

**MARGINAL STRUCTURAL COX PROPORTIONAL
HAZARDS MODEL FOR DATA WITH
MEASUREMENT ERRORS**

by

Ming Geng

B.S. Public Health, Hunan School of Medicine, Changsha, China,
1992

M.S. Clinical Psychology, Central South University, Changsha,
China, 2001

M.S. Biostatistics, University of Pittsburgh, 2006

Submitted to the Graduate Faculty of
the Graduate School of Public Health in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH
GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Ming Geng

It was defended on

November 10th, 2014

and approved by

Chung-Chou H. Chang, Ph.D., Professor

Departments of Medicine and Biostatistics

School of Medicine and Graduate School of Public Health

University of Pittsburgh

Abdus S. Wahed, Ph.D., Associate Professor

Department of Biostatistics

Graduate School of Public Health

University of Pittsburgh

Yan Lin, Ph.D., Research Assistant Professor

Department Biostatistics

Graduate School of Public Health

University of Pittsburgh

Jonathan G. Yabes, Ph.D., Assistant Professor

Departments of Medicine and Biostatistics

School of Medicine and Graduate School of Public Health

University of Pittsburgh

Andrew C. Thomas, Ph.D., Visiting Assistant Professor

Department of Statistics

Carnegie Mellon University

Dissertation Director: Chung-Chou H. Chang, Ph.D., Professor

Departments of Medicine and Biostatistics

School of Medicine and Graduate School of Public Health

University of Pittsburgh

Copyright © by Ming Geng
2014

MARGINAL STRUCTURAL COX PROPORTIONAL HAZARDS MODEL FOR DATA WITH MEASUREMENT ERRORS

Ming Geng, PhD

University of Pittsburgh, 2014

Abstract

In causal inference, interest often lies in estimating the joint effect of treatment on outcome at different time points. Marginal structural models are particularly useful for this purpose when a time-dependent confounder exists in the causal path between the treatment assigned in the previous time and the outcome. These models provide a consistent estimate when treatment is measured perfectly. In practice however, treatment may be subject to measurement error. Many studies have shown that measurement error in treatment can result in underestimating its effect. One approach proposed in the literature for correcting this problem is the marginal structural measurement-error model. It requires using a validation data set in which both the true treatment and the observed treatment are available to correct the bias. In this study, we developed a new method which combines the marginal structural Cox proportional hazards model, the regression calibration method, and the Bayesian method to account for measurement error in treatment without the need for a validation data set. Moreover, instead of fitting a traditional pooled logistic regression model, a weighted Cox proportional hazards model is implemented to reduce bias. The performance of our proposed method was assessed through the simulation study. Our simulation results show that the bias is reduced even with an approximate value of the parameter of the prior distribution. Our sensitivity analysis also shows that the estimated treatment effect is robust to the choice of the prior distribution. We applied our proposed method to estimate the effect of highly active antiretroviral therapy on the incidence of acquired immune deficiency syndrome or death

among HIV-positive patients using a data set in which the observed treatment assignment was subject to misclassification. **Public Health Significance:** Measurement errors can happen in medical studies despite good intentions. In general, either a validation data set or the replicates of the observed predictor are needed to correct for bias in estimation. Our study provides a new method in causal inference for correcting bias caused by measurement errors when investigators only have the main data set in which the observed treatment is measured only once at each time point.

KEY WORDS : Bayesian, marginal structural Cox model; misclassification; time-dependent confounder, treatment causal effect.

TABLE OF CONTENTS

PREFACE	xi
1.0 INTRODUCTION	1
1.1 CAUSAL EFFECT AND MARGINAL STRUCTURAL MODELS	2
1.2 MEASUREMENT ERROR IN PREDICTOR	5
1.3 AVAILABLE METHOD AND THE MOTIVATION OF OUR STUDY	9
2.0 OUR PROPOSED METHOD	13
2.1 INTRODUCTION	13
2.2 NOTATIONS AND MODELS	14
2.3 ESTIMATION	16
2.4 VARIANCE CALCULATION	17
2.5 SIMULATION	18
2.5.1 Simulation Setting	18
2.5.2 Simulation Results	20
2.6 EXAMPLE	22
2.7 CONCLUSIONS AND DISCUSSION	24
APPENDIX: TABLES AND FIGURES	27
BIBLIOGRAPHY	40

LIST OF TABLES

1	Simulation results with beta prior distribution of parameter comparing the bias, the Monte Carlo standard error (MCSE), the standard deviation estimates (SD), the coverage probability (95%) of the different analysis methods when the event risk is low (15%)	34
2	Simulation results with beta prior distribution of parameter comparing the bias, the Monte Carlo standard error (MCSE), the standard deviation (SD), the coverage probability (95%) of the different analysis methods when the event risk is moderate (25%)	35
3	Simulation results with beta prior distribution of parameter comparing the bias, the Monte Carlo standard error (MCSE), the standard deviation (SD), the coverage probability (95%) of the different analysis methods when the event risk is high (35%)	36
4	Simulation results with uniform prior distribution of parameter comparing the bias, the Monte Carlo standard error (MCSE), the standard deviation estimates (SD), the coverage probability (95%) of the different analysis methods when the event risk is low (15%)	37
5	Simulation results with uniform prior distribution of parameter comparing the bias, the Monte Carlo standard error (MCSE), the standard deviation (SD), the coverage probability (95%) of the different analysis methods when the event risk is moderate (25%)	38

6	Simulation results with uniform prior distribution of parameter comparing the bias, the Monte Carlo standard error (MCSE), the standard deviation (SD), the coverage probability (95%) of the different analysis methods when the event risk is high (35%)	39
7	Effect of HAART on the time to developing clinical AIDS or death among HIV-positive patients	39

LIST OF FIGURES

1	Difference between causation and association.	27
2	Conditions under which standard methods fail to give consistent estimate. . .	28
3	Underlying idea of marginal structural models.	29
4	Relation between the true treatment and observed treatment.	30
5	Variables relationship in simulated data set.	31
6	Comparison of dispersion of the estimated treatment effect in different sample size in scenario 1	32
7	Comparison of dispersion of the estimated treatment effect in different sample size in scenario 2	33

PREFACE

This dissertation is about using marginal structural Cox proportional hazard model to estimate the causal effect of possibly misclassified treatment on survival outcome without the need for a validation data set. This research was conducted under the supervision of Professor (Joyce) Chung-Chou H. Chang in the Departments of Biostatistics and Medicine, University of Pittsburgh.

I would like to express my sincerest gratitude to my advisor, Dr. Chung-Chou H. Chang for her excellent guidance and caring. It was her who provided me with an excellent atmosphere for realizing my shortcoming in research and presentation and patiently corrected my writing.

I would like to thank my committee members, Dr. Abdus S. Wahed, Dr. Andrew. C. Thomas, Dr. Yan Lin, Dr. Jonathan G. Yabes for their insightful suggestions.

I would like to thank Dr. Bo Fu, who was always available to help me in programming. I would not have been able to perform simulations successfully without his helps.

I would like to thank Dr. Wenjin Li for his caring and financial support during my study.

I would like to thank my parents for their love, caring and selfless support from a childhood. Without them, I would not be able to overcome the difficulties in my life and become a confident person.

Finally, I would like to thank my niece, Wenru Zhou, who teaches me love and forgiveness.

1.0 INTRODUCTION

In the current research in medicine, researchers often collect data from a longitudinal study in which both treatment assignments and covariate values vary over time. In such events, a confounder that predicts the subsequent treatment assignment and the outcome could be time-dependent. Moreover, this time-dependent confounder could be predicted by the treatment assignment at the previous time point. [Robins \(1997\)](#) showed that in this complex longitudinal data, the standard methods, such as regression, stratification, and matching, might result in a biased estimate of the joint effect of treatment on the outcome at various time points regardless this time-dependent confounder is adjusted or not. [Robins \(1998\)](#) later developed marginal structural models in order to estimate the causal effect of treatment on the outcome consistently.

It is well known that incorrectly recorded treatment assignments may result in a biased estimated treatment effect and the solutions for correcting the bias can be found in a large body of literature. However, less attention has been given to estimating the causal effect of treatment when treatment is measured imperfectly in the above mentioned complex longitudinal data. [Cole et al. \(2010\)](#) proposed a marginal structural measurement-error model (will be described in detail in Section 1.3) and applied their method to obtain a bias-corrected estimator of treatment effect on the survival. In order to apply their method, a validation data set is needed in which both true treatment and misclassified treatment are recorded.

In most of the situations, a validation data set does not exist. The objective of our study is to develop a marginal structural Cox model to estimate the causal effect of treatment on a survival outcome when the treatment assignment might be misclassified and the validation data set is not available. In Section 1.1, we will review causal effects and the marginal structural models proposed by [Robins \(1998\)](#). In Section 1.2, we will review methods for

adjusting measurement errors. In Section 1.3, we will provide a detailed description of the marginal structural measurement-error model proposed by [Cole et al. \(2010\)](#) which is the basis of our proposed model.

1.1 CAUSAL EFFECT AND MARGINAL STRUCTURAL MODELS

Let $Y_{a=1}$ and $Y_{a=0}$ be the outcome that would have been observed had a subject been treated ($a = 1$) and untreated ($a = 0$), respectively. $Y_{a=1}$ and $Y_{a=0}$ are referred to as the counterfactual outcomes. Since in reality everyone can only be either treated or untreated, every subject at most can contribute to only one counterfactual outcome.

According to the definition of [Robins and Hernán \(2008\)](#), causation can be demonstrated by comparing the marginal expected value of the counterfactual outcome $E(Y_{a=1})$, in which every subject in the population were treated, with that of the counterfactual outcome $E(Y_{a=0})$, in which every subject in the same population were untreated. If $E(Y_{a=1}) \neq E(Y_{a=0})$, the treatment A has an average causal effect on the outcome Y in that population. Since for every subject only one of the counterfactual outcomes is observed, the causal effect cannot be estimated directly. In contrast, association can be demonstrated by comparing the conditional expected value of the observed outcome in the treated group $E(Y|A = 1)$ with that in the untreated group $E(Y|A = 0)$. The associational effect can be estimated directly. Figure 1 shows the difference between causation and association.

In order to consistently estimate the average causal effect of the treatment A on the outcome Y in a population, one needs to link the causation to the association. The identifiability conditions are necessary but not sufficient for this linkage.

1. Consistency: if $\bar{A} = \bar{a}$, then $Y_{\bar{a}} = Y$ for that subject, where \bar{A} is the treatment history, \bar{a} is its possible value. Consistency means that $Y_{\bar{a}}$, a subject's counterfactual outcome under treatment history \bar{a} , equals to his observed outcome Y if his observed treatment history \bar{A} happens to be \bar{a} .
2. Conditional exchangeability: $Y_{\bar{a}} \perp A(t) | \bar{A}(t-1) = \bar{a}(t-1), \bar{L}(t) = \bar{l}(t)$ for all \bar{a} and $\bar{l}(t)$, where $\bar{L}(t)$ is the covariate history up to time t and $\bar{l}(t)$ is its possible value. Conditional

exchangeability means that the counterfactual outcome is independent of the treatment given the treatment and covariate histories. The assumption also implies there is no unmeasured confounder. Under this condition, the equations $E\{Y_{a(t)=1}|A(t) = 1\} = E\{Y_{a(t)=1}|A(t) = 0\} = E\{Y_{a(t)=1}\}$ and $E\{Y_{a(t)=0}|A(t) = 1\} = E\{Y_{a(t)=0}|A(t) = 0\} = E\{Y_{a(t)=0}\}$ will hold.

3. Positivity: if $P\{\bar{a}(t-1), \bar{l}(t)\} \neq 0$, then $P\{a(t)|\bar{a}(t-1), \bar{l}(t)\} > 0$ for all $a(t)$, where P denotes the probability. It means that among the subjects who have the same treatment and covariate histories in the underlying population, the probability of the assignment of any available treatment should be larger than 0. For example, if in the underlying population all patients who have the same treatment and covariate histories are treated, then the probability of not receiving treatment is 0, thus positivity is violated.

Positivity ensures that for a given subject, his counterparts, who have the same treatment history $\bar{A}(t-1)$ and covariate history $\bar{L}(t)$ but different current treatment $A(t)$, are available in the underlying population. Conditional exchangeability ensures that this subject and his counterparts would have the same expected value of the counterfactual outcome had they received the same treatment, so the information of the missing counterfactual outcome for a subject can be retrieved from the counterparts. Consistency ensures that the counterfactual outcomes can be identified from the observed outcomes. In short, conditional exchangeability ensures that every subject is exchangeable with his counterparts, and positivity ensures the availability of his counterparts, and consistency ensures that the counterfactual outcomes are observable.

