_0 \\ \hat{\boldsymbol{\gamma}}_{ML} - \boldsymbol{\gamma}_0 \\ \hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0 \end{pmatrix} \rightarrow_d N(\mathbf{0}, \boldsymbol{V}_{ML})$$

Theorem 2.4.3a: For the general relative risk model, under regularity conditions, $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \rightarrow N(0, \Sigma_{\boldsymbol{\beta}})$. Here $\boldsymbol{\beta}^*$ is the solution for the following estimating equation

$$0 = E\int_0^\tau \left[\frac{\frac{\partial h(t,Z,\hat{M}(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0),V;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}}{h(t,Z,\hat{M}(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0),V;\boldsymbol{\beta})} - \frac{E\frac{\partial h(t,Z,\hat{M}(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0),V;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}Y(t)}{Eh(t,Z,\hat{M}(\boldsymbol{\alpha}_0,\boldsymbol{\gamma}_0),V;\boldsymbol{\beta})Y(t)} \right] dN(t).$$

Theorem 2.4.3b: For the additive hazard model, under regularity conditions, $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \to N(0, \Sigma_{\boldsymbol{\beta}})$. Here $\boldsymbol{\beta}$ is the solution for the following estimating equation

$$0 = E\int_0^\tau [\frac{\partial h(t, Z, \hat{M}(\boldsymbol{\alpha}_0, \boldsymbol{\gamma}_0), V; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} - \frac{E\frac{\partial h(t, Z, \hat{M}(\boldsymbol{\alpha}_0, \boldsymbol{\gamma}_0), V; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} Y(t)}{EY(t)}][dN(t) - Y(t)h(t, Z, \hat{M}(\boldsymbol{\alpha}_0, \boldsymbol{\gamma}_0), V; \boldsymbol{\beta})dt]$$

Theorem 2.4.4a: For the general relative risk model, under regularity conditions, $\sqrt{n}(\hat{\Lambda}(t) - \Lambda^*(t)) \to GP(0, K)$, where GP means gaussian process and K is the covariance process. Here

$$\Lambda_0^*(t) = \int_0^t \frac{EdN(s)}{Eh(t, Z, \hat{M}(\boldsymbol{\alpha}_0, \boldsymbol{\gamma}_0), V; \boldsymbol{\beta}^*))Y(s)},$$

or

$$\tilde{\Lambda}_0^*(t) = \int_0^t \frac{EdN(s)}{Eh(t, Z, M, V; \boldsymbol{\beta}^*))Y(s)},$$

The second one need to use higher order approximation.

Theorem 2.4.4b: For the additive hazard model, under regularity conditions, $\sqrt{n}(\hat{\Lambda}(t) - \Lambda^*(t)) \to GP(0, K)$, where GP means gaussian process and K is the covariance process. Here

$$\Lambda_0^*(t) = E\int_0^t \frac{dN(s) - h(t, Z, \hat{M}(\boldsymbol{\alpha}_0, \boldsymbol{\gamma}_0), V; \boldsymbol{\beta}^*))Y(s)ds}{EY(s)}$$

or

$$\tilde{\Lambda}_0^*(t) = E\int_0^t \frac{dN(s) - h(t, Z, M, V; \boldsymbol{\beta}^*))Y(s)ds}{EY(s)}$$

Colloray 2.4.5: All controlled and natural effects consistent and asymptotically normal distributed to certain value.

The proofs of Theorem 2.4.1-2.4.5 are similar to those from 2.3.1-2.3.5 while changing the single dimensional of M to a finite dimensional $M(t_1), \cdots M(t_k)$.

2.3.5 Simulation

To explore the finite sample property of our proposed estimator for model parameters and whether rare disease assumption is suitable in different situations, we perform simulation studies. For simplicity, in this simulation, we assume there is only one baseline covariate V. We also assume the possible time to observe Q is 1, 3, 6 with time interval 1 to 3 the short period we are interested in. We assume W is measured at time1, 3, 6 within a randomly selected subsample.

The data is generated from model below:

$$M_i(t) = \alpha_0 + \alpha_1 t + \alpha_2 V_i + \alpha_3 Z_i + \alpha_4 Z_i \times t + \alpha_5 Z_i \times V_i + \epsilon_i(t),$$

where $\epsilon_i(t)$ is Gaussian process with variance process $K_m(t,s) = \sigma^2_{m0}\rho^{|t-s|}_{m0}(1 - Z_i) + \sigma^2_{m1}\rho^{|t-s|}_{m1}Z_i$ and is independent of other covariates. Here $\boldsymbol{\alpha} = (1.5, 0.2, 1, -0.1, -0.7, -0.5)$, $\sigma_m = (0.5, 1), \rho_m = (0.2, 0.3)$.

$$Q_i(t) = \gamma_0 + \gamma_1 Z_i + \gamma_2 V_i + \gamma_3 Z_i \times V_i + \gamma_4 M_i(t) + \gamma_5 Z_i \times M_i(t) + \varepsilon_i(t),$$

where $\varepsilon_i(t)$ is Gaussian process with variance process $K_q(t,s) = \sigma^2_q \rho^{|t-s|}_q$ and is independent of other covariates. Here $\boldsymbol{\gamma} = (1, 0.8, 0.7, 0.5, 0.3, 0.2)$, $\sigma_q = 1$ and $\rho_q = 0$.

$$W_i(t) = M_i(t) + \omega_i(t),$$

where $\omega_i(t)$ is Gaussian process with variance process $K_w(t,s) = \sigma^2_w \rho^{|t-s|}_w$ and is independent of other covariates. Here $\sigma_w = 0.1$ and $\rho_w = 0$.

$$\begin{aligned}
\lambda(t|Z, \underline{\tilde{M}(t)}, V) &= \lambda_0(t)\exp\{\beta_1 Z + \beta_3 V\} I(t < 1) \\
&\quad + \lambda_0(t)\exp\{\beta_1 Z + \beta_{21} M(1) + \beta_3 V\} I(1 \le t < 3) \\
&\quad + \lambda_0(t)\exp\{\beta_1 Z + \beta_{22} M(1) + \beta_{23} M(3) + \beta_3 V\} I(t \ge 3)
\end{aligned}$$

$$\lambda(t|Z, \underline{\tilde{M}(t)}, V) = \lambda_0(t)[1 + \beta_1 Z + \beta_3 V] I(t < 1)$$

$$+\lambda_0(t)[1+\beta_1 Z+\beta_{21}M(1)\beta_3 V]I(1\le t<3)$$
$$+\lambda_0(t)[1+\beta_1 Z+\beta_{22}M(1)+\beta_{23}M(3)+\beta_3 V]I(t\ge 3)$$

$$\lambda(t|Z,\underline{\tilde{M}(t)},V) \quad = \quad \lambda_0(t)+\beta_1 Z+[\beta_{21}I(1\le t<3)+\beta_{22}I(t\ge 3)]M(1)+\beta_{23}I(t\ge 3)M(3)+\beta_3 V.$$

Here we have $\boldsymbol{\beta}=(0.3,0.1,0.05,0.15,0.2)$, $\boldsymbol{\beta}=(0.3,0.1,0.05,0.15,0.2)$ and $\boldsymbol{\beta}=(0.003,0.002,0.001,0.0008,0.002)$ respectively for three models. For each parameter setting, we study the relative performance of the following methods.

- Naive: Ignore measurement error issue and treat $Q(t)$ as $Z(t)$.
- True I: Use true parameter to estimate $\hat{M}^{(1)}$.
- True2 I: Use true parameter to estimate $\hat{M}^{(1)}$, $\hat{\Sigma}_M^{(1)}$.
- RC I: Use regression method to estimate $\hat{M}^{(1)}$.
- RC2 I: Use regression method to estimate $\hat{M}^{(1)}$, $\hat{\Sigma}_M^{(1)}$.
- MLE I: Use maximum likelihood from biomarker subsample to estimate $\hat{M}^{(1)}$.
- MLE2 I: Use maximum likelihood from biomarker subsample to estimate $\hat{M}^{(1)}$, $\hat{\Sigma}_M^{(1)}$.
- FMLE I: Use maximum likelihood from all sample to estimate $\hat{M}^{(1)}$.
- FMLE2 I: Use maximum likelihood from all sample to estimate $\hat{M}^{(1)}$, $\hat{\Sigma}_M^{(1)}$.

We assume a constant censoring time at 7. For Cox model, we list the results in Table 2.7 and Table 2.8 for first order and second order calibration. The calibration estimation results are shown in Table 2.9 and Table 2.10.

From the result, we find that the Naive method is severely biased while the other three calibration methods have similar performance. The second order approximation did not improve upon the first order calibration in terms of bias and SD.

Method	$\beta_1 = 0.3$		$\beta_{21} = 0.1$		$\beta_{22} = 0.05$		$\beta_{23} = 0.15$	
	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$
Censoring type I								
Naive	69.0	62.0	-100.9	19.9	-50.1	36.7	-151.8	29.1
True I	-9.8	77.4	-10.0	78.4	-5.1	60.1	-6.0	73.4
RC I	-5.3	73.4	-10.1	82.8	-3.0	63.7	-5.4	75.1
MLE I	-5.1	73.9	-10.2	78.4	-3.6	63.7	-5.4	74.6
FMLE I	-5.1	73.6	-10.1	83.0	-3.2	63.5	-5.5	75.0
Censoring type II								
Naive	76.5	61.4	-102.1	19.1	-53.6	38.0	-149.7	30.5
True I	-5.7	73.1	-8.6	86.5	-4.2	54.3	-1.4	71.6
RC I	-0.9	68.9	-8.6	85.8	-2.2	56.3	-1.0	71.7
MLE I	-0.5	69.4	-9.0	86.3	-2.8	56.6	-0.9	71.6
FMLE I	-0.6	69.0	-8.8	86.3	-2.4	56.0	-1.0	71.8

Table 2.7: Simulation result comparing different simple methods for Cox model.

Method	$\beta_1 = 0.3$		$\beta_{21} = 0.1$		$\beta_{22} = 0.05$		$\beta_{23} = 0.15$	
	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$
Censoring type I								
True2 I	-13.5	78.8	-9.9	78.2	-5.6	60.5	-6.1	73.6
RC2 I	-9.4	74.8	-9.9	82.4	-3.5	64.0	-5.5	71.6
MLE2 I	-9.0	75.3	-10.0	82.6	-4.1	63.9	-5.5	74.8
FMLE2 I	-9.1	74.9	-9.9	82.5	-3.7	63.8	-5.5	71.6
Censoring type II								
True2 I	-9.7	73.9	-8.4	86.2	-4.7	54.6	-1.3	71.7
RC2 I	-5.1	69.7	-8.4	85.4	-2.5	56.4	-1.0	71.8
MLE2 I	-4.6	70.2	-8.7	85.9	-3.2	56.7	-0.9	71.7
FMLE2 I	-4.8	69.8	-8.6	85.9	-2.8	56.1	-1.0	71.9

Table 2.8: Simulation result comparing different second order calibration methods for Cox model.

Method	$\beta_1 = 0.3$		$\beta_{21} = 0.1$		$\beta_{22} = 0.05$		$\beta_{23} = 0.15$	
	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$
Censoring type I								
Naive	-21.7	85.3	-100.8	25.0	-46.6	43.2	-153.1	34.8
True I	-20.1	112.1	11.1	153.7	0.2	116.6	-4.5	144.7
RC I	-12.0	109.5	14.2	156.5	0.3	116.6	-4.8	147.2
MLE I	-12.5	108.7	12.4	154.2	-0.9	115.7	-5.5	145.4
FMLE I	-12.1	108.9	13.6	156.9	-0.3	116.0	-5.4	146.7
Censoring type II								
Naive	-26.6	82.2	-101.0	23.7	-51.1	43.1	-151.3	34.3
True I	-25.2	110.0	13.1	156.8	-9.6	104.6	6.7	156.3
RC I	-17.3	103.2	17.4	163.5	-3.5	112.8	7.6	159.5
MLE I	-17.4	103.1	14.3	157.5	-4.4	111.9	7.1	158.6
FMLE I	-17.3	102.9	16.5	163.8	-3.9	112.5	7.0	158.7

Table 2.9: Simulation result comparing different simple methods for linear intensity function model.

Method	$\beta_1 = 0.003$		$\beta_{21} = 0.002$		$\beta_{22} = 0.001$		$\beta_{23} = 0.0008$	
	Bias$\times 10^4$	SD$\times 10^4$	Bias$\times 10^4$	SD$\times 10^4$	Bias$\times 10^4$	SD$\times 10^4$	Bias$\times 10^4$	SD$\times 10^4$
Censoring type I								
Naive	-0.5	4.8	-20.2	1.7	-9.8	2.6	-8.3	2.2
True I	-9.2	5.8	-1.7	7.6	0.1	4.4	-1.8	6.5
RC I	-7.8	5.5	-2.2	7.8	-1.1	4.9	-1.1	6.6
MLE I	-7.8	5.6	-2.2	7.7	-1.3	4.6	-1.1	6.6
FMLE I	-7.8	5.6	-2.2	77.	-1.1	4.9	-1.1	6.7
Censoring type II								
Naive	-0.4	4.8	-20.0	1.7	-10.0	2.6	-8.1	2.3
True I	-8.9	6.0	-1.5	8.0	0.1	5.1	-1.8	6.7
RC I	-7.5	5.6	-2.0	8.1	-1.2	5.4	-1.3	7.1
MLE I	-7.5	5.7	-2.0	8.0	-1.2	5.4	-1.1	7.1
FMLE I	-7.5	5.7	-2.0	8.0	-1.2	5.3	-1.3	7.1

Table 2.10: Simulation result comparing different simple methods for additive hazard model.

	β_1		β_{21}		β_{22}		β_{23}	
Method	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Naive	0.96	(0.85,1.08)	1.00	(1.00,1.00)	1.00	(0.94,1.06)	1.00	(0.95,1.06)
RC I	0.95	(0.84,1.08)	1.06	(0.92,1.21)	0.98	(0.92,1.04)	NA	NA
MLE I	0.96	(0.84,1.08)	1.04	(0.91,1.20)	1.04	(0.94,1.15)	0.95	(0.87,1.04)

Table 2.11: Survival parameters for mediation analysis of DM intervention on breast cancer risk through long term average total energy consumption.

The result shows that for the coefficient of Z, there are some bias especially for additive hazard model. For those coefficients of M, the calibrated methods are less biased and have larger variance than Naive method. In general, it shows that calibration is needed to make correct inference.

2.3.6 Data Analysis: DM Intervention and Year 1 and Year 3 Energy Consumption Effects

Here we apply our method to the WHI DM trial study to evaluate whether the DM intervention decrease breast cancer risk through decreasing energy intake among first 3 years after intervention. We consider the measurement at time 1, 3 and 6. So we have $s = 2$, $k = 3$ $t_1 = 1$, $t_2 = 3$, $t_3 = 6$. Here we group the measurement from year 2-4 as year 3 and year 5-7 as year 6. We only use subjects without missing values and we use inverse probability weighting to handle the missingness. The model is the same as that for simulation except that the baseline covariate V now is multi-dimensional. Due to the limited sample size of biomarker subsample, we did not allow general interaction between V and Z or M. Since V include continuous variables, we did not use moment based estimator. We included age, race, BMI, income as covariate in our model. We use the Cox model for analysis. The results for survival parameters are shown in Table 2.11.

We would like to comment that the NA for regression calibration method is due to the fact that the calibrated values are exactly linearly correlated. From the result, we find that the direct effect are similar for different method while mediator effects are not obvious. However, we notice that the effects of first-year energy expenditure and of three-year energy expenditure have different signs, which might explains why an assumption of constant energy

expenditure for previous section did not work well. Given the limited sample size, there is no evidence that the energy expenditure at either 1 year or 3 years after randomization mediate DM intervention effects on breast cancer risk.

2.4 Time-independent exposure and time-varying error-prone mediator

2.4.1 Model

The model is the same except now all time point is in the model.

$$\lambda(t, z, \boldsymbol{m}|V) = \lambda_0(t)h(t, z, \underline{\tilde{m}(t)}, V; \boldsymbol{\beta}), \tag{2.7}$$

$$\lambda(t, z, \boldsymbol{m}|V) = \lambda_0(t) + h(t, z, \underline{\tilde{m}(t)}, V; \boldsymbol{\beta}). \tag{2.8}$$

Under sequential ignorability, the model for observed variables are

$$\lambda(t|Z, \underline{\tilde{M}(t)}, V) = \lambda_0(t)h(t, Z, \underline{\tilde{M}(t)}, V; \boldsymbol{\beta}), \tag{2.9}$$

$$\lambda(t|Z, \underline{\tilde{M}(t)}, V) = \lambda_0(t) + h(t, Z, \underline{\tilde{M}(t)}, V; \boldsymbol{\beta}). \tag{2.10}$$

2.4.2 Estimation of Parameters

When the measure time points are fixed, the results for short period mediator can directly applied here. So here we consider the measure time points are large enough and can be approximately viewed as continuously updated.

The likelihood based model parameters' estimation for the measurement error models are the same as previous section. The moment-based estimator and simple regression calibration method do not work in this context. Under the special case that $h(t, Z, \underline{\tilde{M(t)}}, V; \boldsymbol{\beta}) = h(t, Z, M(t), V; \boldsymbol{\beta})$, we can consider the following regression calibration method by regressing $W(t)$ on $Q(t)$, Z and V with some parametric coefficient $\boldsymbol{\delta}(t)$. The estimating equations are similar as in the previous section. After we obtain the estimator for $\boldsymbol{\beta}$, we can estimate the

controlled effects and natural effects. The asymptotics are similar as in previous section.

2.4.3 Simulation

To explore the finite sample property of our proposed estimator for model parameters and whether rare disease assumption is suitable in different situations, we perform simulation studies. For simplicity, in this simulation, we assume there is only one baseline covariate V and consider time points 0, 1, 3.

The data is generated from model below:

$$M_i(t) = \alpha_0 + \alpha_1 t + \alpha_2 V_i + \alpha_3 Z_i + \alpha_4 Z_i \times t + \alpha_5 Z_i \times V_i + \epsilon_i(t),$$

where $\epsilon_i(t)$ is Gaussian process with variance process $K_m(t,s) = \sigma^2_{m0}\rho^{|t-s|}_{m0}(1 - Z_i) + \sigma^2_{m1}\rho^{|t-s|}_{m1}Z_i$ and is independent of other covariates. Here $\boldsymbol{\alpha} = (1.5, 0.2, 1, -0.1, -0.7, -0.5)$, $\sigma_m = (0.5, 1)$ and $\rho_m = (0.2, 0.3)$.

$$Q_i(t) = \gamma_0 + \gamma_1 Z_i + \gamma_2 V_i + \gamma_3 Z_i \times V_i + \gamma_4 M_i(t) + \gamma_5 Z_i \times M_i(t) + \varepsilon_i(t),$$

where $\varepsilon_i(t)$ is Gaussian process with variance process $K_q(t,s) = \sigma^2_q \rho^{|t-s|}_q$ and is independent of other covariates. Here $\boldsymbol{\gamma} = (1, 0.8, 0.7, 0.5, 0.3, 0.2)$, $\sigma_q = 1$ and $\rho_q = 0$.

$$W_i(t) = M_i(t) + \omega_i(t),$$

where $\omega_i(t)$ is Gaussian process with variance process $K_w(t,s) = \sigma^2_w \rho^{|t-s|}_w$ and is independent of other covariates. Here $\sigma_w = 0.1$ and $\rho_w = 0$. We assume the models for Cox model, linear intensity function model and additive hazard model as follows:

$$\begin{aligned}
\lambda(t|Z, \underline{M(t)}, V) &= \lambda_0(t)\exp\{\beta_1 Z + \beta_2 M(t) + \beta_3 V\}, \\
\lambda(t|Z, \underline{M(t)}, V) &= \lambda_0(t)[1 + \beta_1 Z + \beta_2 M(t) + \beta_3 V], \\
\lambda(t|Z, \underline{M(t)}, V) &= \lambda_0(t) + \beta_1 Z + \beta_2 M(t) + \beta_3 V.
\end{aligned}$$

Here we have $\boldsymbol{\beta} = (0.3, 0.1, 0.2)$, $\boldsymbol{\beta} = (0.3, 0.1, 0.2)$ and $\boldsymbol{\beta} = (0.003, 0.002, 0.002)$ respectively

	Censoring I				Censoring II			
	$\beta_1 = 0.3$		$\beta_2 = 0.1$		$\beta_1 = 0.3$		$\beta_2 = 0.1$	
Method	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$
Naive	30.1	64.0	-99.2	14.7	24.7	62.3	-100.9	14.4
True I	-10.7	71.8	-2.3	63.7	-14.3	66.7	4.1	54.0
RC I	-10.8	71.5	-4.9	64.2	-14.7	66.5	4.5	57.5
MLE I	-10.5	71.6	-2.9	64.3	-14.4	66.6	1.9	54.4
FMLE I	-10.6	71.4	-2.7	64.2	-14.5	66.4	4.3	57.6
True II	-10.7	71.8	-2.3	63.7	-14.3	66.7	4.1	54.0
RC II	-10.8	71.5	-4.9	64.2	-14.7	66.5	4.5	57.5
MLE II	-10.5	71.6	-2.9	64.3	-14.4	66.6	1.9	54.4
FMLE II	-10.6	71.4	-2.7	64.2	-14.5	66.4	4.3	57.6

Table 2.12: Simulation result comparing different simple methods for Cox model

for three models. For computational consideration, we did not use full likelihood method and did not perform second order calibration We compare the relative performance of the following methods:

- Naive: Ignore measurement error issue and treat $Q(t)$ as $Z(t)$.
- True I: Use true parameter to estimate $\hat{M}^{(1)}$.
- True II: Use true parameter to estimate $\hat{M}^{(2)}$.
- RC I: Use regression method to estimate $\hat{M}^{(1)}$.
- RC II: Use regression method to estimate $\hat{M}^{(2)}$.
- MLE I: Use maximum likelihood from biomarker subsample to estimate $\hat{M}^{(1)}$.
- MLE II: Use maximum likelihood from biomarker subsample to estimate $\hat{M}^{(2)}$.
- FMLE I: Use full maximum likelihood from biomarker subsample to estimate $\hat{M}^{(1)}$.
- FMLE II: Use full maximum likelihood from biomarker subsample to estimate $\hat{M}^{(2)}$.

