1

Introduction

Causal mediation refers to a process in which one variable influences another through its impact on intermediate variables. This involves a sequence where a first variable affects a second, which then affects a third, and so on, forming a causal chain that links an initiating cause to a terminal outcome. For example, attending college may reduce the risk of unemployment, which leads to higher earnings, which may then improve mental health in adulthood. Participating in a job training workshop can enhance practical skills and boost self-confidence, helping individuals to secure new employment. Consuming media that highlights the economic costs of immigration can trigger fear and anxiety, which may then erode support for certain immigration policies.

All these examples illustrate a process whereby an initial cause impacts intermediate variables that, in turn, influence an outcome. The initiating cause is often referred to as an exposure, treatment, or independent variable, while the intermediaries that transmit the effect of this cause to the outcome are known as mediators. The impact of the initial cause on the outcome that operates through its influence on a mediator or set of mediators is commonly termed an indirect effect. Conversely, the impact of the initiating cause on the outcome that bypasses the mediators of interest is called a direct effect (Hayes 2022; MacKinnon 2008; VanderWeele 2015).

Causal mediation analysis encompasses methods that evaluate direct and indirect effects and that delineate the intermediate mechanisms connecting a cause to an outcome. It seeks to uncover not just *whether* but also *why* a putative cause influences an outcome by quantifying the processes through which an effect of interest operates. While many scientific studies aim only to establish the existence, direction, or magnitude of a

https://doi.org/10.1017/9781009222198.001 Published online by Cambridge University Press

causal effect from one variable to another, mediation analysis transcends this narrow focus: it additionally probes the underlying mechanisms in an effort to illuminate exactly how causes produce their effects.

This book provides a comprehensive introduction to causal mediation analysis, tailored for applications in the social sciences. It uses precise notation and clear visual diagrams to guide readers from basic definitions of effects, via minimally sufficient assumptions, to cutting-edge estimation procedures. With an emphasis on both theoretical foundations and practical implementation, the text blends technical material with many empirical illustrations, all implemented using widely available data and software. Our goal is to furnish readers with the knowledge and skills needed to conduct a defensible analysis of causal mediation across a wide range of scenarios. This includes the ability to define, identify, and estimate a variety of different direct and indirect effects, interpret these effects accurately, and critically assess the assumptions required for valid inference. Throughout the book, we highlight both the promise and the pitfalls of causal mediation analysis, carefully outlining what can – and cannot – be inferred from empirical data, under what assumptions, and with what level of certainty.

Although the concept of mediation is straightforward – intermediate variables transmit the effect of an initial cause to an outcome – investigating this process empirically can be rather complex. This book serves as an introductory guide, but it does presume some knowledge of graduate-level statistics and research methods, as well as familiarity with regression and general linear models (Fox 2015; Wooldridge 2020). Prior exposure to the foundations of causal inference, including the counterfactual framework and graphical causal models (Angrist and Pischke 2009; Hernan and Robins 2020; Morgan and Winship 2014), would be helpful but not essential. We introduce these topics early on and revisit them throughout the book. To make the material as accessible as possible, technical definitions of important concepts, models, and assumptions are paired with intuitive explanations and practical examples wherever possible.

1.1 MEDIATION IN THE SOCIAL SCIENCES

Causal mediation analysis is central to research in the social sciences, including sociology, psychology, political science, public health, criminology, economics, and public policy evaluation. Across these disciplines, examples of mediation abound: environmental stimuli prompt learning that shapes behavior; family background influences educational

https://doi.org/10.1017/9781009222198.001 Published online by Cambridge University Press

attainment, which subsequently affects earnings in adulthood; marriage facilitates greater investment in children, which improves their performance at school; microbial exposures result in infections that lead to death; tobacco taxes decrease smoking, which then reduces cancer incidence; and so forth. This section introduces several areas of social scientific inquiry involving causal mediation, topics that we will revisit and elaborate on throughout the book.

1.1.1 The Effects of Education on Mental Health

The mental health of a population serves as a "social mirror," reflecting the cumulative human consequences of different environmental contexts, economic conditions, and personal histories (Avison et al. 2007). On an individual level, mental health problems can disrupt work and relationships, thereby exacerbating isolation and further reducing life satisfaction. These challenges also impose significant economic costs, which are evidenced by lower productivity, higher healthcare spending, and strained public resources.

