Causal Reasoning and Inference in Epidemiology

Vanessa Didelez

Chapter in the upcoming 3rd edition of Handbook of Epidemiology (eds. Wolfgang Ahrens & Iris Pigeot), Springer New York

Key words: causal discovery; causal mediation; directed acyclic graphs; g-formula; instrumental variables; potential outcomes; target trial emulation; propensity score; treatment-confounding feedback
Abstract The chapter gives an introduction to basic concepts of causal reasoning and to key methods for causal inference as developed and used in epidemiology over the last two decades. It begins with some considerations about formulating causal questions, e.g., using the idea of a target trial, and their mathematical formalization via potential outcomes. We then address typical assumptions and methods for estimating causal effects under a variety of causal models. The methods include estimating point and time-dependent treatment effects, e.g., via inverse probability of treatment weighting, or using natural experiments such as instrumental variable estimation. While confounding bias is always a concern when using observational data, we discuss how the principle of target trial emulation offers some protection against other, often self-inflicted biases such as immortal time bias. The chapter concludes with the special topics of causal mediation analysis and causal discovery, which both aim at investigating direct and indirect causal pathways.

1 Introduction

A core aim of epidemiological research is to investigate how to retain or improve health, e.g., by assessing the potential benefits of behavioral changes, the impact of new health policies, and the effectiveness (or harm) of preventive measures, new treatments, or modified treatment strategies. Such questions go beyond modeling of how things currently are, for which we use 'associations', but instead address how things would be (or would have been) under different options for action, changes, or interventions — they are therefore 'causal' questions. It is also typical of epidemiology to rely on observational (i.e., mostly non-interventional or non-experimental) data, such as cohort studies, or routinely collected data, such as electronic health records (EHR). The field of causal inference provides a methodological framework for defining and quantifying causal effects, under specific assumptions, with the ultimate aim of obtaining actionable results, such as recommendations or guidelines for decision makers, e.g., individuals, doctors, or health authorities.

Two key elements characterize the field of causal inference: (a) An explicit and formal definition of the causal target of inference (often decoupled from parameters of statistical models and measures of association); and (b) a formal statement (and scrutiny) of the assumptions under which this causal target is identifiable from observable data. For the statistical analysis, it turns out that standard methods often require rather restrictive assumptions or are not sufficiently flexible. Hence, these two key elements are usually followed by a third element: (c) A wide range of specialized statistical methods for causal inference that are tailored to the specific causal targets, their assumptions, and critical evaluation. However, the specialized statistical methodology alone does not make a causal analysis; without (a) and (b) it will be difficult for a causal analysis to be convincing.

The literature provides some guidance on how to bring the above points (a)–(c) to life in any given application: Petersen and van der Laan (2014) propose a *causal roadmap* beginning with the data generating causal model and ending with a possible causal interpretation of the statistical findings (for a recent synthesis see Dang et al (2023)); Hernán and Robins (2016) stress the importance of framing and designing an observational study so as to *emulate a target trial*; Goetghebeur et al (2020) list eight steps of *formulating causal questions* to obtain *principled statistical answers*; Young (2024) emphasizes the importance of a motivating

Vanessa Didelez
Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, e-mail: didelez@leibniz-bips.de

causal story. Inspired by the above, the present chapter gives an introduction to causal reasoning and causal inference in epidemiology. Figure 1 summarizes the key aspects of a causal analysis that will be discussed. We focus on the main ideas and principles with no claim to mathematical detail or completeness, for which we refer to the original literature and key textbooks (Pearl, 2009; Rosenbaum, 2010; Hernán and Robins, 2020).

Summary. Section 2 discusses aspects relevant to formulating causal research questions in terms of hypothetical interventions, and more formally using potential outcomes. In Section 3, causal models using directed acyclic graphs (DAGs) are briefly introduced (see also Chapter on Causal Directed Acyclic Graphs by R. Foraita et al in this book) which are helpful to justify typical assumptions, e.g., to rule out unmeasured confounding. Bringing causal questions, models and assumptions together, Section 4 presents the main principles of statistical estimation of causal effects. Section 5 addresses important issues of designing observational studies in a way to avoid common sources of bias beyond confounding. The special topic of causal mediation analysis, investigating direct and indirect causal effects, is covered in Section 6, while the exploratory tool of causal discovery, i.e., data-driven approaches for finding causal structures, is the subject of Section 7.

2 Formulating Causal Research Questions

It is a matter of debate whether or not it is possible to conduct valid (or even useful) causal inference when premised on an ill-defined causal question (Greenland, 2017; Hernán, 2018a; Young, 2024). At the least, it seems difficult to speak meaningfully about bias (cf. Section 3.2, and the Chapter Sensitivity Analysis and Bias Analysis by S. Greenland in this book) if the target of inference is not well-defined. We therefore begin with some points of consideration when formulating causal research questions. Even though typical of epidemiological research, it is not self-evident how to specify a causal question (Fox et al, 2019; Hernán et al, 2019; Young, 2024) and some journals still discourage declaring a research question as being 'causal' (Hernán, 2018a). However, in line with many voices in causal inference, we posit here that in order to obtain a causal answer we must begin with an explicit causal question, not only for general transparency but also as a crucial prerequisite for determining an appropriate statistical analysis and for a critical appraisal of the findings.

To elicit and clarify the causal nature of a research question, as opposed to descriptive or purely predictive questions (Hernán et al, 2019), one can follow a number of strategies depending on the field of application. It can help to clearly define the decision problem(s) one aims to inform (Dawid, 2021). For example, in the context of a suspected teratogenic medication, we may ask what actionable advice women might seek who are planning to become pregnant; this may be different for those who already are pregnant. It can also be helpful to discuss what, possibly hypothetical, intervention one might introduce. For example, parents could be given advice by local kindergarden teams on bed-time routines so as to increase the sleep duration of their children. Such interventions are *hypothetical* because, typically, they were not carried out in the observational data at hand. To quote Hernán and Robins (2020), causal inference addresses "What if...?" questions[1].

It is furthermore good practice to specify the research question by formulating a hypothetical trial that would directly inform our decision problem or choice of interventions, even if this trial is not practically feasible for ethical, financial or other reasons (Didelez, 2016). This concept of a *target trial* is introduced in Section 2.1; the practical emulation of a target trial with available data is discussed later in Section 5. Much of the methodological literature on causal inference has focused on the mathematical formulation of different types of causal effects and their subsequent statistical estimation. We address the former in Section 2.3 after introducing formal notation and terminology in Section 2.2; methods of estimation are deferred to Section 4.

2.1 Decision Problems and Target Trials

It is widely accepted that randomization, and thus randomized controlled trials (RCTs), deliver causally interpretable results. However, RCTs have a number of further characteristics, in addition to randomization, that facilitate a clear and meaningful interpretation. Following the design principles of an RCT even when using observational data has been shown to resolve seemingly conflicting findings, for instance, comparing

[1] While Pearl and Mackenzie (2018) view causality as being about "Why?" questions.

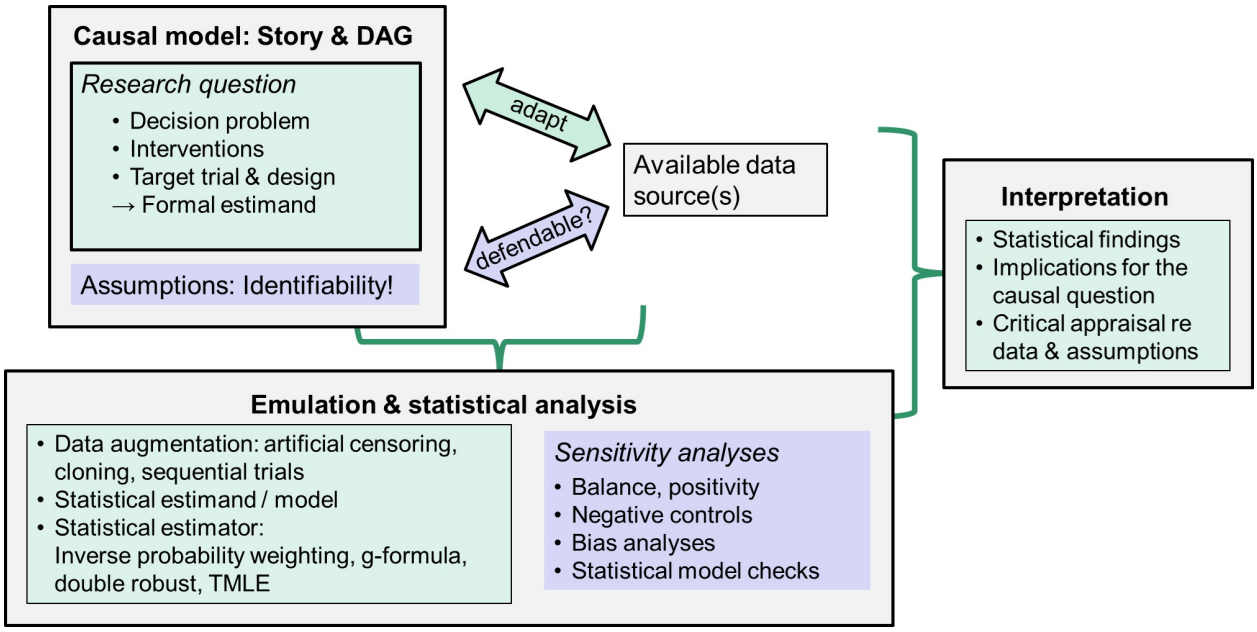

Fig. 1 Overview of key aspects and steps in a causal analysis.

earlier observational studies versus later RCTs on adverse effects of hormone replacement therapy (Hernán et al, 2008). Hernán and Robins (2016) list the following key components to be specified for a target trial: Eligibility criteria, treatment strategies to be compared, assignment to treatment strategies, length of follow-up, outcome of interest, causal contrast and analysis plan. The thought process required to specify all these components for an observational study prompts the researcher to borrow other strengths from RCTs, apart from randomization.

What to compare. One key component is the explicit formulation of the trial arms which should reflect the decision options or hypothetical interventions to be compared. For instance, in the above sleep example, we may determine that we want to compare 'no intervention' (known as 'natural course') with a hypothetical intervention imposing a bed-time routine that increases current sleep duration by 10%. In other examples, we may suspect that a certain medication causes severe adverse events, but it may be challenging to define a suitable comparison group, particularly when a 'no treatment' group is not practical or relevant (e.g., among groups of very ill patients that are never left untreated and therefore always take an active medication). It is also part of the trial arm definition to state whether we are interested in 'point treatments', i.e., one-off treatments, or whether these are to be sustained over a given period of time. They could even be adaptive, i.e., functions of the individual's development, like increasing or decreasing the dosage in response to the patient's progress. To complete the definition of treatment arms, we need the definition of the outcome, which deserves careful thought so as to be meaningful and well-defined under various circumstances. For example, when interested in the risk of malformation as a pregnancy outcome, we have to consider that it may not always be possible or meaningful for every pregnancy to determine whether there was a malformation or not, e.g., when there was an early abortion or miscarriage.

Alignment at time-zero. A further important aspect of a target trial is the alignment of certain components of the trial at baseline or *time-zero*. The elements to be aligned are (i) the inclusion and exclusion criteria, which must be assessed before (ii) the assignment to trial arms, and finally (iii) the start of follow-up with the moment of randomization. This sequence is natural in RCTs but not in observational studies, e.g., when using EHRs there is often no unique time-zero. Importantly, the alignment facilitates the clear formulation of a research question in the context of time-dependent treatments. It has been illustrated in various settings that a failure to align these elements in the design of observational studies can lead to considerable bias (Hernán et al, 2016; García-Albéniz et al, 2017; Dickerman et al, 2019; Braitmaier et al, 2024).

Many treatments in epidemiology are long-term behaviors or exposures, e.g., 'smoking', so that it is tempting to summarize them into a single binary indicator, e.g., 'current smoker' or 'former smoker'. However, a treatment definition such as 'former smoker' is not well-defined. It would be infeasible to design a trial where participants are randomized to be 'former smokers' as we would need to randomize the past. Thus, it violates

the alignment and lumps together possibly very long periods of time ignoring that, in this example, those who started smoking and remained smokers may be very different from those who started and soon stopped again. To obtain alignment and ask a more precise research question we might formulate a time-window during which smoking should start, e.g., adolescence, and then consider sustained or discontinued smoking separately, while 'never smoking' or 'later start of smoking' may serve as control arms.

A causal story. Formulating a target trial to make the research aim transparent has been adopted especially in pharmacoepidemiology, a field that lends itself to this framework due to its typical research questions and data sources such as EHRs or health claims data. In other fields the target trial principle has sometimes been regarded as limiting the research questions that can be asked (Pearce and Vandenbroucke, 2023). However, target trial examples have been developed for a large variety of questions and designs including, for example, dietary studies (Chiu et al, 2021). Experience shows that the ensuing clarity of the analysis much improves scientific practice (De Stavola et al, 2023; Didelez et al, 2024). Still, a target trial has to be preceded by a 'causal story' that generates the research question. This story may be quite distant from what has actually been measured in the available data which, in turn, may result in a perceived difficulty of mapping the question to a target trial. For instance, common exposures in epidemiology include biomarkers, air pollution, body mass index (BMI) or birth weight – none of these can directly be randomized in a real trial. But there are ways to target and modify them indirectly via interventions, e.g., traffic rules, physical activity programs in school, or support during pregnancy, and the causal effects of such interventions are of much relevance to health outcomes. In the sleep example, there is no practical intervention that 'fixes' sleep duration at a given value, but the causal story could be extended to a hypothetical intervention imposing a bed-time routine that increases sleep duration by 10%. Other situations are even more complicated: For example, to explain the inverse association between cancer and dementia, Rojas-Saunero et al (2024) suggest constructing a target trial with a hypothetical drug that increases Pin1 expression (a protein involved in the cell cycle) specifically in the brain tissue even though such a drug does not (yet) exist and Pin1 expression was not measured in the available data. To quote Greenland (2017), 'practical concerns should prod us to extend our analysis to potentially modifiable components of the initial study factors, for which potential outcomes can be made precise'. Explicitly addressing the gap between available data and causal research questions will guide sensitivity analyses and improve data collection in future studies (Young, 2024).

Bibliographic notes: Further discussion of the issue around well-defined causal research questions with examples can be found in Hernán and Taubman (2008) or Hernán (2016). There is also a debate on the difficulty of specifying causal questions in the field of social epidemiology, with exposures such as socioeconomic status or race (Kaufman, 2019; Galea and Hernán, 2019; Jackson and Arah, 2019). Ramspek et al (2021) provide a review on the common practice of not clearly distinguishing between prediction and causality (see also van Geloven et al (2020)). The popular but ambiguous notion of 'risk factor', which may or may not be intended to have a causal meaning, is dissected by Huitfeldt (2016). Practical difficulties in determining the research question in observational pharmacoepidemiology are highlighed by Luijken et al (2023).

2.2 Terminology: Association and Causation

In this section we introduce some basic concepts of a formal framework for articulating causal research questions. This allows precision when it comes to showing (mathematically) which method is valid under which assumptions. While realistic research questions can be quite complex, as illustrated with some of the above examples, we keep the formalization basic for the sake of simplicity and to keep a focus on key principles. Note that we usually speak of treatments rather than exposures as the objects of causal inference; for instance, 'smoking' can also be a 'treatment'.

In one aspect, however, we will be less simple than most introductory texts. Throughout we keep in mind the time-dependence of treatment variables, $A_1, A_2, \ldots$, because few typical exposures in epidemiology are point treatments, i.e., occur at a single point in time. In fact, timing or duration of treatments or exposures play an important role in many studies. Moreover, the causal assumptions and methodology in the context of sequential treatments are distinct from point treatments because relevant events can occur *between* sequential treatments. These events may be causally affected by earlier and in turn affect later treatments and outcomes, which requires more complex assumptions and methods. When we do consider a point treatment we will retain the subscript A_1 as a reminder that this is often only the first in a sequence of treatments. The key challenges that occur for time-dependent treatments can already be explained with just two sequential treatments A_1 and A_2, so we mostly restrict the exposition to this case. Further, we denote the outcome variable

by Y and other variables or covariates by a set $\mathbf{X}$ or $X_1, X_2, \ldots, X_p$, where the latter may sometimes have a temporal order; unmeasured variables are denoted by U. For simplicity, all variables are considered discrete.

Probabilistic versus causal dependence. We use the notation $P(Y = y \,|\, A_1 = a_1)$ and $E(Y \,|\, A_1 = a_1)$ to denote conditional probabilities and conditional expectations, respectively. These describe associative dependencies. For instance, if $P(Y = y \,|\, A_1 = 0) \neq P(Y = y \,|\, A_1 = 1)$ then both variables are associated and A_1 is predictive of Y (and, in fact, vice versa). Multivariable regression models are usually models for $E(Y \,|\, A_1, X_1, X_2, \ldots, X_p)$, as addressed in the Chapter on Regression Methods for Epidemiological Analyses by S. Greenland in this book, where $(X_1, X_2, \ldots, X_p)$ may or may not be ordered in time. In our context, multivariable regressions aim at describing the dependence of Y on treatment A_1 after accounting for (i.e., conditioning on) the covariates $(X_1, X_2, \ldots, X_p)$. Further, we use the notation $Y \perp\!\!\!\perp A_1 \,|\, \mathbf{X}$ for conditional independence (Dawid, 1979); in this instance, conditional independence would imply that $P(Y = y \,|\, A_1 = a_1, \mathbf{X} = \mathbf{x}) = P(Y = y \,|\, \mathbf{X} = \mathbf{x})$. Of note, unlike causal relations probabilistic dependence is symmetric so the same conditional independence also implies that $P(A_1 = a_1 \,|\, Y = y, \mathbf{X} = \mathbf{x}) = P(A_1 = a_1 \,|\, \mathbf{X} = \mathbf{x})$.

An intuitive way to formally distinguish association from causation is via the 'do-operator' (Pearl, 1995, 2009). Here, $\mathrm{do}(A_1 = 0)$ denotes an intervention that fixes A_1 at zero. Hence, $E(Y \mid \mathrm{do}(A_1 = 0))$ now denotes the expected value of Y under such an intervention on A_1. That 'association does not imply causation' is formally expressed as

$$P(Y = y \,|\, A_1 = 0) \neq P(Y = y \mid \mathrm{do}(A_1 = 0)).$$

Moreover, if $P(Y = y \mid \mathrm{do}(A_1 = a_1)) = P(Y = y)$, for all a_1, we have that A_1 does not cause Y (at population level).

Potential outcomes. The above do-operator is not common in the epidemiological literature. Instead, *potential outcomes* (POs) are more popular as a means to formalize causal quantities. We denote by $Y(a_1)$ the potential outcome if A_1 were set to a_1 by some intervention[2] (Rubin, 1974, 2005); if A_1 is a binary treatment vs. control variable we have two potential outcomes, $Y(a_1 = 1)$ and $Y(a_1 = 0)$ (written $Y(1)$ and $Y(0)$ if unambiguous). POs for different treatment values can never be observed jointly because only one value of treatment can be realized for any given individual at any time; they are therefore commonly known as 'counterfactual' outcomes. For a discussion of the subtle differences between hypothetical, potential and counterfactual outcomes see Dawid (2007). For many practical purposes, we can consider $E(Y \mid \mathrm{do}(A_1 = 0))$ as expressing the same notion as $E(Y(a_1 = 0))$, namely the expectation of the outcome under this particular intervention on A_1. Under the assumption of consistency (introduced below), we have that the observable outcome Y equals the potential outcome $Y(A_1)$ that occurs if treatment takes 'the natural value' that it would take without intervention. If $Y(a_1) = Y$, for all a_1, then A_1 does not cause Y (at individual level).

When we are interested in *joint interventions* on multiple, possibly time-dependent treatments, say (A_1, A_2), we consider the POs $Y(a_1, a_2)$. For instance, $Y(a_1 = 0, a_2 = 0)$ expresses the PO under 'never treat' whereas $Y(a_1 = 1, a_2 = 1)$ is the PO under 'always treat' (never / always being only two points in time, here); this may further be contrasted with $Y(a_1 = 1, a_2 = 0)$ for the PO under 'only early treatment' or $Y(a_1 = 0, a_2 = 1)$ for the PO under 'late start of treatment'. Note that when we reduce a time-dependent treatment sequence to the first treatment we effectively consider $Y(a_1) = Y(a_1, A_2)$, i.e., the outcome where the value of the first treatment is imposed by an intervention and the second treatment is left to take its natural value.

Cross-world assumptions and single world models. The language of POs allows us to express beliefs or assumptions such as '$P(Y(0) = 1 \,|\, Y(1) = 0) = 0.2$' stating that the probability of the outcome event occurring without treatment is 0.2 given that, under treatment, the outcome event did not occur. This is a genuinely counterfactual statement in the sense that it refers to a joint model for the POs $(Y(1), Y(0))$. Such models or assumptions are termed *cross-world* because they combine two parallel worlds, one where treatment is taken and the other where it is not taken (Dawid, 2000). Richardson and Robins (2013b) suggest a framework for causal models that is restricted to a *single world* and does not allow cross-world assumptions, as these are empirically never verifiable and therefore regarded as unscientific by some authors (Richardson and Robins, 2013a; Stensrud et al, 2021; Andrews and Didelez, 2021). These single world models go back to the so-called *finest fully randomized causally interpretable structured tree graphs (FFRCISTGs)* of Robins (1986).

Bibliographic notes: Binary treatments and interventions that fix a treatment at a given value are still rather simplistic, even if we allow for sequential treatments. Frameworks for more complex treatment strategies exist and are often more relevant in practical applications, e.g., shift interventions as mentioned in Section 2.1 in the context of increasing sleep duration by 10%, or dynamic or adaptive interventions such

[2] Other common notations are Y^{a_1} or Y_{a_1}.

as 'start treatment when a certain condition is met' (Cain et al, 2010; Morzywołek et al, 2022). These are 'generalized' treatments and are sometimes denoted by the index g, and thus give rise to the potential outcome $Y(g)$ denoting the outcome when following, say, a specified dynamic strategy g. Other popular formal causal modeling frameworks are based on *structural equation models (SEMs)*, e.g., the non-parametric SEMs or structural causal models which impose additional structure (Pearl, 2009; Peters et al, 2017). Perhaps the most parsimonious framework, imposing the least structure, is that of influence diagrams proposed by Dawid (2002) (see also Dawid and Didelez (2010) or Dawid (2015)).

2.3 Basic Notions of Causal Effects

With the terminology of Section 2.2 we now present some basic notions of causal effects, often called 'causal estimands'. These are defined without refering to, or requiring, a parametric model (Hernán, 2004). In the following, we make a number of simplifications, such as binary treatments and either one, A_1, or two possibly sequential treatments, (A_1, A_2). However, we keep in mind that practically relevant causal effects usually involve much more complex intervention(s), as in the above examples.

