2017

# Marginal Structural Cox Model for Survival Data with Treatment-Confounder Feedback

Yanan Zhang
*University of South Carolina*

Marginal Structural Cox Model for Survival Data with Treatment-Confounder Feedback

by

Yanan Zhang

Bachelor of Medicine

Shandong University, 2012

---

Submitted in Partial Fulfillment of the Requirements

For the Degree of Mater of Science in Public Health in

Biostatistics

The Norman J. Arnold School of Public Health

University of South Carolina

2017

Accepted by:

Jiajia Zhang, Director of Thesis

James W. Hardin, Reader

Xuemei Sui, Reader

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

# ACKNOWLEDGEMENTS

I would like to thank my thesis committee chair, Dr. Jiajia Zhang for her generous help in both research and life. Thank you for teaching me and encouraging me so patiently. Also, I want to thank my two committee members, Dr. Xuemei Sui and Dr. James W. Hardin. In addition to this thesis, I have worked with Dr. Sui and Dr. Zhang on another research project which led to a publication. I have improved a lot under your guidance. Many thanks to Dr. Hardin. He shared with me a lot of learning materials and gave me useful suggestions. I appreciate the advice and encouragement of my committee members. It made my study in USC very happy and productive.

I also would like to thank the professors of Department of Epidemiology and Biostatistics, my fellow classmates, friends, and family. Thank you all for the support over the last two years. This experience has been priceless.

# ABSTRACT

In an observational longitudinal study, there can be time-varying exposure/treatment and time-varying confounders. When the confounders affect the exposure and prior exposure also has an impact on levels of confounders, there is treatment confounder feedback. To admit estimation of unbiased causal effects, these conditions need to be hold, exchangeability, positivity, consistency. The traditional method of conditioning on potential confounders does not meet these 3 conditions. Therefore, parameter estimates from traditional Cox model are biased casual effect estimates when the treatment confounder feedback exists. The marginal structural Cox model can be used to address this issue. By calculating and including inverse probability (IP) weights, the impact of confounding can be removed. Estimates from models with IP weights are interpreted as the causal effect that comparing always in treatment group vs. never in treatment group.

In this study, first, I introduced basic concepts of causal inference, treatment confounder feedback and the marginal structural model; detailed steps of calculating IP weights and model fitting. In simulation study, I compared the time-dependent Cox models and the marginal structural Cox model; Also, for the marginal model, results using three types of IP weights were compared: un-stabilized weight, stabilized weight, and stabilized weight considering censoring. Performance metrics of each method

were evaluated based on their bias, percentage bias, empirical standard deviation, standard error and coverage probability of 95% confidence intervals. Aerobics Center Longitudinal Study (ACLS) data were used to explore the causal effect of cardiorespiratory fitness on hypertension incidence. Overweight or obese is a risk factor of hypertension. We hypothesized that cardiorespiratory fitness may help lower BMI via physical exercise, while reduced BMI or improved overweight status may promote cardiorespiratory fitness. Thus, there exists cardiorespiratory (treatment) overweight (confounder) feedback, and the marginal structural Cox model may deepen our understanding of association between hypertension and CRF through ACLS data.

TABLE OF CONTENTS

LIST OF TABLES

# CHAPTER 1

# INTRODUCTION

## 1.1 CAUSAL INFERENCE

To explore the effect of an intervention on an outcome, ideally, we want to have outcomes of subjects with intervention and outcomes of the same subjects without intervention. Then the intervention effect would be the difference under two intervention conditions. Usually, we don't observe both outcomes. So, there have been many different proposed techniques, like randomization and matching to compare 2 groups of subjects (with and without exposure). Under various approaches, the difference between treatment groups is reasonable to represent the true effect of the intervention.

Let random variable A be a binary treatment, and let a represent the value of A (a=1 as treated, and a=0 as untreated). Further, let Y represent the observed outcome and L represent a vector of confounders. $Y^{a=1}$ is defined as the counterfactual outcome over all subjects in the population had they been treated. $Y^{a=0}$ is defined as the counterfactual outcome over all subjects in the population had they been untreated. The average causal effect in population is defined as $E[Y^{a=1}]-E[Y^{a=0}]$. Because definition is not conditional on other variables, it is also called marginal causal effect. (1)

In randomized clinical trials, randomization can eliminate the effect of confounding, so results can be explained as causal effect. In an observational study, when the confounding is controlled, the conditional effect is a consistent estimator of the causal effect, $E[Y=1|a=1] = E[Y^{a=1}]$. To get an unbiased estimate of the casual effect from an observational study, three assumptions need to hold (2):

1) Exchangeability:

Participants with treatment would have the same outcome as those without treatment had they not received the treatment. Similarly, participants without treatment would have the same outcome (as those with treatment) had they received the treatment. In other words, the observed treatment status is independent of counterfactual outcome, $Y^a$ is independent of A, for all a.

2) Positivity:

It is impossible to get the average effect of treatment, if all participants are in treatment group, or all people are untreated. A positive probability of accepting all treatment levels is required. $Pr[A=a|L=l] >0$, for all l with $Pr[L=l]$ not equal to 0.

3) Consistency:

The observed outcome under the observed treatment status equals the counterfactual outcome of the observed treatment status. If A=a, then $Y=Y^a$.

## 1.2 TIME-VARYING TREATMENT AND CONFOUNDERS

In a longitudinal study, we have time-varying treatment and time-varying confounders. Robins first introduced the effects of time-varying treatments in observational studies in 1986. (3) Suppose there are m+1 visits in the longitudinal

study. Let $A_j$ be the time-varying treatment at time j (j=0, 1, 2, ..., m), $L_j$ be the vector of time-varying confounders, Y be the outcome observed at time m+1. Allowing an overbar to denote the history, $\bar{A}_j$ is then the history of treatment, $\bar{A}_j$= ($A_0$, $A_1$, ...$A_j$); $\bar{L}_j$ is the history of confounders $\bar{L}_j$= ($L_0$, $L_1$, ...$L_j$). The lowercase letters represent the value of random variables. For binary treatment, if the treatment=1 at all visits, that is $\bar{A}_j$ = (1, 1, ..., 1), then it is "always treat". On the other hand, if $\bar{A}_j$ = (0, 0, ..., 0), then it is "never treat". The marginal causal effect is defined as the difference between the above 2 counterfactual effects (always treat vs. never treat), $E\left(Y^{\bar{a}=\bar{1}}\right) - E(Y^{\bar{a}=\bar{0}})$. (4)

Extending to time-varying treatments, to achieve valid causal inferences, all three conditions described in previous section need to hold. At each visit, if the time-varying treatment is unconfounded conditional on previous treatment history and confounders history, then exchangeability holds. Thus, at each time point, conditional exchangeability holds. It can also be called sequential exchangeability.

Sequential positivity is defined as,

$\Pr\left(A_j = a_j \middle| \bar{A}_{j-1} = \bar{a}_{j-1}, \bar{L}_j = \bar{l}_j\right) > 0$ , for all $\bar{a}_j$ and $\bar{l}_j$, if $Pr(\bar{A}_{j-1} = \bar{a}_{j-1}, \bar{L}_j = \bar{l}_j) \neq 0$.

Sequential consistency is defined as,

If $\bar{A} = \bar{a}$, then $Y = Y^{\bar{a}}$. If $\bar{A}_{j-1} = \bar{a}_{j-1}$, then $\bar{L}_j = \bar{L}_j^{\bar{a}}$. (4)

## 1.3 TREATMENT CONFOUNDER FEEDBACK

In a longitudinal study, variables are observed repeatedly. At each time point, there is a set of observations including treatment and potential confounders. However, their relationship might complicate the analysis. When the confounders affect the

treatment and previous treatment also affects levels of confounders, there is treatment confounder feedback. (5)

For example, in the Aerobics Center Longitudinal Study (ACLS), to investigate the effect of cardiorespiratory fitness (CRF) on hypertension incidence, participants were enrolled and were followed up from 1974 to2003. Details of ACLS study were described in Chapter 4. At each visit, information collected included age, sex, CRF, body mass index (BMI), smoking and heavy drinking status, family history of hypertension, diagnosis of hypertension, diabetes and hypercholesterolemia, etc. Overweight or obese is a risk factor of hypertension. The improved overweight or obese status will help increase CRF level. On the other hand, increased CRF level caused by the increased physical activity help reduce body fat percentage, BMI and change the overweight status. Therefore, there is treatment confounder feedback.

## 1.4 TIME DEPENDENT COX MODEL AND WHY FAIL

In the survival analysis setting, to estimate the effect of treatment on the outcome, one may estimate a time-dependent Cox model with baseline confounders. Another way is modeling the time-dependent Cox model with time-varying confounders.