In the fields of sociology, economics, and public health, education is widely thought to improve mental health (Heckman et al. 2018; Hout 2012; Mirowsky 2003). For example, numerous studies document a negative correlation between higher education and depression, with strong indications that this relationship is causal (Adams et al. 2003; Warren 2009; Yan and Williams 1999). A more advanced education therefore appears to reduce depression, but important questions remain. In particular, through what mechanisms does education exert its beneficial effects on mental health? What processes explain the lower levels of depression that follow from higher levels of educational attainment?

Education may improve mental health through a number of different pathways. For example, higher education might prevent depression by insulating individuals from the economic and psychological stressors associated with unemployment. Similarly, adults with higher levels of education often experience greater demand for their skills in the labor market, leading to higher-paying jobs compared to their counterparts with lower levels of education (Card 1999; Hout 2012). The greater incomes that accrue to the highly educated may subsequently improve mental health by helping to fulfill their basic needs and providing better access to healthcare, while also boosting self-esteem and enhancing personal agency (Mirowsky and Ross 1990; Silverstone and Salsali 2003). The influence of education on mental health might even extend beyond its

https://doi.org/10.1017/9781009222198.001 Published online by Cambridge University Press

impact on employment and earnings. An advanced education could also reduce depression by providing access to health-related information, promoting healthier lifestyles, or enlarging support networks (Lee 2011; Mirowsky 2003).

Does employment stability transmit the effect of education on depression? How important is the explanatory role of higher income? Can we differentiate the impact of these economic processes from the influence of other factors, such as health information, lifestyle choices, and social support? These questions, which probe the mechanisms through which one variable affects another, are the realm of causal mediation analysis.

1.1.2 Media Discourse, Public Opinion, and Immigration

Throughout history, attitudes toward immigration have significantly influenced politics, governance, and inter-group conflict (Glazer and Moynihan 1970; Omi and Winant 2014; Portes and Rumbaut 2006). Serving often as a catalyst for political polarization, immigration has consistently shaped public opinion, national policy, and social cohesion during periods of heavy population movement. The recent intensification of these debates in North America and Western Europe, spurred by globalization and its concomitant demographic shifts, underscores the persistent contentiousness surrounding immigration.

In the fields of political science, sociology, and psychology, understanding the determinants of attitudes toward immigration is a central concern. These attitudes are highly volatile, with opposition to immigration waxing and waning from one decade to the next (Simon and Lynch 1999). Media discourse and messaging from political elites are widely recognized as key influences on public sentiments. In particular, news stories and political communications that emphasize the perceived negative impacts of immigration on jobs, crime, taxes, public services, or social cohesion tend to amplify public opposition (Brader et al. 2008).

But what psychological processes account for these effects? One perspective contends that media highlighting the harms of immigration may simply change individual beliefs about the severity of the issue (Scheve and Slaughter 2001; Sniderman et al. 2000). According to this view, exposure to negative media framing alters perceptions of the costs and benefits associated with immigration, thereby intensifying opposition. Alternatively, exposure to media emphasizing the harms of immigration could provoke fear, anxiety, or distrust, and this emotional response, particularly potent when the immigrants in question are members of a denigrated

https://doi.org/10.1017/9781009222198.001 Published online by Cambridge University Press

racial group, might fuel opposition to immigration (Brader et al. 2008; Marcus et al. 2000, 2005).

Does exposure to media that highlights the costs of immigration heighten opposition by triggering fear and anxiety? Or does the influence of negative framing stem from changes in beliefs about the costs and benefits of immigration? Disentangling the role of beliefs versus emotions in transmitting the effects of media framing on public opinion is yet another task for causal mediation analysis.

1.1.3 Workforce Development Programs and Employment

Unemployment is a critical social problem with far-reaching impacts on both individuals and communities. High unemployment rates are associated with increased poverty and greater mortality at the population level (Brady 2009; Roelfs et al. 2011). The personal toll of unemployment is also severe, as it can lead to psychological distress, diminished self-esteem, and frayed social relationships (Jahoda 1982; Paul and Moser 2009). Addressing unemployment and facilitating a return to work are therefore crucial not only for the economic vitality of society but also for the personal well-being of its members.

Reemployment programs can help mitigate the negative impacts of job loss, enhance productivity, and foster a sense of purpose and belonging among individuals who are out of work. Research in economics, public policy, and program evaluation has explored a variety of workforce development interventions designed to assist unemployed workers with reentering the labor force (Barnow 1987; Bloom et al. 1997; Holzer 2013). One strategy is to enroll these individuals in community-based workshops that provide training in a variety of skills, such as job search techniques, resume construction, interviewing demeanor, and problem-solving. Additionally, these workshops may offer programming to help boost self-esteem and self-efficacy. Participation has been shown to facilitate reemployment and increase earnings, though the effects can sometimes be modest (Vinokur et al. 1995; Vinokur and Schul 1997).