	Censoring I				Censoring II			
	$\beta_1 = 0.3$		$\beta_2 = 0.1$		$\beta_1 = 0.3$		$\beta_2 = 0.1$	
Method	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$
Naive	-25.4	88.7	-98.1	17.4	-32.6	89.5	-99.3	19.4
True I	-18.3	102.5	-3.0	93.1	-20.3	103.9	12.3	106.9
RC I	-18.2	103.0	-3.0	93.4	-20.6	103.5	13.3	108.0
MLE I	-18.5	102.4	-3.5	93.7	-20.5	103.6	12.4	107.5
FMLE I	-18.2	102.7	-3.2	93.9	-20.4	103.5	12.7	107.8
True II	-18.2	103.1	-3.1	93.3	-20.4	104.0	12.3	107.1
RC II	-18.2	103.2	-3.0	93.1	-20.6	103.6	413.4	108.1
MLE II	-18.3	102.7	-3.3	93.7	-20.4	103.8	12.5	107.7
FMLE II	-18.2	102.9	-3.3	93.6	-20.4	103.6	12.8	107.8

Table 2.13: Simulation result comparing different simple methods for linear intensity function model

	Censoring I				Censoring II			
	$\beta_1 = 0.003$		$\beta_2 = 0.002$		$\beta_1 = 0.003$		$\beta_2 = 0.002$	
Method	Bias$\times 10^4$	SD$\times 10^4$	Bias$\times 10^4$	SD$\times 10^4$	Bias$\times 10^4$	SD$\times 10^4$	Bias$\times 10^4$	SD$\times 10^4$
Naive	-2.7	6.0	-20.1	1.5	-4.1	6.1	-20.2	1.4
True I	-9.5	6.6	-3.0	7.0	-10.1	6.2	-3.2	6.6
RC I	-9.5	6.6	-2.9	7.0	-10.2	6.1	-3.1	6.6
MLE I	-9.5	6.6	-2.9	7.0	-10.2	6.2	-3.1	6.6
FMLE I	-9.5	6.6	-2.9	7.0	-10.2	6.1	-3.1	6.6
True II	-9.5	6.6	-3.0	7.0	-10.1	6.2	-3.2	6.6
RC II	-9.5	6.6	-2.9	7.0	-10.2	6.1	-3.1	6.6
MLE II	-9.5	6.6	-2.9	7.0	-10.2	6.2	-3.1	6.7
FMLE II	-9.5	6.6	-2.9	7.0	-10.2	6.2	-3.1	6.7

Table 2.14: Simulation result comparing different simple methods for additive hazard model

	CDE			CME		
Method	HR	95%CI		HR	95%CI	
Naive	0.95	0.87	1.05	1.00	1.00	1.00
RC I	1.00	0.99	1.01	1.00	0.97	1.02
MLE I	0.99	0.75	1.31	1.00	0.96	1.03

Table 2.15: Survival parameters for mediation analysis of DM intervention on breast cancer risk through long term average total energy consumption.

The result shows that ignoring measurement error tend to give estimate with large bias, especially for mediator effect. Different calibration methods perform similarly. They reduce the bias but have larger variance.

2.4.4 Data Analysis: DM Intervention and Time-Varying Energy Consumption Effects

Here we apply our method to the WHI DM trial study to evaluate whether the DM intervention decrease breast cancer risk through decreasing energy intake over time. We consider the measurement at time 1, 3 and 6. Here we group the measurement from year 2-4 as year 3 and year 5-7 as year 6. We only use subjects without missing values and we use inverse probability weighting to handle the missingness. The model is the same as that for simulation except that the baseline covariate V now is multi-dimensional. Due to the limited sample size of biomarker subsample, we did not allow general interaction between V and Z or M. Since V include continuous variables, we did not use moment based estimator. We included age, race, BMI, income as covariate in our model. We use the Cox model for analysis. The results are shown in Table 2.15.

The result suggests that there are no evidence that the pattern of energy consumption over time mediate the DM intervention effect on breast cancer risk. As we previously observed that the effect at different time point might have different signs, by assuming a constant effect over time, we might cancelled it out and this might explain the neutral result we find here.

2.5 Discussion

In the above methods, the parameters $\boldsymbol{\alpha}$ in model M is estimated only using information in validation sample if we use moment estimator or subsample likelihood estimator. Since these parameters are useful when computing effects related to mediation analysis. We suggest run a regression of $\underline{E_j(t)}$ on Z_j and V_j to update parameters $\boldsymbol{\alpha}$.

When the measure time is continuous and randomly selected. The moment based estimator and regression at each time point will not apply and the likelihood based method will be needed. We may need to use certain smooth technique if using regression method to estimate $\hat{M}$ and $\hat{\Sigma}_M$.

We would like to mention that although the V is required to be time-independent to obtain a causal interpretation, the estimation of $\boldsymbol{\beta}$ as described above works even if V is time-dependent. Also we can integrate over those post-intervention V values to obtain the marginal hazard or survival function.

For the cesnoring, we need to make sure that the true underlying mediator process does not affect the censoring process to yield valid inference. In terms of regression calibration method, we can allow the censoring to be dependent of exposure Z and covariate V. But the causal interpretation need the assumption that censoring is independent of event time conditional on V.

Chapter 3

MEDIATION ANALYSIS WITH MEASUREMENT ERROR IN EXPOSURE ASSESSMENT

In this chapter, we propose methods that can be used, for example, to adress the question of whether a high energy intake increases the risk of diabetes through increasing body mass index, with the energy consumption measurement subject to considerable measurement error.

3.1 Time varying exposure and mediator

Let $Z(t)$ denote the process of the predictor of interest of time t, $M(t)$ denote the mediator process, V denote the baseline covariate and $L(t)$ denote the time-varying confounder at cohort follow-up time t. Let $Q(t)$ denote a surrogate for $Z(t)$ that may have systematic bias and let $W(t)$ denote the biomarker estimate of $Z(t)$ that merely adds random error to $Z(t)$. The event time (e.g. disease incidence time) and censoring time is denoted as T and C, thus the observed composite endpoint is $T^* = T \wedge C$ and $\Delta = I(T \leq C)$. Using the counting process notation, we denote $Y(t) = I(T^* \geq t)$ and $N(t) = I(T \leq t, \Delta = 1)$. We denote $O(t) = \{Z(t), M(t), L(t), Y(t), V\}$. We assume a Markov model which means $O(t) \perp \{O(0), \cdots, O(t-2)\}|O(t-1)$. This assumption is rather strong, but it excludes the previous history of O as confounder between current status of O and the outcome, and thus the model we fit can be viewed as both a marginal and a conditional model. This property ensures that we do not need to use the inverse probability weighting approach to balance the previous histories of covariates when making a causal argument. We assume that the covariate process has a finite number of prespecified jump points, say $t_1, \cdot, t_K$. For simplicity, we take them to be $1, \cdots, K$.

3.1.1 Model I

When $L(t)$ is not affected by $Z(t)$ or $M(t)$, we can treat it as a type of interaction between the baseline covariate and t. We can then assume the following structural models without modeling $L(t)$

$$\begin{aligned}
Z(t) &= f_z(Z(t-1), M(t-1), L(t-1), Y(t)=1, V; \boldsymbol{\beta}_z(t)) + e_z(t) \\
M(t) &= f_m(Z(t), M(t-1), L(t), Y(t)=1, V; \boldsymbol{\beta}_m(t)) + e_m(t) \\
Q(t) &= f_{q0}(V, M(t-1), L(t-1), Y(t)=1; \boldsymbol{\beta}_q(t)) \\
&\quad + f_{q1}(V, M(t-1), L(t-1), Y(t-1)=1; \boldsymbol{\beta}_q(t)) Z(t) + e_q(t) \\
W(t) &= Z(t) + e_w(t) \\
\lambda(t) &= f_y(Z(t), M(t), L(t), V, Y(t-)=1, e_y(t); \boldsymbol{\theta}(t))
\end{aligned}$$

Here we assume $e_z(t)$, $e_m(t)$, $e_q(t)$, $e_w(t)$ and $e_y(t)$ to be independent processes with certain parametric distribution indexed by parameter $\boldsymbol{\sigma}_z$, $\boldsymbol{\sigma}_m$, $\boldsymbol{\sigma}_q$, $\boldsymbol{\sigma}_w$. For continuous variables, we typically assume a Gaussian process. For binary process, we typically assume a logistic distribution and the corresponding $\boldsymbol{\sigma}$ is the empty set or an over-dispersion parameter. However, we should not use this method for the case where $M(t)$ and $Q(t)$ is binary variables. In that case, a misclassification model rather than measurement error model should be used.

3.1.2 Model II

When we simultanteously model $L(t)$, we assume the following structural models

$$\begin{aligned}
Z(t) &= f_z(Z(t-1), M(t-1), L(t-1), Y(t)=1, V; \boldsymbol{\beta}_z(t)) + e_z(t) \\
L(t) &= f_l(Z(t), M(t-1), L(t-1), Y(t)=1, V; \boldsymbol{\beta}_l(t)) + e_l(t) \\
M(t) &= f_m(Z(t), M(t-1), L(t), Y(t)=1, V; \boldsymbol{\beta}_m(t)) + e_m(t) \\
Q(t) &= f_{q0}(V, M(t-1), L(t-1), Y(t)=1; \boldsymbol{\beta}_q(t)) \\
&\quad + f_{q1}(V, M(t-1), L(t-1), Y(t-1)=1; \boldsymbol{\beta}_q(t)) Z(t) + e_q(t) \\
W(t) &= Z(t) + e_w(t)
\end{aligned}$$

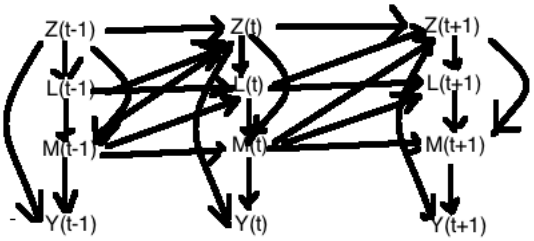
$$\lambda(t) \quad = \quad f_y(Z(t), M(t), L(t), V, Y(t-) = 1, e_y(t); \boldsymbol{\theta}(t))$$

Here we assume $e_z(t)$, $e_l(t)$, $e_m(t)$, $e_q(t)$, $e_w(t)$ and $e_y(t)$ to be independent processes with certain parametric distribution indexed by parameter $\boldsymbol{\sigma}_z$, $\boldsymbol{\sigma}_l$, $\boldsymbol{\sigma}_m$, $\boldsymbol{\sigma}_q$, $\boldsymbol{\sigma}_w$. For continuous variables, we typically assume a Gaussian process and for binary process, we typically assume a logistic distribution and the corresponding $\boldsymbol{\sigma}$ is the empty set or an over-dispersion parameter. The model for exposure $Z(t)$ assumes that current exposure only depends on whether the disease occurred up to t and previous exposure, mediator and covariate levels. The model for time-varying covariate $L(t)$ assumes that the current value of L only depends on whether the disease occurred yet, current exposure level and previous mediator and covariate levels. The model for time-varying covariate $L(t)$ assumes that the current value of L only depends on whether the disease occurred yet, current exposure level and previous mediator and covariate level. The model for time varying mediator $M(t)$ assumes that the current value of L only depends on whether the disease occurred yet, current exposure and covariate level and previous mediator level. The measurement error model for $Q(t)$ requires that the measurement error only depend on previous variables that happen before $Z(t)$ and the model for $W(t)$ is just a classical measurement error model. The last model for $Y(t)$ assume the hazard does not depend on the history of all variables given their current values. We might also write the hazard model as

$$\lambda(t) \quad = \quad f_y(Z(t), M(t), L(t), V; \boldsymbol{\theta}(t))$$

This means that for certain time t, the order of variables are $Z(t) \rightarrow L(t) \rightarrow M(t) \rightarrow Y(t)$ and Markov property for different time unit. These constraints are shown in Figure 3.1.

Figure 3.1: DAG for a three time points case.



3.2 Inference procedure

It is clear that the coefficient of $Z(t)$ can be estimated with traditional regression models if $Z(t)$ is observed. The pertinent estimating equations can be written as

$$
\begin{aligned}
0 &= E[S_l(Z(t), L(t), M(t-1), L(t-1), V, \boldsymbol{\beta}_l(t), \boldsymbol{\sigma}_l)|Y(t)=1], \\
0 &= E[S_m(Z(t), L(t), M(t-1), M(t), V, \boldsymbol{\beta}_m(t), \boldsymbol{\sigma}_m)|Y(t)=1], \\
0 &= E[S_y(Z(t), L(t), M(t), Y(t), dN(t), V, \boldsymbol{\theta}(t))|Y(t)=1],
\end{aligned}
$$

where S_l, S_m and S_y are corresponding unbiased estimating equations given by

$$
\begin{aligned}
S_l &= W_l(Z(t), M(t-1), L(t-1))[L(t) - f_l(Z(t), M(t-1), L(t-1), Y(t)=1, V; \boldsymbol{\beta}_l(t))], \\
S_m &= W_m(Z(t), M(t-1), L(t))[M(t) - f_m(Z(t), M(t-1), L(t), Y(t)=1, V; \boldsymbol{\beta}_l(t))], \\
S_y &= \int_0^\tau W_y(Z(t), M(t), L(t))[dN(t) - f_y(Z(t), M(t), L(t), V; \boldsymbol{\beta}_l(t))Y(t)dt].
\end{aligned}
$$

Here W_l, W_m and W_y are selected weight functions.

Regression calibration methods can be applied by replacing $Z(t)$ by its conditional mean given all other covariates except for the main outcome. ($Y(t)$ for survival model and $M(t)$, $L(t)$ for corresponding mediator and confounder process models) and all covariates that contribute the systematic bias in $Q(t)$. So we might use

$$
\hat{Z}_l(t) = E[Z(t)|Q(t), L(t-1), M(t-1), V, Y(t-)=1] \tag{3.1}
$$

$$\hat{Z}_m(t) = E[Z(t)|Q(t), L(t), L(t-1), M(t-1), V, Y(t-) = 1] \tag{3.2}$$

$$\hat{Z}_y(t) = E[Z(t)|Q(t), L(t), M(t), L(t-1), M(t-1), V, Y(t-) = 1] \tag{3.3}$$

to replace $Z(t)$ in the estimating equations above. Under a rare disease assumption, this can be further simplified to

$$\hat{Z}_l(t) = E[Z(t)|Q(t), L(t-1), M(t-1), V] \tag{3.4}$$

$$\hat{Z}_m(t) = E[Z(t)|Q(t), L(t), L(t-1), M(t-1), V] \tag{3.5}$$

$$\hat{Z}_y(t) = E[Z(t)|Q(t), L(t), M(t), L(t-1), M(t-1), V] \tag{3.6}$$

which can be estimated from a working model for $E[W(t)|Q(t), L(t-1), M(t-1), V]$, $E[W(t)|Q(t), L(t), L(t-1), M(t-1), V]$, $E[W(t)|Q(t), L(t), M(t), L(t-1), M(t-1), V]$ since $e_w(t)$ is independent of other variable processes. To combine the information from different time points, we can stack the data and assume independence working correlations. Mathematically, we have

$$\begin{aligned}
0 &= \sum_t P_n[S_l(\hat{Z}_l(t), L(t), M(t-1), L(t-1), V, \boldsymbol{\beta}_l(t), \boldsymbol{\sigma}_l)] \\
0 &= \sum_t P_n[S_m(\hat{Z}_m(t), L(t), M(t-1), M(t), V, \boldsymbol{\beta}_m(t), \boldsymbol{\sigma}_m)] \\
0 &= \sum_t P_n[S_y(\hat{Z}_y(t), L(t), M(t), Y(t), dN(t), V, \boldsymbol{\theta}(t))].
\end{aligned}$$

Here P_n represents the empirical mean, i.e. $P_n f(O) = n^{-1} \sum_{i=1}^n f(O_i)$.

3.2.1 Estimation of Effects Related to Mediation Analysis

When the effect of Z on M is not modified by any of the risk factor of Y, the indirect effect can be approximately written in a product form of the mediator effect and the effect of Z on M. The direct effect is approximately the sum of effect of Z on Y and the product of effect of Z on L and effect of Z on L. Specifically, we consider the following model

$$Z(t) \quad = \quad \beta_{z0} + \beta_{z1} Z(t-1) + \beta_{z2} M(t-1) + \beta_{z3} L(t-1) + \beta_{z4} V + e_z(t)$$

	Naive				Regression Calibration			
Parameter	Bias	SD	SESE	CR	Bias	SD	SESE	CR
$\theta_1 = 0.1$	0.022	0.027	0.035	88%	0.000	0.022	0.022	95%
$\theta_2 = 0.3$	0.010	0.028	0.029	93%	0.002	0.029	0.029	94%
$\beta_{m1} = 0.3$	0.075	0.003	0.075	0%	0.000	0.003	0.003	95%
Direct=0.2	0.022	0.027	0.035	88%	0.000	0.022	0.022	95%
Indirect=0.09	0.026	0.010	0.028	28%	0.000	0.009	0.009	94%

Table 3.1: Simulation result for Model I

$$
\begin{aligned}
L(t) &= \beta_{l0} + \beta_{l1}Z(t) + \beta_{l2}M(t-1) + \beta_{l3}L(t-1) + \beta_{l4}V + e_l(t) \\
M(t) &= \beta_{m0} + \beta_{m1}Z(t) + \beta_{m2}M(t-1) + \beta_{m3}L(t) + \beta_{m4}V + e_m(t) \\
Q(t) &= \beta_{q0} + \beta_{q1}Z(t) + \beta_{q2}M(t-1) + \beta_{q3}L(t-1) + \beta_{q4}V + e_q(t) \\
W(t) &= Z(t) + e_w(t) \\
\lambda(t) &= \lambda_0(t)\exp\{\theta_1 Z(t) + \theta_2 M(t) + \theta_3 L(t) + \theta_4 V\}.
\end{aligned}
$$

Then the mediator effect is θ_2, the direct effect is $\theta_1 + \beta_{l1}\theta_3$ and the indirect effect is $(\beta_{l1}\beta_{m3} + \beta_{m1})\theta_2$.

3.2.2 Simulation

Here, we evaluate the performance of regression calibration method in terms of estimating parameters and effects related to mediation analysis. We considered a two time point model. Our covariate is only measured at baseline, time 1 and time 2. The simulation results are shown in Table 3.1 and Table 3.2. The model parameters are specified as below:

$$\boldsymbol{\beta}_z = (0, 1, 0.1, 0.1, 0.1), \boldsymbol{\beta}_l = (1, \beta_{l1}, 0, 0, 0.5), \boldsymbol{\beta}_m = (0, 0.3, 1, 0.2, 0.1), \boldsymbol{\beta}_q = (0.1, 0.7, 0.8, 0.6, 0.3),$$

where $\beta_{l1} = 0$ for model I and $\beta_{l1} = 0.5$ for model II. The sample size is 50,000 with disease rate about 5%-10%.

From the results shown in Table 3.1 and Table 3.2, we find that the regression calibration provides an estimator that has little bias and correct coverage rate while ignoring

	Naive				Regression Calibration			
Parameter	Bias	SD	SESE	CR	Bias	SD	SESE	CR
$\theta_1 = 0.1$	0.013	0.031	0.034	93%	0.001	0.028	0.028	96%
$\theta_2 = 0.3$	0.009	0.027	0.029	93%	0.001	0.028	0.028	96%
$\theta_3 = 0.2$	0.011	0.027	0.029	94%	-0.001	0.028	0.028	95%
$\beta_{l1} = 0.5$	0.113	0.003	0.113	0%	0.000	0.005	0.005	95%
$\beta_{m1} = 0.3$	0.044	0.004	0.044	0%	0.000	0.004	0.004	95%
$\beta_{m3} = 0.2$	0.039	0.003	0.040	0%	0.000	0.005	0.005	95%
Direct=0.2	0.060	0.031	0.067	52%	0.001	0.028	0.028	95%
Indirect=0.12	0.032	0.013	0.034	34%	0.001	0.011	0.011	96%

Table 3.2: Simulation result for Model II

measurement error tend to have large bias.

3.2.3 Data Analysis:Time-Varying Diet and Time Varying BMI

We applied the method of this Chapter to the WHI study to evaluate how BMI may confound and mediate the association between total energy expenditure and risk of diabetes. Tinker *et al.* (2011) found that energy intake is significantly associated with the diabetes risk and suggests that BMI might be a dominating mediator. However, since that study only used energy intake and BMI at a certain fixed time point, it is impossible to fully differentiate the confounding, mediator and measurement error effect of BMI. Here, we use both baseline and year 3 data to separate these effects when analyzing OS data. For the DM comparison group, we have more data and can use a model with time point 0, 1, 3, 6. We use the log transformed total energy consumption from FFQ and double-labeled water as $Z(t)$ and $W(t)$. We consider BMI at follow-up time t as $M(t)$. Previous diabetes before year 0 are excluded from the analysis and dietary modifcation trial exclusion criteria are applied to OS cohort. to enhance comparability between the two cohorts The results are shown in Table 3.3.

From the result, we find that the short term BMI might serve as a confounder between the short term energy intake and disease risk. Also both the indirect effect and direct effect exist and are in similar magnitude which suggest the BMI is also serve as a mediator

Associated with	DM Comparison			OS		
20% Energy Change	HR	95% CI		HR	95% CI	
Uncalibrated						
Not adjust BMI	1.03	1.00	1.05	1.03	1.02	1.05
Total Effect	1.01	0.99	1.03	1.02	1.00	1.04
Direct Effect	1.01	0.98	1.03	1.01	0.99	1.03
Indirect Effect	1.00	0.99	1.01	1.00	1.00	1.01
Calibrated						
Not adjust BMI	2.65	2.14	3.29	2.63	2.15	3.20
Total Effect	1.17	0.79	1.74	1.34	0.94	1.90
Direct Effect	1.10	0.74	1.64	1.16	0.86	1.55
Indirect Effect	1.09	0.61	1.93	1.26	0.62	2.60

Table 3.3: WHI data analysis results

between the total energy consumption and risk of diabetes. However, due to the limited biomarker subsample, the variance is large and the results are not statistically significant. An important time-varying covariate is physical activity. However, it is more likely that the physical activity cause dietary change rather than the opposite way and also the physical activity measurement itself has measurement error and is correlated to that for energy consumption. So we will leave this question for next chapter.

3.3 Time-independent exposure

In general, the model considered in this section is a special case of section 3.1 with $Z(t) = Z$. When the exposure is time-independent, the mediator can be either time-dependent or not. An example for such Z is that we may consider Z as long term averaged dietary variable. The causal direction can be specified without loop if and only if we assume that $M(t)$ and $L(t)$ will not affect dietary variables. If we further assume that $M(t)$, $L(t)$ not cause systematic bias in $Q(t)$, then all the method derived in section 3.1 can be directly applied here. However, even now Z is time-independent, we cannot use regression calibration when $Q(t)$ have systematic bias respect to $M(t)$. Although the model $Y|M(t), L(t), Z, V$ can still be estimated by regression calibration method, the model $M(t)|L(t), Z, V$ and $L(t)|Z, V$ cannot be well estimated as shown below.

True association=10				
Adjust Baseline	Calibrate	Bias	SD	MSE
No	Use $M(t)$	2.75	1.01	8.59
No	Use $M(0)$	0.04	1.22	1.48
Yes	Use $M(t)$	3.05	1.11	10.54
Yes	Use $M(0)$	0.17	1.23	1.55
True association=0				
Adjust Baseline	Calibrate	Bias	SD	MSE
No	Use $M(t)$	1.15	0.24	1.38
No	Use $M(0)$	0.00	0.15	0.02
Yes	Use $M(t)$	1.17	0.24	1.43
Yes	Use $M(0)$	0.00	0.15	0.02

Table 3.4: Simulation when the true measurement error depend only on $M(0)$ from 1000 simulations.

