

Inverse Probability of Treatment Weighting Under Violations of Positivity

Tomas D. Morley

July 31, 2018

Abstract

Acknowledgements

Contents

1	Introduction	5
1.1	Outline	6
1.2	Software	6
2	Marginal structural models	8
2.1	Counterfactuals and causality	8
2.2	Confounding	8
2.3	Directed Acyclic Graphs: graphical representations of causality	10
2.4	Time dependent confounding	10
2.5	Inverse Probability of Treatment Weighting	12
2.5.1	Applications of IPTW	14
2.6	Assumptions	15
2.6.1	No unmeasured confounders	15
2.6.2	Consistency	15
2.6.3	No model misspecification	15
2.6.4	No measurement error	15
2.6.5	Positivity.	15
3	Simulating from marginal structural models	18
3.1	Algorithm	19
3.2	Algorithm with positivity violations.	19
4	Violations of Positivity	21
4.1	Extended discussion of algorithm linking to positivity	21
4.2	Simulation scenarios	21
5	Simulation study	22
5.1	Data Structure	22
5.2	Number of positivity compliant doctors.	22
5.3	Varying levels of threshold.	22
5.4	Simulation Algorithm	22
5.4.1	Algorithm	22
5.4.2	Discussion of how algorithm works	24
5.5	Constructing IPT weights	24
5.6	Simulation Set-up	25
5.7	Results	25
6	Discussion and Conclusion	25
6.1	Limitations	25

1 Introduction

Marginal structural models (MSMs) are a popular class of models for performing causal inference in the presence of time dependent confounders. These models have an important application in areas of research such as epidemiology, social sciences and economics where randomised trials are prohibited by ethical or financial considerations, and hence confounding cannot be ruled out by randomization. Under these circumstances confounding can obscure the causal effect of treatment on outcome. An example of this, common in epidemiological studies, occurs when prognostic variables inform treatment decisions while at same time being predictors of the outcome of interest. In a longitudinal setting this is further complicated when the confounder itself is determined by earlier treatment. One consequence is that regression adjustment methods do not control for confounding in the longitudinal case and other techniques are required.

The Inverse probability of treatment weighting (IPTW) estimator is a technique which leads to consistent estimates in the presence of has been applied to censoring, missing data and survey design problems. The central idea is that by weighting the observed data in order to create a pseudo population is constructed in which treatment is assigned at random. Subsequent analysis where we ignore the confounder is then possible. which inference on the target population can be achieved. For example, when there is missing data weights can be used to create a pseudo-population in which there is no missingness. In the context of MSMs, the IPT weights relate to a pseudo-population in which there is no longer any confounding between the confounder and treatment and causal inferences can be made.

Underlying the IPTW method for estimating MSMs are four assumptions: 1) consistency 2) exchangeability 3) positivity 4) and correct model specification. Exchangeability, also known as the no unmeasured confounding assumption, is closely linked to causality?? Several studies have considered violations of exchangeability and corrected model specification. Positivity has received less attention because in typical observational study positivity violations are not suspected explain why. In the clinical context that we consider, protocols (give some examples, like Platt 2012) threaten to violate the positivity assumption and we investigate whether MSMs are robust against positivity. The focus of this thesis will be on violations of the positivity assumption. Positivity means that within every strata spanned by the confounders, there must be a positive probability of patients being exposed or unexposed to treatment. For example, in a medical context, if treatment protocols demand that treatment is initiated whenever a prognostic variable falls below a pre-defined threshold, there will only be exposed and no unexposed patients in this strata of the confounding prognostic variable. make decisions based on protocols positivity can be. In the absence of structural positivity violations, there is always the threat that random zeroes arise in some strata of the confounder especially when the sample size is small or the number of confounding variables is large. In each case the sparsity of data within the strata of the confounder results in a high chance that positivity is violated. Positivity violations increase the bias and variance of estimates of the causal effect but the extent of the damage is not well known. The central aim of this thesis will be to investigate positivity violations when fitting MSMs to longitudinal data. To our knowledge positivity violations have not been systematically studied in the literature from a simulation point of view. We quantify the bias and variance introduced due to positivity violations and hope to provide practical advice to researchers tempted to fit MSMs to overcome confounding without realising the potential consequences of positivity violations in their data.

Throughout this thesis we focus on clinical applications as examples. In the literature on

marginal structural models the causal effect of Zidovudine on the survival of HIV positive men is often cited as an example. In this example a patients white blood cell (CD4) count is a prognostic variable that influences a doctor's decision to initiate treatment while at the same time being a predictor of survival. As a result CD4 count is a confounder. In the longitudinal setting previous treatments influence CD4 count. As such studies often depend on protocols which means that positivity in some levels of the confounder make this a suitable example for our purposes.

The structure of this thesis is as follows. In section 2 of part 1, the model considered in this thesis and its important aspects are explained. In part 2 simulating from this statistical model is discussed in detail. In part 3 the model under dynamic strategies is considered and comparisons are drawn with the static case. In part 4 we entertain violations of positivity in the data, this section represents the novelty in this thesis. Part 5 conducts a simulation study. Part 6 includes a discussion, conclusions and suggestions for future work.

A second consequence is that simulating data from a specific marginal structural models is more challenging when the data is to exhibit time dependent confounding.

Look through literature for applications of MSMs

Allow for the joint determination of outcomes and treatment status or omitted variables related to both treatment status and outcomes (Angrist 2001).

A covariate L is a confounder if it predicts the event of interest and also predicts subsequent exposure. Explain how this actually happens, as U_0 is a common ancestor of A through L and also Y , also that there is selection bias, and L is sufficient to adjust for confounding see Havercroft algorithm code page bottom.

1.1 Outline

In this thesis we assess the impact of violations of the positivity assumption on the performance of marginal structural models.

The first chapter introduces marginal structural models and inverse probability of treatment weighting. Particular attention is given to the role of the positivity assumption in MSMs and the trade-off between finer confounding control and positivity. Chapter 2 explains the issues surrounding simulating data from a specific MSM with a longitudinal structure that captures the issues which arise in time dependent confounding. This chapter includes a literature review of algorithms that have been developed to simulate from a given marginal structural model. A particular simulation algorithm which is versatile enough that it can be used to introduce violations of positivity is then selected and explained in more detail. Chapter 3 presents simulation results and key findings. Chapter 4 uses real world data in which positivity violations arise as a result of treatment protocols in a chemotherapy trial. The last chapter concludes and provides limitations and directions for future work.

1.2 Software

All simulations and analysis carried out in this thesis use the Python programming language and are provided with this thesis. Several modules were used to extend the base Python language and these are highlighted in the code where appropriate. The *survey* and *ipw* packages written in the R programming language were used to provide functionality not currently available as a Python module. These packages are

freely available through the Comprehensive R Archive Network (CRAN). Combining R functionality in Python code is made possible through the *rpy2* Python module.

All functions used for this thesis are provided in appendices. Appendix ? contains the code for simulating data from the chosen marginal structural model. Appendix ? contains the code used to generate the results and graphs in this thesis.

