'Self-Controlled' Structural Nested Models

Zach Shahn[1] and James Robins[2]
[1]IBM Research, Yorktown Heights
[2]Harvard School of Public Health, Boston

1 Introduction

'Self-controlled' methods (such as the case-crossover (CC) (Maclure, 1990), case time-control (CTC) (Suissa, 1995), and self-controlled case series (SCCS) (Farrington, 1995)) are popular in epidemiology and other fields. The reason for their popularity is that they can adjust for all baseline confounders, even unobserved baseline confounders such as subjects' full genetic codes or deep-seeded fears, by exploiting longitudinal data to in some sense use subjects as their own controls.

To achieve this feat, however, self-controlled methods rely on strong assumptions that lead to various drawbacks. In their original incarnations, the CC, CTC, and SCCS all: (1) estimate instantaneous hazard ratios, i.e. the effect of exposure at time t on the probability of the outcome occurring at time t; (2) assume that effects are constant across subjects and time; (3) are unable to adjust for observed postbaseline confounders that are influenced by past exposures; and (4) only handle binary exposures and outcomes. Each method also makes additional idiosyncratic non-standard assumptions: the CC assumes no time trends in exposure; the CTC assumes similar time trends in exposure among cases and controls; and the SCCS assumes no outcome dependent exposures. Further, the CTC posits a conditional logistic regression model for exposure with subject specific effects, while the SCCS posits a Poisson process for the outcome, also with subject specific effects.

Various modifications of these seminal methods have partially relaxed some of the above restrictions. For example, Schuemie et al. (2015) present a SCCS model that breaks down the effect of exposure over time into independent acute and cumulative components. However, to our knowledge, the Self-Controlled Structural Nested Mean Model (sc-SNMM) we present here is the first self-controlled method that: (1) allows for general time-varying effects of exposure; (2) can model heterogeneity as a function of observed baseline or time varying covariates; (3) can adjust for observed time varying confounders that are influenced by past exposure; and (4) can handle continuous or binary exposures and continuous, binary, or survival outcomes.

These advantages are not at the cost of stronger assumptions relative to existing methods. Beyond the standard causal assumptions of consistency and sequential exchangeability (conditional on all baseline confounders and observed time varying confounders), the sc-SNMM requires: (1) correct specification of a fixed effects model either for exposure or a transformation of the outcome; and (2) that unobserved baseline confounders are not also effect modifiers. Assumption (1) is no stronger than the similar assumptions required by the CTC and SCCS. And while (2) is a major effect homogeneity assumption, it is weaker than the similar implicit homogeneity assumptions of existing methods. Thus, the sc-SNMM relaxes several important restrictions on existing self-controlled methods at no additional cost.

The organization of the paper is as follows. In Section 2, we lay out notation and review the counterfactual causal framework for time varying treatments in which sc-SNMMs are defined. In Section 3, we define and explain sc-SNMMs. In Section 4, we derive estimators for the parameter of an sc-SNMM under a fixed effects model for either exposure or (a transformation of) the outcome. In Section 5, we present self controlled structural nested cumulative failure time models (sc-SNCFTMs), which are modifications of sc-SNMMs for survival outcomes. In Section 6, we present a data analysis estimating the effect of thiazolidinedinodes (TZDs), a second line diabetes therapy, on bone fractures from a large medical claims database.

2 Notation and Counterfactual Framework

Suppose we observe a cohort of N subjects indexed by $i \in \{1, \ldots, N\}$. For now, assume that each patient is observed at regular intervals from baseline time 0 through end of followup time K, and there is no loss to follow-up. Let A_m denote treatment received at time m, Y_m the outcome of interest at time m, and L_m a vector of time varying covariates (including past values of the outcome) at time m. For arbitrary time varying variable X, we denote by $\bar{X}_m$ the history of X through time m, i.e. $(X_0, \ldots, X_m)$, and we denote by $\underline{X}_m$ the future of X from time m through time K, i.e. $(X_m, ..., X_K)$. Finally, we let α denote an unmeasured, possibly continuous and multidimensional, baseline variable.