Individual, conditional, and average causal effects of a point treatment. Consider just a single treatment A_1 and potential outcome $Y(a_1)$ with $a_1 \in \{0, 1\}$. Using the index i to denote an individual, we define an *individual causal effect (ICE)* as a contrast of the individual potential outcomes $Y_i(1)$ versus $Y_i(0)$, for instance $ICE = Y_i(1) - Y_i(0)$. Given that potential outcomes cannot simultaneously be observed for the same individual under different treatment settings, it is self-evident that no data or study design (not even randomization) can inform us about individual effects without being supplemented by special assumptions that are fundamentally untestable (Dawid, 2000; Hernán and Robins, 2020). We might view the index i as replacing a description of *all* characteristics of an individual that could possibly be relevant. Hence, a less stringent type of causal effect is obtained by averaging within meaningful subgroups, i.e., by conditioning on a finite set of measured baseline (or pre-treatment) covariates $\mathbf{Z}$ such as sex, age and income. The *conditional average (causal) treatment effect (CATE)* is given as

$$CATE(\mathbf{Z} = \mathbf{z}) = E(Y(1) - Y(0) \,|\, \mathbf{Z} = \mathbf{z}),$$

or in words, the average difference in outcome if all those in subgroup $\mathbf{Z} = \mathbf{z}$ (e.g., women aged 40-45 with low income) were treated versus if they were not treated. If the average is taken over the whole eligible population, we obtain a (population) *average causal effect (ACE)*,

$$ACE = E(Y(1) - Y(0)).$$

In words this is the average difference in outcome if all those eligible were treated versus if all of them were not treated. Note that if the *CATE* differs for different choices of subgroups, $\mathbf{Z} = \mathbf{z}$, then there is *effect modification*, or moderation, by (some variable in) $\mathbf{Z}$. The *ACE* is then a weighted average of the *CATE*s over the subgroups. Also note that all of these are *total* effects, in the sense that they describe differences between the distributions (or just their expectations) of $Y(1)$ and $Y(0)$ regardless of exactly how the differences are brought about.

A particular subgroup is that of the 'treated', meaning those who would naturally take treatment if offered, i.e., where $A_1 = 1$ without intervention. The average *causal effect of treatment on the treated (ATT)* is defined as $E(Y(1) - Y(0) \,|\, A_1 = 1)$. This causal effect is relevant for policy makers when considering if, say, a preventive measure such as a cancer screening program were to be offered because the effect would only concern those who would take advantage of the program, a subgroup which might be quite different from those who would not (Wang et al, 2017).

In all of the above, other contrasts can be used to replace the difference, such as the ratio, e.g., the causal relative risk for binary outcomes. Odds ratios and hazard ratios are popular effect parameters, but their causal interpretation involves some subtelties: The population odds- or hazard ratio is not usually a weighted average of the conditional or subgroup odds or hazard ratios due to noncollapsibility (Greenland et al, 1999; Hernán, 2010; Stensrud et al, 2017; Greenland and Pearl, 2011; Daniel et al, 2021). Furthermore, the point treatment can, in principle, be replaced by a set of multiple treatments, though the contrast will then typically involve more than two settings (Linden et al, 2016).

Multiple, sequential, controlled direct and interacting treatment effects. For all of the above types of causal effects, we may replace the single treatment A_1 by multiple treatments or a sequence of temporally ordered treatments $A_1, A_2, \ldots$, so as to obtain individual, conditional, and population causal effects

contrasting *joint interventions* on the treatments. For instance, with two treatments, $ACE = E(Y(a_1 = 1, a_2 = 1) - Y(a_1 = 0, a_2 = 0))$ would be the population effect of 'always' or sustained treatment versus 'never treat' (we will also write this as $ACE = E(Y(1,1) - Y(0,0))$). It is clear that when comparing any possible treatment sequence, there are a large number of potentially relevant contrasts. A special case is the contrast $E(Y(1,1) - Y(0,1))$ where the second treatment is fixed at the same value, while the first is varied. This can be interpreted as the *controlled direct effect (CDE)* of A_1, fixing $A_2 = 1$, because any effect that A_1 may have on A_2, and thus indirectly on Y, will be cut off by the intervention fixing A_2 (see also Section 6, below). An example where such an effect could be interesting is the question whether certain behaviors in early childhood, e.g., high versus low physical activity, still have a remaining effect on adolescent health if we could hypothetically ensure high physical activity levels in later childhood.

A further special case of joint intervention effects is that of *causal interactions* (this is regardless of any temporal ordering of A_1 and A_2). A causal interaction occurs if the effect of intervening on A_2 is different depending on what we fix A_1 to be, i.e., if $E(Y(1,1) - Y(1,0)) \neq E(Y(0,1) - Y(0,0))$ (VanderWeele, 2015). Note that this is not the same as effect modification (cf. above) for which we simply condition on other observed variables. The difference is that for a causal interaction, both A_1 and A_2 must actually have a causal effect on Y, whereas to be an effect modifier it would suffice to condition on a relevant subgroup indicator or, say, a proxy for an unmeasured cause of Y.

Attributable fractions. Many more types of causal effects have been proposed in the literature. For instance a contrast of a possibly sustained intervention versus no intervention, e.g., $E(Y(a_1 = 1, a_2 = 1))$ versus $E(Y)$, is referred to as the *population intervention* effect (Westreich, 2017). It quantifies how much the average outcome would be different if a particular intervention were to be imposed versus retaining the current state of affairs (or versus 'standard of care'). In case of a binary outcome denoting presence of, say, a disease, this type of contrast is often transformed to the *attributable fraction (AF)*:

$$AF = \frac{P(Y = 1) - P(Y(1,1) = 1)}{P(Y = 1)}.$$

Supposing that the intervention reduces the risk of disease, the AF denotes the proportion of diseased that could be prevented by implementing the intervention in the whole population[3]. For example, Börnhorst et al (2023) investigate whether a hypothetical intervention of sustained adherence to guidelines for healthy behaviors in children affected the 13-year risk of overweight/obesity compared to the natural course (i.e., not doing anything differently). The population intervention effect, here, was estimated to be a risk reduction by 5 percentage points. Transforming this, the estimated AF suggested that 17% of the 13-year risk of overweight/obesity in children were attributable to non-adherence to guidelines on sustained health behaviors.

Time-to-event and survival outcomes. When we have a time-to-event or survival outcome, Y, the statistical analysis usually needs to deal with censored data. It is then popular to consider parameters on the hazard scale, typically hazard ratios (HRs), as hazards are simpler to estimate under censoring. The noncollapsibility of hazard ratios was mentioned earlier, which means that a HR version of a population effect does not necessarily equal a weighted average of subgroup HRs. A second issue is the following: We can define a 'causal hazard ratio' as the ratio of the hazard of the distribution of $Y(a_1 = 1)$ and the hazard of the distribution of $Y(a_1 = 0)$. However, the interpretation is subtle because the hazard is based on a conditional probability given survival, $Y(a_1) \geq t$, $a_1 = 0, 1$, which will over time become different in the treated versus control arm. Those surviving, say, one year on the treatment arm may be very different kinds of patients from those surviving one year in the control arm (Hernán, 2010). The issue is, in fact, one of post-treatment selection (Stensrud et al, 2017). However, to resolve the problem, it is often possible to use flexible hazard models just as a tool when analyzing time-to-event data but to report causal effects on the relative or absolute risk scale by suitably transforming the hazard functions (Cole and Hernán, 2004; Toh et al, 2010; Sjölander, 2018; Daniel et al, 2021).

Bibliographic notes: Other types of causal effects include those of shift or stochastic (randomized) interventions, as well as dynamic interventions mentioned above (Díaz and van der Laan, 2013). A special type of subgroup effects are those restricted to a 'principal stratum' (Frangakis and Rubin, 2002). For instance, when assessing the cognitive function in elderly patients, the principal stratum is often defined as those who would be alive (so that cognitive function is well-defined) under treatment *and* under control. Such principal stratum effects have been criticized due to their cross-world nature, and because the particular

[3] Hernán and Robins (2020) distinguish the AF, or excess fraction, from the *etiological* fraction which is defined at individual level.

subgroup is indeterminable from observable data (Robins and Richardson, 2011; Dawid and Didelez, 2012; Stensrud and Dukes, 2022).

3 Causal Models and Assumptions

Causal inference is typically conducted in the context of a causal (or structural) model, i.e., a model for the probabilistic distribution of relevant variables and processes under the intervention(s) of interest and under the absence of an intervention. A notable exception is, perhaps, a randomized trial where, by virtue of randomization, a causal interpretation of an association between treatment and outcome is regarded as acceptable without a causal model. However, even RCTs are usually not conducted without a motivating question and background knowledge that not only provides justification for carrying out the RCT, but also guides many of the RCT's design choices. As such, one could consider this background knowledge as an implicit causal model for RCTs.

In observational studies, a causal model makes the current state of knowledge as well as the assumptions motivating or justifying a particular analysis explicit. It links the observable data to the (assumed) underlying data generating process and allows us to scrutinize and either exclude or be aware of possible sources of bias. Basic causal models specify a minimum of relations (or their absence) between POs, observable and possibly further latent variables. For instance, we could assume a basic model claiming that $Y(a_1) \perp\!\!\!\perp A_1 \,|\, \mathbf{X}$ and that $Y(a_1) = Y$ if $A_1 = a_1$, which implies that conditionally on $A_1 = a_1$ and $\mathbf{X} = \mathbf{x}$ we can equate the distribution of Y with that of $Y(a_1)$. However, to make a convincing argument supporting the assumption that $Y(a_1) \perp\!\!\!\perp A_1 \,|\, \mathbf{X}$, more elaborate causal models can be helpful, such as causal directed acyclic graphs (DAGs) covered in Section 3.1, below. Causal DAGs impose restrictions on the presence or absence of (direct) causal relations and of latent variables without any specific parametric assumptions. They are, therefore, often referred to as 'non-parametric' causal models[4] (Pearl, 2009). At the other extreme are causal models that heavily rely on mechanistic and parametric assumptions, such as linear SEMs which allow us to model the joint distribution of POs and to make cross-world assumptions.

Identifiability. An important stage of a causal analysis is to assess identifiability of the desired causal target of inference under a defensible causal model. Identifiability is given if the observable data are in principle sufficient to uniquely determine the causal effect. Here, 'in principle' means disregarding finite sample issues. Identifiability has been formally proved for a variety of causal effects under a variety of causal models. Note that assumptions yielding identifiability are specific to causal inference and more fundamental than further statistical modeling assumptions typically made when it comes to statistical estimation, such as assuming a logistic, linear or proportional hazards model.

3.1 Causal Directed Acyclic Graphs

Causal DAGs feature prominently in epidemiology as tools to guide causal inference because they offer a transparent and intuitive way to communicate assumptions (Robins, 2001; Staplin et al, 2017). Hernán advised: 'Draw your assumptions before you draw your conclusions!' (Hernán, 2024). An introduction to (causal) DAGs is given in the Chapter on Causal Directed Acyclic Graphs by R. Foraita et al, in this book, and an overview of the different graph-based causal modeling approaches can be found in Didelez (2018). Here, we only give a brief recap of the main concepts.

A DAG has nodes representing variables (measured or unmeasured) and directed edges. Strictly speaking, an edge $V \longrightarrow W$ between some variables V and W expresses that we cannot rule out the presence of a *CDE* of V on W fixing (by an intervention) other parent nodes of W, i.e., being 'direct' is relative to the set of other variables in the graph. The presence of an edge is sometimes misunderstood as there must be a (direct) effect, but when expressing the key assumptions of a causal model, it is more important that the *absence* of an edge must indicate certainty about the absence of a direct effect. For instance, in Figure 2(i), the absence of edges between A_1 and X_2 and from X_1 to Y may be crucial for the validity of an analysis as they imply that adjusting for either X_1 or X_2 is sufficient to eliminate any confounding under this causal model, i.e., we do not need to measure both (cf. below). With this caveat in mind, we will nevertheless loosely say that, in a causal DAG, a directed path from one variable to another indicates that there is a (possibly indirect) causal effect of the former on the latter, while any other open path between two variables induces a non-causal

[4] This should not be confused with non-parametric statistical models or methods.

association. Whether a path is open or not can be established using d-separation (Pearl, 2009). For instance in Figure 2(i), the path $A_1 \longleftarrow X_1 \longrightarrow X_2 \longrightarrow Y$ is open, but would be blocked if we condition on either X_1 or X_2 (or both), e.g., by stratification. In contrast, in Figure 2(ii) the path $A_1 \longleftarrow X_1 \longrightarrow S \longleftarrow X_2 \longrightarrow Y$ is blocked, but would be opened by conditioning on S, e.g., by selection. These two basic examples are at the heart of most sources of structural bias threatening the validity of causal inference: Confounding and selection bias (cf. Section 3.2).

Single world intervention graphs (SWIGs). It may not be obvious how to link DAGs on nodes that seemingly represent factual variables with typical causal assumptions because POs do not explicitly appear in causal DAGs. Richardson and Robins (2013b) devise a sound way to combine the two and make POs explicit in the graph, which they call *single world intervention graphs (SWIGs)*. These are augmented DAGs where the treatment nodes are split into two nodes, the 'natural' variable (without intervention) and the specific value imposed by the intervention. An example is given in Figure 2(iii): The intervened node, a_1, has no parent nodes, representing that an intervention severs the observational dependencies, while all descendants of such a node inherit the interventional value as POs. The node representing the natural treatment, A_1, inherits all of the parent nodes. Via d-separation we can now check on this SWIG that $Y(a_1) \perp\!\!\!\perp A_1 \,|\, X_1$ as well as $Y(a_1) \perp\!\!\!\perp A_1 \,|\, X_2$. As the name says, SWIGs are intended as models that cannot express, and therefore do not make, any cross-world assumptions. We refer the interested reader to a series of papers for further details (Robins et al, 2022; Shpitser et al, 2022).

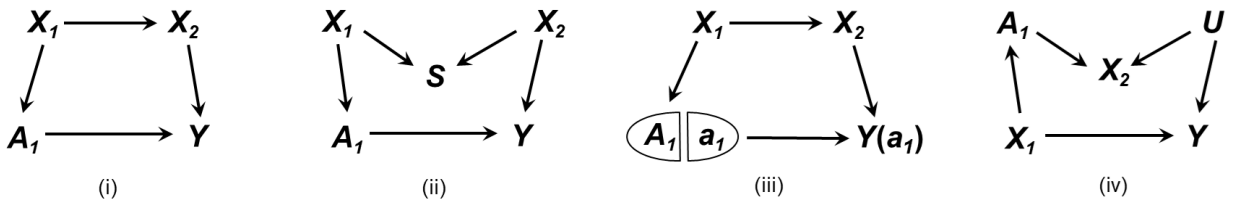

Fig. 2 Examples of DAGs with point treatment A_1 and outcome Y: (i) Two covariates (X_1, X_2) on one backdoor path; (ii) M-DAG illustrating backdoor path opened by selection on S; (iii) Single world intervention graph (SWIG) for (i); (iv) Post-treatment selection bias in, e.g., regression of Y on (A_1, X_1, X_2).

Bibliographic notes: A popular and useful tool for constructing and analyzing (causal) DAGs is the software 'dagitty' (Textor et al, 2016) with its latest extension (Ankan et al, 2021). A detailed critical discussion of the assumptions implied by causal DAG models is given by Dawid (2010). Care is also needed when attempting to use DAGs for representing dynamic processes in continuous time with feedback (Aalen et al, 2016).

3.2 Structural Bias

Bias refers to a *systematic* difference between the intended target of inference (the estimand) and the average result of a chosen analysis. Note that random differences due to sampling variation, typically due to small sample sizes, are not indicative of bias. Further, we distinguish between statistical bias of a chosen estimator (which often vanishes with sufficiently large samples under the assumed statistical model) and *structural bias* which is induced by problems with a chosen study design, database, or analysis plan. Many sources of structural bias affect any kind of statistical analysis, such as measurement error or non-ignorable drop-out. Here, we focus on those types of structural bias that are most typical for causal analysis: Confounding and selection bias. We distinguish between simple confounding and time-dependent confounding, where the latter requires special attention in case of treatment-confounding feedback.

Confounding. Confounding of the effect of a point treatment on an outcome occurs when there are some common causes of both (Greenland et al, 1999). Formally this implies that $Y(a_1) \not\perp\!\!\!\perp A_1$ or $P(Y \,|\, A_1 = a_1) \neq P(Y \,|\, \mathrm{do}(A_1 = a_1))$. Graphically it can easily be detected in a causal DAG by tracing back any common ancestors of the treatment and outcome nodes. Randomization 'severs' edges into the treatment node and thus ensures balance (on average) regarding all baseline characteristics and hence also for those relevant to the outcome. Note that we have defined 'confounding' and not a 'confounder'; as discussed in much detail by VanderWeele and Shpitser (2013), it is not trivial to define the concept of a

confounder as this may be relative to other variables in the analysis. For example, in Figure 2(i), confounding is present because X_1 is a common cause of treatment and outcome; however, if we adjust for X_2 then there is no residual confounding (under this causal DAG model). VanderWeele and Shpitser (2013) also show that common associational 'definitions' of a confounder are not correct. Confounding must, indeed, be considered in the context of a causal model.

Selection. Selection occurs when either by design of the study or by the method of analysis we deliberately or inadvertently condition on certain characteristics (Hernán et al, 2004; Didelez et al, 2010a). This does not necessarily lead to bias, but needs to be scrutinized. Typical study designs based on selection are, e.g., case-control studies (Didelez and Evans, 2018) or test-negative designs (Shrier et al, 2023). Formally, the problem can be analyzed by introducing a binary indicator S for selection (Didelez et al, 2010a). Suppose we want to investigate aspects of $E(Y(a_1) \,|\, \mathbf{Z} = \mathbf{z})$ like the *CATE*, and we have established that identification holds given data on $(Y, A_1, \mathbf{X})$. However, due to selection we can only 'access' the conditional distribution of these variables given selection, $P(Y, A_1, \mathbf{X} \,|\, S = 1)$. For identification we then need to establish, under defensible assumptions, that relevant aspects of the unconditional distribution can be obtained from the distribution conditional on selection. Graphically, the problem can often be detected when there is direct or indirect (through a descendant) selection on a so-called *collider* node which opens an otherwise blocked path (Cole et al, 2010; Hernán and Monge, 2023).

Many design-based biases are a form of selection bias, such as some kinds of immortal-time or prevalent-user bias (Ray, 2003; Suissa, 2008; Hernán et al, 2016; Shrier and Suissa, 2022). Consider the DAG in Figure 2(ii) as an example. With d-separation we establish that, here, there is no open non-causal path between A_1 and Y. However, conditionally on S the non-causal path $A_1 \longleftarrow X_1 \longrightarrow S \longleftarrow X_2 \longrightarrow Y$ is opened. Due to the shape of the DAG, this is known as *M-bias* (Greenland, 2003). An example for a design that induces M-bias is illustrated by Braitmaier et al (2024): When investigating the effect of screening colonoscopy on colorectal cancer (CRC), it is common to define exposure A_1 as a summary of past attendances at such screenings X_1, with the outcome Y being incident CRC. At the same time, persons with prior CRC diagnosis are excluded from the study ($S = 0$) but a prior diagnosis is most likely to occur if there was latent CRC (X_2) and a screening took place in the past. Thus, the eligibility criteria together with the particular exposure definition which looks into the past mean that this study design systematically reduces the number of latent CRCs in the 'exposed' but not in the 'unexposed' subjects leading to an overestimation of the causal effect.

A selection effect could further occur, for instance, when the analysis wrongly adjusts (conditions or stratifies) for 'colliders'. While it is well-known that one should not adjust for consequences of treatment as this may mask the causal effect, there is a second issue: In Figure 2(iv), suppose the analyst is not aware that X_2 is affected by the treatment and decides to simply adjust for all covariates, X_1 and X_2, e.g., by fitting a regression model for $E(Y \,|\, A_1, X_1, X_2)$. Note that in this example, A_1 has no causal effect at all on Y, as indicated by the absence of any directed path. However, while conditioning on X_1 blocks one non-causal path, conditioning on the collider X_2 opens the non-causal path $A_1 \longrightarrow X_2 \longleftarrow U \longrightarrow Y$. A real example can be found in an analysis of risk factors for COVID-related death (Williamson et al, 2020): The analysis has been interpreted as suggesting that smoking A_1 has a protective effect on the outcome Y. However, one could plausibly explain this by the simultaneous inclusion of covariates X_2 indicating respiratory diseases which are themselves consequences of prior smoking and may have other common causes U, not included in the analysis, with the outcome. For instance, smokers who have a respiratory disease would be less likely to have other unmeasured lung problems, U, than non-smokers with respiratory diseases and thus have a reduced risk of COVID-related death (van Geloven et al, 2024). A similar causal structure has been proposed to explain the 'birth weight paradox', where infants with low birth weight have a lower mortality if born to smokers than those born to non-smokers (Hernández-Díaz et al, 2006).

While selection bias is sometimes avoidable by choosing a more suitable design or analysis, it may be that without further data or information it cannot fully be avoided. Criteria for bounding selection bias in such situations have been given (Smith and VanderWeele, 2019; Zetterstrom and Waernbaum, 2022).

Treatment-confounding feedback. To understand the issues with treatment-confounding feedback, we need to combine the two types of structural biases above. Consider two sequential treatments A_1 and A_2, where a covariate X_2 is observed after the first treatment decision but before the second one; $\mathbf{X}_1$ may be a set of general baseline covariates. A historical example is that of early observational studies with HIV patients, where the timing of initiation of anti-retroviral treatment (early $A_1 = A_2 = 1$, or later $A_1 = 0, A_2 = 1$) was usually decided dynamically depending on the CD4-count X_2 which reacts to the presence or absence of earlier treatment (Hernán et al, 2000). A plausible causal model is depicted in Figure 3(i). Here, X_2 is confounding the effect of later treatment on the outcome while being a consequence of earlier treatment. Moreover, the change in CD4-count further depends on the underlying unobserved general health U of the patient. In this model, conditioning on X_2 blocks some non-causal paths between A_2 and Y. However, it

also blocks a causal path from A_1 to Y potentially masking the overall effect of A_1. Further, as a collider it opens the non-causal path $A_1 \longrightarrow X_2 \longleftarrow U \longrightarrow Y$, thus introducing two potential sources of bias. Due to its subtle role in this collider path, U has been termed a 'phantom' variable (Bates et al, 2022). A regression model for $E(Y \mid A_1, A_2, X_2)$ would mix the latter two non-causal associations regarding the effect of A_1, while a regression for $E(Y \mid A_1, A_2)$ would be biased due to not accounting for confounding by X_2 of A_2 and the outcome. However, as we will address further below, the joint causal effect of intervening on A_1 and A_2 is still identified in such a treatment-confounding feedback case, provided that X_2 is measured, just not by a single regression model.