The identifiability conditions hold in a marginal or conditional randomized experiment. If the conditions hold in an observational study, this observational study is equivalent to a conditional randomized experiment, therefore, it is possible to estimate the causal effect of treatment on outcome. However, [Robins and Hernán \(2008\)](#) demonstrated that one might obtain a biased estimated causal effect of treatment on outcome using the conventional methods, such as regression, stratification, and matching, when time-dependent confounder exists in the causal path between the treatment assigned at the previous time point and the outcome. According to the definition of [Robins and Hernán \(2008\)](#), if the identifiabil-

ity conditions hold, yet $E(Y_{\bar{a}}|L_0) \neq E(Y|\bar{A} = \bar{a}, L_0)$, where L_0 is the vector of baseline covariates, a time-dependent confounder exists. Figure 2 illuminates this situation. L_1 predicts the subsequent treatment A_1 and the outcome Y given the baseline variables. Thus, $E(Y_{\bar{a}}|L_0) \neq E(Y|\bar{A} = \bar{a}, L_0)$ and L is the time-dependent confounder. Treatment A_0 also predicts the time-dependent confounder L_1 . If one wants to consistently estimate the causal effect of A_1 on the outcome Y , he should adjust the confounder L_1 . However, L_1 is the collider of the treatment A_0 and the unmeasured factor U . By conditioning on L_1 , subjects with a certain association between U and A_0 are more likely to present in a certain level of L_1 , thus a “selection bias” [Robins and Hernán \(2008\)](#) is introduced. Also, L_1 is in the causal path of A_0 and Y . By adjusting for L_1 one can remove the treatment effect of A_0 on the outcome Y . Therefore, conventional methods cannot provide a consistent estimator of the joint effect of A_0 and A_1 on the outcome Y .

To address the above mentioned issue, [Robins \(1998\)](#) developed a new method, the marginal structural models, for estimating the causal effect of treatment consistently in this complex longitudinal data. He proved that the estimated parameter from the marginal structural models is a regular asymptotically linear estimator of the true parameter, and thus it is an asymptotically unbiased estimator. In order to perform a marginal structural model, a pseudo-population is created using the inverse probability treatment weighting (IPTW) method by which subject i at time j is assigned a weight W_{ij} which equals to the inverse conditional probability of receiving his current actual treatment given the treatment and covariate histories. Thus, in pseudo-population, every subject not only represents himself but also the $W_{ij} - 1$ subjects, who have the same treatment and covariate histories but different current treatment. That is, in the pseudo-population the treated sub-population includes the exactly the same patients as the untreated sub-population does. This exactly meets the definition of causation given by [Robins and Hernán \(2008\)](#). [Robins and Hernán \(2008\)](#) stated that $E(Y_a)$ in the pseudo-population equals to that in the actual population, and that the time-dependent confounder L which exists in the actual population is no longer a confounder in the pseudo-population. After creating the pseudo-population, one can fit a model, such as, linear model, logistic model, or Cox proportional hazards model in the pseudo-population. Figure 3 illustrates the underlying idea of marginal structural models.

In our study, what we are interested in is the hazard of the event, not the expected value of the outcome.

1.2 MEASUREMENT ERROR IN PREDICTOR

The problem caused by measurement error in predictor is not rare in medical study since variables can be measured imperfectly for a variety of reasons. Usually, the word “measurement error” is used to describe the continuous variables which are measured with error, whereas “misclassification” is used to describe the categorical variables. In order to correct for the bias in the estimator caused by measurement error, one needs an outcome model which relates the outcome with the true predictor and a measurement error model which relates the true predictor with the observed predictor. [Carroll et al. \(2006\)](#) described two commonly used measurement error models, the Classical measurement error model and the Berkson measurement error model, as well as other measurement error models that are less frequently used. Let X denote the observed predictor, Z denote its underlying true value, and U denote the measurement error. The classical measurement error model assumes that $X = Z + U_C$, $U_C \sim N(0, \sigma_C^2)$, whereas the Berkson measurement error model assumes that $Z = X + U_B$, $U_B \sim N(0, \sigma_B^2)$. Although the difference in the forms is subtle, they are very different models. If the error prone predictor is measured from each individual, the classical measurement error model is suitable. If all subjects in a small group are assigned with the same value for the predictor, then the Berkson measurement error is preferable.

The effect of measurement error in the predictor depends on the outcome model as well as the measurement error model. [Carroll et al. \(2006\)](#) showed that in simple linear regression, the classical type of measurement error results in an attenuated estimated coefficient of the mismeasured predictor, whereas the Berkson type of measurement error does not. [Carroll \(2011\)](#) proved that in multiple linear regression, the classical type of measurement error biases estimators for the mismeasured predictor as well as those covariates measured perfectly, unless they are independent. [Carroll \(2011\)](#) also showed that if there is more than one covariate measured with error, the estimated coefficients could be biased toward or away

from 0. [Reeves et al. \(1998\)](#) proved in logistic regression model, both the classical type and the Berkson type of measurement errors bias the estimated coefficient of the associated variable toward 0. [Li and Ryan \(2004\)](#) proved that in Cox proportional hazards model the measurement error attenuated the estimated coefficients for both the mismeasured predictor and the variable measured correctly, even though they were independent.

[Carroll et al. \(2006\)](#) also illustrated that the variance of the estimated coefficient of the mismeasured predictor is larger than that of the estimated true predictor effect, thus the measurement error in the predictor causes the loss of statistical power.

The problem of measurement errors has received a great deal of attention and different techniques have been developed to allow for the measurement error in predictor. Regression calibration method [Carroll et al. \(1995\)](#) is one of the commonly used methods. Instead of using the true predictor Z , this method use $E(Z|X)$ as predictor and then fit a standard model. If the validation data set in which both X and Z are recorded is available, one can fit model to regress Z on X in the validation data set and calculate the estimator of $E(Z|X)$ in the main data set, and then fit a standard model using $E(Z|X)$ as the predictor to obtain the bias-corrected estimated predictor effect. If only the replicates of the observed predictor are available, one can still estimate $E(Z|X)$. [Carroll et al. \(2006\)](#) and [Bartlett \(2011\)](#) described how to estimate $E(Z|X)$ in detail. Regression calibration method is potentially applicable for any model, given one can estimate $E(Z|X)$ correctly. It can reduce the bias in the estimate in logistic regression model and Cox proportional hazards model, but might still give a biased estimate when the measurement error is large.

[Rosner et al. \(1989\)](#) developed another version of regression calibration method in logistic regression. Bias is corrected using the formula $\hat{\theta} = \hat{\beta}/\hat{\gamma}_1$ under the assumptions that the disease is rare, the measurement error is not large and non-differential, meaning that the outcome is independent of the observed predictor given the true predictor, where $\hat{\theta}$ and $\hat{\beta}$ was the estimated effect of the true predictor, and the estimated effect of the observed predictor on the outcome, respectively, and $\hat{\gamma}_1$ was the estimated coefficient in the linear model $Z = \gamma_0 + \gamma_1 X + \epsilon, \epsilon \sim N(0, \sigma^2)$. The measurement error could be either a random error or systematic error. That is, the mean of the measurement error does not have to be 0. They extended their method to the case where there was more than one predictor measured

with error [Rosner et al. \(1990\)](#), and to the case where only the replicates of the mismeasured predictor were available [Rosner et al. \(1992\)](#).

[Prentice \(1982\)](#) first discussed the measurement errors problem in Cox proportional hazards model. Assuming the measurement error is non-differential, he derived the induced hazard function $\lambda\{t; X(t)\} = \lambda_0 E_{\{T \geq t, X(t)\}} \exp\{Z(t)\theta\}$, where λ_0 is the baseline hazard, and θ is the parameter of interest. However, the relative risk $E_{\{T \geq t, X(t)\}} \exp\{Z(t)\theta\}$ in the induced hazard function depends on the baseline hazard, but this dependency can be ignored when the risk of event is low. He then gave the partial likelihood $L(\theta) = \prod_{i=1}^k [\prod_{l \in F(t_i)} E \exp\{z_l(t_i)\theta\}] / [\sum_{l \in R(t_i)} E \exp\{z_l(t_i)\theta\}]^{m_i}$ under the assumption that the censoring is non-informative. By specifying the conditional distribution of $f(Z|X)$ in each risk set, one then can obtain the bias-corrected coefficient of the true predictor. [Spiegelman et al. \(1997\)](#) indicated that the bias-corrected estimate derived from Prentice's method also has the form $\hat{\theta} = \hat{\beta} / \hat{\gamma}_1$ under the assumptions that the linear measurement error model in Rosner's method is valid, the measurement error is small and non-differential, the censoring is non-informative, and the event risk is low. Thus, the relationship $\hat{\theta} = \hat{\beta} / \hat{\gamma}_1$ holds in both logistic regression and Cox proportional hazards model.

Another approach for correcting the bias in estimate caused by measurement error in predictor is Bayesian method. [Tadesse et al. \(2005\)](#) developed Bayesian error-in-variable model which solves the problem of measurement error in predictor through the Bayesian approach. In general, in order to perform the Bayesian analysis, one needs to specify the outcome model $f(Y|Z, \theta)$ which could be the Cox model, and the measurement error model such as the Berkson measurement error model $f(Z|X)$. The likelihood is the product of these two models. By specifying the prior distribution of the parameter θ , one can have the joint posterior distribution of θ and Z and the conditional posterior distribution for θ and Z , respectively, and estimate the bias-corrected coefficient from the corresponding conditional posterior distribution. The true predictor Z is estimated by drawing it from its conditional posterior distribution according to other known information. [Tadesse et al. \(2005\)](#) assumed the baseline hazard was a constant within each disjoint small interval and used the full likelihood of Cox model to perform the analysis. [Sinha et al. \(2003\)](#) proposed the use of the [Breslow \(1974\)](#) type of partial likelihood to obtain the posterior distribution if the outcome

model is the Cox proportional hazard model, thus simplified the computational process of implementing Cox model through the Bayesian approach.

Nakamura (1992) applied the corrected score function method developed by Nakamura (1990) to the Cox proportional hazards model and obtained asymptotically unbiased estimate under the assumption that the measurement errors are additive and normally distributed. He corrected the bias by adding a term to the naive score function $\Lambda\theta$, where Λ is the known measurement error variance and θ is the unknown parameter of interest; therefore, the corrected score function was defined as $U_i^*(\theta, X, Y) = U_i(\theta, X, Y) + \Lambda\theta$ and $U^*(\theta, X, Y) = \sum U_i^*(\theta, X, Y)$, where $U_i(\theta, X, Y)$ is the naive score function for subject i derived from the partial likelihood with X as the predictor. Nakamura showed that $E\{U^*(\theta, X, Y)\} \approx U(\theta, Z, Y)$ if $\theta^T \Lambda \theta$ is small. The estimator from the corrected score function is an approximately unbiased estimator.

Tsiatis and Davidian (2001) developed the conditional score method in survival analysis accounting for the time-dependent covariate that was mistimed and mismeasured. They used the classical measurement error model and assumed that the true predictor follows random coefficient linear model; therefore, $X_i(t_{ij}) = Z_i(t_{ij}) + e_{ij}$ and $Z_i(u) = \alpha_{0i} + \alpha_{1i}u$, where $X_i(t_{ij})$ is the observed predictor for subject i at time j , and $Z_i(t_{ij})$ is its underlying true value, and e_{ij} is the normally distributed measurement error with mean 0. Their outcome model is $\lambda_i(u) = \lambda_0(u) \exp\{\theta Z_i(u) + \eta^T W_i\}$, where θ and η are the parameters of interest and W is a covariate vector measured precisely. Conventionally, $\alpha_i = (\alpha_{0i}, \alpha_{1i})$ were assumed to be normally distributed. The essence of the conditional score method is to relax this assumption by finding the complete sufficient statistic of the random effect α_i and “condition away” the dependence of the hazard function on α_i according to its complete sufficient statistic. They then provided the estimating equations for solving θ and η based on this conditional intensity process, given the complete sufficient statistic and W . Unlike Nakamura who assumed a known measurement error variance, Tsiatis et al provided a method to estimate the variance of the measurement error from the data. However, Tsiatis method is asymptotically equivalent to Nakamura’s method, if Z is time-independent and the variance of measurement error is known.