Let $A_j$ be the time-dependent exposure, $L_0$ be the baseline covariate vectors, and $L_j$ be the time-varying covariate vectors. $\lambda_0(j)$ is the baseline hazard function. Model with baseline covariates,

$$\lambda\big(j|Z(j)\big) = \lambda_0(j)\exp\big(\boldsymbol{\beta Z(j)}\big) = \lambda_0(j)\exp\big(\boldsymbol{\beta_1 A_j} + \boldsymbol{\beta_2 L_0}\big)$$

Model with time-varying covariates,

$$\lambda\big(j|Z(j)\big) = \lambda_0(j)\exp\big(\boldsymbol{\beta Z(j)}\big) = \lambda_0(j)\exp\big(\boldsymbol{\beta_3 A_j} + \boldsymbol{\beta_4 L_j}\big)$$

$\exp(\beta_1)$ is the constant hazard ratio of the exposure conditional on levels of baseline covariates. $\exp(\boldsymbol{\beta}_3)$ is the hazard ratio of the exposure conditional on levels of covariates at time j.

To estimate the parameter $\boldsymbol{\beta}$, the partial likelihood method is used. The partial likelihood can be treated as the product of conditional probability that subject i fails from the risk set at time $\boldsymbol{J_i}$. The partial likelihood is:

$$L(\boldsymbol{\beta}) = \prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} \left[ \frac{\boldsymbol{exp}\,\{\boldsymbol{\beta Z_i(J_i)}\}}{\sum_{\boldsymbol{t\in R(j\geq J_i)}} \boldsymbol{exp}\,\{\boldsymbol{\beta Z_t(J_i)}\}} \right]^{\boldsymbol{\delta_i}}$$

where $R(j \geq J_i)$ is the risk set at time $J_i$, and $\delta_i$ (1=censored and 0=uncensored) is the censor status at time $J_i$. It is the product of conditional probability that subject i fails from the risk set at time $J_i$.

By solving the derivative of log[L(β)=0], the solution, denoted as $\hat{\beta}$, is the maximum likelihood estimate of parameter β. The hazard at time j depends on the variable of treatment and confounders at that time. The regression effect of treatment A and confounders L are constant over time.

The time dependent Cox model with baseline confounders only uses the baseline information, and the confounding can't be fully controlled. Thus, the estimation effect from this model would be biased.

For the time-dependent Cox model using time-varying confounders, by simply adjusting for the time-varying confounders, at each visit, sequential exchangeability would not hold. This is because the confounder at time j is influenced by previous exposure or treatment. (5)

Marginal structural Cox models (MSMs) can be used to get the causal inference of treatment in the presence of treatment confounder feedback. An early variant of an MSM was developed by Dr. Marian Pugh in 1993, to solve the problem of missing data. (6) Dr. James Robins and Miguel Hernán from Harvard first published the general approach in 1999. (7) The idea is to apply weights to eliminate the confounding on treatment and/or censoring (in survival setting), thus allowing unbiased causal effects to be estimated. Marginal means the model estimates the marginal distribution instead of the conditional distribution. Structural refers to the causal inference.

## 1.5 OUTLINE OF THESIS

The main aim of this thesis is to understand the marginal structural Cox model and to apply this model in real data analysis.

In Chapter 2, we will explain the inverse probability (IP) weighting and the marginal structural Cox model. In Chapter 3, there is a simulation study and comparisons are made between the results of a traditional time-dependent Cox model and a marginal structural Cox model. In Chapter 4, we will apply this method using the Aerobics Center Longitudinal Study (ACLS) data. In Chapter 5, there is discussion and conclusion.

# CHAPTER 2

# IP WEIGHTING AND MARGINAL STRUCTURAL COX MODEL

## 2.1 IP WEIGHTING

Inverse probability (IP) weighting is a method commonly used in survey sampling to adjust for the sample selection process and get unbiased estimates. (8, 9) Each observation is weighted by the reciprocal of the predicted probability of the observed exposure status. There are 2 properties of IP weighting. First, using IP weights, exposure is unconfounded. Second, the effect of exposure on outcome is the same as in the true study population. (10) To briefly explain the idea of IP weighting, I simulated data with variable exposure, sex and outcome Y (sample size=2000). The distribution of sex is not balanced for two exposure groups.

First, binary variable exposure (Yes=1 vs. No=0) was generated based on binomial distribution (n=2000, p=0.5). For exposed group, sex (women=1 vs. men=0) was generated based on binomial with p=0.3; while for unexposed group, sex was on binomial with p=0.7. Error was generated based on normal distribution with mean being 0 and variance being 0.01. Outcome Ys equal to the sum of intercept (true value 0.5), effect of sex (true value 0.2 times sex), effect of exposure (true value -0.6 times

exposure) and random error. $Y = 0.5 + 0.2 * female - 0.6 * exposure + error$

As Table 2.1, in unweighted count, the sex distribution between two exposure groups was different, so sex might be a confounder. IPW weight was calculated as 1/Pr [exposure| sex]. For example, the weight for men in unexposed group was 3.29, which meant contributing 3.29 times observation in pseudo population, that it 306*3.29=1006. By adjusting for weights, the distribution of sex between two groups was balanced. If we conduct linear regression of Y on exposure, in unweighted case, the estimated effect of exposure was biased (point estimate -0.68). However, after adjusting for weights, the estimated effect was not biased (point estimate -0.60).

Table 2.1 Sex distribution between unexposed and exposed groups

| | Men | Women |
|---|---|---|
| Unweighted count | | |
| Unexposed | 306 (30.42%) | 700 (70.42%) |
| Exposed | 700 (69.58%) | 294 (29.58%) |
| Weights | | |
| Unexposed | 3.29 | 1.42 |
| Exposed | 1.44 | 3.38 |
| Weighted count | | |
| Unexposed | 1006 (50%) | 994 (50%) |
| Exposed | 1006 (50%) | 994 (50%) |

There are different ways of calculating the weight. Weights should be chosen so that 1) the exposure is unconfounded; 2) effects in pseudo population are the same as in the true study population; and 3) weights are ‘’as close as possible to 1” to prevent extreme weights and reduce variance. (10)

In the context of time-dependent treatment, the weights are time-varying at different observed times or visits for the same subject. From the start (j=0) to the end of follow-up (j=m), IP weighting is based on the overall probability of the subject receiving his or her own observed history of treatment $A_{j=0}$to $A_{j=m}$, the product of visit specific probabilities. According to published papers, different calculations lead to weights that are either un-stabilized or stabilized. (1, 10, 11) An un-stabilized weight is the inverse of the estimated probability that a subject received the observed treatment, given the baseline covariates $L_0$ (not time-varying), history of treatments up to visit time j-1, $\bar{A}_{j-1}$, and history of confounder up to visit time j, $\bar{L}_j$. If the probability is small, then the inverse of probability would be large and large weights lead to unstable results.

$$w_m{}^T = \prod_{j=0}^{m} \frac{1}{\Pr(A_j = a_j | \bar{A}_{j-1} = \bar{a}_{j-1}, L_0 = l_0, \bar{L}_j = \bar{l}_j)}$$

Stabilized weights of treatment have the same denominator, instead of 1 as the numerator, it uses the estimated probability that a subject received the observed treatment, given the baseline covariates and history of treatments up to visit time j-1. History of confounder was not included. In this way, the variability of stabilized weights is smaller and the resulting calculations of the weights are much closer to 1.

$$sw_m{}^T = \prod_{j=0}^{m} \frac{\Pr(A_j = a_j | \bar{A}_{j-1} = \bar{a}_{j-1}, L_0 = l_0)}{\Pr(A_j = a_j | \bar{A}_{j-1} = \bar{a}_{j-1}, L_0 = l_0, \bar{L}_j = \bar{l}_j)}$$

The right censoring is very common to be seen in survival data, which is caused by lost to follow up or the end of study. Most of time, we assume that censoring is informative. Applying the same idea of dealing confounders by weights, weights can

solve the problems of informative censoring. The denominator of stabilized weights for censoring is the probability of subjects not censored at time j, given their treatment history till j-1, baseline covariates, and the history of confounder up to time j-1. The numerator is the probability without further conditional on the history of the confounder. The final stabilized weights are product of stabilized weights of treatment and censor, ${sw_m}^T \times {sw_m}^C$.

$$ {sw_m}^C = \prod_{j=0}^{m} \frac{\Pr(C_j = 0 | \bar{C}_{j-1} = 0, \bar{A}_{j-1} = \bar{a}_{j-1}, L_0 = l_0)}{\Pr(C_j = 0 | \bar{C}_{j-1} = 0, \bar{A}_{j-1} = \bar{a}_{j-1}, L_0 = l_0, \bar{L}_{j-1} = \bar{l}_{j-1})} $$

The treatment can be binary, multinomial, and continuous. The inverse probability weights can be calculated from different models. For example, to fit the pooled logistic regression for binary treatment; to fit the multinomial regression for categorical treatment. Next, we introduce the detailed calculation of stabilized weights of treatment and censoring by fitting four pooled logistic models.