2 Marginal structural models

Marginal structural models (MSMs) are a class of models for the estimation of causal effects from observational data [Robins et al. \(2000\)](#). More specifically, marginal structural models allow us to make causal statements in the presence of time dependent confounding. Several key assumptions need to be met to achieve this. In this section we describe MSMs and focus on those assumptions. This section explains the central concepts behind marginal structural models and introduces the notation that will be used in this thesis. Marginal structural models use only observed data and a set of assumptions to investigate causal effects. Positivity, the focus of this paper, is tied closely to confounder and, as we shall see, there is a trade-off between finer confounder control and positivity. The algorithm adapted in this paper simulates longitudinal data and hence we discuss time dependent confounding and the extra bits required for this. In particular, we draw attention to the core parts of MSMs needed for this study - linking to positivity.

- describe other methods like g formula as alternative to MSMs
- MSMs are models for some aspect (like the mean) of the distribution of counterfactuals.
- many types like a marginal structural cox model (maybe let this follow on after weights part.

2.1 Counterfactuals and causality

In the counterfactual framework ([Neyman \(1923\)](#), [Rubin \(1978\)](#), [Robins \(1986\)](#)) the causal effect of treatment X on outcome Y for one subject can be defined as the difference between that subject's outcome had they been exposed and unexposed to X . In other words, one outcome is necessarily counterfactual because in reality the same subject cannot be both exposed and unexposed to X . If we denote the outcome when exposed as $Y_{x=1}$ and the outcome when not exposed as $Y_{x=0}$ then the causal effect for one subject can be expressed as $Y_{x=1} - Y_{x=0}$. For example, suppose a subject with a headache takes ibuprofen ($X = 1$), a popular treatment for headaches. After a suitable amount of time, say one hour, the headache either remains $Y_{x=1} = 1$ or has passed $Y_{x=1} = 0$. The outcome which is not observed is the counterfactual outcome that prevails had the subject, contrary to fact, not taken ibuprofen. In other words, we don't observe $Y_{x=0}$.

Often we are interested in the average causal effect for a population rather than for one subject. Suppose sixty subjects are suffering from a headache and every subject was given ibuprofen. After one hour each subject will either have a headache ($Y_{x=1} = 1$) or their headache will have passed ($Y_{x=1} = 0$). The average outcome across all subjects is $\mathbb{E}(Y_{x=1})$ or equivalently when Y is a dichotomous variable, $\mathbb{P}(Y_{x=1})$. The relevant causal comparison is now between $\mathbb{P}(Y_{x=1})$ and $\mathbb{P}(Y_{x=0})$, the latter being the counterfactual had none of sixty subjects been exposed. We do not observe the quantity $Y_{x=0}$ for any subject, and consequently we do not observe the quantity $\mathbb{P}(Y_{x=0})$.

Dawid 1979 - we can vary the individual I and treatment X but this is largely a conceptual entity because only one treatment can in fact be applied to any unit. - Can also use the odds ratio as the comparison instead of the risk difference. - only have the data and observed outcomes, not the counterfactuals.

2.2 Confounding

Continuing the example, if group A consists of thirty of the headache sufferers who all took ibuprofen, we would ideally compare the quantity $\mathbb{P}(Y_{x=1}|X = 1) = \mu_{A_{x=1}}$ with the

quantity $\mathbb{P}(Y_{x=0}|X = 1) = \mu_{A_{x=0}}$. As $\mu_{A_{x=0}}$ is not actually observed, we could instead compute the observable quantity $\mathbb{P}(Y_{x=0}|X = 0) = \mu_{B_{x=0}}$ from the remaining thirty subjects who did not use ibuprofen and belong to group B. Replacing the comparison between $\mu_{A_{x=1}}$ and $\mu_{A_{x=0}}$ with the comparison between $\mu_{A_{x=1}}$ and $\mu_{B_{x=0}}$ will have a causal interpretation if $\mu_{A_{x=0}} = \mu_{B_{x=0}}$. In other words, if a subject from group B can be viewed as an analogue of a subject from group A had they, contrary to fact, not received ibuprofen.

If $\mu_{A_{x=1}} \neq \mu_{B_{x=0}}$ then the comparison $\mu_{A_{x=1}} - \mu_{B_{x=0}}$, a measure of association, is confounded for $\mu_{A_{x=1}} - \mu_{A_{x=0}}$, a measure of causal effect (Greenland et al. (1999)). For example, if all the subjects in group A are male it would be reasonable to ask whether their sex influenced their decision to take ibuprofen. Suppose that males also tend to have headaches of a shorter duration so that at the end of one hour they are less likely to have a headache than females. The result is that both the decision to take ibuprofen and the probability of having a headache at the end of one hour are dependent on the sex of the subject. This obscures the causal effect of ibuprofen on headaches because there is a spurious association between X and Y through the subject's sex. We cannot establish whether the outcome is due to a causal relationship between ibuprofen and headache alleviation, a relationship between sex and headache alleviation or a mixture of the two.

Closely related to confounding, exchangeability is the assumption that the distribution of the counterfactual outcomes Y_x is independent of the actually observed treatment X . When exchangeability holds, subjects from group A and group B are exchangeable in the sense that were they all to remain untreated the distribution of the counterfactual outcomes Y_x would be the same in the two groups Daniel et al. (2013). Imagine exchanging a subject from group A with a subject from group B where both receive the treatment prevailing in their new group. Under exchangeability, the average outcome in the two groups is unchanged Imbens and Rubin (2008). However, exchanging subjects between group A and group B introduces females into group A and males into group B. As males have a higher probability that $Y = 0$, exchanging subjects changes the distribution of the counterfactuals. The relationship between confounding and exchangeability is why the assumption of exchangeability is also called the assumption of "no unmeasured confounding".

- set-up why comparisons are possible within strata and then averages across strata and why this means that naive methods for addressing time fixed confounding work
- Hint that the finer the confounding control the more accurate the analysis but this has consequences for positivity.
- from pearl 2001: namely, that if we compare treated vs. untreated subjects having the same
- values of the selected factors, we get the correct treatment effect in that subpopulation of subjects.
- Explain clearly why it is a bias.
- Also explain why we want to be judicious in our choice of number confounders to control for. Can't include everything (Dawid 1979 on this)
- Explain that structural parameters only coincide with associational parameters under exchangeability.
- Define what the naive analysis is - analysis without adjustment.
- knowing the value of Z gives us no more information about the distribution of the counterfactuals Y_x
- Explain that in a randomized experiment exchangeability is guaranteed because X is automatically not related to any other variables.

- Randomization ensures that missing values occur by chance. So the counterfactual values that we don't see for some observations are missing randomly and not due to confounding through a covariate.
- Any residual confounding cannot be due to the variables that we have conditioned on.

2.3 Directed Acyclic Graphs: graphical representations of causality

Causal relationships, like those described in the previous section, can be represented using graphs. A graph consists of a finite set of vertices v and a set of edges e . The vertices of a graph correspond to a collection of random variables which follow a joint probability distribution $P(v)$. Edges in e consist of pairs of distinct vertices and denote a certain relationship that holds between the variables Pearl (2009). The absence of an edge between two variables indicates that the variables are independent of one another. The direction of the causal relationship is denoted by an arrow and is acyclic because causal relationships between two variables only proceed in one direction. There are no feedback loops or mutual causation because in a causal framework a variable cannot be a cause of itself directly or indirectly Hernán et al. (2004).