Here we run a simple simulation to show that why the simple regression calibration method may not work well and a longitudinal method is required. Although we would like to obtain good estimations using the unobserved data, we should be careful when adding the consequence of that variable into our estimating equation. The simulation shows that adding such variables in the estimating equation might cause a large bias.

As shown in Table 3.4-3.5, if the true systematic bias depend on time-varying covariate, here the current BMI, then simple calibration on either baseline or current BMI will give a biased result.

Hence, we need either use the biomarker subsample only to estimate parameters in these two models or combine with the information from those only with $Q(t)$. If only biomarker subsample is used, then the method to estimate the model $M(t)|L(t), Z, V$ and $L(t)|Z, V$ follow exactly same formula as given in Huang and Prentice (2011). Specifically, we can estimate the covariance matrix for $(M(t), L(t), W(t), V)$ and then use the fact that $Cov(W(t), (M(t), L(t), V)) = Cov(Z, (M(t), L(t), V))$ and $Var(Z) = Var(W(t)) - Var(e_w(t))$ to obtain the covariance matrix for $(M(t), L(t), Z, V)$. Then the $M(t)|L(t), Z, V$ and $L(t)|Z, V$ followed by conditional normal distribution forms. Since we only have two replicates, to estimate the variance of the measurement error for biomarker, we need to

True association=10				
Adjust Baseline	Calibrate	Bias	SD	MSE
No	Use $M(t)$	-0.44	1.80	3.45
No	Use $M(0)$	0.46	2.74	7.72
Yes	Use $M(t)$	8.47	7.44	127.2
Yes	Use $M(0)$	1.28	3.35	12.86
True association=0				
Adjust Baseline	Calibrate	Bias	SD	MSE
No	Use $M(t)$	0.03	0.51	0.26
No	Use $M(0)$	0.00	0.29	0.09
Yes	Use $M(t)$	0.04	0.68	0.46
Yes	Use $M(0)$	0.00	0.30	0.09

Table 3.5: Simulation when the true measurement error depend only on $M(t)$ from 1000 simulations.

assume certain correlation among the two measurements.

In general, we did not recommend using this type of model since it is hard to argue that $M(t)$ will not affect later diet and by restricting Z to be time-independent, we artificially put some constraints on causal direction. Also the biomarker subsample tend to reflect the short term dietary patterns. Another view of the long term and short term diet is to consider $Z(t)$ represent the average diet among previous τ years. When τ is small, such as 1, it becomes the time-varying case and when τ is large enough, such as 10, it represents the time-independent case. However, we need to point out that when τ is moderate, such as 3, then the Markov assumption in section 3.1 might not hold and we might need to include a period of τ for process of $M(t)$ and $L(t)$ in both the risk model and measurement error model. Such analysis will require a relatively larger sample size for the biomarker subsample to estimate related coefficients and it will be necessary to modify the correlation between $W(t)$ to account for this.

Chapter 4

MEDIATION ANALYSIS WITH MEASUREMENT ERROR IN BOTH EXPOSURE AND MEDIATOR

In this chapter, we develop methods for data analysis when measurement error exists in both exposure and mediator. Motivation is provided by interest in understanding the relationship between total energy intake, physical activity (measured as physical activity related energy expenditure) and disease risk. It is hypothesized that on the one hand, a larger short term physical activity might cause larger short term energy intake and thus possibly higher disease risk while on the other hand, greater physical activity might have a preventive direct effect that leads to lower disease risk. The cancellation of these two kinds of effects might cause the marginal association between physical activity and disease risk to be weak, but a mediation analysis could bring out both direct and mediator effects.

4.1 Longitudinal exposure and mediators

4.1.1 Models

Similar to Chapter 4, we use $Z(t)$ to denote the exposure process, $M(t)$ to denote the potential mediator process, $L(t)$ to denote time-varying covariates and V to denote the baseline covariates. We use T and C to denote the time to disease and the censoring time. So the observed composite endpoint will be $T^* = min(T, C)$, $\Delta = I(T \leq C)$. Using traditional counting process notation, we have $N(t) = I(T^* \leq t, \Delta = 1)$, $Y(t) = I(T^* \geq t)$. We use $Q(t)$ and $X(t)$ to denote the self-reported exposure and self-reported mediator and we use $W_z(t)$ and $W_m(t)$ to denote the biomarker value for the exposure and mediator respectively. A rather general model can be written in the following form:

$$Z(t) = f_z(Z(t-1), M(t-1), L(t-1), Y(t), V, e_z(t); \boldsymbol{\beta}_z(t)) \quad (4.1)$$

$$L(t) = f_l(Z(t), M(t-1), L(t-1), Y(t), V, e_l(t); \boldsymbol{\beta}_l(t)) \quad (4.2)$$

$$M(t) = f_m(Z(t), M(t-1), L(t-1), Y(t), V, e_m(t); \boldsymbol{\beta}_m(t)) \tag{4.3}$$

$$\begin{pmatrix} Q(t) \\ X(t) \end{pmatrix} = \begin{pmatrix} f_{q0}(V, L(t-1), Y(t); \boldsymbol{\beta}_q(t)) \\ f_{x0}(V, L(t-1), Y(t); \boldsymbol{\beta}_x(t)) \end{pmatrix} \tag{4.4}$$

$$+F_{qx}(V, L(t-1), Y(t); \boldsymbol{\beta}_{qx}(t)) \begin{pmatrix} Z(t) \\ M(t) \end{pmatrix} + \begin{pmatrix} e_q(t) \\ e_x(t) \end{pmatrix} \tag{4.5}$$

$$\begin{pmatrix} W_z(t) \\ W_m(t) \end{pmatrix} = \begin{pmatrix} Z(t) \\ M(t) \end{pmatrix} + e_w(t) \tag{4.6}$$

$$dN(t) = f_y(Z(t), M(t), L(t), L(t-1), V, Y(t) = 1, e_y(t); \boldsymbol{\theta}(t)) \tag{4.7}$$

Here we assume that $e_z(t)$, $e_l(t)$, $e_m(t)$, ($e_q(t)$,$e_x(t)$), $e_w(t)$ and $e_y(t)$ are independent processes with certain parametric distribution indexed by parameter $\boldsymbol{\sigma}_z$, $\boldsymbol{\sigma}_l$, $\boldsymbol{\sigma}_m$, $\boldsymbol{\Sigma}_{qx}$, $\boldsymbol{\sigma}_w$. The model forms f_z, f_l,f_m, f_y, f_{q0}, f_{q1} and F_{qx} are pre-specified. The simplest special case would assume linear models.

Here the model for $Z(t)$ is a transition model for the exposure process which depicts how the current exposure is affected by previous exposure, mediator, time-varying covariates, baseline covariates for study subjects still at risk. The model for $L(t)$ is a transition model for the time-varying covariates process and we allow it to be affected by current exposure status and previous mediator status. Also, we allow it to depend on baseline covariate. The model for $M(t)$ is a transition model for the mediator process. It shares the same mathematical form as $L(t)$ and can be affected by previous time-varying covariates and current exposure status as well as baseline covariates. When F_{qx} is nonlinear while f_z, f_l, f_m, f_{q0} and f_{q1} are linear, we have a partial linear model for the self-reported data. Specifically, given any covariate V and $L(t-1)$, $Q(t)$ and $X(t)$ is in a linear form with respect to $Z(t)$ and $M(t)$. When F_{qx} is constant, and f_{q0} and f_{q1} are linear, which are frequently assumed working model, the conditional mean for $Z(t)$ and $M(t)$ given $Q(t)$, $X(t)$, V, $L(t-1)$ will follow a linear form. The partly linear model assumes that conditional on previous time-varying covariates and baseline covariates, the self-reported data is a linear transformation of the underlying variables of interest. The model for biomarker (4.6) assumes a classical measurement error model. The model (4.7) is certain type of survival model that models the

hazard function in relation to current exposure, mediator and both current and previous covariates. A directed acyclic graph representation of the hypothesized data generation mechanism is given in Figure 4.1.

Here we would like to comment that although the $L(t)$ and $M(t)$ have different interpretation in our model, the distinction between them are purely based on our scientific knowledge. Mathematically, $M(t)$ and $L(t)$ are symmetric and cannot be distinguished using the data listed above, even when the true values rather than the self-reported estimates are observed. Also, the model order for different variables and possible lag periods needs to be carefully specified in real applications.

4.1.2 Estimation

We will focus our estimation method on the Cox regression model with specific functional forms for $(f_z, f_l, f_m, f_{q0}, f_{x0}, F_{qx}, f_y)$. However, the regression calibration method can be extended to other response variable models. Specificaly, we consider the model

$$\begin{aligned}
Z(t) &= (Z(t-1), M(t-1), L(t-1), V)\boldsymbol{\beta}_z(t) + e_z(t) \\
L(t) &= (Z(t), M(t-1), L(t-1), V)\boldsymbol{\beta}_l(t) + e_l(t) \\
M(t) &= (Z(t), M(t-1), L(t-1), V)\boldsymbol{\beta}_m(t) + e_m(t) \\
\begin{pmatrix} Q(t) \\ X(t) \end{pmatrix} &= \begin{pmatrix} f_{q0}(V, L(t-1); \boldsymbol{\beta}_q(t)) \\ f_{x0}(V, L(t-1); \boldsymbol{\beta}_x(t)) \end{pmatrix} \\
&\quad + F_{qx}(V, L(t-1); \boldsymbol{\beta}_{qx}(t)) \begin{pmatrix} Z(t) \\ M(t) \end{pmatrix} \\
\begin{pmatrix} W_z(t) \\ W_m(t) \end{pmatrix} &= \begin{pmatrix} Z(t) \\ M(t) \end{pmatrix} + e_w(t) \\
\lambda(t|\underline{Z}(t), \underline{M}(t), \underline{L}(t), V) &= \lambda_0(t) exp\{(Z(t), M(t), L(t), L(t-1), V)\boldsymbol{\theta}(t)\}
\end{aligned}$$

where $\underline{Z}(t)$, $\underline{M}(t)$ and $\underline{L}(t)$ denote the history up to time t for these processes. Now we can consider the induced hazard model as in previous chapters. Although we can condition on the whole process of $Q(t)$ and $X(t)$, the previous results showed that doing so will not gain

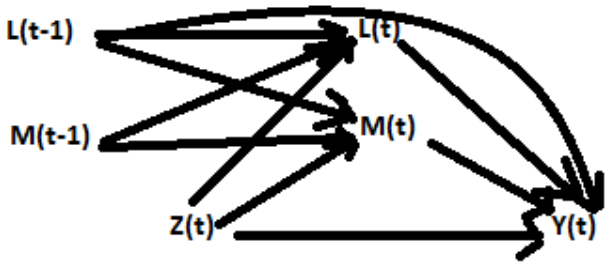


Figure 4.1: Our hypothesized DAG.

much efficiency and will lead to complex calibration models. Therefore, we only consider the calibration method that predict the calibrated exposure using current self-reported data. So we consider the following induced hazard model, under a rare disease assumption, as follows:

$$\begin{aligned} & \lambda(t|Q(t), X(t), \underline{L}(t), V) && (4.8)\\ = \; & E[\lambda_0(t) exp\{(Z(t), M(t), L(t), V)\boldsymbol{\theta}(t)\}|Q(t), X(t), \underline{L}(t), V, T \geq t] && (4.9)\\ \approx \; & E[\lambda_0(t) exp\{(Z(t), M(t), L(t), V)\boldsymbol{\theta}(t)\}|Q(t), X(t), \underline{L}(t), V] && (4.10)\\ \approx \; & \lambda_0(t) exp\{(\hat{Z}(t), \hat{M}(t), L(t), V)\boldsymbol{\theta}(t)\} && (4.11) \end{aligned}$$

where $\hat{Z}(t) = E[Z(t)|Q(t), X(t), L(t), L(t-1)]$ and $\hat{M}(t) = E[M(t)|Q(t), X(t), L(t), L(t-1)]$ This approximation can be relaxed to a second order approximation written as

$$\lambda_0(t) exp\{(\hat{Z}(t), \hat{M}(t), L(t), V)\boldsymbol{\theta}(t) + \boldsymbol{\theta}(t)^T \hat{\Sigma} \boldsymbol{\theta}(t)/2\}$$

where $\hat{\Sigma} = Var[(Z(t), M(t), L(t), V)|Z(t), M(t), L(t), L(t-1), V]$. This approximation becomes equality when a joint normality assumption of $Z(t), M(t)|Q(t), X(t), L(t), L(t-1)$ holds.

Under the rare disease assumption, approximately, we have

$$\begin{aligned} \begin{pmatrix} \hat{Z}(t) \\ \hat{M}(t) \end{pmatrix} &= E[\begin{pmatrix} W_z(t) \\ W_m(t) \end{pmatrix} |Q(t), X(t), L(t), L(t-1), V, Y(t) = 1] \\ &\approx E[\begin{pmatrix} W_z(t) \\ W_m(t) \end{pmatrix} |Q(t), X(t), L(t), L(t-1), V]. \end{aligned}$$

So we can first fit a multivariate regression model with independence working correlation assumption for

$$E[\begin{pmatrix} W_z(t) \\ W_m(t) \end{pmatrix} |Q(t), X(t), L(t), L(t-1), V]$$

The detailed modeling form will depend on the specification of $f_{q0}(\cdot)$, $f_{x0}(\cdot)$ and $F_{qx}(\cdot)$. When both functions are linear, a simple linear regression approximation is appropriate.

For second-order approximation, we can also fit a variance model to estimate

$$Var(\begin{pmatrix} W_z(t) \\ W_m(t) \end{pmatrix} | Q(t), X(t), L(t), L(t-1), V)$$

However, to obtain

$$Var(\begin{pmatrix} Z(t) \\ M(t) \end{pmatrix} | Q(t), X(t), L(t), L(t-1), V)$$

we need to know

$$Var(e_w(t)|Q(t), X(t), L(t), V) = Var(e_w(t))$$

which can be estimated only when we have biomarker data at multiple time points. If the variance matrix above only depends on $L(t-1)$ and V, then the Cox regression coefficients for $Z(t)$, $M(t)$ and $L(t)$ are approximately valid.

Then we can use the following estimating equations to first order approximation:

$$0 = \sum_i \int_0^\tau (A_i(t) - \bar{A}(t))^T dN_i(t)$$

where $A_i(t) = (\hat{Z}_i(t), \hat{M}_i(t), \hat{L}_i(t), \hat{L}_i(t-1), V_i)$, and

$$\bar{A}(t;\boldsymbol{\theta}) = \frac{\sum_j A_j(t) exp\{(\hat{Z}_j(t), \hat{M}_j(t), \hat{L}_j(t), \hat{L}_j(t-1), V_j)\boldsymbol{\theta}(t)\}Y_j(t)}{\sum_j exp\{(\hat{Z}_j(t), \hat{M}_j(t), \hat{L}_j(t), \hat{L}_j(t-1), V_j)\boldsymbol{\theta}(t)\}Y_j(t)}$$

For second order approximation, we can use the following estimating equation:

$$0 = \sum_i \int_0^\tau (B_i(t) + \hat{\Sigma}_i\boldsymbol{\theta}(t) - \bar{B}(t))^T dN(t)$$

where $B_i(t) = (\hat{Z}_i(t), \hat{M}_i(t), \hat{L}_i(t), \hat{L}_i(t-1), V)$ and

$$\bar{B}(t;\boldsymbol{\theta}) = \frac{\sum_j [B_j(t) + \hat{\Sigma}_j \boldsymbol{\theta}(t)] exp\{(\hat{Z}_j(t), \hat{M}_j(t), \hat{L}_j(t), \hat{L}_j(t-1), V_j)\boldsymbol{\theta}(t) + \boldsymbol{\theta}(t)^T \hat{\Sigma}_j \boldsymbol{\theta}(t)/2\} Y_j(t)}{\sum_j exp\{(\hat{Z}_j(t), \hat{M}_j(t), \hat{L}_j(t), \hat{L}_j(t-1), V_j)\boldsymbol{\theta}(t) + \boldsymbol{\theta}(t)^T \hat{\Sigma}_j \boldsymbol{\theta}(t)/2\} Y_j(t)}$$

To estimate the parameters in the model for $L(t)$, we can use

$$\begin{pmatrix} \hat{Z}_l(t) \\ \hat{M}_l(t) \end{pmatrix} = E[\begin{pmatrix} W_z(t) \\ W_m(t) \end{pmatrix} | Q(t), X(t), L(t-1), V].$$

and then run regression of $L(t)$ on $\hat{Z}_l(t)$, $\hat{M}(t)$, $L(t-1)$ and V. This is just the traditional regression calibration method for linear regression.

To estimate the parameters in the model of $M(t)$, we can first estimate the following variance matrix

$$Var(\begin{pmatrix} Z(t) \\ M(t) \end{pmatrix} | L(t-1), V).$$

from the variance matrix of

$$Var(\begin{pmatrix} W_z(t) \\ W_m(t) \end{pmatrix} | L(t-1), V),$$

if we know

$$Var(e_w(t)|L(t-1), V) = Var(e_w(t)).$$

Then with normal approximation, we can obtain parameters for the following conditional distribution

$$M(t)|Z(t), L(t-1), V$$

from

$$Var(\begin{pmatrix} Z(t) \\ M(t) \end{pmatrix} | L(t-1), V).$$

4.1.3 Simulation

Here, we perform simulations to examine moderate sample size properties of the regression calibration method under different data generation scenarios. For simplicity, we use a model with only one time point and without L. We will focus on the parameters in the survival model and would like to explore how the magnitude of measurement error and the correlation between Z and M affect the biases and variances of the estimator. We generate data from the following models

$$
\begin{aligned}
&V \sim N(0,1)\\
&Z \sim (1,V)\beta_z + N(0,\sigma_z^2)\\
&M \sim (1,Z,V)\beta_m + N(0,\sigma_m^2)\\
&Q \sim (1,Z,M,V)\beta_{qz} + N(0,\sigma_{qz}^2)\\
&X \sim (1,Z,M,V)\beta_{qm} + N(0,\sigma_{qm}^2)\\
&W_z \sim Z + N(0,\sigma_{wz}^2)\\
&W_m \sim M + N(0,\sigma_{wm}^2)\\
&\lambda(t) = \lambda_0(t)exp\{Z,M,V)\boldsymbol{\theta}\}
\end{aligned}
$$

The censoring is set to be at $t = 6$ and we have 10000 sample size, from which 500 are randomly selected as biomarker subsample. The disease rate is around 5-6%. We set the parameters as follows:

$$\beta_z = (0,1) \tag{4.12}$$

$$\sigma_z = \sigma_m = 0.5 \tag{4.13}$$

$$\sigma_{wz} = \sigma_{wm} = 0.1 \tag{4.14}$$

$$\boldsymbol{\theta} = c(0.3,-0.3,0.5) \tag{4.15}$$

$$\lambda_0(t) = 0.01. \tag{4.16}$$

We vary β_m to set different correlation between M and Z and vary σ_{qz} and σ_{qm} to

change the measurement error magnitude. We also use different β_{qz} and β_{qm} to simulate four different scenarios: (1) The self reported data only depend on the true value (2) The self reported data only depend on the true value and the covariate (3) The self reported data depend on the true value of both Z and M, but do not depend on V (4) The self reported data depend on all variables. We have $\beta_{qz} = (2, 0.5, 0, 0)$, $\beta_{qm} = (1, 0, 1.5, 0)$, $\beta_{qz} = (2, 0.5, 0, 1)$, $\beta_{qm} = (1, 0, 1.5, -1)$, $\beta_{qz} = (2, 0.5, 0.3, 0)$, $\beta_{qm} = (1, -0.2, 1.5, 0)$, $\beta_{qz} = (2, 0.5, 0.3, 1)$, $\beta_{qm} = (1, -0.2, 1.5, -1)$ respectively.

We compare the following estimation methods: (1) Fitting model with true $Z(t)$ and $M(t)$, (2) Fitting model with self reported data, i.e $Q(t)$ and $X(t)$, (3) Fitting model with $W_z(t)$ and $W_m(t)$ using biomarker subsample, (4) Using separate calibration, i.e. calibrated $Z(t)$ by $Q(t)$ and $M(t)$ by $X(t)$, (5) Using joint calibration as described above. For the joint calibration, we fit separate linear regression model for $Z(t)$ and $M(t)$ and for the GEE calibration method, we allow the error $e_{wz}(t)$ and $e_{wm}(t)$ to be correlated and fit a GEE model on $(Z(t), M(t))$ with an indicator of whether one is Z or M to be interacted with all other variables.

The simulation results are shown in Tables 4.1-4.2. From the results, we can see that the naive estimator performs badly with large bias. The separate calibration is only valid when the measurement error does not depend on other variables (Z and M respectively). We can see that the regression calibration method generally works better than the biomarker subsample analysis for moderate correlation between $Z(t)$ and $M(t)$ and moderate measurement errors. However, a large correlation or large measurement error can lead to a very large variance of the estimator to the extent that it may become useless for data analysis (As shown in Table 4.2). Among these two factors, we find that correlation seems to be more important. Note that the correlation between either the self- reported data, i.e. $Cor(Q, X)$ or the correlation between the biomarker data, i.e. $Cor(W_z, W_m)$ does not indicate the magnitude of $Cor(Z, M)$ or more importantly $Cor(Z, M|V)$. In fact, when measurement error is large, using biomarker subsample gives smaller variance. However, given the limited sample size in the subsample and the low event rate, we still have low power to detect the effect and the interpretation shall need to be with cautious. This is analogous to findings with generalized estimating equation in similar settings, where the addition of relevant, but

noisy, variables to the estimation procedure may reduce performance.

When the measurement error does not depend on V, we have two ways to do the estimation. The first one is to leave V out of the calibration equation and the second one is first include V in the calibration equation and then integrate it out (Prentice and Huang 2011). We compare these two methods and find that both method perform similarly when we have large correlation and large measurement error. Including V first and integrate it out does not help for estimating survival parameters. The details can be found in Table 4.3.

4.1.4 Application to Energy Consumption, Activity-Related Energy Expenditure and Chronic Disease Risk

Diet and physical activity patterns over the lifespan constitute exposures that could explain much of the dramatic variations in chronic disease incidence rates worldwide [33], while also contributing importantly to risk variations within populations. Migrant studies suggest that recent exposures, when markedly different from those prior to migration, may be particularly influential [76]. Yet some decades of intensive analytic epidemiology research [1] [12] has failed to identify diet and activity factors that can explain much of the chronic disease risk variation. This lack of evidence undermines the specificity, reliability, and substance of diet and physical activity public health recommendations, and may detract from needed public health policy initiatives.