**Step 1**: Data preparation. Data are organized into the long format (for each subject, there are multiple rows of observations – one observation per time period). Variables include participants' ID, start time of each visit $T_j$, end time of visit $T_{j+1}$, exposure at the start of the visit interval $A_j$, potential confounders (time-independent confounders $L_0$ and time-varying confounders $L_j$), censor indicator$C_j$, and outcomes $Y_j$.

**Step 2**: Fitting pooled logistic models and getting the estimated probability. Model 1: logit pr($\mathrm{A_j}$ =1) = $\overline{\mathrm{A}}_{\mathrm{j-1}} + \mathrm{L}_0$ . If the observed treatment $\mathrm{A_j} = 1$, then the estimated probability of the observed treatment equals to the probability of treatment

at time j. If $A_j = 0$, then the estimated probability of the observed treatment equals to 1 minus the estimated probability of treatment. Model 2: logit pr($A_j$ =1) = $\bar{A}_{j-1} + L_0 + \bar{L}_j$. Model 3: logit pr($C_j = 0$) = $\bar{A}_{j-1} + L_0$. We assume that once subjects are censored, they will not come back to the study. We estimate the probability that a subject remains uncensored at time j. Model 4: logit pr($C_j$ =1) = $\bar{A}_{j-1} + L_0 + \bar{L}_{j-1}$.

**Step 3**: Combing weights of treatment and censoring. The numerator of the stabilized weights $sw_m{}^T \times sw_m{}^C$, can be estimated by multiplying the estimated probability of the observed treatment at time j (from Model 1) and probability of remaining uncensored till time j (from Model 3). The denominator of the weights can be estimated by multiplying the estimated probability from Model 2 and Model 4.

**Step 4**: So far, weights are calculated at each time point during follow-up. In the final step, we need to calculate the cumulative product over all previous times up to j. For example, weights at time 3, are production of estimated $sw_m{}^T \times sw_m{}^C$ (results of step 3) at time 1, time 2 and time 3. Weight at time 5 are production of weights of the first 5 times.

If the models to estimate the weights are correctly specified, then by incorporated the calculated weight into the final model of interest, the confounding of treatment and censor will be eliminated.

To estimate the inverse probability weights, there is an R package “ipw”. For longitudinal data, the iptm function can compute weights at each time point during

follow-up. The exposure can be continuous, binomial, multinomial, or ordinal. Both stabilized and un-stabilized weights can be estimated.

## 2.2 MARGINAL STRUCTURAL COX MODEL

By using IP weights, the confounding due to time-dependent covariates is removed, and the hazard function of the marginal structural Cox model is as follows,

$$\lambda_{J\bar{a}} = \lambda_0(j)\exp(\beta_1 A_j)$$

Parameter β can be estimated using the partial likelihood method. 95% confidence interval for β can be calculated using bootstrap methods or by computing analytic variance estimates, or using robust variance estimates.

$$\hat{\beta} \pm 1.96 \times \sqrt{\widehat{var(\beta)}}$$

The outcome variable of marginal Cox model is a counterfactual since it uses the pseudo-population. Therefore, it is called structural mean model. The IP weighted estimates causation of the marginal structural model. Parameters can be interpreted as the mean hazard ratio of event if everybody was always treated comparing to if everybody was never being treated.

# CHAPTER 3. SIMULATION STUDY

The aims of this chapter are: Firstly, to compare the estimates from the time dependent Cox models and the marginal structural Cox model. Another aim is to compare the performance of un-stabilized weights and stabilized weights in the marginal structural model. In addition, we will check how estimates vary when the sample size, censoring rate or the true effects change.

## 3.1 GENERATING AND PREPARING DATA

500 samples, each with n subjects (n=500 or n=2500) and 10 visits were generated according to the algorithm described in Young et al (2008). (12) Corresponding SAS code is provided at https://cdn1.sPH.harvard.edu/wp-content/uploads/sites/148/2012/10/simulate_snaftm.txt.

For each sample,

**Step 1**: Simulate the counterfactual $T_0$ from an exponential distribution with scale parameter $\lambda_0$ ($\lambda_0 = 0.01$ or $\lambda_0 = 0.1$). Define $L_{-1} = A_{-1} = Y_0 = 0$. For each j ϵ [0, 9] implement steps 2-4:

**Step 2**: Simulate time varying confounders $L_j$ from

$$logit\left[\left[\Pr\left(L_j = 1\middle|\ \bar{L}_{j-1}, \bar{A}_{j-1}, T0,\ Y_j = 0; \boldsymbol{\beta}\right)\right] = \beta_0 + \beta_1 I(T0 < c) + \beta_2 A_{j-1} + \beta_3 L_{j-1}\right.,$$ set $\boldsymbol{\beta}$= ($\beta_0$, $\beta_1$, $\beta_2$, $\beta_3$)=(log(3/7),2, log(1/2),log(3/2)) and c=30

**Step 3**: Simulate $A_j$ from $logit\left[A_j = 1\middle|\bar{L}_j, \bar{A}_{j-1}, Y_j = 0, \boldsymbol{\alpha}\right] = \alpha_0 + \alpha_1 L_j + \alpha_2 L_{j-1} + \alpha_3 A_{j-1}$, set $\boldsymbol{\alpha}$= ($\alpha_0$,$\alpha_1$,$\alpha_2$,$\alpha_3$)=(log (2/7), 1/2, 1/2, log(4))

**Step 4**: simulate $Y_{j+1}$and possible T

If $T_0 > \int_0^{j+1} \exp\{\varphi_a \times A_m\}\, d_m$ then $Y_{j+1}$=0;

else $Y_{j+1}$=1, T=$j + (T_0 - \int_0^j \exp\{\varphi_a \times A_m\}\, d_m) \exp\{-\varphi_a \times A_j\}$

( $\varphi_a$=0.3, 0 or -0.3)

To explore the effect of sample size on effect estimation, for each sample, we generate n=500 subjects as an example of small sample size, and n=2500 subjects as an example of a large sample size. For step 1, the counterfactual time was generated from an exponential distribution with constant rate of monthly events $\lambda_0$ throughout the follow-up. $\lambda_0 = 0.01$ is for rare incidence of event. $\lambda_0 = 0.1$ is for relatively common occurrence of event. It also defined that before the start of study, there is no confounder $L_{-1} = 0$, subjects are not treated $A_{-1} = 0$, remain uncensored and without event occurrence $Y_0 = 0$.

Step 2 defines time varying confounders, which are affected by the previous treatment $A_{j-1}$ and confounders $L_{j-1}$. c is an arbitrary cutoff point, which affects the degree to which $T0$ affects $L_j$ for a chosen value of c. For step 3, treatment is affected

by confounders observed this time $L_j$ and previous time $L_{j-1}$, and previous treatment $A_{j-1}$. In step 4, true value of marginal effect of treatment is $\varphi_a$. Three values were simulated respectively, negative effect -0.3, null effect 0 and positive effect 0.3.

From these data generation steps, we see that $L_j$ is associated with outcome $Y_{j+1}$ via indicator variable. $L_j$ predicts future treatment $A_j$ $and$ $A_{j+1}$; $A_{j-1}$ has an impact on $L_j$. There is treatment-confounder feedback.

Data structure

Table 3.1 shows the 'long-format data structure. For each subject, there are at most 10 visits. For example, there are 10 visits for ID=1 and ID=2. Only 1 visit for ID=13, that is because the event occurs at time=0.93188. 7 visits for ID=20 because the event occurs at time=8.513671875. Time was cut into visit intervals from tpoint2 to tpoint, 0-1, 1-2, 2-3, …, 9-10. During each time interval, A is the treatment status at the start of time, Am1 is the previous treatment; L is the confounder, Lm1 and Lm2 are the confounder history of previous 2 visits. Y is the binary outcome. If no event occurs at the end of the 10th time interval, this subject was censored, censor_r=1. Similarly, Ym is the outcome for the end of previous visits. T0 is the generated counterfactual time for censored subjects, T is the observed time for participants without censoring.