For example, figure ? represents the case where interest is in the causal relationship between treatment X and outcome Y . Treatment is assigned according to conditional distributions $P(\text{treatment}|\text{male})$ and $P(\text{treatment}|\text{female})$. Once treatment has been assigned, the outcome Y is determined by both X and Z by the conditional distribution $P(Y|X, Z)$. Pearl2001, Pearl2014. Blocking or screening off Z has the same intuition as explained in the section on confounding. The causal effect of X and Y cannot be different between two subjects because of Z when everyone is that strata has the same value of Z . Blocking is the same as holding Z constant. Intuition to drive forward is that difference in outcome cannot be due to strata when everyone shares that strata.

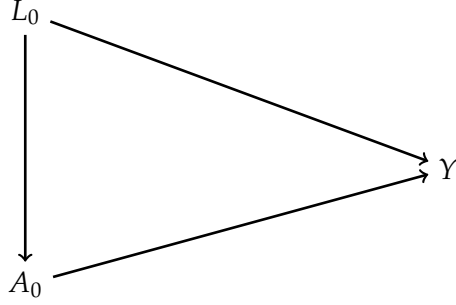
Show the same graph without causal relationship between X and Y . There is a marginal dependence between X and Y through Z , but once we condition on Z this dependence disappears as shown by the lack of an edge between X and Y . Once we condition on Z (i.e. we know that the subject was male or female.) then the marginal dependence disappears. Introduce idea of common cause here as well.

Both treatment and outcome are determined by sex leading to a spurious association between X and Y through Z . This is called a "back door" path between X and Y . Conditioning on Z is represented graphically by blocking the back door and any spurious associations to allow causal estimation

- use example of cause by just removing an arrow from the DAG to illustrate the point that there must be a cause so adding back the causal arrow of interest does not change the fact that part of the cause comes through the confounder.
- common cause and structural approach to selection bias paper.
- the absence of an arrow means no direct effect between two variables cole 2009 illustrating bias paper.

2.4 Time dependent confounding

So far we have considered the time fixed context in which treatment and confounders take on a single value. It was sufficient to block the "back door" path between the treatment and outcome by conditioning on the confounding variable(s). In the headache example, the causal



Causal graph

effect of ibuprofen on headache alleviation was confounded by sex. For most people, sex is a time-fixed covariate because it does not change value over time. To broaden the setting to a time dependent context, we adopt the canonical example of the causal effect of Zidovudine (AZT) on mortality amongst human immunodeficiency virus (HIV)-infected subjects [Hernan et al. \(2000\)](#). In this example, subjects are measured at baseline $t = 0$ and at subsequent visits. In each visit the patient's CD4 lymphocyte count is measured and a treatment decision made. Survival at the end of follow-up is a binary outcome equal to 1 if the patient has died and 0 otherwise.

The time-fixed notation can be extended to include subject histories for time varying variables. Treatment and covariate histories up to visit k can be represented by an over-head bar. For example, $\bar{X}_k = \{X_0, \dots, X_k\}$ represented the vector of treatment decisions while $\bar{Z}_k = \{Z_0, \dots, Z_k\}$ represents the vector of measurements on the time dependent-confounder Z . Time-fixed covariates like sex, or covariates which change linearly over time like age are typically recorded at baseline ($t = 0$) and we denote the collection of baseline covariates as V_0 . The outcome of interest at the end of follow-up is mortality Y which is a binary variable taking the value 1 if the patient is dead and 0 otherwise.

Just as in the time-fixed case, time-dependent confounders lead to spurious associations between X and Y through a "back door" path between X and Y through L . To estimate a causal effect it is necessary to block this path by conditioning on the confounding variables. Figure ? gives an example of this in the time dependent case for two periods ($t = 0, 1$). In the first period a treatment decision is made based on the measured confounder Z_0 . In the second period ($t = 1$) a new treatment decision is made based on both Z_0 and Z_1 . Conditioning on \bar{Z} under this DAG leads to a consistent estimate of the causal effect because doing so blocks all paths between X_0 and X_1 and Y except the causal path of interest.

However, the time-dependent context also admits structures like the middle pane of figure ? with the addition of a causal relationship between X_0 and Z_1 . It is now possible for current treatments to be a determinant of future confounders which are in turn determinants of future treatment [Robins \(2000\)](#). As a result the effect of A_0 on Y is mediated through L_1 in the path $A_0 \rightarrow L_1 \rightarrow Y$. Blocking this path by conditioning on Z also blocks some portion of the effect of A_0 on Y and will lead to a biased estimate.

A second danger in the time-dependent context arises when Z is a common effect of treatment and an unmeasured variable U which also influences the outcome Y . There is no direct association Figure ? shows the same two structures with the addition of a an unmeasured variable U which

influences Z and Y . Conditioning on Z . Selection bias precludes unbiased estimation [Hernán et al. \(2004\)](#). There is a mediating relationship between A and Z in which case there is a spurious relationship between A and Y again? This is less intuitive and so examples are best according to [Cole et al. \(2010\)](#). We can say that Z is a common effect of A and U , once we condition on Z we create a dependence of A on U . U is a cause of Y and hence there is an association between A and Y . This association is present even when there is no direct causal path between A and Y .

Hazard ratios and selection bias [Hernán \(2010\)](#). Actual application will look at toxicity of treatment. Some people will be susceptible and drop out leaving more people in the untreated arm of the study.

general point about selection bias is that the general population is not a valid control group.

Conclusion, 1) clearly a different technique is required for analysis 2) the nature of time dependent case needs to be described fully enough to explain why we choose the simulation algorithm that we choose and any holes in it. In subsequent sections our choice of simulation algorithm will be motivated by the structure of time dependent confounding as well as the viability of introducing positivity violations which are propagated through the time dependent structure. Explain meaning of a collider and that a collider that is conditioned on will not block confounding. Essentially with this kind of data we cannot use confounding or stratification methods.

- introduce survival analysis?
- Intuition from Pearl 2009 book pp. 17 on sprinklers.
- Simpson's paradox linked to making comparisons within strata
- explain that we are often interested in parsimonious models so cannot have all covariates U that will create associations between X and Y
- explain why we do not need to worry about the path between A_0 and A_1
- explain why mediation is likely to occur in example.
- explain why saturated models cannot be used because they will have
- intuitive examples of selection bias.
- Saturated models are not an option because they would be computationally intensive and so we use parametric models which also links to positivity because we smooth over zeroes in certain strata.
- Explain why hazard ratios have a built in selection bias after giving some examples of why selection bias arises. It is because it is selective on patients reaching the time period in question. Is this also the reason why summary methods create selection bias.
- give an intuitive explanation for why CD4 count is a predictor of subsequent treatment and of death.
- Because treatment is randomized (at baseline) in expectation the proportion of men and women in each group is the same.

2.5 Inverse Probability of Treatment Weighting

The previous section has highlighted how standard approaches for controlling for confounding in a time dependent context may lead to biased estimates. In this section we describe a technique called inverse probability of treatment weighting that can be used to obtain unbiased estimates of the causal effect of treatment on outcome in the presence of time dependent confounding.