Hu et al. (1998) have developed a likelihood-based approach to account for measurement errors in the predictor in the Cox proportional hazards model. Let X be the observed predictor, Z be the underlying true value, and Y be the outcome. The likelihood is $L(Y, X) = \prod_{i=1}^n [\int f_C(y_i|x_i, z_i) \times f_M(x_i, z_i) dz]$, where f_C is the full likelihood of Cox model and f_M is the joint distribution of X and Z . Under the non-differential assumption, $f_C(y_i|x_i, z_i) = f_C(y_i|z_i)$. They use the classical type measurement error model $X_i = Z_i + U_i, U_i \sim N(0, \sigma_u^2)$, so $f_M(x, z) = f(x|z) \times f(z)$. $f(x|z)$ is assumed known. They provided three inferential methods, fully parametric method in which $f(z)$ was assumed to be a parametric distribution with the unknown parameters which have to be estimated, and a nonparametric method in which $f(z)$ was approached using the nonparametric maximum likelihood estimation of a mixing distribution method, and a semiparametric method in which $f(z)$ was assumed to be a smooth function and approximated by the Hermite series. Details of the nonparametric maximum likelihood estimation of a mixing distribution method and the Hermite series are beyond the scope of our study and therefore are not presented. Parameters were estimated by maximizing the likelihood. The fully parametric method was not robust to the misspecification of $f(z)$, and the nonparametric method introduced many additional parameters which have to be estimated. Their simulation results showed that the semiparametric method was superior to the fully parametric method and the nonparametric method in multiple scenarios.

1.3 AVAILABLE METHOD AND THE MOTIVATION OF OUR STUDY

All the approaches to correcting measurement errors described in Section 1.2 are regression-adjustment methods. Robins (1998) has shown that regression-adjustment methods fail to give consistent estimators of the joint causal effect of treatment at different time points when there is a time-dependent confounder, which is predicted by the previous treatment assignments. Cole et al. (2010) developed a marginal structural measurement-error model to estimate the causal effect of antiretroviral therapy on incidence of acquired immunodeficiency syndrome (AIDS) or death among patients with human immunodeficiency virus

(HIV) infection. The treatment information in their study was collected from self-reported questionnaires and not all patients provided the correct information about the treatments they actually received. Their validation data showed that only 84% of patients who received treatments reported that they did, and 80% of patients who did not receive treatments reported that they did not. [Cole et al. \(2010\)](#) combined the marginal structural Cox model proposed by [Robins \(1998\)](#) with the regression calibration method proposed by [Rosner et al. \(1989\)](#) in estimating causal effect of treatment to reduce the bias caused by treatment misclassification. The method of [Cole et al. \(2010\)](#) took a three-step approach in estimating the effect of treatment. In the first step, they fitted a marginal structural Cox proportional hazards model using the self-reported treatment assignments as the predictor in order to obtain the estimated self-reported treatment causal effect $\hat{\beta}$. In this step, they fitted a weighted pooled logistic regression model instead of a weighted time-dependent Cox model because that most of the software packages did not allow for time-dependent weights in fitting a Cox model. In the second step, they adopted the idea of [Rosner et al. \(1989\)](#) and fitted a weighted pooled linear regression model $E(Z_i|X_i) = \gamma_0 + \gamma_1 X_i$ using a validation data set to estimate γ_1 , where X_i was the self-reported treatment for patient i , and Z_i was the true treatment which was taken from the patient medical record. In the third step, Cole et al obtained a bias-corrected treatment effect using the formula $\hat{\theta} = \hat{\beta}/\hat{\gamma}_1$. [Rosner et al. \(1989\)](#) showed that $\hat{\beta}$ is an “approximate maximum likelihood estimate” of $\theta\gamma_1$ and thus one can expect $\hat{\beta}/\hat{\gamma}_1$ to be a consistent estimate of θ .

If a validation data set is not available, the method proposed by [Cole et al. \(2010\)](#) cannot be applied directly. Without validation data, one cannot combine either the regression calibration method or the likelihood-based method with the marginal structural Cox model to correct the bias in the estimated treatment causal effect because a validation data set is required in estimating the relation between the true treatment and the observed treatment.

Although the corrected score function method and the conditional score method for measurement error adjustments described in Section 1.2 are easy to implement, we cannot combine either of them with the marginal structural Cox model to obtain a bias-corrected estimator of treatment causal effect because these two measurement error correction methods assume that measurement error is normally distributed with mean 0 and a known variance

or the variance can be estimated from the data. These requirements are not suitable for our study. For example, the mean of measurement error in our study, which is $P_1 - P_2$ may not be 0, where P_1 is the probability of a subject reporting that he was treated but actually was not treated, and P_2 is the probability of a subject reported that he was not treated but actually was treated.

In the following paragraphs, we will first review the standard marginal structural Cox model in estimating casual effect of treatment when data without treatment misclassification are available and then review the Bayesian measurement-error correction method in estimating treatment associational effect. Finally, we will describe the difficulties one may encounter in estimating causal effect of treatment when he/she tries to use the combination of the Bayesian method and the marginal structural Cox model.

To estimate causal effect of treatment when data without treatment misclassification are available, a marginal structural Cox model can be used. Using this model, we will first create a pseudo population that removes the effect of time-dependent confounders on treatment by assigning a subject with a weight. The weight is equal to the inverse conditional probability of receiving the current treatment given the previous treatment and covariate histories. A time-dependent Cox model will then be fitted to this pseudo-population. This is equivalent to fitting a weighted time-dependent Cox model to the actual population. When treatment is not misclassified, weights will be calculated from the true treatment. [Robins \(1998\)](#) proved that treatment effect estimated from the marginal structural Cox model is unbiased.

If we are interested in estimating treatment associational effect θ (not treatment causal effect), while treatment is possibly misclassified, we can apply the Bayesian measurement-error correction method to the actual population to reduce bias. To do so, we will first create a Cox partial likelihood of survival outcome given θ and true treatment Z . Then, we will create a measurement-error model of true treatment Z given observed treatment X . When a prior distribution of θ is specified, we will be able to obtain the posterior distribution of θ by combining the partial likelihood, the measurement-error model, and the prior distribution.

Intuitively, we can construct a weighted partial likelihood of survival outcome given θ and true treatment Z if we are interested in estimating treatment causal effect when treatment is possibly misclassified. If we are to do so, we could apply the Bayesian method described

in the previous paragraph to obtain a bias-corrected estimator. However, the weights should be calculated from the true treatment Z which is not possible when a validation data is unavailable. Therefore, combining Bayesian measurement-error method and the marginal structural Cox model to estimate treatment causal effect is not feasible.

To overcome the above-mentioned problems, we propose to replace the second step of the method of Cole et al by the Bayesian method when estimating γ_1 . Once the estimated γ_1 is obtained, we will use the formula $\hat{\theta} = \hat{\beta}/\hat{\gamma}_1$ to obtain a bias-corrected treatment effect, which is the last step of the method proposed by [Cole et al. \(2010\)](#).

2.0 OUR PROPOSED METHOD

2.1 INTRODUCTION

Robins (1998) introduced the marginal structural models which included the marginal structural Cox proportional hazards model. He proved that the estimator from the marginal structural models is a regular linear estimator and therefore it is an unbiased estimator. Hernán et al. (2000) demonstrated how to use the marginal structural Cox proportional hazards model with an example of estimating the causal effect of Zidovudine on mortality for patients with HIV infection.

For each treatment history \bar{a} the hazard of the marginal structural Cox proportional hazards model has the form

$$\lambda_{T_{\bar{a}}}(t|V) = \lambda_0(t) \exp[\beta_1 g\{\bar{a}(t)\} + \beta_2 V],$$

where \bar{a} denotes treatment history, $g(\cdot)$ is a known function, $T_{\bar{a}}$ denotes a subject's counterfactual event time had he followed treatment history \bar{a} , V denotes the vector of baseline covariates, $\lambda_{T_{\bar{a}}}(t|V)$ denotes hazard of the event at time t for subjects with covariate V in the population had all subjects followed \bar{a} through time t , $\lambda_0(t)$ denotes a unspecified baseline hazard of the event, and β_1 and β_2 denote parameters which need to be estimated.

In practice, fitting the marginal structural Cox proportional hazards model is equivalent to fitting a weighted time-dependent Cox proportional hazards model, $\lambda_T(t|\bar{A}(t), V) = \lambda_0(t) \exp[\gamma_1 g\{\bar{A}(t)\} + \gamma_2 V]$ by assigning subject i at time j a weight $w_{ij} = w_{ij}^A \times w_{ij}^C$, where $A(t)$ is the actual treatment at time t , $\bar{A}(t)$ is the actual treatment history up to time t , $\lambda_T(t|\bar{A}(t), V)$ is the conditional hazard of the event given the actual treatment history $\bar{A}(t)$ and baseline covariate V , and γ_1 and γ_2 are the parameters to be estimated. The weight W_{ij}^A

denotes the inverse probability treatment weight and W_{ij}^C denotes the inverse probability censoring weight and can be calculated as follows, respectively

$$w_{ij}^A = \prod_{k=0}^j \frac{\text{pr}\{A(k) = a_i(k) | \bar{A}(k-1) = \bar{a}_i(k-1), V = v_i\}}{\text{pr}\{A(k) = a_i(k) | \bar{A}(k-1) = \bar{a}_i(k-1), \bar{L}(k) = \bar{l}_i(k)\}},$$

$$w_{ij}^C = \prod_{k=0}^j \frac{\text{pr}\{C(k) = 0 | \bar{C}(k-1) = 0, \bar{A}(k-1) = \bar{a}_i(k-1), V = v_i\}}{\text{pr}\{C(k) = 0 | \bar{C}(k-1) = 0, \bar{A}(k-1) = \bar{a}_i(k-1), \bar{L}(k-1) = \bar{l}_i(k-1)\}}.$$

2.2 NOTATIONS AND MODELS

Let Z be the true treatment, X be the patient self-reported treatment, D_k be the set of subjects who experienced the event at time k , R_k be the set of subjects who were at risk right before time k , $g(\bar{x}_{ik})$ be the proportion of treated time reported by patient i among the first k times, W_{ik} be the weight for subject i at time k , γ_1 be the coefficient in the model $E(Z|X) = \gamma_0 + \gamma_1 X$ indicating the relation between Z and X , θ be the true treatment effect which we are interested in and will be estimated, and β be the self-reported treatment effect.

In order to estimate the causal effect of the treatment Z , we fit a weighted time-dependent Cox model directly and suggest using the Breslow-type [Breslow \(1974\)](#) weighted partial likelihood for β :

$$L_p(\beta) = \prod_{k=1}^K \frac{\exp\{\beta \sum_{D_k} W_{ik} g(\bar{x}_{ik})\}}{[\sum_{R_k} W_{ik} \exp\{\beta g(\bar{x}_{ik})\}]^{\sum_{D_k} W_{ik}}}.$$

Since the relationship $\theta = \beta/\gamma_1$ holds in Cox proportional hazards model under certain conditions [Spiegelman et al. \(1997\)](#), by knowing the partial likelihood of β , the partial likelihood for θ becomes

$$\begin{aligned} L_p(\theta) &= \frac{\partial(\gamma_1 \theta)}{\partial \theta} \prod_{k=1}^K \frac{\exp\{\gamma_1 \theta \sum_{D_k} W_{ik} g(\bar{x}_{ik})\}}{[\sum_{R_k} W_{ik} \exp\{\gamma_1 \theta g(\bar{x}_{ik})\}]^{\sum_{D_k} W_{ik}}} \\ &= \gamma_1 \prod_{k=1}^K \frac{\exp\{\gamma_1 \theta \sum_{D_k} W_{ik} g(\bar{x}_{ik})\}}{[\sum_{R_k} W_{ik} \exp\{\gamma_1 \theta g(\bar{x}_{ik})\}]^{\sum_{D_k} W_{ik}}} \\ &= \gamma_1 \prod_{k=1}^K \frac{\exp[\theta \sum_{D_k} W_{ik} \{\gamma_1 g(\bar{x}_{ik})\}]}{(\sum_{R_k} W_{ik} \exp[\theta \{\gamma_1 g(\bar{x}_{ik})\}])^{\sum_{D_k} W_{ik}}}, \end{aligned}$$

which is the Breslow-type partial likelihood with $\gamma_1 g(\bar{x}_{ik})$ as the regressor.

To estimate γ_1 in $L_P(\theta)$ when a validation data set is not available, we define the correct-classification rates and the marginal probability of being actually treated as follows

$$\eta_1 = P(X = 1|Z = 1),$$

$$\eta_2 = P(X = 0|Z = 0),$$

where η_1 and η_2 can be viewed as the probability of correctly identifying those treated patients and the probability of correctly identifying those untreated patients, respectively.

We also define the marginal probability of being actually treated as

$$\pi = P(Z = 1).$$

[McGlothlin et al. \(2008\)](#) have defined these probabilities in order to correct the effect of misclassification in outcome in the logistic regression model when the outcome was misclassified and one of the predictors was measured with error.

