Causal modeling in epidemiological practice
van der Wal, W.M.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 1

General Introduction

1.1 An informal introduction to causal effects

The murder of archduke Franz Ferdinand by a Bosnian Serb is often seen as the start of the First World War (Keegan, 1999). One could wonder what might have been the course of events, when this murder would not have happened. All else being equal, e.g. with the same treaties, war plans, ethnic tensions and economic circumstances in place, compare the (possible) course of history in two different situations:

1. Franz Ferdinand was assassinated on 28 June 1914.

2. Franz Ferdinand was not assassinated on 28 June 1914.

Contrasting these two situations is a natural, intuitive process (Kahneman & Miller, 1986). However, the comparison will always be hypothetical, since only situation 1 is observed, and therefore situation 2 is unknown. Nevertheless, the difference in outcome between the two situations could be considered as the causal effect of Franz Ferdinand’s murder. Expert historians could try to estimate what would be the difference in the number of soldiers and civilians being killed as a result of military actions during the next four years. It is clear that such estimates are difficult to produce, and uncertain.

1.2 Causal effects in epidemiological practice

One can debate at length on the metaphysical nature of causality (see e.g. Hume, 1748). At the other extreme, one could try to avoid the subject (as was done by Pearson, 1911). However, in medical practice, we are not concerned with metaphysical questions, only with physical questions. Also, we are clearly interested in determining the causal effect of possible actions on the well-being of patients
(Gillies, 2005). In medicine we are often faced with problems such as having to choose between treatment options for an individual patient. In public health, we are often faced with treatment decisions or behavioral interventions for certain groups, such as whether the elderly should be vaccinated against a certain disease, and the resulting implications for e.g. life expectancy and quality of life relative to the financial costs associated with the treatment. When facing such questions, the counterfactual framework for causal inference provides a theoretical foundation for the identification and estimation of causal effects (Robins, 1999; Höfler, 2005). This framework depends on the concept of counterfactuals or potential outcomes, which are the outcomes that could have been observed, had a certain treatment been given.

### 1.3 Counterfactual framework

Within the counterfactual framework, causal effects are defined as contrasts between different potential outcomes, both at the individual and at the population level (see e.g. Hernán, 2004; Hernán & Robins, 2006a). For instance, with a dichotomous treatment $A$, the potential outcome$^1$ for individual $i$ when receiving treatment level 0 is $Y_{i,a=0}$, and the potential outcome for individual $i$ when receiving treatment level 1 is $Y_{i,a=1}$. The causal effect of $A$ on $Y$ for individual $i$ can then be expressed e.g. as a difference or ratio between $Y_{i,a=0}$ and $Y_{i,a=1}$. At the population level, the effect of a dichotomous treatment $A$ on an outcome $Y$ can be defined as a contrast between the distribution of potential outcomes when everybody would have received treatment level 0, which is $f(y_a=0)$, compared to the distribution of potential outcomes, when everybody would have received treatment level 1, which is $f(y_a=1)$. Depending on the type of outcome, this causal effect could for instance be defined as a mean difference, a risk difference, a risk ratio, odds ratio, hazard ratio, etc. The comparison between counterfactual outcome distributions could also be made nonparametrically, e.g. by comparing Kaplan-Meier estimates of cumulative survival (Cole & Hernán, 2004).

Both individual causal effects and population causal effects cannot be estimated directly, since one can only observe one potential outcome for each unit under study simultaneously. However, population causal effects can be estimated indirectly by using a substitution step (Maldonado & Greenland, 2002). Performing a clinical trial is such a substitution. In a trial$^2$, treatment arms are comparable, in terms of all covariates. In other words, the distribution of covariates is the same in each treatment arm. More importantly, when there is no selection bias, treatment arms are comparable to the whole population. Therefore, the outcomes observed in the different treatment arms are estimates of counterfactuals.

$^1$Note that Pearl’s “DO” operator presents an alternative notation for counterfactuals (Pearl, 2000).

$^2$For now assuming that there is no covariate unbalance between treatment arms due to chance (see section 1.5).
For instance, with a dichotomous outcome $Y$ and dichotomous exposure $A$, under randomisation

$$P(Y_a = 1|A = 1) = P(Y_a = 1|A = 0) = P(Y_a = 1).$$

E.g. this means that the counterfactual outcome (regardless of whether it is observed or not) for exposure level 0, $Y_{a=0}$, is the same for individuals who have actually received exposure level 0 (as indicated by the condition “$A = 0$”), and individuals who have actually received exposure level 1 (as indicated by the condition “$A = 1$”). A similar argument can be made for the counterfactual for exposure level 1, $Y_{a=1}$. More generally, the potential outcome $Y_a$ is independent of the exposure level that was actually observed:

$$Y_a \perp A \quad (1.1)$$

(Hernán & Robins, 2006a). A contrast in outcome between the treatment arms can therefore be interpreted as a causal effect.

