More Than Just Associations: An Introduction to Causal Inference for Sport Science

Master thesis

From

Simon Nolte

German Sport University Cologne

Cologne 2024

Thesis supervisor:

Dr. Oliver Jan Quittmann

Institute of Movement and Neurosciences

Affirmation in lieu of an oath

Herewith I affirm in lieu of an oath that I have authored this Bachelor thesis independently and did not use any other sources and tools than indicated. All citations, either direct quotations or passages which were reproduced verbatim or nearby-verbatim from publications, are indicated and the respective references are named. The same is true for tables and figures. I did not submit this piece of work in the same or similar way or in extracts in another assignment.

_____

Personally signed

Abstract

Zusammenfassung (German Abstract)

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Relevance

Empirical research is acquiring knowledge through systematic observations by analyzing data. Data analysis typically encompasses three primary tasks: description, prediction, and causal inference (Carlin & Moreno-Betancur, 2023; Hernán et al., 2019). Description means characterizing features in a subset of a population. Prediction means forecasting outcomes based on available data. Causal inference means making claims about causality — what would have happened under different circumstances.

Most research in sport science is of causal nature. We want to understand how sports works with the ultimate goal to intervene: If we understand why certain people or teams are winning a competition, we can use that knowledge to adjust training and tactics. Likewise, in health contexts, we seek for sport intervention that change an individual's fitness to ultimately increase well-being compared to if no intervention were undertaken. Ultimately, we are interested in potential outcomes — what would have happened if the team had played different or if the individual had undergone a different training. This exactly is causal thinking.

Research has devised a framework for conducting studies that can infer causality without knowledge of the exact underlying causal mechanisms: the randomized controlled trial (RCT). But in sport science, RCTs are often not feasible, because of the difficulty or undesirability of implementing randomized interventions, particularly in the context of elite sports (Bullock et al., 2023). Consequently, causality must often be inferred through alternative designs, such as observational studies. The field of causal inference offers tools for this particular task.

An association on its own does not inherently indicate causality, echoing the famous adage: "correlation does not imply causation." Associations observed in data may indeed stem from causality, but they can also arise from different types of bias, resulting in spurious associations. Conversely, causation does not necessarily imply correlation. Genuine causal relationships might remain obscured within the data. Distinguishing between associations and causal relationships necessitates looking beyond the data itself.

Causal data analysis requires something that is not relevant to most description and prediction tasks: A scientific model informed by expert-domain knowledge, that depicts the causal nature of the phenomena under investigation. This causal model serves as the foundation for all causal inference. By adhering to the rules implied by the causal model, we can analyze our data in a manner that allows for the estimation of causal effects. Methods of causal inference are vital to estimate causal effects from observational data. But they can also aid in designing and analyzing experiments, and even provide benefits for description and prediction analyses.

As all statistical analyses, causal modeling is not free of assumptions. Those are assumptions about the underlying data, but also about the underlying data generative process (the

world in which the data have been created). Causal modeling requires to think more clearly about these assumptions before conducting an analysis, and is in general more transparent in communicating them (Grosz et al., 2020). In a way, this is a more honest way of doing inference than relying on non-causal language when inferring causality was the actual research goal (Hernán, 2018).

I will start by establishing a working definition of causality and by providing an overview of causal inference as a research field, with its history and popular frameworks. Following this, I will outline recent applications of causal inference across various disciplines with a focus on the (sparse) literature of causal inference in sport science.

1.2 Previous Research

What causality actually means is a merely philosophical question (Illari & Russo, 2014). For the sake of this thesis, we use the framework of potential outcomes to define causality (Rubin, 1974). If we intervene on a variable and this leads to changes compared to if we had not intervened, we can define the intervention as causing the outcome. A causal effect is therefore defined by the comparison between two states, what has actually happened, and what would have potentially happened under different intervention. The intervention itself does not need to be actually possible to conduct, it can be purely hypothetical. For example, e.g., if we define the causal effect of biological sex on endurance performance we are actually asking: If we could intervene on an individual's sex (by changing it), what difference in endurance performance would we expect. We can state this without actually being able to change biological sex (when defined via chromosome[1]).

It can be easy to define causal effects, but difficult to estimate them. For estimation, we can only use real data and not hypothetical. We still want to estimate the difference between potential outcomes, with the caveat that for each unit of observation we only have one actual outcome available. Essentially, causal inference can be viewed as a missing data problem (Ding & Li, 2018). The most straightforward way to deal with this problem is using a randomization controlled research design[2], but often this is impossible or impractical.

Fisher (1925) was the first to suggest randomization as the basis to inference of causal effects in experiments. Randomized controlled designs quickly became the gold standard of experimental research (Cochran & Cox, 1957). Possibly until the 1970s it remained the common view that causal effects can only validly studied in randomized experiments, and not in observational studies. But based on the earlier invention of potential outcome notation by Neyman

[1] For the mathematical notation of (conditional) independence, see the appendix.

[2] There are of cause examples, where causality can be bidirectional. For example in feedback loops, such as the price and demand models in economy, changes in price cause changes in demand and the other way around. But even in this case one can argue that these are essentially two different paths of causality, that happen sequentially if observed with enough precision. For this thesis we will not deal with feedback systems, but stick with simpler models that assume purely directional causality.

(1923), Rubin (1974) provided a framework for estimating causal effect from both experimental and observation data. This framework later termed the 'Rubin Causal Model' (Holland, 1986) remains one of the predominant approaches to causal inference from observational data (see Appendix for the mathematical notation of this framework).

Another approach to causal inference is the use of graphical models. Pioneered by Pearl (1993, 1995), directed acyclic graphs (DAGs) have become a popular tool to assist estimating causal effects. They serve as an easy tool to aid estimating causal effect (Shrier & Platt, 2008). The graph-based approach has been criticized for being unnecessary (Rubin, 2022) or requiring a vast of (often not considered) assumptions (Dawid, 2010), yet it is popular in many fields (Morgan & Winship, 2014). Other approaches to causal inference aim to bring the potential outcome framework into a graph form (Richardson & Robins, 2013), or are less structural in that they neither require potential outcomes nor graphs (Dawid, 2000). Discussion about the different frameworks of causal inference can be found elsewhere. In this thesis I will often follow Pearl's graph-based approach (Pearl, 2009), because it is in my view the most intuitive and accessible way of learning causal inference[3], but I will also consider ideas and specific methods from the potential outcome framework (Angrist & Pischke, 2009).