Previous analysis shows that without adjustment of total energy expenditure, physical activity did not show preventive effect on many major chronic diseases. Unfortunately, both the energy intake and physical activity-related energy expenditure data in the WHI are self-reported with substantial measurement error. We applied our method to WHI data to study how energy intake mediates the relationship between physical activity related energy consumption and obesity risk. Here $Z(t)$ denote the log of physical activity related energy expenditure, $M(t)$ denote the log of total energy consumption, $L(t)$ denote BMI and V include personal characteristics and is chosen different for different outcomes (See table 4.7 for details). The log of self-reported energy intake from the FFQ is $Q(t)$ and the log of self-reported physical activity related energy expenditure is $X(t)$. Using the method given in

Method	Variable	Bias$\times 10^3$	SD$\times 10^3$	sMSE$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$	sMSE$\times 10^3$
		Error does not depend on m or z			Error does not depend on m or z		
True	z	0	18	18	0	18	18
True	m	0	17	17	-2	18	18
True	v	-1	43	43	5	42	42
Naive	z	-187	15	188	-188	15	189
Naive	m	176	10	176	175	10	175
Naive	v	106	39	113	-127	41	133
Subsample	z	3	82	82	3	84	84
Subsample	m	-6	83	84	-6	82	83
Subsample	v	6	192	192	16	204	204
Separate Calibration	z	-16	45	48	-17	46	49
Separate Calibration	m	32	22	39	30	22	37
Separate Calibration	v	-31	67	74	-26	64	70
Joint Calibration	z	-5	48	48	-6	50	50
Joint Calibration	m	8	26	27	6	27	28
Joint Calibration	v	-16	67	69	-11	66	67
GEE Calibration	z	-5	48	48	-5	50	50
GEE Calibration	m	8	26	27	6	27	28
GEE Calibration	v	-16	67	69	-11	66	67
		Error does not depend on v			Error depends on all		
True	z	0	17	17	0	18	18
True	m	0	17	17	-1	18	18
True	v	-1	43	43	1	42	42
Naive	z	-216	15	217	-216	15	217
Naive	m	150	10	151	150	10	150
Naive	v	104	39	111	-126	43	134
Subsample	z	3	84	84	3	80	80
Subsample	m	1	76	76	-4	85	85
Subsample	v	-1	201	201	7	196	197
Separate Calibration	z	-80	42	90	-83	43	94
Separate Calibration	m	-26	23	35	-26	24	35
Separate Calibration	v	96	61	114	102	61	118
Joint Calibration	z	-5	45	45	-9	46	47
Joint Calibration	m	8	25	26	7	26	27
Joint Calibration	v	-16	66	68	-9	65	66
GEE Calibration	z	-5	45	45	-9	46	47
GEE Calibration	m	8	25	26	6	26	27
GEE Calibration	v	-16	66	68	-9	65	66

Table 4.1: Comparison of different method for estimating parameters in survival model with small measurement error and small correlation between Z and M. TRUE means using Z in the regression. Naive means using Q and X in the regression. Subsample means using W in the regression and only use those with W measured. Separate Calibration means calibrate Q with W_z and X with W_m. Joint Calibration means calibrate Q and X by W_z and W_m together. Generalized estimating equation (GEE) Calibration means a GEE rather than linear regression is used to construct calibration equation. sMSE denote the square root of mean square error.

Method	Variable	Bias$\times 10^3$	SD$\times 10^3$	sMSE$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$	sMSE$\times 10^3$
		Error does not depend on m or z			Error does not depend on m or z		
True	z	-1	114	114	-1	117	117
True	m	1	81	81	-1	83	83
True	v	-2	119	119	4	121	121
Naive	z	-299	21	300	-300	20	300
Naive	m	278	18	279	279	18	280
Naive	v	-235	66	244	-257	62	264
Subsample	z	-24	509	509	-25	517	518
Subsample	m	23	377	377	18	363	363
Subsample	v	-18	575	576	0	559	559
Separate Calibration	z	-198	2267	2276	-267	3031	3043
Separate Calibration	m	149	123	194	156	126	200
Separate Calibration	v	-102	2263	2265	-46	3114	3115
Joint Calibration	z	-114	13930	13930	-205707	6472781	6476049
Joint Calibration	m	74	7518	7518	107220	3373553	3375257
Joint Calibration	v	-35	1334	1335	-12757	401400	401603
GEE Calibration	z	-114	13930	13930	-205779	6475062	6478331
GEE Calibration	m	74	7518	7518	107258	3374742	3376446
GEE Calibration	v	-35	1334	1335	-12761	401542	401745
		Error does not depend on v			Error depends on all		
True	z	-2	114	114	2	117	117
True	m	2	83	83	-1	81	81
True	v	-1	123	123	-1	119	119
Naive	z	-304	19	304	-303	20	304
Naive	m	278	18	279	277	18	278
Naive	v	-234	68	243	-253	65	261
Subsample	z	-48	539	541	-29	515	516
Subsample	m	23	371	372	14	355	355
Subsample	v	4	555	555	-3	563	563
Separate Calibration	z	-385	459	599	-376	472	603
Separate Calibration	m	141	134	195	134	129	186
Separate Calibration	v	103	517	527	107	524	535
Joint Calibration	z	4356	134689	134760	855	23746	23761
Joint Calibration	m	-2401	74655	74694	-503	13760	13769
Joint Calibration	v	425	13898	13904	140	3552	3554
GEE Calibration	z	4356	134689	134760	855	23746	23761
GEE Calibration	m	-2401	74655	74694	-503	13760	13769
GEE Calibration	v	425	13898	13904	140	3552	3554

Table 4.2: Comparison of different method for estimating parameters in survival model with large measurement error and large correlation between Z and M. TRUE means using Z in the regression. Naive means using Q and X in the regression. Subsample means using W in the regression and only use those with W measured. Separate Calibration means calibrate Q with W_z and X with W_m. Joint Calibration means calibrate Q and X by W_z and W_m together. Generalized estimating equation (GEE) Calibration means a GEE rather than linear regression is used to construct calibration equation. sMSE denote the square root of mean square error.

Method	Variable	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$	Bias$\times 10^3$	SD$\times 10^3$
		Case I		Case II		Case III		Case IV	
n=10000, subsample=500, large correlation (0.8) between z and m, small error in q (0.5)									
True	z	0	114	0	116	0	114	0	117
True	m	0	60	0	60	0	60	0	62
Naive	z	-227	78	-107	50	-217	68	-132	41
Naive	m	112	18	17	34	95	30	-24	42
Subsample	z	-17	509	-17	515	-39	537	-24	42
Subsample	m	11	268	13	276	19	283	10	266
Sep cal int V	z	-155	197	-166	189	-217	148	-214	157
Sep cal int V	m	83	92	89	88	109	73	108	77
Joint Cal int V	z	21	377	-1	362	8	329	19	342
Joint Cal int V	m	-11	190	0	182	-4	167	-10	172
Sep cal no V	z	-234	70	3	78	-196	86	56	86
Sep cal no V	m	131	28	9	38	103	43	-28	44
Joint cal no V	z	12	377	1555	723	-1	327	1380	646
Joint cal no V	m	-6	190	-790	367	0	165	-696	328
n=10000, subsample=500, large correlation (0.8) between z and m, moderate error in q (1)									
True	z	0	115	0	117	0	115	0	112
True	m	0	60	0	62	0	62	0	60
Naive	z	-274	40	-210	32	-261	37	-209	27
Naive	m	117	14	78	23	109	19	54	28
Subsample	z	-34	498	-25	503	-34	493	-37	505
Subsample	m	18	264	13	269	21	262	17	269
Sep cal int V	z	-216	357	-203	369	-296	225	-298	231
Sep cal int V	m	111	160	106	166	146	105	147	107
Joint Cal int V	z	22	804	69	864	16	713	36	719
Joint Cal int V	m	-12	404	-35	434	-8	358	-18	360
Sep cal no V	z	-248	84	-115	66	-233	68	-85	66
Sep cal no V	m	143	22	77	29	128	30	49	34
Joint cal no V	z	14	799	1947	3519	9	684	1507	1272
Joint cal no V	m	-7	401	-990	1802	-4	343	-760	643
n=10000, subsample=500, small correlation (0.3) between z and m, moderate error in q (1)									
True	z	0	17	0	16	0	17	0	17
True	m	0	16	0	16	0	16	0	16
Naive	z	-215	15	-185	13	-230	15	-209	12
Naive	m	150	9	101	10	112	9	63	10
Subsample	z	3	78	4	77	7	75	1	81
Subsample	m	-1	74	-1	71	-2	74	-1	72
Sep cal int V	z	-16	40	-17	39	-63	39	-61	40
Sep cal int V	m	23	22	24	21	-13	22	-13	23
Joint Cal int V	z	-3	45	-4	42	-5	41	-3	41
Joint Cal int V	m	2	26	3	26	2	24	3	24
Sep cal no V	z	-43	40	15	35	-75	37	-9	35
Sep cal no V	m	87	18	-11	21	11	19	-95	23
Joint cal no V	z	-9	47	124	45	-11	43	124	44
Joint cal no V	m	5	28	-67	33	6	25	-51	30

Table 4.3: Comparison of integrating out V and leaving V out of calibration equation in estimating parameters in survival model with different measurement error magnitude and correlation between Z and M. Case I means error does not depend on m, z, v; case II means error does not depend on m, z; case III means error does not depend on v; case IV means error depends on all variables. Naive means using Q and X in the regression. Subsample means using W in the regression and only use those with W measured. 'Sep' means calibrate Q with W_z and X with W_m. 'Joint' means calibrate Q and X by W_z and W_m together.'int V' means include V in calibration equation and then integrate it out. 'no V' means exclude V from calibration equation.

Neuhouser *et al.* (2013)[31], we calculated self-reported physical activity by first computing the total METs by by combining the information of recreational physical activity from personal habit questionnaire and observational study questionnaire for yard work, home work, sitting and sleeping time. Then we multiply the total METs by weight to obtain an approximated estimate of activity-related energy expenditure (AREE). The biomarker for the total energy expenditure (TEE) is obtained from double labeled water assessments [**?**] while resting energy expenditure (REE) is obtained from indirect calorimetry [14]. The difference between TEE and REE is considered as the biomarker for AREE. The approximation of using the product of weight and METs to estimate activity related energy expenditure from self-reported physical activity data might not be adequate. Another limitation of the data is that for the biomarker subsample, only the recreational physical activity information is available. Thus when building the calibration equation, we can only use data collected at an earlier time point to compute mets for non-recreational physical activity and assume it did not change until the time biomarker subsample is collected. Since the physical activity data is only available for the OS cohort, we analyze this subgroup with time point 0 and 3. Missing data rates were generally low for modeled variables, and study subjects were excluded from analyses for a particular outcome if any of the corresponding modeled covariates was missing. Women having prior CVD, prior invasive cancer, or prior treated diabetes at enrollment were excluded from respective CVD, cancer and diabetes analyses.

The results for diabetes are shown in Table 4.4. From the table, we find a large variance for the estimator when joint calibration method is performed. To explain this result, we tried to estimate the correlation between $Z(t)$ and $M(t)$ using following method. Since we have reliability sample, so the biomarker data is available for two time points. For simplicity, we denote this two time points as 0 and 1. We assume

$$Var(e_{wz}(0), e_{wz}(1), e_{wm}(0), e_{wm}(1)) = \begin{pmatrix} \sigma^2_{wz} & \rho_{wz}\sigma^2_{wz} & \rho\sigma_{wz}\sigma_{wm} & 0 \\ \rho_{wz}\sigma^2_{wz} & \sigma^2_{wz} & 0 & \rho\sigma_{wz}\sigma_{wm} \\ \rho\sigma_{wz}\sigma_{wm} & 0 & \sigma^2_{wm} & \rho_{wm}\sigma^2_{wm} \\ 0 & \rho\sigma_{wz}\sigma_{wm} & \rho_{wm}\sigma^2_{wm} & \sigma^2_{wm} \end{pmatrix}$$

For this kind of structure, we assume there is correlation ρ_{wz} and ρ_{wm} between each variables at two time points and a correlation ρ between the error for two variables at the same time. The correlation between two variables at different time is assumed to be 0. Unfortunately, it is impossible to identify all 5 parameters from the data with only two time points. To estimate ρ, σ_{wz} and σ_{wm}, we need to specify ρ_{wz} and ρ_{wm}. We can estimate σ^2_{wz} and σ^2_{wm} by $Var(W_z(1) - W_z(0))/2/(1 - \rho_{wz})$ and $Var(W_m(1) - W_m(0))/2/(1 - \rho_{wm})$ respectively. Then we can obtain ρ by $Cov(W_z(1) - W_z(0), W_m(1) - W_m(0))/2/\sigma_{wm}/\sigma_{wz}$. After getting this information, we can obtain $Var(Z, M, V) = Var(W_z, W_m, V) - Var(e_z, e_m, V)$ and then obtain the conditional variance matrix for $Var(Z, M|V)$. The correlation between Z and M given V can be obtained directly from there.

If we consider relatively long term effect where $Z(t)$ and $M(t)$ be considered as long term average, then it is likely that ρ_{wz} and ρ_{wm} is positive. However, when we consider them to be short term average, ρ_{wz} and ρ_{wm} can be negative. We find that with a negative ρ_{wz} and ρ_{wm}, we tend to obtain a reasonable ρ, σ_{wz} and σ_{wm} and result in a smaller correlation between Z and M.

The result shows that these two variables are at least moderately correlated (0.6) even when we assume a highly negative correlation (-0.8) between the measurement error at two time points. The measurement error is large with the ratio of variance of measurement error to the variance of exposure or mediator is around 6-8, which means the variance of self-reported data given all covariates is 6-8 times the variance of underlying log transformed TEE or AREE given all other covariates. The wide confidence interval for the HRs estimated from joint calibration approach can be explained by the fact that given the information from personal habit questionnaire (PHQ), adding the self-reported total energy from FFQ into the calibration equation for AREE did not explain more variation (R2 increase less than 0.01). To better recover the true activity related energy expenditure and total energy intake, we add log transformed self-reported total fat, protein and carbohydrate into the calibration equation. We observe an increased R2 of 0.035 for AREE and a narrower CI for the HR estimates. However, we need to make an additional assumption that given the total calories, the macronutrient component from which the energy derive from does not contribute to the risk of diabetes.

Uncal: Energy	1.06(1.04,1.07)
Uncal: PA	1.01(1.00,1.02)
Uncal BMI: Energy	1.03(1.02,1.04)
Uncal BMI: PA	0.98(0.98,0.99)
Cal: Energy	4.17(2.53,6.87)
Cal: PA	0.60(0.42,0.85)
Cal BMI: Energy	50.3(0.01,Inf)
Cal BMI: PA	0.24(0.01,5.53)
Better cal: Energy	2.73(2.06,3.64)
Better cal: PA	0.84(0.68,1.05)
Better cal BMI: Energy	0.93(0.52,1.67)
Better cal BMI: PA	1.04(0.86,1.26)
Uncal: BMI	1.07(1.06,1.08)
Cal: BMI	0.89(0.61,1.29)
Better cal: BMI	1.07(1.04,1.11)

Table 4.4: Estimated Hazard Ratio (95% Confidence Interval) associated with 20% higher uncalibrated and calibrated total energy intake, activity related energy expenditure or resting energy expenditure for treated diabetes Incidence, in the Womens Health Initiative Observational Study (OS), from Baseline (1994-1998) Through September 17, 2012. Better calibration denotes calibration using micronutrient data. PA means association with activity-related energy expenditure.

Since the activity related energy expenditure (AREE) is not directly measured in the biomarker subsample, its measurement error might be in a complex form with total energy expenditure (TEE). So a possible way is to instead modeling resting energy expenditure (REE=TEE-AREE) and total energy expenditure (TEE) to make the measurement error of these two component to approximately independent. Also, it might be interesting to use AREE and REE to make scientific interpretation easier. However, we find that using any of these two variables face the same problem of large variance in the estimator from joint calibration.

The above analysis suggest that the current data available might not support a joint analysis of physical activity and energy intake. In fact, we noticed that calibration equation for both these two variables mainly depend on covariates with only minor dependency on self reported physical activity. But given these variables, self-reported energy intake from the FFQ data did not provide much more information. This suggest that we need better assessment of energy intake and physical activity so that we can understand the complex

relationship between these two variables and the disease risk.

From table 4.4, we can see that if we assume the BMI affect diabetes risk, then we cannot obtain any significant association between either TEE or AREE and diabetes risk using regression calibration method. The weak positive association from the uncalibrated analysis might due to the bias in self-reported data. However, when we assume the BMI effect is explained through long term total energy intake, we observe strong positive association between energy intake and diabetes risk. There is also a strong negative association between physical activity and diabetes risk, however, the result is not significant due to the large variation of the estimator.

Finally, we applied our method to other outcomes including various cancer and cardiovascular disease categories. The result is shown in Table 4.5 and Table 4.6. From the results, we can see that for the calibration method, if we did not make the assumption that BMI is no longer a risk factor for the disease after TEE and AREE are included in the model, then we did not observe any significant association between either TEE or AREE and any disease risk. If we assume the influence of BMI is fully captured and it is purely used to help assess the long term total energy consumption and long term physical activity, then we find that higher total energy consumption is associated with higher total cancer risk, breast cancer risk, endometrial cancer risk and total cardiovascular disease (CVD). We also find that higher physical activity is associated with lower risk of cardiovascular disease.

HRs for a 20% increment in TE are in the vicinity of 1.5 for most coronary disease outcomes, while corresponding HRs for a 20% increment in AREE are in the vicinity of 0.8. HRs are particularly extreme for coronary death and for congestive heart failure. Similar HRs are observed for ischemic stroke, whereas HR patterns appear to be quite different, though non-significant, for hemorrhagic stroke. In contrast uncalibrated TE and AREE, associations with CVD incidence were weak or non-existent. Total and site-specific cancer findings were similar to those for CVD. Positive associations of calibrated TE, and inverse associations of calibrated AREE, with total invasive cancers, obesity-related cancer, breast cancer, and rectal cancer were found. Additionally, TE hazard ratios were elevated for colon cancer, endometrial cancer, and kidney cancer even though corresponding AREE HRs were not significant; and the pancreatic cancer HR was lower at higher AREE, even though the

TE hazard ratio was not significant. A 20% increment in calibrated TE was associated with an HR of about 1.7 for obesity-related cancers, with corresponding HR of about 0.8 for a 20% increment in AREE. Corresponding estimated HRs were essentially null without measurement error correction. Using macronutrients proportions can improve precision, but did not change point estimates much. Many BMI associations are still significant when better calibration method is used. For those outcomes, the analysis without BMI included should be carefully interpreted.

Part of the work in 4.1.4 is included in a manuscript with WHI-related co-authors, which was approved by the WHI Publications and Presentation Committee prior to journal submission.

	All Invasive Cancer	Obesity related cancer	Breast	Colon	Rectum
Uncal: Energy	1.01(1.00,1.02)	1.02(1.00,1.03)	1.01(0.99,1.02)	1.00(0.96,1.03)	1.01(0.91,1.11)
Uncal: PA	0.99(0.99,1.00)	1.00(0.99,1.01)	1.00(0.99,1.01)	1.00(0.97,1.03)	0.99(0.92,1.06)
Uncal BMI: Energy	1.01(1.00,1.02)	1.01(1.00,1.02)	1.00(0.99,1.02)	0.99(0.96,1.02)	1.00(0.91,1.10)
Uncal BMI: PA	0.99(0.98,1.00)	0.99(0.98,1.00)	0.99(0.98,1.01)	0.99(0.97,1.02)	0.98(0.92,1.05)
Cal: Energy	1.43(1.11,1.82)	1.71(1.26,2.31)	1.47(1.15,1.89)	1.86(1.06,3.24)	2.75(1.02,7.40)
Cal: PA	0.84(0.70,1.01)	0.79(0.63,0.98)	0.82(0.69,0.98)	0.83(0.63,1.09)	0.51(0.25,1.06)
Cal BMI: Energy	12.02(0.11,1372)	10.12(0.12,855)	23.91(1.05,546)	0.85(0.11,6.56)	19.55(0.00,Inf)
Cal BMI: PA	0.37(0.06,2.29)	0.40(0.07,2.28)	0.29(0.09, 0.89)	0.99(0.61,1.62)	0.27(0.01,6.01)
Better cal: Energy	1.22(1.10,1.37)	1.38(1.21,1.58)	1.23(1.08,1.40)	1.33(0.88,1.99)	1.50(0.81,2.78)
Better cal: PA	0.95(0.87,1.02)	0.93(0.85,1.02)	0.95(0.87,1.03)	0.99(0.83,1.17)	0.81(0.56,1.18)
Better cal BMI: Energy	1.00(0.75,1.34)	1.01(0.73,1.40)	0.90(0.66,1.24)	0.53(0.17,1.66)	0.51(0.08,3.20)
Better cal BMI: PA	0.98(0.90,1.06)	0.98(0.89,1.07)	1.00(0.91,1.10)	1.12(0.90,1.39)	0.96(0.59,1.57)
Uncal: BMI	1.01(1.01,1.02)	1.02(1.02,1.02)	1.01(1.01,1.01)	1.02(1.01,1.03)	1.02(0.99,1.05)
Cal: BMI	0.91(0.73,1.11)	0.92(0.75,1.12)	0.88(0.77,1.00)	1.04(0.94,1.15)	0.91(0.56,1.47)
Better cal: BMI	1.02(1.00,1.03)	1.02(1.00,1.04)	1.02(1.00,1.04)	1.06(0.99,1.13)	1.08(0.96,1.21)
	Ovary	Endometrial	Bladder	Kidney	Pancreas
Uncal: Energy	1.00(0.96,1.05)	1.08(1.04,1.13)	1.03(0.97,1.10)	1.05(0.98,1.12)	0.95(0.89,1.01)
Uncal: PA	1.01(0.98,1.05)	1.01(0.98,1.05)	0.96(0.92,1.00)	1.02(0.96,1.07)	0.97(0.92,1.01)
Uncal BMI: Energy	1.00(0.95,1.05)	1.06(1.02,1.10)	1.03(0.97,1.09)	1.03(0.97,1.10)	0.94(0.89,1.00)
Uncal BMI: PA	1.02(0.98,1.06)	1.00(0.97,1.03)	0.95(0.92,0.99)	1.00(0.95,1.05)	0.96(0.92,1.01)
Cal: Energy	0.85(0.41,1.75)	2.72(1.29,5.74)	1.80(0.82,3.98)	2.94(1.20,7.22)	2.06(0.86,4.92)
Cal: PA	1.12(0.68,1.83)	0.77(0.45,1.30)	0.68(0.39,1.19)	0.62(0.31,1.26)	0.61(0.33,1.11)
Cal BMI: Energy	0.45(0.01,25.6)	17.58(0.10,3012)	150.8(0.00,Inf)	41.0(0.00,Inf)	0.04(0.98,1.00)
Cal BMI: PA	1.40(0.34,5.73)	0.40(0.05,3.11)	0.14(0.00, 8.96)	0.24(0.01,7.28)	2.38(0.14,41.3)
Better cal: Energy	0.95(0.66,1.38)	2.29(1.72,3.05)	1.43(0.92,2.24)	1.78(1.06,2.99)	1.27(0.77,2.08)
Better cal: PA	1.03(0.85,1.25)	0.88(0.74,1.03)	0.81(0.62,1.04)	0.93(0.65,1.32)	0.87(0.68,1.12)
Better cal BMI: Energy	1.02(0.46,2.27)	2.22(1.11,4.44)	1.18(0.39,3.56)	0.69(0.17,2.79)	0.44(0.14,1.40)
Better cal BMI: PA	1.02(0.83,1.26)	0.88(0.74,1.04)	0.83(0.60,1.14)	1.11(0.72,1.73)	1.02(0.75,1.39)
Uncal: BMI	1.00(0.98,1.02)	1.05(1.03,1.06)	1.01(0.99,1.03)	1.05(1.02,1.06)	1.02(1.00,1.04)
Cal: BMI	1.03(0.86,1.23)	0.92(0.72,1.16)	0.81(0.48,1.38)	0.88(0.56,1.38)	1.20(0.84,1.72)
Better cal: BMI	1.00(0.94,1.06)	1.00(0.96,1.05)	1.02(0.95,1.09)	1.07(0.98,1.16)	1.08(1.00,1.16)
	Lung	Lymphoma	Leukemia		
Uncal: Energy	0.99(0.96,1.02)	1.08(1.03,1.13)	1.01(0.95,1.06)		
Uncal: PA	0.97(0.95,0.99)	1.00(0.96,1.03)	0.98(0.94,1.02)		
Uncal BMI: Energy	0.99(0.96,1.02)	1.08(1.03,1.13)	1.00(0.95,1.06)		
Uncal BMI: PA	0.97(0.95,1.00)	0.99(0.96,1.03)	0.97(0.93,1.02)		
Cal: Energy	1.14(0.70,1.86)	0.99(0.44,2.26)	1.48(0.65,3.37)		
Cal: PA	0.79(0.56,1.10)	1.16(0.65,2.05)	0.74(0.43,1.30)		
Cal BMI: Energy	7.5(0.07,797)	48.4(0.06,Inf)	0.97(0.00,202)		
Cal BMI: PA	0.41(0.07,2.34)	0.32(0.03,3.25)	0.86(0.15, 5.11)		
Better cal: Energy	0.90(0.68,1.19)	0.94(0.64,1.38)	1.00(0.62,1.62)		
Better cal: PA	0.93(0.80,1.09)	1.20(0.95,1.51)	0.99(0.76,1.29)		
Better cal BMI: Energy	0.66(0.34,1.26)	1.25(0.43,3.63)	0.57(0.18,1.85)		
Better cal BMI: PA	0.97(0.81,1.17)	1.17(0.90,1.51)	1.04(0.77,1.42)		
Uncal: BMI	0.99(0.98,1.00)	1.01(0.99,1.03)	1.00(0.98,1.03)		
Cal: BMI	0.92(0.73,1.15)	0.84(0.61,1.14)	1.02(0.79,1.32)		
Better cal: BMI	1.02(0.98,1.07)	0.98(0.91,1.05)	1.04(0.97,1.13)		

Table 4.5: Estimated Hazard Ratio (95% Confidence Interval) associated with 20% higher uncalibrated and calibrated total energy intake, activity related energy expenditure or resting energy expenditure for various cancer categories, in the Womens Health Initiative Observational Study (OS), from Baseline (1994-1998) through September 30, 2010. Better calibration denotes calibration using micronutrient data. PA means association with activity-related energy expenditure.