The counterfactual definition of causal effects is not limited to medical treatments. For etiological reasons, it could be of interest to estimate the causal effect of the development of a certain disease on a relevant outcome. An example is the causal effect of active tuberculosis on mortality in HIV positive individuals, as studied in chapter 7 of this thesis. The causal effect of interest is then defined as a contrast between the outcome distribution when everybody in the population would suffer from this disease, compared to the outcome distribution when everybody in the population would not suffer from this disease, all else being equal. Such an effect often cannot be estimated using a clinical trial, due to practical limitations and ethical objections. For instance, active tuberculosis cannot be induced directly. Instead, observational data has to be used, in which confounding is often present (see section 1.5).

Furthermore, the counterfactual definition of causal effects is not limited to discrete exposures. With a continuous exposure $A$, the causal effect on $Y$ defined in terms of counterfactuals can be quantified by describing the relationship between outcome $Y$ and exposure level $a$, when all in the population of interest would receive that specific level $a$. This could be done using a marginal structural model, as described below in section 1.4.

### 1.4 Marginal structural models

Using the counterfactual framework, the causal effect of an exposure on an outcome of interest can be described quantitatively using a **marginal structural model** (MSM, see e.g. Hernán et al., 2000; Robins et al., 2000). The term “marginal” refers to the comparison that is made between potential outcomes, regardless of the correlation between the potential outcomes within subjects (Robins et al., 2000). As also pointed out by Robins et al. (2000), the term “structural” refers
to the comparison made between potential outcomes, as is common in the econom- 
metric and social science literature.

As an example, the MSM

\[ E(Y_a) = \beta_0 + \beta_1 a, \] (1.2)

could describe the marginal causal effect of a dichotomous or continuous exposure \( A \) on a continuous outcome \( Y \). The response variable \( Y_a \) is the potential outcome that would have been observed, when every unit would have received the same specific treatment level \( a \) (Robins et al., 2000). Parameter \( \beta_1 \) then quantifies the marginal causal effect of \( A \) on \( Y \). This parameter is equal to the difference in the mean of \( Y \) between either the two treatment levels for a dichotomous \( A \), or for a one unit increase for a continuous \( A \), comparing the fictitious situations where everyone would receive a certain exposure level. The parameters of such a MSM could be estimated by fitting the standard regression model

\[ E(Y) = \beta_0 + \beta_1 A \] (1.3)

on observations \( \{Y_i, A_i\} \), from a randomized trial with no selection bias. MSMs depend on the assumptions listed in section 1.6. The use of MSMs to estimate marginal causal effects is illustrated throughout this thesis, see for instance equations (3.7) and (7.7).

MSMs provide answers to questions such as:

1. What is the ratio of the mortality risks, comparing when everybody in the population would have been treated, versus when nobody in the population would have been treated?

Often, a researcher is interested in a more specific question, such as:

2. What is the ratio in mortality risks, comparing when everybody in a specific age group would have been treated, versus when nobody in a specific age group would have been treated?

Questions such as 2) are conditional on a certain covariate, in this case age group. To quantify such a conditional causal effect, we could fit a conditional marginal structural model (see e.g. VanderWeele, 2009), such as:

\[ \lambda_{T_a}(t|V) = \lambda_0(t) \exp\{\beta_1 a + \beta_2 V + \beta_3 aV\}. \] (1.4)

Model (1.4) is a Cox proportional hazards model, with \( T_a \) the observed death time, depending on whether all subjects would have been treated \( (a = 1) \) or untreated \( (a = 0) \) during the study, and \( V \) an indicator\(^3\) for age group (e.g. \( V = 0: 18-64 \) years old, \( V = 1: \geq 65 \) years old). This model describes the mortality risk both when all subjects would have been treated \( (a = 1) \) and when nobody would have been treated \( (a = 0) \), with separate risks for each age group.