Causal inference, whether in the framework of potential outcomes or graphical representations is considered one of the most influental statistical ideas of the past decades (Gelman & Vehtari, 2021). While the potential outcome framework dominates contemporary economic research (Imbens, 2020), graph-based causal inference has gained wide popularity in other fields, such as epidemiology (Greenland et al., 1999; Tennant et al., 2021), psychology (Rohrer, 2018), and sociology (Morgan & Winship, 2014). These fields share similar challenges with sport science: They study complex systems (i.e., humans) and often have to rely on observational data for inference. Despite its potential value, the use of causal inference in sport science is so far limited.

It is unsurprisingly that the most active research areas of causal inference in sport science are at the intersection to the field of epidemiology (Lynch et al., 2020), mostly in the area of injury research. For researching the prevention of injuries, calls to use causal modeling are frequent (Kalkhoven, 2024; Nielsen et al., 2020; Shrier, 2007), but its actual use is rare (Rommers et al., 2021). Shrier (2007) and Hopkins (2008) were the first to propose graphical causal models for sport science. The unusual presentation in form of a slideshow by Hopkins (2008), the narrow scope on injuries by Shrier (2007), and the lack of an accessible and focused reasoning by both may have limited the impact of their ideas. Recently, Steele et al. (2020) undertook a new try to highlight the need of causal thinking and modeling in sport science. Embedded in a general model of sport research (Bishop, 2008), they used an example of strength training to introduce key elements of causal inference such as potential outcomes and causal graphs. But they rather focused on the process of answering a specific research question (in part

[3] Exposure here is the medical term for what is often named the "independent variable" in a statistical model. It is the variable that we image our intervention on, so it does not need to be an actual *exposure* in the strict sense of the word.

utilizing causal inference tools) rather than explicitly introducing causal inference to sport science.

In a recent extensive debate revolving around the causal effect of muscle hypertrophy on strength, all author groups agreed on the difficulties of distinguishing associations and causal relations, and the challenge of adequately controlling experiments or using observational data for causal statements (Balshaw et al., 2017; Buckner et al., 2017; Dankel et al., 2018; Loenneke et al., 2019; Taber et al., 2019). Yet none of them mentioned causal inference as a potential way to deal with these problems until a later publication by Nuzzo et al. (2019), again exemplifying the potential usefulness, but currently low dissemination of causal inference methods in sport science. In a recent article, Kalkhoven (2024) calls for the use of graphical causal models in sports injury research. Kalkhoven (2024) concludes his text with an appeal to all sport scientist to engage with the field of causal inference. This thesis will provide sport-scientist with an accessible, field-specific introduction to causal inference.

1.3 Aim

The aim of this thesis is to bring the methods of causal inference to sport science. The overachieving goal is to demonstrate the utility and necessity of causal inference methods for data analysis in sport science. I start with demonstrate key concepts of causal models using directed acyclic graphs by introducing confounders, colliders, and conditioning rules. I then revisit two published observation studies from the field of endurance running from a causal inference perspective. Finally, I will discuss opportunities that causal inference brings to sport science as well as challenges and limitations of adopting such approaches.

I aim to make the thesis as accessible as possible to readers who are new to causal inference. Detailed methodologies of modeling and mathematical formulations will be included in the appendices. My objective is to ensure that the thesis is understandable for any sport scientist with some basic statistical education. Instead of critiquing current statistical practices in sport science, the objective of this work is to showcase the effectiveness of methods that extend beyond these practices.

2 Theoretical Background

2.1 Causality, Associations, and (In)dependence

In the preceding section we defined causality as a concept involving hypothetical interventions. When intervening on a variable X results in changes in another variable Y we assert that X causes Y. From a statistical standpoint, X and Y become dependent[4]. Conversely, an association, implies that X and Y share information; Knowledge about one variable implies knowledge about the other variable, and *vice versa*. Crucially, associations lack directionality, whereas causality is typically understood as directional[5]. Causality can be one reason for associations to arise, but other reasons for associations exist, for example a shared common cause. Consequently, both causal relations and spurious relations can produce associations and render variables dependent. It is the underlying causal model that can distinguish between mere associations and causal relationships.

2.2 Graphical Causal Models

Graphical models provide a straightforward framework for conceptualizing causal systems. Pioneered by Pearl (1995), they offer a visual representation of causal relationships, which eases development and comprehension of causal models. A graphical causal model visualizes the exposure, outcome, covariates, and their (assumed) causal relationship. In the following, we will typically denote the exposure[6] as X, the outcome as Y, and covariates with other letters. Variables in graphical causal model are linked by arrows. An arrow between X and Y means that a direct causal relationship between the two is possible (see Figure 1) . The direction of the arrow indicates the direction of causality. As depicted in Figure 1, $X \rightarrow Y$ means that X causes Y (and not the other way around). In accordance with our definition of causality, this implies that intervening on X should results in a change in Y.

The direction of causality has to be determined by theoretical knowledge; it cannot be found in the data alone. Suppose that in our first example in Figure 1, X represents biological sex and Y denotes endurance performance. It seems apparent that a causal relationship exists between them (though it is undoubtedly much more complicated than that depicted in this simple model). However, the fact that it is sex that causes performance — and not

[4] For the mathematical notation of (conditional) independence, see the appendix.

[5] There are of cause examples, where causality can be bidirectional. For example in feedback loops, such as the price and demand models in economy, changes in price cause changes in demand and the other way around. But even in this case one can argue that these are essentially two different paths of causality, that happen sequentially if observed with enough precision. For this thesis we will not deal with feedback systems, but stick with simpler models that assume purely directional causality.

[6] Exposure here is the medical term for what is often named the “independent variable” in a statistical model. It is the variable that we image our intervention on, so it does not need to be an actual *exposure* in the strict sense of the word.

the other way around — is based purely on theoretical knowledge and understanding of the world. There are neither randomized trials available (because you cannot randomly assign sex), nor are controlled interventions feasible (because you cannot easily intervene on sex) possible. Ultimately, the direction of causality is an assumption by the researcher.

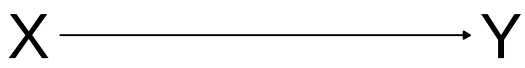

Figure 1: A simple graphical causal model with two variables. The variable X (exposure) is assumed to cause the variable Y (outcome). No other variables are believed to influence this process.

Causal systems in the world are typically more complex than consisting of only exposure and outcome, and thus the graphical causal models depicting them are more complex as well. A slightly more complex graph is displayed in Figure 2. X and Y are not directly linked anymore, but are connected indirectly via B. This sequence $X \rightarrow B \rightarrow Y$ is called a *causal path*. We will later see, that some models also have non-causal paths.