	CHD	Nonfatal MI	Coronary Death	CHF	CABG and PCI
Uncal: Energy	1.00(0.98,1.02)	1.00(0.98,1.03)	0.97(0.94,1.02)	1.04(1.01,1.08)	1.01(0.99,1.04)
Uncal: PA	0.99(0.97,1.01)	0.99(0.97,1.01)	0.97(0.94,1.00)	0.97(0.95,1.00)	1.01(0.99,1.03)
Uncal BMI: Energy	1.00(0.97,1.02)	1.00(0.98,1.02)	0.97(0.93,1.01)	1.03(1.00,1.06)	1.01(0.99,1.03)
Uncal BMI: PA	0.99(0.97,1.00)	0.99(0.97,1.01)	0.96(0.93,0.99)	0.96(0.93,0.98)	1.01(0.99,1.02)
Cal: Energy	1.57(1.18,2.09)	1.49(1.12,1.99)	2.22(1.34,3.65)	3.51(2.05,6.02)	1.43(1.16,1.74)
Cal: PA	0.78(0.64,0.96)	0.80(0.66,0.98)	0.63(0.45,0.88)	0.57(0.40,0.82)	0.90(0.78,1.04)
Cal BMI: Energy	2.62(0.15,46.7)	3.07(0.16,60.1)	1.31(0.05,32.4)	31.3(0.02,Inf)	1.47(0.28,7.65)
Cal BMI: PA	0.67(0.27,1.67)	0.64(0.22,1.83)	0.74(0.27, 2.03)	0.29(0.02,3.45)	0.89(0.51,1.55)
Better cal: Energy	1.34(1.08,1.66)	1.24(0.99,1.55)	1.78(1.23,2.58)	2.42(1.71,3.41)	1.23(1.03,1.47)
Better cal: PA	0.88(0.76,1.01)	0.92(0.79,1.06)	0.74(0.59,0.92)	0.77(0.60,0.99)	1.00(0.89,1.13)
Better cal BMI: Energy	0.80(0.45,1.43)	0.72(0.38,1.36)	0.82(0.35,1.96)	0.77(0.26,2.25)	0.72(0.44,1.19)
Better cal BMI: PA	0.97(0.82,1.16)	1.02(0.84,1.24)	0.85(0.66,1.11)	0.96(0.70,1.33)	1.12(0.96,1.31)
Uncal: BMI	1.02(1.01,1.03)	1.02(1.01,1.02)	1.03(1.02,1.04)	1.06(1.05,1.07)	1.02(1.01,1.02)
Cal: BMI	0.97(0.84,1.12)	0.96(0.82,1.13)	1.03(0.87,1.21)	0.89(0.59,1.35)	1.00(0.92,1.08)
Better cal: BMI	1.04(1.00,1.07)	1.04(1.00,1.08)	1.06(1.00,1.11)	1.08(1.01,1.15)	1.04(1.01,1.07)
	Total Stroke	Ischemic Stroke	Hemorrhagic Stroke	CHD and Stroke	All CVD
Uncal: Energy	0.97(0.95,1.00)	0.98(0.95,1.01)	0.94(0.89,0.99)	0.99(0.97,1.00)	1.00(0.98,1.01)
Uncal: PA	0.99(0.97,1.01)	0.99(0.97,1.01)	1.03(0.99,1.08)	0.99(0.98,1.00)	1.00(0.99,1.01)
Uncal BMI: Energy	0.97(0.95,0.99)	0.98(0.95,1.00)	0.94(0.89,1.00)	0.98(0.97,1.00)	0.99(0.98,1.01)
Uncal BMI: PA	0.99(0.97,1.01)	0.98(0.96,1.00)	1.04(0.99,1.10)	0.99(0.97,1.00)	0.99(0.98,1.00)
Cal: Energy	1.36(1.02,1.81)	1.55(1.13,2.12)	0.47(0.20,1.13)	1.49(1.16,1.90)	1.49(1.22,1.83)
Cal: PA	0.83(0.67,1.02)	0.78(0.62,0.96)	1.37(0.79,2.36)	0.80(0.67,0.94)	0.83(0.72,0.95)
Cal BMI: Energy	0.43(0.04,4.31)	0.89(0.11,7.44)	0.01(0.00,184)	1.23(0.27,5.5)	1.27(0.40,4.07)
Cal BMI: PA	1.19(0.55,2.59)	0.93(0.46,1.86)	5.09(0.17,150)	0.85(0.53,1.35)	0.87(0.58,1.30)
Better cal: Energy	1.40(0.91,1.43)	1.21(0.93,1.58)	0.63(0.34,1.16)	1.24(1.03,1.48)	1.24(1.06,1.46)
Better cal: PA	0.94(0.82,1.09)	0.93(0.78,1.10)	1.12(0.80,1.56)	0.91(0.80,1.03)	0.95(0.85,1.05)
Better cal BMI: Energy	0.58(0.32,1.07)	0.54(0.26,1.14)	0.86(0.23,3.27)	0.66(0.41,1.08)	0.66(0.41,1.01)
Better cal BMI: PA	1.08(0.90,1.29)	1.09(0.87,1.36)	1.05(0.69,1.60)	1.03(0.89,1.20)	1.08(0.94,1.22)
Uncal: BMI	1.01(1.00,1.02)	1.02(1.01,1.02)	0.97(0.94,1.00)	1.02(1.01,1.02)	1.02(1.01,1.02)
Cal: BMI	1.06(0.93,1.21)	1.03(0.92,1.15)	1.24(0.71,2.16)	1.01(0.94,1.09)	1.01(0.95,1.07)
Better cal: BMI	1.05(1.01,1.09)	1.06(1.01,1.10)	0.98(0.90,1.06)	1.04(1.01,1.07)	1.04(1.02,1.07)

Table 4.6: Estimated Hazard Ratio (95% Confidence Interval) associated with 20% higher uncalibrated and calibrated total energy intake, activity related energy expenditure or resting energy expenditure for cardiovascular diseases, in the Womens Health Initiative Observational Study (OS), from Baseline (1994-1998) Through September 30, 2010. Better Calibration denotes calibration using micronutrient data. PA means association to activity-related energy expenditure. Outcome abbreviations: CHF: cogestive heart failure; CHD, coronary heart disease; MI, myocardial infarction; CABG, coronary artery bypass graft; PCI, percutaneous coronary intervention.

	Age	Race	Income	Education	Smoking	Alcohol	Hormone Use	Hypertension
CVD:	x	x	x	x	x	x	x	x
All Cancer:	x	x	x	x	x	x	x	x
Obesity-Related Cancer:	x	x	x	x	x	x	x	x
Breast Cancer:	x	x	x	x	x	x	x	
Colon Cancer:	x	x	x		x	x	x	
Rectum Cancer:	x	x	x		x	x		
Ovary Cancer:	x	x	x				x	
Endometrial Cancer:	x	x	x				x	x
Bladder Cancer:	x	x	x		x	x		
Kidney Cancer:	x	x	x		x	x		x
Pancreas Cancer:	x	x	x		x	x		
Lung Cancer:	x	x	x		x	x		
Lymphoma:	x	x	x					
Leukemia:	x	x	x		x			
Diabetes:	x	x	x	x	x	x	x	x
	CVD His	Cancer His	Diabetes His	CVD Rel	Breast Rel	Colon Rel	Polyps	Diabetes Rel
CVD:	e	x	x	x				x
All Cancer:		e	x					
Obesity-Related Cancer:		e	x					
Breast Cancer:		e			x			
Colon Cancer:		e				x	x	
Rectum Cancer:		e				x		
Ovary Cancer:		e			x			
Endometrial Cancer:		e						
Bladder Cancer:		e						
Kidney Cancer:		e						
Pancreas Cancer:		e						
Lung Cancer:		e						
Lymphoma:		e						
Leukemia:		e						
Diabetes:	x	x	e					x

Table 4.7: Covariates Included in the Disease Risk Models for Various Outcomes. (Here, CVD denotes all of the cardiovascular disease categories considered. The notation x means that the variable is included in the disease model, and e means that the variable was used to exclude women from analysis. 'Rel' is an abbreviation for first degree relative having the disease.)

Chapter 5

MEDIATION ANALYSIS WITHOUT SEQUENTIAL IGNORABILITY

In this section, we provide estimators for an additive hazard model without requiring the sequential ignorability assumption which means we allow confounding of a particular form between the mediator and outcome. Compared with the Cox model, the additive hazard models ([26]) have the convenient property that the parameters are collapsible and that under certain assumption, the indirect effect can be written as the product of the intervention effect on mediator and the mediator effect. Under sequential ignorability assumption, Martinussen *et al.* (2011) [28] proposed an estimator for direct effect using additive hazard model with time varying effect.

5.1 Models and assumptions

In this section, we first give the models in the potential outcome notation. Then we give the detailed expression of the 'causal' parameters of interest in terms of parameters in the model. We present the assumptions used in this paper and relate the potential outcome model to the derived model for the observed variables and unmeasured confounders.

5.1.1 Model

In this section, we only consider time-independent exposure and mediator. As before, we denote the exposure/intervention as Z and denote the mediator as M. We denote the measured and unmeasured confounder with $\boldsymbol{V}$ and $\boldsymbol{U}$ respectively. We denote the potential event time and censoring time as $T(z,m)$ and $C(z,m)$ respectively. We denote the potential mediator as $M(z)$ and we denote the observed composite endpoint as T^* and Δ.

We propose to use the following forms to model the potential cumulative hazard for the event time.

$$\Lambda(t,z,m|\boldsymbol{V}_i,\boldsymbol{U}_i) = \Theta_z(t)z + \Theta_m(t)m + G(z,m,\boldsymbol{V}_i,\boldsymbol{U}_i,t), \tag{5.1}$$

where $G(z, m, \boldsymbol{V}_i, \boldsymbol{U}_i, t)$ can be an unknown function. We do not need a baseline term, as it can be absorbed into the unknown time-varying function G.

We use the same measurement error as in Chap 2.1:

$$
\begin{aligned}
M_i(z) &= g_z(z, \boldsymbol{V}_i; \boldsymbol{\alpha}) + \epsilon(z)_i, \\
Q_i(t) &= g_{q0}(t, Z_i, \boldsymbol{V}_i; \boldsymbol{\gamma}) + g_{q1}(t, Z_i, \boldsymbol{V}_i; \boldsymbol{\gamma}) M_i + \varepsilon_i(t), \\
W_i(t) &= M_i + \omega_i(t).
\end{aligned}
$$

Here $\epsilon(z)_i$ is normally distributed with variance $\sigma_z^2(z, \boldsymbol{V}_i; \boldsymbol{\theta})$ and is independent of other covariates, $\varepsilon_i(t)$ is Gaussian process with variance process $K_q(t, s) = K(t, s; \boldsymbol{\theta}_q, Z_i, \boldsymbol{V}_i)$ and is independent of other covariates, and $\omega_i(t)$ is Gaussian process with variance process $K_w(t, s) = K(t, s; \boldsymbol{\theta}_w, Z_i, \boldsymbol{V}_i)$ and is independent of other covariates. All gs are known link functions. Here we assume $W_i(t)$ is only available within a biomarker subsample in data analysis.

Then we can write the controlled mediator effect (CME) and controlled direct effect (CDE) as follow:

$$
\begin{aligned}
CME(z, m_1, m_2, t) &= \Theta_z(t)(m_1 - m_2) \\
CDE(z, t) &= \Theta_z(t).
\end{aligned}
$$

We assume that the sequential ignorability assumption hold if $\boldsymbol{U}$ is observed, which ensure

$$
T(z, m) \perp M(z) | Z, \boldsymbol{V}, \boldsymbol{U}.
$$

We also assume that there is no unmeasured confounder between exposure and both mediator and outcome, which is typically satisfied when Z is intervention in randomized trial. If we assume a linear model $g_z(z, \boldsymbol{V}_i; \boldsymbol{\alpha}) = (z, \boldsymbol{V}_i)\boldsymbol{\alpha}$, then we can write the natural direct effect (NDE) and natural indirect effect (NIE) as follow:

$$
NDE(z, t) = \Theta_z(t)
$$

$$NIE(z,t) \quad = \quad \Theta_z(t)\boldsymbol{\alpha}_1.$$

Under this model, one just needs to estimate $\Theta_m(t)$ and $\Theta_z(t)$ and $\boldsymbol{\alpha}_1$ to compute the quantities needed to assess mediation. Due to the assumption that Z is randomized, $\boldsymbol{\alpha}_1$ can be easily obtained from a regression of M on Z. So what we focus on this chapter is how to estimate $\Theta_m(t)$ and $\Theta_z(t)$.

Here we want to point out that if we have $\Theta_z(t) = \theta_z t$ and $\Theta_m(t) = \theta_m t$, then the model below is an additive hazard model with the form

$$\lambda(t, z, m|\boldsymbol{X}_i, \boldsymbol{U}_i) = \theta_z z + \theta_m m + g(z, m, \boldsymbol{V}_i, \boldsymbol{U}_i, t).$$

However, when the model has time-varying effects, we have $\Theta_r(t) = \int_0^t \theta_r(s)ds$ and $\Theta_z(t) = \int_0^t \theta_z(s)ds$, which describe the cumulative effects until a certain time t. The additional flexibility provided by allowing time-varying effects will typically be needed in application.

5.1.2 Assumptions

In this section, we listed all assumptions we make for this Chapter's results.

1. Assumption 1: The stable unit treatment value assumption (SUTVA) includes two sub-assumptions: (1) there is only one set of potential outcomes for each participant, which does not depend on the treatment assignment or mediator value of other participants; and (2) the observed outcome is just one realization of the potential outcome with observed intervention level r and mediator level z. The first part of the SUTVA assumption means that there is not multiple versions of the intervention, and there is no interference between participants. This allows us to use a single model for the sample, and interactions between participants do not need to be accounted for in the model.

2. Assumption 2: Consistency and positivity for the survival outcome (or say its cumu-

lative hazard) and censoring time. Mathematically, this means:

$$T = \sum_{z,m} T(z,m) I(Z = z, M = m),$$

and

$$C = \sum_{z,m} C(z,m) I(Z = z, M = m),$$

where $I(\cdot)$ is the indicator function. Also, we need that the probability of observing any joint level of (z, m) given covariate V to be greater than 0. This assumption means what we observed is one realization from the counterfactual space, and this assumption is need to relate our model to the observed variables.

3. Assumption 3: Randomization (i.e. ignorability). The observed intervention assignment Z is independent of the mediator and all potential outcomes (defined by different levels of intervention and mediator) conditional on covariates. For the traditional model, the intervention is not affected by any unmeasured covariates, and thus there is no confounding between intervention and mediator or outcome. A mathematical representation of this can be written as $\boldsymbol{U} \perp Z | \boldsymbol{V}$.

4. Assumption 4: Model correctly specified. It is required that the model shown in equation 5.1 is correctly specified up to unknown parameters. The only exception is that we allow misspecification of function $g(v, u, t)$. This means that observed covariates should not modify the direct and mediator effects. This assumption is partly testable and can be relaxed.

5. Assumption 5: The same conditional distribution (for the unknown function with unmeasured confounder). Mathematically, we require $g(z, m, \boldsymbol{V}_i, \boldsymbol{U}_i, t)$ or $G(z, m, \boldsymbol{V}_i, \boldsymbol{U}_i, t)$ to have the same means and the variances condition on $\boldsymbol{V}_i$ and $\boldsymbol{U}_i$. This can be implied by the assumption that no unmeasured confounder serve as effect modifier for the direct effect of intervention or the mediator effect. This is a crucial assumption

that cannot be tested from the data. But this assumption is often weaker than the sequential ignorability assumption.

6. Assumption 6: For each time t at which direct and mediator effects are to be estimated, the probability of survival to that time is larger than some $\delta > 0$. This is just a useful and trivial assumption for 'survival' analysis. If all failure time are censored before a certain time, then we cannot identify what happened after that time. Other regularity conditions for survival analysis are given later in the proof.

7. Assumption 7: Independent Censoring. The censoring time is independent of event time given all measured covariates. As we know that the event time is a function of unmeasured confounder $\boldsymbol{U}$, so this assumption requires that the hazard function for the censoring time not associated with $\boldsymbol{U}$, which implies no unmeasured confounder between M and C.

8. Assumption 8: Identification Requirement. For locally identifiability, we need that there exist functional form $A(Z, \boldsymbol{V})$ such that $EA(Z, \boldsymbol{V})|\boldsymbol{V} = 0$ and the following matrix

$$A(Z, \boldsymbol{V})S(t)\exp(\Theta_z(t)Z + \Theta_m(t)M)(Z, M)$$

   is either positive definite or negative definite at the true parameters. For globally identifiability, we need the above matrix to be positive definite or negative definite at all parameter values. Since $S(t)$ contains information of U, this assumption is not testable from the data. But in the simulation, we have assessed this assumption for different models and parameters and the result suggest this assumption holds under our simulation settings.

9. Assumption 9: No systematic measurement error with respect to any unmeasured confounders.

Here we would like to discuss these assumptions. Assumption 1 is a commonly used assumption in causal inference. Though the no interference assumption might not hold when we study the group behavior or infectious disease, it is likely to be true when the outcome is chronic disease like breast cancer. Assumption 2 is a necessary assumption that allows us to relate the potential outcome to the real outcome. Assumption 3 automatically holds for randomized clinical trial. Assumption 4 is an important but untestable assumption. But since this is just a model assumption and our model is only a semiparametric model with restriction on marginal hazard functions, our model is still flexible enough to approximate the truth in many applications. Assumption 5 is another untestable assumption, but compared with the sequential ignorability assumption, which assumes that there is no unmeasured confounder, this assumption becomes weaker since it may be more likely to have a main effect than an interaction effect. Assumption 6 is required for most survival analysis since if all individuals are censored before certain time t, then there is no way to know what would happen after that time unless a parametric model is assumed for extrapolation. Assumption 7 is also a commonly used assumption for survival analysis. Assumption 8 is untestable from the observed data because $S(t|Z, M, \boldsymbol{V}, \boldsymbol{U})$ depends on unobserved $\boldsymbol{U}$ and the expectation depend on nuisance parameter η. But in simulation, we have assessed this assumption for different models and parameters, and the simulation results suggested that the locally identifiability assumption holds under those settings. Assumption 8 places a modest requirement on the covariate. In conclusion, compared with the additive hazard regression method, we relax the sequential ignorability assumption to Assumption 5 with the price of requiring Assumption 8. Assumption 9 means that the measurement error is ignorable conditional on the measured covariates and this assumption is likely to be true in many applications.

5.2 Method

In this section, we first give the estimator when Z is observed and then extend to the case with measurement error by regression calibration under a rare disease assumption. Here we list some regularity conditions as follows:

- $\boldsymbol{\Theta}(t)$ belong to a compact range of support and with the true value $\boldsymbol{\Theta}_0(t)$ in the interior.

- The joint density of observed variables have third order partial derivatives for $\boldsymbol{\Theta}$.

- Z, M and $\boldsymbol{V}$ are bounded.

- $S_C(t|z, m, \boldsymbol{v})$ and $S(t|z, m, \boldsymbol{v})$ is bounded from 0.

- Global identifiability, i.e. $E\left[I(T > t)exp(\Theta_z(t)Z + \Theta_m(t)M)\right](Z, M)$ is positive definite for all $\boldsymbol{\Theta}(t)$.

5.2.1 Identifiability

Theorem 1:

The parameter $\boldsymbol{\Theta}$ in model 6.1 is locally identifiable under assumption 1-8 and is globally identifiable under the stronger version of assumption 8 described above.