\(^3\)Note that in practice the use of continuous age is often preferable to the use of categorized age.
1.5 Observational data and confounding

The parameters of a model regressing an outcome of interest on an exposure of interest are unbiased estimates of the parameters of a corresponding MSM when fitted on data from a perfect randomized experiment with no selection bias. However, in observational studies confounding is often present. Confounders are covariates that have an effect on the exposure allocation, as well as on the outcome of interest (Pearl, 2000). Unadjusted effect estimators will be biased estimators for the causal effect when such confounding is present (see e.g. Greenland & Morgenstern, 2001).

Adjusting for confounding is possible using various methods, e.g. as listed below in sections 1.7 and 1.8. Importantly, when interpreting any “adjusted” estimate as a causal effect, it is always necessary to make the assumption of no unmeasured confounding (Robins et al., 2000) (see section 1.6). As will be shown in sections 1.7 and 1.8 below, the specific causal structure of the data generating mechanism in an observational study has consequences for the manner in which confounding adjustment can be performed. A useful tool to visualize and examine assumptions regarding the data generating mechanism are directed acyclic graphs (DAGs, see e.g. Robins, 2002; Hernán et al., 2004; Hernán & Robins, 2006b). DAGs illustrate the assumed causal structure between variables, with nodes representing variables, and causal effects depicted by unidirectional edges. The term “acyclic” refers to the fact that in DAGs, no variable can have either a direct or indirect effect on itself. In this thesis, DAGs are used to illustrate Berkson’s bias (Figure 1.2) and to explore the causal structure of the interaction between CD4 count, active tuberculosis and AIDS-related mortality over time after HIV seroconversion (Figure 7.1).

1.6 Assumptions

(Cole & Hernán, 2008) describe the assumptions that are necessary when fitting MSMs:

1. The assumption of consistency states that a subject’s counterfactual outcome under his or her observed exposure level is precisely his or her observed outcome (Robins et al., 1999; Cole & Hernán, 2008). This means that the exposure, the causal effect of which is estimated, needs to be clearly defined. Defining an exposure clearly is often surprisingly difficult, both in observational and in experimental studies (Hernán & Robins, 2011). E.g. in a trial, there is a difference between exposure allocation (which is randomized), and the treatment itself. In an observational study, the effect that is estimated could include unforeseen aspects of the treatment that is received. E.g. a

---

4Note that it is suggested in the literature to adjust for covariates in randomized experiments to increase both efficiency and power (see e.g. Friedman et al., 1998; Moore & Van der Laan, 2007).
treatment decision could be made based on the condition of the patient, and could be explained to the patient as such, which could have an effect on the outcome of interest through certain psychological mechanisms.

2. It is also necessary to assume positivity, i.e. that every level of exposure $A$ has a positive probability of being allocated in every stratum defined by confounders $L$ (Cole & Hernán, 2008; Wang et al., 2006). When a certain level of exposure $A$ cannot be allocated in a certain stratum defined by confounders $L$, the causal effect in that stratum cannot be estimated. When the causal effect cannot be estimated in all strata, the average causal effect cannot be estimated. This assumption is also known as the assumption of experimental treatment assignment (ETA).

3. The assumption of conditional exchangeability means that within strata defined by measured covariates $L$, potential outcomes $Y_a$ are independent of the observed exposure level $A$:

\[ Y_a \perp A | L \]  

(Hernán & Robins, 2006a). In practice, this only holds when there are no unmeasured confounders. Assessment of the validity of this assumption can only be done based on the available medical and biological knowledge. However, it is possible to conduct a sensitivity analysis to quantify the impact of possible unmeasured confounding on effect estimates (Arah et al., 2008; Groenwold, 2009; Klungsoyr et al., 2009; Chiba, 2010).

Note than when correcting for measured confounders, it is thereby also assumed that the specific form of correction fully adjusts for the confounding, and that the correction method itself does not introduce bias (as can be the case with standard conditioning, see section 1.7).

### 1.7 Adjusting for confounding: conditioning

A widely used method to correct effect estimates for confounding is **conditioning**. Conditioning amounts to the pooling of associations, estimated within strata defined by confounders. In this manner an overall adjusted effect estimate is obtained. Such a pooled estimate can be obtained either by using stratification methods, or by including confounders as covariates (which could either be discrete or continuous) in a regression model (see e.g. Fitzmaurice, 2004; Ranstam, 2008). As shown in sections 1.7.1, 1.7.2 and 1.7.3, conditioning has some important drawbacks.