Inverse probability of treatment weighting is a technique that can be used to obtain unbiased estimates of the causal effect of treatment on outcome in the presence of time dependent

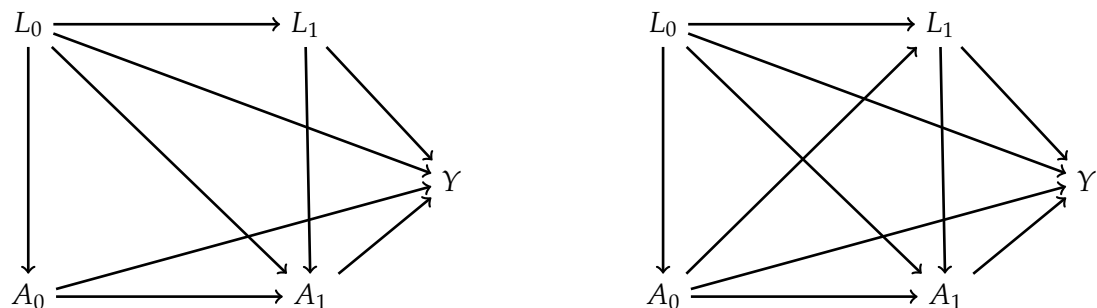


Figure 1 DAG

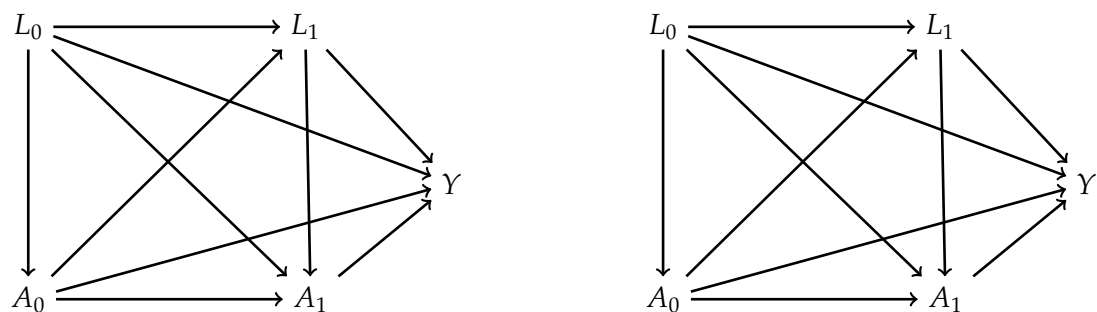


Figure 1 DAG

confounding. The intuition behind the technique is that by re-weighting the data a pseudopopulation is created in which the treatment is independent of any measured confounders. Regression analysis on the pseudopopulation can be carried out without the need to control for confounders eliminating the problems which arose in the previous section due to conditioning on Z . Crucially, in the pseudopopulation, the causal effect of X on Y remains unchanged. As a result, it is possible to estimate the true causal effect of X on Y .

Construction of weights

- why does it work and naive methods do not?
- How does it break the link?
- Dangers associated with stratification and controlling for methods highlighted already explain why we need other methods - explain a few of these like G-estimation, SNTM etc.
- creates a pseudo-population in which we have something similar to an experimental setting
- Because we need weights this means we need a model for the weights - model can be non parametric or parametric depending on data used.
- Contrast IPTW methods with stratification methods.

Horvitz and Thompson (1952) Areas where IPTW has been used (time dependent confounding, comparing dynamic regimes, missing data)

Inverse probability of treatment weighting is a technique that re-weights subject observations to a population where assignment of treatment is at random. An early example of this technique is the ? weighted estimator of the mean. In the context of marginal structural models, a weight is calculated for each subject which can be thought of informally as the inverse of the probability that a subject receives their own treatment **Robins et al. (2000)**. The result of applying these

weights is to re-weight the data to create a pseudo-population in which treatment is independent of measured confounders [Cole and Hernán \(2008\)](#). Crucially, in the pseudo population the counterfactual probabilities are the same as in the true study population so that the causal RD, RR or OR are the same in both populations [Robins et al. \(2000\)](#).

$$w_{t,i} = \frac{1}{\prod_{\tau=0}^t p_{\tau}(A_{\tau,i} \mid \bar{A}_{\tau-1,i}, \bar{L}_{\tau,i})}$$

stabilized weights

$$sw_{it} = \frac{\prod_{\tau=0}^t p_{\tau}(A_{\tau,i} \mid \bar{A}_{\tau-1,i})}{\prod_{\tau=0}^t p_{\tau}(A_{\tau,i} \mid \bar{A}_{\tau-1,i}, \bar{L}_{\tau,i})}$$

The use of IPTW is valid under the four assumptions of consistency, exchangeability, positivity and no misspecification of the model [Cole and Hernán \(2008\)](#).

Informally a patients weight through visit k is proportional to the inverse of the probability of having her own exposure history through visit k (Cole and Hernan 2008)

The weight is informally proportional to the participants probability of receiving her own exposure history

As these weights have high instability we need to stabilize them. The unstabilized weights can be driven by only a small number of observations. Why are they unstable?

- true weights are unknown but can be estimated from the data.
- A_t is no longer affected by L_t , and crucially the causal effect of \bar{A} on Y remains unchanged

Be more specific about what is contained in the weights. The denominator depends on the measured confounders L the numerator does not.

- weighted regression and MSM are equivalent.

Point out that we need baseline variables in the conditional statments in the num and denom of the weights otherwise we break the relationship between outcome and baselines in the new pseudo-population. If the baseline variables are not confounders, then we do not want to break this relationship. Baseline covariates also help to stabalize the weights (how?)

importantly, changing the relationship between L and A , won;t change the relationship between L and Y . This means that an intervention in A does not affect the relationship between L and Y . So we remove the link between L and A and assign to A the value of treatment on or off. Once we place the patient on treatment, regardless of the relationship which had existed before hand between the covariate and treatment, a new relationship between A and Y exists in which the covariate has no say.

2.5.1 Applications of IPTW

Different from application of MSMs - IPW is a technique which has been applied to MSMs - i.e. MSMs are on example of how IPTW can be used.

- Crucial that the relationship between Y and X remains the same in the new population. I.e. the marginal structural model is the same.

2.6 Assumptions

This section presents a more formal discussion of the key assumptions underlying MSMs. Most attention will be given to positivity. Conditions under which IPTW work are largely untestable (westreich 2012)

2.6.1 No unmeasured confounders

$$Y_x \perp\!\!\!\perp X$$

2.6.2 Consistency

Cole and Frangakis (2009) ?

2.6.3 No model misspecification

Correct specification of the inverse probability weighting (IPW) model is necessary for consistent inference from a marginal structural Cox model (MSCM).

2.6.4 No measurement error

2.6.5 Positivity.

The final assumption underlying MSMs, and the central topic of this thesis, is the positivity assumption. MSMs are used to estimate average causal effects in the study population, and one must therefore be able to estimate the average causal effect in every subset of the population defined by the confounders Cole and Hernán (2008). The positivity assumption requires that there be exposed and unexposed individuals in every strata of the confounding covariates. For example, when treatment is Zidovudine and CD4 count is the confounder, there must be a positive probability of some patients being exposed and unexposed at every level of CD4 count. Positivity can be expressed formally as $Pr(A = a | L) > 0$ for all $a \in A$, which extends straightforwardly to the time dependent case where the positivity assumption must hold at every time step conditional on previous treatment, time dependent confounders and any baseline covariates:

$$Pr(A_{it} = a_{it} | L_{it}, A_{i,t-1}, V_{i0}) > 0$$

Models for the risk $P(Y = 1 | A = a, L = l)$ are commonly studied in epidemiological applications. Applying basic probability rules reveals that the risk can be re-written with the term $Pr(A = a | L = l)$ in the denominator:

$$P(Y = 1 | A = a, L = l) = \frac{P(Y = 1, A = a, L = l)}{Pr(A = a, L = l)} = \frac{P(Y = 1, A = a, L = l)}{Pr(A = a | L = l)Pr(L = l)}$$

This model is only estimable when $Pr(A = a | L = l) \neq 0$. Therefore, when positivity does not hold it is not possible to estimate the model. In the context of MSMs a similar problem emerges. Although weighting via IPTW allows naive estimation of (?) without including the confounders, the weights in (?) involved the term $Pr(A = a | L = l)$ in the denominator. This means that the weights are inestimable whenever positivity is violated. In order to estimate the causal effect of A on Y , weights must be estimable in every subset of the population otherwise the average causal effect in the study population cannot be estimated.