Table 3.1 Example of long-formatted data structure

| ID | A | Am1 | L | Lm1 | Lm2 | Y | Ym | T | T0 | IT0 | tpoint | tpoint2 | censor_r |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | | 21.140 | 1 | 1 | 0 | 0 |
| 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | | 21.140 | 1 | 2 | 1 | 0 |
| 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | | 21.140 | 1 | 3 | 2 | 0 |
| 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | | 21.140 | 1 | 4 | 3 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | | 21.140 | 1 | 5 | 4 | 0 |
| 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | | 21.140 | 1 | 6 | 5 | 0 |
| 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | | 21.140 | 1 | 7 | 6 | 0 |
| 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | | 21.140 | 1 | 8 | 7 | 0 |
| 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | | 21.140 | 1 | 9 | 8 | 0 |
| 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | | 21.140 | 1 | 10 | 9 | 1 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | 181.648 | 0 | 1 | 0 | 0 |
| 2 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | | 181.648 | 0 | 2 | 1 | 0 |
| 2 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | | 181.648 | 0 | 3 | 2 | 0 |
| 2 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | | 181.648 | 0 | 4 | 3 | 0 |
| 2 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | | 181.648 | 0 | 5 | 4 | 0 |
| 2 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | | 181.648 | 0 | 6 | 5 | 0 |
| 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | | 181.648 | 0 | 7 | 6 | 0 |
| 2 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | | 181.648 | 0 | 8 | 7 | 0 |
| 2 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | | 181.648 | 0 | 9 | 8 | 0 |
| 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | | 181.648 | 0 | 10 | 9 | 1 |
| 13 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0.932 | 0.932 | 1 | 1 | 0 | 0 |
| 20 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 8.514 | 8.514 | 1 | 1 | 0 | 0 |
| 20 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 8.514 | 8.514 | 1 | 2 | 1 | 0 |
| 20 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 8.514 | 8.514 | 1 | 3 | 2 | 0 |
| 20 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 8.514 | 8.514 | 1 | 4 | 3 | 0 |
| 20 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 8.514 | 8.514 | 1 | 5 | 4 | 0 |
| 20 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 8.514 | 8.514 | 1 | 6 | 5 | 0 |
| 20 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 8.514 | 8.514 | 1 | 7 | 6 | 0 |
| 20 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 8.514 | 8.514 | 1 | 8 | 7 | 0 |
| 20 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 8.514 | 8.514 | 1 | 9 | 8 | 0 |

## 3.2 COMPUTING WEIGHT

A pooled logistic regression was fitted to estimate IP weights. Four models were fitted. Model 1 is for the numerator of $sw_m{}^T$, logit pr(A=1) = Am1; model 2 for denominator of $sw_m{}^T$, logit pr(A=1) = Am1+L+Lm1; Model 3 for numerator of $sw_m{}^C$, logit pr(censor_r=0) =Am1; model 4 for denominator of $sw_m{}^C$, logit pr(censor_r=0) =Am1 + Lm1 +Lm2.

Predictions of treatment A from model 1 and 2 are estimated. To get the probability of observed treatment, we did the following calculation. If treatment A=1, then probability of observed treatment equals the prediction; if treatment A=0, then probability of observed treatment equals 1 minus the prediction. The time varying weight is the production of previous weight from the visit 1 to the end of current visit. Prediction of censoring is estimated from model 3 and 4. The time-varying weight for censoring is the production of previous weight from the first visit to the end of current visit.

Three different weights are calculated,

$$w_m{}^T = \prod_{j=0}^{m} \frac{1}{predictions\ from\ model\ 2},$$

$$sw_m{}^T = \prod_{j=0}^{m} \frac{predictions\ from\ model\ 1}{predictions\ from\ model\ 2},$$

$$sw_m{}^T \times sw_m{}^C = \prod_{j=0}^{m} \frac{predictions\ from\ model\ 1\ and\ 3}{predictions\ from\ model\ 2\ and\ 4}.$$

The SAS code is attached, please see Appendix A.

## 3.3 MODEL FITTING

The marginal Cox models with 3 different weights were fitted using function Coxph() in the R package 'survival'. The dependent variable was the time and event status, and independent variable was treatment. Cluster () was specified to obtain

robust sandwich variance estimates of the coefficients. The Efron approximation was used for handling ties (multiple events at the same discrete time point).

To compare results from the marginal structural Cox model, the time-dependent Cox models were also fitted. One was a time-dependent model with baseline covariates L0. Another model was fitted using time-dependent confounder.

In the longitudinal data, one subject has several observations. To get the variance of estimators, the interclass correlation needs to be considered because the observations are not independent. Due to the computational difficulty of getting the exact estimates of variance, the robust standard errors were estimated based on the modified sandwich variance estimator. Based on the normal approximation, the 95% confidence intervals can be computed by ± 1.96 times the robust standard error. The variance could also have been obtained by bootstrapping. However, that takes a considerable amount of time to run, so in this study, the robust variance estimates were used.

## 3.4 PERFORMANCE METRICS

The performance of different models was assessed by the following measures:

Bias: $\sum_{i=1}^{N}(\widehat{\varphi}_i - \varphi)\ /N$, the average difference between N (N=1000 for large sample size and N=500 for small sample size) estimated parameters and true value.

Percentage Bias: $\frac{Bias}{True\ value\ \varphi} \times 100\%$. Only for true value -0.3 and 0.3.

Empirical Standard Deviations: Standard deviation of N estimated parameters.

Standard Error: the average of N estimated standard errors of parameters.

Means of Standard Error: $Bias^2 + \text{Empirical Standard Deviation}^2$

Coverage Probability of 95% confidence intervals: proportion of N samples in which the true parameters are contained in the 95% confidence interval.

## 3.5 RESULTS

Results are listed in Table 3.2, 3.3 and 3.4.

**Models**

The marginal Cox model with stabilized weights ${sw_m}^T$ and ${sw_m}^T \times {sw_m}^C$ performed better than the model using un-stabilized weights. The model with two stabilized weights had smaller bias and empirical standard deviations, and the coverage probability reached around 95%. For the two marginal structural models with stabilized weights, bias, empirical standard deviations, standard error, MSE and coverage probability were comparable. Censoring in the generated data was not informative. When there is informative censoring, ${sw_m}^T \times {sw_m}^C$ is expected to behave better than ${sw_m}^T$.

For the two time-dependent Cox models, the model using time-varying covariates had smaller bias and bigger coverage than model using baseline covariates. However, the bias was still big and the real coverage probability didn't reach 95%.

Comparing estimates of marginal Cox model with time-dependent Cox models, marginal structural Cox models had smaller bias. The marginal models with stabilized weights and two time-dependent Cox models had similar empirical standard

deviations. Their empirical standard deviations were comparable to model based standard error, which indicated that the model fitted well. MSE was a combination of bias and empirical standard deviation. Marginal Cox model with stabilized weights ${sw_m}^T$ and ${sw_m}^T \times {sw_m}^C$ had the smallest MSE.

**Sample size**

Comparing to the performance metrics in small sample size (n=500), results from large sample (n=2500) had smaller bias, empirical SD, model based SE, and MSE.

**Incidence rate of event**

The censoring rate in the simulated data is about 90% for rare event $\lambda_0 = 0.01$, and about 30%-40% for common event, $\lambda_0 = 0.1$. The coverage probability was larger for bigger incidence rate or smaller censoring rate. Especially, in the two time-varying Cox models, the coverage improved a lot when incidence rate increase from 0.01 to 0.1.

**True effect of treatment**

Performance metrics of models were consistent in simulated data with three different true effect of treatment, null effect, positive and negative effects. Different true effects of exposure don't; impact the performance of models.

Table 3.2 Performance Metrics of models with null true effect

| True effect=0, $\lambda_0$=0.01, sample size=500, censor rate=0.904 | | | | | | |
|---|---|---|---|---|---|---|
| **Models** | **Bias** | **%bias** | **StDev** | **SE** | **MSE** | **Coverage** |
| Marginal Cox Model | | | | | | |
| w | 0.064 | NA | 0.968 | 0.697 | 0.941 | 0.844 |
| sw | 0.025 | NA | 0.326 | 0.329 | 0.107 | 0.950 |
| swc | 0.025 | NA | 0.327 | 0.331 | 0.108 | 0.950 |
| Time dependent Cox Model | | | | | | |
| Baseline L0 | 0.338 | NA | 0.298 | 0.296 | 0.203 | 0.804 |
| Time dependent Lm | 0.293 | NA | 0.300 | 0.296 | 0.176 | 0.834 |
| True effect=0, $\lambda_0$=0.01, sample size=2500, censor rate=0.904 | | | | | | |
| Marginal Cox Model | | | | | | |
| w | 0.008 | NA | 0.416 | 0.390 | 0.173 | 0.944 |
| sw | 0.007 | NA | 0.152 | 0.146 | 0.023 | 0.942 |
| swc | 0.008 | NA | 0.153 | 0.147 | 0.024 | 0.936 |
| Time dependent Cox Model | | | | | | |
| Baseline L0 | 0.332 | NA | 0.131 | 0.130 | 0.127 | 0.286 |
| Time dependent Lm | 0.283 | NA | 0.132 | 0.131 | 0.098 | 0.444 |
| True effect=0, $\lambda_0$=0.1, sample size=500, censor rate=0.369 | | | | | | |
| Marginal Cox Model | | | | | | |
| w | 0.025 | NA | 0.478 | 0.391 | 0.229 | 0.880 |
| sw | 0.008 | NA | 0.125 | 0.122 | 0.016 | 0.950 |
| swc | 0.009 | NA | 0.124 | 0.122 | 0.016 | 0.954 |
| Time dependent Cox Model | | | | | | |
| Baseline L0 | 0.042 | NA | 0.118 | 0.115 | 0.016 | 0.926 |
| Time dependent Lm | 0.039 | NA | 0.118 | 0.115 | 0.015 | 0.936 |
| True effect=0, $\lambda_0$=0.1, sample size=2500, censor rate=0.368 | | | | | | |
| Marginal Cox Model | | | | | | |
| w | 0.001 | NA | 0.207 | 0.199 | 0.043 | 0.942 |
| sw | 0.000 | NA | 0.057 | 0.054 | 0.003 | 0.942 |
| swc | 0.000 | NA | 0.057 | 0.054 | 0.003 | 0.944 |
| Time dependent Cox Model | | | | | | |
| Baseline L0 | 0.035 | NA | 0.054 | 0.051 | 0.004 | 0.878 |
| Time dependent Lm | 0.031 | NA | 0.054 | 0.051 | 0.004 | 0.896 |

Note: w refers to un-stabilized weight; sw is stabilized weight; swc is stabilized weight considering censoring. StDev is empirical standard deviations.