With these three probabilities, we can calculate the following two probabilities:

$$\psi_1 = P(Z = 1|X = 1) = \frac{\eta_1 \times \pi}{\eta_1 \times \pi + (1 - \eta_2) \times (1 - \pi)},$$

$$\psi_2 = P(Z = 0|X = 0) = \frac{\eta_2 \times (1 - \pi)}{\eta_2 \times (1 - \pi) + (1 - \eta_1) \times \pi},$$

where ψ_1 and ψ_2 can be viewed as the probability of self-reported treated that are actually treated and the probability of self-reported untreated that are actually untreated.

Simulation evidences in the studies of [Cole et al. \(2006\)](#) and [Spiegelman et al. \(2000\)](#) suggest that it is appropriate to use the model $E(Z_i|X_i) = \gamma_0 + \gamma_1 X_i$ to represent the relation between the self-reported treatment X and the true treatment Z . From this model, $E(Z|X) = \gamma_0 + \gamma_1 X$ for $X = 0$ and $X = 1$. Note that $E(Z|x = 0) = P(z = 1|x = 0) = 1 - P(z = 0|x = 0) = 1 - \psi_2$ and $E(Z|x = 1) = P(z = 1|x = 1) = \psi_1$. It is evident that $\gamma_1 = E(Z|x = 1) - E(Z|x = 0) = \psi_1 + \psi_2 - 1$. Figure 4 depicts these relations. Once π is specified, one can estimate η_1 and η_2 , then ψ_1 and ψ_2 , and then γ_1 .

If η_1 and η_2 are unknown, we can postulate prior distributions of η_1 and η_2 to estimate the uncertainties. We then will combine these prior distributions with the observed data to estimate the parameter through the Bayesian approach.

2.3 ESTIMATION

We let parameters η_1 and η_2 follow the prior distribution. $\eta_1 \sim \text{beta}(a_1, b_1)$, $\eta_2 \sim \text{beta}(a_2, b_2)$, respectively, and let parameter $\theta \sim N(\mu_0, \sigma_0^2)$, where the information about a_1, b_1, a_2, b_2 can be calculated based on the mean and the variance of η_1 and η_2 , and the mean and the variance of η_1 and η_2 can be estimate from the median and the range of η_1 and η_2 . [Hozo et al. \(2005\)](#) developed a method to estimate the mean and variance from the median and the range. We assume this information can be obtained from a pilot study or from experts. Since little is known about θ , we set $\mu_0 = 0$ and use a large value of σ_0^2 (e.g. $\sigma_0^2 = 100$).

If η_1 , η_2 , and θ are independent, the joint posterior distribution of η_1 , η_2 , and θ is

$$P(\theta, \eta_1, \eta_2 | g(\bar{x}_{ik}), t) \propto \gamma_1 \prod_{k=1}^K \frac{\exp\{\gamma_1 \theta \sum_{D_k} W_{ik} g(\bar{x}_{ik})\}}{[\sum_{R_k} W_{ik} \exp\{\gamma_1 \theta g(\bar{x}_{ik})\}]^{\sum_{D_k} W_{ik}}} \times \\ \eta_1^{a_1} (1 - \eta_1)^{b_1 - 1} \times \eta_2^{a_2} (1 - \eta_2)^{b_2 - 1} \times \exp\left\{-\frac{(\theta - \mu_0)^2}{\sigma_0^2}\right\}$$

where $\gamma_1 = \frac{\eta_1 \times \pi}{\eta_1 \times \pi + (1 - \eta_2) \times (1 - \pi)} + \frac{\eta_2 \times (1 - \pi)}{\eta_2 \times (1 - \pi) + (1 - \eta_1) \times \pi} - 1$

In order to estimate η_1 , η_2 , and θ , we randomly sampled η_1 , η_2 , and θ from their joint posterior distribution, and estimated η_1 , η_2 , and θ by taking their average values, respectively. The joint posterior distribution of η_1 , η_2 , and θ is not a standard density, so it is difficult to sample η_1 , η_2 , and θ from it. Markov chain Monte Carlo (mcmc) is particularly useful in this situation. The essence of mcmc is to create a Markov chain which will typically converge to the desired posterior distribution over the simulation after a large number of iterations, then the draws from Markov chain are approximately from the posterior distribution. We use the *mcmc* package in R 3.0.2 to implement the simulation. *mcmc* is a package which allows users to create Markov chain via metropolis algorithm without writing their own code.

Let $P(\theta|x)$ be the posterior distribution of the parameter θ . Metropolis algorithm works as follows.

1. Choose initial value θ^0 for θ with $P(\theta^0|x) > 0$.
2. At iteration t , draw the candidate θ^* from the proposal distribution $J_t(\theta^*|\theta^{t-1})$.
3. Calculate the acceptance ratio $r = P(\theta^*|x)/P(\theta^{t-1}|x)$.
4. Accept θ^* with probability $\min(r, 1)$, otherwise $\theta^t = \theta^{t-1}$.

5. Repeat step 2 to step 4 N times, where N is a large number (e.g. $N = 10,000$).

In order to use *mcmc* package, one has to specify the initial values of the parameters and the logarithm of unnormalized posterior distribution which equals to the logarithm of likelihood plus logarithm of prior distribution. At iteration t , the software draws a candidate of parameters from the proposed distribution which is a multivariate normal distribution, calculates the acceptance ratio, and determines whether the candidate of parameters should be accepted. The output is an $m \times p$ matrix with each element the mean of batch, where batch is a segment of the Markov chain with certain length (e.g. the k^{th} batch can be $\theta^{k+1}, \dots, \theta^{k+b}$, m is the number of batches and p is the dimension of parameters. The overall estimator of each parameter is obtained by calculating the associated column mean of the matrix. For example, if there are two parameters, θ and γ , to be estimated, the output will be as follows

$$\begin{pmatrix} \theta_1 & \gamma_1 \\ \dots & \dots \\ \theta_m & \gamma_m \end{pmatrix},$$

where the batch mean θ_1 is the average value of the first batch $\theta^1, \dots, \theta^b$, and θ_m is the average value of the last batch $\theta^{(m-1)b+1}, \dots, \theta^{(m-1)b+b}$. The overall estimated θ is $\hat{\theta} = (\theta_1 + \dots + \theta_m)/m$. Parameter γ is estimated with the same method.

2.4 VARIANCE CALCULATION

In order to make the hypothesis test feasible, we constructed the variance estimator of the estimated treatment causal effect $\hat{\theta}$ from our proposed method. [Rosner et al. \(1989\)](#) provided a formula to calculate the variance of $\hat{\theta}$ where $\hat{\theta} = \hat{\beta}/\hat{\gamma}_1$ and $\hat{\beta}$ is the estimated self-reported treatment effect. $\hat{\beta}$ and $var(\hat{\beta})$ can be obtained from the standard marginal structural Cox model with the self-reported treatment as the predictor. In their study, $\hat{\gamma}_1$ and $var(\hat{\gamma}_1)$ are estimated by fitting a linear regression model using a validation data set. Because we do not have a validation data set, we are not able to use their strategy to construct the variance estimator. We estimated $\hat{\gamma}_1$ and $var(\hat{\gamma}_1)$ based on the Bayesian Central Limit Theorem.

The idea is described below. Let $\alpha = (\theta, \gamma_1)$, $\hat{\alpha}$ be the MLE of α , H be the observed Fisher information matrix, and $\Sigma = H^{-1}$. The Bayesian Central Limit Theorem states that when sample size is large, the posterior distribution of α , $P(\alpha|x)$, is approximately normally distributed $N(\hat{\alpha}, \Sigma)$ where x denotes the data. In our study, the joint likelihood function of θ and γ_1 is

$$L_p(\theta) = \gamma_1 \prod_{k=1}^K \frac{\exp\{\gamma_1 \theta \sum_{D_k} W_{ik} g(\bar{x}_{ik})\}}{[\sum_{R_k} W_{ik} \exp\{\gamma_1 \theta g(\bar{x}_{ik})\}]^{\sum_{D_k} W_{ik}}},$$

θ and γ_1 cannot be estimated simultaneously from this partial likelihood directly because that two different sets of θ and γ_1 may lead to the same likelihood, and thus they are not identifiable. Therefore, we estimated θ using the proposed method we described above and substituted θ by its estimator $\hat{\theta}$, then we calculated the mean and the variance of γ_1 , and then substituted these values into the formula of

$$var(\hat{\theta}) = (1/\hat{\gamma}_1^2)var(\hat{\beta}) + (\hat{\beta}^2/\hat{\gamma}_1^4)var(\hat{\gamma}_1)$$

provided by [Rosner et al. \(1989\)](#).

2.5 SIMULATION

2.5.1 Simulation Setting

Figure 5 was adapted based on ‘‘Appendix Figure 1’’ in [Cole et al. \(2010\)](#). It illustrates the relationship among variables in our simulated data sets. Because that the true treatment Z has causal effects on the time-dependent confounder L and the self-reported treatment X and that L has causal effect on survival time T , Z has indirect causal effect on T . Self-reported treatment X does not have causal effect on T given Z . We generated simulation data sets based on Figure 5. A detailed specification is described as follows:

1. True treatment at time 0: $Z_0 \sim Bernoulli(0.5)$.
2. True treatment at time 1: $Z_1|L = 1 \sim Bernoulli(0.7)$, $Z_1|L = 0 \sim Bernoulli(0.3)$.
3. Self-reported treatment at time 0: $X_0|Z_0 = 1 \sim Bernoulli(\eta_1)$, $X_0|Z_0 = 0 \sim Bernoulli(1 - \eta_2)$.

4. Self-reported treatment at time 1: $X_1|Z_1 = 1 \sim \text{Bernoulli}(\eta_1)$, $X_1|Z_1 = 0 \sim \text{Bernoulli}(1 - \eta_2)$.
5. Time-dependent confounder: $L|Z_0 = 1 \sim \text{Bernoulli}(0.7)$, $L|Z_0 = 0 \sim \text{Bernoulli}(0.3)$. $P(L = 1) \simeq 0.5$.
6. Survival time: $T \sim \text{Exp}(\lambda)$ with $\lambda = \lambda_0 \exp(\alpha L)$, where $\alpha = -0.7$, λ_0 was set to make the risk of event 0.15, 0.25, 0.35.
7. Censoring time was generated from the uniform distribution. The parameters were set to make censoring rate 0, 0.15, 0.3.
8. $\eta_1 = 0.9$, $\eta_2 = 0.8$, and $\pi = 0.5$, where η_1 and η_2 are the probabilities of actually treated and untreated patients who are correctly identified. π is the marginal probability of being actually treated.
9. The parameter of the true treatment effect on the survival time T was calculated using the method of [Cole et al. \(2010\)](#). As described above, the true treatment Z_0 has an indirect effect on the survival time T through L , therefore, in this step, we treat L as the true predictor of the survival time T and treat Z_0 as the “misclassified” predictor. We use θ to denote the parameter of treatment effect, which has to be calculated in this step, and use α to denote the effect of L , which we have already set to be -0.7 . In order to estimate γ_1 in the method of [Cole et al. \(2010\)](#), we simulated 100 data sets with each containing 100,000 observations and fitted the linear model $E(L|Z_0) = \gamma_0 + \gamma_1 Z_0$ in each data set. The estimated γ_1 is the average of γ_1 obtained from the above linear model fitted in each data set. $\gamma_1 = 0.3997$. $\theta = \alpha * \gamma_1 = -0.7 * 0.3997 = 0.27979 \approx -0.28$.

We simulated 500 data sets each containing 1000, 2000, and 3000 subjects. For each of the simulated data set, we performed two analyses to estimate the causal effect of treatment on mortality and compared results obtained from these two methods. First, we performed the naive analysis using the observed treatment assignments as the predictor without correcting for possibly misclassified treatment assignments. This was fitted via a standard marginal structural Cox model. Then we applied the proposed method to obtain the bias-corrected causal effect of treatment on mortality. The estimated treatment effect is calculated as the average of the estimated treatment effect obtained from each data set. We also calculated the standard error of the estimated treatment effect which is the square root of the average

of the variance estimated from each data set, and the coverage probability which is the proportion that the confidence interval contains the true parameter, which is -0.28 in our simulation study.

In order to examine the impacts of the prior distributions of η_1 and η_2 on the estimators of the treatment causal effect, we performed the sensitive analysis. Specifically, besides the beta distribution, we also assume η_1 and η_2 follow uniform distribution. The parameters of uniform distribution can be calculated based on the mean and the variance of η_1 and η_2 . For these two types prior distribution, we shifted down η_1^* , the mean of the prior distribution of η_1 , by 5%, 10% and shifted up η_2^* , the mean of the prior distribution of η_2 , by 5%, 10% from the original settings of η_1 and η_2 , respectively. We also shifted η_1^* and η_2^* in the same direction (down or up) by 5% simultaneously.