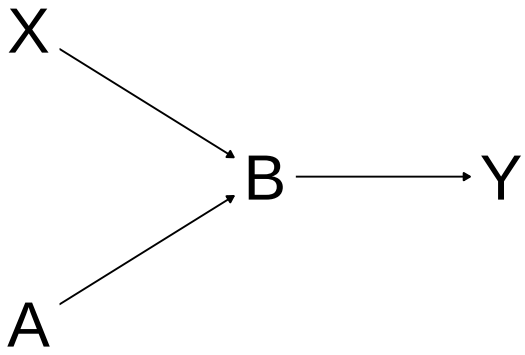

Figure 2: A more complex graphical causal model with four variables. X and A both cause B, which in turn causes Y.

The graph in Figure 2 is called a directed acyclic graph (DAG). It is directed, because all paths have arrows (establishing the direction of causality). It is acyclic, because there are no circular paths in it. Finally, it is a graph. All graphs in this thesis will be DAGs, as many of the concepts presented herein require this, and most research problems can be adequately formulated using them. More important than which arrows a DAG contains is which it arrows are absent. A DAG should depict all *potential* causal relations relevant to the research question. If two variables are not connected, we explicitly assume that they do not causally relate to each

other[7]. For example, in Figure 2, there is no direct link between X and A, or between X and Y.

DAGs tell a story. For example, we can assign the variables in Figure 2 to a simple model of endurance performance. Let X be the biological sex, A the nutrition status, B the physiological capacity to perform endurance tasks, and Y the endurance performance in a competition. Our model assumes that sex and nutrition both directly affect the physiological capacity, which subsequently affects performance. Conversely, it assumes that sex and nutrition are not causally related, and that neither direct affects performance; rather their effect are indirect mediated through physiological capacity.

2.3 Modeling Causal Systems & Error Terms

DAGs serve as an abstract concept to describe research problems. This level of abstraction allows to plan a study and its data analysis on a conceptual level. However, for the actual data analysis or demonstration purposes, a DAG has to be filled with data and functions. One way to fill a DAG, is to think of it as a linear regression model (or more precise, as a linear structural equation model[8]). For instance, the easiest DAG in the form $X \rightarrow Y$ can be analyzed as the linear regression $Y \sim X + \epsilon$. This assumes, that Y is an additive linear combination of other variables. In this thesis, we will analyze all DAGs as linear models, keeping in mind, that other types of models (e.g., non-linear relationships, interactions) are possible. A special role in these linear models has the error term ϵ.

If we knew the true causal model and could measure all variables perfectly, we could exactly determine all causal effects. In reality, this is impossible. One of the main reasons for this is the presence of unobserved factors (errors), that influence our relevant variables in the model. This could also be factors like random measurement error or biological variability. Furthermore, since we can always only investigate causal effects in a sample of the population, our research will only result in an estimate of the true causal effect we seek to determine (the estimand).

Just like in any statistical analysis, we aim to obtain unbiased and precise estimates. Unbiasedness means that, on average, our estimate will correspond to the true value of the estimand. Precision means, that the estimate will have a small variance, or in other words,

[7] In other words, if two variables are connected they might or might not have a causal relation. If two variables are not connected we assume that they definitely have no causal relation. This is a strong assumption in many scenarios, but when reasoned properly the foundation of causal inference.

[8] A linear structural equation model (SEM) is essentially a linear regression model with additional causal assumptions (Bollen & Pearl, 2013). All DAGs (and many of the research question from the potential outcome framework of causal inference) can be rewritten as a linaer SEM, assuming the additional constraints of linearity and additive components, though SEMs can in theory also be generalized to a non-linear setting (Bollen & Pearl, 2013). The analysis of DAGs via linear SEM profs to bring insights into causal systems, both in theory (e.g., Ding & Miratrix, 2015) and in practice [e.g.,].

that repeated measurement will yield similar estimates. Random error terms add imprecision, but not bias to our model[9]. We will later encounter scenarios that introduce bias.

To illustrate the concepts of precision and bias of causal effect estimates, I will use toy data simulations in the following. These simulations generate samples of data (with n = 100) corresponding to the simulated linear model including random error terms. Each simulated sample (k = 1000) is modeled to yield a single causal effect estimate. I then visualize the distribution of simulated effect estimates. Further details of the simulation procedure can be found in Section A.2. Figure 3 demonstrates how unobserved factors (random error terms) add uncertainty to a causal effect estimate. In the absence of random error terms, each sample would give the exact true causal effect. However, in the presence of random error terms, some samples will give estimates that deviate from the true causal effect.

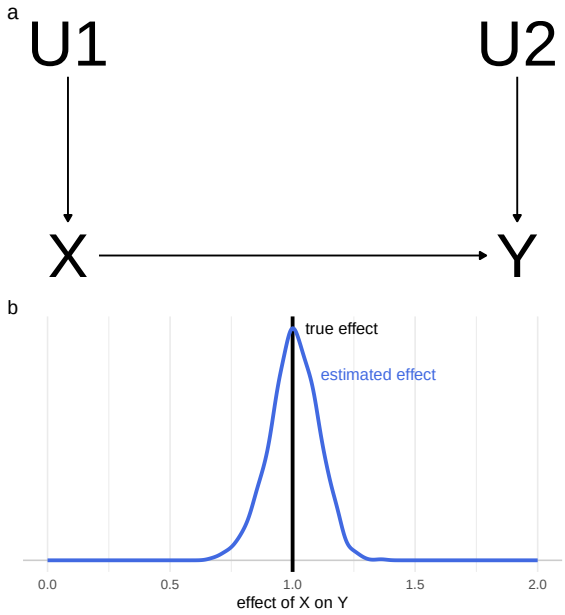

Figure 3: A simple causal path, with random error. (a) X causes Y, but both variables are influenced by other unobserved variables (random error). (b) A simulation of the model. The density plot shows the distribution of k = 1000 simulations of the model with random error terms. The random error adds uncertainty to the estimate of the causal effect, but no bias (i.e., on average, the true causal effect can be correctly estimated).

Precision in causal effect estimates is higher in simpler models. This is primarily because simpler models have fewer random error terms. This can be demonstrated by comparing a simple causal relation with a causal path (a chain). Along a causal path, information is typically lost, even if the causal effects remain unchanged. This loss of information is caused by the additional error terms of intermediate variables (see Figure 4). Chains therefore introduce uncertainty into causal effect estimates, but do not induce bias[10].

[9] At least this is an extremely common assumption. See the appendix for the mathematical notation.
[10] View the appendix for a mathematical proof.