Treatment-confounding feedback thus constitutes a particular challenge requiring non-standard methods. The importance of acknowledging treatment-confounding feedback for time-dependent exposures in many, if not most, epidemiological analysis was uncovered and advocated by Robins (1986) and in much work since. An early example of Robins' is the healthy worker survivor effect: Workers who fall sick due to adverse work conditions tend to drop out or change employment so that those who remain in work tend to be especially resistant. Combined with naive methods that are unsuitable for analyzing treatment-confounding feedback, this could make the adverse work conditions appear beneficial instead of harmful.

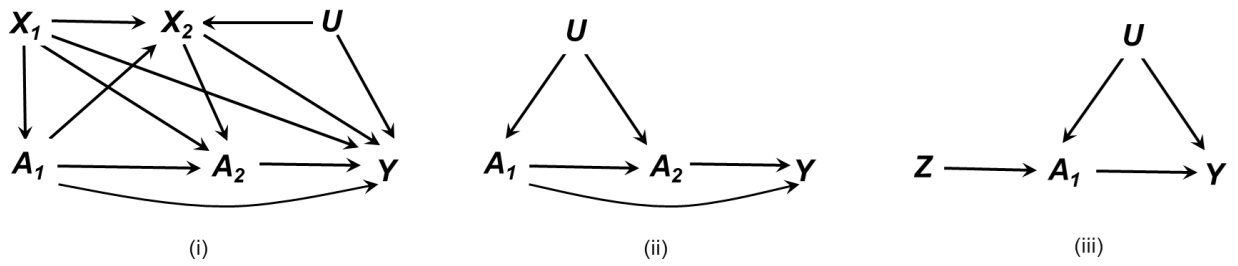

Fig. 3 (i) Example DAG with sequential treatments (A_1, A_2) and treatment-confounding feedback via X_2; (ii) DAG where the joint effect of (A_1, A_2) is identified, but not of A_1 alone when U is unmeasured; (iii) IV-assumptions with Z as IV for the effect of A_1 on Y and unmeasured confounder U.

3.3 Structural Assumptions

Structural assumptions enable identifiability: They ensure that the observable data are *in principle sufficient* to infer the desired causal quantity, except for sample size issues and parametric assumptions. In the field of causal inference, there are mainly two types of such assumptions: (i) For one, we could assume that enough information is available in the data to adjust for confounding. Or, (ii) we could acknowledge unmeasured confounding and instead rely on natural experiments or specific designs and modeling assumptions. These are complemented by approaches that gauge the amount of bias or produce bounds for causal quantities under much weaker assumptions. In this section we address standard assumptions of type (i) for point and sequential treatments; and as an example for (ii), we briefly describe instrumental variable (IV) designs, which yield bounds on causal effects or can be supplemented by specific parametric assumptions to obtain point identification. As mentioned before, the plausibility of structural assumptions will be strengthened if they can be derived from a defensible causal model. Moreover, the model may help to derive some testable implications to further support plausibility.

(Causal) consistency. Causal *consistency*[5] is a key assumption that allows us to link the observational data to the situation under the targeted hypothetical intervention. Formally, causal consistency postulates that

$$\text{when we observe } A_1 = a_1 \text{ then } Y = Y(a_1),$$

or $Y = Y(A_1)$. Similarly for multiple treatments: When $(A_1, A_2) = (a_1, a_2)$ is observed, then $Y = Y(a_1, a_2)$. In words this means that when treatment was observed to be a_1, e.g., high physical activity (PA), then the observable outcome Y equals the potential outcome $Y(a_1)$ which would be observed if an intervention, e.g., an enhanced PA program, had been applied to enforce $A_1 = 1$. Violations occur if the observational situation is very different from the envisaged interventional setting, e.g., if an intervention would be too invasive,

[5] In the context of causal inference, this is often just called 'consistency'. It must not be confused with, and is completely unrelated to, the identically named property of a statistical estimator.

creating a completely different situation, or if there is too much ambiguity in how the intervention could be imposed (VanderWeele and Hernán, 2013). In the PA example, for instance, it may be implausible that children not participating in the enhanced PA program would all have low PA as the absence of such a program does not prevent them from becoming physically active. Indeed, for many typical epidemiological studies it is implausible that one can intervene directly on the exposure, such as BMI or sleep duration. In such cases it may be worthwhile to augment the causal model with additional variables that could be intervened upon so as to make additional assumptions explicit. For instance, if we envisage changing BMI by increasing PA, then we must be aware that such an intervention also affects other aspects, such as cardiovascular fitness and muscle tone, which may be relevant to the outcome. Views on how strict one should be about causal consistency differ (Cole and Frangakis, 2009; VanderWeele and Hernán, 2013; Hernán, 2016), but being as explicit as possible about the relationship between available data and targeted intervention can help guide analysis choices (Young, 2024).

Conditional exchangeability. We begin with the case of a point treatment A_1. *Conditional exchangability* requires that a set $\mathbf{X}$ of pre-treatment covariates is sufficient to adjust for confounding. More formally this is expressed as

$$Y(a_1) \perp\!\!\!\perp A_1 \,|\, \mathbf{X}.$$

There are many other names for this assumption in the literature: 'no unmeasured confounding' (presupposing that $\mathbf{X}$ is measured), 'conditional ignorability', 'unconfoundedness' (given $\mathbf{X}$), or 'selection on observables' (again, presupposing that $\mathbf{X}$ is observable). Intuitively, the assumption means that within strata of $\mathbf{X}$, the observed treatment can be regarded 'like randomized' as it retains no further information on the PO. In other words, treated and untreated subjects are 'exchangeable' at baseline given $\mathbf{X}$. The covariates $\mathbf{X}$ are then called *sufficient for adjustment*.

To see mathematically how this assumption helps, consider

$$P(Y(a_1) = y) = \sum_{\mathbf{x}} P(Y(a_1) = y \,|\, \mathbf{X} = \mathbf{x}) P(\mathbf{X} = \mathbf{x})$$

by total probability[6]. With conditional exchangeability we now have

$$P(Y(a_1) = y \,|\, \mathbf{X} = \mathbf{x}) = P(Y(a_1) = y \,|\, A_1 = a_1, \mathbf{X} = \mathbf{x})$$

which equals $P(Y = y \,|\, A_1 = a_1, \mathbf{X} = \mathbf{x})$ by causal consistency. In summary, we obtain

$$P(Y(a_1) = y) = \sum_{\mathbf{x}} P(Y = y \,|\, A_1 = a_1, \mathbf{X} = \mathbf{x}) P(\mathbf{X} = \mathbf{x}). \tag{1}$$

This yields identification because the right-hand side of equation (1) is now a statistical estimand which can be estimated from observational data on $(Y, A_1, \mathbf{X})$. Equation (1) is the simplest case of the so-called *g-formula* (Robins, 1986), and is also known as standardization (Davis, 1984; Keiding and Clayton, 2014). Note that (unconditional) exchangeability holds if $\mathbf{X}$ is empty, which can be ensured by randomizing A_1.

A famous graphical rule in point-treatment settings is the *backdoor criterion* which establishes conditional exchangeability given $\mathbf{X}$ (Pearl, 1995). In a causal DAG representing the plausible causal structure, the criterion states that the set of nodes representing $\mathbf{X}$ must not be descendants of A_1 and block all backdoor paths from treatment to outcome, i.e., any paths between A_1 and Y that begins with an edge out of the former, $A_1 \longleftarrow \cdots$; see Chapter on Causal Directed Acyclic Graphs by R. Foraita et al in this book. Intuitively, any unblocked backdoor path would potentially induce a non-causal association and therefore needs to be blocked. In the example of Figure 2(i), the single backdoor path $A_1 \longleftarrow X_1 \longrightarrow X_2 \longrightarrow Y$ is blocked by either X_1 or X_2 or both but not by the empty set (illustrating that a sufficient adjustment set is not unique). In contrast, in Figure 2(ii), the single backdoor path $A_1 \longleftarrow X_1 \longrightarrow S \longleftarrow X_2 \longrightarrow Y$ is blocked by the empty set. However, conditioning on S, e.g., through selection or by analysis opens this path (inducing a non-causal association) so that, again, either X_1 or X_2 would need to be included for identification. Alternatively, making the POs explicit, conditional exchangeability can be read off from a SWIG by checking if the node $Y(a_1)$ is d-separated from the node A_1 by $\mathbf{X}$. For instance, in Figure 2(iii) they are separated by either X_1 or X_2.

The assumption of conditional exchangeability and the backdoor criterion can be extended to multiple treatments and joint interventions. Suppose that (A_1, A_2) are two not necessarily temporally ordered treatments. Conditional exchangeability then requires that $\mathbf{X}$ must be prior to both (A_1, A_2) and satisfy $Y(a_1, a_2) \perp\!\!\!\perp (A_1, A_2) \,|\, \mathbf{X}$. Using the backdoor criterion, the set $\mathbf{X}$ would then need to block all backdoor

[6] At this stage, we use that $\mathbf{X}$ is pre-treatment, i.e., not itself affected by treatment so that $\mathbf{X}(a_1) \equiv \mathbf{X}$.

paths out of (A_1, A_2), jointly. Equation (1) is then extended by including A_2. Note that joint interventions can be identified even if individual ones are not; see Figure 3(ii): Here, the (total) effect of A_1 alone is not identified if U is unmeasured due to the open backdoor path $A_1 \longleftarrow U \longrightarrow A_2 \longrightarrow Y$ that cannot be blocked by A_2 as this is itself affected by A_1, but the joint effect of (A_1, A_2) is identified in this DAG.

Importantly, for time-dependent treatments conditional exchangeability is often too stringent an assumption, and the backdoor criterion does not capture all causal structures under which the causal effects are identifiable. Conditional exchangeability does not hold in case of treatment-confounding feedback as in Figure 3(i), and the backdoor criterion does not apply because, here, X_2 is a descendant of (and thus affected by) A_1. In particular, equation (1) cannot be extended to sequential treatments by simply replacing A_1 with (A_1, A_2). The following assumption is therefore more appropriate for sequential treatments.

Sequential conditional exchangeability. Let us consider the setting with two sequential treatments A_1 and A_2; assume $\mathbf{X}_1$ is a set of pre-A_1 covariates and $\mathbf{X}_2$ are pre-A_2 but possible causally affected by A_1. Formally, *sequential conditional exchangeability* is given if

$$Y(a_1, a_2) \perp\!\!\!\perp A_1 \,|\, \mathbf{X}_1 \quad \text{and} \quad Y(a_1, a_2) \perp\!\!\!\perp A_2 \,|\, (\mathbf{X}_1, \mathbf{X}_2, A_1 = a_1).$$

In Figure 3(i), sequential conditional exchangeability holds without the need to adjust for U as long as X_1 and X_2 are given. The assumption has also been described as 'sequential randomization' given $(\mathbf{X}_1, \mathbf{X}_2)$ because it requires that observationally A_1 and A_2 be 'like randomized' within strata of $\mathbf{X}_1$ and within strata of $(\mathbf{X}_1, \mathbf{X}_2, A_1)$, respectively (Robins, 1997). The graphical check is more involved (Pearl and Robins, 1995; Robins, 1997): Loosely speaking, we need $\mathbf{X}_1$ to block all backdoor paths out of A_1 (except if they are blocked by future treatment nodes) and $(A_1, \mathbf{X}_1, \mathbf{X}_2)$ to block all backdoor paths out of A_2.

With causal consistency, sequential conditional exchangeability allows the representation

$$\begin{aligned} P(Y(a_1, a_2) = y) = \sum_{\mathbf{x}_1, \mathbf{x}_2} & P(Y = y \,|\, A_1 = a_1, A_2 = a_2, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2) \\ & P(\mathbf{X}_2 = \mathbf{x}_2 \,|\, A_1 = a_1, \mathbf{X}_1 = \mathbf{x}_1) P(\mathbf{X}_1 = \mathbf{x}_1). \end{aligned} \tag{2}$$

This is known as the *g-formula* (Robins, 1986) and can be interpreted as sequential version of standardization for point treatments; see equation (1). The right-hand side of equation (2) yields a statistical estimand in terms of observables. We see from (2) that the g-formula makes use of $P(Y \,|\, A_1, A_2, \mathbf{X}_1, \mathbf{X}_2)$ which could be modeled as a regression of the outcome Y on A_1, A_2 and all the covariates; however, for the weighted averaging an additional model for $P(\mathbf{X}_2 \,|\, A_1, \mathbf{X}_1)$ is required (averaging over $\mathbf{X}_1$ could be done empirically without model). Fitting parametric models for these two conditional distributions in (2) to obtain an estimate for $P(Y(a_1, a_2) = y)$ is known as the *parametric g-formula* (cf. Section 4, below). Note that standard regression-based methods neglect the role of $P(\mathbf{X}_2 \,|\, A_1, \mathbf{X}_1)$.

Positivity. Causal consistency and (sequential) conditional exchangeability are not yet sufficient for the identifiability of $P(Y(a_1))$ or $P(Y(a_1, a_2))$. A further important assumption demands that a sufficient variety of treated and untreated subjects are observable. For a point treatment, this means that for the chosen adjustment set $\mathbf{X}$ there is a positive probability

$$0 < P(A_1 = a_1 \,|\, \mathbf{X} = \mathbf{x}) \text{ for } \textit{all} \text{ values of } (a_1, \mathbf{x}).$$

We can see that there will be insufficient information to estimate the right-hand side of equation (1) if there are no observations for some combinations $(A_1 = a_1, \mathbf{X} = \mathbf{x})$[7]. For example, in a pharmacoepidemiological study, some types of patients may never receive the more aggressive treatment while other types of patients may never receive the less aggressive treatment. For those two groups, it is not only empirically impossible to compare the two treatments, it may also not make much medical sense. For the sequential treatment case, positivity similarly demands that each treatment value has positive probability in any stratum of past covariates in the adjustment set.

No interference. A common, and often implicit, assumption is that of *no interference* which essentially states that the treatment of one individual in the population does not affect the outcome of another individual. An obvious violation occurs in the context of infectious diseases where vaccination of some individuals (grandchildren) may protect other individuals (their grandparents). More generally, the relevance of social networks must be considered from the outset for certain research questions. For instance, to investigate adolescent uptake of smoking or alcohol consumption and its consequences, it could be important to take

[7] Strictly, we only need positivity for those $\mathbf{x}$ that can occur, i.e., with $P(\mathbf{X} = \mathbf{x}) > 0$

the structure of the peer network into account. In fact, there may be subtle issues of selection bias, as it cannot easily be disentangled if adolescents choose their friends because they have similar attitudes (e.g., towards smoking) or if the attitudes are chosen to match those of their friends (Shalizi and Thomas, 2011). For overviews on the topic of causal inference under interference we refer to Hudgens and Halloran (2008) or Ogburn et al (2020); in the remainder of this chapter we assume no interference.

Instrumental variable assumptions. As one example of a natural experiment we briefly explain *instrumental variables (IVs)* which are especially popular in econometrics (Angrist and Pischke, 2009; Wooldridge, 2010). Assume that we are interested in a causal effect of A_1 on Y but that unmeasured confounding cannot be ruled out, i.e., it is implausible that any set of measured covariates is sufficient to fully adjust for confounding. An IV Z can be thought of as an imperfect randomization device that induces some variation in A_1 without being itself confounded. Examples in epidemiology are, for instance, *Mendelian randomization*, where a genetic predisposition to an exposure is exploited as an IV (Davey Smith and Ebrahim, 2003; Didelez and Sheehan, 2007; Lawlor et al, 2008), or in parmacoepidemiology the *physician's preference* for a certain version of drug that influences the actually prescribed drug irrespective of patient characteristics (Brookhart et al, 2006; Swanson et al, 2015).

Formally, the IV must be associated with the treatment, $A_1 \not\perp\!\!\!\perp Z$, and it must not itself have an effect other than through A_1 on, nor be confounded with, the outcome Y (possibly after accounting for observed covariates). In a causal DAG, Figure 3(iii), these assumptions are reflected by the absence of any edge between Z and the unmeasured confounder U, reflecting that Z is like randomized; further, the only causal path from Z to Y is through A_1 and there is no other open path between Z and Y.

For a basic idea of why an IV is helpful, consider the special case where the edge $A_1 \longrightarrow Y$ is removed, i.e., where treatment has no causal effect on the outcome. The IV and the outcome are then marginally independent. This is analogous to the case of an RCT with imperfect adherence, where the random assignment and the outcome are independent if treatment has no causal effect. Thus, a valid test of the null hypothesis of no causal effect is given by testing for an association between IV and outcome.

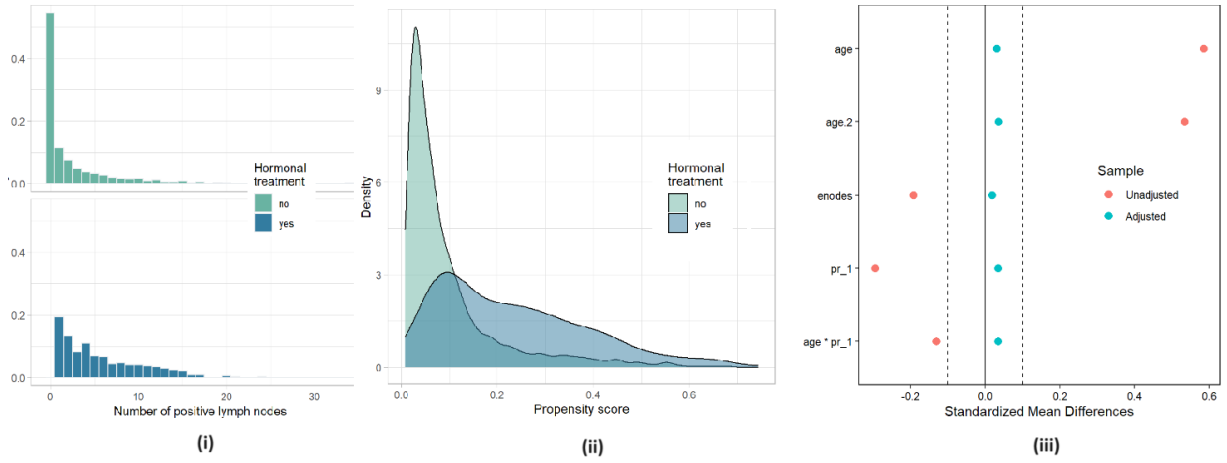

Fig. 4 Diagnostics in causal analyses using the example of hormonal treatment for breast cancer patients. (i) Histograms of covariate 'number of positive lymph nodes' by treatment status reveals a positivity violation: Patients with value zero never receive treatment. (ii) Smoothed density plot of propensity score by treatment group reveals that there is clear confounding but, despite (i), overlap appears satisfactory. (iii) Balancing ('Love') plot of the absolute standardized mean difference between the treated and untreated subjects for each covariate included in the propensity score without versus with adjustment.

Checking assumptions. Of the above assumptions, only positivity can be checked empirically as it presupposes a sufficient adjustment set $\mathbf{X}$ and then only concerns the distribution of observable variables. There are many visualizations that can be used to check positivity, such as histograms or box plots of all covariates in $\mathbf{X}$ by treatment group (see Figure 4(i)). The most common method for inspecting positivity, or in fact the amount of overlap between treated and untreated subjects, is to fit a model for the *propensity score* $P(A_1 = 1 \,|\, \mathbf{X} = \mathbf{x})$ and plot the predicted values separately for treated and untreated subjects (see Figure 4(ii)); a positivity violation would be evident if one group has values near zero or one, but details could be blurred as the fitted propensity score model may extrapolate (cf. Section 4). Alternatively, one can use machine learning approaches to determine the (possibly high-dimensional) subset of $\mathbf{x}$-values for which positivity holds (Oberst et al, 2020).

The other structural assumption given above, and many alternative ones found in causal inference, are not empirically testable as they involve POs; one would need observational and interventional data from otherwise entirely comparable settings to check them[8]. Sometimes the assumptions do have testable implications that can be exploited in the sense that if the implications do not hold then the assumptions can be falsified, but vice versa they cannot be verified. In Figure 2(i) we can test, for instance, if $Y \perp\!\!\!\perp X_1 \,|\, (A_1, X_2)$, which would allow us to rule out X_1 for the adjustment set. This strategy can be generalized, but must still be considered with care as a non-significant statistical test does not guarantee that there is no dependence (De Luna et al, 2011; VanderWeele and Shpitser, 2013; Witte and Didelez, 2019). Tennant et al (2021) discuss possibilities of empirically scrutinizing the assumptions encoded in a causal DAG. Further, the IV assumptions impose empirical constraints on the observables in form of bounds on certain frequencies when all variables are discrete; these can and should be checked to support an IV analysis (Balke and Pearl, 1994; Palmer et al, 2011; Guo et al, 2023; Sachs et al, 2023).

Bibliographic notes: More details on the backdoor criterion and a complete characterization can be found in Perković et al (2018). Also, when there is a choice of sufficient adjustment sets, we can graphically determine which one is the most efficient in the sense that it leads to most precise estimation (Witte et al, 2020). For instance, in Figure 2(i) it is more efficient to adjust for X_2 than for X_1. A different graphical structure is exploited by the frontdoor criterion which yields identification under unmeasured confounding if a perfect unconfounded mediator of the treatment effect is given (Pearl, 1995). The frontdoor criterion has rarely been used in practice, so far, but Piccininni et al (2023) provide a convincing real example investigating the causal effect of deploying mobile stroke units on patients' 3-month functional outcomes. Graphical rules dealing with identification of joint or dynamic causal effects under treatment-confounding feedback can be found in various versions, an early reference being Pearl and Robins (1995); see also Dawid and Didelez (2010). Natural experiments other than IVs, with different model assumptions, are exploited by designs such as difference-in-differences (Kennedy-Shaffer, 2024), regression discontinuity (Geneletti et al, 2015), or interrupted time-series (Linden, 2018).

4 Principles of Causal Effect Estimation

After determining the causal research question of interest (possibly by referring to a target trial), formalizing it as a causal estimand, and referring to a causal model with which identifiability assumptions can be derived or justified, the next step is the statistical estimation. This is essentially a statistical task, but the field of causal inference has produced specialized methods to tackle the specific challenges involved, especially for sequential treatments. We restrict the exposition to the most basic principles with references to the vast literature on this topic. A useful practical introduction including software is, e.g., given by Brumback (2021).