How do these interventions succeed in increasing employment? Do they equip participants with tangible skills that employers value and help them to find new work, or do they primarily instill confidence, which sustains individuals through a prolonged job search and fosters resilience against setbacks? Could the workshops be modified to enhance their impact on employment by concentrating more on psychological well-being or practical skills? Causal mediation analysis can reveal how

https://doi.org/10.1017/9781009222198.001 Published online by Cambridge University Press

employment programs achieve their impacts and inform the development of more effective interventions in the future.

1.2 THE PROMISE OF CAUSAL MEDIATION ANALYSIS

Causal mediation analysis is central to research in the social sciences for several reasons (Glynn 2021; MacKinnon et al. 2007; VanderWeele 2015). First, it enriches our understanding of social phenomena. For example, discerning why participation in a job training workshop leads to increased employment and earnings – whether through improved self-efficacy, the acquisition of tangible skills, or both – deepens our understanding of how the social world operates. Such knowledge not only satisfies a fundamental human curiosity but also supports informed decision-making and lays a foundation for future learning and discovery.

Second, causal mediation analysis is essential for evaluating and refining theory. Many theories in the social sciences posit causal chains linking different variables or events. Sometimes, competing models propose different mechanisms for the same effect, from an initial cause to a terminal outcome. By analyzing these hypothesized mechanisms, we can test, refine, and adjudicate between theories more rigorously. This enhances the reliability and applicability of our theoretical models, which can then support more accurate predictions and better decision-making outside the social science laboratory.

Third, causal mediation analysis can help improve interventions and public policies. For example, if evidence suggests that a job training workshop succeeds at increasing employment primarily by boosting participant self-efficacy, future iterations of the program could focus more on confidence-building exercises to potentially enhance its effectiveness. Conversely, mediation analysis might reveal components of the program that are ineffective or even detrimental, such as activities that inadvertently heighten anxiety or despair, which may then inhibit the search for a new job. These components could be discarded or marginalized in an effort to maximize the program's beneficial impact. In general, knowledge of the causal mechanisms through which an intervention affects an outcome can aid in optimizing its impact and in diagnosing failures to achieve intended results.

Relatedly, mediation analysis can also inform the development of new policy interventions by identifying more feasible targets for change. While addressing the root cause of a problem is ideal, it is often impractical. Consider the example of education and mental health. If research shows that

https://doi.org/10.1017/9781009222198.001 Published online by Cambridge University Press

education primarily improves mental health by increasing income, then directly boosting income through targeted financial assistance to workers or families might be a more practical solution for realizing the same benefits compared to encouraging further education, which can be logistically complicated. Mediation analysis can point toward more promising and actionable points of intervention when directly addressing the initial cause proves impractical or infeasible.

In addition, causal mediation analysis can contribute to understanding how the effects of interventions might vary across different populations, time periods, or settings. By identifying specific mechanisms that drive causal effects, researchers can better predict where and for whom these interventions are likely to be more versus less effective. This information can be useful when tailoring interventions to specific target populations or transporting them from one setting to another.

Finally, mediation analysis can bolster the credibility of inferences about the causal effect of one variable on another. In the social sciences, where conducting randomized experiments can be logistically challenging or unethical, researchers often depend on a variety of different study designs to investigate causation. By triangulating these different types of evidence, we can develop a more robust understanding of causal relationships between variables. Mediation analysis contributes to this pluralistic approach by tracing the entire process – from the initial cause, through intermediate variables, to the final outcome – thereby strengthening the support for causal inferences about any one link in the chain.

Despite the promise of causal mediation analysis for deepening understanding, evaluating theory, refining interventions, and bolstering the credibility of causal inferences, tracing causal processes remains challenging in practice. In most applications, the assumptions required to learn about causal mediation from empirical data are considerably stronger than those needed to infer relationships between only an initial cause and a terminal outcome (VanderWeele 2015). Furthermore, designing experiments to study mediation poses additional challenges. While randomized experiments are effective and widely used for assessing the impacts of one variable on another, their capacity to reveal how these effects are transmitted through intermediate factors is comparatively limited, as experiments specifically aimed at uncovering mediation are more complex and difficult to implement.