We adopt the counterfactual framework for time-varying treatments (Robins, 1986) which posits that corresponding to each treatment regime $\bar{a}_m$, each subject has a counterfactual or potential outcome $Y_{m;\bar{a}_m}$ that would have been observed had that subject received treatment regime $\bar{a}_m$.

Throughout, we make the causal assumptions:

$$\textbf{Consistency: } Y_{m;\bar{A}_m} = Y_m \ \forall m, \bar{A}_m \tag{1}$$

$$\alpha - \textbf{Sequential Exchangeability:} A_m \perp\!\!\!\perp Y_{m;\bar{a}_K} | \bar{L}_m, \bar{A}_{m-1}, \alpha \ \forall m, \bar{a}_K \tag{2}$$

Consistency is a standard assumption stating that observed outcomes are equal to counterfactual outcomes corresponding to observed treatments. This assumption is necessary to link observed data to the parameters of a causal model for counterfactual outcomes.

α-Sequential exchangeability essentially states that there are no unobserved confounders except for α. This is a weaker version of the standard sequential exchangeability assumption (Robins, 1986) that essentially states that there is no unobserved confounding at all.

3 The Causal Model

The self-controlled Structural Nested Mean Model (sc-SNMM) is defined as

$$\gamma_m(\bar{A}_m, \bar{L}_m; \Psi) = E[\underline{Y}_{m+1;\bar{A}_m,0} | \bar{A}_m, \bar{L}_m, \alpha] - E[\underline{Y}_{m+1;\bar{A}_{m-1},0} | \bar{A}_m, \bar{L}_m, \alpha] \tag{3}$$

where γ_m is a known $K - m$ dimensional function equal to 0 whenever $\Psi = 0$. An sc-SNMM models the effect of receiving treatment a_m at time m followed by no further treatment compared to no treatment at time m or afterward conditional on treatment and covariate history through time m and unobserved baseline covariate α. $\gamma_m(\bar{A}_m, \bar{L}_m; \Psi)$ is sometimes referred to as a 'blip

function' because it models the effect of a final blip of treatment at time m.

(3) is very similar to a standard SNMM (Robins, 1994), the difference being the presence of unobserved baseline covariate α in the conditioning events on the right hand side. That $\gamma_m(\bar{A}_m, \bar{L}_m; \Psi)$ does not include α as an argument implies the additional assumption

No Effect Modification by Unobserved Baseline Confounders (4)

That is, α is not an effect modifier. This is a major assumption, but still weaker than the effect homogeneity assumptions implicit in other self-controlled methods.

From a policy perspective, analysts are not necessarily interested in the effect of final blips of treatment at various times. The more common reason to fit a (sc)- SNMM is that given the parameter of interest Ψ, one can estimate the expected counterfactual outcome trajectory in the population in the total absence of treatment, i.e. $E[\bar{Y}_{K;\bar{0}}]$.

4 An Estimator of Ψ

Define $H_m(\Psi)$ to be the $K - m$ dimensional vector with components

$$Y_t - \sum_{j=m}^{t} \gamma_{j,m}(\bar{A}_j, \bar{L}_j; \Psi) \tag{5}$$

for t in $m+1, \ldots, K$. If Ψ^* is the true value of Ψ, $H_m(\Psi^*)$ has the properties that

$$E[H_m(\Psi^*)|\bar{A}_m, \bar{L}_m, \alpha] = E[Y_{m+1;\bar{A}_{m-1},0}|\bar{A}_m, \bar{L}_m, \alpha] \tag{6}$$

and

$$E[H_m(\Psi^*)|\bar{A}_m, \bar{L}_m, \alpha] = E[H_m(\Psi^*)|\bar{A}_{m-1}, \bar{L}_m, \alpha] \tag{7}$$