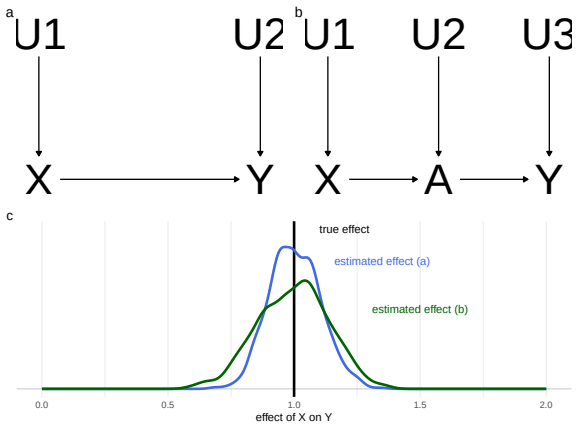

Figure 4: Random errors in a causal path. (a) X causes Y directly. Both variables are influenced by random errors. (b) X causes Y via A. All three variables are influenced by random errors. (b) A simulation of the effect of X on Y in both models. The chain introduces additional uncertainty in the effect estimate, but no bias.

For an example from sport science think of two different causal effects. First, the effect of a running intervention on mitochondrial density. Second, the effect of a running intervention on endurance performance. Even if we assume, that the effect in the second case is entirely mediated trough mitochondrial density (i.e., $intervention \rightarrow density \rightarrow performance$), the effect on endurance performance is harder to estimate. The primary reason is, that endurance performance will be influenced by additional unobserved factors, that will not influence mitochondrial density, for example motivation, pacing, or day-to-day variability.

Examining the causal model in Figure 4, we have to reconsider that the arrows drawn in a DAG are just as noteworthy as the arrows not drawn. In this example, all unobserved error terms are parent nodes, meaning that they are not influenced by any other relevant variable, including each other. This is a general assumption regarding unobserved error terms: We assume random errors to be uncorrelated. As soon as errors influence each other (directly or via other variables), we should model them explicitly[11].

2.4 Conditioning

Causal paths can be blocked by conditioning on intermediate variables. Take for example the causal path $X \rightarrow A \rightarrow Y$. Let X be the stroke volume of the heart, A the maximum oxygen uptake, and Y the endurance performance in a competition. We assume that all of the causal effect of stroke volume on endurance performance is mediated via maximum oxygen uptake. However, if we condition on maximum oxygen uptake, no relationship between stroke volume

[11] The assumption of uncorrelated error terms is also common in applied statistics outside of causal inference. If error terms are correlated this complicates the estimation of effects. We can model correlated error terms in a DAG by creating a node for an unobserved variable. Another way to investigate the consequences of correlated error terms in linear SEMs is by drawing them from a multivariate normal distribution with an appropriate covariance matrix (e.g. in Ding & Miratrix, 2015).

and endurance performance remains. Conditioning on the intermediate variable A effectively blocks the causal path between X and Y, rendering the causal effect of stroke volume on endurance performance non-existing.

Several ways to condition on variables exist[12]. An experimental approach is to stratify the sample by the variable. For instance, if we would only investigate athletes with a similar maximum oxygen uptake, we would anticipate that the relationship between stroke volume and endurance performance would deminish. A modeling approach of conditioning on a variable is to include it in the statistical model. For example, modeling $Y \sim A + X + \epsilon$ would effectively block the causal effect of X on Y (see Figure 5)[13].

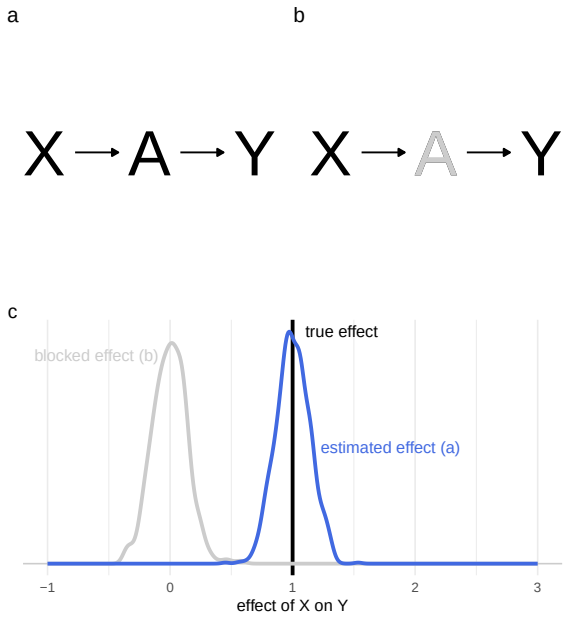

Figure 5: A causal path blocked by conditioning. (a) X causes Y via A. (b) The causal path is blocked, because the analysis conditions on A. As all affects of X on Y trail through A, no causal effect remains. (c) A simulation of the effect of X on Y in both models. Blocking removes the true causal effect entirely.

One of the main goals of causal inference using graph-based methods is identification — to identify which variables should be conditioned on. This process is crucial to provide unbiased and accurate effect estimates. Depending on the structure of the model, certain variables can introduce bias if left unconditioned, while others bias the estimate if conditioned on. The following section will further elucidate these concepts by introducing confounders and colliders.

[12]The mathematical notation of conditioning is straightforward (see Appendix). The experimental ways to condition are diverse and include methods that can be used during experimental design or data analysis.
[13]Other popular ways of conditioning include matching, …

2.5 Confounders and Colliders

Confounders are variables that causally influence both the exposure and the outcome (see Figure 6 a). The confounder creates a spurious (non-causal) association between the exposure and the outcome. Conceptually, a confounder gives a set of similar information (knowledge) to both exposure and outcome. This leads to both sharing common information, irrespective of their true causal relationship. This leads to bias in the causal effect estimate.

Confounders can be controlled for by conditioning on them in the model. This removes the entire bias and preserves the true causal relationship. Let's take an example illustrated in Figure 6. We are interested in the relationship between the (average) 5000-m time trial speed and the (average) 100-m sprint speed. We assume, that being fast in an endurance task reduces the ability to sprint fast, and thus decreases the 100-m speed. Therefore, we are interested in the causal relationship between X (endurance speed) and Y (sprinting speed). Note that this is a very simplistic causal model, as we could also model the unobserved ability to sprint and ability to perform endurance tasks, as well as their potential causes.

Our model has a collider A, representing biological sex. Based on expert knowledge, we understand that sex causally influences both sprinting and endurance performance, mainly via anthropometry and physiology. Sex thus biases the causal relationship between sprinting and endurance performance. To remove this bias, the analysis must control for sex. For a discrete variable like sex is typically documented as, controlling for means in practice stratifying the analysis by it. Assuming our causal model is correct — which holds of course not true in our toy example here — controlling for sex gives us the true (unbiased) causal relationship between endurance and sprinting performance.