Table 3.3 Performance Metrics of models with true effect being 0.3

| True effect=0.3, $\lambda_0$=0.01, sample size=500, censor rate=0.888 | | | | | | |
|---|---|---|---|---|---|---|
| **Models** | **Bias** | **%bias** | **StDev** | **SE** | **MSE** | **Coverage** |
| Marginal Cox Model | | | | | | |
| w | 0.075 | 24.867 | 0.868 | 0.667 | 0.759 | 0.862 |
| sw | 0.003 | 1.004 | 0.317 | 0.310 | 0.100 | 0.952 |
| swc | 0.004 | 1.324 | 0.318 | 0.312 | 0.101 | 0.952 |
| Time dependent Cox Model | | | | | | |
| Baseline L0 | 0.334 | 111.336 | 0.288 | 0.278 | 0.194 | 0.784 |
| Time dependent Lm | 0.283 | 94.348 | 0.287 | 0.279 | 0.163 | 0.828 |
| True effect=0.3, $\lambda_0$=0.01, sample size=2500, censor rate= 0.888 | | | | | | |
| Marginal Cox Model | | | | | | |
| w | 0.009 | 3.164 | 0.442 | 0.376 | 0.195 | 0.894 |
| sw | 0.005 | 1.811 | 0.140 | 0.138 | 0.020 | 0.952 |
| swc | 0.006 | 2.130 | 0.140 | 0.139 | 0.020 | 0.950 |
| Time dependent Cox Model | | | | | | |
| Baseline L0 | 0.335 | 111.590 | 0.122 | 0.123 | 0.127 | 0.214 |
| Time dependent Lm | 0.284 | 94.806 | 0.123 | 0.123 | 0.096 | 0.366 |
| True effect=0.3, $\lambda_0$=0.1, sample size=500, censor rate= 0.306 | | | | | | |
| Marginal Cox Model | | | | | | |
| w | 0.008 | 2.773 | 0.440 | 0.386 | 0.194 | 0.914 |
| sw | 0.002 | 0.558 | 0.126 | 0.117 | 0.016 | 0.926 |
| swc | 0.002 | 0.706 | 0.126 | 0.117 | 0.016 | 0.926 |
| Time dependent Cox Model | | | | | | |
| Baseline L0 | 0.038 | 12.555 | 0.117 | 0.110 | 0.015 | 0.926 |
| Time dependent Lm | 0.034 | 11.280 | 0.117 | 0.111 | 0.015 | 0.934 |
| True effect=0.3, $\lambda_0$=0.1, sample size=2500, censor rate= 0.307 | | | | | | |
| Marginal Cox Model | | | | | | |
| w | -0.007 | -2.196 | 0.217 | 0.194 | 0.047 | 0.928 |
| sw | -0.002 | -0.770 | 0.053 | 0.052 | 0.003 | 0.952 |
| swc | -0.002 | -0.706 | 0.053 | 0.052 | 0.003 | 0.956 |
| Time dependent Cox Model | | | | | | |
| Baseline L0 | 0.033 | 11.157 | 0.049 | 0.049 | 0.004 | 0.896 |
| Time dependent Lm | 0.030 | 9.928 | 0.050 | 0.049 | 0.003 | 0.908 |

Note: w refers to un-stabilized weight; sw is stabilized weight; swc is stabilized weight with censoring. StDev is empirical standard deviations.

Table 3.4 Performance Metrics of models with true effect being -0.3

| True effect=-0.3, $\lambda_0$=0.01, sample size=500, censor rate= 0.916 | | | | | | |
|---|---|---|---|---|---|---|
| **Models** | **Bias** | **%bias** | **StDev** | **SE** | **MSE** | **Coverage** |
| Marginal Cox Model | | | | | | |
| w | 0.094 | 31.365 | 1.006 | 0.725 | 1.021 | 0.834 |
| sw | 0.006 | -2.067 | 0.357 | 0.354 | 0.127 | 0.960 |
| swc | 0.007 | -2.208 | 0.358 | 0.355 | 0.128 | 0.956 |
| Time dependent Cox Model | | | | | | |
| Baseline L0 | 0.331 | -110.373 | 0.323 | 0.319 | 0.214 | 0.814 |
| Time dependent Lm | 0.280 | -93.411 | 0.322 | 0.320 | 0.182 | 0.864 |
| True effect=-0.3, $\lambda_0$=0.01, sample size=2500, censor rate= 0.917 | | | | | | |
| Marginal Cox Model | | | | | | |
| w | 0.013 | -4.197 | 0.474 | 0.414 | 0.225 | 0.914 |
| sw | 0.002 | 0.584 | 0.163 | 0.158 | 0.026 | 0.938 |
| swc | 0.001 | 0.411 | 0.164 | 0.159 | 0.027 | 0.936 |
| Time dependent Cox Model | | | | | | |
| Baseline L0 | 0.319 | -106.379 | 0.146 | 0.141 | 0.123 | 0.376 |
| Time dependent Lm | 0.270 | -90.109 | 0.146 | 0.142 | 0.094 | 0.538 |
| True effect=-0.3, $\lambda_0$=0.1, sample size=500, censor rate= 0.422 | | | | | | |
| Marginal Cox Model | | | | | | |
| w | 0.020 | 6.692 | 0.467 | 0.400 | 0.218 | 0.900 |
| sw | 0.001 | 0.316 | 0.133 | 0.129 | 0.018 | 0.952 |
| swc | 0.001 | 0.385 | 0.133 | 0.129 | 0.018 | 0.952 |
| Time dependent Cox Model | | | | | | |
| Baseline L0 | 0.031 | -10.222 | 0.123 | 0.121 | 0.016 | 0.946 |
| Time dependent Lm | 0.027 | -8.882 | 0.123 | 0.121 | 0.016 | 0.944 |
| True effect=-0.3, $\lambda_0$=0.1, sample size=2500, censor rate= 0.422 | | | | | | |
| Marginal Cox Model | | | | | | |
| w | 0.004 | 1.171 | 0.209 | 0.203 | 0.044 | 0.942 |
| sw | 0.000 | -0.118 | 0.054 | 0.057 | 0.003 | 0.966 |
| swc | 0.000 | -0.133 | 0.054 | 0.057 | 0.003 | 0.968 |
| Time dependent Cox Model | | | | | | |
| Baseline L0 | 0.033 | -10.972 | 0.050 | 0.054 | 0.004 | 0.926 |
| Time dependent Lm | 0.029 | -9.715 | 0.050 | 0.054 | 0.003 | 0.938 |

Note: w refers to un-stabilized weight; sw is stabilized weight; swc is stabilized weight considering censoring. StDev is empirical standard deviations.

To summarize, First, the estimate from the time dependent Cox model using time-varying confounders was better than the estimate from model using baseline confounders. Although the time-dependent Cox model is commonly used in practice, the estimates remained biased when there was treatment confounder feedback based on the simulation results.

Second, when there is treatment confounder feedback, the marginal structural Cox model should be applied to get unbiased estimates of the casual inference effect. Estimates with stabilized weights had smaller bias and variability, and larger coverage probability than those using un-stabilized weights. When existence of informative censoring, estimates with the stabilized weight considering censoring are expected to perform better than those without considering censoring.

# CHAPTER 4. ACLS DATA

Hypertension is a very common chronic disease and affects the health of numerous people. The risk of developing high blood pressure includes, age, race, family history, being overweight or obese, not being physical active, smoking, too much sodium diet, too little potassium in diet, heavy drinking, stress and some chronic diseases. (13)

Cardiorespiratory fitness (CRF) measures the ability of the circulatory and respiratory systems to supply oxygen to skeletal muscles during sustained physical activity. Studies have shown that CRF is inversely associated with the risk of hypertension. (14, 15)

Both CRF and overweight can work as independent risk factors of hypertension. Increased CRF level which caused by increased physical activity can help reduce body fat percentage, body mass index (BMI), thus improve overweight or obese status. The reduced BMI also help increase cardiorespiratory fitness level subsequently. Therefore, there exists a treatment-confounder feedback. That is, CRF-overweight feedback.