Suppose that we somehow had access to $E[A_m|\bar{A}_{m-1}, \bar{L}_m, \alpha]$. Then we could obtain an unbiased estimate of Ψ by solving the estimating equations

$$\sum_i \sum_m H_{i,m}(\Psi)(d_m(A_{i,m}) - E[d_m(A_m)|\bar{A}_{i,m-1}, \bar{L}_{i,m}, \alpha_i]) = 0, \tag{8}$$

where d_m is some function with dimension $dim(\Psi)$. Unfortunately, it is not possible to non-parametrically obtain a consistent estimate of $E[A_m|\bar{A}_{m-1}, \bar{L}_m, \alpha]$ as we do not observe α. However, suppose we specify the following parametric model:

$$A_m = \mu_\alpha + f_m(\bar{A}_{m-1}, \bar{L}_m; \theta_m) + \epsilon_m \tag{9}$$

where $E[\epsilon_m|\bar{A}_{m-1}, \bar{L}_m, \mu_\alpha] = 0$ and μ_α is a subject level fixed effect. We can estimate θ by solving the estimating equation

$$\sum_i \sum_m g_m(\bar{A}_{i,m-2}, \bar{L}_{i,m-1})\{(A_{i,m} - f_m(\bar{A}_{i,m-1}, \bar{L}_{i,m}; \theta_m)) - (A_{i,m-1} - f_{m-1}(\bar{A}_{i,m-2}, \bar{L}_{i,m-1}; \theta_{m-1}))\} = 0 \tag{10}$$

In (10), the difference of differences in the braces cancels out μ_α, which remains constant across time for each subject under model (9), allowing us to estimate θ. Now define

$$E_{\hat{\theta}}[A_m|\bar{A}_{m-1}, \bar{L}_m, \alpha] \equiv \frac{1}{K} \sum_{j \neq m} (A_j - f_j(\bar{A}_{j-1}, \bar{L}_j; \hat{\theta}_j)) + f_m(\bar{A}_{m-1}, \bar{L}_m; \hat{\theta}_m) \tag{11}$$

Since, given θ, $E[A_j - f_j(\bar{A}_{j-1}, \bar{L}_{j-1}; \theta_j)] = \mu_\alpha$ for each j, the first term on the RHS of (11) is an unbiased estimate of μ_α, making (11) an unbiased estimate of $E[A_m|\bar{A}_{m-1}, \bar{L}_m, \alpha]$ in the limit as $N \to \infty$. (However, note that (11) is not a consistent estimator since K does not go to ∞). If $d(A_m) \equiv d(\bar{A}_m, \bar{L}_m)$ is linear in A_m, then $E_{\hat{\theta}}[d(A_m)|\bar{A}_{m-1}, \bar{L}_m, \alpha] \equiv d(E_{\hat{\theta}}[A_m|\bar{A}_{m-1}, \bar{L}_m, \alpha])$ is an unbiased estimate of $E[d(A_m)|\bar{A}_{m-1}, \bar{L}_m, \alpha]$.

Consider the estimating equations

$$\sum_i \sum_m H_{i,0m}(\Psi)(d_m(A_{i,m}) - E_{\hat{\theta}}[d_m(A_m)|\bar{A}_{i,m-1}, \bar{L}_{i,m}, \alpha_i]) = 0, \tag{12}$$

where $H_{st}(\Psi)$ denotes the $(t-s)^{th}$ component of $H_s(\Psi)$ defined in (5), i.e. $Y_t - \sum_{k=s}^{t} \gamma_{kt}(\bar{A}_k, \bar{L}_k; \Psi)$. (12) is similar to the estimating equations (8) that we would use to estimate Ψ if α were observed and conditional expected exposures $E[d_m(A_m)|\bar{A}_{i,m-1}, \bar{L}_{i,m}, \alpha_i]$ were known. We have just replaced $H_m(\Psi)$ by $H_{0m}(\Psi)$ and $E[d_m(A_m)|\bar{A}_{i,m-1}, \bar{L}_{i,m}, \alpha_i]$ by $E_{\hat{\theta}}[d_m(A_m)|\bar{A}_{i,m-1}, \bar{L}_{i,m}, \alpha_i]$ as defined in (11). Below, we will show that the estimator solving (12) is asymptotically unbiased for Ψ.