In all simulations, we set the number of batch to 1000 and the length of batch to 10.

2.5.2 Simulation Results

We performed simulations in different scenarios with combinations of different event rates, censoring rates, and values of η_1^* and η_2^* . Simulation results that are presented in Table 1 to Table 3 were estimated from the sample containing 3,000 observations in 45 scenarios with beta prior distribution for η_1 and η_2 . Simulation results that are presented in Table 4 to table 6 were estimated from the sample with the same size in the same scenarios but with uniform prior distribution for η_1 and η_2 . The naive estimate is always biased severely toward 0. Either high event risk ($\geq 35\%$) or high censoring rate ($\geq 30\%$) bias the estimated treatment effect toward 0 regardless the prior distribution, but the bias is moderate ($\leq 10\%$) if only one of these two factors exists. However, if both the event risk and the censoring rate are high, our proposed method gives a biased estimate but is still better than the naive estimate. In general, if the high risk of event and the high censoring rate are not present at the same time, our method reduces bias of the estimate greatly in the following settings:

1. let η_1^* and η_2^* be the original settings of η_1 and η_2 , respectively.
2. shift η_1^* down from η_1 by 5% and η_2^* up from η_2 by 5%.
3. shift η_1^* down from η_1 by 10% and η_2^* up from η_2 by 10%.

However, when both of η_1^* and η_2^* are shifted from the original settings of η_1 and η_2 in the same direction by 5% simultaneously, our proposed method gave a biased estimate which is less biased as compared with the naive estimate.

In all scenarios, the estimated coverage probability is lower than 95%, which is the target value we expected to achieve. The coverage probability increases when the event risk and the censoring rate are high. The coverage probability also increases with the sample size becoming large. For example, the coverage probability is 0.724 for the sample with size 1000 in the setting that the event risk is equal to 15%, censoring rate is equal to 0, η_1^* is equal to 0.9, and η_2^* is equal to 0.8, it increases to 0.796 if the sample size is 2000, and increases to 0.86 if the sample size is 3000.

Our simulation results also show that if the sample size is small (e.g. 1000), although the estimated treatment effect generally shows the trend we described above using the large sample (e.g. 3000), the estimated treatment effect in some scenarios is abnormal, reflecting that the simulation results are not very stable. For example, in the setting that both the event risk and censoring rate are equal to 15%, η_1^* is equal to 0.9, and η_2^* is equal to 0.8, the estimated treatment effect is -0.253 , which is biased, whereas the corresponding estimator from a large sample is very close to the true parameter. In the scenario that the event risk and censoring rate are 15%, shifting up η_1^* and η_2^* by 5%, results in a estimated treatment effect of -0.263 , which is “accurate”, whereas, shifting up η_1^* and η_2^* by 5% simultaneously bias the estimated treatment effect toward 0 in all settings when the sample size is large.

When the sample size is small (e.g. $< 1,000$), the estimated treatment effect from each simulation data set could spread widely around the true parameter, about 10% of the estimators were positive, which were in the opposite direction of the true parameter value. When the sample size becomes larger (e.g. 2,000), it is less likely that the estimated treatment effect from data sets are positive, but they still spread widely around the true parameter. The estimated treatment effect becomes more concentrate around the true parameter with the sample size increasing. Figure 6 and Figure 7 show this phenomenon in 2 different scenarios: scenario 1 was set the event risk to 0.15, censoring rate to 0, η_1^* to 0.9, and η_2^* to 0.8; scenario 2 was set the event risk to 0.35, censoring rate to 0.3, η_1^* to 0.9, and η_2^* to 0.8.

The estimates of η_1 and η_2 from the posterior distribution are almost equivalent to the mean of the prior distribution of η_1 and η_2 , respectively, showing that the data have little impact on the estimates.

2.6 EXAMPLE

Human immunodeficiency virus (HIV) infection presents one of the challenges in disease treatment worldwide. Center for Disease Control and Prevention (CDC) estimates that in the United States there were 1,148,200 people who had been infected with HIV by the end of 2009, 15,529 people who diagnosed of AIDS died in 2010, and 32,052 people were diagnosed with AIDS in 2011 [CDC \(2011\)](#). Highly active antiretroviral therapy (HAART) was introduced in 1996, and many studies showed that it delays the progression to AIDS [Wong et al. \(2004\)](#) and also prolongs the survival among HIV infected patients [Wong et al. \(2004\)](#), [Tam et al. \(2002\)](#). However, these studies either failed to take the time-dependent confounder into account [Wong et al. \(2004\)](#), [Tam et al. \(2002\)](#) or assumed that once patients received the treatment, they would stay in the treatment [Sterne et al. \(2005\)](#). Moreover, none of these studies calibrated the recorded treatment for possible measurement errors. Our interest lies in estimating the net effect of HAART on the onset of AIDS or death among HIV-positive patients realistically by calibrating for the measurement error in the observed treatment and accounting for the dropout of patients from the treatment.

We applied our proposed method to the data which are from the multicenter AIDS Cohort Study (MACS) and Womens Interagency HIV Study (WIHS). Data in these two studies were collected from physical examination, blood test, and questionnaire which were conducted every six months on the patients. The period of analysis is from January 1996, soon after the Food and Drug Agency (FDA) approved HAART in December 6, 1995, to the end of 2007.

In our analysis, we included 651 patients who were alive, HIV-positive, without clinical AIDS, and not using non-HAART antiretroviral therapy at the first visit of 1996. Among them, 205 participants developed clinical AIDS or died, and 581 participants reported that

they administrated HAART. There were 651 patients accounted for 2,136 person-visits which is equal to 1,068 person-visits/year. The median follow-up time is 1.5 years.

The baseline covariates include the age at the first visit, gender, CD4 cell counts (< 200 , $200 - 350$, $351 - 500$, and > 500), and viral load (≤ 4000 , $4001 - 10,000$, and $> 10,000$). The treatment history in the weight calculation at time k includes X_{k-1} , X_{k-2} , X_{k-3} , and the covariate histories are the restricted cubic splines of CD4 cell counts and logarithm to base 10 of viral load at time $k - 1$ with the knots at 5, 33, 67, 95 percentiles.

The parameters of the beta prior distribution of the classification rates were calculated based on the validation sub-study described in [Cole et al. \(2010\)](#). Specifically, η_1 , the probability of actually treated patients who were correctly identified, was equal to 0.84 and its standard deviation was equal to 0.02. The parameters of the prior distribution of η_1 were 252 and 48. η_2 , the probability of actually untreated patients who are correctly identified was 0.8 and its standard deviation was 0.04. The parameters of the prior distribution of η_2 were 80 and 20. We used the proportion of the truly treated patient as the prevalence of treatment. In order to assess the impact on the choice of the prior distribution of the correct classification rates, we shifted these two means in different direction by 5% and 10%. For comparison, we also performed analyses using the standard marginal structural Cox model. The results are presented in Table 7.

Both the proposed method and the standard marginal structural Cox model identified that HAART prolongs the time to developing clinical AIDS or the survival time among HIV-positive patient. However, the standard marginal structural Cox proportional hazards model biased the estimated coefficient down to 0 severely. [Cole et al. \(2010\)](#) summarized the previous similar studies and stated that the range of the estimated treatment effect was from 0.14 to 0.54. The estimate from the proposed method fell in that range, but the naive estimate did not. The results also show that the choice of the prior distribution of the correct-classification rates has little impact on the estimated treatment effect.

In practice, one may only have the range and the median of η_1 and η_2 , the method of [Hozo et al. \(2005\)](#) can be used to estimate the mean and the variance of η_1 and η_2 , which are required for estimating the parameters of the prior distribution. The range of η_1 and η_2 can be validated using the data from pilot study or the information from literatures. If none

of the information is available, one still can expect to obtain the estimated treatment effect with small bias given the prior mean of η_1 and η_2 are not biased in the same direction.

In this example, the sample size was 651 patients, but our simulation results suggest that small sample size (e.g. 1000) might result in unstable estimate. Our goal is to demonstrate the application of the proposed method and this is the only data set which is available for us. In order to obtain a relative stable estimates, we set the number of batches to 100000, and the batch length to 10. We repeated the analysis 10 times and all the estimated coefficients of treatment effect were very close to -0.82 .

2.7 CONCLUSIONS AND DISCUSSION

In this study, we incorporated the Bayesian method and the regression calibration method into the marginal structural Cox models to develop a new approach which will allow us to estimate the causal effect of treatment on the survival with possibly misclassified treatment. Our proposed method extends the work of [Cole et al. \(2010\)](#) which required a validation data set whereas a validation data set is not needed in our study. We use the prior distribution to capture the uncertainty about the correct-classification rates. Our simulation results show that even though the mean of the prior distribution of the correct-classification rates is different from the true underlying correct-classification rates, our proposed method still provides an estimate which is close to the true treatment effect in various scenarios. Therefore, no parameters are assumed known exactly when applying our proposed method. This will have a significant impact on statistical analysis. Before our studies, it was difficult for investigators to account for measurement error without having a validation set because that a validation data set was often used for correcting bias in the estimate caused by measurement error in the predictor. Our proposed method offers a new approach in causal inference which is particularly useful when the investigator has little information about the measurement error and a validation data set is unavailable.

It is worth noting that our simulation results also show that our proposed method will greatly reduce bias in estimating the true treatment effect at a cost of precision. Nevertheless,

just as what [Carroll \(2011\)](#) indicated in his lecture, no strategy can relieve this caveat thus far. The mean squared error (MSE) of the estimate could be larger in our proposed method than that in the naive method. Thus, we do not suggest to use our proposed method if the effect of the predictor is expected to be small since bias corrected is finite, but the variance of the estimate is inflated.

The bias of the estimation from our proposed method is larger when the risk of event is high ($\geq 35\%$), as compared with that when the risk of event is low. This result coincides with the studies of [Prentice \(1982\)](#) and [Rosner et al. \(1989\)](#), which required that the event risk is low or the disease is rare. [Spiegelman et al. \(1997\)](#) also stated that the relationship $\hat{\theta} = \hat{\beta}/\hat{\gamma}_1$ holds in a Cox model under the assumption of low event rate. [Andersen and Liestøl \(2003\)](#) did a simulation study to assess the impact of high risk of event on the estimate of treatment effect and found that high event rate could be another attenuating factor in estimation under the Cox model framework, but the effect is moderate.

Our simulation results show that small sample size might result in unstable estimates and the estimated treatment effect obtained from simulation data sets could spread widely around the true parameter. Therefore, one would expect a large variance in the estimate and thus a lower statistical power when the sample size is small. This feature might restrict the application of our proposed method to data with a large sample. However, just as what we showed in the above example, the investigator might obtain a relatively stable estimate by increasing the number of batches.

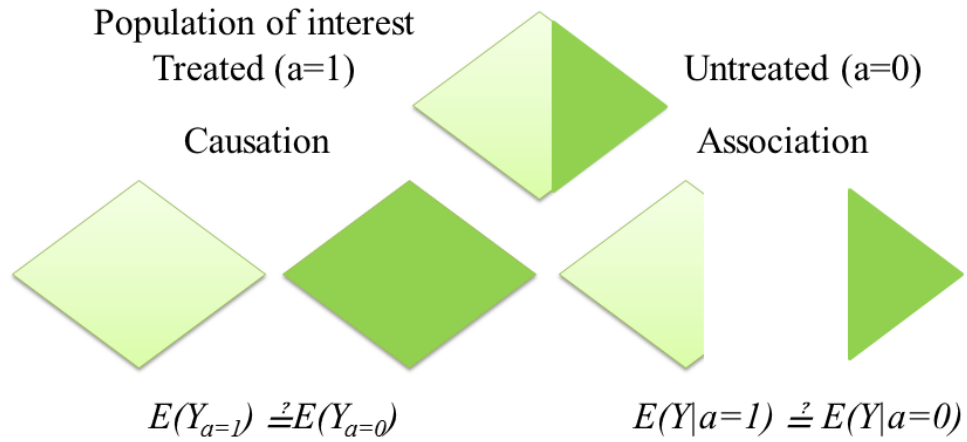
The difference between the naive estimated treatment effect and the estimate from our proposed method is close to $\hat{\beta}(1 - 1/\gamma_1)$, but not close to $\hat{\beta}(1 - 1/\hat{\gamma}_1)$ in the scenarios which the event risk and censoring are low and the proposed method works well. For example, in the scenario that event risk is equal to 0.15, the censoring rate is equal to 0, and both η_1^* and η_2^* are equal to 0.85, the difference between the naive estimate and our estimate is 0.077, $\hat{\beta}(1 - 1/\gamma_1)$ is 0.085, but $\hat{\beta}(1 - 1/\hat{\gamma}_1)$ is -0.02 . This is because $\hat{\gamma}_1$ obtained by maximizing $L_p(\theta)$ with θ replaced by $\hat{\theta}$ is 1.13, which should be about 0.707. In general, our method does not estimate γ_1 accurately. However, the $var(\hat{\theta})$ calculated using the information of $\hat{\beta}, var(\hat{\beta}), \hat{\gamma}_1, var(\hat{\gamma}_1)$ is close to the empirical estimate when the event risk is low (e.g. $< 35\%$.)