Nevertheless, understanding mediation is crucial for social science inquiry. Moreover, reasoning and learning about causal chains are central

https://doi.org/10.1017/9781009222198.001 Published online by Cambridge University Press

components of human intelligence in general. From as early as age three, individuals begin to interpret the world around them in terms of causal processes (MacKinnon 2008; Shultz 1982). Successfully navigating our social and natural environments is almost unimaginable without the ability to learn and apply knowledge of mediation – forecasting and anticipating the sequences of events triggered by an initial cause, its intermediate consequences, and ultimate outcomes. Although the scientific investigation of mediation is fraught with challenges and uncertainties, it remains indispensable to human understanding, decision-making, and action, not only in the social sciences but also in everyday life.

1.3 A BRIEF HISTORY

Over the past several decades, a "causal revolution" has transformed research in statistics and the social sciences, driven by seminal works from Donald Rubin (1974, 1983), James Robins (1986), and Judea Pearl (1995, 2009). Rubin sparked the revolution by introducing the counterfactual framework for conceptualizing causation and explicitly distinguishing it from statistical association. Robins then extended this framework to a wider range of applications, such as those involving longitudinal data.

The counterfactual framework established a rigorous notation for defining causal relationships in terms of "potential outcomes," which envision different, hypothetical states of the world to determine what would have occurred if a variable or set of variables had differed from their actual state. With this approach, if an outcome would have differed had some exposure been other than it was, that exposure is considered a cause of the outcome (VanderWeele 2015). The introduction of potential outcomes recast causal inference as a missing data problem. Since we can never observe the outcome that would have occurred had an exposure differed from its factual state, we are always missing a key piece of information needed to infer causality with certainty. Thus, the challenge of causal inference lies in trying to learn about outcomes that cannot be observed, using only the data that is observable.

Following the introduction of the counterfactual framework, Pearl accelerated the causal revolution by developing a graphical framework paired with a "causal calculus" for analyzing causation. This approach allows researchers to visually represent causal relationships, deduce potential outcomes from a broader causal system, and derive testable implications for observable data. These innovations also help to identify the specific conditions under which we can learn about counterfactual

https://doi.org/10.1017/9781009222198.001 Published online by Cambridge University Press

outcomes from the empirical data available, along with the necessary calculations to achieve this learning. Together, graphical causal models and Pearl's calculus have enhanced our ability to precisely determine when and how statistical associations in observed data imply causation – or when they do not.

Before the causal revolution, efforts to quantify mediation started with Wright's (1921, 1934) development of linear path analysis. This approach uses simple diagrams and linear models to estimate and visualize direct and indirect relationships among a set of variables. Linear path analysis gained traction in the social sciences during the 1960s and 1970s, where it was further developed by Duncan (1966), Goldberger (1972), and Blalock (1971). Later, Alwin and Hauser (1975) systematized procedures for decomposing a total effect into direct and indirect components using linear models, while Sobel (1982) developed procedures for drawing statistical inferences from these types of decompositions. In a highly influential article, Baron and Kenny (1986) introduced a closely related methodology, which involves estimating a series of linear models and applying statistical tests to certain coefficients in order to evaluate mediation. More recently, texts by Hayes (2022) and MacKinnon (2008) have widely disseminated these approaches across the social sciences, focusing largely on methods that stem from the tradition of analyzing mediation with linear models.

All of these approaches, from Wright's (1934) early work to Duncan (1966), and from Baron and Kenny (1986) to contemporary texts like Hayes (2022) and MacKinnon (2008), agree that mediation is fundamentally a causal concept. For example, Baron and Kenny (1986, p. 1173) define mediation as "the generative mechanism through which the focal independent variable is able to influence the dependent variable of interest." Similarly, Alwin and Hauser (1975, p. 39) describe indirect effects as "those parts of a variable's total effect which are transmitted … by variables specified as intervening between the cause and effect of interest." Likewise, MacKinnon (2008, p. 8) explains that "in a mediation model, the independent variable causes the mediator which then causes the dependent variable," while Hayes (2022, p. 7) describes mediation as occurring when "variation in [an exposure] X causes variation in one or more mediators M, which in turn causes variation in [an outcome] Y."

Despite the broad consensus that mediation inherently involves causation, most earlier works did not adopt a formal notation or conceptual framework that clearly distinguished between causation and mere association. Prior approaches to mediation analysis in the social sciences have

https://doi.org/10.1017/9781009222198.001 Published online by Cambridge University Press

typically defined and analyzed mediation within the confines of statistical models that blur the distinction between causal versus associational relationships, forgoing the innovations of the causal revolution described previously.