Proof: Here, we will show the above conditions combined with assumption 1-8 is sufficient condition for locally (globally if stronger version of assumption 8 hold) identifiability of the parameter $\boldsymbol{\Theta}$ from the estimating equation 5.3 with some $A(Z, \boldsymbol{V}, t)$. For simplicity, we simply denote the summand of the right hand side of equation 5.3 as $F_i(\boldsymbol{\Theta})$. Since the estimation of censoring survival distribution does not depend on $\boldsymbol{\Theta}$ and can be identified and consistently estimated. So for the proof below, we will treat $S_C(t|z, m, \boldsymbol{v})$ as known. We will use the following steps to show the identifiability for each time point t and thus we will omit the notation t below. For any two parameters $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$ that result in same distribution of observed variables, we will have

$$0 = E_{\boldsymbol{\Theta}_1} F_i(\boldsymbol{\Theta}_1) = E_{\boldsymbol{\Theta}_2} F_i(\boldsymbol{\Theta}_1).$$

Also, we know that $E_{\boldsymbol{\Theta}} F_i(\boldsymbol{\Theta}) = 0$ for any $\boldsymbol{\Theta}$. So we have

$$0 = E_{\boldsymbol{\Theta}_2} F_i(\boldsymbol{\Theta}_1) = E_{\boldsymbol{\Theta}_2} F_i(\boldsymbol{\Theta}_2).$$

So to prove that $\boldsymbol{\Theta}_1 = \boldsymbol{\Theta}_2$, we will just need to show $0 = E_{\boldsymbol{\Theta}_2} F_i(\boldsymbol{\Theta})$ have unique solution. It is sufficient to show that the derivative of $E_{\boldsymbol{\Theta}_2} F_i(\boldsymbol{\Theta})$ is positive definite. For local identifiability, as we assume the derivative is positive definite at $\boldsymbol{\Theta}^0$ and the derivative is continuous with respect to $\boldsymbol{\Theta}$, so there exists a region contain $\boldsymbol{\Theta}^0$ such that the derivative above is positive definite at any value within the value, this is sufficient for the locally identifiability hold. For the global identifiability, the assumption 8 tells us that the derivative is positive definite at any $\boldsymbol{\Theta}$.

5.2.2 Estimation

Under Assumption 1, 2, 4 and 5. The derived model given observed and unobserved covarites can be written in the additive hazard model form as:

$$\Lambda(t|Z_i, M_i, \boldsymbol{V}_i, \boldsymbol{U}_i) = \Theta_z(t)Z_i + \Theta_m(t)M_i + G^{Z_i M_i}(\boldsymbol{V}_i, \boldsymbol{U}_i, t) \tag{5.2}$$

The derived models clearly showed the assumption of unmeasured confounders to be in an additive term in either cumulative hazard or log transformed survival time. We will construct unbiased estimating equation based on these models.

We can rewrite the derived model 5.2 as the following form:

$$S_i(t|Z_i, M_i, \boldsymbol{V}_i, \boldsymbol{U}_i) \exp(\Theta_z(t)Z_i + \Theta_m(t)M_i) = \exp(-G(\boldsymbol{V}_i, \boldsymbol{U}_i, t))$$

By Assumption 3, we have $Z_i \perp \boldsymbol{U}_i | \boldsymbol{V}_i$ and thus we have for any time t

$$0 = E[A(Z_i, \boldsymbol{V}_i, t) S_i(t|Z_i, M_i, \boldsymbol{V}_i, \boldsymbol{U}_i) \exp(\Theta_z(t)Z_i + \Theta_m(t)M_i)]$$

where $A(\cdot,\cdot,\cdot) = (a_1(\cdot,\cdot,\cdot), a_2(\cdot,\cdot,\cdot))^T$ that satisfy $E[A(Z_i, \boldsymbol{V}_i, t)|\boldsymbol{V}_i] = \mathbf{0}$.

For the empirical version, when there is no censoring, we will use

$$0 = \sum_{i=1}^{n} [A(Z_i, \boldsymbol{V}_i, t) I(T_i > t) \exp(\Theta_z(t)Z_i + \Theta_m(t)M_i)].$$

and therefore obtain $\Theta_z(t)$ and $\Theta_m(t)$. When there is censoring, we will use the inverse

censoring probability weighting (IPCW) technique proposed by Robins and Rotnizky [46]. The weighted estimating equation can be written as

$$0 = \sum_{i=1}^{n} [A(Z_i, \boldsymbol{V}_i, t) \frac{I(T_i^* > t)}{\hat{S}_C(t|M_i, \boldsymbol{V}_i)} \exp(\Theta_z(t) Z_i + \Theta_m(t) M_i)]. \tag{5.3}$$

where $\hat{S}_C(\cdot, \cdot, \cdot)$ is a regular asymptotic linear (RAL) estimator for the censoring survival function. Here we want to comment that requirement of such an estimator is not a very strong assumption. Such estimator exist for most semi-parametric models such as Cox model, additive hazard model. Some kernel based estimator like nearest neighbor estimator (NNE) as shown in Cai *et al.* [6] also satisfy the requirement. Since by Assumption 7, there is no unmeasured confounder between mediator and censoring time. So the estimation from either semi-parametric regression or non-parametric kernel estimation will work. To gain efficiency, we can substract the projection of $[I(T > t) \exp(\Theta_z(t) Z + \Theta_m(t) M)]$ on $\boldsymbol{V}$ under the true $\boldsymbol{\Theta}$. However, this projection includes the unknown functional form $G(\boldsymbol{V}_i, \boldsymbol{U}_i, t)$ with unobserved covariate $\boldsymbol{U}$, so we can only use some working model $\tilde{G}(\boldsymbol{V}_i, t)$ and thus the estimating equation becomes

$$0 = \sum_{i=1}^{n} \{A(Z_i, \boldsymbol{V}_i, t) \left[I(T_i > t) \exp(\Theta_z(t) Z_i + \Theta_m(t) M_i) - \tilde{G}(\boldsymbol{V}_i, t) \right]\}. \tag{5.4}$$

The corresponding weighted version will be

$$0 = \sum_{i=1}^{n} \{A(Z_i, \boldsymbol{V}_i, t) \left[\frac{I(T_i^* > t)}{\hat{S}_C(t|Z_i, \boldsymbol{V}_i)} \exp(\Theta_z(t) Z_i + \Theta_m(t) M_i) - \tilde{G}(\boldsymbol{V}_i, t) \right]\}. \tag{5.5}$$

In practice, we can model the $\tilde{G}$ with some unknown parameters, and write this in the form as $\tilde{G}(\boldsymbol{V}_i, t, \boldsymbol{\beta}(t))$. We can use an extra set of estimating equation to estimate $\boldsymbol{\beta}(t)$, which can be written as the following form:

$$0 = \sum_{i=1}^{n} B(\boldsymbol{V}_i, \boldsymbol{\beta}(t)) \left[\frac{I(T_i^* > t)}{\hat{S}_C(t|Z_i, \boldsymbol{V}_i)} \exp(\Theta_z(t) Z_i + \Theta_m(t) M_i) - \tilde{G}(\boldsymbol{V}_i, t) \right] \tag{5.6}$$

To estimate $\Theta_z(t)$ and $\Theta_m(t)$, we can use one of the following two methods: The first solves two sets of estimating equations together by either iteratively updating $\boldsymbol{\Theta}(t)$ and

$\boldsymbol{\beta}(t)$ or directly solving them simultaneously. The second uses an one step estimator, which means we first solve $\Theta^{(0)}$ by assuming $\boldsymbol{\beta}(t) = 0$, then run a regression to solve $\boldsymbol{\beta}(t)$ treating $\boldsymbol{\Theta}^{(0)}$ as known and finally update $\boldsymbol{\Theta}$ with the estimated $\boldsymbol{\beta}(t)$. The second approach is computationally simple and robust in most situations. Also, it does not require the identifiability for $\boldsymbol{\beta}(t)$ to yield a consistent estimator for $\boldsymbol{\Theta}$. However, the analytic format asymptotic variance is not easy to compute by the second approach and the bootstrap will be used for this estimator. In contrast, the first method has good theoretical asymptotic behavior as it can be viewed as M-estimator and the standard sandwich formula applies to its variance computation. However, the solution might not be unique with finite sample size and the iterative algorithm might not converge. Also it should be pointed out that an incorrect working model might decrease the estimation efficiency and we have not found an acceptable way to select models without potential harm to the estimation.

5.2.3 Asymptotic Theorems

For simplicity, we assume the model $Z|V$ is known by design and will not take into consideration for the variation in the estimation of parameters involved in that model. We give the asymptotic results for our proposed estimator in the following theorems.

Theorem 2:

We denote the true value of parameters $\boldsymbol{\Theta} = (\Theta_z(t), \Theta_m(t))^T$ by $\boldsymbol{\Theta}^0(t)$. Under the assumption of identifiability and regularity condition as for Theorem 1, we have

$$\sqrt{n}\left(\hat{\boldsymbol{\Theta}}(t) - \boldsymbol{\Theta}^0(t)\right) \longrightarrow_d GP(\mathbf{0}, K)$$

where GP means gaussian process and the covariance process $K(t_1, t_2) = E\psi_i(t_1)\psi_i(t_2)^T$. Here $\psi_i(t)$ is the influence function derived from estimating equation 5.3 and estimating equation for censoring parameters.

Proof:

Here we prove the asymptotic linearity of the estimator for a fixed time point t and thus we simply denote $\boldsymbol{\Theta}(t)$ by $\boldsymbol{\Theta}$ and the true value as $\boldsymbol{\Theta}^0$. As the estimation of the survival distribution of C does not depend on $\boldsymbol{\Theta}$ and we assumed that a regular asymptotic linear

(RAL) estimator is used. By definition of RAL, we have

$$\sqrt{n}(\hat{S}_C(t|z,m,\boldsymbol{v}) - S_C(t|z,m,\boldsymbol{v})) = n^{-1/2}\sum_i \tilde{\phi}_i(t|z,m,\boldsymbol{v}) + o_p(1).$$

By the delta method and the assumption that $S_C(t)$ is bounded away from 0, we have

$$\begin{aligned}\sqrt{n}(\hat{S}_C^{-1}(t|z,m,\boldsymbol{v}) - S_C^{-1}(t|z,m,\boldsymbol{v})) &= -n^{-1/2}\sum_i S_C^{-2}(t|z,m,\boldsymbol{v})\phi_i(t|z,m,\boldsymbol{v}) + o_p(1)\\ &= n^{-1/2}\sum_i \phi_i(t|z,m,\boldsymbol{v}) + o_p(1).\end{aligned}$$

For simplicity, we denote the estimating equation (5.3) for $\boldsymbol{\Theta}$ as below:

$$n^{-1}\sum_i \hat{S}_C^{-1}(t|Z_i,M_i,\boldsymbol{V}_i)F_i(\boldsymbol{\Theta})$$

where $F_{ij}(\boldsymbol{\Theta}) = F(T^*_{ij}, Z_{ij}, M_{ij}, \boldsymbol{V}_{ij}, \boldsymbol{\Theta})$. Also, we denote $\hat{S}_C^{-1}(t|Z_i,M_i,\boldsymbol{V}_i)$ as $\hat{S}_{Ci}$ and $S_C^{-1}(t|Z_i,M_i,\boldsymbol{V}_i)$ as S_{Ci} for notation simplicity.

We will first prove the consistency of $\hat{\boldsymbol{\Theta}}$.

$$\begin{aligned}0 &= n^{-1}\sum_i \hat{S}_{Ci}^{-1}(t)F_i(\hat{\boldsymbol{\Theta}})\\ &= n^{-1}\sum_i S_{Ci}^{-1}(t)F_i(\hat{\boldsymbol{\Theta}}) + n^{-1}\sum_i[\hat{S}_{Ci}^{-1}(t) - S_{Ci}^{-1}(t)]F_i(\hat{\boldsymbol{\Theta}})\end{aligned}$$

Since $\hat{S}_C^{-1}(t|z,m,\boldsymbol{v}) - S_C^{-1}(t|z,m,\boldsymbol{v})$ converge to 0 uniformly over t, z, m and $\boldsymbol{v}$ and $F_{ij}(\hat{\boldsymbol{\Theta}})$ is bounded in probability, we know that the second term converge to 0 in probability. By regularity condition, we can take derivative over $F_i(\hat{\boldsymbol{\Theta}})$ and thus the first term become

$$\begin{aligned}n^{-1}\sum_i S_{Ci}^{-1}(t)F_i(\hat{\boldsymbol{\Theta}}) &= n^{-1}\sum_i S_{Ci}^{-1}(t)F_i(\boldsymbol{\Theta}^0) + n^{-1}\sum_i S_{Ci}^{-1}(t)\frac{\partial F_i(\boldsymbol{\Theta}^*)}{\partial \boldsymbol{\Theta}}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^0)\\ &= o_p(1) + \frac{\partial E(S_{Ci}^{-1}(t)F_i(\boldsymbol{\Theta}^*))}{\partial \boldsymbol{\Theta}}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^0)\end{aligned}$$

where $\boldsymbol{\Theta}^*$ is between $\boldsymbol{\Theta}^0$ and $\hat{\boldsymbol{\Theta}}$. Under the locally identifiability assumption, we have

$\frac{\partial E(S_{Ci}^{-1}(t)F_i(\mathbf{\Theta}^*))}{\partial \mathbf{\Theta}}$ is positive definite, Since the derivative is continuous respect to $\mathbf{\Theta}$, we have $\frac{\partial E(S_{Ci}^{-1}(t)F_i(\mathbf{\Theta}^*))}{\partial \mathbf{\Theta}}$ is positive definite when $\mathbf{\Theta}^*$ is within certain range of $\mathbf{\Theta}^0$. So we have there exist a sequence of root $\hat{\mathbf{\Theta}}$ such that $\hat{\mathbf{\Theta}} - \mathbf{\Theta}^0 = o_p(1)$. Or if we use the assumption of global identifiability, the derivative is positive definite for all $\mathbf{\Theta}$ and thus we have $\hat{\mathbf{\Theta}} - \mathbf{\Theta}^0 = o_p(1)$. So we obtain the consistency of $\hat{\mathbf{\Theta}}$. Note that $\hat{S}(t|z, m, \boldsymbol{v})$ and $E_n(F_i(\mathbf{\Theta}))$ are $\sqrt{n}-$consistency, so for any $d < 1/2$ $n^d(\hat{\mathbf{\Theta}} - \mathbf{\Theta}^0) = o_p(1)$.

Now we find the asymptotic expansion of $\mathbf{\Theta}$.

$$\begin{aligned} 0 &= n^{-1/2}\sum_i \hat{S}_{Ci}^{-1}(t)F_i(\hat{\mathbf{\Theta}}) \\ &= n^{-1/2}\sum_i S_{Ci}^{-1}(t)F_i(\mathbf{\Theta}^0) + n^{-1/2}\sum_i S_{Ci}^{-1}(t)\frac{\partial F_i}{\partial \mathbf{\Theta}}(\hat{\mathbf{\Theta}} - \mathbf{\Theta}^0) \\ &\quad + n^{-1/2}\sum_i [\hat{S}_{Ci}^{-1}(t) - S_{Ci}^{-1}(t)]F_i(\hat{\mathbf{\Theta}}) + o(\sqrt{n}\|\hat{\mathbf{\Theta}} - \mathbf{\Theta}^0\|) \end{aligned}$$

Here we just need to expand the third term as below

$$\begin{aligned} & n^{-1/2}\sum_i [\hat{S}_{Ci}^{-1}(t) - S_{Ci}^{-1}(t)]F_i(\hat{\mathbf{\Theta}}) \\ = \; & n^{-1/2}\sum_i [\hat{S}_{Ci}^{-1}(t) - S_{Ci}^{-1}(t)]F_i(\mathbf{\Theta}^0) + o_p(1) \\ = \; & n^{-3/2}\sum_i \sum_l [\phi_l(t|Z_i, M_i, \boldsymbol{V}_i)]F_i(\mathbf{\Theta}^0) + o_p(1) \\ = \; & n^{-1/2}\sum_l [n^{-1}\sum_i \phi_l(t|Z_i, M_i, \boldsymbol{V}_i)F_i(\mathbf{\Theta}^0)] + o_p(1) \end{aligned}$$

Let $E[\phi_l(t|Z_i, M_i, \boldsymbol{V}_i)F_i(\mathbf{\Theta}^0)] = \psi_l$, then we have the quantity above is

$$n^{-1/2}\sum_l \psi_l + o_p(1)$$

With the identification assumption, we know that

$$E[S_{Ci}^{-1}(t)\frac{\partial F_i}{\partial \mathbf{\Theta}}] = [\frac{\partial E S_{Ci}^{-1}(t)F_i}{\partial \mathbf{\Theta}}] \equiv \mathcal{A}$$

is positive definite, so we have

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}) &= \mathcal{A}^{-1}(n^{-1/2}\sum_i[S_{Ci}^{-1}(t)F_i(\boldsymbol{\Theta})+\psi_i]+o_p(1)\\ &= n^{-1/2}\sum_i\Psi_i\end{aligned}$$

where $\Psi_i = S_{Ci}^{-1}(t)F_i(\boldsymbol{\Theta})+\psi_i$. So we have proved that our estimator is asymptotic linear at each time point t. For any finite time points, the $o_p(1)$ term in the proof above will be uniform and thus we can have that for any finite dimension, the distribution is joint normal. Then it follows that we have $\boldsymbol{\Theta}(t), t\in[t_1,t_2]$ converge to a Gaussian Process.

The variance covariance matrix process can be estimated by its empirical version, given by

$$\hat{K}(t_1,t_2) = n^{-1}\sum_i\hat{\psi}_i(t_1)\hat{\psi}_i(t_2),$$

where $\hat{\psi}_i(t)$ can be obtained by replace parameters in $\psi_i(t)$ by their consistent estimators. We argue that $\hat{K}$ uniformly converges to K in probability using Glivento-Cantelli Theory.

Theorem 3:

We denote the true value of parameters $\boldsymbol{\Theta} = (\Theta_z(t),\Theta_m(t))^T$ by $\boldsymbol{\Theta}^0(t)$. We assume under the working model, there is unique solution $\boldsymbol{\beta}^0(t)$ to $EG_2(\boldsymbol{\beta},\boldsymbol{\theta}_0)=\mathbf{0}$. Under the assumption of identifiability and regularity condition, we have

$$\sqrt{n}\begin{pmatrix}\hat{\boldsymbol{\Theta}}(t)-\boldsymbol{\Theta}^0(t)\\ \hat{\boldsymbol{\beta}}(t)-\boldsymbol{\beta}^0(t)\end{pmatrix}\longrightarrow_d GP(\mathbf{0},K)$$

where $K(t_1,t_2) = E\psi_i(t_1)\psi_i(t_2)^T$ and $\psi_i(t)$ is the influence function derived from estimating equations 5.5-5.6 and estimating equation for censoring parameters.

Proof:

The proof is similar to that for theorem 2 and we will just list the general steps here. For

notation simplicity, we denote the estimating equations 5.5-5.6 as

$$\begin{aligned} 0 &= \sum_{i=1}^{n} S_{1i}(\boldsymbol{\Theta}, \boldsymbol{\beta}, \hat{S}(t|z)) \\ 0 &= \sum_{i=1}^{n} S_{2i}(\boldsymbol{\Theta}, \boldsymbol{\beta}, \hat{S}(t|z)) \end{aligned}$$

Since we have shown the estimator in theorem 2, denoted as $\hat{\boldsymbol{\Theta}}^{(0)}$ consistent to $\boldsymbol{\Theta}$, so we have

$$\sum_{i=1}^{n} S_{1i}(\hat{\boldsymbol{\Theta}}^{(0)}, \boldsymbol{\beta}, \hat{S}(t|z)) \longrightarrow_p 0$$

for any $\boldsymbol{\beta}$. We also have

$$\sum_{i=1}^{n} S_{2i}(\boldsymbol{\Theta}, \boldsymbol{\beta}, \hat{S}(t|z)) \longrightarrow_p ES_{2i}(\boldsymbol{\Theta}^0, \boldsymbol{\beta}, S(t|z)).$$

Under following additional assumptions:

- For the $A(z, v, t)$ we choose, equation 5.3 is bounded in probability for all $\boldsymbol{\Theta}$;
- The joint density of observed variables have third order partial derivatives for both $\boldsymbol{\Theta}$ and $\boldsymbol{\beta}$;
- The estimating equation for $\boldsymbol{\beta}$ have unique solution under $\boldsymbol{\Theta}^0$;

we know that $\hat{\boldsymbol{\beta}}^{(1)} \longrightarrow_p \boldsymbol{\beta}^0$. Using same argument, we can show for k-th iteration, $\hat{\boldsymbol{\Theta}}^{(k)}$ and $\hat{\boldsymbol{\beta}}^{(k)}$ consistent to their true values. So if the limiting exist, then the final estimator $\hat{\boldsymbol{\Theta}}$ and $\hat{\boldsymbol{\beta}}$ is consistent estimator and jointly solve estimating equations (6) and (7). Then the asymptotic normality and linear expansion can be obtained using standard Taylor expansion as used for sandwich estimator.

The variance covariance matrix process can be estimated by its empirical version, given

by

$$\hat{K}(t_1, t_2) = n^{-1} \sum_i \hat{\psi}_i(t_1)\hat{\psi}_i(t_2)$$

where $\hat{\psi}_i(t)$ can be obtained by replace parameters in $\psi_i(t)$ by their consistent estimator. By empirical process theory, we know $\hat{K}$ uniformly converges to K in probability. Due to the complex form of the influence function, in application, we recommend using a bootstrap procedure to compute SE unless the censoring is assumed to be completely random, in which case, the inverse weighting can be ignored and simple sandwich form can be obtained from the estimating equations 5.5-5.6.

5.2.4 Efficiency Considerations

Although we can choose any weight functions $\mathbf{A}(Z, \boldsymbol{V}, t) = (a_1(Z, \boldsymbol{V}, t), a_2(Z, \boldsymbol{V}, t))$ to obtain a consistent estimator, the variance for the corresponding estimators are very different when choosing different weight functions. When we do not estimate $\boldsymbol{\beta}$, the most efficient choice will be let

$$\mathbf{A}(Z, \boldsymbol{V}, t) = \tilde{\mathbf{A}}(Z, \boldsymbol{V}, t) - E[\tilde{\mathbf{A}}(Z, \boldsymbol{V}, t)|\boldsymbol{V}]$$

where

$$\begin{aligned}
\tilde{\mathbf{A}}(Z, \boldsymbol{V}, t) &= E\left[\frac{\partial I(T > t)\exp(\Theta_z(t)Z + \Theta_m(t)M)}{\partial \boldsymbol{\Theta}}|Z, \boldsymbol{V}\right] \\
&= E\left[(Z, M)^T I(T > t)\exp(\Theta_z(t)Z + \Theta_m(t)M)|Z, \boldsymbol{V}\right] \\
&= E\left[(Z, M)^T \exp(-G(\boldsymbol{V}, \boldsymbol{U}))|Z, \boldsymbol{V}\right].
\end{aligned}$$

Though we know that Z and $\boldsymbol{U}$ are independent conditional on $\boldsymbol{V}$, we don't have an explicit form for the correlation between Z and $\boldsymbol{U}$ conditional on Z and $\boldsymbol{V}$. Thus, we will not able to know the optimal weight in this case. However, one may use the estimated $\boldsymbol{\Theta}$ to update the weight by running a regression model for $(Z, M)^T I(T > t)\exp(\Theta_z(t)Z + \Theta_m(t)M)$ on

Z and $\boldsymbol{V}$.