4.1 Inference under Exchangeability

A first intuitively plausible approach to causal inference under the 'no unmeasured confounding' assumption is direct matching on the covariates in the adjustment set (Rosenbaum, 1989). Matching constructs an independence between treatment and covariates, i.e., balances the covariates, in the matched sample (see Figure 5(i)). While attractive as a nearly non-parametric approach, this quickly becomes difficult with continuous or other high dimensional covariates requiring some bias-variance trade-off, and faces special challenges with time-dependent treatments (Stuart, 2010). Therefore we address, here, the basic principles underlying alternative approaches to account for (possibly time-dependent) confounding (for an excellent overview see Daniel et al (2013)).

First, we note that under causal consistency, conditional (sequential) exchangeability and positivity, average causal effects of point and sequential treatments, respectively, are identified. The former is identified by equation (1) and the latter by equation (2). These equations suggest estimating $P(Y = y \,|\, A_1 = a_1, \mathbf{X} = \mathbf{x})$ or $P(Y = y \,|\, A_1 = a_1, A_2 = a_2, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2)$ via regression models and plug these into the respective formulas, where we still need to average over the respective covariate distributions either empirically or via modeling. Such approaches adjust for confounding by incorporating the adjustment set into a *model for the outcome*. Alternative approaches incorporate the covariates into a *treatment model*. The underlying rationale

[8] However, such an empirical verification is in principle feasible, unlike cross-world assumptions mentioned earlier which are not even in principle testable (Robins and Richardson, 2011).

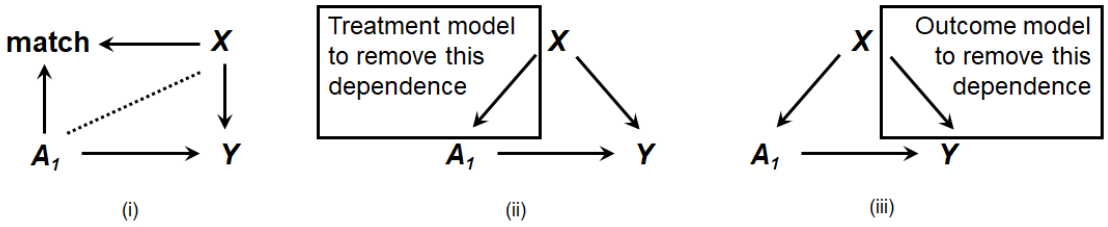

Fig. 5 Illustration of estimation principles under exchangeability for a point treatment A_1 with outcome Y and sufficient adjustment set $\mathbf{X}$. (i) Matching to balance $\mathbf{X}$. (ii) 'Remove' $\mathbf{X} \longrightarrow A_1$ by using a treatment model. (iii) 'Remove' $\mathbf{X} \longrightarrow Y$ by using an outcome model.

is that equation (1) is equivalent to

$$P(Y(a_1) = y) = \sum_{\mathbf{x}} \frac{P(Y = y, A_1 = a_1, \mathbf{X} = \mathbf{x})}{P(A_1 = a_1 \,|\, \mathbf{X} = \mathbf{x})} \tag{3}$$

and equation (2) is equivalent to

$$P(Y(a_1, a_2) = y) = \sum_{\mathbf{x}_1, \mathbf{x}_2} \frac{P(Y = y, A_1 = a_1, A_2 = a_2, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2)}{P(A_1 = a_1 \,|\, \mathbf{X}_1 = \mathbf{x}_1) P(A_2 = a_2 \,|\, A_1 = a_1, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2)}. \tag{4}$$

These alternative identifying formulas motivate why the treatment models given past covariates, which occur in the denominators, are useful for causal inference. Figures 5(ii,iii) illustrate the two basic principles of estimation under exchangeability (conditionally on $\mathbf{X}$), exploiting a treatment model or an outcome model. In the remainder, we address some of the ways these two principles can be put into practice for the estimation of causal effects.

The propensity score. In case of a binary point treatment, $A_1 \in \{0, 1\}$, the quantity

$$\pi(\mathbf{x}) = P(A_1 = 1 \,|\, \mathbf{X} = \mathbf{x})$$

is known as the *propensity score (PS)* (Rosenbaum and Rubin, 1983). The PS can be used in multiple ways for causal inference because it has the desirable property of being a so-called *balancing score*, i.e.,

$$A_1 \perp\!\!\!\perp \mathbf{X} \,|\, \pi(\mathbf{X}).$$

This property implies that given $\pi(\mathbf{x})$ the possibly high-dimensional adjustment set $\mathbf{X}$ carries no further information on the treatment, i.e., $\mathbf{X}$ are conditionally balanced between treated and control group given units with the same PS. This is exploited by methods such as PS matching or PS stratification (Stuart, 2010; Goetghebeur et al, 2020). It is also implicit in the use of the PS for *inverse probability of treatment weighting (IPTW)*, whereby the treated are weighted with $1/\pi(\mathbf{x})$ and the controls with $1/(1 - \pi(\mathbf{x}))$ thus making treatment independent of $\mathbf{X}$ in the weighted population (Robins et al, 2000).

In practice, the PS is usually not known and must be estimated from data. It is popular to use a simple logistic regression to estimate the PS, but more flexible models have also been suggested (Watkins et al, 2013; Shortreed and Ertefaie, 2017). Of note, the aim of modeling the PS is not to predict treatment as accurately as possible, but to achieve balance of the covariates in the adjustment set in the matched or weighted population (Sjölander, 2009; Waernbaum, 2012; Naimi et al, 2021). Including too many unnecessary predictors in the model for π, e.g., IVs like Z in Figure 3(iii), can even lead to *bias amplification* (Greenland and Pearl, 2011; Stokes et al, 2022)

PS methods are popular, presumably due to their simplicity and transparency but also because they are easily combined with standard methods. For instance, after matching treated and controls on the PS, the remaining analysis can proceed almost as if the matched sample came from an RCT (standard errors need to be considered (Austin and Small, 2014)), or any survival curve estimator can be adjusted for baseline confounding by IPTW (Cole and Hernán, 2004; Denz et al, 2023). However, some of the simplicity of PS methods is lost when it comes to continuous, multiple and especially time-dependent treatments. As addressed below, IPTW becomes more complicated in the latter case.

Diagnostic checks as in Figure 4 are often based on the treatment and covariates, or use the PS, and do not involve the outcome. Positivity and amount of overlap can be checked by simply considering the covariate distribution by treatment groups. Further, a basic impression of successful balancing by adjustment, e.g.,

by matching or weighting, is obtained by comparing the standardized mean differences in each covariate between the treatment groups before and after adjustment in a so-called *Love plot* (Love, 2002). However, it has to be remembered that this can only assess balance on measured covariates, not on unmeasured ones. The assumption of conditional exchangeability itself cannot be checked by such plots.

Sequential treatments: IPTW and marginal structural models. Suppose we want to model the joint effect of two sequential treatments (A_1, A_2) on an outcome Y under the causal model of Figure 3(i). We may wish to specify a (semi-parametric) model $\mu(\theta)$ for $E(Y(a_1, a_2))$, e.g., a linear, logistic, or proportional hazards regression model. Such models are called *marginal structural models (MSMs)* because they model the effect of (sequential) interventions on (A_1, A_2) marginally over the time-dependent covariates $\mathbf{X}_2$ (Robins et al, 2000; Havercroft and Didelez, 2012). The representation given in equation (4) suggests the following approach: Fit $\mu(\theta)$ to the *weighted* data using, for each individual, the denominator of equation (4), i.e., the inverse of the product of the treatment models given past covariates

$$\text{weights } = \{P(A_1 = a_1 \,|\, \mathbf{X}_1 = \mathbf{x}_1)P(A_2 = a_2 \,|\, A_1 = a_1, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2)\}^{-1},$$

where $(a_1, a_2, \mathbf{x}_1, \mathbf{x}_2)$ are the values actually observed for that individual. The estimator $\hat{\theta}$ obtained by fitting $\mu(\theta)$ to the weighted data can be shown to consistently estimate the parameter θ if the MSM is correctly specified. In practice, the weights are usually not known and need to be estimated. For consistent estimation, the treatment models must then also be correctly specified. In particular, in the weighted data, the covariates must be balanced at all treatment times which can be checked empirically by suitably modifying the Love plots (Jackson, 2019). Due to the multi-step estimation, standard errors are typically computed via individual-level bootstrap (Enders et al, 2018).

There are some practical problems with (sequential) IPTW, e.g., issues with extreme weights when the denominator is near zero. This is often related to (near) violations of the sequential positivity assumption. Thus, MSMs tend to yield rather unstable inference, with large standard errors, alerting the analyst that there are subgroups for which there is little information in the data to compare the treatment sequences of interest. As a remedy, it is common to use *stabilized weights* which additionally reweight according to the marginal treatment distribution in the population (Hernán et al, 2000). MSMs can be used for non-binary sequential treatments, but due to the stability issues such applications are rare in practice. Improved performance is obtained by augmenting the IPTW estimator yielding a doubly robust estimator (Orellana et al, 2010); see below.

There have been numerous practical applications of MSMs and IPTW for time-dependent treatments in epidemiology over the last two decades. The earliest examples investigated immediate versus delayed initiation of antiretroviral therapy for HIV patients before RCTs were available (Hernán et al, 2000; Sterne et al, 2005; Cain et al, 2011) and are still used today, e.g., in a South African HIV cohort study (Bell Gorrod et al, 2020). A recent psychosocial application of MSMs investigates the longitudinal effects of community violence exposure on youths' internalizing and externalizing symptoms (Kennedy et al, 2023).

Sequential treatments: The parametric g-formula. Consider again the estimation of the joint effect of two sequential treatments (A_1, A_2) on an outcome Y under the causal model of Figure 3(i). Now, we do not specify an explicit model for $E(Y(a_1, a_2))$, but make use of the representation given in equation (2), instead. An obvious approach would be to posit and fit models for all 'ingredients' of this formula, including in particular the possibly high-dimensional conditional distribution $P(\mathbf{X}_2 \,|\, A_1, \mathbf{X}_1)$. All these are then combined for the estimation of $E(Y(a_1, a_2))$ under those treatment sequences (a_1, a_2) of interest, e.g., 'never treat' or 'always treat'. This idea forms the basis of the *parametric g-formula* approach (Taubman et al, 2009). It is 'parametric' as it relies on fully parametric models for all time-varying covariates in $\mathbf{X}_2$ and for the outcome, Y, while MSMs are semi-parametric models. Note that in realistic cases with more than two time points, there are also more time-varying covariates to be modeled. Moreover, in many typical realistically complex applications of the g-formula there are no explicit ways of obtaining $E(Y(a_1, a_2))$ so Monte-Carlo approximations are used instead (Young et al, 2011). This means that simulated data are generated from the fitted models, under the specified hypothetical treatment sequences- The simulated outcomes can be seen as predicted POs $\hat{Y}(a_1, a_2)$ from which summaries and estimates of causal contrasts are computed. In addition, and as a check of model fit, the so-called 'natural course' under no intervention can be estimated with the g-formula and should approximate the observational outcomes' average trajectory.

The g-formula is straightforward to adapt for more complex, especially dynamic hypothetical interventions. For instance, we may want to compare an intervention that increases physical activity (PA) only for those individuals and at those time points where the 'natural' level is below some threshold. Hence, we first need to predict the natural value (under past hypothetical treatments) and then 'change' it if necessary. Alternatively, an intervention could be considered that shifts the natural level of PA. Similarly, a hypothetical strategy

could determine the dosage of a treatment as a function $g(\cdot)$ of some recently measured blood parameter. Such an adaptive strategy would require restricting equation (2) to values of the second treatment, say $a_2 = g(x_2)$, that correspond to the desired hypothetical treatment rule. It is also possible to represent uncertainty about the exact treatments under an intervention; for instance, a delayed treatment may not occur deterministically at a specific point in time, but during some plausible interval. Such more general treatments can be incorporated by, e.g., an additional term $\tilde{P}(A_2 = a_2 \,|\, \mathbf{X}_1, \mathbf{X}_2)$ in equation (2) representing a stochastic intervention on A_2. The resulting extension is called the *generalized g-formula* (Young et al, 2014). Hypothetical interventions that shift, possibly stochastically, the natural value of treatment have the advantage of not requiring full positivity because they remain 'closer' to the observed treatment patterns. Otherwise, the g-formula relies on the parametric models to extrapolate over treatment-covariate values with positivity violations which can yield misleading results. Positivity should therefore be checked in a separate step when using the g-formula.

Despite the seemingly hopeless task of correctly specifying a very large number of models for all time-dependent covariates, an increasing variety of convincing analyses using the parametric g-formula can be found in epidemiological research. Chiu et al (2021) investigate the 20-year risk of all-cause mortality under the sustained implementation of food-based goals of the American Heart Association using data from three observational studies, carefully discussing the strong assumptions underlying their analysis. Börnhorst et al (2023) employ the parametric g-formula to assess the effect of hypothetical behavioral interventions in early childhood on the 13-year risk of overweight/obesity based on a children's cohort study. In a sensitivity analysis addressing ambiguity about the causal model in view of large intervals between the study waves, the authors could show that results remained largely robust towards different choices of lagged dependencies.

Doubly robust estimation and double/debiased machine learning. In view of the instability of IPTW, on the one hand, and the parametric assumptions of the g-formula, on the other hand, it is possible to combine the two approaches. Indeed, a special class of estimators combines the treatment model and the outcome model, and these estimators have the property of *double-robustness* (Bang and Robins, 2005): It suffices that either the treatment or the outcome model are correctly specified for the estimator to be (statistically) consistent for the causal estimand.

In view of the options to use either the treatment model, as for IPTW, or the outcome model, as for the g-formula, one may ask if they can be chosen in a data-driven manner in order to reduce the risk of misspecification. For instance, machine learning algorithms such as random forests allow non-linearities and interactions without prespecification. However, it can be shown that simple methods of causal inference when combined with data-driven model selection do not perform well; this is because they need to be regularized to prevent over-fitting, which in turn leads to bias specifically when covariates are very imbalanced in some subgroups (Naimi et al, 2021). In contrast, exploiting machine learning approaches to fit both models for doubly robust estimators leads to a much better performance. The approach is known as double/debiased-machine learning (Chernozhukov et al, 2018; Knaus, 2022) and is also implicit in targeted learning (van der Laan and Rose, 2011).

Bibliographic notes: Popular alternatives for dealing with extreme weights in IPTW are 'overlap weighting' (Li et al, 2019) or generalized balancing scores (Chattopadhyay et al, 2020); see overview by Matsouaka et al (2024). Moreover, different estimands based on shifting the PS have been proposed to deal with positivity issues (Kennedy, 2019) or to extend IPTW for continuous treatments (Kennedy et al, 2017).

While most of traditional causal inference deals with average treatment effects and its variants, much interest and work has recently emerged for heterogenuous treatment effects especially in connection with machine learning (Wager and Athey, 2018; Künzel et al, 2019); see overview by Shah et al (2021). We have not touched on so-called *structural nested models*, which are used especially in survival analysis, and *g-estimation* which is closely related to doubly robust estimation (Robins et al, 1992; Vansteelandt and Sjölander, 2016).

4.2 Inference without Exchangeability

Assuming conditional exchangability, as in the previous section, amounts to believing that the information in the data is, in principle, sufficient to fully adjust for confounding. In some settings, this might be a 'heroic' assumption. The social sciences, e.g., econometrics, psychology, or sociology tend to not find exchangeability defensible for their typical observational data sources, such as panel data, so that other principles for the identification and inference for causal effects are pursued. Here, we address IVs as one example to give the reader a flavor, but various other designs can be found. Introductions to IVs for epidemiology are given by

Greenland (2000) and Hernán and Robins (2006).

IV: Testing the null and bounding causal effects. The IV assumptions detailed in Section 3.3 (see Figure 3(iii)) do not fully identify causal effects under unmeasured confounding. However, they are sufficient to test the null hypothesis of no causal effect (Didelez and Sheehan, 2007): Under the IV assumptions, if we find that there is no association between the IV and the outcome then there is no causal effect of the treatment A_1. The principle of using an IV as an indication for the presence of a causal effect is, in fact, at the heart of the original idea by Katan (1986) when formulating the first idea of Mendelian randomization. As a slight caveat, we have to be aware that 'no causal effect' can also occur if positive and negative effects in different strata of the unmeasured confounding U cancel each other out.

Further, when all variables are discrete, bounds on the average causal effect can be derived using an IV (Balke and Pearl, 1994; Palmer et al, 2011; Sachs et al, 2023). Lower and upper bounds for a causal effect mean that the observable data are consistent with any causal effect value in this interval, the unmeasured confounding prevents us from pin-pointing one single value (not be confused with confidence intervals which are about statistical uncertainty). Moreover, a similar principle leads to an empirical test of the IV conditions which can sometimes be falsified (Guo et al, 2023).

IV: The Wald ratio estimator. A simple semi-parametric assumption motivates a popular IV-estimator. Supplementing the causal model of Figure 3(iii) by the *additive structural mean model*

$$E(Y - Y(0)|A_1 = a_1, Z) = \beta a_1,$$

where β is the ATT, an estimating equations for β is obtained by exploiting the fact that under the IV assumption, we must have $Y(0) \perp\!\!\!\perp Z$. This principle of estimation is known as *g-estimation* (Hernán and Robins, 2006). In the present case, it can be shown that $\beta = \mathrm{Cov}(Y, Z)/\mathrm{Cov}(A_1, Z)$ (Wooldridge, 2010; Didelez et al, 2010b). A consistent estimator for the parameter β is thus given by the ratio of coefficients from a linear regression of Y on Z and of A_1 on Z. This procedure has desirable statistical properties of robustness and efficiency (Vansteelandt and Didelez, 2018), and it can also be used when data on the outcome and the treatment are not jointly available, e.g., in two-sample Mendelian randomization (Lawlor, 2016).

There are numerous extensions and variations on the principle of the Wald ratio estimator which has different interpretation under different assumptions (Didelez et al, 2010b). For instance, it has been extended to the case where multiple candidate IVs are available, but not all of them are valid (Bowden et al, 2015). Returning to our general message that treatments (or exposures) in epidemiology rarely occur at a single point in time, Morris et al (2022) suggest an interpretation of the Wald ratio estimator (or its estimand) when A_1 is just a single measurement of a time-dependent treatment process. The authors augment the underlying causal model to include a latent 'liability' through which the IV affects the treatment process over time, so that the effect of a hypothetical intervention on this liability becomes the target of inference. Their data example addresses the effect of BMI on blood-pressure: The authors consider hypothetically intervening on the liability in a way such that average BMI is increased by one standard deviation at the age of 53 years, and this is estimated to increase mid-life blood pressure by 12.78 mmHg. However, further work is still needed to develop robust and practical IV-based approaches for time-dependent treatments in epidemiology, e.g., for life course studies.

Other natural experiments. The presence of an IV can be seen as a natural experiment where the IV represents an imperfect randomization which partly but not fully determines the treatment. Other study designs using natural experiments are difference-in-differences (DiD) and its generalization, the synthetic control method for comparative case studies, regression discontinuity, or interrupted time-series. These approaches exploit changes of exposures, treatments, or policies that are not simultaneously adopted across time or space. An early example was an analysis prompted by the introduction of tobacco control through a tax in California, without such a tax being introduced in other states (Abadie et al, 2010). Another example exploited that during the COVID pandemic, different regions adopted the face-mask policy at different points in time allowing the identification of causal effects under other assumptions than exchangeability (Mitze et al, 2020). The methods still require some form of 'common trend' assumption. While promising, one possible downside of such designs is that the implicit or explicit causal estimands tend to be very specific to the considered settings and the generalizability of study findings might be unclear. Methods based on natural experiments are increasingly used in (pharmaco)epidemiology (Bonander et al, 2021; Kennedy-Shaffer, 2024).

Quantitative bias analysis. Suppose a causal analysis using methods from Section 4.1 has been carried out but some doubts exist about residual confounding that may not have been accounted for. The presence of unmeasured confounding can be investigated by using so-called negative controls. For instance, if

an alternative outcome variable is available, which cannot possibly be causally affected by the exposure or treatment, but would be affected by the same (unobserved) confounding factors, then this helps detect or rule out residual confounding; for an overview see Lipsitch et al (2010). A variety of methods have been proposed, known as quantitative bias analysis, to obtain a correction of a biased estimate. The correction terms or factors require some assumptions, e.g., based on domain knowledge, about the unmeasured confounder(s) so that the confounding can be 'removed' from the estimate (Fox et al, 2021) (cf. Chapter Sensitivity Analysis and Bias Analysis by S. Greenland in this book). Related to this, the so-called *e-value* quantifies how strong confounding bias by unmeasured factors would need to be so as to 'explain away' an estimated causal effect (VanderWeele and Ding, 2017). For a critical discussion of the e-value see Sjölander and Greenland (2022).

Bibliographic notes: Useful introductions to IVs in the context of Mendelian randomization are given by Didelez and Sheehan (2007), to quasi-experimental methods for pharmacoepidemiology by Kennedy-Shaffer (2024), and to quantitative bias analysis by Fox et al (2021). The reporting of IV analyses is addressed by Davies et al (2013) or Skrivankova et al (2021). *Triangulation* is advocated as a way of obtaining more reliable answers to research questions through integrating results from several different approaches, where each approach has different key sources of potential bias that are unrelated to each other; see overview by Lawlor et al (2016). Moreover, a promising new avenue for dealing with unmeasured confounding is *proximal inference*, where proxies for the unmeasured common causes are systematically exploited; an introduction is given by Tchetgen Tchetgen et al (2020).

4.3 Causal Interpretation of Multivariable Regressions

In most introductory courses to statistical methods, multivariable regressions play a key role. It is therefore worth addressing the question when, if at all, these yield valid causal inference and which causal estimands can be obtained. This is also discussed in Chapter on Regression Methods for Epidemiological Analyses by S. Greenland in this book.

Regression adjustment for a point treatment. Consider a single point treatment A_1 and an outcome Y. It is common to adjust for covariates $\mathbf{X} = (X_1, \ldots, X_p)$ in a multivariable regression, which is essentially a model for $E(Y \mid A_1, X_1, \ldots, X_p)$, say, a linear or logistic regression. With the causal concepts, assumptions, and methods covered so far, we can conclude that this regression describes the conditional treatment effect (*CATE*) of A_1 on Y given $\mathbf{X}$ if the model is correctly specified, causal consistency and positivity hold, and $\mathbf{X}$ is a sufficient adjustment set; in particular these covariates must not themselves be affected by A_1. This *CATE* may or may not correspond to a single coefficient in the regression model depending on whether there is effect modification by $\mathbf{X}$. If there is no effect modification and the model is a linear regression, then the coefficient of A_1 equals the average causal effect (*ACE*). Note that this is not the case in a logistic or proportional hazards regression due to lack of collapsibility (Greenland et al, 1999; Daniel et al, 2021).