We will need the following result:

Lemma 1. $E[H_{0t}|\bar{A}_m, \bar{L}_m, \alpha] = E[H_{0t}|\bar{A}_{m-1}, \bar{L}_m, \alpha]$ $\forall m, t$.

Proof. See appendix. □

Now, note that as $\hat{\theta} \to \theta$, $A_m - E_{\hat{\theta}}[A_m|\bar{A}_{m-1}, \bar{L}_m, \alpha] \to (\epsilon_m - \frac{1}{K}\sum_{j \neq m} \epsilon_j)$ where ϵ is as defined in (9). We then have that

Proposition 1. $E[H_{0m}(\Psi)(A_m - E_{\hat{\theta}}[A_m|\bar{A}_{m-1}, \bar{L}_m, \alpha])] \to 0$ *as* $\hat{\theta} \to \theta$ $\forall m$.

Proof.
$E[H_{0m}(\Psi)(A_m - E_{\hat{\theta}}[A_m|\bar{A}_{m-1}, \bar{L}_m, \alpha])]$
$\to E[H_{0m}(\Psi)(\epsilon_m - \frac{1}{K}\sum_{j \neq m} \epsilon_j)]$ as $\hat{\theta} \to \theta$
$= E[E[H_{0m}(\Psi)\epsilon_m|\bar{A}_{m-1}, \bar{L}_m, \alpha]] - \frac{1}{K}\sum_{j \neq m} E[E[H_{0m}(\Psi)\epsilon_j|\bar{A}_{j-1}, \bar{L}_j, \alpha]]$
$= E[E[H_{0m}(\Psi)|\bar{A}_{m-1}, \bar{L}_m, \alpha]E[\epsilon_m|\bar{A}_{m-1}, \bar{L}_m, \alpha]] - \frac{1}{K}\sum_{j \neq m} E[E[H_{0m}(\Psi)|\bar{A}_{j-1}, \bar{L}_j, \alpha]E[[\epsilon_j|\bar{A}_{j-1}, \bar{L}_j, \alpha]]$
(by Lemma 1)
$= 0$ □

For $d(A_m) \equiv d(\bar{A}_m, \bar{L}_m)$ linear in A_m, the argument in the proof of Proposition 2 also shows that $E[H_{0m}(\Psi)(d(A_m) - E_{\hat{\theta}}[d(A_m)|\bar{A}_{m-1}, \bar{L}_m, \alpha])] = 0$, thus demonstrating that (12) are asymptotically unbiased estimating equations for Ψ.

To recap, we can consistently estimate Ψ by estimating equations (12) under: model (3); assumptions (1), (2), and (4); and exposure fixed effects model (9).

4.1 Estimation Based on an Alternative Fixed Effects Model

Instead of the fixed effects exposure model (9), suppose we specified the following parametric model for $H_{0m}(\Psi)$:

$$H_{0m}(\Psi) = \lambda_\alpha + h_m(\bar{A}_{m-1}, \bar{L}_m; \phi_m) + \epsilon_m^H \tag{13}$$

where $E[\epsilon_m^H|\bar{A}_{m-1}, \bar{L}_m, \lambda_\alpha] = 0$ and λ_α is a subject level fixed effect.

Then consider the estimating equations

$$\sum_i \sum_m \{(H_{i,0m}(\Psi) - h_m(\bar{A}_{i,m-1}, \bar{L}_{i,m}; \hat{\phi}_m)) - (H_{i,0m-1}(\Psi) - h_{m-1}(\bar{A}_{i,m-2}, \bar{L}_{i,m-1}; \hat{\phi}_{m-1}))\} d(\bar{A}_{i,m-1}, \bar{L}_{i,m-1}) = 0 \tag{14}$$

for arbitrary function $d(\bar{A}_{i,m-1}, \bar{L}_{i,m-1})$.