In this study, we want to explore that effect of CRF on hypertension incidence by using the marginal structural Cox model.

## 4.1 DATASET DESCRIPTION

Started in 1970, the Aerobics Center Longitudinal Study (ACLS) is a prospective cohort study aiming to investigate health outcomes associated with cardiorespiratory fitness and physical activity. In our study, 14290 participants who have completed a baseline examination at the Cooper Clinic (Dallas, Texas) during 1974–2003 were included. All participants were free of hypertension at baseline; at least 2 visits are available for each subject; they were able to achieve at least 85% of age-predicted maximal heart rate (220 minus age in years) at each visit; were free of history of heart attack, stroke, cancer, and abnormal ECG at baseline; subjects whose BMI less than 18.5 or greater than 80 were excluded; all have complete data on blood pressure, glucose, cholesterol, fitness, and BMI.

The study protocol was approved annually by the Institutional Review Board of the Cooper Institute and all participants provided written consent to participate in this follow-up study.

**Exposure/ Treatment**

Cardiorespiratory fitness (CRF) level was assessed as the duration of a symptom-limited maximal treadmill exercise test using a modified Balke protocol. (16, 17) The treadmill speed was 88 $m \cdot min^{-1}$ for the first 25 min. During this time, the grade was 0% for the first minute, 2% the second minute and increased 1% for each minute. After 25 min, the grade remained constant while the speed increased 5.4 $m \cdot min^{-1}$ each minute until test termination. Patients were encouraged to give a maximal effort

during the test. Maximal metabolic equivalents (METs, 1 MET = 3.5 ml $O_2$ uptake · $kg^{-1}$ · $min^{-1}$) were estimated from the final treadmill speed and grade. Maximal treadmill time was measured in minutes.

Subjects were divided into 3 groups, low (lowest 20%), middle (middle 40%) and high (upper 20%), according to the quantile of maximal treadmill time in each sex- and age-group (20-39, 40-49, 50-59, 60+) specific distribution from the overall ACLS population. The exposure or treatment was time-varying.

**Outcome and censoring**

All participants were followed from the date of their baseline examination until their occurrence of hypertension or December 31, 2003. Hypertension was defined as physician diagnosed high blood pressure or blood pressure >=140/90 mmHg. If a subject was diagnosed as hypertension, then the event occurs. While if a subject remained not being diagnosed as hypertension at the end of study, then this person was defined as censored. In log format data structure, the start and end of visits were from 0-1, 1-2, 2-3, …, until 25-26.

**Confounders**

The baseline clinical examination included anthropometry, resting blood pressure and ECG, fasting blood chemistry analysis, personal and family health history, and a maximal graded exercise test. Examination methods and procedures followed a standard manual of operations, as described previously. (16)

After checking the value of variables, sex and family history of hypertension were treated as fixed confounders. Since age, BMI, diabetes, hypercholesterolemia, smoking and heavy drinking changed their values during follow-up, they were treated as time-varying confounders.

If participants reported the parental hypertension during all study periods, the family history of hypertension of this participant was defined as 'Yes'. Otherwise, 'No'. Body mass index [BMI = weight (kg) / height (m) $^2$] was computed from measured height and weight. Overweight or obese was defined if the BMI > 25 Kg/m$^2$. Diabetes was defined as physician diagnosed diabetes, insulin use, or glucose>=126 mg/dL; and hypercholesterolemia was defined as by total cholesterol ≥240 mg/dl, or physician diagnosed hypercholesterolemia. Information on smoking habits (current smoker or not), heavy drink (alcohol drinks >14 per week or not) was obtained from a standardized questionnaire.

## 4.2 ANALYSIS USING MARGINAL STRUCTURAL COX MODEL

### 4.2.1 Data Preparation

Data were organized into long format as described in Chapter 3. Each subject had at least 2 rows of observations. Variables list were ID, number of visit, start of each visit, end of each visit, occurrence of hypertension, censor indicator, time dependent variables (CRF levels, overweight or obese, age, diabetes, hypercholesterolemia, smoking and heavy drinking status), time independent variables (sex and family history of hypertension), and history of previous CRF levels.

Means and standard deviation were used to describe the baseline continuous variables. Frequency and proportion was used to describe discrete variables. Baseline differences between three CRF groups were tested using ANOVA and Chi square test.

### 4.2.2 IP Weights And Marginal Structural Cox Model

Calculation of IP weights was as we described in Chapter 3. Specific, a cumulative logit model to the ordinal data was fitted to estimate the numerator of ${sw_m}^T$. The history of exposure is the previous CRF level (low CRF is the reference group). Covariates includes sex, family history of hypertension, BMI, age, smoking and heavy drinking status, diabetes and hypercholesterolemia). Here, only baseline covariates $\mathrm{L_0}$ were used.

Model 1: cumlogit ($\mathrm{CRF_j}$ ) = $\mathrm{CRF\_middle_{j-1}} + \mathrm{CRF\_high_{j-1}} + \mathrm{L_0}$

Model 2 was used to estimate the denominator of ${sw_m}^T$. The time-varying covariates $\mathrm{L_j}$ were included. cumlogit ($\mathrm{CRF_j}$ ) = $\mathrm{CRF\_middle_{j-1}} + \mathrm{CRF\_high_{j-1}} + \mathrm{L_j}$

Following the methods described in Chapter 3, prediction of fitness level was estimated from model 1 and 2; the stabilized time-varying weights, ${sw_m}^T$ was calculated.

The marginal Cox model was estimated using Coxph() function in the R package survival. The time-dependent Cox models using baseline covariates and time-varying covariates were also fitted. The estimated parameters from the marginal structural

Cox model can be explained as the causal effects of cardiorespiratory fitness on hypertension incidence. Robust variance estimates were obtained.

### 4.2.3 Results

Baseline characteristics are shown in Table 4.1. Among the total 14,290 participants, there were 1,280 subjects in low CRF group, 5,079 in middle group and 7,931 in high CRF group. Subjects who had higher CRF were elder, had less body weight, BMI, lower blood pressure and total cholesterol. The maximal METs and treadmill time duration were higher with increasing CRF level. People in high CRF group had large proportions of women, not current smokers, not diagnosed with diabetes and hypercholesterolemia. The proportions of heavy drinking and family history of hypertension were the highest in high CRF group.

Table 4.1 Descriptive statistics of baseline variables in ACLS study 1974-2003

| Variables | Low CRF (n=1280) | Middle CRF (n=5079) | High CRF (n=7931) | p value |
|---|---|---|---|---|
| Age (year) | 41.5±41.1 | 42.5±42.3 | 43.7±43.6 | <.0001 |
| Weight (Kg) | 86.8±86 | 81±80.7 | 75.9±75.6 | <.0001 |
| Body mass index (Kg/m$^2$) | 27.8±27.6 | 25.8±25.7 | 24.2±24.2 | <.0001 |
| Maximal METs | 8.7±8.6 | 10.6±10.6 | 13.3±13.2 | <.0001 |
| Treadmill time duration (min) | 11.6±11.4 | 15.7±15.7 | 21.3±21.2 | <.0001 |
| Systolic blood pressure (mm Hg) | 116.3±115.9 | 115.5±115.2 | 115.4±115.2 | 0.0105 |
| Diastolic blood pressure (mm Hg) | 77.8±77.5 | 77±76.8 | 76.3±76.2 | <.0001 |
| Total cholesterol (mg/dL) | 212.7±210.9 | 207.2±206.3 | 200.4±199.4 | <.0001 |
| Fasting blood glucose (mg/dL) | 100.7±99.9 | 98.1±97.7 | 98.9±96.2 | 0.7303 |
| Female (%) | 13.52 | 15.57 | 19.58 | <.0001 |
| Current smoker (%) | 29.38 | 18.67 | 9.39 | <.0001 |
| Heavy drink (>14 per week, %) | 4.92 | 5.55 | 7.7 | <.0001 |
| Diabetes (%) | 7.27 | 4.11 | 2.71 | <.0001 |
| Hypercholesterolemia (%) | 28.67 | 24.59 | 20.05 | <.0001 |
| Family history of hypertension (%) | 17.81 | 23.13 | 28.08 | <.0001 |

All participants had at least 2 visits. More than a half had three to five visits. And near 4% subjects were followed up for over 10 visits. 3869 (27.1%) subjects had hypertension occurred during follow-up and 77.9% of subjects were censored. (Please see Table 4.2)