In comparison to the formal causal estimation approaches, we note the following differences: When estimating the *ACE*, regardless of effect modification or not, we would additionally average over the covariate distribution (see g-formula, equation (1)) and thus obtain a single parameter estimate, or we estimate directly the causal effect by PS matching or IPTW without constructing a model for how Y depends on $\mathbf{X}$. The latter is sometimes seen as an advantage because misspecification of the dependence of Y on the possibly high-dimensional $\mathbf{X}$ can bias the causal effect estimation, while with PS matching or weighting the empirical checks for balance are, arguably, simpler.

Regression adjustment for sequential treatments. Consider now two sequential treatments (A_1, A_2) and sets of covariates $\mathbf{X}_1$ and $\mathbf{X}_2$. Assume causal consistency, conditional sequential exchangeability as in Figure 3(i), and positivity. Here, a (correctly specified) multivariable regression, say, a model for $E(Y \mid A_1, A_2, \mathbf{X}_1, \mathbf{X}_2)$, does *not* typically have an immediate causal interpretation due to the treatment-confounding feedback problem, whereby $\mathbf{X}_2$ confounds A_2 and Y while being itself affected by A_1. Thus, a multivariable regression does not correctly identify the joint effect if treatment-confounding feedback is present. Such a single regression does not correspond to equation (2) which would identify the joint effect in such a situation as long as sequential exchangeability given measured time-dependent covariates holds.

We could impose stronger assumptions than sequential exchangeability. Consider, for instance, the special case that U is empty in the causal model of Figure 3(i). A regression $E(Y \mid A_1, A_2, \mathbf{X}_1, \mathbf{X}_2)$ then describes the conditional association of Y and A_1 given $(A_2, \mathbf{X}_1, \mathbf{X}_2)$, which would be informative for the conditional controlled direct effect of A_1 on the outcome fixing both mediators $(A_2, \mathbf{X}_2)$. The regression further describes the conditional association of Y and A_2 given $(A_1, \mathbf{X}_1, \mathbf{X}_2)$, which would be informative for the conditional

causal effect of A_2 on the outcome. The multivariable regression would not describe a joint effect of (A_1, A_2) as this needs to allow the effect of A_1 to 'flow' through $\mathbf{X}_2$. Even more restrictive, we could assume that $\mathbf{X}_2$ is empty, i.e., that any confounding between A_1 or A_2 and Y is fully captured by baseline covariates $\mathbf{X}_1$, even though this will be implausible in many longitudinal settings. However, in this special case the multivariable regression does contain information on the joint conditional effect of (A_1, A_2) given $\mathbf{X}_1$, or on the conditional controlled direct effect of A_1 controlling for A_2 and the conditional average treatment effect of A_2 given $(A_1, \mathbf{X}_1)$. Both interpretations are not symmetric in A_1 and A_2 due to the fact that A_1 may affect A_2 (and $\mathbf{X}_2$) but not vice versa; if we wanted the total effect of A_1 (possibly given $\mathbf{X}_1$) we would need to omit A_2 (and $\mathbf{X}_2$) from the model. The intricacies of the correct causal interpretation of a multivariable regression under specific assumptions are linked to the 'table 2 fallacy' described by Westreich and Greenland (2013).

Comparing multivariable regression with the above formal causal methods, the main difference is that with time-dependent IPTW or with the g-formula we do not need to assume that U or $\mathbf{X}_2$ in Figure 3(i) are empty for the identification of effects of joint interventions. Moreover, we have the possibility to combine the outcome model and the treatment model for doubly robust estimation. Nevertheless, multivariable regressions are still key ingredients in the estimation of causal effects, e.g., when using the g-formula of equations (1) or (2). But they are then used as a mere tool which must be complemented by further computations to obtain the targeted causal parameter.

5 Target Trial Emulation

In Section 2.1 we introduced the concept of a target trial to help elicit and communicate transparently one's causal research question. It requires specifying the following trial components: Eligibility, treatment strategies, assignment to treatment strategies, length of follow-up, outcome of interest, causal contrast and analysis plan. The next step is to design the *emulation* of the target trial (TTE) with available data. For this we need to consider how these target trial components are to be implemented with available data.

The methods of Section 4 address one aspect: How to emulate randomization by adjusting for (or otherwise dealing with) confounding. There are many methods to choose from for the emulation of randomization, and a very large part of methodological research in causal inference is dedicated to refining, generalizing, and improving robustness of such methods. We stress that the careful emulation of the *other* components of a target trial are just as important. While residual confounding is always a threat in observational studies, other sources of bias are, in contrast, often avoidable or can be minimized by choosing a suitable design for the emulation. In this section, we therefore consider those other aspects of TTE and common techniques to minimize design-related biases.

5.1 Basic Techniques to Avoid Self-Inflicted Bias

Here we address a selection of principles and techniques that can be found in the context of emulating a target trial. However, this is not a complete list, nor are these techniques confined to TTE.

Eligibility, treatment strategies and positivity. The population of interest is defined by the inclusion and exclusion criteria. The emulation step determines how this can be implemented with the available data. In view of the earlier structural assumption, an important point to note is that eligibility should be defined so that positivity for all treatment arms is plausible, i.e., for everyone in the eligible population it should be possible to be assigned to any of the trial arms. Moreover, the available database should provide empirical support for positivity. This can be achieved by prior descriptive analysis of, e.g., drug-user profiles illustrating typical drug usage over time. For instance, it would not be sensible to define a trial arm with 'five years of continual use of drug ABC' if nobody in the database ever used the drug for more than three years. For an example consider the analysis by Börnhorst et al (2021), where second line treatment for type 2 diabetes patients is investigated and the trial arms had to be adapted to the actual usage so as to have sufficient empirical support for a comparison.

Alignment at time-zero. Any real trial has a clear baseline (time-zero) at the point when randomization takes place. For every individual in an RCT, it is therefore unambiguous which measurements were before and which were after time-zero. As no actual randomization or assignment to a trial arm occurs in observational data, time-zero needs to be determined as part of the emulation design. There are essentially two cases,

Table 1 Potential limitations of study designs: RCTs versus observational studies with or without explicit TTE (adapted from Braitmaier and Didelez (2022)).

Potential limitation	***RCT***	***Observational studies with RWD***	
		with TTE	***without TTE***
Prevalent user bias (Ray, 2003)	Low risk: Randomization marks treatment start	Low risk: Avoided by alignment at time zero	High risk, if exposure assignment uses pre-baseline information
Immortal time bias (Suissa, 2008)	Low risk: Assignment to arms with randomization	Low risk: Avoided by alignment at time zero	High risk, if exposure assignment uses post-baseline information
Unclear research question (Didelez, 2016)	Low risk in both, due to definition of (hypothetical) interventions determining treatment arms		High risk, if no explicit intervention on exposure described
Baseline confounding	Low risk: Avoided by baseline randomization	High risk in both; can be corrected for by appropriate adjustment (if data are sufficiently informative)	
Time-dependent confounding (Hernán and Hernández-Díaz, 2012)	High risk in all, if (differential) loss to follow-up, treatment switching (non-adherence), artificial censoring, among others; can be adjusted for if data are sufficiently informative (e.g., using inverse probability weights, g-formula)		
Lack of external validity	High risk: Trial population is highly selected and possibly differs substantially from target population	Depends on data source: Registry or claims data contain information on subgroups routinely excluded from RCTs and are more informative for real-life use or exposure. However, volunteer bias (among others) can occur in other data sources	
Economic and time costs	High risk	Low risk in both, if existing (routine) data can be used	

(i) there is a natural unique time-zero for each individual, or (ii) time-zero is not unique for some or all individuals.

Case (i) occurs, for example, when comparing two treatment options that are both indicated at a fixed time, such as at first diagnosis or at a specific progression stage. In the analysis of Börnhorst et al (2021), time-zero is at diagnosis of type 2 diabetes in the target trial and at the first recorded dispensation of first-line antidiabetic treatment in the emulation, from which point onwards patients were followed with respect to the timing of their second-line treatment.

Case (ii) occurs when the treatment or the control arm definitions are not linked to a unique point in time. For example, Braitmaier et al (2022) consider the effect of screening colonoscopy which is offered in Germany any time from the age of 55 years onwards. Moreover, this is to be compared with no attendance at a screening colonoscopy which is also not specific to any particular time or age. It is tempting to declare the 'exposed' as those who attend a screening colonoscopy at some point during their lifetime and as 'not exposed' those who never attend a screening colonoscopy; but this would clearly induce *immortal time bias* (Suissa, 2008; Hernán et al, 2016). The issue is that for such an exposure definition we would 'look into the future' and violate the requirement that treatment assignment occurs at time-zero. Instead, one may opt for constructing a unique time-zero, e.g., when the eligibility criteria are met for the first time. Or, alternatively, it is common to design a *sequence of emulated trials* and allow individuals to enter multiple trials, which could be pooled in the analysis. Using such a sequence has the advantage of efficiently making use of all the data as illustrated in an empirical comparison by García-Albéniz et al (2017). A correct assessment of standard errors can be achieved by individual-level bootstrap. The biases induced by designs without alignment at time-zero are further illustrated by Braitmaier et al (2024) with the example of screening colonoscopy.

Artificial censoring, cloning and weighting. A related issue occurs when comparing sustained or dynamic treatment strategies, where it may not be obvious how to achieve alignment at time-zero. Consider, for instance, a (control) strategy of never attending screening colonoscopy; it may seem that we need to look into the future of a 55-year old patient's data in order to asses that they comply with this strategy. However, this could re-introduce bias (García-Albéniz et al, 2017). One technique to address this problem is that of *artificial censoring*: Each individual is assigned to, and retained on, any treatment strategy arm(s) as long as their history complies with the definition of that strategy, but once they deviate from the strategy they are artificially censored. 'Artificial' refers to the fact that data on the individual may still have been measured beyond that point in time, but as far as the strategy in question is concerned they are censored.

For instance, a person who becomes eligible for screening colonoscopy at the age of 55, but does not attend such a procedure until the age of 60 years is retained in the treatment arm of sustained 'never-screen' for those 5 years, together with the information about any diagnosis of colorectal cancer during that time. But they are artificially censored for this strategy at the age of 60 years because from then onwards they do not contribute any information to the 'never-screen' strategy anymore. As there may be reasons for the deviation, artificial censoring is not 'random' and must be accounted for by weighting to avoid re-introducing selection bias. Screening attendance, for example, could have been prompted by a general check-up after a diagnosis of another disease. Avoiding selection bias due to artificial censoring requires re-weighting of the uncensored individuals. The weights are given as the inverse probability of adherence to the strategy in question conditioned on relevant past covariate information. These past covariates need to be sufficient to reasonably assume that in the re-weighted population artificial censoring is like randomized. The inverse weighting for artificial censoring is analogous to the IPTW for sequential treatments (Hernán, 2018b), i.e., it relies on the structural assumption of conditional sequential exchangeability and any time-dependent confounding must be accounted for.

For the comparison of particular treatment strategies, an individual's data might also be compatible with *multiple* trial arms. Hernán (2018b) considers the question of how to assess the effect of 'duration of treatment' on a survival outcome. Obviously those who take treatment for longer live longer or, vice versa, one has to live longer to be in a position to take the treatment for longer — again, a potential source of immortal-time. Thus, when comparing short (e.g., 6 months) and long (e.g., 12 months) duration of treatment, all patients who take treatment for 8 months, contribute to both arms until 6 months and would then be artificially censored for the first arm; they would also be artificially censored at 8 months for the second arm, while uncensored patients are re-weighted. The 'trick' of assigning an individual to all strategy arms with which they comply for as long as possible is a data augmentation technique known as 'cloning' (Hernán, 2018b). It makes full use of the available data and avoids bias which could otherwise be introduced by looking into the future and systematically assigning certain patients to one and others to the other arm. Alternatively, cloning can be avoided by randomly assigning individuals to a single strategy out of those with which they comply (without taking their future data into account). As before, the artificial censoring at the point when an individual deviates from an strategy must be accompanied by suitable inverse weighting of those individuals who are not censored with the conditional probability of adhering given past covariates. This can be thought of as up- or down-weighting the uncensored individuals according to how representative they are for those artificially censored ones.

The above 'clone-censor-weight' approach is limited to comparisons of relatively simple and few treatment strategy arms and to survival or time-to-event outcomes. For more complex dynamic treatments, there may not be sufficient information left if artificial censoring occurs early. In such situations, the g-formula could be a feasible alternative.

Emulation for complex treatment strategies. Let us consider the study by Chiu et al (2021) as an example for a more complex TTE. The authors investigate the 20-year risk of all-cause mortality under the sustained implementation of several food-based goals of the American Heart Association using data from three prospective observational studies. The goals were represented in 14 different treatment strategies, separate and joint ones on, e.g., minimum servings of fruit and vegetables and maximum servings of processed meat etc. As mentioned earlier, the strategies depend on the natural value: The intervention is only applied if and when the individual naturally consumes insufficient fruits and vegetables. To implement such rather complex strategies with observational data and to account for treatment-confounding feedback, the parametric (generalized) g-formula was used. Moreover, the target trial principle was especially useful in this context as it helped to make the analyses of the three cohorts as comparable as possible, while strong differences remained due to rather different study populations. The results showed that especially in the studies with older populations full adherence to all six dietary goals considerably reduced the mortality risk.

5.2 Some Caveats

TTE is sometimes confused or equated with trial *replication* (Wang et al, 2023). Trial replication typically refers to the replication of existing RCTs, where actual trial protocols exist and the challenge lies in emulating these as closely as possible with observational data. The aim is typically to identify databases and methods that yield similar results to RCTs. However, for replication the emulation will typically be confined to the observational analogue of an 'intention-to-treat' (ITT) analysis, i.e., the effect of a point treatment

representing *initiation* of treatment versus control[9]. This is because the randomization in RCTs can only guarantee unconfounded inference for the ITT contrast as non-adherence or post-randomization switching are not randomized. Moreover, replication necessarily needs to restrict the target population to be comparable to those in the RCT so that the effects can reasonably be expected to be similar.

In contrast, the strength and potential of TTE is that it offers an option, often the only one, for addressing important decision questions when RCTs are not (yet) available. Examples include the early studies on COVID vaccine effectiveness and safety (Dagan et al, 2021), or drug safety studies during pregnancy (Yland et al, 2022). Moreover, TTE does not need to be restricted to ITT contrasts; the clone-censor-weight approach or the parametric g-formula (and similar methods) allow us to consider sustained ('per-protocol') or more complex strategies. These methods could also be used with data from RCTs to obtain per-protocol effects of sustained treatment strategies if sufficient information on time-dependent covariates is recorded (Hernán and Hernández-Díaz, 2012).

For a convincing TTE there are high demands on data quality and informativeness. The data should provide sufficient information on the relevant baseline and time-dependent confounding. Alternatively, instrumental variables or other natural experiments need to be found. Typically, for many causal research questions longitudinal information at sufficiently frequent time points is crucial. However, even when such desirable data are not available, it is worthwhile to formulate and attempt to emulate a target trial: The process highlights possible sources of bias and will better inform future data collection. The above study by Chiu et al (2021), for instance, had to rely on cohort data with large time gaps between measurements; the authors carefully outlined the additional assumptions needed to address this issue pointing out that more detailed data are currently not available.

Finally, we re-emphasize that even though a carefully designed TTE using high-quality and detailed observational data can eliminate many types of design-based and time-related biases, it can never guarantee that there is no residual confounding bias. In Table 1, we contrast some of the strengths and weaknesses of RCTs versus observational studies with and without TTE.

Bibliographic notes: Early work using the idea of TTE, without using the term, was given by Hernán et al (2008); the term was later coined by Hernán and Robins (2016). An overview on how to structure and report a TTE is provided by Hansford et al (2023).

6 Causal Mediation Analysis

It comes very naturally to speak of direct and indirect (causal) pathways or effects, in epidemiology, and in life course epidemiology in particular. However, the formal treatment of (in)direct effects is surprisingly complicated, so we dedicate this section entirely to the topic of causal mediation. We aim to give the reader a flavor of the key issues, with no claim to mathematical precision for which we refer to the original articles; see the textbook by VanderWeele (2015).

6.1 Controlled Direct Effects

When we consider a causal DAG model such as Figure 6(i) we could say that treatment A_1 potentially has a direct effect symbolized by the edge $A_1 \longrightarrow Y$, but also a potentially indirect effect through A_2 symbolized by the path $A_1 \longrightarrow A_2 \longrightarrow Y$. Indeed, if we consider the strict meaning of a causal DAG, the edge $A_1 \longrightarrow Y$ indicates that under a hypothetical intervention holding fixed A_2, varying the value of A_1 still affects the distribution of Y (remember that the notion of 'direct' effect is relative to the nodes in the DAG or the variables in our causal model); moreover, the edge $A_1 \longrightarrow A_2$ means that manipulating A_1 by an intervention may affect the distribution of A_2 and the edge $A_2 \longrightarrow Y$ indicates that manipulating A_2 under a hypothetical intervention holding fixed A_1 may affect the distribution of Y. All of these interpretations refer to the notion of *controlled direct (causal) effect (CDE)* which is encoded in the edges of a causal DAG (Spirtes et al, 2000).

The *CDE* can be regarded as a special case of a joint intervention effect. Consider the PO for two treatments $Y(a_1, a_2)$. In this context, the total effect of A_1 on Y is a contrast of $E(Y(a_1 = 1, A_2))$ versus $E(Y(a_1 = 0, A_2))$ allowing the second treatment to adopt its natural value (which may depend on a_1).

[9] Here we neglect the fact that an individual may not even initiate the treatment to which they are assigned in an RCT due to non-adherence.

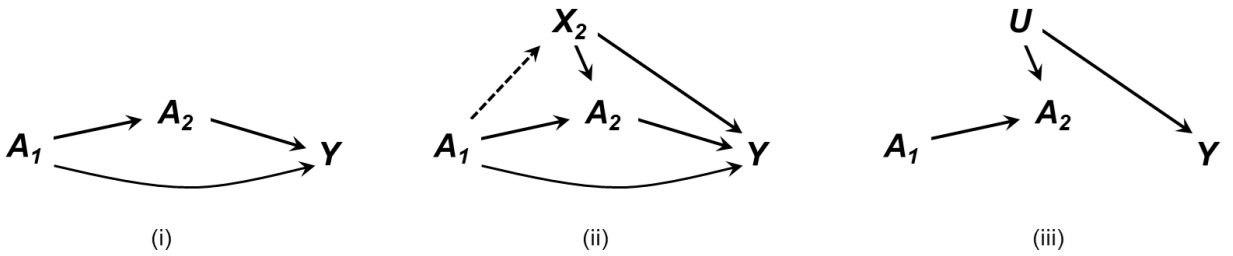

Fig. 6 DAGs illustrating causal mediation of treatment A_1 possibly mediated by mediator A_2. (i) Simple but unrealistic case of no confounding between any pair of the three nodes. (ii) Mediator-outcome confounding by X_2, possibly induced by treatment if dashed edge present. (iii) Under the null of no direct or indirect effect we still have $Y \not\perp\!\!\!\perp A_1 \mid A_2$ due to the collider at A_2.

Instead, a contrast between $E(Y(a_1 = 1, a_2))$ versus $E(Y(a_1 = 0, a_2))$ is the effect of intervening on A_1 not mediated through A_2 because we fix the latter by a second intervention at the constant value a_2 (see illustration in Figure 7(i)). Such an effect may be of interest if the research question aims to establish that, while A_1 may have a large total effect on Y, A_2 is thought to be the actual 'critical' element, which would be the case if there was little or no direct effect of A_1 after we intervene to fix A_2. For instance, we may have the hypothesis that diet affects certain health outcomes only through its effect on the gut microbiome in the sense that if a medication were available to keep the gut microbiome at a desirable balance, then different diets would not much affect those health outcomes anymore.

As a special case of a sequential joint intervention, the *CDE* is identified and estimation can proceed in similar ways as for sequential treatments. In equations (2) or (4) contrasts can be chosen where a_2 remains at the same value. It is worth emphasizing that the structural assumptions are also the same. In particular (conditional) sequential exchangeability is still required, i.e., confounding between the mediator A_2 and the outcome Y must be accounted for, including the case where this confounding is itself affected by A_1. The case of Figure 6(i) is rather unrealistic as this structure cannot be guaranteed even by randomizing treatment A_1. The structure of Figure 6(ii) is more realistic, where there is mediator-outcome confounding by X_2, for which we would need to adjust to obtain the *CDE*.

For further illustration, consider the DAG in Figure 6(iii), where there is no direct or indirect causal effect of A_1 on Y. Under this causal model, if the analyst inadvertently attempts to estimate the direct effect of A_1 by including the suspected mediator A_2 in a regression model, a non-causal association due to the collider at A_2 will typically bias their findings (Cole and Hernán, 2002). Here, with U being unmeasured, only the total effect of A_1 can be identified. Note also that in Figure 6(ii), the *CDE* of A_1 controlling for A_2 subsumes the possible indirect effect via X_2 if the latter is not also controlled for.

Structural equation models and path analysis. SEMs assume (more or less explicitly) functional dependencies for every *CDE* represented by a directed edge in a causal DAG (Pearl, 2009). As such, we can mathematically define any 'path-wise effect' from one node to another by a cascade of *CDE*s. Particularly simple formulas for path-wise effects are obtained in the case of linear SEMs. However, the methodology has attracted some criticism as it typically relies on restrictive and often implausible parametric modeling assumptions including, e.g., absence of interactions or effect modification. More importantly, it is not self-evident how to interpret the resulting objects as actionable causal effects in the sense of contrasts of hypothetical interventions (VanderWeele, 2015).

Beyond controlled direct effects? The notion of *CDE* has a clear practical interpretation in terms of hypothetical interventions. However, it also has certain features of which one has to be aware for a correct interpretation. For instance, the *CDE* of A_1 controlling for a possible mediator A_2 does not reflect if and how much the treatment A_1 affects the suspected mediator A_2: The *CDE* of A_1 on Y controlling for A_2 can be different from its total effect simply through effect modification by A_2. In other words, even without any causal effect of A_1 on A_2, and thus without any 'indirect effect', the total effect may not equal a controlled direct effect.

This brings us to a second issue, namely that there is no corresponding notion of 'controlled *indirect* effect': While controlling the mediator by intervention isolates the direct path(s) (those not through A_2) there is no corresponding way of isolating the indirect path by controlling something else via some hypothetical intervention, except perhaps if all paths are indirect and can separately be interrupted by controlling the corresponding mediators.

Lastly, the *CDE* may not be viewed as a useful notion when manipulation of the mediator A_2 is not considered possible or not relevant to the research question, as is the case, e.g., in some examples in psychology

(Imai et al, 2013; Lok, 2016). It then becomes interesting to elicit in what way direct or indirect causal pathways are relevant to the research question, i.e., what actionable inference is targeted. As before, we emphasize that a clear research question is a prerequisite to choosing a suitable statistical analysis.