#### 1.7.1 Adjusting away the effect

An important drawback of conditioning is that effects can be adjusted away. Suppose that the effect of a time-varying exposure $A$ on an outcome $Y$ is confounded
by time-varying covariate $L$. When $A$ has a causal effect on $L$ at a later time point, $A$ has an indirect effect on $Y$ through $L$. This is illustrated in Figure 7.1 in chapter 7 with active tuberculosis as exposure, CD4 count as intermediate confounder and mortality the outcome of interest. When “keeping constant” $L$ by including it as a covariate in a regression model, the indirect effect of $A$ on $Y$ through $L$ is adjusted away by conditioning on CD4 (Robins, 1997; Robins et al., 1999). The resulting effect estimate includes only the direct effect$^5$ of $A$ on $Y$. In general, the problem of adjusting away the effect will occur when conditioning on any variable that is also intermediate for the effect of the exposure.

### 1.7.2 Non-collapsibility

Conditional effects (e.g. an effect within strata defined by covariates, pooled over the strata), are only equal$^6$ to marginal effects (e.g. a contrast between the fictitious situations “everybody treated”, versus “nobody treated”) on the linear or loglinear scale (Greenland et al., 1999). Therefore, effect estimates from conditional models, e.g. regression models including other covariates in addition to the exposure variable, are only true estimates of marginal effects with a model that utilizes a linear or loglinear link function.

We will illustrate this phenomenon below. Let the population distributions of dichotomous point treatment exposure $A$, dichotomous confounder $L$ and dichotomous outcome $Y$ be parameterized as follows:

\[
\logit P(L = 1) = 0, \tag{1.6}
\]

\[
\logit P(A = 1) = -1 + 2L, \tag{1.7}
\]

\[
\logit P(Y = 1) = -1 + A + \beta L. \tag{1.8}
\]

The conditional effect of exposure $A$ on $Y$ on the parameter scale is

\[
\sum_l \left\{ \ln \left[ \frac{e^{-1+1+\beta l}}{e^{-1+0+\beta l}} \right] P(L = l) \right\} = \sum_l \{1 \times P(L = l)\} = 1,
\]

and therefore the oddsratio corresponding to the conditional effect of $A$ on $Y$ is

\[
\Rightarrow \text{conditional oddsratio for effect } A \text{ on } Y = \exp(1). \tag{1.9}
\]

The oddsratio corresponding to the marginal effect of $A$ on $Y$ can be computed as follows:

\[
\logit P(Y_a = 1|L) = -1 + a + \beta L
\]

---

$^5$Note that this resulting effect estimate could differ from the marginal direct effect of $A$ due to noncollapsibility, see section 1.7.2.

$^6$Note that this section concerns “true effects” i.e. as in the population, and without confounding or selection bias.
\[ P(Y_a = 1) = \sum_l \left\{ \frac{e^{-1+a+\beta L}}{1 + e^{-1+a+\beta L}} P(L = l) \right\} = f_1(a, \beta) \]

\[ \Rightarrow \text{marginal oddsratio for effect } A \text{ on } Y = \frac{f_1(1, \beta)/(1 - f_1(1, \beta))}{f_1(0, \beta)/(1 - f_1(0, \beta))}, \tag{1.10} \]

which is in general unequal to (1.9), i.e. the marginal effect is unequal to the conditional effect. The same holds true on the parameter scale\(^7\), i.e. the natural logarithms of (1.9) and (1.10) are in general unequal. The relationship between the effect of \( L \) on \( Y \) as quantified by \( \beta \), (1.9) and (1.10), is plotted in figure 1.1. In this example, it is clear that the conditional and marginal effect of \( A \) on \( Y \), as quantified by the oddsratio or the natural logarithm of the oddsratio, are only equal when \( \beta = 0 \), i.e. when covariate \( L \) does not have an effect on \( Y \).

![Figure 1.1: Oddsratios corresponding to both the conditional and the marginal causal effect of exposure \( A \) on outcome \( Y \) plotted against \( \beta \), the conditional effect of covariate \( L \) on \( Y \), from the example explained in section 1.7.2.](image)

\(^7\)Note that for (1.10) by “parameter” we refer to the parameter of the MSM quantifying the marginal causal effect of \( A \) on \( Y \), which is equal to the logarithm of (1.10), while for (1.9) we refer by “parameter” to the parameter corresponding to the conditional effect of \( A \) on \( Y \), which is equal to 1.
1.7.3 Collider stratification