Table 4.2 Characteristics of ACLS follow-up

| Characteristics | Frequency and proportion |
|---|---|
| Time of visits | |
| 2 | 14290 (100%) |
| 3-5 | 7404 (51.8%) |
| 6-10 | 2158 (15.1%) |
| ≥11 | 532 (3.7%) |
| Hypertension occurrence counts | 3869 (27.1%) |
| Censoring | 11132 (77.9%) |

The marginal structural Cox model with un-stabilized weights did not converge. The estimates using the stabilized weights can be explained as: taking people who were continuously in CRF low group as reference, the hazard of hypertension was of no significant different from those who were continuously in CRF middle group. There was on average a 23% decrease in hazard of hypertension among subjects who were always in CRF high group. The 95% CI was 5%-38%. (Table 4.3)

Table 4.3 Hazard ratios and 95% confidence intervals from the marginal structural Cox model using stabilized weight

| Variables | Hazard ratios and 95% CIs | Pr |
|---|---|---|
| CRF middle | 1.11 (0.83-1.48) | 0.4985 |
| CRF high | 0.77 (0.62-0.95) | 0.0158 |
| Overweight/obese | 1.55 (1.33-1.80) | <.0001 |
| Age | 1.02 (1.01-1.03) | <.0001 |
| Sex (Female) | 0.42 (0.20-0.92) | 0.0297 |
| Family history of hypertension | 1.20 (1.01-1.42) | 0.0431 |
| Smoking | 0.95 (0.78-1.16) | 0.6322 |
| Heavy drinking | 1.19 (1.01-1.39) | 0.0337 |
| diabetes | 0.75 (0.51-1.10) | 0.1378 |
| hypercholesterolemia | 1.04 (0.90-1.20) | 0.5811 |

## 4.3 COMPARING RESULTS OF DIFFERENT MODELS

### 4.3.1 The Time-Dependent Cox Model

One common approach is to fit a time dependent Cox model using baseline covariates. We can also fit another model using time-varying confounders. In this part, I will compare the estimates resulting from different models.

### 4.3.2 Results Comparisons

Results shown in Table 4.4 were from the time dependent Cox model using baseline covariates. After controlling for other covariables at baseline, the hazard ratio of hypertension was 0.79 (0.68-0.91) for middle CRF group and 0.62 (0.53-0.71) for high CRF group.

Table 4.4 Results of the time-dependent Cox PH model using baseline covariates

| Variables | Hazard ratios and 95% CIs | Pr |
|---|---|---|
| CRF middle | 0.79 (0.68-0.91) | 0.0015 |
| CRF high | 0.62 (0.53-0.71) | <.0001 |
| Overweight/obese | 1.33 (1.24-1.42) | <.0001 |
| Age | 1.02 (1.02-1.03) | <.0001 |
| Sex (Female) | 0.66 (0.60-0.73) | <.0001 |
| Family history of hypertension | 1.16 (1.08-1.24) | <.0001 |
| Smoking | 0.87 (0.80-0.95) | 0.0026 |
| Heavy drinking | 1.19 (1.05-1.34) | 0.0064 |
| diabetes | 0.94 (0.81-1.08) | 0.376 |
| hypercholesterolemia | 1.05 (0.98-1.13) | 0.1548 |

Results of the time-dependent Cox model using time-varying covariates shown that comparing to participants in low cardiorespiratory fitness level, those in middle and high CRF group had reduced risk of hypertension, the hazard ratios and 95% CIs being 0.79 (0.68-0.92) and 0.64 (0.56-0.74), respectively.

In addition, except the effect of hypercholesterolemia, the hazard ratios of other covariates of these two time-dependent models were also similar. Being overweight or obese, getting elder, with family history of hypertension and drinking heavily were significantly associated with the increased hazard of hypertension. While women had lower risk of developing hypertension than men, when controlling for other covariates in this study. The effect of high cholesterol on hypertension was detected in the time-dependent Cox model using time-varying covariates. However, the estimate effect of smoking was not as what we expected. The effect estimate of smoking on hypertension was not significant in the marginal structural Cox model.

Table 4.5 Results of time-dependent Cox PH model using time-varying covariates

| Variables | Hazard ratios and 95% CIs | Pr |
|---|---|---|
| CRF middle | 0.79 (0.68-0.92) | 0.0019 |
| CRF high | 0.64 (0.56-0.74) | <.0001 |
| Overweight/obese | 1.61 (1.50-1.72) | <.0001 |
| Age | 1.04 (1.03-1.04) | <.0001 |
| Sex (Female) | 0.67 (0.61-0.75) | <.0001 |
| Family history of hypertension | 1.13 (1.06-1.21) | 0.0002 |
| Smoking | 0.73 (0.66-0.82) | <.0001 |
| Heavy drinking | 1.45 (1.30-1.62) | <.0001 |
| diabetes | 0.99 (0.82-1.18) | 0.8885 |
| hypercholesterolemia | 1.25 (1.17-1.34) | <.0001 |

# CHAPTER 5. DISCUSSION AND CONCLUSION

Results of the two time-dependent Cox models were similar, which indicated that in this ACLS data, the changing of covariates didn't affect the estimations a lot. Although estimates from these two models were statistically significant, they didn't have the causal interpretations since there was treatment confounder feedback. When there is no treatment confounder feedback, estimates of effect can be obtained from the time-dependent Cox model.

To address exposure-confounder feedback, inverse probability weights were calculated and applied into the marginal structural Cox model. In this study, there was significant decrease of hazard of hypertension for people who were always in high CRF comparing to those who were always in low CRF group. These estimates assumed that only the cardiorespiratory fitness of the previous visit had direct impact on the current CRF level.

The validity of effect estimates depends on assumptions of no measurement errors and no model misspecification. (18) These two conditions are hard to realize in the observational studies. For example, family history of hypertension was defined as any reported parental hypertension during follow-up period. Parental hypertension can be diagnosed several years after subjects entered into this study.

Therefore, the family history of hypertension may be underestimated. When calculating the IP weights, there are possibility of having model misspecification. If the unmeasured confounders had significant effect on the levels of exposure or affect the censoring status, then the calculated IP weights can't remove all confounding. Under this circumstance, the estimates of casual effect would be biased.

In a clinical trial, participants would be re-visited after a certain amount of time, such as 3 months. The time interval between two visits would be regular. However, in the ALCS study, the follow-up was not based on the same intervals. For example, after a subject entering into study, the second visit was 2 years later, the third visit was 5 years after. Ignoring different visit intervals and only using information of number of visit, assumes that the effect of overweight status at visit 1 on visit 2 (2 years ago) is same as the effect of overweight at visit 2 on visit 3 (5 years ago), which is not biologically reasonable. Considering the durations between visits can improve the calculation of the inverse probability weight.

The Ipw R package has function ipwtm() to estimate time-varying inverse probability weights. The exposure can be binomial, multinomial, ordinal or continuous. Estimation of weights can be calculated by using all visits, or only visits until the exposure level first switches form one level to another. After this switch, weights are held constant. Currently, only for binary exposure, all visits can be used. For some clinical trials where after patients initiating the new treatment, they will keep taking it. In this case, it makes sense that weights calculated until the first switch and are constant for the rest visits. However, in the ACLS data, the level of cardiorespiratory

fitness can change during the whole follow-up. The ipwtm() function can't be used to calculate the time-varying weights. Further work should be done to expand this R package to support such data.

In conclusion, to get unbiased estimates of causal effects from the observational study, exchangeability, positivity, consistency, no measurement error and no model misspecification need to be hold. Marginal Cox model can be applied in longitudinal data to deal with the treatment-confounder feedback. Estimates and variance with stabilized weights perform better than the un-stabilized weights.

# REFERENCES


1. Robins JM, Hernán MÁ, Brumback B. Marginal Structural Models and Causal Inference in Epidemiology. Epidemiology. 2000;11(5):550-60.

2. Hernán MÁ, Robins JM. Causal inference. Part I: Causal inference without models. Chapter 3: Observational studies. Boca Raton: Chapman & Hall/CRC, forthcoming 2017. p.27-34. [Updated 2017 March 5; cited 2017 Aug 15] Available from: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

3. Robins J. 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. 1986;7(9):1393-512.

4. Hernán MÁ, Robins JM. Causal inference, Part III: Causal inference from complex longitudinal data. Chapter 19: Time-varying treatments. Boca Raton: Chapman & Hall/CRC, forthcoming 2017. [Updated 2017 March 5; cited 2017 Aug 15] Available from: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

5. Hernán MÁ, Robins JM. Causal inference. Part III: Causal inference from complex longitudinal data. Chapter 20: Treatment confounder feedback. Boca Raton: Chapman & Hall/CRC, forthcoming 2017. [Updated 2017 March 5; cited 2017 Aug 15] Available from: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

6. Pugh MG, Harvard School of Public Health. Inference in the Cox proportional hazards model with missing covariate data; 1993.