Colliders pose a more subtle form of bias. A collider is a variable that is causally influenced by both the exposure and the outcome (see Figure 7 a). Per se, colliders do not yield harm. But when they are conditioning on they introduce bias into a model[14]. This collider bias can be understood by the following: A collider combines knowledge from both its source, the exposure and the outcome, and thus also of their causal relationship. If this combined knowledge is being removed from a model by conditioning on the collider, then some of the actual causal relationship between exposure and outcome is also removed.

Consider the causal relationship between X as the post-lactate in a ramp test and Y as the maximum oxygen uptake in the same ramp test. Essentially, our question is if more lactate causes a different (higher or lower) maximum oxygen uptake. In our model, both lactate and VO2max influence the maximum velocity achieved in the ramp test. This appears reasonable, as individuals with a more capable glycolytic or oxidatative energy metabolism are likely to outperform their counterparts that have neither in term of the maximum velocity. The maximum velocity attained is thus the collider B. Conditioning on it will bias our model.

[14] Equally, conditioning on a descendant of a collider introduces bias (though generally not that large as when conditioning on the collider itself).

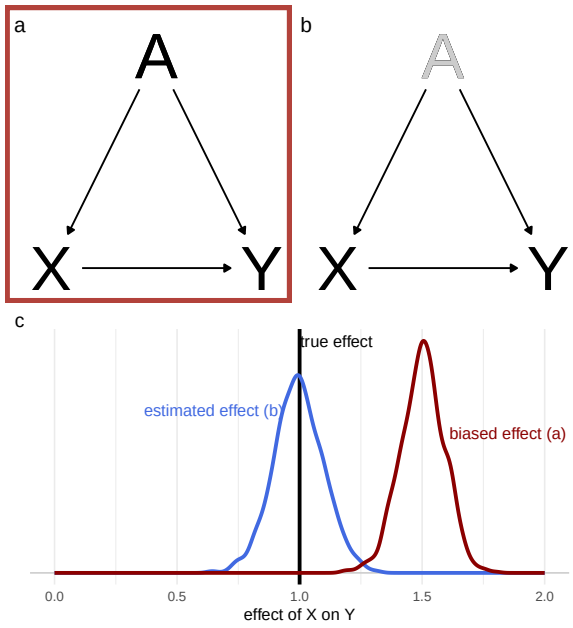

Figure 6: A graphical example of confounding. Both X and Y share a common cause A. (a) This confounder biases determining the causal effect of X on Y. (b) Conditioning on A removes the bias in the analysis. (c) A simulation of the effect of X on Y in both models.

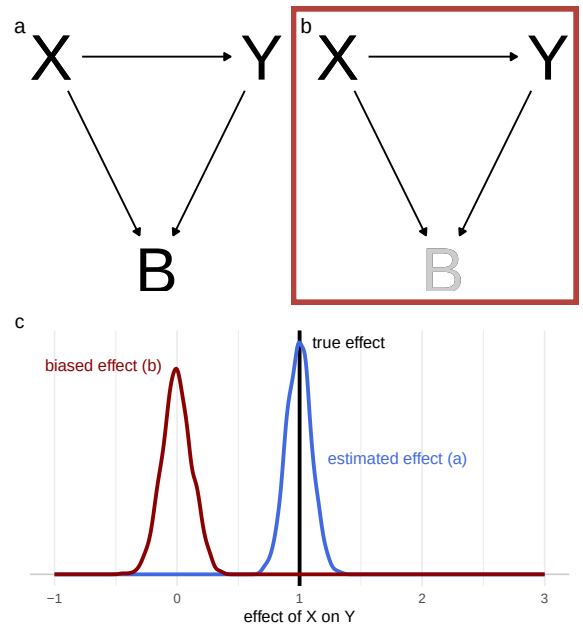

Figure 7: A graphical example of collider bias. Both X and Y directly affect the collider B. (a) As long as B is not conditioned on, the causal effect of X on Y is unbiased. (b) Conditioning on B will introduce bias in the model. (c) A simulation of the effect of X on Y in both models.

2.6 Conditioning Rules: The Backdoor Criterion

Building on the concepts of confounders and colliders, we can derive more general rules for determining the optimal conditioning set for a given causal model. The most famous of these conditioning rules is the backdoor criterion (CITE!!!). It works by first identifying all non-causal paths (backdoor paths), and second blocking all of them.

A non-causal path is any path between X and Y that starts with an arrow pointing into X. A non-causal path is open, if it has no collider or no variable conditioned on in it. It can be blocked (closed) by conditioning on a non-collider. For example, in Figure 6 $X \rightarrow Y$ is a causal path, whereas $X \leftarrow A \rightarrow Y$ is a non-causal path. The non-causal path can be blocked by conditioning on A, fulfilling the backdoor criterion, and thus providing an unbiased estimate of the causal effect of X on Y.

On the contrary, non-causal paths are blocked by default if they contain a collider. For example, in Figure 8 one non-causal path$X \leftarrow A \rightarrow B \leftarrow Y$ exists, but is is blocked by default because B is a collider. Therefore the backdoor criterion is satisfied and no conditioning is required (i.e., the minimal sufficient conditioning set is empty). If one would decide to condition on B in this scenario (for example if A were unobserved, and we chose to condition on all observed covariates), this would reopen the backdoor-path, biasing the estimate.

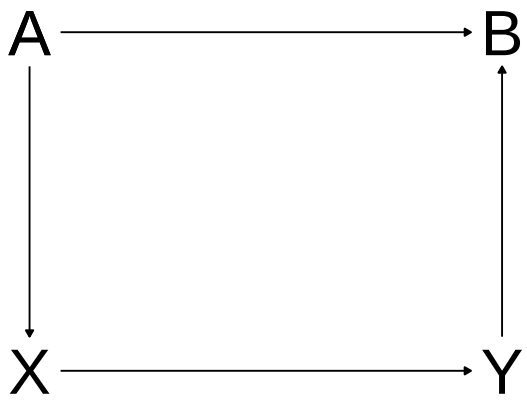

Figure 8: A graphical example of a backdoor path closed by default. The non-causal path via A and B contains a collider and is therefore closed. Conditioning on B would reopen the backdoor path.