6.2 Causal Mediation: What are the Questions?

Summarizing the above, investigating joint interventions on (A_1, A_2), including the *CDE*, does not quite reflect an intuitive notion of causal mediation as these effects are agnostic to whether A_2 is itself causally affected by A_1 or not, or even vice versa. At this stage, it may be worth reflecting on some of the motivations typically given for causal mediation analyses.

Where best to intervene? A popular motivation for a causal mediation analysis is the wish to find out where best to intervene given several possible and interconnected targets for hypothetical interventions. However, this is not a question that requires assessing direct or indirect effects; it requires assessing the effect of the different options for interventions and comparing their effects on the relevant outcomes. In our running example with two treatments A_1 and A_2, this amounts to comparing, say $E(Y(a_1 = 1) - Y(a_1 = 0))$ versus $E(Y(a_2 = 1) - Y(a_2 = 0))$, which are each causal effects of two different point treatments.

As an example, Do et al (2024) investigate the role of emotional behaviors in sweet and fatty food choices and aimed at comparing whether hypothetically intervening on psychosocial well-being (PWB) or on emotion-driven impulsiveness (EDI) is more promising. In the underlying causal model, EDI is thought to be a potential mediator of PWB. However, this means primarily that PWB needs to be included in the set of covariates used for adjustment when estimating the causal effect of EDI. Taking the various causal effect estimates at face value (putting aside the wide confidence intervals) the analysis would suggest that both hypothetical interventions have small effects and an intervention on EDI would be slightly more effective than on PWB in reducing the consumption of sweet and fatty foods. A causal mediation analysis with the same data investigating mediation of the PWB effect via EDI suggests that roughly 50% of the effect is mediated — from this result it is not clear how to decide where best to intervene.

In another example, Aitken et al (2018) investigate material, psychosocial and behavioral factors as mediators between disability acquisition and mental health. Among others, the conclusion states that income support would be especially important for those with disability; however, the actual analysis does not target the effect of hypothetical income support for those with (compared to those without) disabilities. Instead of the performed mediation analysis, one could have assessed the potential of separate hypothetical interventions on either material or psychosocial or behavioral factors for those with disabilities.

Many interrelated outcomes. A further typical motivation for causal mediation analyses is that an intervention on the treatment is thought to have effects on many different, possibly interrelated outcomes. For instance, returning to the HIV example, anti-retroviral treatment affects AIDS-free survival 'through' various biological changes, e.g., it increases the CD4-count. A careful analysis of the various effects of a treatment by considering separately or jointly the components of a multivariate outcome does not in itself require any notion of mediation or (in)direct effects. Such an analysis, on its own, could suggest new hypotheses to be considered or help better assess the harms and benefits of an intervention.

H_0: There is no 'direct' effect. Some research questions essentially appear to seek confirmation or refutation of the hypothesis that there is no actual direct effect of a treatment or exposure. Put differently, the question is whether all mediators are captured by measured information so that the given set of mediators explains at least most of the total effect; an example may be the above gut microbiome example. At first glance, the *CDE* should suffice to address the question whether there is no or only a small direct effect after controlling for a set of mediators. However, such questions often occur in contexts where 'controlling' the mediators by some intervention is not practically feasible.

Examples are typically related to questions of fairness, e.g., it is well-known that deprivation or low socioeconomic status (SES) is associated with worse health outcomes (Li et al, 2016; Naimi et al, 2016; Schlüter et al, 2022). In this setting, an interesting question might be: If those with low SES had the same access to the same quality and intensity of health care as those with high SES, would the association disappear? Again, we find that the question is about an intervention in the suspected mediators, albeit a very hypothetical intervention and one that must be informed by the distribution of the mediators in the comparator group with high SES (called 'randomized intervention' in the next subsection).

Understanding mechanisms. The most common motivation for causal mediation analyses is the desire to better understand 'causal mechanisms'. This aim is in stark contrast to the typical mediation analyses, where (i) strong prior assumptions about the causal mechanisms are required in the first place, e.g., in SEMs, and (ii) the main result reported is typically that a certain proportion of the effect of A_1 is mediated though A_2. Suppose this proportion is near zero, then we do not know if the treatment A_1 has no effect on the mediator A_2, if A_2 has no effect on the outcome Y, or if the change that A_1 effects in A_2 is just not sufficient to subsequently produce much change in Y. Without a detailed description of the individual causal relations it appears difficult to obtain any insights into the causal mechanisms.

Separable treatments. In the seminal paper on direct and (in)direct effects, Pearl (2001) motivates his particular notion of indirect effect by a hypothetical experiment, where a novel type of cigarette would contain no nicotine, and the entire effect of nicotine on myocardial infarction is thought to go through hypertensive status while the other toxins in cigarettes have no effect on hypertension. In the novel cigarettes, the indirect pathway through hyptertension would be comparable to nonsmokers and thus 'interrupted', so that the effect of smoking these new versus no cigarettes would be the direct effect of smoking not mediated through hypertension. A similar example is given by Shrier and Suzuki (2022), describing research questions where the effect of exposure on the mediator can be 'broken' while leaving other causes of the mediator intact and its value therefore random. As argued by Robins and Richardson (2011), this is, however, the effect of a 'modified treatment' where components of the original treatment are separated, e.g., nicotine versus other toxins in cigarettes (see more details in Robins et al (2022)). More formally, the treatment A_1 is made-up of different components (A_1^Y, A_1^M) which could (hypothetically) be manipulated separately and are thought to separately activate different causal paths via or avoiding the mediator (Robins and Richardson, 2011; Robins et al, 2022).

A similar example is that of placebo-controlled blinded clinical trials, where the two components of a treatment are the 'awareness' of being treated and the 'active ingredient'. Both components are separated in such designs, with awareness being separately manipulated by using placebos that have no active ingredient. Such a design isolates the direct effect of the active ingredient but could also be modified to isolate the indirect 'placebo effect' (Didelez, 2013). Separable treatments occur in various contexts, e.g., in drug development, vaccine, or dementia studies (Stensrud and Dukes, 2022; Andrews et al, 2023; Stensrud et al, 2024).

Bibliographic notes: The *CDE* is useful to clarify certain approaches of dealing with competing events in time-to-event analyses, e.g., where the competing event is eliminated by an intervention (Young et al, 2020; Rojas-Saunero et al, 2023). Stensrud and Dukes (2022) discuss a number of examples from clinical trials, where it was suggested that some of the research questions relating to so-called intercurrent events in trials could be answered by causal mediation analyses. However, the authors show that many of the verbal descriptions of the research questions can more appropriately be formulated as joint (sequential) interventions, as controlled direct effects or separable treatment effects. The notion of fairness has been linked to path specific effects (Kusner et al, 2017; Nabi and Shpitser, 2018; Chiappa, 2019).

6.3 Alternative Notions of (In)Direct Causal Effects

Other notions of direct and indirect causal effects have been proposed to supplement the notion of *CDE*, but also to provide an alternative to the restrictive parametric assumptions underlying typical SEM approaches. Interestingly, these alternatives have in common that they are identified by a version of the generalized g-formula, but each have a different interpretation and are identified under different structural assumptions. Let us consider the causal model of Figure 6(ii), without the dashed edge, where the treatment A_1 is not affected by any confounding, e.g., because A_1 was randomized, and X_2 is the only common cause of the mediator A_2 and the outcome Y. In this case, the generalized g-formula can be given as follows.

$$\Phi(a_1, \tilde{a}_1) = \sum_{a_2, x_2} E(Y \mid A_1 = a_1, A_2 = a_2, X_2 = x_2) \\ P(A_2 = a_2 \mid A_1 = \tilde{a}_1, X_2 = x_2) P(X_2 = x_2). \quad (5)$$

Note that in equation (5) we average over the conditional distribution of A_2 given $A_1 = \tilde{a}_1$ (and $X_2 = x_2$), where $\tilde{a}_1$ is possibly different from a_1 in the conditional expectation of Y. We will now consider three alternative mediational concepts and how they relate to the above formula.

Natural direct and indirect effects. The notions of *natural direct and indirect effects (NDE, NIE)* have been propagated by Pearl (2001), but go back to Robins and Greenland (1992). Recall that $A_2(a_1)$ is the PO for the mediator A_2 under an intervention on the treatment A_1; then, $Y(a_1, A_2(\tilde{a}_1))$ is the *nested counterfactual* outcome if A_1 were set to a_1 and A_2 were set to its 'natural' value, $A_2(\tilde{a}_1)$, that it would take had A_1 been set to $\tilde{a}_1$ (see illustration in Figure 7(ii)). When $a_1 \neq \tilde{a}_1$ the nested counterfactual is a cross-world counterfactual because the same treatment A_1 is set to two different values simultaneously. Within this setup, varying a_1 yields the natural direct effect, while varying $\tilde{a}_1$ yields the natural indirect effect of A_1 on Y relative to the mediator A_2. Under specific structural assumptions $E(Y(a_1, A_2(\tilde{a}_1)))$ is identified by $\Phi(a_1, \tilde{a}_1)$ from equation (5). These assumptions are (informally): Conditional exchangeability regarding A_1 and A_2, A_1 and Y, and A_2 and Y (given A_1), possibly using different sets of covariates for each, and an additional cross-world independence assumption $Y(a_1, a_2) \perp\!\!\!\perp A_2(\tilde{a}_1)$ (given covariates). The latter would be violated if, e.g., in the DAG of Figure 6(ii) the dashed edge $A_1 \longrightarrow X_2$ is present even when X_2 is measured and can be adjusted for. This problem is known as 'treatment-induced confounding' (or 'recanting witness') of the mediator and the outcome (Avin et al, 2005; Andrews and Didelez, 2021).

The literal interpretations as well as the actionable use of the NDE and NIE are awkward as one can hardly imagine any actual manipulations of A_1 and A_2 that would result in these effects, even in idealized experimental trials; Robins and Richardson (2011) therefore call them 'non-manipulable' effects. An exception is discussed in Imai et al (2013), who consider special crossover study designs that are sometimes possible, e.g., in psychological experiments. Due to the cross-world nature and underlying assumption, the NDE and NIE have been described as unscientific (Robins and Richardson, 2011; Naimi et al, 2014). Moreover, it has been shown that the NDE or NIE are never identified in settings where Y is a survival outcome and mediation is through a longitudinal process (Didelez, 2019). This is related to the above issue of treatment-induced confounding, where it has also been suggested that the natural effects are not just unidentifiable but even not well-defined (Shrier, 2024). Nevertheless, these mediational concepts have attracted much interest and have been widely applied in epidemiological studies (VanderWeele, 2015).

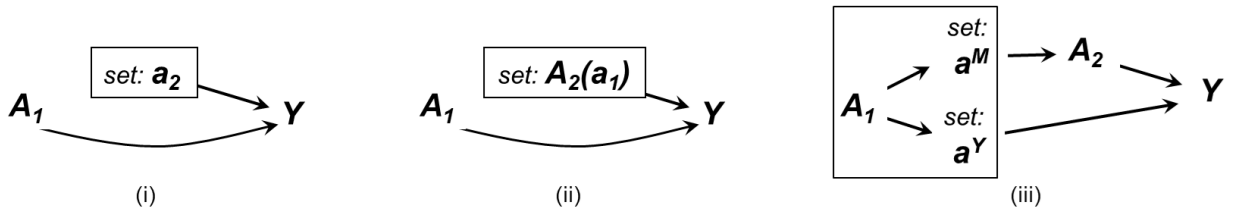

Fig. 7 DAGs illustrating types of interventions on mediator: (i) Controlled direct effect if a_2 is fixed; or randomized interventional direct effect if a_2 is randomly chosen from $P(A_2 = a_2 \mid \mathrm{do}(A_1 = \tilde{a}_1))$. (ii) Natural causal mediation effects setting mediator to its PO $A_2(\tilde{a}_1)$. (iii) Augmented DAG for separable treatment $A_1 = (A_1^M = a^M, A_1^Y = a^Y)$.

Randomized interventional effects. Formula (5) can be interpreted as the expected outcome under a joint intervention fixing $A_1 = a_1$ while the intervention on A_2 is such that the value is randomly drawn from its conditional distribution given an intervention setting $A_1 = \tilde{a}_1$ (Geneletti, 2007). Since perfect manipulation of the mediator is often regarded as infeasible, such stochastic interventions on A_2 have been proposed as an alternative basis to define (in)direct causal effects (VanderWeele et al, 2014; Lok, 2016; Díaz et al, 2021). As such it can be seen as generalization of the *CDE* (see Figure 7(i)). In particular, it has been applied in settings of fairness, where the aim might be to investigate, for instance, how health outcomes would change if access to the quality and intensity of health care for those with low SES were distributed like it is for those with high SES (Li et al, 2016).

The approach has the advantage of requiring weaker structural assumptions for identifiability, analogously to sequential treatments (Didelez et al, 2006), and no cross-world assumption is needed. In exchange, the distribution of the PO $A_2(\tilde{a}_1)$ must be identified. However, the randomized interventional (in)direct effects have been shown to not satisfy certain criteria one might expect of indirect effects (Miles, 2023). They have also been described as a technical work-around without the subtly different interpretation being correctly reflected in the analyses (Sarvet et al, 2023).

Separable treatment effects. Suppose that the research question can be represented as one of separable treatment effects with components (A_1^Y, A_1^M), where A_1^M is thought to act through the mediator A_2 (see Figure 7(iii)), as in the earlier nicotine-free cigarette example or placebo-controlled trial. As target of inference consider $E(Y(A_1^Y = a_1, A_1^M = \tilde{a}_1))$, i.e., the PO under separate manipulation of the treatment components.

One can view manipulation of a_1 while holding $\tilde{a}_1$ fixed as the direct separable effect (not through A_2), and manipulation of $\tilde{a}_1$ while holding a_1 fixed as the indirect separable effect. Given certain structural assumptions this is identified by the generalized g-formula (5). These assumptions are (informally) that A_1^Y and A_1^M have no direct effects on A_2 and Y, respectively, and that X_2 sufficiently accounts for any confounding between A_2 and Y. Again, no cross-world assumption is required, but we need to envisage settings where components (A_1^Y, A_1^M) are meaningful and separately manipulable.

Formalizing direct and indirect effects using separable treatments can be seen as a way to elaborate a research question to potentially modifiable factors when those represented by the measured variables are not (Rojas-Saunero et al, 2024). The concept has been extended to settings with survival outcomes and longitudinal mediating processes (Didelez, 2019), and has also been modified for dealing with competing events (Stensrud et al, 2022).

Bibliographic notes: Recent reviews on the reporting of (causal) mediation studies found many shortcomings, e.g., regarding the statement of assumptions or their justification, confounding control, sensitivity analyses, or simply stating the effect sizes (Cashin et al, 2019; Rizzo et al, 2022). Moreover, most studies used cross-sectional data. However, causal mediation should be considered over time and much recent work has considered time-varying settings (VanderWeele and Tchetgen Tchetgen, 2017; Vansteelandt et al, 2019; Didelez, 2019; Janvin et al, 2024), including separable effects for dynamic path analysis (Aalen et al, 2020). Further, the topic of multiple mediators has attracted much interest (Daniel et al, 2015; Steen et al, 2017) of which time-varying mediation can sometimes be regarded as special case.

7 Causal Discovery

In the previous section we mostly considered the situation where the causal research question is rather specific such that it can be formulated as a target trial, and mainly of the form 'what is the effect of one, multiple or sequential treatments on one or more (possibly time-dependent) outcomes'. Moreover, inference on causal effects typically relies on sufficient causal background knowledge, often represented by a causal DAG expressing the domain knowledge, to determine sufficient adjustment sets or motivate, say, an IV analysis. Usually, the construction of the causal DAG in epidemiology is expert-driven.

However, some causal questions may be more exploratory in nature, e.g., because the development of a research question is at an earlier stage and specific expert knowledge to draw a causal DAG is not yet available. This is where *causal discovery* (or causal search, (causal) structure learning) can make a contribution (Spirtes et al, 2000). The aim of causal discovery is to construct a causal DAG in a *data-driven* way, where some expert input or constraints such as temporal ordering can still be incorporated. The underlying principle reverses the fact that causal structures imply specific patterns of marginal and conditional independencies: Investigating and detecting such independencies in the data can help to (partially) recover the DAG. Basic examples are shown in Figure 8, where (i) shows DAGs that cannot be distinguished from each other because they encode the same conditional independencies; while (ii) shows examples that can be distinguished from (i) but have different intepretations depending on whether we allow latent (unmeasured) variables or not.

Causal discovery relies on a number of stringent assumptions and the methodology should mainly be viewed as an exploratory tool. There are also several sources of uncertainty which have to be considered: Not all causal structures are identifiable from data, e.g., if all variables are interconnected then we may not be able to find sufficient independencies to make much progress. Further, the outputs of DAG-search algorithms are often rather unstable in the sense that small changes to the data result in large changes in the discovered (set of) DAGs. An analysis should therefore investigate and report the stability of its findings by suitable resampling procedures, such as the bootstrap (Pigeot et al, 2015). We do not go into the details of any causal discovery algorithms here; some basic approaches are addressed in the Chapter on Causal Directed Acyclic Graphs by R. Foraita et al in this book. Recent methodological developments facilitate the application to typical epidemiological data by tackling missing values, mixed measurement scales and temporal structure (Witte et al, 2022; Foraita et al, 2024).

Despite the caveats regarding causal discovery, the methodology is an important one as it provides a data-driven alternative to the traditional entirely expert-driven construction of (causal) DAGs which is likely to be accompanied by 'confirmation bias', i.e., the analysis is set up so as to support what was believed beforehand. Injecting some empirical checks against data might sometimes generate original hypotheses (Tennant et al, 2021; Didelez, 2024; **?**). Especially in life course studies, causal discovery has the potential to suggest novel direct or indirect, potentially causal pathways (or the lack thereof). In view of the very formal approaches to (in)direct effects of Section 6, we note that due to the exploratory nature of causal discovery we speak

rather informally of (in)direct causal relations, here.

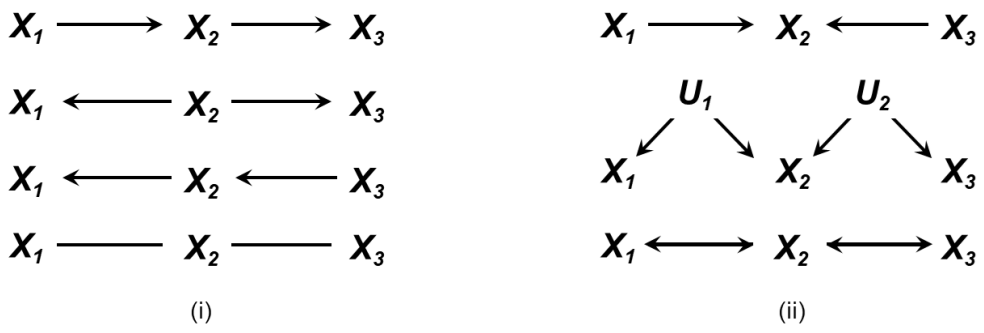

Fig. 8 Basic DAG patterns. (i) Equivalent DAGs, all representing $X_1 \perp\!\!\!\perp X_3 \,|\, X_2$; the whole set of DAGs is represented as undirected graph in the last line. (ii) DAGs without (top), with explicit (middle) and with implicit (bottom) latent variables, the latter are shown with bidirected edges; in all three graphs we have $X_1 \perp\!\!\!\perp X_3$ but $X_1 \not\perp\!\!\!\perp X_3 \,|\, X_2$.

Causal discovery in life course epidemiology. Recent examples of applications of causal discovery in life course epidemiology produced interesting insights, especially into the lack of direct causal relations between many, but not all, early life factors and later outcomes. The studies of Petersen et al (2021) and Petersen et al (2023) investigate the development of depression in Danish men based on a birth cohort followed to the age of 65 years. In particular, they find that some of the directed edges detected by the data-driven DAG-search algorithm are plausible even though the experts did not predict them, e.g., from social class (at birth) to childhood and youth' intelligence scores. Similarly, Foraita et al (2024) analyze the development of risk factors, health behaviors and health outcomes in a European children's cohort by determining plausible causal structures supported by the data. The authors find unexpectedly few and weak potentially causal links between early childhood behaviors and teenage health. For example, modifiable behaviors in a child's life such as diet, physical activity or sleep quality have no direct effects but statistically robust indirect effects on adolescent BMI. According to domain experts, the findings could indicate support for the somewhat disregarded 'systems hypothesis' claiming that the general environment has a strong influence, that is difficult to break, on individual health. In a careful application of causal discovery to postoperative length of stay in cardiac surgery, Lee et al (2023) also find that many of the known predictors of length of stay were in fact indirect causes and few were direct causes.

Assumptions of causal discovery. To give a basic idea, we address some of the typical key assumptions underlying many algorithms for causal discovery. Due to the variety of approaches to causal discovery there are many more subtle assumptions specific to the different methods; we refer to the Chapter on Causal Directed Acyclic Graphs by R. Foraita et al, in this book, and further literature for more details (Heinze-Deml et al, 2018; Glymour et al, 2019; Vowels et al, 2022).

First and foremost, the variables included in the dataset and input into a search algorithm must be chosen in a way so that they can meaningfully be connected by a causal DAG. This would, for instance, often be implausible with cross-sectional data on time-varying processes (Aalen et al, 2016). The field of 'causal representation learning' has emerged in this context, aiming to automatically determine semantically meaningful variables from raw data (Schölkopf et al, 2021). For instance, a distinction must be made between high-dimensional low-level data, such as sensory measurements from wearables, versus high-level variables, such as hours of vigorous physical activity per day; the latter may be more suitable for causal discovery than the former.

Given data on a carefully chosen set of variables, a common assumption is *faithfulness* (Spirtes et al, 2000): It requires that any conditional or marginal independence in the data distribution must correspond to a d-separation in the underlying DAG. This assumption reverses the (causal) Markov assumption and means that associations induced by different pathways cannot cancel each other out. While this is unlikely to occur exactly, 'near' cancellations are not unlikely and can still lead to bad performance of data-driven DAG-searches with finite samples.

A final important and rather restrictive assumption typical for many causal-search algorithms is *causal sufficiency*. It requires that the set of variables includes all common causes of every pair of variables, i.e., there is no latent confounding of any variables. In many applications, this is rather implausible. Causal sufficiency often holds approximately, at best, e.g., if the case can be made that all confounding domains

are represented by some measured variables, and that measurements are sufficiently frequent to adequately capture any time-dependent confounding and feedback processes. While causal discovery can be relaxed by some algorithms (cf. methods used by Lee et al (2023)), the computational costs are then often very high.