Another aspect to be considered is the choice of model form for $\Theta_z(t)$ and $\Theta_m(t)$. For illustration purpose, we assume $\Theta_z(t) = \theta_z t$ and $\Theta_m(t) = \theta_m t$. Our goal is to estimate $\boldsymbol{\theta} = (\theta_z, \theta_m)^T$. It is obvious that we can choose any t_0 and compute $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\Theta}}(t_0)/t_0$. So the first question is how to choose t_0 to make most efficiency estimation. As will shown in the simulation, this can be done by separating the data into two part, where part of the data will be used to find the best t_0 that minimize the estimated variance for $\boldsymbol{\theta}$. Another possibility is to use several time points for analysis, we can run a regression model for $\hat{\boldsymbol{\Theta}}(t)$ on t with weight of the inverse of $var(\hat{\boldsymbol{\Theta}}(t))$ or we can use the weighted average estimating equation as below with arbitrary weight function $w(t)$:

$$\sum_i \int_0^\tau w(t) S_i(t) dt = 0.$$

Now consider the regression calibration method, which entails replacing Z in the estimating equations above by $\hat{Z} = E[Z|Q, \boldsymbol{V}]$. The right hand side can be computed by either $E[Z|Q, \boldsymbol{V}; \hat{\alpha}, \hat{\gamma}]$ with the parameters estimated from likelihood estimator as in chapter 3 or via a working model for $E[Z|Q, \boldsymbol{V}]$ and run regression of W on Q and $\boldsymbol{V}$ directly.

Since the above regression calibration method can be considered as a first order approximation, we further consider an expected estimating equation method. We can estimate the distribution of $Z|Q, \boldsymbol{V}$ by different methods. Assuming a joint normal distributon of $Z, Q|\boldsymbol{V}$, we obtain the following form, which can also be view as the second order approximation if the normality assumption does not hold.

$$0 = \sum_{i=1}^{n} \{A(Z_i, \boldsymbol{V}_i, t) \left[\frac{I(T_i^* > t)}{\hat{S}_C(t|Z_i, \boldsymbol{V}_i)} \exp(\Theta_z(t) Z_i + \Theta_m(t) \hat{M}_i + \Theta_m(t)^2 \widehat{Var}[M|Q, Z, \boldsymbol{V}]/2) - \tilde{G}(\boldsymbol{V}_i, t) \right] \}. \tag{5.7}$$

$$0 = \sum_{i=1}^{n} B(\boldsymbol{V}_i, \boldsymbol{\beta}(t)) \left[\frac{I(T_i^* > t)}{\hat{S}_C(t|Z_i, \boldsymbol{V}i)} \exp(\Theta_z(t) Z_i + \Theta_m(t) \hat{M}_i + \Theta_m(t)^2 \widehat{Var}[M|Q, Z, \boldsymbol{V}]/2) - \tilde{G}(\boldsymbol{V}_i, t) \right] \tag{5.8}$$

5.3 Simulation

In this section, we show the small sample performance of the estimators discribed above. And we want to compare these estimators with the traditional regression method under both the settings where sequential ignorability does or does not hold. We describe the parameter setting and the result below and in the tables, we refer our model as the Robust method to reflect its robustness to the covariate effect misspecification and certain type of unmeasured confounder. The model we considered is a binary intervention Z and a continuous mediator M follow the model:

$$E(M|Z,V,U) = \gamma_z Z + \gamma_v V + \gamma_{zv} ZV + \gamma_u U,$$

and the outcome follow a survival model with cumulative hazard function

$$\Lambda(z,m,t) = \beta_v V_i t + \theta_z zt + \theta_m mt + \beta_u U_i t, \tag{5.9}$$

where V follows standard uniform distribution and U is unmeasured confounder which is independent uniform distributed over 0 to 1. Here we use uniform distribution rather than the standard normal distribution to automatically satisfy the positivity requirement for the hazard function. We generate ε from an exponential distribution with rate 1 and the residual for M is set to be have variance 0.1. We set other parameters as $\theta_z = 1$,$\theta_m = 1$, $\gamma_z = 0$,$\gamma_v = 0$,$\gamma_{zv} = 1$, $\gamma_u = 0.15$. We choose $\beta_v = 0$ and $\beta_v = 0.2$ to explore the case with and without covariate effect. We choose $\beta_u = 0$ and $\beta_u = 2$ for the case that sequential ignorability assumption holds and does not hold. We use both linear (wrong) and exponential (correct) working model for $\tilde{G}(V)$. The censoring rate is set to be 30% and the censoring following an exponential distribution that are independent of other covariates. When we use certain workin model $G(\boldsymbol{V};\boldsymbol{\beta},t)$, we use the estimator from 1 iteration if the derivative of the joint estimating equation is close to singular or the iteration procedure does not converge. In the simulation, we choose the sample size to be 1000 and 4000 and run 1000 simulations. For a traditional additive hazard model analysis, we use the estimator from Lin and Ying [26]. We choose the time point $t = 0.4$ in this simulation. The expected event rate

before $t = 0.4$ is about 10%. For each simulation, we use Bootstrap with 1,000 replications to estimate SE and then construct 95% CI with asymptotic normal approximation. The detailed simulation results are given in Table 5.1.

n	Method	Working Model	Bias	MSE	95% CR	Bias	MSE	95% CR
$\beta_u = 0$, $\beta_v = 0$				θ_z			θ_m	
1000	Proposed	Zero	0.00	0.12	95%	0.01	0.48	94%
		Linear	0.00	0.04	95%	0.00	0.16	95%
	Traditional	NA	0.00	0.01	96%	0.00	0.04	95%
4000	Proposed	Zero	0.01	0.03	93%	-0.01	0.12	94%
		Linear	0.00	0.01	95%	0.00	0.04	96%
	Traditional	NA	0.00	0.003	93%	0.00	0.01	91%
$\beta_u = 0$, $\beta_v = 0.2$				θ_z			θ_m	
1000	Proposed	Zero	0.00	0.13	96%	0.01	0.51	95%
		Linear	0.00	0.05	95%	0.00	0.19	96%
	Traditional	NA	0.01	0.01	95%	0.00	0.06	95%
4000	Proposed	Zero	0.00	0.03	93%	0.00	0.13	94%
		Linear	0.00	0.01	94%	0.01	0.05	96%
	Traditional	NA	0.00	0.004	91%	0.01	0.02	90%
$\beta_u = 2$, $\beta_v = 0$				θ_z			θ_m	
1000	Proposed	Zero	0.01	0.22	96%	0.01	0.83	95%
		Linear	0.01	0.14	96%	0.00	0.49	96%
	Traditional	NA	-0.52	0.32	35%	1.12	1.42	24%
4000	Proposed	Zero	0.00	0.06	92%	0.01	0.21	94%
		Linear	0.00	0.04	96%	0.01	0.12	95%
	Traditional	NA	-0.52	0.29	0%	1.12	1.29	0%
$\beta_u = 2$, $\beta_v = 0.2$				θ_z			θ_m	
1000	Proposed	Zero	0.01	0.24	96%	0.02	0.89	95%
		Linear	0.00	0.16	96%	0.01	0.55	96%
	Traditional	NA	-0.51	0.32	42%	1.16	1.45	30%
4000	Proposed	Zero	0.00	0.06	95%	0.01	0.21	95%
		Linear	0.00	0.04	97%	0.02	0.12	97%
	Traditional	NA	-0.52	0.28	1%	1.12	1.30	0%

Table 5.1: Comparison of different estimators under scenario with and without sequential ignorability, $t_0 = 0.4$ from 1,000 simulations

As we fixed t in different situation, the result from different β_u and β_v are not directly comparable. So we only compare different estimators within each setting. From the simulation result in table 5.1, we find that when sequential ignorability assumption does not hold, the traditional regression method yield biased result and low coverage rate for its 95% nominal confidence interval. On the contrary, the proposed method using zero and linear

working model yield unbiased result and correct coverage. It is interesting to find that the exponential working model gives a bad result. It seems due to the fact that the estimating equations have partial derivative that is close to singular when exponential working model is used. Compared with zero working model, linear working model often have gain in efficiency whether the true model is zero or not. We also try to adjust using different functional forms, but they are not as stable as the linear one and often show large variance. In general, we recommend use the zero working model when there are sufficient data. The proposed method give smaller MSE compared to the traditional regression method, especially with sample size 4000. When the sequential ignorability assumption holds, both proposed method and traditional regression gives approximately unbiased result with correct coverage rate. The proposed method lose some efficiency compared to the traditional regression method, this can be considered as the price to be paid in order to have estimators being robust to certain type of unmeasured confounders.

Next we assess different ways to combine information from different time points to estimate $\mathbf{\Theta}$. We just include results from situation where sequential ignorability assumption fail. For methods that combine information across t values, we combine θ estimates at different times linearly and weight them by the inverse of their variance.

From the result in Table 5.2, we can see that both methods have some improvement in the efficiency compared to use one time point only. The price is that the computational burden is much heavier if we want to combine information for many different time points. The combine method is more efficient compared to select the best time point t.

For the model with measurement error, we use the following models

$$\begin{aligned} Q &= 1.1M + 0.5V + 0.2Z + N(0, 0.2^2) \\ W &= M + N(0, 0.1^2) \end{aligned}$$

To make the rare disease assumption approximately hold, the baseline hazard is set to be 1 and the true parameters are set as $\theta_z = \theta_m = 0.1$. We have a total sample size of 10000 with 500 sample size where W is available. We compare the different methods in the following table. The result suggest that the proposed method has better coverage rate and less bias

n	Method	Working Model	Bias	MSE	95% CR	Bias	MSE	95% CR
$\beta_u = 2$, $\beta_v = 0$				θ_z			θ_m	
1000	Select t	Zero	-0.04	0.18	96%	0.03	0.65	96%
		Linear	-0.01	0.13	97%	-0.03	0.44	97%
	Combine t	Zero	-0.02	0.17	94%	0.01	0.63	91%
		Linear	-0.02	0.08	97%	-0.02	0.32	96%
4000	Select t	Zero	0.00	0.03	97%	0.01	0.11	99%
		Linear	0.00	0.03	95%	0.02	0.11	93%
	Combine t	Zero	0.01	0.03	93%	0.01	0.12	90%
		Linear	0.00	0.02	91%	0.01	0.07	93%
$\beta_u = 2$, $\beta_v = 0.2$				θ_z			θ_m	
1000	Select t	Zero	-0.03	0.18	96%	0.01	0.65	92%
		Linear	-0.02	0.13	92%	0.00	0.43	97%
	Combine t	Zero	-0.01	0.16	93%	-0.01	0.58	92%
		Linear	-0.01	0.07	94%	-0.04	0.28	95%
4000	Select t	Zero	0.01	0.04	97%	0.00	0.12	97%
		Linear	0.00	0.03	95%	0.04	0.11	95%
	Combine t	Zero	0.01	0.03	93%	0.01	0.12	91%
		Linear	0.00	0.02	94%	0.02	0.07	93%

Table 5.2: Comparing different ways to combine information for different time point t (0.3,0.4,0.5,0.6,0.7) from 1,000 simulations

t_0	Method	Working Model	Bias	MSE	95% CR	Bias	MSE	95% CR
$\beta_u = 0.2$, $\beta_v = 0.1$				θ_z			θ_m	
0.5	Proposed RC	Zero	-0.02	0.012	94%	0.04	0.047	94%
		Linear	-0.02	0.002	93%	0.04	0.009	93%
	Proposed Second Order RC	Zero	-0.02	0.012	94%	0.04	0.047	94%
		Linear	-0.02	0.002	93%	0.03	0.009	93%
	Traditional RC	NA	-0.05	0.003	2%	0.11	0.012	1%
$\beta_u = 0.2$, $\beta_v = 0.1$				θ_z			θ_m	
1	Proposed RC	Zero	-0.02	0.004	93%	0.04	0.015	93%
		Linear	-0.02	0.001	92%	0.04	0.005	92%
	Proposed Second Order RC	Zero	-0.02	0.004	93%	0.04	0.015	93%
		Linear	-0.02	0.001	92%	0.04	0.005	92%
	Traditional RC	NA	-0.05	0.003	2%	0.11	0.012	1%

Table 5.3: Comparison of different estimators under scenario without sequential ignorability, from 1,000 simulations

than the traditional regression method.

5.4 Data analysis

To illustrate our proposed method, we applied our method to the Women's Health Initiative (WHI) data [23]. The DM component (recruited 1993-1998) of the WHI study is a randomized controlled trial, which assign 48,835 postmenopausal women to either low fat eating pattern (40%) or self-selected dietary behavior (60%)[42]. The intervention was applied group of 10-15 women and included both nutritional and behavioral component. The principal nutritional goal of the intervention is to reduce the energy from fat from baseline level of about 35% to 20%. Women who were on a low-fat diet (<32% estimated energy from fat) were excluded from the study. A secondary goal was to increase the daily vegetable and fruit intake. The average follow-up time is 8.5 years with disease outcomes self-reported semi-annually followed by physician adjudication. Previous study showed that the intervention group in the DM trial have a borderline significant lower breast cancer rate, especially for women having a relatively high proportion of energy from fat [37]. Also there is result suggesting that DM trial intervention women have lower total energy consumption, that is associated with lower cancer risk. So it is natural to ask the question whether the randomization effects on breast cancer risk are mediated by the year 1 total energy consumption after randomization. We exclude women with missing weight information at either 1 year or baseline. The sample size used for analysis is 44,551 with event rate 6.6%. As is suggested in the simulation, we use linear working model to adjust the baseline covariate. One problem with the data is that mediator is only available for those survive for at least one year. To overcome this, we can use the rare disease assumption and use year 1 as the new baseline and the intervention is still approximately randomized for those 1 year survivors. A better way is to consider time-varying effect and assume the mediator effect is 0 before year 1 and then we can choose $t_2 > t_1 > 1$ and obtain $\hat{\Theta}(t_2)$ and $\hat{\Theta}(t_1)$. Then have $\hat{\theta} = \frac{\hat{\Theta}(t_2)-\hat{\Theta}(t_1)}{t_2-t_1}$. We find that the time-independent effect assumption seems implausible in the WHI example and thus we focus on estimating the time-varying cumulative parameter $\boldsymbol{\Theta}(t)$. For illustration purpose, we choose $t = 6$ and we use the baseline BMI to construct estimating equation.

The analysis results are shown in table 5.4. From the results, we did not find significant

direct or mediator effects. This finding is consistent with our result in chap 2. Better self-reported data and strong instrumental variables are needed to decrease the variance of our proposed estimator and yield some signals for the effect of total energy expenditure.

	Direct Effect $\times 10^2$		Mediator Effect $\times 10^2$	
Method	Estimate	95% CI	Estimate	95% CI
Naive	0.2	(-0.1,0.4)	0.0	(-0.4, 0.4)
RC	0.1	(-0.1,0.4)	-0.0	(-2.9, 2.7)
Proposed Naive	0.2	(-7.6,7.9)	0.3	(-5.3, 6.0)
Proposed RC	-1.6	(-4.1,1.0)	10.6	(-3.2, 24.4)

Table 5.4: Data analysis for assessing whether DM intervention decrease breast cancer risk through decreasing total energy expenditure

5.5 Discussion

Here we would like to discuss the relationship between our proposed additive hazard model and the Cox model. Since we allow the effects to be time varying, when both Z and M are binary, the Cox model can be represented as our additive hazard model with an interaction term between Z and M. So we can assess the robustness to model specification by performing sensitivity analysis with certain type of interaction effects.

Another issue with the additive hazard model is the positivity restriction on the hazard function, but since we have an unspecified term $G(z, m, \boldsymbol{V}_i, \boldsymbol{U}_i, t)$, we will always find certain form to make the positivity hold. So our estimator may not as sensitive to occational large covariate values as the original additive hazard regression.

From the simulation, we have find that with moderate strong interaction between the covariate and intervention on mediator and the sequential ignorability holds, the proposed method has some loss of efficiency (variance ratio about 4:1) compared with the traditional regression method. In most real application, we can only have covariate have weak interaction with the intervention on mediator and thus will lead to much larger variance as shown in data analysis part. Advanced method that can combine several weak instrumental variable to a strong one is needed to allow broader use of the proposed method.

Chapter 6

EXTENSIONS

In this last chapter, we discuss some limitation of the current work and describe possible future extensions.

6.1 Handling internal time varying covariate

In our data example, the available total energy consumption history for a study subject can be very different depending on whether or not the study disease occurs, and thus this variable should be viewed as an 'internal' process [21]. However, our estimation method requires 'external process' [21]. So we first discuss how to approximate the internal process, so that it can be treated as external.

In this section, we show that under certain latent model, the interval observation process can be viewed as an shifted external process with a certain time period lag. Consider that we have a covariate process $Z(t)$, such as the total energy intake at follow-up time t. We might imagine the total energy intake decreases sharply before occurence of severe disease. This implies that we need treat $Z(t)$ as an internal process. One possibility is that we allow for the same phenomenon using an approximating external process $Z^*(t)$. Then we can consider a model $\lambda(t|\underline{Z}(\infty)) = \lambda(t|\underline{Z}(t-\delta))$ for some $\delta > 0$ for that external process. Here $\underline{Z}(\infty) = \{Z(t), t \in [0,\infty)\}$ is the whole process of Z. After we obtain the joint distribution of $(T, \underline{Z}(\infty))$, we define the internal process as $Z(t) = Z^*(t)I(t \leq T-\delta) + Z^*(t)\frac{T-t}{\delta}I(T-\delta < t \leq T)$. One disadvantage of such a procedure is that the covariate value at time t, $Z(t)$ will depend on $I(T > t+\delta)$. However given the $I(T > t+\delta)$ happened after observing $Z(t)$, this relationship cannot be causal.

Here we try to consider a latent health status variable $H(t)$ that leads to association between $I(T > t+\delta)$ and $Z(t)$. First consider discrete time model, assume we have $\delta = 1$

and the following models

$$\begin{aligned} Pr(T = t|T \geq t, \underline{Z}(t), \underline{H}(t)) &= \lambda(t|\underline{Z}(t), \underline{H}(t)) \\ Pr(Z(t) = z|\underline{Z}(t-1), \underline{H}(t), T \geq t) &= f(z|\underline{Z}(t-1), \underline{H}(t)) \\ Pr(H(t) = h|\underline{Z}(t-1), \underline{H}(t-1), T \geq t) &= g(h|\underline{Z}(t-1), \underline{H}(t-1)) \end{aligned}$$

Consider a special case as below:

$$\begin{aligned} \lambda(t|\underline{Z}(t), \underline{H}(t)) &= \lambda(t|Z(t-1), H(t)) \\ f(z|\underline{Z}(t-1), \underline{H}(t)) &= f_t(z|Z(t-1), H(t)) \\ g(h|\underline{Z}(t-1), \underline{H}(t-1)) &= g_t(h|Z(t-1), H(t-1)) \end{aligned}$$

Consider the extreme case that $H(t)$ is binary and that 1 means 'healthy' and 0 means 'unhealthy'. If for unhealthy people, we assume they will die in next year, i.e. $\lambda(t|Z(t-1), 0) = 1$. then we have

$$\begin{aligned} \lambda(t|\underline{Z}(t)) &= \lambda(t|\underline{Z}(t), H(t-1) = 1) \\ &= \int \lambda(t|Z(t-1), h) Pr(H(t) = h|H(t-1) = 1, Z(t), Z(t-1)) dh \\ &= \lambda(t|Z(t-1), 1) Pr(H(t) = 1|H(t-1) = 1, Z(t), Z(t-1), T \geq t) \\ &\quad + Pr(H(t) = 0|H(t-1) = 1, Z(t), Z(t-1), T \geq t) \end{aligned}$$

$$\begin{aligned} & Pr(H(t) = 1|H(t-1) = 1, Z(t), Z(t-1)) \\ = & \frac{Pr(Z(t)|H(t) = 1, H(t-1), Z(t-1)) Pr(H(t) = 1|H(t-1), Z(t-1))}{\sum_h Pr(Z(t)|H(t) = h, H(t-1), Z(t-1)) Pr(H(t) = h|H(t-1), Z(t-1))} \\ = & \frac{f_t(Z(t)|Z(t-1), 1) g_t(1|Z(t-1), 1)}{f_t(Z(t)|Z(t-1), 1) g_t(1|Z(t-1), 1) + f_t(Z(t)|Z(t-1), 0) g_t(0|Z(t-1), 1)} \\ = & F(Z(t), Z(t-1)) \end{aligned}$$

Thus, we have

$$\begin{aligned}\lambda(t|\underline{Z}(t)) &= \lambda(t|Z(t-1),1)F(Z(t),Z(t-1)) \\ &\quad +(1-F(Z(t),Z(t-1))) \\ &= G(Z(t),Z(t-1))\end{aligned}$$

However, the marginal model that integral over $Z(t)$ still has

$$\begin{aligned}\lambda(t|\underline{Z}(t-1)) &= \int[\lambda(t|Z(t-1),1)F(z,Z(t-1))+(1-F(z,Z(t-1)))] \\ &\quad \times[f_t(z|Z(t-1),1)g_t(1|Z(t-1),1)+f_t(z|Z(t-1),0)g_t(0|Z(t-1),1)]dz \\ &= \int(\lambda(t|Z(t-1),1)-1)f_t(z|Z(t-1),1)g_t(1|Z(t-1),1)dz \\ &\quad +\int[f_t(z|Z(t-1),1)g_t(1|Z(t-1),1)+f_t(z|Z(t-1),0)g_t(0|Z(t-1),1)]dz \\ &= \int\lambda(t|Z(t-1),1)f_t(z|Z(t-1),1)g_t(1|Z(t-1),1)dz \\ &\quad +\int f_t(z|Z(t-1),0)g_t(0|Z(t-1),1)dz \\ &= \lambda(t|Z(t-1),1)g_t(1|Z(t-1),1)+g_t(0|Z(t-1),1) \\ &= G(Z(t-1))\end{aligned}$$

So we have shown that under this special case, the marginal hazard only depends on $Z(t-1)$. This shows that if we model the internal process with a certain lag period, then the internal process can be viewed as part of an external process when constructing a partial likelihood function. We assume that there exists a certain model having marginal hazard in proportional hazards form [40].

6.2 Combine different source of data

The basic idea for combining two different sources of data is to make a plausible assumption to build some relationship between two sets of parameters in the two data sets. For example, we might consider the control group have the same parameters as the observational study for the self-reported dietary data model while allowing different parameters in disease risk models. As long as the parameters are identifiable in each data set, after adding some

constraints they will still be identifiable. The number of constraints to add depends on insight concerning the homogeneity of the two data sets. Increasing number of constraints will improve efficiency but make the estimation method less robust.

6.3 Time-varying covariates effect on exposure or measurement error

When we allow BMI/weight to affect future dietary profile, the natural direct and indirect effect will not be well defined. Instead of using the potential outcome $\underline{M(t, \underline{z(\infty)})}$, and needs two sequence of potential outcomes $M(t, \underline{z(t)m(t-)})$ and $Z(t, \underline{z(t-)m(t-)})$. To estimate the direct or mediator effect comparing certain profiles, one needs to adjust for preceding variables at each time point t to make groups comparable, which can be realized by weighting by the inverse of the conditional density at each time point. In the chapters where we assume sequential ignorability, we can allow the measurement error depend on the time-varying quantities as long as there are some lag periods. However, for the chapter without sequential ignorability, we can only allow the systematic bias in the error-prone variable to be depend on baseline covariates. Corresponding methods that can handle a time-varying unmeasured confounder are not currently available.