The backdoor criterion helps to determine the variables that need to be conditioned on in graphical causal models of various complexity to obtain an unbiased estimate. These variables form the so-called minimal sufficient conditioning set. Conditioning on more variables than sufficient can increase precision in certain cases, but often brings the risk of introducing new bias or reducing precision. When certain variables in a DAG are unobserved, these can not be conditioned on. In this case it is possible that no minimal sufficient conditioning set exists that fulfills the backdoor criterion. Therefore; an unbiased estimation of the causal

effect given the assumed causal model is impossible. We will come back to the question of selecting conditioning variables in the DISCUSSION.

3 Methods

I conducted all analyses in this thesis using R version 4.3.1 (R Core Team, 2023) in the RStudio IDE version 2023.09.1.494 (Posit team, 2023). The thesis was written in Quarto version 1.3.450 (Allaire et al., 2023). The default settings and attached packages are documented in Appendix Section A.3. The DAGs in this thesis were drawn using the ggdag R package (Barrett, 2024), which is based on the software daggity (Textor et al., 2016). All source code of this project is available at GitHub.

4 Results

4.1 Example 1: Use of Different Running Shoes and Injury Risk

Malisoux et al. (2015)

4.2 Example 2: Determinants of Marathon Pacing

March et al. (2011)

5 Discussion

5.1 General Applications of Causal Inference in Sport Science

5.1.1 Causality in Observational Data

5.1.2 Identification of variables to Condition on

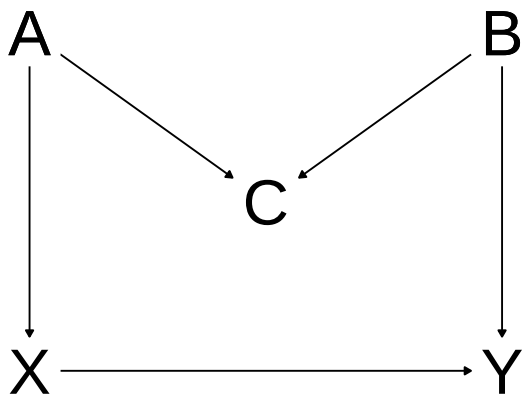

Figure 9: A graphical example of M bias. C is caused by both A and B, which also effect X and Y, respectively. In this scenario, a non-causal path exists, but it is closed, because C is a collider. When C is conditioned on, this would open the backdoor path, because the conditiong creates a spurious relationship between X and Y, so that both act together as a confounder.

5.2 Applicability of Special Causal Inference Methods in Sports

5.2.1 Instrumental Variables

5.2.2 Regression Discontinuity

5.2.3 Difference-in-difference

5.2.4 Synthetic Control

5.3 Challenges and Limitations

5.3.1 Need for Theoretical Models

5.3.2 Complex Systems

5.3.3 Small Samples

5.3.4 Data Quality

5.4 Perspectives and Further Possibilities

5.4.1 Modeling Missing Data and Measurement Error

5.4.2 Heterogenous Effects

5.4.3 Longitudinal Data

5.4.4 Understanding Big Data

5.4.5 Communicating Causality

5.5 Causal Modeling Workflows in Sport Science Practice

6 Conclusion

References

Allaire, J. J., Teague, C., Scheidegger, C., Xie, Y., & Dervieux, C. (2023). *Quarto*. https://doi.org/10.5281/zenodo.5960048

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion* (Illustrated Edition). Princeton Univers. Press. https://doi.org/10.1515/9781400829828

Balshaw, T. G., Massey, G. J., Maden-Wilkinson, T. M., & Folland, J. P. (2017). Muscle size and strength: debunking the "completely separate phenomena" suggestion. *European Journal of Applied Physiology*, *117*(6), 1275–1276. https://doi.org/10.1007/s00421-017-3616-y

Barrett, M. (2024). *ggdag: Analyze and create elegant directed acyclic graphs*. https://github.com/r-causal/ggdag

Bishop, D. (2008). An Applied Research Model for the Sport Sciences. *Sports Medicine*, *38*(3), 253–263. https://doi.org/10.2165/00007256-200838030-00005

Bollen, K. A., & Pearl, J. (2013). *Eight Myths About Causality and Structural Equation Models* (S. L. Morgan, Ed.; pp. 301–328). Springer Netherlands. https://doi.org/10.1007/978-94-007-6094-3_15

Buckner, S. L., Dankel, S. J., Mattocks, K. T., Jessee, M. B., Grant Mouser, J., & Loenneke, J. P. (2017). Muscle size and strength: another study not designed to answer the question. *European Journal of Applied Physiology*, *117*(6), 1273–1274. https://doi.org/10.1007/s00421-017-3615-z

Bullock, G. S., Ward, P., Hughes, T., Thigpen, C. A., Cook, C. E., & Shanley, E. (2023). Using Randomized Controlled Trials in the Sports Medicine and Performance Environment: Is It Time to Reconsider and Think Outside the Methodological Box? *Journal of Orthopaedic & Sports Physical Therapy*. https://doi.org/10.2519/jospt.2023.11824

Carlin, J. B., & Moreno-Betancur, M. (2023). *On the uses and abuses of regression models: A call for reform of statistical practice and teaching*. arXiv. https://doi.org/10.48550/arXiv.2309.06668

Cochran, W. G., & Cox, G. M. (1957). *Experimental designs* (2nd ed). John Wiley & Sons.

Dankel, S. J., Buckner, S. L., Jessee, M. B., Grant Mouser, J., Mattocks, K. T., Abe, T., & Loenneke, J. P. (2018). Correlations Do Not Show Cause and Effect: Not Even for Changes in Muscle Size and Strength. *Sports Medicine*, *48*(1), 1–6. https://doi.org/10.1007/s40279-017-0774-3

Dawid, A. P. (2000). Causal inference without counterfactuals. *Journal of the American Statistical Association*, *95*(450), 407–424. https://doi.org/10.2307/2669377

Dawid, A. P. (2010). Beware of the DAG! *Causality: Objectives and Assessment*, 59–86.

Ding, P., & Li, F. (2018). Causal inference: A missing data perspective. *Statistical Science*, *33*(2), 214–237. https://doi.org/10.1214/18-STS645

Ding, P., & Miratrix, L. W. (2015). To Adjust or Not to Adjust? Sensitivity Analysis of M-Bias and Butterfly-Bias. *Journal of Causal Inference*, *3*(1), 41–57. https://doi.org/10.1515/jci-

2013-0021

Fisher, R. A. (1925). *Statistical methods for research workers*. Oliver; Boyd.

Gelman, A., & Vehtari, A. (2021). What are the Most Important Statistical Ideas of the Past 50 Years? *Journal of the American Statistical Association*, *116*(536), 2087–2097. https://doi.org/10.1080/01621459.2021.1938081

Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal diagrams for epidemiologic research. *Epidemiology*, *10*(1), 37.

Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The Taboo Against Explicit Causal Inference in Nonexperimental Psychology. *Perspectives on Psychological Science*, *15*(5), 1243–1255. https://doi.org/10.1177/1745691620921521

Hernán, M. A. (2018). The c-word: Scientific euphemisms do not improve causal inference from observational data. *American Journal of Public Health*, *108*(5), 616–619. https://doi.org/10.2105/AJPH.2018.304337

Hernán, M. A., Hsu, J., & Healy, B. (2019). A second chance to get causal inference right: A classification of data science tasks. *CHANCE*, *32*(1), 4249. https://doi.org/10.1080/09332480.2019.1579578

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*(396), 945960. https://doi.org/10.1080/01621459.1986.10478354

Hopkins, W. (2008). Research designs: Choosing and fine-tuning a design for your study. *Sportscience*, *12*(1), 13.

Illari, P., & Russo, F. (2014). *Causality: Philosophical theory meets scientific practice*. OUP Oxford.

Imbens, G. W. (2020). Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics. *Journal of Economic Literature*, *58*(4), 1129–1179. https://doi.org/10.1257/jel.20191597

Kalkhoven, J. T. (2024). Athletic Injury Research: Frameworks, Models and the Need for Causal Knowledge. *Sports Medicine*. https://doi.org/10.1007/s40279-024-02008-1

Loenneke, J. P., Buckner, S. L., Dankel, S. J., & Abe, T. (2019). Exercise-Induced Changes in Muscle Size do not Contribute to Exercise-Induced Changes in Muscle Strength. *Sports Medicine*, *49*(7), 987–991. https://doi.org/10.1007/s40279-019-01106-9

Lynch, B. M., Dixon-Suen, S. C., Ramirez Varela, A., Yang, Y., English, D. R., Ding, D., Gardiner, P. A., & Boyle, T. (2020). Approaches to Improve Causal Inference in Physical Activity Epidemiology. *Journal of Physical Activity & Health*, *17*(1), 80–84. https://doi.org/10.1123/jpah.2019-0515

Malisoux, L., Ramesh, J., Mann, R., Seil, R., Urhausen, A., & Theisen, D. (2015). Can parallel use of different running shoes decrease running-related injury risk? *Scandinavian Journal of Medicine & Science in Sports*, *25*(1), 110–115. https://doi.org/10.1111/sms.12154

March, D. S., Vanderburgh, P. M., Titlebaum, P. J., & Hoops, M. L. (2011). Age, sex, and finish time as determinants of pacing in the marathon. *The Journal of Strength & Conditioning Research*, *25*(2), 386. https://doi.org/10.1519/JSC.0b013e3181bffd0f

Morgan, S. L., & Winship, C. (2014). *Counterfactuals and causal inference: Methods and*

principles for social research (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781107587991

Neyman, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles. *Ann. Agricultural Sciences*, 151.

Nielsen, R. O., Simonsen, N. S., Casals, M., Stamatakis, E., & Mansournia, M. A. (2020). Methods matter and the 'too much, too soon' theory (part 2): what is the goal of your sports injury research? Are you describing, predicting or drawing a causal inference? *British Journal of Sports Medicine*, *54*(22), 1307–1309. https://doi.org/10.1136/bjsports-2020-102144

Nuzzo, J. L., Finn, H. T., & Herbert, R. D. (2019). Causal Mediation Analysis Could Resolve Whether Training-Induced Increases in Muscle Strength are Mediated by Muscle Hypertrophy. *Sports Medicine*, *49*(9), 1309–1315. https://doi.org/10.1007/s40279-019-01131-8

Pearl, J. (1993). Comment: Graphical models, causality and intervention. *Statistical Science*, *8*(3), 266269. https://doi.org/10.1214/ss/1177010894

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, *82*(4), 669–688. https://doi.org/10.2307/2337329

Pearl, J. (2009). *Causality*. Cambridge university press. https://doi.org/10.1017/CBO9780511803161

Posit team. (2023). *RStudio: Integrated development environment for R*. Posit Software, PBC. http://www.posit.co/

R Core Team. (2023). *R: A language and environment for statistical computing*. https://www.R-project.org/

Richardson, T. S., & Robins, J. M. (2013). Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality. *CSSS Technical Report*, *128*, 1–150.

Rohrer, J. M. (2018). Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. *Advances in Methods and Practices in Psychological Science*, *1*(1), 27–42. https://doi.org/10.1177/2515245917745629

Rommers, N., Rössler, R., Shrier, I., Lenoir, M., Witvrouw, E., D'Hondt, E., & Verhagen, E. (2021). Motor performance is not related to injury risk in growing elite-level male youth football players. A causal inference approach to injury risk assessment. *Journal of Science and Medicine in Sport*, *24*(9), 881–885. https://doi.org/10.1016/j.jsams.2021.03.004

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701. https://doi.org/10.1037/h0037350

Rubin, D. B. (2022). Interview with Don Rubin. *Observational Studies*, *8*(2), 77–94. https://doi.org/10.1353/obs.2022.0009

Shrier, I. (2007). Understanding causal inference: The future direction in sports injury prevention. *Clinical Journal of Sport Medicine*, *17*(3), 220. https://doi.org/10.1097/JSM.0b013e3180385a8c

Shrier, I., & Platt, R. W. (2008). Reducing bias through directed acyclic graphs. *BMC Medical*

Research Methodology, *8*(1), 70. https://doi.org/10.1186/1471-2288-8-70

Steele, J., Fisher, J., & Crawford, D. (2020). Does increasing an athletes' strength improve sports performance? A critical review with suggestions to help answer this, and other, causal questions in sport science. *Journal of Trainology*, *9*(1), 20. https://doi.org/10.17338/trainology.9.1_20

Taber, C. B., Vigotsky, A., Nuckols, G., & Haun, C. T. (2019). Exercise-Induced Myofibrillar Hypertrophy is a Contributory Cause of Gains in Muscle Strength. *Sports Medicine*, *49*(7), 993–997. https://doi.org/10.1007/s40279-019-01107-8

Tennant, P. W. G., Murray, E. J., Arnold, K. F., Berrie, L., Fox, M. P., Gadd, S. C., Harrison, W. J., Keeble, C., Ranker, L. R., Textor, J., Tomova, G. D., Gilthorpe, M. S., & Ellison, G. T. H. (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. https://doi.org/10.1093/ije/dyaa213

Textor, J., Zander, B. van der, Gilthorpe, M. S., Liśkiewicz, M., & Ellison, G. T. (2016). Robust causal inference using directed acyclic graphs: The R package 'dagitty'. *International Journal of Epidemiology*, *45*(6), 1887–1894. https://doi.org/10.1093/ije/dyw341

A Appendix

A.1 Mathematical Background

A.1.1 Probability Theory

A random variable is a property we cannot absolutely predict. The probability of the random variable X is given by $Pr(X)$. An event is the assignment of a value to a random variable. The probability of event A given that event B has occurred is the conditional probability of A given B and is denoted by $Pr(A|B)$. The events A and B are statistically independent if the observation of B does not alter the probability of A, or $Pr(A|B) = Pr(A)$. Another way to note independence is $A \perp\!\!\!\perp B$. Two events are conditionally independent if they are independent given a third event C, implying that $Pr(A|B, C) = Pr(A|C)$. This conditional independence can also be denoted as $A \perp\!\!\!\perp B|C$.The expected value of a random variable X is the weighted probability of the values it can take denoted by $E(X)$.