Although the estimated treatment effect from our proposed method is much less biased as compared with that from the naive estimate, the bias is still significant after shifting the mean of the prior distribution of η_1 and η_2 from the original settings of η_1 and η_2 in the same direction by 5% simultaneously. Moreover, the estimates of η_1 and η_2 from their posterior distributions were not updated from the data, so the estimates depended on their prior settings.

Besides misclassification in the treatment, there are other factors which could bias toward 0 in estimating the treatment causal effect on survival. [Andersen and Liestøl \(2003\)](#) showed that in longitudinal studies, non-synchronously updating the measurements of variables and ageing can also attenuate the estimated predictor effect. In this study, we only took treatment misclassification into consideration.

In applying the method of [Cole et al. \(2010\)](#), there are two sources of restriction in low event rate. The first one requires low event rate in the regression calibration method. [Rosner et al. \(1989\)](#) proved that $\hat{\theta} = \hat{\beta}/\hat{\gamma}_1$ holds in logistic regression when the event rate is low, and the measurement error is small and non-differential. [Spiegelman et al. \(1997\)](#) stated that the same relationship holds in Cox proportional hazards model under the assumptions that the event rate is low, the measurement error is small and non-differential, and the censoring is non-informative. The second source of requirement of low event rate was established when, using a pooled logistic regression model to approximate a weighted time-dependent Cox model. [D'Agostino et al. \(1990\)](#) showed that the estimator from the pooled logistic regression is close to that from the Cox proportional hazards model when the event rate is low. [Cole et al. \(2010\)](#) stated that their method works well when the event rate in each time interval is less than 10%. In applying our proposed method, we can fit a marginal structural Cox model directly via the input data with the counting process style of [Fleming and Harrington \(1991\)](#). This will then remove the restriction of low event rate from the second source so that the allowable event rate is expected to exceed 10% when using our proposed method. The simulation results of our proposed method show that the allowable event rate increases to at least 25%.

APPENDIX: TABLES AND FIGURES



Adapted based on figure 1.1 in Causal inference, Miguel A. Hernán, James Robins 2011

http://www.tc.umn.edu/~alonso/hernanrobins_v1.10.11.pdf

1

Figure 1: Difference between causation and association.

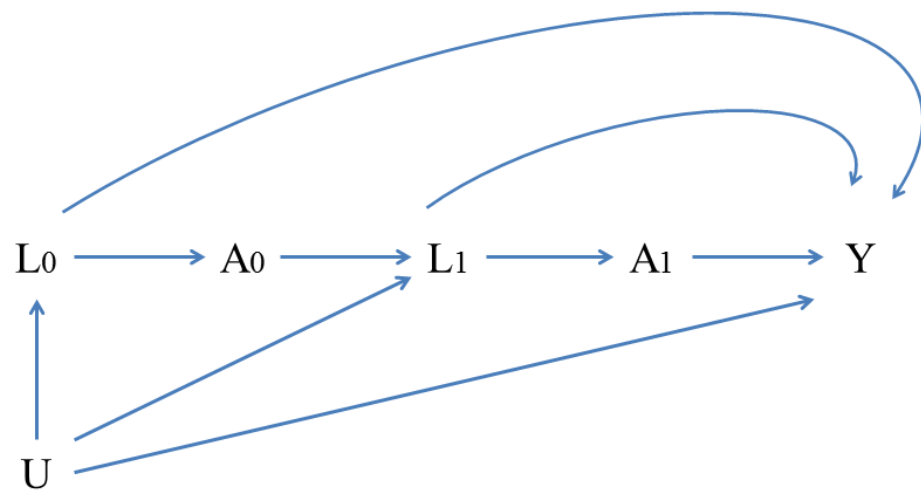
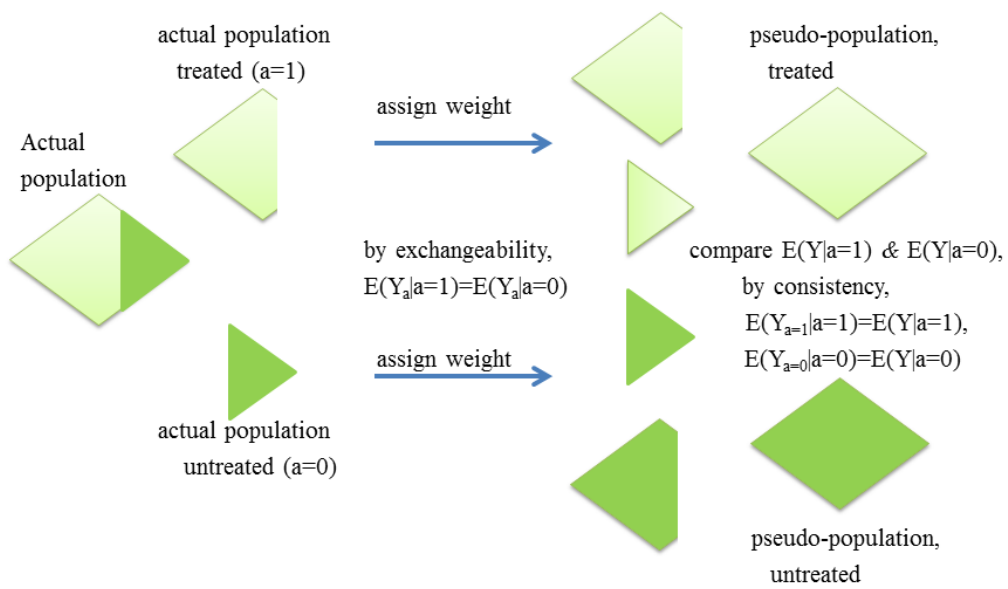


Figure 2: Conditions under which standard methods fail to give consistent estimate.



1

Figure 3: Underlying idea of marginal structural models.

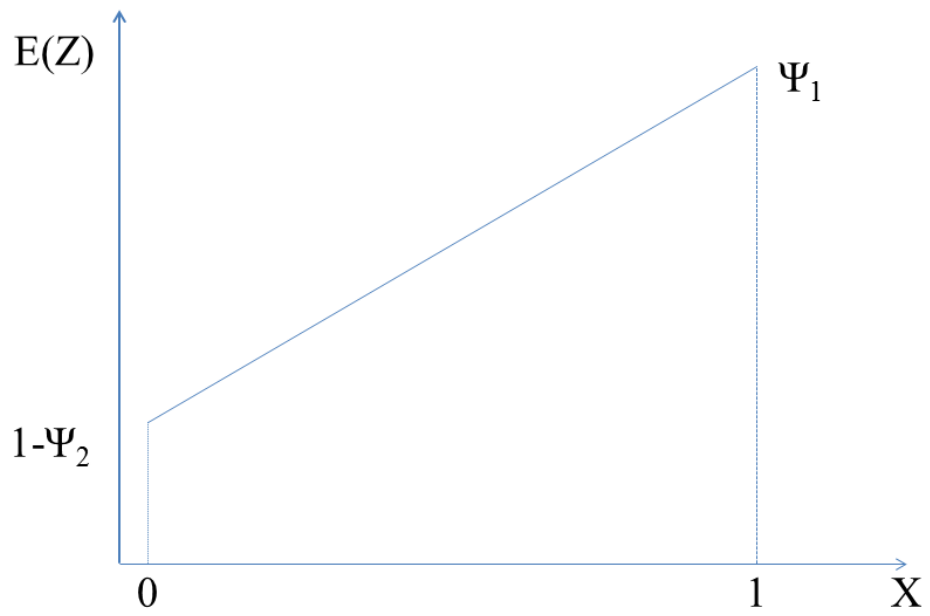


Figure 4: Relation between the true treatment and observed treatment.

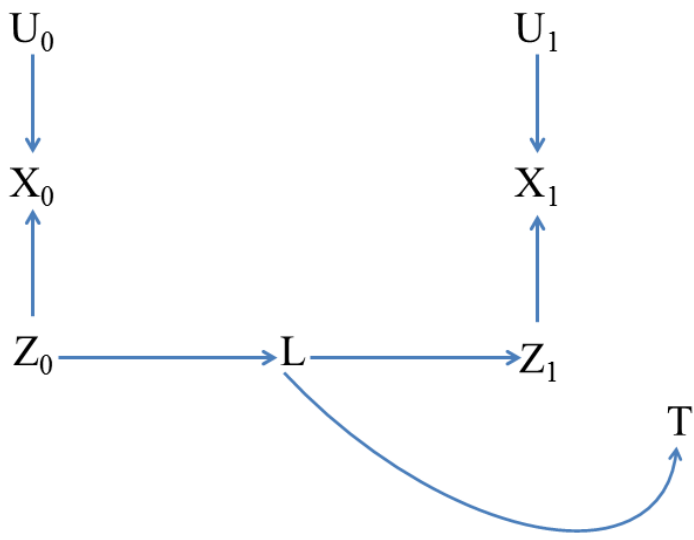


Figure 5: Variables relationship in simulated data set.

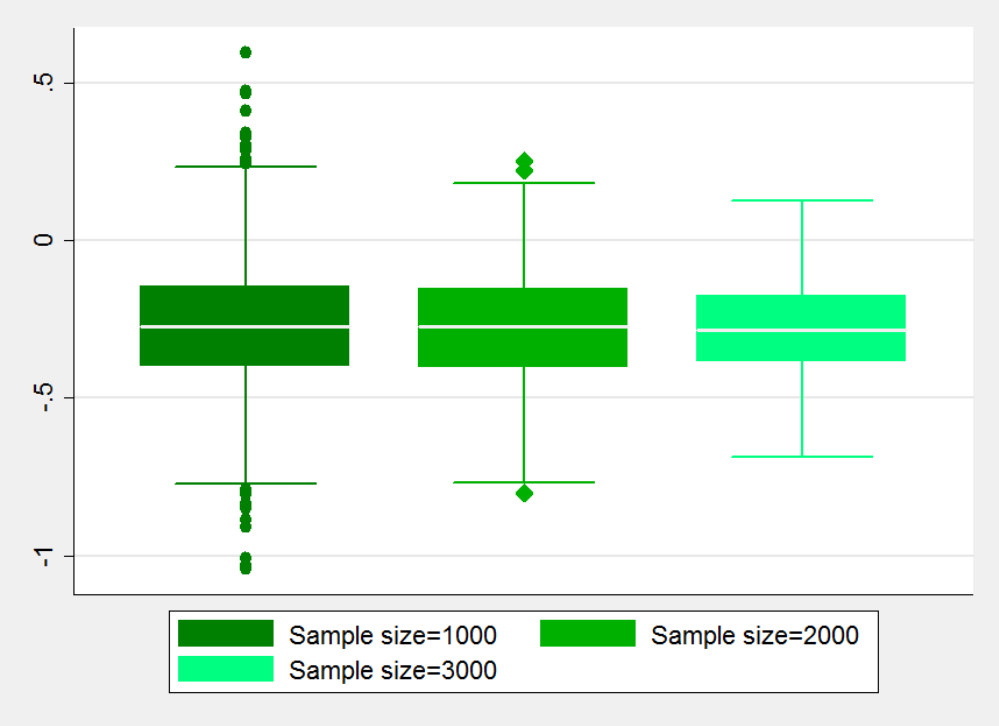


Figure 6: Comparison of dispersion of the estimated treatment effect in different sample size in scenario 1