Proposition 2. *The estimating equations (14) are asymptotically unbiased for Ψ if the fixed effects model (13) is correctly specified.*

Proof.
$E[(H_{0m}(\Psi) - h_m(\bar{A}_{m-1}, \bar{L}_m; \phi_m)) - (H_{0m-1}(\Psi) - h_{m-1}(\bar{A}_{m-2}, \bar{L}_{m-1}; \phi_{m-1}))d(\bar{A}_{m-1}, \bar{L}_{m-1})] =$
$E\{$
$E[H_{0m}(\Psi) - h_m(\bar{A}_{m-1}, \bar{L}_m; \phi_m)|\bar{A}_{m-1}, \bar{L}_{m-1}, \alpha]d(\bar{A}_{m-1}, \bar{L}_{m-1}) -$
$E[H_{0m-1}(\Psi) - h_{m-1}(\bar{A}_{m-2}, \bar{L}_{m-1}; \phi_{m-1})|\bar{A}_{m-2}, \bar{L}_{m-1}, \alpha]E[d(\bar{A}_{m-1}, \bar{L}_{m-1})|\bar{A}_{m-2}, \bar{L}_{m-1}, \alpha]$
$\}$
$= 0$ □

Thus, we can consistently estimate Ψ under: model (3); assumptions (1), (2), and (4); and transformed outcome fixed effects model (13).

5 Survival Outcomes

We now consider survival settings, where existing self-controlled methods are most frequently applied. Survival outcomes are events that can only occur once, such as death, a first myocardial infarction, or a first anything. Because survival outcomes are so often harmful, they are frequently referred to as 'failures'. Let $Y_k = 1$ denote failure before k, so $Y_k = 1$ implies that $Y_j = 1$ for all $k \leq j \leq K$. Whenever failure has already occurred and $Y_j = 1$, we define, by convention, $L_j = A_j = 0$. To estimate effects of time-varying treatments in this scenario in the absence of unobserved baseline confounding, Piccioto et al. (2012) proposed a Structural Nested Cumulative Failure Time Model (SNCFTM). Here, we propose a self-controlled SNCFTM (or sc-SNCFTM):

$$e^{\gamma_{m,k}(\bar{A}_m, \bar{L}_m; \Psi)} = \begin{cases} \frac{E[Y_{k;\bar{A}_m,0}|\bar{A}_m, \bar{L}_m, Y_m=0, \alpha]}{E[Y_{k;\bar{A}_{m-1},0}|\bar{A}_m, \bar{L}_m, Y_m=0, \alpha]}, \text{ if } Y_m = 0 \\ 1, \text{ if } Y_m = 1 \end{cases} \tag{15}$$

where $\gamma_{m,k}(\bar{A}_m, \bar{L}_m; \Psi) = 0$ when $\Psi = 0$ or $A_m = 0$. (15) is a model for the conditional multiplicative effect of one last blip of treatment at time m on the cumulative risk of failure by time k. The absence of α as an argument of $\gamma_{m,k}(\bar{A}_m, \bar{L}_m; \Psi)$ again implies assumption (4) that unobserved confounders are not effect modifiers.