7. Robins JM. Association, Causation, And Marginal Structural Models. Synthese. 1999;121(1):151-79.8.

8. Hernán MÁ, Robins JM. Causal inference. Part II: Causal inference with models. Chapter 12: IP weighting and marginal structural models. Boca Raton: Chapman & Hall/CRC, forthcoming 2017. p12-6. [Updated 2017 March 5; cited 2017 Aug 15] Available from: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

9. Kish L. Weighting for unequal pi, Journal of official statistics, 1992;8(2):183-200

10. Cynthia SC, Louis AS, Abigail BG, Elizabeth JA, Terry MT. The Basics of Propensity Scoring and Marginal Structural Models. Technical Report #84. 2013 August 1; Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota.

11. Hernán MÁ, Brumback B, Robins JM. Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men. Epidemiology. 2000;11(5):561-70.

12. Young JG, Hernán MA, Picciotto S, Robins JM. Simulation from structural survival models under complex time-varying data structures. JSM Proceedings, Section on Statistics in Epidemiology, Denver, CO: American Statistical Association. 2008.

13. Mayo clinic staff. Diseases and Conditions: High blood pressure (hypertension).[cited 2017 Aug 15] Available from: http://www.mayoclinic.org/diseases-conditions/high-blood-pressure/basics/risk-factors/con-20019580.

14. Barlow CE, LaMonte MJ, FitzGerald SJ, Kampert JB, Perrin JL, Blair SN. Cardiorespiratory Fitness Is an Independent Predictor of Hypertension Incidence among Initially Normotensive Healthy Women. American Journal of Epidemiology. 2006;163(2):142-50.

15. Rankinen T, Church TS, Rice T, Bouchard C, Blair SN. Cardiorespiratory fitness, BMI, and risk of hypertension: the HYPGENE study. Med Sci Sports Exerc. 2007;39(10):1687-92.

16. Blair SN, Kohl HW, Iii, Paffenbarger RS, Jr, Clark DG, et al. PHysical fitness and all-cause mortality: A prospective study of healthy men and women. JAMA. 1989;262(17):2395-401.

17. Balke B, Ware RW. An experimental study of PHysical fitness of Air Force personnel. United States Armed Forces medical journal. 1959;10(6):675-88.

18. Hernán MÁ, Robins JM. Causal inference, Part II: Causal inference with models. Chapter 13: Standardization and the parametric g-formula. Boca Raton: Chapman & Hall/CRC, forthcoming 2017. p.30. [Updated 2017 March 5; cited 2017 Aug 15] Available from: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

# APPENDIX A SOURCE CODES

SAS code, calculating IP weights:

```
%macro weight();
/* Model 1 */
proc logistic data=sim desc noprint;
      model A = am1;
      output out=tr_top p=ptr_num;
run;
/* Model 2 */
proc logistic data=sim desc noprint;
      model A = am1 l lm1;
      output out=tr_bot p=ptr_den;
run;
/* Model 3 */
proc logistic data=sim noprint;
      model censor_r = am1;
 output out=cen_top p=pcen_num;
run;
/* Model 4 */
proc logistic data=sim noprint;
      model censor_r = am1 lm1 lm2;
 output out=cen_bot p=pcen_den;
run;
proc sort data=tr_top; by id tpoint; run;
proc sort data=tr_bot; by id tpoint; run;
proc sort data=cen_top; by id tpoint; run;
proc sort data=cen_bot; by id tpoint; run;
data main_w;
      merge tr_top tr_bot cen_top cen_bot;
      by ID tpoint;
if a=1 then ptr_num=ptr_num;
if a=0 then ptr_num=1-ptr_num;
if a=1 then ptr_den=ptr_den;
if a=0 then ptr_den=1-ptr_den;
if first.id then do;
tr_num=1;
```

```
tr_den=1;
cen_num=1;
cen_den=1;
end;
retain tr_num tr_den cen_num cen_den;
tr_num=tr_num*ptr_num;
tr_den=tr_den*ptr_den;
cen_num=cen_num*pcen_num;
cen_den=cen_den*pcen_den;

wc=1/(tr_den*cen_den);
swc=(tr_num*cen_num)/(tr_den*cen_den);
w=1/(tr_den);
sw=(tr_num)/(tr_den);
run;
%mend;
/*%weight();*/
%macro data(n=, subjects=,   psi1= , lam=, out=);
%let c=1;
%do i=1 %to &n;
%simulate(subjects=&subjects,  psi1=&psi1 , lam=&lam);
%weight();
proc export data=main_w
  outfile=%unquote(%str(%')C:\sim\&out\a&c%str(.)csv%str(%'))
  dbms=csv
  replace;
run;
%let c=%eval(&c+1);
%end;
%mend;
%data(n=500, psi1=0 , lam=0.01, subjects=500,  out=p00_lmd001_n500);
%data(n=500, psi1=0 , lam=0.1, subjects=500,  out=p00_lmd01_n500);
%data(n=500, psi1=0 , lam=0.01, subjects=2500,  out=p00_lmd001_n2500);
%data(n=500, psi1=0 , lam=0.1, subjects=2500,  out=p00_lmd01_n2500);
%data(n=500, psi1=-0.3 , lam=0.01, subjects=500,  out=p-03_lmd001_n500);
%data(n=500, psi1=-0.3 , lam=0.1, subjects=500,  out=p-03_lmd01_n500);
%data(n=500, psi1=-0.3 , lam=0.01, subjects=2500,  out=p-03_lmd001_n2500);
%data(n=500, psi1=-0.3 , lam=0.1, subjects=2500,  out=p-03_lmd01_n2500);
```

**R code, fitting marginal Cox models:**

```
library('survival')

simu<-500 #1fixed number

truev<- 0 #2true treatment effect, fai
```

```
est<-matrix(0, simu,9)
est1<-matrix(0, simu,6)
for(i in 1:simu){
x <- read.csv(paste("C:/sim/p00_lmd001_n500/a", i, ".csv", sep=""))#3location
a <- subset(x, tpoint==1, select = c(id, L))
a$L0 <-a$L
a <- subset(a, select=c(id, L0))
x <-merge(x, a, by = "id")
m1<-CoxPH(Surv(tpoint2,tpoint,Y)~A+cluster(id), data=x, weights=w)
m2<-CoxPH(Surv(tpoint2,tpoint,Y)~A+cluster(id), data=x, weights=wc)
m3<-CoxPH(Surv(tpoint2,tpoint,Y)~A+cluster(id), data=x, weights=sw)
m4<-CoxPH(Surv(tpoint2,tpoint,Y)~A+cluster(id), data=x, weights=swc)
est[i,c(1,2)] <- c(m1$coef,m1$var)
est[i,c(3,4)] <- c(m2$coef,m2$var)
est[i,c(5,6)] <- c(m3$coef,m3$var)
est[i,c(7,8)] <- c(m4$coef,m4$var)
est[i,9] <- sum(x$Y)/500 #4number of id, 500 or 2500
m5<-CoxPH(Surv(tpoint2,tpoint,Y)~A+L0+cluster(id), data=x)
m6<-CoxPH(Surv(tpoint2,tpoint,Y)~A+L+cluster(id), data=x)
m7<-CoxPH(Surv(tpoint2,tpoint,Y)~A+L+Lm1+cluster(id), data=x)
est1[i,c(1,2)] <- c(m5$coef[1],m5$var[1,1])
est1[i,c(3,4)] <- c(m6$coef[1],m6$var[1,1])
est1[i,c(5,6)] <- c(m7$coef[1],m7$var[1,1])
}
betabar<-c (mean(est[,1]),mean(est[,3]),mean(est[,5]),mean(est[,7]))
emp_sd<-c(sqrt(var(est[,1])),sqrt(var(est[,3])),sqrt(var(est[,5])),sqrt(var(est[,7])))
sd <-c (mean(est[,2]),mean(est[,4]),mean(est[,6]),mean(est[,8]))
bias<-betabar-truev
```

```
per_bias<-100*bias/truev

#std_bias<-100*bias/betastd

MSE<-bias^2+emp_sd^2

coverage<-c(

sum(as.numeric((truev<est[,1]+qnorm(0.975)*sqrt(est[,2]) & truev>est[,1]-
qnorm(0.975)*sqrt(est[,2])))/simu ,

sum(as.numeric((truev<est[,3]+qnorm(0.975)*sqrt(est[,4]) & truev>est[,3]-
qnorm(0.975)*sqrt(est[,4])))/simu ,

sum(as.numeric((truev<est[,5]+qnorm(0.975)*sqrt(est[,6]) & truev>est[,5]-
qnorm(0.975)*sqrt(est[,6])))/simu ,

sum(as.numeric((truev<est[,7]+qnorm(0.975)*sqrt(est[,8]) & truev>est[,7]-
qnorm(0.975)*sqrt(est[,8])))/simu

)

censor_rate<- 1-mean(est[,9])
```