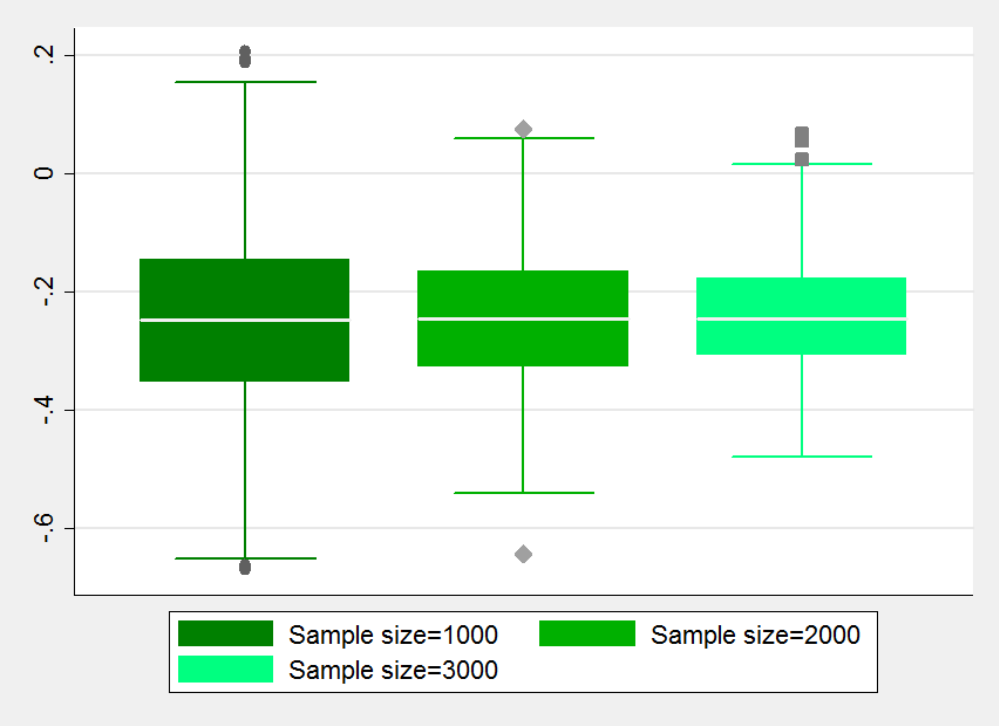


Figure 7: Comparison of dispersion of the estimated treatment effect in different sample size in scenario 2

Table 1: Simulation results with beta prior distribution of parameter comparing the bias, the Monte Carlo standard error (MCSE), the standard deviation estimates (SD), the coverage probability (95%) of the different analysis methods when the event risk is low (15%)

<i>risk, censoring, η_1^*, η_2^*</i>	Proposed method				Naive method			
	Bias	MCSE	SD	CP	Bias	MCSE	SD	CP
0.15, 0.00, 0.90, 0.80	0.0025	0.1476	0.1496	0.8600	0.0760	0.1073	0.1078	0.8860
0.15, 0.00, 0.85, 0.85	-0.0011	0.1521	0.1511	0.8640	0.0757	0.1096	0.1079	0.8880
0.15, 0.00, 0.80, 0.90	0.0049	0.1498	0.1489	0.8720	0.0776	0.1073	0.1080	0.8980
0.15, 0.00, 0.85, 0.75	-0.0470	0.1811	0.1773	0.8840	0.0758	0.1118	0.1077	0.8920
0.15, 0.00, 0.95, 0.85	0.0498	0.1275	0.1278	0.8080	0.0855	0.1073	0.1079	0.8740
0.15, 0.15, 0.90, 0.80	0.0188	0.1476	0.1487	0.8500	0.0881	0.1073	0.1118	0.8860
0.15, 0.15, 0.85, 0.85	0.0119	0.1543	0.1511	0.8560	0.0850	0.1118	0.1118	0.8880
0.15, 0.15, 0.80, 0.90	0.0174	0.1610	0.1491	0.8120	0.0870	0.1185	0.1120	0.8640
0.15, 0.15, 0.85, 0.75	-0.0449	0.1878	0.1792	0.8700	0.0761	0.1163	0.1120	0.8840
0.15, 0.15, 0.95, 0.85	0.0297	0.1230	0.1361	0.8780	0.0695	0.1006	0.1118	0.9340
0.15, 0.30, 0.90, 0.80	0.0184	0.1610	0.1532	0.8360	0.0870	0.1185	0.1163	0.8900
0.15, 0.30, 0.85, 0.85	0.0080	0.1655	0.1569	0.8340	0.0818	0.1185	0.1162	0.8860
0.15, 0.30, 0.80, 0.90	0.0067	0.1521	0.1556	0.8480	0.0790	0.1118	0.1159	0.8960
0.15, 0.30, 0.85, 0.75	-0.0294	0.1789	0.1796	0.8560	0.0860	0.1118	0.1159	0.9120
0.15, 0.30, 0.95, 0.85	0.0408	0.1498	0.1350	0.8020	0.0781	0.1252	0.1161	0.8700

Bias=estimated treatment effect $-(-0.28)$; η_1^* : mean of the prior distribution of η_1 ; η_2^* : mean of the prior distribution of η_2 ;

Table 2: Simulation results with beta prior distribution of parameter comparing the bias, the Monte Carlo standard error (MCSE), the standard deviation (SD), the coverage probability (95%) of the different analysis methods when the event risk is moderate (25%)

<i>risk, censoring, η_1^*, η_2^*</i>	Proposed method				Naive method			
	Bias	MCSE	SD	CP	Bias	MCSE	SD	CP
0.25, 0.00, 0.90, 0.80	0.0171	0.1096	0.1276	0.9060	0.0827	0.0813	0.0825	0.8380
0.25, 0.00, 0.85, 0.85	0.0027	0.1140	0.1309	0.9220	0.0850	0.0728	0.0825	0.8380
0.25, 0.00, 0.80, 0.90	0.0111	0.1118	0.1284	0.9100	0.0827	0.0763	0.0825	0.8380
0.25, 0.00, 0.85, 0.75	-0.0305	0.1342	0.1494	0.9260	0.0850	0.0796	0.0824	0.8320
0.25, 0.00, 0.95, 0.85	0.0520	0.0984	0.1111	0.8760	0.0819	0.0850	0.0825	0.8440
0.25, 0.15, 0.90, 0.80	0.0202	0.1051	0.1304	0.9100	0.0821	0.0850	0.0856	0.8680
0.25, 0.15, 0.85, 0.85	0.0224	0.1163	0.1298	0.8980	0.0863	0.0872	0.0856	0.8200
0.25, 0.15, 0.80, 0.90	0.0194	0.1163	0.1296	0.9100	0.0819	0.0872	0.0855	0.8480
0.25, 0.15, 0.85, 0.75	-0.0248	0.1297	0.1528	0.9240	0.0829	0.0850	0.0855	0.8500
0.25, 0.15, 0.95, 0.85	0.0598	0.0939	0.1119	0.8660	0.0875	0.0805	0.0855	0.8440
0.25, 0.30, 0.90, 0.80	0.0194	0.1118	0.1285	0.8880	0.0785	0.0872	0.0834	0.8260
0.25, 0.30, 0.85, 0.85	0.0225	0.1140	0.1287	0.8920	0.0872	0.0830	0.0834	0.8200
0.25, 0.30, 0.80, 0.90	0.0229	0.1118	0.1284	0.9040	0.0815	0.0872	0.1159	0.8340
0.25, 0.30, 0.85, 0.75	-0.0297	0.1230	0.1521	0.9380	0.0758	0.0805	0.0834	0.8540
0.25, 0.30, 0.95, 0.85	0.0529	0.0984	0.1121	0.8720	0.0783	0.0872	0.0834	0.8520

Bias=estimated treatment effect $-(-0.28)$; η_1^* : mean of the prior distribution of η_1 ; η_2^* : mean of the prior distribution of η_2 ;

Table 3: Simulation results with beta prior distribution of parameter comparing the bias, the Monte Carlo standard error (MCSE), the standard deviation (SD), the coverage probability (95%) of the different analysis methods when the event risk is high (35%)

$risk, censoring, \eta_1^*, \eta_2^*$	Proposed method				Naive method			
	Bias (mcse)	MCSE (mcse)	SD (mcse)	CP (mcse)	Bias (mcse)	MCSE (mcse)	SD	CP
0.35, 0.00, 0.90, 0.80	0.0232	0.0872	0.1167	0.9460	0.0788	0.0693	0.0702	0.8040
0.35, 0.00, 0.85, 0.85	0.0178	0.0962	0.1179	0.9200	0.0767	0.0760	0.0701	0.7940
0.35, 0.00, 0.80, 0.90	0.0285	0.0872	0.1156	0.9300	0.0831	0.0671	0.0701	0.7680
0.35, 0.00, 0.85, 0.75	-0.0138	0.0984	0.1357	0.9580	0.0834	0.0671	0.0700	0.7900
0.35, 0.00, 0.95, 0.85	0.0566	0.0783	0.1015	0.8940	0.0789	0.0693	0.0700	0.7980
0.35, 0.15, 0.90, 0.80	0.0186	0.0917	0.1195	0.9280	0.0737	0.0738	0.0719	0.8160
0.35, 0.15, 0.85, 0.85	0.0197	0.0894	0.1204	0.9520	0.0761	0.0693	0.0718	0.8240
0.35, 0.15, 0.80, 0.90	0.0362	0.0894	0.1163	0.9260	0.0872	0.0693	0.0719	0.7580
0.35, 0.15, 0.85, 0.75	-0.0114	0.1073	0.1374	0.9480	0.0839	0.0716	0.0718	0.7840
0.35, 0.15, 0.95, 0.85	0.0602	0.0783	0.1030	0.8760	0.0804	0.0693	0.0719	0.7940
0.35, 0.30, 0.90, 0.80	0.0357	0.0939	0.1187	0.9220	0.0844	0.0738	0.0737	0.7820
0.35, 0.30, 0.85, 0.85	0.0311	0.0962	0.1199	0.9000	0.0827	0.0760	0.0737	0.7860
0.35, 0.30, 0.80, 0.90	0.0263	0.0939	0.1206	0.9120	0.0774	0.0738	0.0737	0.8120
0.35, 0.30, 0.85, 0.75	-0.0158	0.1073	0.1415	0.9440	0.0787	0.0738	0.0737	0.7920
0.35, 0.30, 0.95, 0.85	0.0582	0.0783	0.1052	0.9020	0.0766	0.0716	0.0737	0.8300

Bias=estimated treatment effect-(-0.28); η_1^* : mean of the prior distribution of η_1 ; η_2^* : mean of the prior distribution of η_2 ;

Table 4: Simulation results with uniform prior distribution of parameter comparing the bias, the Monte Carlo standard error (MCSE), the standard deviation estimates (SD), the coverage probability (95%) of the different analysis methods when the event risk is low (15%)

<i>risk, censoring, η_1^*, η_2^*</i>	Proposed method				Naive method			
	Bias	MCSE	SD	CP	Bias	MCSE	SD	CP
0.15, 0.00, 0.90, 0.80	0.0001	0.1497	0.1512	0.8740	0.0763	0.1082	0.1080	0.8960
0.15, 0.00, 0.85, 0.85	0.0078	0.1640	0.1502	0.8320	0.0829	0.1179	0.1081	0.8580
0.15, 0.00, 0.80, 0.90	0.0110	0.1448	0.1487	0.8580	0.6032	0.1044	0.1080	0.8820
0.15, 0.00, 0.85, 0.75	-0.0356	0.1745	0.1739	0.8560	0.0822	0.1079	0.1078	0.8700
0.15, 0.00, 0.95, 0.85	0.0383	0.1326	0.1314	0.8280	0.0782	0.1098	0.1080	0.8860
0.15, 0.15, 0.90, 0.80	0.0198	0.1510	0.1495	0.8460	0.0890	0.1096	0.1118	0.8840
0.15, 0.15, 0.85, 0.85	0.0135	0.1498	0.1518	0.8460	0.0864	0.1077	0.1119	0.8860
0.15, 0.15, 0.80, 0.90	0.0034	0.1477	0.1540	0.8640	0.0773	0.1073	0.1120	0.9000
0.15, 0.15, 0.85, 0.75	-0.0454	0.1920	0.1791	0.8640	0.0763	0.1192	0.1116	0.8980
0.15, 0.15, 0.95, 0.85	0.0402	0.1390	0.1332	0.8040	0.0777	0.1149	0.1119	0.8800
0.15, 0.30, 0.90, 0.80	0.0014	0.1572	0.1564	0.8520	0.0751	0.1148	0.1161	0.9140
0.15, 0.30, 0.85, 0.85	0.0109	0.1626	0.1558	0.8440	0.0848	0.1159	0.1160	0.8960
0.15, 0.30, 0.80, 0.90	0.0112	0.1673	0.1550	0.8300	0.0837	0.1218	0.1162	0.8740
0.15, 0.30, 0.85, 0.75	-0.0335	0.1988	0.1800	0.8240	0.0833	0.1237	0.1162	0.8620
0.15, 0.30, 0.95, 0.85	0.0469	0.1324	0.1333	0.8160	0.0832	0.1114	0.1160	0.9040