When using conditioning to adjust for confounding, a form of selection bias can be induced. As an illustration, suppose that time-varying exposure $A$ can change and is measured at two time points (see Figure 1.2). We indicate the value of $A$ at the first and second time point with $A_0$ and $A_1$, respectively. Furthermore, covariate $L$ is intermediate for the effect of $A_0$ on $A_1$, and $U$ is a (either measured or unmeasured) variable that affects both $L$ and the outcome $Y$. When conditioning on covariate $L$ (as indicated by the box in Figure 1.2), an association is induced between $A_0$ and $U$ (as indicated by the dotted line), as explained in Hernán et al. (2004). The phenomenon that an association is introduced between two variables by conditioning on a common effect, was first described by Berkson (1946), and is known as collider stratification (see e.g. Greenland, 2003; Whitcomb et al., 2009). For instance, Berkson pointed out that diseases that are unassociated in the general population could be associated among hospitalized patients, when both diseases affect the probability of hospital admission. The induced association between $A_0$ and $U$ also causes an association between $A_0$ and $Y$. When there was no association between $A_0$ and $Y$, an association is introduced between $A_0$ and $Y$ by conditioning on $L$. When there was an association between $A_0$ and $Y$, the association is changed by conditioning on $L$. Therefore, an estimate of the causal effect of $A$ on $Y$, from a regression model that includes $L$ as a covariate, or by stratification on $L$, could be biased. Note that this problem could also occur with an $L$ that is intermediate for the effect of $A$ (as illustrated in section 1.7.1). Different forms of selection bias due to collider stratification, such as informative censoring, are described in Hernán et al. (2004).

![Figure 1.2: Illustration of collider stratification, as explained in section 1.7.3.](image)

1.8 Fitting marginal structural models

In sections 1.7.1, 1.7.2 and 1.7.3 it is shown that when using conditioning to adjust for confounding, the resulting estimators are not in general unbiased estimators of marginal causal effects. The parameters of MSMs can be estimated using recently developed methods to correct for confounding, as well as for forms of selection bias such as informative censoring, that do not share the drawbacks of conditioning. These methods are described below in sections 1.8.1, 1.8.2 and 1.8.3.
1.8.1 Inverse probability weighting

As was shown by Robins (1998), the parameters of MSMs can be estimated using inverse probability weighting (IPW) to correct for confounding. This amounts to the fitting of a model regressing the outcome of interest on the exposure of interest using observational data, with each observation weighted by the inverse of the probability of the observed exposure level given the observed value of the confounders.

**IPW in a point treatment**

Before turning to a more general setting, let us first illustrate IPW in a point treatment setting, i.e. a cross-sectional study with an exposure, outcome and possible confounders measured at one time point. Using IPW, we can adjust for a set of confounders $L$ when estimating the effect of discrete exposure $A$ by weighting each observation $i$ by the inverse probability weight

$$w_i = \frac{1}{P(A_i = a_i | L_i = l_i)}.$$  \hspace{1cm} (1.11)

We indicate the observed exposure and confounder status with $a$ and $l$, respectively. The denominator of (1.11) contains the probability of the observed exposure level given the observed values of covariates $L$. Weighting by $w_i$ creates a pseudopopulation in which $L$ no longer predicts $A$. Weighting observations $i$ by $w_i$, one can then unbiasedly fit a causal model, for instance the MSM

$$E(Y_a) = \beta_0 + \beta_1 a,$$  \hspace{1cm} (1.12)

with a continuous outcome $Y$. The response variable $Y_a$ is the potential outcome that would have been observed in all units under study, when every unit would have received the same specific treatment level $a$ (Robins et al., 2000). As in (1.2), parameter $\beta_1$ then quantifies the causal effect of $A$ on $Y$.

To increase statistical efficiency and attain better coverage of confidence intervals, it is recommended to use stabilized weights (Hernán et al., 2000; Cole & Hernán, 2008), e.g.

$$sw_i = \frac{P(A_i = a_i)}{P(A_i = a_i | L_i = l_i)}.$$  \hspace{1cm} (1.13)

The numerator of (1.13) contains the probability of the observed exposure level, which can be estimated by the observed frequency. To further stabilize the weights, one can condition both in the numerator and in the denominator of (1.13) on a set of time-fixed covariates $V$, that are not confounders for the effect of $A$ on $Y$, but that are related to $A$. For instance, when a researcher believes that sex does not influence the outcome of interest, but the distribution of exposure level varies between both sexes, sex could be included in $V$. 
IPW in a longitudinal study