Interpretation of causal discovery. In view of the special assumptions of DAG-search algorithms, the discovered (set of) DAGs should be interpreted with care, especially if instability is also an issue. Directed paths may suggest novel indirect causal relationships that would not typically be found in any regression-based approaches. However, a key focus should also be on the absence of edges and paths. These can be interpreted as evidence for the absence of (in)direct causal relations even when causal sufficiency is violated. A caveat, here, is that searching for independencies in the data requires sufficiently large sample sizes to ensure the power to detect weak causal relations. Again, we emphasize that assessing the stability of interesting causal structures, such as presence or absence of certain paths, plays an important role (cf. analysis by Foraita et al (2024)).

Most recent causal discovery algorithms use machine learning, such as various types of neural networks or variational autoencoders, with the aim to relax restrictive assumptions on the shape of the associations (Vowels et al, 2022). However, these advanced approaches to causal discovery, while promising, have yet to be tried and tested in epidemiological studies (Petersen et al, 2024).

Bibliographic notes: It is tempting to first estimate the DAG, and then proceed as if this DAG had been known: Estimate any relevant causal effects using adjustment sets identified from this DAG. This is suggested by Maathuis et al (2010) and is known as IDA algorithm ('Intervention when the DAG is Absent'). However, the following two issues occur: (i) When the DAG is not fully identifiable then the adjustment set is not unique and more than one estimate may be obtained. (ii) There is an issue of post-selection inference because any default standard errors or confidence intervals do not account for the estimation of the DAG. First ideas on how to address post-selection inference are developed by Gradu et al (2022) and Chang et al (2024).

8 Conclusions

Historically, in epidemiology, the causal interpretation of statistical associations would refer to Hill's famous criteria (Hill, 1965). In this chapter we presented an approach to formal causal inference combing potential outcomes, dating back to the 70s (Rubin, 1974), methods for generalized time-dependent treatments, dating back to the 80s (Robins, 1986) and causal DAGs, dating back to the 90s (Pearl, 1995). VanderWeele (2020) discusses the relation between Hill's criteria and the more formal approach: The former is inductive while the latter is deductive, but both have common principles. For instance, a stronger association is less likely to be explained away by confounding as typically confirmed by formal bias analyses; or, the requirement for a causal association to be sufficiently specific is reflected in the idea of using negative controls as a check for unobserved confounding. However, which associations are more or less relevant seems hard to decide without a clear causal question, embedded in a causal model, as the starting point. The formal framework for causal reasoning and inference has been much extended and enthusiastically adopted since the 90s, and remains a very active field of research. It has considerably strengthened observational epidemiological research and will certainly continue to do so.

References

Aalen OO, Røysland K, Gran JM, Kouyos R, Lange T (2016) Can we believe the DAGs? A comment on the relationship between causal DAGs and mechanisms. Statistical Methods in Medical Research 25(5):2294–2314

Aalen OO, Stensrud MJ, Didelez V, Daniel R, Røysland K, Strohmaier S (2020) Time-dependent mediators in survival analysis: modeling direct and indirect effects with the additive hazards model. Biometrical Journal 62(3):532–549

Abadie A, Diamond A, Hainmueller J (2010) Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. Journal of the American Statistical Association 105(490):493–505

Andrews RM, Didelez V (2021) Insights into the cross-world independence assumption of causal mediation analysis. Epidemiology 32(2):209–219

Andrews RM, Shpitser I, Didelez V, Chaves PHM, Lopez OL, Carlson MC (2023) Examining the causal mediating role of cardiovascular disease on the effect of subclinical cardiovascular disease on cognitive impairment via separable effects. The Journals of Gerontology: Series A 78(7):1172–1178

Angrist JD, Pischke JS (2009) Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press

Ankan A, Wortel IM, Textor J (2021) Testing graphical causal models using the R package 'dagitty'. Current Protocols 1(2):e45

Austin PC, Small DS (2014) The use of bootstrapping when using propensity-score matching without replacement: A simulation study. Statistics in Medicine 33(24):4306–4319

Avin C, Shpitser I, Pearl J (2005) Identifiability of path-specific effects. In: Proceedings of the 19th International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, pp 357–363

Balke AA, Pearl J (1994) Counterfactual probabilities: Computational methods, bounds and applications. In: Mantaras R, Poole D (eds) Proceedings of the 10th Conference on Uncertainty in Artificial Inteligence, AUAI Press, pp 46–54

Bang H, Robins JM (2005) Doubly robust estimation in missing data and causal inference models. Biometrics 61(4):962–973

Bates S, Kennedy E, Tibshirani R, Ventura V, Wasserman L (2022) Causal inference with orthogonalized regression adjustment: Taming the phantom. arXiv preprint arXiv:220113451

Bell Gorrod H, Schomaker M, Maartens G, Murphy R (2020) Increased mortality with delayed and missed switch to second-line antiretroviral therapy in south africa. Journal of Acquired Immune Deficiency Syndromes 84(1):107–113

Bonander C, Humphreys D, Degli Esposti M (2021) Synthetic control methods for the evaluation of single-unit interventions in epidemiology: A tutorial. American Journal of Epidemiology 190(12):2700–2711

Börnhorst C, Reinders T, Rathmann W, Bongaerts B, Haug U, Didelez V, Kollhorst B (2021) Avoiding time-related biases: A feasibility study on antidiabetic drugs and pancreatic cancer applying the parametric g-formula to a large German healthcare database. Clinical Epidemiology 13:1027–1038

Börnhorst C, Pigeot I, De Henauw S, Formisano A, Lissner L, Molnár D, Tornaritis M, Veidebaum T, Vrijkotte T, Wolters M, Didelez V (2023) The effects of hypothetical behavioral interventions on the 13-year incidence of overweight/obesity in children and adolescents. International Journal of Behavioural Nutrition and Physical Activity 20(100):epub

Bowden J, Davey Smith G, Burgess S (2015) Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. International Journal of Epidemiology 44(2):512–525

Braitmaier M, Didelez V (2022) Emulierung von "Target Trials" mit Real-World-Daten (emulation of target trials with real-world data). Prävention und Gesundheitsförderung pp 1–8

Braitmaier M, Schwarz S, Kollhorst B, Senore C, Didelez V, Haug U (2022) Screening colonoscopy similarly prevented distal and proximal colorectal cancer: A prospective study among 55–69-year-olds. Journal of Clinical Epidemiology 149:118–126

Braitmaier M, Schwarz S, Didelez V, Haug U (2024) Misleading and avoidable: design-induced biases in observational studies evaluating cancer screening–the example of site-specific effectiveness of screening colonoscopy. medRxiv preprint medRxiv:2024042924306522

Brookhart MA, Wang PS, Solomon DH, Schneeweiss S (2006) Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. Epidemiology 17(3):268–275

Brumback BA (2021) Fundamentals of Causal Inference with R. Chapman and Hall/CRC

Cain LE, Robins JM, Lanoy E, Logan R, Costagliola D, Hernán MA (2010) When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. The International Journal of Biostatistics 6(2):321–342

Cain LE, Logan R, Robins JM, Sterne JAC, Sabin C, Bansi L, Justice A, Goulet J, van Sighem A, de Wolf F, Bucher V H C von Wyl, Esteve A, Casabona J, del Amo J, Moreno S, Seng R, Meyer L, Pérez-Hoyos S, Muga R, Lodi S, Lanoy E, Costagliola D, Hernán MA (2011) When to initiate combined antiretroviral therapy to reduce mortality and AIDS-defining illness in HIV-infected persons in developed countries: An observational study. Annals of Internal Medicine 154(8):509–515

Cashin AG, Lee H, Lamb SE, Hopewell S, Mansell G, Williams CM, Kamper SJ, Henschke N, McAuley JH (2019) An overview of systematic reviews found suboptimal reporting and methodological limitations of mediation studies investigating causal mechanisms. Journal of Clinical Epidemiology 111:60–68

Chang TH, Guo Z, Malinsky D (2024) Post-selection inference for causal effects after causal discovery. arXiv preprint arXiv:240506763

Chattopadhyay A, Hase CH, Zubizarreta JR (2020) Balancing vs modeling approaches to weighting in practice. Statistics in Medicine 39(24):3227–3254

Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, Robins J (2018) Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal 21(1):C1–C68

Chiappa S (2019) Path-specific counterfactual fairness. In: Proceedings of the AAAI Conference on Artificial Intelligence, AAAI Press, Menlo Park, CA, vol 33, pp 7801–7808

Chiu YH, Chavarro JE, Dickerman BA, Manson JE, Mukamal KJ, Rexrode KM, Rimm EB, Hernán MA (2021) Estimating the effect of nutritional interventions using observational data: The American Heart Association's 2020 dietary goals and mortality. The American Journal of Clinical Nutrition 114(2):690–703

Cole SR, Frangakis CE (2009) The consistency statement in causal inference: A definition or an assumption? Epidemiology 20(1):3–5

Cole SR, Hernán MA (2002) Fallibility in estimating direct effects (with discussion). International Journal of Epidemiology 31(1):163–165

Cole SR, Hernán MA (2004) Adjusted survival curves with inverse probability weights. Computer Methods and Programs in Viomedicine 75(1):45–49

Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, Poole C (2010) Illustrating bias due to conditioning on a collider. International Journal of Epidemiology 39(2):417–420

Dagan N, Barda N, Kepten E, Miron O, Perchik S, Katz MA, Hernán MA, Lipsitch M, Reis B, Balicer RD (2021) BNT162b2 mRNA COVID-19 vaccine in a nationwide mass vaccination setting. New England Journal of Medicine 384(15):1412–1423

Dang LE, Gruber S, Lee H, Dahabreh IJ, Stuart EA, Williamson BD, Wyss R, Díaz I, Ghosh D, Kıcıman E, Alemayehu DA, Hoffman KL, Vossen CY, Huml RA, Ravn H, Kvist K, Pratley R, Shih MC, Pennello G, Martin D, Waddy SP, Barr CE, Akacha M, Buse JB, van der Laan M, Petersen M (2023) A causal roadmap for generating high-quality real-world evidence. Journal of Clinical and Translational Science 7(1):e212

Daniel RM, Cousens S, De Stavola B, Kenward MG, Sterne J (2013) Methods for dealing with time-dependent confounding. Statistics in Medicine 32(9):1584–1618

Daniel RM, De Stavola BL, Cousens SN, Vansteelandt S (2015) Causal mediation analysis with multiple mediators. Biometrics 71(1):1–14

Daniel RM, Zhang J, Farewell D (2021) Making apples from oranges: Comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets. Biometrical Journal 63(3):528–557

Davey Smith G, Ebrahim S (2003) Mendelian randomization: Can genetic epidemiology contribute to understanding environmental determinants of disease? International Journal of Epidemiology 32(1):1–22
Davies NM, Smith GD, Windmeijer F, Martin RM (2013) Issues in the reporting and conduct of instrumental variable studies: A systematic review. Epidemiology 24(3):363–369
Davis JA (1984) Extending Rosenberg's technique for standardizing percentage tables. Social Forces 62:679–708
Dawid A (2021) Decision-theoretic foundations for statistical causality. Journal of Causal Inference 9(1):39–77
Dawid AP (1979) Conditional independence in statistical theory (with discussion). Journal of the Royal Statistical Society Series B: Statistical Methodology 41(1):1–31
Dawid AP (2000) Causal inference without counterfactuals (with discussion). Journal of the American Statistical Association 95:407–448
Dawid AP (2002) Influence diagrams for causal modelling and inference. International Statistical Review 70:161–89
Dawid AP (2007) Counterfactuals, hypotheticals and potential responses: A philosophical examination of statistical causality. In: Russo F, Williamson J (eds) Causality and Probability in the Sciences, Texts in Philosophy, vol 5, College Publications, London, pp 503–532
Dawid AP (2010) Beware of the DAG! In: Causality: objectives and assessment, Proceedings of Machine Learning Research, pp 59–86
Dawid AP (2015) Statistical causality from a decision-theoretic perspective. Annual Review of Statistics and Its Application 2(1):273–303
Dawid AP, Didelez V (2010) Identifying the consequences of dynamic treatment strategies: A decision-theoretic overview. Statististical Surveys 4:184–231
Dawid AP, Didelez V (2012) "Imagine a can opener" – The magic of principal stratum analysis. The International Journal of Biostatistics 8(1):1–10
De Luna X, Waernbaum I, Richardson TS (2011) Covariate selection for the nonparametric estimation of an average treatment effect. Biometrika 98(4):861–875
De Stavola BL, Gomes M, Katsoulis M (2023) Transparency and rigor: Target trial emulation aims to achieve both. Epidemiology 34(5):624–626
Denz R, Klaaßen-Mielke R, Timmesfeld N (2023) A comparison of different methods to adjust survival curves for confounders. Statistics in Medicine 42(10):1461–1479
Díaz I, van der Laan MJ (2013) Assessing the causal effect of policies: An example using stochastic interventions. The International Journal of Biostatistics 9(2):161–174
Díaz I, Hejazi NS, Rudolph KE, van Der Laan MJ (2021) Nonparametric efficient causal mediation with intermediate confounders. Biometrika 108(3):627–641
Dickerman BA, García-Albéniz X, Logan RW, Denaxas S, Hernán MA (2019) Avoidable flaws in observational analyses: An application to statins and cancer. Nature Medicine 25(10):1601–1606
Didelez V (2013) Discussion of 'Experimental designs for identifying causal mechanisms' by Imai, Tingley, Yamamoto. Journal of the Royal Statistical Society Series A: Statistics in Society 176:39
Didelez V (2016) Commentary: Should the analysis of observational data always be preceded by specifying a target experimental trial? International Journal of Epidemiology 45(6):2049–2051
Didelez V (2018) Causal concepts and graphical models. In: Maathuis M, Drton M, Lauritzen SL, Wainwright M (eds) Handbook of Graphical Models, Handbooks of Modern Statistical Methods, Chapman and Hall/CRC
Didelez V (2019) Defining causal mediation with a longitudinal mediator and a survival outcome. Lifetime Data Analysis 25:593–610
Didelez V (2024) Invited commentary: Where do the causal DAGs come from? American Journal of Epidemiology kwae028:epub
Didelez V, Evans RJ (2018) Causal inference from case-control studies. In: Borgan O, Breslow N, Chatterjee N, Gail MH, Scott A, Wild CJ (eds) Handbook of Statistical Methods for Case-Control Studies, Chapman and Hall/CRC, pp 87–115
Didelez V, Sheehan NA (2007) Mendelian randomisation as an instrumental variable approach to causal inference. Statistical Methods in Medical Research 16(4):309–330
Didelez V, Dawid A, Geneletti S (2006) Direct and indirect effects of sequential treatments. In: Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence, AUAI Press, Arlington, Virginia, pp 138–146
Didelez V, Kreiner S, Keiding N (2010a) Graphical models for inference under outcome-dependent sampling. Statistical Science 25(3):368–387
Didelez V, Meng S, Sheehan NA (2010b) Assumptions of IV methods for observational epidemiology. Statistical Science 25(1):22–40
Didelez V, Haug U, Garcia-Albeniz X (2024) Re: Are target trial emulations the gold standard for observational studies? Epidemiology 35(1):e3
Do S, Didelez V, Börnhorst C, Coumans JM, Reisch LA, Danner UN, Russo P, Veidebaum T, Tornaritis M, Molnár D, et al (2024) The role of psychosocial well-being and emotion-driven impulsiveness in food choices of European adolescents. International Journal of Behavioral Nutrition and Physical Activity 21(1)
Enders D, Engel S, Linder R, Pigeot I (2018) Robust versus consistent variance estimators in marginal structural Cox models. Statistics in Medicine 37(24):3455–3470
Foraita R, Witte J, Börnhorst C, Gwozdz W, Pala V, Lissner L, Lauria F, Reisch L, Molnár D, De Henauw S, Moreno L, Veidebaum T, Tornaritis M, Pigeot I, Didelez V (2024) A longitudinal causal graph analysis investigating modifiable risk factors and obesity in a European cohort of children and adolescents. Scientific Reports 14(6822):epub
Fox MP, Edwards JK, Platt R, Balzer LB (2019) The critical importance of asking good questions: The role of epidemiology doctoral training programs. American Journal of Epidemiology 189(4):261–264
Fox MP, MacLehose RF, Lash TL (2021) Applying Quantitative Bias Analysis to Epidemiologic Data. Springer
Frangakis CE, Rubin DB (2002) Principal stratification in causal inference. Biometrics 58(1):21–29
Galea S, Hernán MA (2019) Win-win: Reconciling social epidemiology and causal inference. American Journal of Epidemiology 189(3):167–170
García-Albéniz X, Hsu J, Hernán MA (2017) The value of explicitly emulating a target trial when using real world evidence: Amultiple imputation and test-wise deletion for causal discovery with incomplete cohort datan application to colorectal cancer screening. European Journal of Epidemiology 32:495–500

van Geloven N, Swanson SA, Ramspek CL, Luijken K, van Diepen M, Morris TP, Groenwold RH, van Houwelingen HC, Putter H, le Cessie S (2020) Prediction meets causal inference: The role of treatment in clinical prediction models. European Journal of Epidemiology 35:619–630

van Geloven N, Keogh RH, van Amsterdam W, Cinà G, Krijthe JH, Peek N, Luijken K, Magliacane S, Morzywołek P, van Ommen T, Sperrin M, Didelez V (2024) The risks of risk assessment: Causal blind spots when using prediction models for treatment decisions. arXiv preprint arXiv:240217366

Geneletti S (2007) Identifying direct and indirect effects in a non-counterfactual framework. Journal of the Royal Statistical Society Series B: Statistical Methodology 69(2):199–215

Geneletti S, O'Keeffe AG, Sharples LD, Richardson S, Baio G (2015) Bayesian regression discontinuity designs: incorporating clinical knowledge in the causal analysis of primary care data. Statistics in Medicine 34(15):2334–2352

Glymour C, Zhang K, Spirtes P (2019) Review of causal discovery methods based on graphical models. Frontiers in Genetics 10(524)

Goetghebeur E, le Cessie S, De Stavola B, Moodie EE, Waernbaum I (2020) Formulating causal questions and principled statistical answers. Statistics in Medicine 39(30):4922–4948

Gradu P, Zrnic T, Wang Y, Jordan MI (2022) Valid inference after causal discovery. arXiv preprint arXiv:220805949

Greenland S (2000) An introduction to instrumental variables for epidemiologists. International Journal of Epidemiology 29(4):722–729

Greenland S (2003) Quantifying biases in causal models: Classical confounding vs collider–stratification bias. Epidemiology 14:300–306

Greenland S (2017) For and against methodologies: Some perspectives on recent causal and statistical inference debates. European Journal of Epidemiology 32:3–20

Greenland S, Pearl J (2011) Adjustments and their consequences — collapsibility analysis using graphical models. International Statistical Review 79(3):401–426

Greenland S, Robins JM, Pearl J (1999) Confounding and collapsibility in causal inference. Statistical Science 14(1):29–46

Guo K, Diemer EW, Labrecque JA, Swanson SA (2023) Falsification of the instrumental variable conditions in Mendelian randomization studies in the UK Biobank. European Journal of Epidemiology 38(9):921–927

Hansford HJ, Cashin AG, Jones MD, Swanson SA, Islam N, Douglas SR, Rizzo RR, Devonshire JJ, Williams SA, Dahabreh IJ, Dickerman BA (2023) Reporting of observational studies explicitly aiming to emulate randomized trials: A systematic review. JAMA Network Open 6(9):e2336023

Havercroft W, Didelez V (2012) Simulating from marginal structural models with time-dependent confounding. Statistics in Medicine 31(30):4190–4206

Heinze-Deml C, Maathuis MH, Meinshausen N (2018) Causal structure learning. Annual Review of Statistics and Its Application 5(1):371–391

Hernán MA (2004) A definition of causal effect for epidemiological research. Journal of Epidemiology and Community Health 58(4):265–271

Hernán MA (2010) The hazards of hazard ratios. Epidemiology 21(1):13–15

Hernán MA (2016) Does water kill? A call for less casual causal inferences. Annals of Epidemiology 26(10):674–680

Hernán MA (2018a) The C-word: Scientific euphemisms do not improve causal inference from observational data. American Journal of Public Health 108(5):616–619

Hernán MA (2018b) How to estimate the effect of treatment duration on survival outcomes using observational data. British Medical Journal 360

Hernán MA (2024) Causal diagrams: Draw your assumptions before your conclusions. `https://www.harvardonline.harvard.edu/course/causal-diagrams-draw-your-assumptions-your-conclusions`, [Online; accessed 25-April-2024]

Hernán MA, Hernández-Díaz S (2012) Beyond the intention-to-treat in comparative effectiveness research. Clinical Trials 9(1):48–55

Hernán MA, Monge S (2023) Selection bias due to conditioning on a collider. British Medical Journal 381

Hernán MA, Robins JM (2006) Instruments for causal inference: An epidemiologist's dream? Epidemiology 17(4):360–372

Hernán MA, Robins JM (2016) Using big data to emulate a target trial when a randomized trial is not available. American Journal of Epidemiology 183(8):758–764

Hernán MA, Robins JM (2020) Causal Inference: What If. Chapman and Hall/CRC

Hernán MA, Taubman SL (2008) Does obesity shorten life? The importance of well-defined interventions to answer causal questions. International Journal of Obesity 32(3):8–14

Hernán MA, Brumback B, Robins JM (2000) Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 11(5):561–570

Hernán MA, Hernández-Díaz S, Robins JM (2004) A structural approach to selection bias. Epidemiology 15(5):615–625

Hernán MA, Alonso A, Logan R, Grodstein F, Michels KB, Stampfer MJ, Willett WC, Manson JE, Robins JM (2008) Observational studies analyzed like randomized experiments: An application to postmenopausal hormone therapy and coronary heart disease. Epidemiology 19(6):766–779

Hernán MA, Hsu J, Healy B (2019) A second chance to get causal inference right: A classification of data science tasks. CHANCE 32(1):42–49

Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I (2016) Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. Journal of Clinical Epidemiology 79:70–75

Hernández-Díaz S, Schisterman EF, Hernán MA (2006) The birth weight 'paradox' uncovered? American Journal of Epidemiology 164(11):1115–1120

Hill AB (1965) The environment and disease: Association or causation? Proceedings of the Royal Society of Medicine pp 295–300

Hudgens MG, Halloran ME (2008) Toward causal inference with interference. Journal of the American Statistical Association 103(482):832–842

Huitfeldt A (2016) Is caviar a risk factor for being a millionaire? British Medical Journal 355

Imai K, Tingley D, Yamamoto T (2013) Experimental designs for identifying causal mechanisms. Journal of the Royal Statistical Society Series A: Statistics in Society 176(1):5–51