We define

$$H_{mk}(\Psi) = \begin{cases} Y_k exp\{-\sum_{j=m}^{k-1} \gamma_{j,k}(\bar{A}_j, \bar{L}_j; \Psi)\}, \text{ if } Y_m = 0 \\ 1, \text{ if } Y_m = 1 \end{cases} \tag{16}$$

It can be shown that for Ψ^* the true value of Ψ,

$$E[Y_{k;\bar{A}_{m-1},0}|\bar{L}_m, \bar{A}_m, \alpha] = E[H_{mk}(\Psi^*)|\bar{L}_m, \bar{A}_m, \alpha] \tag{17}$$

$$= E[H_{mk}(\Psi^*)|\bar{L}_m, \bar{A}_{m-1}, \alpha] \tag{18}$$

By arguments similar to those in Section 4, it follows that under fixed effects model for exposure (9), the estimating equations

$$\sum_i \sum_m \{(d_m(A_{i,m}) - E_{\hat{\theta}}[d_m(A_m)|\bar{A}_{i,m-1}, \bar{L}_{i,m}, \alpha_i]) \times \sum_{k=m+1}^{K} H_{i,0k}(\Psi)\} = 0 \tag{19}$$

are asymptotically unbiased for Ψ under model (15).

6 Data Analysis

Thiazolidinediones (TZDs) are second-line therapies for type 2 Diabetes known to cause bone fractures. We fit a sc-SNMM to estimate the effect of TZDs on bone fractures among a cohort of diabetes patients initially prescribed metformin (the widely recommended first line therapy). We obtained our data from a large medical claims database.

First prescription of metformin after diagnosis with diabetes was taken to be baseline. A_m indicated use of a TZD in month m after baseline, Y_m indicated the occurrence of a fracture in month m after baseline, and L_m denoted time varying confounders at month m after baseline. The time-varying confounders comprised indicators of the occurrence of 13 common diabetes complications in the past 3 months and indicators of the occurrence of the same complications at any point after baseline. We chose these confounders because switching from metformin to a TZD may be prompted by development of diabetes complications, which in turn may be associated with fractures. There are surely other time-varying confounders we neglected to include.

For illustrative purposes, we specified a simple linear blip model in which past diabetic neuropathy (DN) was the only effect modifier:

$$\gamma_{mt}(\bar{A}_m, \bar{L}_m; \Psi) = a_m\Psi_1 + a_m DN_m\Psi_2 + a_m(t-m)\Psi_3 \tag{20}$$

This is a model of the conditional effect on the likelihood of a fracture at month t of one last month of taking a TZD at month m compared to never taking a TZD at month m or later. Ψ_2 determines how this effect is modified by past DN, and Ψ_3 determines how this effect decays or grows over time since the last blip of TZD.

We also specified a linear fixed effects model for probability of a patient taking a TZD at month m as a function of past and current diabetes complications and whether the patient was on a TZD the previous month:

$$A_m = \mu_\alpha + (L_{1m}^{recent}, \ldots, L_{13m}^{recent}, L_{1m}^{ever}, \ldots, L_{13m}^{ever}, A_{m-1})\theta \tag{21}$$

We then estimated θ by solving estimating equations (10), plugged our estimate $\hat{\theta}$ into the estimating equations (12) for Ψ, and computed bootstrap standard errors for Ψ. The results were: $\hat{\Psi}_1 = .002$ ($se = .0009$); $\hat{\Psi}_2 = .001$ ($se = .0006$); $\hat{\Psi}_3 = -.0002$ ($se = .00008$). Assuming correct specification of the blip model (20) and the fixed effect exposure model (21), the interpretation of our estimate of Ψ is that a final blip of TZD treatment increases the probability of a fracture at future months, moreso in patients with DN, and the magnitude of the effect decays gradually over time since the final treatment.

The fracture rate in the observed data was .14 per 5 person-years. Using $\hat{\Psi}$, we estimated

that this rate would have declined by .002 fractures (se = .0004) per 5 person-years in the cohort had TZDs never been prescribed. The small relative size of this effect is partly due to the fact that only about 1 in 1000 patients in the cohort were ever prescribed TZDs.

7 Discussion and Future Work

We have introduced sc-SNMMs and sc-SNCFTMs for estimating causal effects from longitudinal data in the presence of baseline confounding. Other methods that exploit longitudinal data to eliminate baseline confounding have been dubbed 'self-controlled' (Maclure et al., 2012). The self-controlled methods we present here are the first that can handle general time-varying effects (e.g. arbitrary cumulative or lagged effects of exposure) and adjust for time-varying confounders that are influenced by past treatment.