IPW is especially useful with longitudinal data, when effects could be adjusted away and bias due to collider stratification could be induced by using conditioning. Suppose that a discrete exposure $A$ may change over time, and the decision to allocate a certain exposure level is made and recorded within each unit $i$ at time points $t_{ij}$. Time-varying confounders for the effect of $A_{ij}$ on the outcome of interest, measured right before each time point $t_{ij}$ in each unit $i$, are contained in $L_{ij}$. In addition, $L_{ij}$ can also contain time-fixed confounders. Let $\overline{A}_{ij}$ and $\overline{L}_{ij}$ indicate the observed longitudinal history, i.e. all measurements up to time point $t_{ij}$ within unit $i$, of $A$ and $L$ respectively. $V_i$ are measured time-fixed covariates, that are not confounders but that are associated with the exposure. One can adjust for time-varying confounders $L$ by weighting observations at $t_{ij}$ by the stabilized weights

$$sw_{ij} = \prod_{k=0}^{j} \frac{P(A_{ik} = a_{ik}|\overline{A}_{ik-1} = \overline{a}_{ik-1}, V_i = v_i)}{P(A_{ik} = a_{ik}|\overline{A}_{ik-1} = \overline{a}_{ik-1}, \overline{L}_{ik} = \overline{l}_{ik}, V_i = v_i)}.$$  \hspace{1cm} (1.14)

Equation (1.14) is a product over all time points from baseline up to time point $t_{ij}$, within each unit $i$. The factors in the numerator of (1.14) contain the probability of the observed exposure status at each time point, $a_{ik}$, given the observed exposure history up to the previous time point, $\overline{a}_{ik-1}$, and the observed time-fixed covariates, $v_i$. The factors in the denominator of (1.14) contain the probability of the observed exposure status at each time point, given the observed exposure history up to the previous time point, the observed history of time-varying confounders up to each time point, $\overline{l}_{ik}$, and the observed time-fixed covariates. Note that time-fixed covariates $V$ are included both in the numerator and denominator of (1.14) to stabilize the weights. To estimate the causal effect of $A$ on the exposure of interest, one can estimate the parameters of a MSM using the observations made at time points $t_{ij}$, weighted by $sw_{ij}$, as was done e.g. by Hernán et al. (2000).

Notes

The numerators and denominators of the inverse probability weights should be estimated using correctly specified exposure allocation models. For instance, the denominators of (1.11) and (1.13) can be estimated by using exposure allocation models regressing $A$ on $L$. Similarly, the numerator of (1.13) can be estimated by using exposure allocation models regressing $A$ on the constant only. The elements in the denominator of (1.14) can be estimated by using an exposure allocation model regressing time-varying exposure $A_{ij}$ on follow-up time $t_{ij}$, the history of $A$ up to but not including $t_{ij}$, $\overline{A}_{i(j-1)}$, the observed history of confounders $L_{ij}$ up to and including $t_{ij}$, $\overline{L}_{ij}$, and the time-fixed covariates $V_i$. The elements in the numerator of (1.14) can be estimated from a similar model, not including $\overline{L}$. 

When the assumptions listed in section 1.6 are met, and the exposure allocation models are correctly specified, i.e. equal to the population allocation mechanism, IPW weighted estimates are asymptotically unbiased (Haight et al., 2003). However, IPW estimators are sensitive to practical violations of the ETA assumptions, leading to small sample bias (Neugebauer & Van der Laan, 2005; Robins et al., 2000; Yu & Van der Laan, 2004), as explained in detail in chapter 8.

Some additional sidenotes are that 1) IPW in both a point treatment and a longitudinal setting can be generalized to a continuous exposure variable, as explained in chapters 2 and 3, 2) When using IPW, observations can have weights unequal to each other, which introduces clustering in the data. When this is not taken into account, the standard error of the causal effect estimate could be underestimated. Therefore, when using IPW to estimate the parameters of a MSM, it is necessary to use a robust standard error estimator for inference (Hernán et al., 2000), 3) A correction for forms of selection bias such as informative censoring can be made using the same IPW techniques (see e.g. Hernán et al., 2000, 2002; Cole & Hernán, 2004) and 4) different IPW corrections, e.g. both for confounding and informative censoring, can be easily combined by multiplying the weights, as we have done e.g. in chapter 3.

The use of IPW to estimate the parameters of MSMs is illustrated throughout this thesis, for instance in chapters 3, 4 and 5. We explore finite sample behavior and the efficiency of IPW compared to standard conditional models and an improved iterative IPW method in chapter 8.