In practice, positivity can arise when random zeroes or structural zeroes are present in some levels of the confounding covariates. Random zeroes arise when, by chance, no individuals or all individuals, receive treatment within a certain strata as defined by the covariates. For example, [Cole and Hernán \(2008\)](#) studies positivity violations in individuals in strata defined by CD4 count and viral load. By increasing the levels of CD4 count the chances of random zeroes also increases and [Cole and Hernán \(2008\)](#) show that the IPT weights rapidly lose their stability with the consequence that causal effects are no longer estimable. Researchers applying IPTW methods must actively check that there are both treated and untreated individuals at every level of their covariates within cells defined by their covariates because parametric methods will smooth over positivity violations and not provide any indication of nonpositivity. Increasingly refined covariates are attractive because they provide better control of confounding, but the point that [Cole and Hernán \(2008\)](#) make is that this control needs to be traded off against increased occurrence of random zeroes and subsequent instability of IPT weights.

More relevant to this thesis are violations of the positivity assumption due to structural zeroes. These occur when an individual cannot possibly be treated or if an individual is always treated within some levels of the confounding covariate, as is the case in the clinical protocol example motivating this thesis. Several studies give examples of structural violations of the positivity assumption in epidemiological contexts. In [Cole and Hernán \(2008\)](#) structural zeroes arise when the health effects due to exposure to a chemical are confounded by health status proxied by being at work. If individuals can only be exposed to the chemical at work then all individuals not at work will be unexposed. A second example is liver disease as a contraindication of treatment. If individuals with liver disease cannot be treated then all individuals in the "liver disease = 1" strata will be untreated. In [Messer et al. \(2010\)](#) structural zeroes arise in the context of rates of preterm birth and racial segregation, whereas [Cheng et al. \(2010\)](#) find structural zeroes in the context of fetal position and perinatal outcomes. Our motivating example is most closely related to liver disease as a contraindication, except that the clinical protocols require that patients with low CD4 count always be treated instead of never being treated, as in the case in the liver disease example.

Although in many epidemiological settings the positivity assumption is guaranteed by experimental design, studying positivity violations is relevant because, as our own motivating example and the examples above suggest, structural violations do occur, and random zeroes are always possible especially at finer levels of confounding covariates. Studying the finite sample properties of MSMs under violations to positivity is therefore an important issue which is yet to be dealt with systematically in the literature. As [Westreich and Cole \(2010\)](#) points out, positivity violations, positivity violations by a time varying confounder pose an analytic challenge and they suggest g-estimation or g-computation may be a way forward. A good start to dealing with the time varying confounder case is to see how well MSMs work when positivity is violated. This is also a novelty of this thesis.

6. estimated weights with a mean far from one, or very extreme values indicate either non-positivity or model misspecification of the weight model.
7. It is not always true that we want more finely tuned covariates for confounder control because the bias and variance of the effect estimate may increase with the number of categories. This is similar to the positivity masking example.
8. Our results are equally valid for other circumstances in which positivity may arise.
9. Also think about how the number of categories of exposure increases the chance that one level of exposure will have a positivity.

10. Westreich and Cole 2010 have suggested that methodological approaches are needed to weigh the resultant biases incurred when trading of confounding and positivity. The framework we use is flexible enough to allow this in a simulation setting.

If the structural bias occurs within levels of a time-dependent confounder then restriction or censoring may lead to bias whether one uses weighting or other methods (Cole and Hernan 2008). In fact, weighted estimates are more sensitive to random zeroes (Cole, Hernan, 2008) Introducing violations of positivity can be achieved by censoring observations.

But to give an intuitive example, think about how it links back to a situation where sicker patients receive treatment compared to others. So in the "sick" strata of the CD4 count **ALL** patients receive treatment which inflates the IPTW. This also affects how we think about the associational versus causal models. The causal effect might be 50/50 but because sicker patients get treatment the mortality ratio in the treated group is likely to be higher.

The trade-off between positivity and confounding bias is emphasized in Cole2008

Why is practicality important? Cole paper highlights practical advice to practitioners. positivity can be violated in a practical setting because of two few strata, it can be the result of protocols in a clinical setting and it can be seen as a trade-off between exchangeability (and we need more measured predictors to maintain exchangeability) and positivity where more predictors leads to more likely a zero problem.

- Dynamic strategies evaluated using MSMs will have rules like, start treatment if CD4 falls below a certain threshold. See Didelez presentation on this
- Explain that there is positivity in estimation of the structural model and also in the weight model. The reason why positivity is more important in the weight model is because when the weights are unstable the estimates can be very wrong as a result.

Have been used for missing data problems. see pp.442 of Hernan, Brumback, Robins 2001 for a list of papers linked to this

A model that parameterises $P(Y \mid do(A = a))$ is called a marginal structural model (MSM) as it is marginal over any covariates and structural in the sense that it represents an interventional rather than observational model.

3 Simulating from marginal structural models

In order to assess the impact of violations of the positivity assumption on the performance of the IPTW estimator we simulate data from a specific marginal structural model in a series of monte carlo simulations. In this chapter we start by describing monte carlo simulations in general terms. Next, we consider several important criteria that a simulation model must exhibit in the context of MSMs. Several algorithms are briefly compared. An appropriate simulation algorithm for the purposes of this thesis means we can simulate data from a specific MSM against which we can test the model. The algorithm should also generate data with the time-dependent confounding structure described earlier. Specifically, it should be possible for treatment to determine confounder and confounder to determine treatment in turn. Selection bias should also be possible. In this section we briefly review several simulation algorithms which have been applied to the study of MSMs. We then focus on the algorithm suggested in ? and explain why it satisfies our requirements and how positivity violations are introduced and propagated through the model. This chapter also includes a broader view of monte carlo simulations.

Monte Carlo Simulations In statistical research interest often lies in the estimation of a population parameter θ . When only a sample X from the population is available, statistical methods are applied to estimate the population parameter $\hat{\theta}$. If a second sample is drawn from the population it will result in second estimate $\hat{\theta}_2$ and so on for more samples.

Analogously, a "true" model can be specified and a sample of data generated according to that model. The success of a technique can be compared to the true model. Any single sample generated

- simulate from a specific "true" or "known" model.
- estimate method
- variability hence do this many times
- aggregate results and calculate SE
- to study finite sample properties we simulate a finite sample (of size $n = 1000$ for example.)