Bias=estimated treatment effect $-(-0.28)$; η_1^* : mean of the prior distribution of η_1 ; η_2^* : mean of the prior distribution of η_2 ;

Table 5: Simulation results with uniform prior distribution of parameter comparing the bias, the Monte Carlo standard error (MCSE), the standard deviation (SD), the coverage probability (95%) of the different analysis methods when the event risk is moderate (25%)

<i>risk, censoring, η_1^*, η_2^*</i>	Proposed method				Naive method			
	Bias	MCSE	SD	CP	Bias	MCSE	SD	CP
0.25, 0.00, 0.90, 0.80	0.0135	0.1083	0.1287	0.9100	0.0790	0.0805	0.0825	0.8400
0.25, 0.00, 0.85, 0.85	0.0107	0.1056	0.1300	0.9260	0.0784	0.0793	0.0823	0.8420
0.25, 0.00, 0.80, 0.90	0.0191	0.1102	0.1270	0.9020	0.0821	0.0822	0.0823	0.8220
0.25, 0.00, 0.85, 0.75	-0.0233	0.1328	0.1476	0.9320	0.0832	0.0853	0.0824	0.8120
0.25, 0.00, 0.95, 0.85	0.0521	0.0937	0.1115	0.8700	0.0833	0.0800	0.0824	0.8420
0.25, 0.15, 0.90, 0.80	0.0130	0.1099	0.1297	0.9160	0.0755	0.0832	0.0836	0.8480
0.25, 0.15, 0.85, 0.85	0.0276	0.1071	0.1278	0.8980	0.0893	0.0809	0.0836	0.8180
0.25, 0.15, 0.80, 0.90	0.0183	0.1105	0.1296	0.9060	0.0810	0.0837	0.0837	0.8460
0.25, 0.15, 0.85, 0.75	-0.0240	0.1268	0.1503	0.9160	0.0818	0.0827	0.0837	0.8300
0.25, 0.15, 0.95, 0.85	0.0498	0.0922	0.1138	0.8920	0.0808	0.0795	0.0837	0.8720
0.25, 0.30, 0.90, 0.80	0.0308	0.1082	0.1266	0.8960	0.0872	0.0830	0.0833	0.8240
0.25, 0.30, 0.85, 0.85	0.0255	0.1079	0.1282	0.9020	0.0849	0.0807	0.0832	0.8080
0.25, 0.30, 0.80, 0.90	0.0262	0.1042	0.1274	0.9100	0.0844	0.0806	0.0832	0.8300
0.25, 0.30, 0.85, 0.75	-0.0161	0.1284	0.1491	0.9180	0.0846	0.0837	0.0845	0.8260
0.25, 0.30, 0.95, 0.85	0.0583	0.1006	0.1141	0.8500	0.0871	0.0868	0.0866	0.8280

Bias=estimated treatment effect $-(-0.28)$; η_1^* : mean of the prior distribution of η_1 ; η_2^* : mean of the prior distribution of η_2 ;

Table 6: Simulation results with uniform prior distribution of parameter comparing the bias, the Monte Carlo standard error (MCSE), the standard deviation (SD), the coverage probability (95%) of the different analysis methods when the event risk is high (35%)

$risk, censoring, \eta_1^*, \eta_2^*$	Proposed method				Naive method			
	Bias (mcse)	MCSE (mcse)	SD (mcse)	CP (mcse)	Bias (mcse)	MCSE (mcse)	SD	CP
0.35, 0.00, 0.90, 0.80	0.0263	0.0903	0.1162	0.9300	0.0823	0.0699	0.0702	0.7860
0.35, 0.00, 0.85, 0.85	0.0251	0.0863	0.1172	0.9420	0.0823	0.0665	0.0701	0.7980
0.35, 0.00, 0.80, 0.90	0.0235	0.0884	0.1163	0.9260	0.0785	0.0689	0.0701	0.7920
0.35, 0.00, 0.85, 0.75	-0.0133	0.0994	0.1356	0.9600	0.0836	0.0658	0.0701	0.8040
0.35, 0.00, 0.95, 0.85	0.0561	0.0788	0.1022	0.8860	0.0808	0.0697	0.0701	0.8020
0.35, 0.15, 0.90, 0.80	0.0330	0.0871	0.1172	0.9120	0.0847	0.0684	0.0719	0.8020
0.35, 0.15, 0.85, 0.85	0.0273	0.0900	0.1190	0.9180	0.0825	0.0697	0.0718	0.8020
0.35, 0.15, 0.80, 0.90	0.0193	0.0900	0.1191	0.9380	0.0732	0.0712	0.0719	0.8300
0.35, 0.15, 0.85, 0.75	-0.0077	0.1070	0.1362	0.9520	0.0851	0.0719	0.0718	0.7540
0.35, 0.15, 0.95, 0.85	0.0646	0.0814	0.1019	0.8740	0.0848	0.0731	0.0718	0.7820
0.35, 0.30, 0.90, 0.80	0.0371	0.0911	0.1197	0.9200	0.0866	0.0720	0.0750	0.7960
0.35, 0.30, 0.85, 0.85	0.0264	0.0959	0.1205	0.9280	0.0786	0.0761	0.0729	0.0802
0.35, 0.30, 0.80, 0.90	0.0257	0.0904	0.1204	0.9320	0.0759	0.0723	0.0729	0.8340
0.35, 0.30, 0.85, 0.75	-0.0026	0.1131	0.1396	0.9300	0.0795	0.0765	0.0730	0.9300
0.35, 0.30, 0.95, 0.85	0.0701	0.0796	0.1040	0.8700	0.0884	0.0729	0.0749	0.7940

Bias=estimated treatment effect $-(-0.28)$; η_1^* : mean of the prior distribution of η_1 ; η_2^* : mean of the prior distribution of η_2 ;

Table 7: Effect of HAART on the time to developing clinical AIDS or death among HIV-positive patients

Model	Hazard ratio	Confidence interval
Model 1	0.613	0.457-0.821
Model 2	0.436	0.228-0.833
Model 3	0.403	0.198-0.819
Model 4	0.402	0.203-0.796

Model 1: standard marginal structural Cox model;
 Model 2: the proposed method with $\eta_1^* = 0.84$ and $\eta_2^* = 0.8$;
 Model 3: the proposed method with $\eta_1^* = 0.79$ and $\eta_2^* = 0.85$;
 Model 4: the proposed method with $\eta_1^* = 0.74$ and $\eta_2^* = 0.90$;

BIBLIOGRAPHY

- P. K. Andersen and K. Liestøl. Attenuation caused by infrequently updated covariates in survival analysis. *Biostatistics*, 4:633–649, 2003.
- J. Bartlett. Correcting for covariate measurement error using regression calibration, 2011. URL <http://csm.lshtm.ac.uk/files/2011/10/Jonathan-14-10-2011.pdf>.
- N. E. Breslow. Covariance analysis of censored survival data. *Biometrics*, 30:89–100, 1974.
- R. J. Carroll. Measurement error in covariates, 2011. URL http://www.stat.tamu.edu/~carroll/talks/NCI_MEM_Call.pdf.
- R. J. Carroll, D. Ruppert, and L. A. Stefanski. *Measurement error in nonlinear models*. Chapman and Hall, London, 1995.
- R. J. Carroll, D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu. *Measurement error in nonlinear models: a modern perspective*. Chapman and Hall/CRC, New York, 2006.
- CDC. Statistics overview, 2011. URL <http://www.cdc.gov/hiv/statistics/basics/>.
- S. R. Cole, H. Chu, and S. Greenland. Multiple-imputation for measurement-error correction. *International Journal of Epidemiology*, 35:1074–1081, 2006.
- S. R. Cole, L. P. Jacobson, P. C. Tien, L. Kingsley, J. S. Chmiel, and K. Anastos. Using marginal structural measurement-error models to estimate the long-term effect of antiretroviral therapy on incident aids or death. *American Journal of Epidemiology*, 171:113 – 122, 2010.
- R. B. D’Agostino, M-L. Lee, A. J. Belanger, L. A. Cupples, K. Anderson, and W. B. Kannel. Relation of pooled logistic regression to time dependent cox regression analysis: the framingham heart study. *Statistics in Medicine*, 9:1501–1515, 1990.
- T. R. Fleming and D. P. Harrington. *Counting processes and survival analysis*. Wiley, New York, 1991.
- M. A. Hernán, B. Brumback, and J. M. Robins. Marginal structural models to estimate the causal effect of zidovudine on the survival of hiv-positive men. *Epidemiology*, 11:561 – 570, 2000.

- S. P. Hozo, B. Djulbegovic, and I. Hozo. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Medical Research Methodology*, 5:5–13, 2005.
- P. Hu, A. A. Tsiatis, and M. Davidian. Estimating the parameters in the cox model when covariate variables are measured with error. *Biometrics*, 54:1407 – 1419, 1998.
- Y. Li and L. Ryan. Survival analysis with heterogeneous covariate measurement error. *Journal of the American Statistical Association*, 99:724 – 735, 2004.
- A. McGlothlin, J. D. Stamey, and Jr J. W. Seaman. Binary regression with misclassified response and covariate subject to measurement error: a bayesian approach. *Biometrical Journal*, 50:123 – 134, 2008.
- T. Nakamura. Corrected score function of errors-in-variables models: Methodology and application to generalized linear model. *Biometrika*, 77:127 – 137, 1990.
- T. Nakamura. Proportional hazards model with covariates subject to measurement error. *Biometrics*, 48:829 – 838, 1992.
- R. L. Prentice. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 69:331 – 342, 1982.
- G. K. Reeves, D. R. Cox, S. C. Darby, and E. Whitley. Some aspects of measurement error in explanatory variables for continuous and binary regression models. *Statistics in Medicine*, 17:2157 – 2177, 1998.
- J. M. Robins. Causal inference from complex longitudinal data. In: *M. Merikane, ed. Latent Variable Modelling and Application to Causality. Lecture Notes in Statistics*, 120:69 – 117, 1997.
- J. M. Robins. Marginal structural models. *1997 Proceedings of the American Statistical Associations, Section on Bayesian Statistical Science*, pages 1 – 10, 1998.
- J. M. Robins and M. Á. Hernán. Estimation of the causal effects of time-varying exposures, 2008. URL http://isites.harvard.edu/fs/docs/icb.topic156289.files/General/chapter_12jun07-1.pdf.
- B. Rosner, W. C. Willett, and D. Spiegelman. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine*, 8:1051 – 1069, 1989.
- B. Rosner, D. Spiegelman, and W. C. Willett. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *American Journal of Epidemiology*, 132:734 – 745, 1990.
- B. Rosner, D. Spiegelman, and W. C. Willett. Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. *American Journal of Epidemiology*, 136:1400 – 1413, 1992.

- D. Sinha, J. G. Ibrahim, and M-H. Chen. A bayesian justification of cox's partial likelihood. *Biometrika*, 90:629–641, 2003.
- D. Spiegelman, A. McDermott, and B. Rosner. Regression calibration method for correcting measurement-error bias in nutritional epidemiology. *American Journal of Clinical Nutrition*, 65 (suppl):1179s – 86s, 1997.
- D. Spiegelman, B. Ronser, and R. Logan. Estimation and inference for logistic regression with covariate misclassification and measurement error in main study/validation study designs. *Journal of the American Statistical Association*, 95:51–61, 2000.
- J. A. Sterne, M. A. Hernán, B. Ledergerber, K. Tilling, R. Weber, P. Sendi, M. Rickenbach, J. M. Robins, M. Egger, and Swiss HIV Cohort Study. Long-term effectiveness of potent antiretroviral therapy in preventing aids and death: a prospective cohort study. *Lancet*, 366:378–384, 2005.
- M. G. Tadesse, J. G. Ibrahim, R. Gentleman, S. Chiaretti, J. Ritz, and R. Foa. Bayesian error-in-variable survival model for the analysis of genechip arrays. *Biometrics*, 61:488 – 497, 2005.
- H. K. Tam, Z.-F. Zhang, L. P. Jacobson, J. B. Margolick, J. S. Chmiel, C. Rinaldo, and R. Detels. Effect of highly active antiretroviral therapy on survival among hiv-infected men with kaposi sarcoma or non-hodgkin lymphoma. *International Journal of Cancer*, 98: 916–922, 2002.
- A. A. Tsiatis and M. Davidian. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika*, 88:447 – 458, 2001.
- K. H. Wong, K. C. W. Chan, and S. S. Lee. Delayed progression to death and to aids in a hong kong cohort of patients with advanced hiv type 1 disease during the era of highly active antiretroviral therapy, 2004. URL <http://cid.oxfordjournals.org/content/39/6/853.full.pdf+html>.