Jackson JW (2019) Diagnosing covariate balance across levels of right-censoring before and after application of inverse-probability-of-censoring weights. American Journal of Epidemiology 188(12):2213–2221
Jackson JW, Arah OA (2019) Invited commentary: Making causal inference more social and (social) epidemiology more causal. American Journal of Epidemiology 189(3):179–182
Janvin M, Young JG, Ryalen PC, Stensrud MJ (2024) Causal inference with recurrent and competing events. Lifetime Data Analysis 30(1):59–118
Katan M (1986) Apoupoprotein E isoforms, serum cholesterol, and cancer. The Lancet 327(8479):507–508
Kaufman JS (2019) Commentary: Causal inference for social exposures. Annual Review of Public Health 40:7–21
Keiding N, Clayton D (2014) Standardization and control for confounding in observational studies: A historical perspective. Statistical Science 29(4):529–558
Kennedy EH (2019) Nonparametric causal effects based on incremental propensity score interventions. Journal of the American Statistical Association 114(526):645–656
Kennedy EH, Ma Z, McHugh MD, Small DS (2017) Nonparametric methods for doubly robust estimation of continuous treatment effects. Journal of the Royal Statistical Society Series B: Statistical Methodology 79(4):1229–1245
Kennedy TM, Kennedy EH, Ceballo R (2023) Marginal structural models for estimating the longitudinal effects of community violence exposure on youths' internalizing and externalizing symptoms. Psychological Trauma: Theory, Research, Practice, and Policy 15(6):906–912
Kennedy-Shaffer L (2024) Quasi-experimental methods for pharmacoepidemiology: Difference-in-differences and synthetic control methods with case studies for vaccine evaluation. American Journal of Epidemiology kwae019:epub
Knaus MC (2022) Double machine learning-based programme evaluation under unconfoundedness. The Econometrics Journal 25(3):602–627
Künzel SR, Sekhon JS, Bickel PJ, Yu B (2019) Metalearners for estimating heterogeneous treatment effects using machine learning. Proceedings of the National Academy of Sciences 116(10):4156–4165
Kusner MJ, Loftus J, Russell C, Silva R (2017) Counterfactual fairness. Advances in Neural Information Processing Systems 30
van der Laan MJ, Rose S (2011) Targeted Learning: Causal Inference for Observational and Experimental Data. Springer
Lawlor DA (2016) Commentary: Two-sample Mendelian randomization: Opportunities and challenges. International Journal of Epidemiology 45(3):908–915
Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G (2008) Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. Statistics in Medicine 27(8):1133–1163
Lawlor DA, Tilling K, Davey Smith G (2016) Triangulation in aetiological epidemiology. International Journal of Epidemiology 45(6):1866–1886
Lee JJ, Srinivasan R, Ong CS, Alejo D, Schena S, Shpitser I, Sussman M, Whitman GJ, Malinsky D (2023) Causal determinants of postoperative length of stay in cardiac surgery using causal graphical learning. The Journal of Thoracic and Cardiovascular Surgery 166(5):e446–e462
Li F, Thomas LE, Li F (2019) Addressing extreme propensity scores via the overlap weights. American Journal of Epidemiology 188(1):250–257
Li R, Daniel R, Rachet B (2016) How much do tumor stage and treatment explain socioeconomic inequalities in breast cancer survival? Applying causal mediation analysis to population-based data. European Journal of Epidemiology 31:603–611
Linden A (2018) A matching framework to improve causal inference in interrupted time-series analysis. Journal of Evaluation in Clinical Practice 24(2):408–415
Linden A, Uysal SD, Ryan A, Adams JL (2016) Estimating causal effects for multivalued treatments: A comparison of approaches. Statistics in Medicine 35(4):534–552
Lipsitch M, Tchetgen Tchetgen E, Cohen T (2010) Negative controls: A tool for detecting confounding and bias in observational studies. Epidemiology 21(3):383–388
Lok J (2016) Defining and estimating causal direct and indirect effects when setting the mediator to specific values is not feasible. Statistics in Medicine 35(22):4008–4020
Love TE (2002) Displaying covariate balance after adjustment for selection bias. In: Joint Statistical Meetings, vol 11
Luijken K, van Eekelen R, Gardarsdottir H, Groenwold RH, van Geloven N (2023) Tell me what you want, what you really really want: Estimands in observational pharmacoepidemiologic comparative effectiveness and safety studies. Pharmacoepidemiology and Drug Safety 32(8):863–872
Maathuis MH, Colombo D, Kalisch M, Bühlmann P (2010) Predicting causal effects in large-scale systems from observational data. Nature Methods 7(4):247–248
Matsouaka RA, Liu Y, Zhou Y (2024) Overlap, matching, or entropy weights: What are we weighting for? Communications in Statistics – Simulation and Computation pp 1–20
Miles CH (2023) On the causal interpretation of randomised interventional indirect effects. Journal of the Royal Statistical Society Series B: Statistical Methodology 85(4):1154–1172
Mitze T, Kosfeld R, Rode J, Wälde K (2020) Face masks considerably reduce COVID-19 cases in Germany. Proceedings of the National Academy of Sciences 117(51):32293–32301
Morris TT, Heron J, Sanderson EC, Davey Smith G, Didelez V, Tilling K (2022) Interpretation of Mendelian randomization using a single measure of an exposure that varies over time. International Journal of Epidemiology 51(6):1899–1909
Morzywołek P, Steen J, Vansteelandt S, Decruyenaere J, Sterck S, Van Biesen W (2022) Timing of dialysis in acute kidney injury using routinely collected data and dynamic treatment regimes. Critical Care 26(1):365–377
Nabi R, Shpitser I (2018) Fair inference on outcomes. In: Proceedings of the AAAI Conference on Artificial Intelligence, AAAI Press, Menlo Park, CA, vol 32
Naimi AI, Kaufman JS, MacLehose RF (2014) Mediation misgivings: Ambiguous clinical and public health interpretations of natural direct and indirect effects. International Journal of Epidemiology 43(5):1656–1661
Naimi AI, Schnitzer ME, Moodie EE, Bodnar LM (2016) Mediation analysis for health disparities research. American Journal of Epidemiology 184(4):315–324
Naimi AI, Mishler AE, Kennedy EH (2021) Challenges in obtaining valid causal effect estimates with machine learning algorithms. American Journal of Epidemiology 192(9):1536–1544

Oberst M, Johansson F, Wei D, Gao T, Brat G, Sontag D, Varshney K (2020) Characterization of overlap in observational studies. In: Chiappa S, Calandra R (eds) Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research, vol 108, pp 788–798

Ogburn EL, Shpitser I, Lee Y (2020) Causal inference, social networks and chain graphs. Journal of the Royal Statistical Society Series A: Statistics in Society 183(4):1659–1676

Orellana L, Rotnitzky A, Robins JM (2010) Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part i: main content. The International Journal of Biostatistics 6(2):1–47

Palmer TM, Ramsahai RR, Didelez V, Sheehan NA (2011) Nonparametric bounds for the causal effect in a binary instrumental-variable model. The Stata Journal 11(3):345–367

Pearce N, Vandenbroucke JP (2023) Are target trial emulations the gold standard for observational studies? Epidemiology 34(5):614–618

Pearl J (1995) Causal diagrams for empirical research. Biometrika 82(4):669–688

Pearl J (2001) Direct and indirect effects. In: Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, pp 411–420

Pearl J (2009) Causality: Models, Reasoning, and Inference, 2nd edn. Cambridge University Press

Pearl J, Mackenzie D (2018) The Book of Why: The New Science of Cause and Effect. Basic Books

Pearl J, Robins J (1995) Probabilistic evaluation of sequential plans from causal models with hidden variables. In: Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, pp 444–453

Perković E, Textor J, Kalisch M, Maathuis MH (2018) Complete graphical characterization and construction of adjustment sets in Markov equivalence classes of ancestral graphs. Journal of Machine Learning Research 18(220):1–62

Peters J, Janzing D, Schölkopf B (2017) Elements of Causal Inference: Foundations and Learning Algorithms. MIT Press

Petersen AH, Osler M, Ekstrøm CT (2021) Data-driven model building for life-course epidemiology. American Journal of Epidemiology 190(9):1898—-1907

Petersen AH, Ekstrøm CT, Spirtes P, Osler M (2023) Constructing causal life course models: Comparative study of data-driven and theory-driven approaches. American Journal of Epidemiology 192(11):1917–1927

Petersen AH, Ekstrøm CT, Spirtes P, Osler M (2024) Causal discovery and epidemiology: A potential for synergy. American Journal of Epidemiology p kwae101

Petersen ML, van der Laan MJ (2014) Causal models and learning from data: Integrating causal modeling and statistical estimation. Epidemiology 25(3):418–426

Piccininni M, Kurth T, Audebert HJ, Rohmann JL (2023) The effect of mobile stroke unit care on functional outcomes: An application of the front-door formula. Epidemiology 34(5):712–720

Pigeot I, Sobotka F, Kreiner S, Foraita R (2015) The uncertainty of a selected graphical model. Journal of Applied Statistics 42(11):2335–2352

Ramspek CL, Steyerberg EW, Riley RD, Rosendaal FR, Dekkers OM, Dekker FW, van Diepen M (2021) Prediction or causality? A scoping review of their conflation within current observational research. European Journal of Epidemiology 36:889–898

Ray WA (2003) Evaluating medication effects outside of clinical trials: New-user designs. American Journal of Epidemiology 158(9):915–920

Richardson TS, Robins JM (2013a) Single world intervention graphs: A primer. In: 2nd Workshop on Causal Structure Learning, Conference on Uncertainty in Artificial Inteligence

Richardson TS, Robins JM (2013b) Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. Center for the Statistics and the Social Sciences, University of Washington Series Working Papers 128(30)

Rizzo RR, Cashin AG, Bagg MK, Gustin SM, Lee H, McAuley JH (2022) A systematic review of the reporting quality of observational studies that use mediation analyses. Prevention Science 23(6):1041–1052

Robins JM (1986) A new approach to causal inference in mortality studies with sustained exposure periods — application to control for the healthy worker survivor effect. Mathematical Modelling 7:1393–1512

Robins JM (1997) Causal inference from complex longitudinal data. In: Berkane M (ed) Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics vol. 120, Springer, New York, pp 69–117

Robins JM (2001) Data, design and background knowledge in etiologic inference. Epidemiology 11(3):313–320

Robins JM, Greenland S (1992) Identifiability and exchangeability of direct and indirect effects. Epidemiology 3(2):143–155

Robins JM, Richardson TS (2011) Alternative graphical causal models and the identification of direct effects. In: Shrout P (ed) Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures, Oxford University Press

Robins JM, Blevins D, Ritter G, Wulfsohn M (1992) G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of AIDS patients. Epidemiology 3(4):319–336

Robins JM, Hernán MA, Brumback B (2000) Marginal structural models and causal inference in epidemiology. Epidemiology 11(5):550–560

Robins JM, Richardson TS, Shpitser I (2022) An interventionist approach to mediation analysis. In: Geffner H, Dechter R, Halpern JY (eds) Probabilistic and Causal Inference: The Works of Judea Pearl, ACM, pp 713–764

Rojas-Saunero LP, Young JG, Didelez V, Ikram MA, Swanson SA (2023) Considering questions before methods in dementia research with competing events and causal goals. American Journal of Epidemiology 192(8):1415–1423

Rojas-Saunero LP, van der Willik KD, Schagen SB, Ikram MA, Swanson SA (2024) Towards a clearer causal question underlying the association between cancer and dementia. Epidemiology 35(3):281–288

Rosenbaum PR (1989) Optimal matching for observational studies. Journal of the American Statistical Association 84(408):1024–1032

Rosenbaum PR (2010) Design of Observational Studies. Springer

Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. Biometrika 70(1):41–55

Rubin DB (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology 66(5):688–701

Rubin DB (2005) Causal inference using potential outcomes: Design, modeling, decisions. Journal of the American Statistical Association 100(469):322–331

Sachs MC, Jonzon G, Sjölander A, Gabriel EE (2023) A general method for deriving tight symbolic bounds on causal effects. Journal of Computational and Graphical Statistics 32(2):567–576

Sarvet AL, Stensrud MJ, Wen L (2023) Interpretational errors in statistical causal inference. arXiv preprint arXiv:231207610

Schlüter DK, Keogh RH, Daniel RM, Agbla SC, Taylor-Robinson D (2022) How do growth and nutrition explain social inequalities in lung function in children with cystic fibrosis? A longitudinal mediation analysis using interventional disparity effects with time-varying mediators and intermediate confounders. medRxiv

Schölkopf B, Locatello F, Bauer S, Ke NR, Kalchbrenner N, Goyal A, Bengio Y (2021) Toward causal representation learning. Proceedings of the IEEE 109(5):612–634

Shah V, Kreif N, Jones AM (2021) Machine learning for causal inference: Estimating heterogeneous treatment effects. In: Hashimzade N, Thornton MA (eds) Handbook of Research Methods and Applications in Empirical Microeconomics, Edward Elgar Publishing, pp 438–487

Shalizi CR, Thomas AC (2011) Homophily and contagion are generically confounded in observational social network studies. Sociological Methods & Research 40(2):211–239

Shortreed SM, Ertefaie A (2017) Outcome-adaptive lasso: Variable selection for causal inference. Biometrics 73(4):1111–1122

Shpitser I, Richardson TS, Robins JM (2022) Multivariate counterfactual systems and causal graphical models. In: Geffner H, Dechter R, Halpern JY (eds) Probabilistic and Causal Inference: The Works of Judea Pearl, ACM, pp 813–852

Shrier I (2024) Letter: Natural effects with a recanting witness: Non-identifiability or meaningless estimand? Epidemiology pp (e–pub), DOI 10.1097/EDE.0000000000001768

Shrier I, Suissa S (2022) The quintessence of causal DAGs for immortal time bias: Time-dependent models. International Journal of Epidemiology 51(3):1028–1029

Shrier I, Suzuki E (2022) The primary importance of the research question: Implications for understanding natural versus controlled direct effects. International Journal of Epidemiology 51(4):1041–1046

Shrier I, Stovitz SD, Textor J (2023) Identifiability of causal effects in test-negative design studies. International Journal of Epidemiology 52(6):1968–1974

Sjölander A (2009) Propensity scores and M-structures. Statistics in Medicine 28(9):1416–1420

Sjölander A (2018) Estimation of causal effect measures with the r-package stdreg. European Journal of Epidemiology 33(9):847–858

Sjölander A, Greenland S (2022) Are e-values too optimistic or too pessimistic? Both and neither! International Journal of Epidemiology 51(2):355–363

Skrivankova VW, Richmond RC, Woolf BA, Davies NM, Swanson SA, VanderWeele TJ, Timpson NJ, Higgins JP, Dimou N, Langenberg C, et al (2021) Strengthening the reporting of observational studies in epidemiology using Mendelian randomisation (STROBE-MR): Explanation and elaboration. British Medical Journal 375:n2233

Smith LH, VanderWeele TJ (2019) Bounding bias due to selection. Epidemiology 30(4):509–516

Spirtes P, Glymour C, Scheines R (2000) Causation, Prediction and Search, 2nd edn. MIT Press

Staplin N, Herrington WG, Judge PK, Reith CA, Haynes R, Landray MJ, Baigent C, Emberson J (2017) Use of causal diagrams to inform the design and interpretation of observational studies: An example from the study of heart and renal protection (SHARP). Clinical Journal of the American Society of Nephrology 12(13):546–552

Steen J, Loeys T, Moerkerke B, Vansteelandt S (2017) Flexible mediation analysis with multiple mediators. American Journal of Epidemiology 186(2):184–193

Stensrud MJ, Dukes O (2022) Translating questions to estimands in randomized clinical trials with intercurrent events. Statistics in Medicine 41(16):3211–3228

Stensrud MJ, Valberg M, Røysland K, Aalen OO (2017) Exploring selection bias by causal frailty models: The magnitude matters. Epidemiology 28(3):379–386

Stensrud MJ, Hernán MA, Tchetgen Tchetgen EJ, Robins JM, Didelez V, Young JG (2021) A generalized theory of separable effects in competing event settings. Lifetime Data Analysis 27(4):588–631

Stensrud MJ, Young JG, Didelez V, Robins JM, Hernán MA (2022) Separable effects for causal inference in the presence of competing events. Journal of the American Statistical Association 117(537):175–183

Stensrud MJ, Nevo D, Obolski U (2024) Distinguishing immunologic and behavioral effects of vaccination. Epidemiology 35(2):154–163

Sterne JA, Hernán MA, Ledergerber B, Tilling K, Weber R, Sendi P, Rickenbach M, Robins JM, Egger M (2005) Long-term effectiveness of potent antiretroviral therapy in preventing AIDS and death: A prospective cohort study. The Lancet 366(9483):378–384

Stokes T, Steele R, Shrier I (2022) Causal simulation experiments: Lessons from bias amplification. Statistical Methods in Medical Research 31(1):3–46

Stuart E (2010) Matching methods for causal inference: A review and a look forward. Statistical Science 25(1):1–21

Suissa S (2008) Immortal time bias in pharmacoepidemiology. American Journal of Epidemiology 167(4):492–499

Swanson SA, Miller M, Robins JM, Hernán MA (2015) Definition and evaluation of the monotonicity condition for preference-based instruments. Epidemiology 26(3):414–420

Taubman SL, Robins JM, Mittleman MA, Hernán MA (2009) Intervening on risk factors for coronary heart disease: An application of the parametric g-formula. International Journal of Epidemiology 38(6):1599–1611

Tchetgen Tchetgen EJ, Ying A, Cui Y, Shi X, Miao W (2020) An introduction to proximal causal learning. arXiv preprint arXiv:200910982

Tennant PW, Murray EJ, Arnold KF, Berrie L, Fox MP, Gadd SC, Harrison WJ, Keeble C, Ranker LR, Textor J, Tomova GD, Gilthorpe MS, Ellison GTH (2021) Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: Review and recommendations. International Journal of Epidemiology 50(2):620–632

Textor J, Van der Zander B, Gilthorpe MS, Liśkiewicz M, Ellison GT (2016) Robust causal inference using directed acyclic graphs: The R package 'dagitty'. International Journal of Epidemiology 45(6):1887–1894

Toh S, Hernández-Díaz S, Logan R, Robins JM, Hernán MA (2010) Estimating absolute risks in the presence of nonadherence: An application to a follow-up study with baseline randomization. Epidemiology 21(4):528–539

VanderWeele T (2015) Explanation in Causal Inference: Methods for Mediation and Interaction. Oxford University Press

VanderWeele TJ (2020) Hill's causal considerations and the potential outcomes framework. Observational Studies 6(2):47–54

VanderWeele TJ, Ding P (2017) Sensitivity analysis in observational research: introducing the e-value. Annals of Internal Medicine 167(4):268–274

VanderWeele TJ, Hernán MA (2013) Causal inference under multiple versions of treatment. Journal of Causal Inference 1(1):1–20

VanderWeele TJ, Shpitser I (2013) On the definition of a confounder. The Annals of Statistics 41(1):196–220

VanderWeele TJ, Tchetgen Tchetgen EJ (2017) Mediation analysis with time varying exposures and mediators. Journal of the Royal Statistical Society Series B: Statistical Methodology 79(3):917–938

VanderWeele TJ, Vansteelandt S, Robins JM (2014) Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. Epidemiology 25(2):300–306

Vansteelandt S, Didelez V (2018) Improving the robustness and efficiency of covariate-adjusted linear instrumental variable estimators. Scandinavian Journal of Statistics 45(4):941–961

Vansteelandt S, Sjölander A (2016) Revisiting g-estimation of the effect of a time-varying exposure subject to time-varying confounding. Epidemiologic Methods 5(1):37–56

Vansteelandt S, Linder M, Vandenberghe S, Steen J, Madsen J (2019) Mediation analysis of time-to-event endpoints accounting for repeatedly measured mediators subject to time-varying confounding. Statistics in Medicine 38(24):4828–4840

Vowels MJ, Camgoz NC, Bowden R (2022) D'ya like DAGs? A survey on structure learning and causal discovery. ACM Comput Surv 55(4):1–36

Waernbaum I (2012) Model misspecification and robustness in causal inference: Comparing matching with doubly robust estimation. Statistics in Medicine 31(15):1572–1581

Wager S, Athey S (2018) Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association 113(523):1228–1242

Wang A, Nianogo RA, Arah OA (2017) G-computation of average treatment effects on the treated and the untreated. BMC Medical Research Methodology 17(3):1–5

Wang SV, Schneeweiss S, Initiative RD (2023) Emulation of randomized clinical trials with nonrandomized database analyses: Results of 32 clinical trials. JAMA 329(16):1376–1385

Watkins S, Jonsson-Funk M, Brookhart MA, Rosenberg SA, O'Shea TM, Daniels J (2013) An empirical comparison of tree-based methods for propensity score estimation. Health Services Research 48(5):1798–1817

Westreich D (2017) From patients to policy: Population intervention effects in epidemiology. Epidemiology 28(4):525–528

Westreich D, Greenland S (2013) The table 2 fallacy: Presenting and interpreting confounder and modifier coefficients. American Journal of Epidemiology 177(4):292–298

Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, Curtis HJ, Mehrkar A, Evans D, Inglesby P, et al (2020) Factors associated with COVID-19-related death using OpenSAFELY. Nature 584(7821):430–436

Witte J, Didelez V (2019) Covariate selection strategies for causal inference: Classification and comparison. Biometrical Journal 61(5):1270–1289

Witte J, Henckel L, Maathuis MH, Didelez V (2020) On efficient adjustment in causal graphs. Journal of Machine Learning Research 21(246):1–45

Witte J, Foraita R, Didelez V (2022) Multiple imputation and test-wise deletion for causal discovery with incomplete cohort data. Statistics in Medicine 41(23):4716–4743

Wooldridge JM (2010) Econometric Analysis of Cross Section and Panel Data. MIT Press

Yland JJ, Chiu YH, Rinaudo P, Hsu J, Hernán MA, Hernández-Díaz S (2022) Emulating a target trial of the comparative effectiveness of clomiphene citrate and letrozole for ovulation induction. Human Reproduction 37(4):793–805

Young JG (2024) Story-led causal inference. Epidemiology 35(3):289–294

Young JG, Cain LE, Robins JM, O'Reilly EJ, Hernán MA (2011) Comparative effectiveness of dynamic treatment regimes: An application of the parametric g-formula. Statistics in Biosciences 3:119–143

Young JG, Hernán MA, Robins JM (2014) Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data. Epidemiologic Methods 3(1):1–19

Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernán MA (2020) A causal framework for classical statistical estimands in failure-time settings with competing events. Statistics in Medicine 39(8):1199–1236

Zetterstrom S, Waernbaum I (2022) Selection bias and multiple inclusion criteria in observational studies. Epidemiologic Methods 11(1):1–21