Of course, causal inference in the presence of confounding does not come free, and all self-controlled methods must make assumptions. Our estimators of the parameters of sc-SNMMs and sc-SNCFTMs are consistent provided: (a) correct specification of a fixed effects model for exposure or the (transformed) outcome; and (b) none of the unobserved baseline confounders are also effect modifiers. The fixed effect models for exposure and outcome are similar to models underlying case-time-control and self-controlled case series estimators, respectively. Our effect homogeneity assumption is weaker than similar implicit assumptions underlying alternative self-controlled methods. We note that the more baseline variables are collected, the weaker our homogeneity assumption becomes as only unobserved confounders are prohibited from being effect modifiers.

One shortcoming of our method is that fixed effect models (9) and (13) are bound to be technically misspecified for binary exposures and outcomes as they are not strictly bounded by 0 and 1. However, the models may still approximately fit and the method can reasonably be applied in such cases, as illustrated in our data analysis. Suitable fixed effect models for binary data would be of great interest beyond sc-SNMMs and sc-SNCFTMs. That our method is well suited to continuous exposures and outcomes is also an advantage over existing self-controlled methods.

Another extension we are pursuing is self-controlled optimal regime Structural Nested Mean Models (sc-opt-SNMMs) for learning optimal treatment strategies in the presence of baseline confounding. We also hope to develop efficient and double robust estimators.

8 References

Farrington, C. P. "Relative incidence estimation from case series for vaccine safety evaluation." Biometrics (1995): 228-235.

Maclure, Malcolm. "The case-crossover design: a method for studying transient effects on the risk of acute events." American journal of epidemiology 133.2 (1991): 144-153.

Maclure, Malcolm, et al. "When should case?only designs be used for safety monitoring of medical products?." Pharmacoepidemiology and drug safety 21.S1 (2012): 50-61.

Picciotto, Sally, et al. "Structural nested cumulative failure time models to estimate the effects of interventions." Journal of the American Statistical Association 107.499 (2012): 886-900.

Robins, James. "A new approach to causal inference in mortality studies with a sustained exposure period: application to control of the healthy worker survivor effect." Mathematical Modelling 7.9-12 (1986): 1393-1512.

Robins, James M. "Correcting for non-compliance in randomized trials using structural nested mean models." Communications in Statistics-Theory and methods 23.8 (1994): 2379-2412.

Schuemie, Martijn J., et al. "Detecting adverse drug reactions following long-term exposure in longitudinal observational data: The exposure-adjusted self-controlled case series." Statistical methods in medical research 25.6 (2016): 2577-2592.

Suissa, Samy. "The case-time-control design." Epidemiology (1995): 248-253.

9 Appendix

Proof of Lemma 1

Proof. **Case 1**: $t \geq m$
$E[H_{0t}|\bar{A}_m, \bar{L}_m, \alpha] = E[Y_t - \sum_{j=0}^{t} \gamma_{jt}(\bar{A}_j, \bar{L}_j; \Psi)|\bar{A}_m, \bar{L}_m, \alpha]$
$= E[H_{mt}|\bar{A}_m, \bar{L}_m, \alpha] - E[\sum_{j=0}^{m-1} \gamma_{jt}(\bar{A}_j, \bar{L}_j; \Psi)|\bar{A}_m, \bar{L}_m, \alpha]$
$= E[H_{mt}|\bar{A}_{m-1}, \bar{L}_m, \alpha] - \sum_{j=0}^{m-1} \gamma_{jt}(\bar{A}_j, \bar{L}_j; \Psi)$ by (7)
$= E[H_{0t}|\bar{A}_{m-1}, \bar{L}_m, \alpha]$
Case 2: $t < m$
The result is trivial. □