1.8.2 G-computation

An alternative to IPW to estimate the parameters of MSMs is G-computation (Robins, 1997). To illustrate the basic principle of G-computation, consider first the case of a point treatment study. Within an individual $i$ we let $Y_{i,a}$ indicate all potential outcomes, with exposure $A$ set to all possible levels $a$. The distribution of potential outcomes over all individuals, with $A$ set to all possible levels $a$, is indicated as

$$f(y_a).$$

(1.15)

If this distribution were known, we could evaluate the causal effect of $A$ on $Y$ by applying a chosen contrast to it, such as the risk ratio. However, we only observe the distribution of outcomes for the exposures that are actually received by the individuals in the sample. Evaluating the causal effect of $A$ on $Y$ using this latter distribution, the observed data, would give a biased estimate when confounding is present.

To evaluate the causal effect of $A$ on $Y$ in the presence of a confounder $L$, we can use the following approach (Robins, 1997). The first step is to factorize (1.15)

---

8Note that in chapter 7 we use an alternative notation, using Pearl’s “DO” operator (Pearl, 2000).
over the distribution of possible values \(l\) of \(L\) as

\[
f(y_a) = \int_l f(y_a | L = l) f(l) dl. \tag{1.16}
\]

If \(L\) is discrete, (1.16) can be thought of as the weighted sum of distributions of potential outcomes for all exposure levels, given specific observed values \(l\) of \(L\) - weighted by the frequencies with which those specific values \(l\) occur. In other words, within subgroups of individuals who have the same value of \(L\), there are distributions of potential outcomes, and those distributions are summed together, weighted by the sizes of those subgroups. This practice is known in epidemiology as standardization (see e.g. Clayton & Hills (1993), chapter 14). Note that (1.16) is still unobservable, since individuals only receive one specific exposure level.

To rewrite (1.16) into an observable quantity, we need to assume consistency and no unmeasured confounding (see section 1.6). Under these two assumptions (1.16) is equal to

\[
\int_l f(y | A = a, L = l) f(l) dl, \tag{1.17}
\]

which, if \(L\) is discrete, can be thought of as the weighted sum of distributions of the outcomes for observed exposure levels, given specific observed values \(l\) of \(L\) - weighted by the frequencies with which those specific values \(l\) occur. In other words, within subgroups of individuals who have the same value of \(L\), there are distributions of outcomes for all observed exposure levels, and those distributions are summed together, weighted by the sizes of those subgroups.

Only in simple cases can (1.17) be computed directly, to obtain an estimate of (1.15), allowing for the direct calculation of some contrast applied to (1.15). However, we can approximate (1.15) as follows. We regress the outcome \(Y\) on both exposure \(A\) and confounder \(L\), fitting a “data model” \(Y \sim (A, L)\) on the observed data. Assuming that this data model correctly describes the distribution \(f(y | A = a, L = l)\), and can be fitted using the available data, we can then use it to approximate (1.16) in the following manner. Using the data model, we predict potential outcomes \(y_{i,a} | L = l_i\) for each individual \(i\) under all possible \(a\), with \(l_i\) his or her observed value of \(L\). The distribution of \(L\) in the sample is assumed to be representative\(^9\) of the population distribution of \(L\). In this manner, the distribution of potential outcomes \(f(y_a)\) is approximated. We can then fit a regression model, estimating the parameters of a MSM, using this distribution.

Unlike IPW, G-computation is known to be robust to practical violations of the ETA assumption (see chapter 8) (Neugebauer & Van der Laan, 2005; Yu & Van der Laan, 2004), since, in principle, prediction of the outcome for a specific exposure level can be done for observations with values of \(L\) for which this exposure level was never given in the sample. When the exposure \(A\) is a continuous variable, values

\(^9\)Note that this assumption is always questionable, both in observational and experimental studies. For instance, when doing clinical trials, it has to be assumed that those who participate are representative of the population of interest.
a can be randomly sampled. To improve precision, the sampling of a followed by the prediction of y, and the fitting of a MSM can then be done multiple times, similar to multiple imputation (see e.g. Carpenter et al. (2006)). Inference can be done using the bootstrap.

The implementation of G-computation is challenging in a longitudinal setting. We show that G-computation in a longitudinal setting can be simplified by using individual random effects as latent baseline variables, instead of time-varying founders, in chapter 7. This algorithm can be applied when such baseline variables (e.g. an intercept and slope) fully describe the “true” value of the time-varying confounders, separate from short term fluctuations and measurement error.

1.8.3 Alternative methods to fit MSMs

Doubly robust (DR) estimation (Yu & Van der Laan, 2004) combines both IPW and G-computation in an estimating equation, which can be solved e.g. through the Newton-Raphson algorithm. The resulting estimator is consistent when either the treatment allocation mechanism (used in IPW) or the data model (used in G-computation) is consistently estimated (Van der Laan & Robins, 2002; Yu & Van der Laan, 2004). DR-estimators can be adapted to be consistent when the ETA assumption is practically violated (Neugebauer & Van der Laan, 2005).