Naturally this requires knowing the parameters in advance in order to simulate from the model. In the case of MSMs this requires simulating from a specific MSM

Specific MSM Testing positivity violations requires generating data from a known MSM. For example, suppose that the chosen MSM was $P(Y | do(a)) = \alpha + \beta a$, then the observational data from which this is derived would be Y , A and L . The simulation needs to allow β to be estimated from only that observational data. In order to simulating from a specific MSM we need to generate observational data in such a way that it conforms to a specific MSM. For instance, one specific MSM could be written as $P(Y | do(a)) = \alpha + \beta a$. However, we never see this so we need to generate data in such a way that we get this model from observational data and that we can then return to it.

observational structure The observational structure needs to include the central parts of time dependent confounding - in particular Y and A are common effects of U which creates a confounding relationship. We also need selection bias, some models like hazard ratios have a built in selection bias so this is particularly important.

noncollapsibility Collapsibility means there is no incompatibility between the marginal model and the conditional distributions used to simulate the data. Provide example of this. Explain how this affects the simulation algorithm. Especially hazard ratios which are non-collapsible.

Models are noncollapsible when conditioning on a covariate **related to the outcome** changes the size of the estimate even when the covariate is unrelated to the exposure. Illustrate why this happens with survival models.

Collapsibility comes in because we cannot collapse a model conditional on L into one conditional on only A.

Survival models are non-collapsible. Hence we cannot easily simulate from them. Instead we use U as a sneaky trick. Explain why survival models are non-collapsible

This is particularly important because collapsibility and confounding are often treated as identical concepts when in fact they are not. [Greenland et al. \(1999\)](#)

Simulation algorithms literature review. Several algorithms for simulating from a specific MSM have been suggested in the literature.

[Havercroft and Didelez \(2012\)](#) [Bryan et al. \(2004\)](#) [Westreich et al. \(2012\)](#) [Young and Tchetgen Tchetgen \(2014\)](#)

- Bryan 2004: fixes the vector L at the beginning which means A never affects L which doesn't make no sense. In other words it doesn't have the observational structure we are looking for.
- Do we introduce a form of selection bias into the data when we force positivity according to protocols? At baseline the proportion of people in any CD4 strata was unknown but randomized and hence in expectation it should be the same in treated and untreated.
- Young 2014 - Law of the observed outcome conditional on the measured past. What this paper shows is that the regression results are not correct but the IPTW ones are fine. The issue is comparing IPTW to normal regression results. As we compare IPTW results to IPTW results under positivity it should be fine to use the Havercroft algorithm.

3.1 Algorithm

3.2 Algorithm with positivity violations.

The confounding structure occurs because there is a relationship between A and Y confounded through L through U? Y and A are common effects of U which creates an association between A and Y other than the causal association we want to study in the MSM - refer to structural bias paper.

The parameter of interest could be expected survival or the five year survival probability

Need to specify what the MSM is, give an example of it as a hazard function. Survival is completely determined by the hazard function.

U allows us to get any distribution we like for Y marginal over covariates, Would L itself allow this? probably not. We can somehow get from this the subjects counterfactual survival time.

Importantly, this expression on the left hand side has unobserved counterfactuals, but the right hand side has only observed quantities which would be observed in an actual observational study.

non-collapsibility is an unresolved issue here. So even if we investigate positivity we can still only do so for collapsible models?

Equivalent way of motivating dividing the joint distribution by $\Pr(A | L)$ is through IPTW.

The important point is to resolve the correct, causal parameter on a

How does selection bias relate to U ?

Explain why the algorithm allows positivity to be propagated through the patient's history.

1. derive relationship between MSM and DAG and the correct conditional distributions. Follows from truncated factorisation why we can get $P(Y | do(a))$
2. think of this process as if we had fixed a treatment vector in advance. consistency assumption.
3. HD 2012, with Pearl and truncated factorization formula, show that it is possible to link the counterfactual represented by $P(Y | do(a))$ to observational data generated in an observational way. But the problem arises when the model is non-collapsible or non-linear.

In the one shot case we set $A = 1/0$ because we are interested in the outcome under either of these treatment scenarios. In the time dependent case, A is a vector of 0s and 1s and we want to pretend that we decide in advance that the whole vector A is specified. But A and L have a complex interplay in an observational setting. So we want to pretend that A (a vector of 1s and 0s) is set in advance but at the same time have the observational structure for A and L .

The relationship between Y and L is then dependent on A . There is no relationship between A and U because of the set-up in the DAG. The variable L blocks this relationship.

In their paper [Havercroft and Didelez \(2012\)](#) develop an algorithm that allows simulating data that corresponds to a particular parameterisation of an MSM. This algorithm provides the bedrock of the simulation structure considered in this thesis. Figure 1 represents the system under consideration. The DAG in figure 1 represents the one-shot non-longitudinal case. Factorising the joint distributions of the variables in figure 1 yields

$$P(U, L, W, A, Y) = P(W)P(U)P(W)P(L | U)P(A | L, W)P(Y | U, A)$$

Where, following definition 1.1 we delete $P(A | L, W)$, a probability function corresponding to A , and replace $A = a$ in all remaining functions

$$P(U, L, W, Y | do(A = a)) = \begin{cases} P(U)P(L | U)P(Y | U, A = a) & \text{if } A = a \\ 0 & \text{if } A \neq a \end{cases}$$

The goal is to simulate from a particular MSM. This means parameterising $P(Y | do(A = a))$. Applying the law of total probability over W, U and L yields

$$P(Y | do(A = a)) = \sum_{w,u,l} P(W)P(U)P(L | U)P(Y | U, L, A = a) = \sum_{u,l} P(U)P(L | U)P(Y | U, L, A = a)$$

Making use of the fact that $P(L, U) = P(L | U)P(U) = P(U | L)P(L)$ and summing over either W and U or W and L yields

$$P(Y | do(A = a)) = \sum_l P(Y | L, A = a)P(L) = \sum_u P(Y | U, A = a)P(U)$$

If we can find suitable forms for either $P(Y | L, A = a)$ and $P(L)$ or $P(Y | U, A = a)$ and $P(U)$ that correspond to the MSM $P(Y | do(A = a))$, then, given suitable values for A, L, U it will be possible to simulate from the chosen MSM.

Choosing a functional form for

$$P(Y | do(A = a))$$

depends on convenience. We need a functional form that can be easily represented by $P(Y | L, A = a)P(L)$. non-linear functions will be hard to work into the analysis.

$U \sim U[0, 1]$ is a good choice because we can use the CDF of Y because $U[0, 1]$ is always between 0 and 1

General health is patient specific but comes from a clear distribution and has a nice medical interpretation. In contrast L would be more difficult to include. It is better as a function of U than a value in of itself.

basic idea behind simulation is that we know what the true model is and therefore being able to compare the true model with the results. Explain why simulations work in a general sense don't take it for granted.

- Explain issue that survival models are not collapsible which is why most algorithms don't work. Big reason we choose HD2012 is because of this
- no model misspecification in the HD2012 algorithm
- stay on treatment after treatment starts
- they motivate a logistic model for the hazard function, they use a discrete equivalent to the hazard function (link to citation about farington study.)
- treatment regime is determined by t^* (starting point of treatment because it is a vector of $\{0, 0, 0, 1, 1, 1\}$)

4 Violations of Positivity

The motivation for using the algorithm of [Havercroft and Didelez \(2012\)](#) is that we have control over how L affects Y , so we can introduce positivity using a threshold. In other algorithms there would be a direct link between L and Y , this would be a problem because altering treatment decisions based on L would affect Y directly.

- creating an artificial population in which positivity is violated in specific ways.