6.4 Lag effects

Though the additive hazard model we present already consider the case where we have time lag effect, our current form of Cox model does not allow lag effect. We can introduce lag effect parameter τ_z, τ_m and write the model as

$$\begin{aligned}\lambda(t, r, \underline{z(\infty)}, \underline{m(\infty)}) &= H_0(t)\exp\{\beta_1(t)z(t-\tau_z) + \beta_2(t)m(t-\tau_m) \\ &\quad +\beta_3(t)M_i(0) + \beta_4(t)V_i + \beta_5(t)r\}\end{aligned}$$

When the direct effect and indirect effect have different lag time, the decomposition of the total effect will not have a convinient form even approximately.

Another important lag effects is how the bias in self-reported variables are affected by previous BMI profiles. Our current assumption is that it is only affected by previous time points, which is hard to hold in reality. Fortunately, in our data, the BMI profile does

not change dramatically and control for previous time seems enough to control most of the systematic bias in self-reported data. However, future work is needed to obtain a general framework that can handle any type of lag effects in self-reported bias.

6.5 Validation set only available for later and short period

The asymptotic results developed in the main part assume the biomarker process can be measured at any time point with positive probability. However, this is not the case for the WHI data where the biomarker process is only measured some years after cohort enrollment. In this case, we can only have those $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\theta}}$ converge to some value biased value $\boldsymbol{\alpha}^*$, $\boldsymbol{\gamma}^*$ and $\boldsymbol{\theta}^*$, which can be close to the true value of those parameters under rare disease assumption. However, estimators from different estimation methods will converge to corresponding different limits and the $\boldsymbol{\beta}^*$ might then be sensitive to the bias in these measurement error parameters. Sensitivity analyses for the selection bias for the biomarker subsample as well as the measurement model assumption or the survival model assumption fails can be helpful. One might use simulations with estimated parameters to quantify the possible biases. However, we do not have formal test for goodness of fit of certain of the models. In the future, we need to develop model selection procedures and can apply some targeted inference procedure [64] here.

6.6 Flexible model without sequential ignorability

Currently, we did not obtain a good method for estimating controlled effects under general relative risk model when the sequential ignorability assumption does not hold. Also, due to the need of finding baseline covariate to serve as weight function in estimating equations, we can not allow too many interactions between the covariates and the intervention or mediator. However, if our mediator is continuous variable, such linear effect that does not depend on other covariates might not be true, which limits the application of our method. For survival model part, although we studied the Cox model, additive hazard model and linear intensity model, there are other useful models such as the linear transformation model [73], mean residual life model [7]. Methods for calibrating measurement error for those models will be considered in the future.

6.7 Other related issues

In this dissertation, we focused on the regression calibration method. However, there are other possible ways to handle measurement error. For example, we can use the joint likelihood approach to obtain the nonparametric maximum likelihood estimator (NPMLE). Specifically, one can use the model in Chap 2.1, if Q is measured at time t_1 and t_2 and W is measured at time t_2 for a subset. Then the joint likelihood is proportional to

$$\prod_i \int Pr(V = V_i, \boldsymbol{\Theta}) \times Pr(Z = Z_i | V = V_i, \boldsymbol{\Theta}) \times Pr(M = m | V = V_i, \boldsymbol{\Theta})$$
$$\times Pr(Q_1 = Q_{1i} | M = m, V = V_i, Z = Z_i, \boldsymbol{\Theta}) \times Pr(Q_2 = Q_{2i} | M = m, V = V_i, Z = Z_i, \boldsymbol{\Theta})$$
$$\times Pr(W = W_i | M = m, \boldsymbol{\Theta}) \times \lambda(Y_i, Z_i, m | V)^{\Delta_i} \times \exp\{-\Lambda(Y_i, Z_i, m | V)\} dm.$$

The joint modeling approach maximizes the above quantity to obtain an estimate of $\boldsymbol{\Theta}$, $\boldsymbol{\beta}$ and $\lambda_0(Y_i)$ for those $\Delta_i = 1$. This approach will be computationally intensive with large sample size though it can yield consistent estimator for $\boldsymbol{\Theta}$, $\boldsymbol{\beta}$ under certain regularity conditions. A more complicated form can be written for the model with time-varying mediator M, but the estimation will be more difficult.

Currently, we only consider one mediator process in each analysis, however, besides the total energy intake, we might interested in the joint effect of different micronutrient consumption. Although currently we did not have good biomarker for those components, the search for metabolite biomarkers for these variables is in progress. So in the future, we will need methods that can handle joint estimation of multiple mediator process and probably high-dimensional longitudinal variables. How to define and estimate causal mediation effect in such analyses is still a open question.

As we noticed in our example, missing data is a problem. Currently, we either use the complete data to perform the analysis or simply use the inverse probability weighting with baseline covariates. As an extension, one might use inverse probability weighting including time-varying covariate or multiple imputation to handle missing data problem under a missing at random assumption [48]. However, if the missing also depend on the underlying health status and thus is nonignorable with the measured variables, then joint modeling of

missingness, measurement error, and survival rate might be required.

Although not likely to happen in our application with chronic diseases, if our outcome is time to infection of infectious diseases, then the SUTVA assumption might not hold. In that case, we need to consider the interference between the subject. Instead of denote potential outcome M_i^z, we need a vector $z = (z_i, z_{-i})$ for the index of potential outcomes. Under some simple model structrue, for example, if we assume the potential outcome is just a function of z and $\sum z_{-i}$, then we believe our method can be extended by using multivariate analysis.

6.8 Summary

To summarize, we proposed a general method to calibrate longitudinal self-reported exposures and/or mediators using biomarker subsamples and use these calibrated estimates to study the association of these variables to disease risks under different hazard regression models. Asymptotic theory and simulation are developed to show the large and moderate sample size property of our proposed estimators. We showed of assumptions under which causal inferences can be made including some instances when there are certain kind of unmeasured confounders. Using simulation, we showed that kind of data required to understand the causal mechanism when there are two mismeasured variables (exposure and mediator). Applying our method to WHI data, we showed that total energy expenditure are associated with several kinds of cardiovascular diseases, cancer and diabetes. Besides effect through total energy expenditure, the decrease of physical activity related energy expenditure can lead to higher risk of several kinds of cardiovascular disease, cancer and diabetes. For diabetes, we use the longitudinal BMI data to distinguish the BMI effect as mediator and confounder for the association between the total energy consumption and disease risk.

The work presented here mainly focuses on a relatively ideal situation with all models correctly specified. The general sensitivity analysis regarding different kind of departures from each specific assumption will need to be explored in future. Also the methods assumed joint normality of the self-reported variables and true underlying variables. Methods for extending the current work to models allowing binary or skewed variables will also be of interest.

BIBLIOGRAPHY

[1] World health organization: Diet, nutrition and the prevention of chronic diseases: report of a joint who/fao expert consultation. *WHO Technical Report*, 916:Geneva: World Health Organization, 2003.

[2] B. Ainsworth, W. Haskell, and M. Whitt. Compendium of physical activities: an update of activity codes and met intensities. *Medicine & Science in Sports & Exercise*, 64:348S–354S, 2000.

[3] J. M. Albert and S. Nelson. Generalized causal mediation analysis. *Biometrics*, 67:1028–1038, 2011.

[4] P. S. Albert and J. H. Shih. On estimating the relationship between longitudinal measurements and time-to-event data using a simple two-stage procedure. *Biometrics*, 67:983–987, 2011.

[5] S. A. Bingham. Urine nitrogen as a biomarker for the validation of dietary protein intake. *Journal of Nutrition*, 133(133):suppl:9215–9245, 2003.

[6] T. X. Cai, T. A. Gerds, Y. Zheng, and J. Chen. Robust prediction of t-year survival with data from multiple studies. *Biometrics*, 67:436–444, 2011.

[7] Y. Q. Chen and S. Cheng. Semiparametric regression analysis of mean residual life with censored survival data. *Biometrika*, 92:19–29, 2005.

[8] Y. Y. Chi and J. G. Ibrahim. Bayesian approaches to joint longitudinal and survival models accommodating both zero and nonzero cure fractions. *Statistica Sinica*, 17:445–462, 2007.

[9] S. R. Cole, H. Chu, and S. Greenland. Multiple imputation for measurement error correction. *International Journal of Epidemiology*, 35:1074–1081, 2006.

[10] J. D. Curb, A. Mctiernan, S. R. Heckbert, C. Kooperberg, J. Stanford, M. Nevitt, K. C. Johnson, L. Proulx-Burns, L. Pastore, M. Criqui, and S. Daugherty. Outcomes ascertainment and adjudication methods in the womens health initiative. *Annals of Epidemiology*, 13:S122–S128, 2003.

[11] M. R. Elliott, T. E. Raghunathan, and Y. Li. Bayesian inference for causal mediation effects using principal stratification with dichotomous mediators and outcomes. *Biostatistics*, 11:353–372, 2010.

[12] "World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR)". Food, nutrition, physical activity, and the prevention of cancer: a global perspective. pages Washington, DC: American Institute for Cancer Research, 2007.

[13] R. Gallop, D. S. Small, J. Y. Lin, M. R. Elliott, M. Joffe, and T. R. TenHave. Mediation analysis with principal stratification. *Statistics in Medicine*, 28:1108–1130, 2009.

[14] N. K. Horner, J. W. Lampe, R. E. Patterson, M. L. Neuhouser, S. A. Beresford, and R. L. Prentice. Indirect calorimetry protocol development for measuring resting metabolic rate as a component of total energy expenditure in free-living postmenopausal women. *Journal of Nutrition*, 131:2215–2218, 2001.

[15] Y. Huang and R. L. Prentice. Measurement error modeling and nutritional epidemiology association analyses. *Canadian Journal of Statistics*, 39:498–509, 2011.

[16] Y. Huang and C. Y. Wang. Cox regression with accurate covariates unascertainable a nonparametric correction approach. *Journal of the American Statistical Association*, 95:1209–1219, 2000.

[17] K. Imai, L. Keele, and D. Tingley. A general approach to causal mediation analysis. *Psycholological Methods*, 15(4):309–34, 2010.

[18] K. Imai, L. Keele, and T. Yamamoto. Identification, inference, and sensitivity analysis for causal mediation effects. *Statistical Science*, 25:51–71, 2010.

[19] M. Johnson-Kozlow, C. L. Rock, E. A. Gilpin, K. A. Hollenbach, and J. P. Pierce. Validation of the whi brief physical activity questionnaire among women diagnosed with breast cancer. *American Journal of Health Behavior*, 31:193–202, 2007.

[20] C. Ju and Z. Geng. Criteria for surrogate end points based on causal distributions. *Journal of the Royal Statistical Society Series B*, 72:129–142, 2010.

[21] J. D. Kalbfleisch and R. L. Prentice. *The statistical analysis of failure time data*. Wiley, New York: NY, 2002.

[22] T. Lange and J. V. Hansen. Direct and indirect effects in a survival context. *Epidemiology*, 22:575581, 2011.

[23] R. D. Langer, E. White, C. E. Lewis, J. M. Kotchen, S. L. Hendrix, and M. Trevisan. The women's health initiative observational study: baseline characteristics of participants and reliability of baseline measures. *Annals of Epidemiology*, 13:S107–S121, 2003.

[24] Y. Li and L. Ryan. Inference on survival data with covariate measurement error-an imputation approach. *Scandinavian Journal of Statistics*, 33:169–190, 2006.

[25] X. Liao, D. M. Zuker, and D. Spiegelman. Survival analysis with error-prone time-varying covariates a risk set calibration approach. *Biometrics*, 67:50–58, 2011.

[26] D. Y. Lin and Z. Ying. Semiparametric analysis of the additive risk model. *Biometrika*, 81:61–71, 1994.

[27] K. L. Margolis, L. Qi, R. Brzyski, D. E. Bonds, B. V. Howard, S. Kempainen, L. Simin, J. G. Robinson, M. M. Safford, L. Tinker, L. S. Phillips, and Women Health Initiative Investigators. Validity of diabetes self-reports in the women's health initiative: comparison with medication inventories and fasting glucose values. *Clinical Trials*, 5:240–247, 2008.

[28] T. Martinussen, S. Vansteelandt, M. Gerster, and J. B. Hjelmborg. Estimation of direct effects for survival data by using the aalen additive hazards model. *Journal of the Royal Statistical Society Series B*, 73:773–788, 2011.

[29] S. E. Maxwell and Cole D. A. Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, 12:23–44, 2007.

[30] S. E. Maxwell, D. A. Cole, and M. A. Mitchell. Bias in cross-sectional analyses of longitudinal mediation: Partial and complete mediation under an autoregressive model. *Multivariate Behavioral Research*, 46:816–841, 2011.

[31] M. L. Neuhouser, C. Di, L. Tinker, C. Thomson, B. Sternfeld, Y. Mossavar-Rahmani, M. L. Stefanick, S. Sims, J. D. Curb, M. Lamonte, R. Seguin, K. C. Johnson, and R. L. Prentice. Physical activity assessment: biomarkers and self-report of activity-related energy expenditure in the whi. *American Journal of Epidemiology*, 167, year = 2008, pages = 1247-1259,.

[32] M. L. Neuhouser, L. Tinker, P. A. Shaw, D. Schoeller, S. A. Bringham, L. V. Horn, S. A. Beresford, B. Caan, C. Thomson, S. Satterfield, L. Kuller, G. Heiss, E. Smit, G. Sarto, J. Ockene, M. L. Stefanick, A. Assaf, and R. L. Runswick, S.and Prentice. Use of recovery biomarkers to calibrate nutrient consumption self-reports in the women's health initiative. *American Journal of Epidemiology*, 167:1247–1259, 2008.

[33] D. M. Parkin, S. R. Whelan, J. Ferlay, L. Teppo, and D. B. Thomas. *Cancer incidence in five continents*, volume 155. IARC Scientific Publication, IARC, LYON, 1998.

[34] R. E. Patterson, A. R. Kristal, L. F. Tinker, R. A. Carter, M. P. Bolton, and T. A. Collins. Measurement characteristics of the women's health initiative food frequency questionnaire. *Annals of Epidemiology*, 9:178–187, 1999.

[35] J. Pearl. on measurement bias in causal inference. *Technical Report*, R-357:425–432, 2010.

[36] R. L. Prentice. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 69:331–342, 1982.

[37] R. L. Prentice, B. Caan, R. T. Chlebowski, R. Patterson, L. H. Kuller, J. K. Ockene, K. L. Margolis, M. C. Limacher, J. E. Manson, L. M. Parker, E. Paskett, L. Phillips, J. Robbins, J. E. Rossouw, G. E. Sarto, J. M. Shikany, M. L. Stefanick, C. A. Thomson, L. V. Horn, M. Z. Vitolins, J. Wactawski-Wende, R. B. Wallace, S. Wassertheil-Smoller, E. Whitlock, K. Yano, L. Adams-Campbell, G. L. Anderson, A. R. Assaf, S. A. A. Beresford, H. R. Black, R. L. Brunner, R. G. Brzyski, L. Ford, M. Gass, J. Hays, D. Heber, G. Heiss, S. L. Hendrix, J. Hsia, F. A. Hubbell, R. D. Jackson, K. C. Johnson, J. M. Kotchen, A. Z. LaCroix, D. S. Lane, R. D. Langer, N. L. Lasser, and M. M. Henderson. Low-fat dietary pattern and risk of invasive breast cancer. *The Journal of the American Medical Association*, 295:629–642, 2006.

[38] R. L. Prentice, Y. Huang, L. H. Kuller, L. F. Tinker, L. V. Horn, M. L. Stefanick, G. Sarto, J. Ockene, and K. C. Johnson. Biomarker-calibrated energy and protein consumption and cardiovascular disease risk among postmenopausal women. *Epidemiology*, 22:170–179, 2011.

[39] R. L. Prentice, Y. Mossavar-Rahmani, Y. Huang, L. V. Horn, S. A. A. Beresford, B. Caan, L. Tinker, D. Schoeller, S. Bingham, C. B. Eaton, C. Thomson, K. C. Johnson, J. Ockene, G. Sarto, G. Heiss, and M. L. Neuhouser. Evaluation and comparison of food records, recalls, and frequencies for energy and protein assessment by using recovery biomarkers. *American Journal of Epidemiology*, 174:591–603, 2011.

[40] R. L. Prentice and S. G. Self. Asymptotic distribution theory for cox-type regression models with general relative risk form. *Annals of Statistics*, 11:804–813, 1983.

[41] R. L. Prentice, P. A. Shaw, S. A. Bingham, S. A. Beresford, B. Caan, M. L. Neuhouser, R. E. Patterson, M. L. Stefanick, S. Satterfield, C. A. Thomson, L. Snetselaar, A. Thomas, and L. Tinker. Biomarker-calibrated energy and protein consumption and increased cancer risk among postmenopausal women. *American Journal of Epidemiology*, 169:977–989, 2009.

[42] C. Ritenbaugh, R. E. Patterson, R. T. Chlebowski, B. Caan, L. F. Tinker, B. Howard, and J. Ockene. The women's health initiative dietary modification trial: overview and baseline characteristics of participants. *Ann Epidemiol*, 13:S87–S97, 2003.

[43] J. M. Robins. A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7:1393–1512, 1986.

[44] J. M. Robins and S. Greenland. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3:143–155, 1992.

[45] J. M. Robins, M. A. Hernan, and B. Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11:550–560, 2000.

[46] James M. Robins, Donald Blevins, Grant Ritter, and Michael Wulfsohn. G-estimation of the effect of prophylaxis therapy for *Pneumocystis carinii* pneumonia on the survival of aids patients. *Epidemiology*, 3:319–336, 1992.

[47] D. Rubin. Estimating causal effects of treatments in randomizd and nonrandomized studies. *Journal of Educational Psychology*, 66(5):689, 1974.

[48] D. Rubin. Inference and missing data. *Biometrika*, 63:581–592, 1976.

[49] D. B. Rubin. Multiple imputation for nonresponse in surveys. Michigan, 1987.

[50] D. A. Schoeller and J. M. Hnilicka. Reliability of the doubly labeled water method for the measurement of total daily energy expenditure in free-living subjects. *Journal of Nutrition*, 126:348S–354S, 1996.

[51] P. A. Shaw and Prentice R. L. Hazard ratio estimation for biomarker-calibrated dietary exposures. *Biometrics*, 68:397–407, 2012.

[52] D. S. Small. Mediation analysis without sequential ignorability: using baseline covariates interacted with random assignment as instrumental variables. *Journal of Statistical Research*, 46:91–103, 2012.

[53] X. Song, M. Davidian, and A. A. Tsiatis. A semiparametric lieklihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics*, 58:742–753, 2002.

[54] X. Song and Y. Huang. On corrected score approach for proportional hazards model with covariate measurement error. *Biometrics*, 61:702–714, 2005.

[55] D. Tapsoba, S. M. Lee, and C. Y. Wang. Approximate nonparametric corrected-score method for joint modeling of survival and longitudinal data measured with error. *Biometrical J*, 53:557–577, 2011.

[56] E. J. Tchetgen. On causal mediation analysis with a survival outcome. *The International Journal of Biostatistics*, 7:1–38, 2011.

[57] E. J. Tchetgen and S. H. Lin. Robust estimation of pure natural direct effects with mediator measurement error. *Harvard University Biostatistics Working Paper Series*, Working Papers:96:0, 2012.

[58] E. J. Tchetgen and J. Robins. On parametrization, robustness and sensitivity analysis in a marginal structural cox proportional hazards model for point exposure. *Harvard University Biostatistics Working Paper Series*, Working Papers:141:0, 2012.

[59] E. J. Tchetgen and I. Shpitser. Semiparametric estimation of models for natural direct and indirect effects. *Harvard Working Paper Series*, 1:Paper: 129, 2011.

[60] E. J. Tchetgen and I. Shpitser. Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness and sensitivity analysis. *Harvard University Biostatistics Working Paper Series*, Working Papers:130:0, 2012.

[61] T. R. Ten Have, M. M. Joffe, K. G. Lynch, G. K. Brown, S. A. Maisto, and A. T. Beck. Causal mediation analyses with rank preserving models. *Biometrics*, 63(3):926–934, 2007.

[62] A. A. Tsiatis and Davidian M. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometricka*, 88:447–458, 2001.

[63] A. A. Tsiatis and Davidian M. An overview of joint modeling of longitudinal and time-to-event data. *Statistica Sinica*, 14:793–818, 2004.

[64] M. J. Van der Laan. Targeted maximum likelihood based causal inference: Part i. *International Journal of Biostatistics*, 6, 2010.

[65] T. J. VanderWeele. Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology*, 21:540–551, 2010.

[66] T. J. VanderWeele. Causal mediation analysis with survival data. *Epidemiology*, 22:582–585, 2011.

[67] T. J. VanderWeele. Principal stratification-uses and limitations. *International Journal of Biostatistics*, 7:1–14, 2011.

[68] T. J. VanderWeele, L. Valeri, and E. L. Ogburn. The role of misclassification and measurement error in mediation analysis. *Epidemiology*, 23:561–564, 2011.

[69] C. Y. Wang and M. Pepe. Expected estimating equations to accommodate covariate measurement error. *Journal of the Royal Statistical Society Series B*, 62:509–524, 2000.

[70] C. Y. Wang, S. Wang, L. P. Zhao, and S. T. Ou. Weighted semiparametric estimation in regression analysis with missing covariate data. *Journal of the American Statistical Association*, 92:512–525, 1997.

[71] C.Y. Wang. Corrected score estimator for joint modeling of longitudinal and failure time data. *Statistica Sinica*, 16:235–253, 2006.

[72] D. Zeng and J. Cai. Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *Annals of Statistics*, 33:2132–2163, 2005.

[73] D. Zeng and D. Y. Lin. Efficient estimation of semiparametric transformation models for counting processes. *Biometrika*, 93:627–640, 2006.

[74] C. Zheng and X. H. Zhou. Causal models for mediation analyses: An introduction to structural mean models. *Journal of the Royal Statistical Society Series B*, in press.

[75] W. J. Zheng and M. J. Van der Laan. Causal mediation in a survival setting with time-dependent mediators. *U.C. Berkeley Working Paper Series*, 0:Paper 295, 2012.

[76] R. G. Ziegler, R. N. Hoover, M. C. Pike, A. Hildesheim, A. M. Nomura, D. W. West, A. H. Wu-Williams, L. N. Kolonel, P. L. Horn-Ross, J. F. Rosenthal, and Hyer M. B. Migration patterns and breast cancer risk in asian-american women. *Journal of the National Cancer Institute*, 85:1819–1827, 1993.

VITA

Cheng Zheng is a Ph.D candidate in biostatistics department at the University of Washington. He welcomes your comments to zhengc68@uw.edu.