A.1.2 Potential Outcome Notation

For simplicity, we use a binary variable that takes on the value 0 if a unit i received no treatment and the value 1 if the unit i received treatment. Every unit i has two potential outcomes Y_i^0 and Y_i^1. These outcomes are hypothetical, as each unit only can or cannot receive a treatment and therefore only one of the two potential outcomes is realized. The observed Y_i can be defined as $Y_i = (D_i - 1)Y_i^0 + D_i Y_i^1$ with D_i as the unit-specific treatment indicator. The individual causal effect of the treatment δ_i is defined as a comparison of the two potential outcomes for each unit $\delta_i = Y_i^1 - Y_i^0$. This poses a problem, as we never observe both potential outcomes for a single unit simultaneously, as thus cannot calculate δ_i. The average treatment effect is defined by $E(\delta_i) = E(Y_i^1 - Y_i^0) = E(Y_i^1) - E(Y_i^0)$. Making the strong assumptions that $E(Y_i^1|D = 0) = E(Y_i^1|D = 1)$ and $E(Y_i^0|D = 0) = E(Y_i^0|D = 1)$ we get an unbiased estimate of the average treatment effect by calculating the simple differences in mean $E(Y_i^1|D = 1) - E(Y_i^0|D = 0)$, which are both observed quantities. Or in other words, we obtained an unbiased estimate of the causal treatment effect by comparing the mean of the treatment group and the mean of the untreated group, if we assume that the mean of the treatment group equals the mean that the untreated group would have had if they had received the treatment (and vice versa). This implies, that the assignment of treatment was independent of the potential outcomes, or $(Y^0, Y^1) \perp\!\!\!\perp D$, something that could, for example, be guaranteed by randomization. Often the strict independence of assignment and potential outcomes only holds when conditioning on another variable W that influenced the randomization process. The independence assumption then changes to an assumption of conditional independence $(Y^0, Y^1) \perp\!\!\!\perp D|W$. As long W, which can also be a set of covariates, is observed, we can use appropriate strategies such as sub-classification or matching to get an unbiased estimate of δ_i given the conditional independence assumption.

A.2 Simulations

For demonstrating the basic concepts of causal inference I use simulations of simple linear models. The exposure is normally distributed as $X \sim N(0, 1)$. For the simplest causal inference path of $X \rightarrow Y$, Y is a linear combination of X and an (in reality unobserved) error term $U_1 \sim N(0, 1)$. Therefore, the true causal effect of X on Y equals 1. More complex simulation models work in the same way, with each variable given by a linear combination of its ancestor variables and a random error term.

The causal effect in each simulation is estimated by a linear regression model. For the simplest model of $X \rightarrow Y$, this means estimating the regression coefficient b_1 of $Y = b_0 + b_1 * X + \epsilon$ via ordinary least square estimation with the R Code `lm(Y ~ X, data)`. For each simulation, the estimated regression coefficient is assumed to be the best unbiased estimate of the causal effect, creating a distribution of estimated causal effects.

A.3 Technical Details

A.3.1 Session Info

```
sessionInfo()
```

```
R version 4.4.0 (2024-04-24 ucrt)
Platform: x86_64-w64-mingw32/x64
Running under: Windows 11 x64 (build 22631)

Matrix products: default


locale:
[1] LC_COLLATE=German_Germany.utf8  LC_CTYPE=German_Germany.utf8
[3] LC_MONETARY=German_Germany.utf8 LC_NUMERIC=C
[5] LC_TIME=German_Germany.utf8

time zone: Europe/Berlin
tzcode source: internal

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] patchwork_1.2.0 ggplot2_3.5.1   ggdag_0.2.12    dagitty_0.3-4

loaded via a namespace (and not attached):
 [1] viridis_0.6.5      utf8_1.2.4         generics_0.1.3     tidyr_1.3.1
 [5] stringi_1.8.3      digest_0.6.35      magrittr_2.0.3     evaluate_0.23
 [9] grid_4.4.0         fastmap_1.1.1      rprojroot_2.0.4    jsonlite_1.8.8
[13] ggrepel_0.9.5      gridExtra_2.3      purrr_1.0.2        fansi_1.0.6
[17] viridisLite_0.4.2  scales_1.3.0       tweenr_2.0.3       cli_3.6.2
[21] rlang_1.1.3        graphlayouts_1.1.1 polyclip_1.10-6    tidygraph_1.3.1
[25] munsell_0.5.1      withr_3.0.0        cachem_1.0.8       yaml_2.3.8
[29] tools_4.4.0        memoise_2.0.1      dplyr_1.1.4        colorspace_2.1-0
[33] here_1.0.1         boot_1.3-30        curl_5.2.1         vctrs_0.6.5
[37] R6_2.5.1           lifecycle_1.0.4    stringr_1.5.1      V8_4.4.2
[41] MASS_7.3-60.2      ggraph_2.2.1       pkgconfig_2.0.3    pillar_1.9.0
[45] gtable_0.3.5       glue_1.7.0         Rcpp_1.0.12        ggforce_0.4.2
[49] xfun_0.43          tibble_3.2.1       tidyselect_1.2.1   knitr_1.46
```

```
[53] farver_2.1.1       htmltools_0.5.8.1  igraph_2.0.3       labeling_0.4.3
[57] rmarkdown_2.26     compiler_4.4.0
```

A.3.2 Packages

```
# p_used <- suppressMessages(unique(renv::dependencies(path = "../")$Package))
# p_inst <- as.data.frame(installed.packages())
# out <- p_inst[p_inst$Package %in% p_used, c("Package", "Version")]
# rownames(out) <- NULL
# out
```