4.1 Extended discussion of algorithm linking to positivity

As described in the introduction, one assumption of the model is that there is a non-zero probability of the event occurring at every startum of the covariate.

- When previous covariates like CD4 count are strongly associated with treatment the probabilities in the denominator of the unstabilized weights may vary greatly. Because we are forcing positivity by using a treatment rule when L falls below a threshold and A is then equal to one, we create a strong association between A and L -> hence the unstabilized weights would vary. (Robins et al 2000 pp. 553)
- present the algorithm again with positivity violations.

4.2 Simulation scenarios

thresholding

percentage of compliant doctors.

5 Simulation study

5.1 Data Structure

Include an example simulation graph plot showing the data colored to show where positivity would arise.

5.2 Number of positivity compliant doctors.

5.3 Varying levels of threshold.

We wish to simulate survival data in discrete time $t = 0, \dots, T$ for n subjects. At baseline $t = 0$ all subjects are assumed to be at risk of failure so that $Y_0 = 0$. For each time period $t = 0, \dots, T$ a subject may either be on treatment, $A_t = 1$, or not on treatment, $A_t = 0$. All patients are assumed to be not on treatment before the study begins. Once a patient commences treatment, they remain on treatment in all subsequent periods until failure or the end of follow-up. In each time period L_t is the value of a covariate measured at time t . In the simulated data, L_t behaves in a similar manner to CD4 counts such that a low value of L_t represents a more severe illness and hence a higher probability of both treatment and failure in the following period. In addition to L_t , the variable U_t represents subject specific general health at time t . Although we will simulate U_t , in a real world application U_t is an unmeasured confounder which

Each time period is either a check up visit or is between two check up visits. If t is a check-up visit and treatment has not yet commenced, L_t is measured and a decision is made on whether to commence treatment. Between visits, treatment remains unchanged at the value recorded at the previous visit. Similarly, L_t which is only measured when t is a visit, also remains unchanged.

We represent the history of a random variable with an over bar. For example, the vector representing the treatment history of the variable A is represented by $\bar{A} = [a_0, a_1, \dots, a_m]$ where $m = T$ if the subject survives until the end of follow-up, or $m < T$ otherwise. Prior to baseline both $A = 0$ for all subjects.

- explain what U is and how it relates to the simulation design/algorithm
- Be more specific on Y
- L_t is a measured confounder
- U_t is an unmeasured confounder.

5.4 Simulation Algorithm

5.4.1 Algorithm

Next, we describe the algorithm used to simulate data from our chosen marginal structural model under time dependent confounding. In the following section we discuss in detail how the algorithm works and the salient features for this thesis. The algorithm is taken from [Havercroft and Didelez \(2012\)](#) who generate data on n patients, for k time periods. The outer loop in the following algorithm $i \in 1, \dots, n$, refers to the patients while the inner loop $t \in 1, \dots, T$ refers to the subject specific time periods from baseline to failure or the end of the study. There will be at least one,

and at most T records for each patient.

Algorithm 1: Simulation Algorithm MSM

Result: Marginal Structural Model Under Time Dependent Confounding

```

1 for  $i$  in  $1, \dots, n$  do
2    $U_{0,i} \sim U[0, 1]$ 
3    $\epsilon_{0,i} \sim N(\mu, \sigma^2)$ 
4    $L_{0,i} \leftarrow F_{\Gamma(k, \theta)}^{-1}(U_{i,0}) + \epsilon_{0,i}$ 
5    $A_{-1,i} \leftarrow 0$ 
6    $A_{0,i} \leftarrow \text{Bern}(\text{expit}(\theta_0 + \theta_2(L_{0,i} - 500)))$ 
7   if  $A_{0,i} = 1$  then
8      $T^* \leftarrow 0$ ;
9   end
10   $\lambda_{0,i} \leftarrow \text{expit}(\gamma_0 + \gamma_2 A_{0,i})$ 
11  if  $\lambda_{0,i} \geq U_{0,i}$  then
12     $Y_{1,i} \leftarrow 0$ 
13  else
14     $Y_{1,i} \leftarrow 1$ 
15  end
16  for  $k$  in  $1, \dots, T$  do
17    if  $Y_{t,i} = 0$  then
18       $\Delta_{t,i} \sim N(\mu_2, \sigma_2^2)$ 
19       $U_{t,i} \leftarrow \min(1, \max(0, U_{t-1,i} + \Delta_{t,i}))$ 
20      if  $t \neq 0 \pmod k$  then
21         $L_{t,i} \leftarrow L_{t-1,i}$ 
22         $A_{t,i} \leftarrow A_{t-1,i}$ 
23      else
24         $\epsilon_{t,i} \sim N(100(U_{t,i} - 2), \sigma^2)$ 
25         $L_{t,i} \leftarrow \max(0, L_{t-1,i} + 150A_{t-k,i}(1 - A_{t-k-1,i}) + \epsilon_{t,i})$ 
26        if  $A_{t-1,i} = 0$  then
27           $A_{t,i} \sim \text{Bern}(\text{expit}(\theta_0 + \theta_1 t + \theta_2(L_{t,i} - 500)))$ 
28        else
29           $A_{t,i} \leftarrow 1$ 
30        end
31        if  $A_{t,i} = 1 \wedge A_{t-k,i} = 0$  then
32           $T^* \leftarrow t$ 
33        end
34      end
35       $\lambda_{t,i} \leftarrow \text{expit}(\gamma_0 + \gamma_1[(1 - A_{t,i})t + A_{t,i}T^*] + \gamma_2 A_{t,i} + \gamma_3 A_{t,i}(t - T^*))$ 
36      if  $1 - \prod_{\tau=0}^t (1 - \lambda_{\tau,i}) \geq U_{0,i}$  then
37         $Y_{t+1,i} = 1$ 
38      else
39         $Y_{t+1,i} = 0$ 
40      end
41    end
42  end
43 end

```

Within the inner loop ($t \in 1, \dots, T$) we see that the data is only updated at time $t \neq 0 \pmod k$, where k refers to evenly spaced check-up visits. If t is not a check-up visit the values of A_t and L_t are the same as in $t - 1$. When t is a visit A_t and L_t are updated.

- if treatment has been commenced then a subject may feel extra benefit if more time has elapsed since treatment began
- L_t affects A_t and also Y_t
- explain starting values for A and Y are all zero (except L maybe)

In order to operationalize the Algorithm 1 we need to choose parameters for (\cdot) . In their paper [Havercroft and Didelez \(2012\)](#) use values that simulate data with a close resemblance to the Swiss HIV Cohort Study. We postpone discussion of the parameters in Algorithm 1 to section 2.4. We just need to state that we follow their parameters because this is not the focus of this thesis.

5.4.2 Discussion of how algorithm works

The algorithm of [Havercroft and Didelez \(2012\)](#) works by factorizing the joint density of the histories of the four variables in the analysis.

- Important is that the form of the MSM is not specified until the last stage
- role of $U_{0,i}$
- How does positivity enter the analysis?
- Why this model is important in terms of positivity.