Targeted maximum likelihood estimation (TMLE) (Van der Laan, 2010; Gruber & Van der Laan, 2009) uses so called “super-learning” to estimate the data model (see section 1.8.2) used by G-computation and then uses an “optimal fluctuation function” that is targeted to the causal parameter of interest to update the initial estimate. Similarly to DR estimation, TMLE is doubly robust since it only needs either a consistent initial estimate or a correctly specified optimal fluctuation function. Also, TMLE has been shown to be relatively efficient (Van der Laan & Gruber, 2009). Depending on the amount of information present in the data, the exposure allocation model may or may not need to be estimated. In “collaborative TMLE” (CTLME), a selection is made between different forms of the exposure allocation model. In simulation studies, CTMLE has shown robustness against practical ETA assumption violations (Van der Laan & Gruber, 2009).

DR estimation and TMLE are not used throughout this thesis. These methods are relatively complex and therefore more difficult to implement in standard software (see section 1.8.4). However, their double robustness, the insensitivity to practical non-positivity of specific forms of these methods, and the efficiency of TMLE, make these methods very promising alternatives.

1.8.4 Implementation

The implementation of IPW is relatively easy using standard statistical software. Even in SPSS, weights can be estimated from regression models using a few lines of code. The parameters of MSMs can then be estimated by fitting a regression
on the weighted data. Note however that the use of robust standard errors or bootstrapping is not straightforward in SPSS. Hernán et al. (2000) described how to program IPW in SAS, including inference. Fewell et al. (2004) described how to program IPW in Stata, including inference.

Available software packages to fit MSMs includes CausalGAM, an R package for the estimation of causal effects with generalized additive models in a point treatment with a binary exposure (see http://CRAN.R-project.org/package=CausalGAM), cvDSA, an R package for MSM-based causal inference with point treatment data using data-adaptive estimation with cross-validation and the deletion/substitution/addition (D/S/A) algorithm (see http://www.stat.berkeley.edu/~laan/Software/index.html), tmleLite, an R package for targeted maximum likelihood estimation (Van der Laan, 2010) of marginal additive treatment effect of a binary point treatment (see http://www.stat.berkeley.edu/~laan/Software/index.html) and the SAS macro for doubly robust estimation by Jonsson Funk et al. (2007).

In addition, we have developed the R package ipw (see chapter 2) for estimating inverse probability in a wide range of settings.

1.9 Overview

The outline of the rest of this thesis is as follows. Chapter 2 concerns the implementation of MSM methodology, to increase accessibility to applied researchers. We describe our R package ipw, for estimating inverse probability weights. The package ipw can be used with data from a point treatment situation as well as with a time-varying exposure and time-varying confounders. It can be used with binomial, categorical, ordinal and continuous exposure variables. We illustrate how to use the package to fit MSMs, including inference.

Chapter 3–6 concern the application of MSM methodology to epidemiological studies. In chapters 3 and 4, we use IPW to fit the parameters of a MSM that quantifies the causal effect of renal dialysis type or residual glomerular filtration rate on mortality in end stage renal patients, respectively. In chapter 5 we use IPW to fit the parameters of a MSM to quantify the causal effect of AIDS defining conditions and the time-varying use of highly active antiretroviral therapy on mortality, including their interaction, over calendar time. We contrast this with another possible approach, using a so called history-adjusted marginal structural model (HA- MSM, Van der Laan et al., 2005; Petersen et al., 2007). In chapter 6, we do not estimate a causal effect, but rather examine the course of quality of life over time in end-stage COPD patients qualitatively. However, we illustrate that inverse probability of censoring weighting (IPCW, see section 3.4.4) to correct for informative censoring, is not always necessary in observational studies. Instead, standard conditioning could be used.

In chapter 7 we present an algorithm for estimating the parameters of a MSM through G-computation to quantify the causal effect of a secondary illness on
the progression of a chronic disease. This algorithm is then used to estimate the effect of active tuberculosis on mortality in HIV positive injecting drug users. In *chapter 8* we present an iterative method through which IPW weights can be estimated, resulting in causal effect estimators that suffer substantially less from small sample bias when the ETA assumption is practically violated, as compared to standard IPW.

*Chapter 9* gives a general discussion of this thesis, as well as perspectives for future research.