5.5 Constructing IPT weights

Inverse Probability of Treatment weights can be used to adjust for measured confounding and selection bias in marginal structural models. Link back to pseudo population idea in previous section. This method relies on four assumptions consistency, exchangeability, positivity and no misspecification of the model used to estimate the weights [Cole and Hernán \(2008\)](#). Unstabilized weights are defined as:

$$w_{t,i} = \frac{1}{\prod_{\tau=0}^t p_{\tau}(A_{\tau,i} \mid \bar{A}_{\tau-1,i}, \bar{L}_{\tau,i})}$$

Where the denominator is the probability that the subject received the particular treatment history that they were observed to receive up to time t , given their prior observed treatment and covariate histories (Havercroft, Didelez, 2012). The probabilities $p_{\tau}(A_{\tau,i} \mid \bar{A}_{\tau-1,i}, \bar{L}_{\tau,i})$ may vary greatly between subjects when the covariate history is strongly associated with treatment. In terms of the resulting pseudopopulation, very small values of the unstabilized weights for some subjects would result in a small number of observations dominating the weighted analysis. The result is that the IPTW estimator of the coefficients will have a large variance, and will fail to be normally distributed. This variability can be mitigated by using the following stabilized weights

$$sw_{it} = \frac{\prod_{\tau=0}^t p_{\tau}(A_{\tau,i} \mid \bar{A}_{\tau-1,i})}{\prod_{\tau=0}^t p_{\tau}(A_{\tau,i} \mid \bar{A}_{\tau-1,i}, \bar{L}_{\tau,i})}$$

In the case that there is no confounding the denominator probabilities in the stabilized weights reduce to $p_{\tau}(A_{\tau,i} \mid \bar{A}_{\tau-1,i})$ and $sw_{it} = 1$ so that each subject contributes the same weight. In the case of confounding this will not be the case and the stabilized weight will vary around 1.

In practice, we estimate the weights from the data using a pooled logistic model for the numerator and denominator probabilities. The histories of the treatment and covariates are included in the probabilities. In practice Specifically, following Havercroft and Didelez (2012), we estimate the model where the visit is only the visits every check up time. Between check ups both the treatment and covariate remain the same. Other ways of doing this include a spline function over the months to create a smooth function between the visits. Another difference might be to use a coxph function instead of logistic function

$$\text{logit } p_{\tau}(A_{\tau,i} | \bar{A}_{\tau-1,i}, \bar{L}_{\tau,i}) = \alpha_0 + \alpha_1 k + \alpha_2 a_{k-1} + \cdots + \alpha_k a_0 +$$

We have several options for estimating these weights. We could use a coxph model, or a logistic model.

5.6 Simulation Set-up

We follow the simulation set-up of Havercroft, Didelez (2012) which is based on parameters that closely match the Swiss HIV Cohort Study (HAART).

5.7 Results

- check the distribution of the weights that come out of the model (see Cole 2008). This would allow us to see weight model misspecifications. Not a problem in the simulation case.
- compare the bias, se, MSE, and 95% confidence interval
- compare all of these in the positivity violation and non-positivity violation case.
- explain to some extent monte-carlo standard error.
- we don't confirm the results of the havercroft or Bryan papers, instead refer readers to these papers to see how IPTW outperforms the naive estimators.

6 Discussion and Conclusion

6.1 Limitations

References

- Jenny Bryan, Zhuo Yu, and Mark J. Van Der Laan. Analysis of longitudinal marginal structural models. *Biostatistics*, 2004. ISSN 14654644. doi: 10.1093/biostatistics/kxg041.
- Yvonne W Cheng, Alan Hubbard, Aaron B Caughey, and Ira B Tager. The association between persistent fetal occiput posterior position and perinatal outcomes: an example of propensity score and covariate distance matching. *American journal of epidemiology*, 171(6):656–663, 2010.
- Stephen R. Cole and Constantine E. Frangakis. The consistency statement in causal inference: A definition or an assumption?, 2009. ISSN 10443983.
- Stephen R Cole and Miguel A Hernán. Constructing inverse probability weights for marginal structural models. *American journal of epidemiology*, 168(6):656–664, 2008.

- Stephen R. Cole, Robert W. Platt, Enrique F. Schisterman, Haitao Chu, Daniel Westreich, David Richardson, and Charles Poole. Illustrating bias due to conditioning on a collider. *International Journal of Epidemiology*, 2010. ISSN 03005771. doi: 10.1093/ije/dyp334.
- R. M. Daniel, S. N. Cousens, B. L. De Stavola, M. G. Kenward, and J. A.C. Sterne. Methods for dealing with time-dependent confounding. *Statistics in Medicine*, 2013. ISSN 02776715. doi: 10.1002/sim.5686.
- Sander Greenland, James M Robins, and Judea Pearl. Confounding and collapsibility in causal inference. *Statistical science*, pages 29–46, 1999.
- W G Havercroft and V Didelez. Simulating from marginal structural models with time-dependent confounding. *Statistics in medicine*, 31(30):4190–4206, 2012.
- M A Hernan, B Brumback, and J M Robins. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*, 2000. ISSN 1044-3983. doi: 10.1097/00001648-200009000-00012.
- Miguel A. Hernán. The hazards of hazard ratios, 2010. ISSN 10443983.
- Miguel A Hernán, Sonia Hernández-Díaz, and James M Robins. A structural approach to selection bias. *Epidemiology*, 15(5):615–625, 2004.
- D. G. Horvitz and D. J. Thompson. A Generalization of Sampling Without Replacement from a Finite Universe. *Journal of the American Statistical Association*, 1952. ISSN 1537274X. doi: 10.1080/01621459.1952.10483446.
- Guido W Imbens and Donald B Rubin. Causal Inference. *Boca Raton: Chapman & Hall/CRC, forthcoming.*, (June):39–58, 2008.
- Lynne C Messer, J Michael Oakes, and Susan Mason. Effects of socioeconomic and racial residential segregation on preterm birth: a cautionary tale of structural confounding. *American journal of epidemiology*, 171(6):664–673, 2010.
- J Neyman. On the application of probability theory to agricultural experiments: principles (in Polish with German summary). *Roczniki Nauk Rolniczych*, 10(1):21–51, 1923. ISSN 0883-4237. doi: 10.1214/ss/1177012031. URL <http://projecteuclid.org/euclid.ss/1177010123>.
- Judea Pearl. *Causality*. Cambridge university press, 2009.
- James Robins. A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Mathematical Modelling*, 1986. ISSN 02700255. doi: 10.1016/0270-0255(86)90088-6.
- James M Robins. Marginal structural models versus structural nested models as tools for causal inference. In *Statistical models in epidemiology, the environment, and clinical trials*, pages 95–133. Springer, 2000.
- James M Robins, Miguel Angel Hernan, and Babette Brumback. Marginal structural models and causal inference in epidemiology, 2000.
- Donald B. Rubin. Bayesian Inference for Causal Effects: The Role of Randomization. *The Annals of Statistics*, 1978. ISSN 0090-5364. doi: 10.1214/aos/1176344064.

Daniel Westreich and Stephen R Cole. Invited commentary: positivity in practice. *American journal of epidemiology*, 171(6):674–677, 2010.

Daniel Westreich, Stephen R. Cole, Enrique F. Schisterman, and Robert W. Platt. A simulation study of finite-sample properties of marginal structural Cox proportional hazards models. *Statistics in Medicine*, 2012. ISSN 02776715. doi: 10.1002/sim.5317.

Jessica G. Young and Eric J. Tchetgen Tchetgen. Simulation from a known Cox MSM using standard parametric models for the g-formula. *Statistics in Medicine*, 2014. ISSN 02776715. doi: 10.1002/sim.5994.