In recent years, however, methodological research on mediation analysis has fully embraced the counterfactual framework and graphical causal models (e.g., Pearl 2001; Robins and Greenland 1992; VanderWeele 2011b), leading to significant breakthroughs, important new insights, and a surge of novel methods. Unfortunately, much of this innovative material remains buried within specialist journals, often presented in highly technical language that may be inaccessible to applied researchers who could greatly benefit from these advances. Even in texts aimed at a more general audience that draw heavily on these recent advances, such as VanderWeele's (2015) outstanding treatise on "Explanation in Causal Inference," the authors sometimes still choose to forgo the modern tools of causal analysis, opting instead to "describe as many of the methods and assumptions as possible without requiring specific appeal to the notation of counterfactual-based logic" (p. xi).

In contrast, we consider the counterfactual framework and causal graphs as indispensable tools for mediation analysis. Without an explicit notation and causal calculus, our view is that any approach to analyzing mediation remains incomplete and is liable to generate confusion regarding the objectives of the analysis and the conditions needed to achieve them. Embracing the modern tools of causal analysis broadens the range of theoretical questions about mediation that can be precisely articulated, clarifies the role of empirical data versus unverifiable assumptions in addressing these questions, and unlocks new tools for extracting answers to them.

1.4 OUR DISTINCTIVE APPROACH

Building on the innovations that propelled the causal revolution, our approach to mediation analysis comprises four basic steps:

1. **Define:** clearly articulate a causal estimand using potential outcomes, which should faithfully represent the substantive aims of the research and require no commitment to any particular statistical model.
2. **Identify:** connect the causal estimand, which involves counterfactuals that cannot be observed, to an empirical quantity that can

https://doi.org/10.1017/9781009222198.001 Published online by Cambridge University Press

be learned from observable data under a set of precisely stated assumptions.
3. **Estimate:** collect these data and use them to estimate the empirical quantity of interest, assessing any uncertainty that afflicts the estimation.
4. **Scrutinize:** critically evaluate the assumptions required for the data to yield insights about the causal estimand, which is essential for judging the credibility of the evidence provided.

Following the framework proposed by Lundberg et al. (2021), our first step involves defining a causal estimand, or a set of estimands, using potential outcomes and the counterfactual framework. An estimand is simply the object of scientific inquiry – it represents the theoretical quantity we would aim to determine if it were possible to observe all potential outcomes from each of the counterfactual worlds envisioned. This step translates the substantive goals of the research into precise objects for inference that exist independently of any statistical model and clearly distinguish causation from association.

The second step in our approach involves linking the causal estimand, defined in terms of quantities that cannot be observed, to an empirical estimand that consists solely of observable quantities. Essentially, this step connects the causal estimand of interest to a function of empirical data. Causal estimands defined in terms of counterfactuals can only be linked with empirical quantities through assumptions about how the unobservable data relate to the data we can observe. These assumptions must be articulated with precision and transparency, as the empirical data are informative about the causal estimand only when they hold true. Although there are a variety of ways to express these assumptions, causal graphs are particularly effective and accessible, and we rely on them throughout the book.

The third step involves collecting data and selecting an estimation strategy to learn about the empirical quantity of interest. By decoupling the causal estimand and its empirical counterpart from the estimation process, researchers are afforded a range of options for learning from data, each with distinct strengths and weaknesses. Defining the target of scientific inquiry independently of a statistical model allows researchers the freedom to choose the most suitable method for their data, without being tethered to any single approach. Estimation can proceed without relying on any statistical model whatsoever, if the data permit, or it can involve a variety of models that offer different levels of flexibility in approximating

https://doi.org/10.1017/9781009222198.001 Published online by Cambridge University Press

the true, but unknown, distribution of the observed data. This approach enables researchers to conduct mediation analysis with almost any type of data, no matter the complexity of the relationships or the level of measurement among variables.

It also stands in stark contrast to conventional practices in the social sciences, where researchers analyzing mediation often leap straight to a particular statistical model without separately defining and identifying a causal estimand as the ultimate target of inference. Typically, the target of inference is defined solely within the confines of a particular model for the observed data – for example, as a product of coefficients from several linear models (Baron and Kenny 1986; Hayes 2022) or as a difference in rescaled coefficients from a series of logit models (Karlson et al. 2012; MacKinnon 2008). However, this conventional approach is suboptimal for a number of reasons. In bypassing a model-independent estimand to focus narrowly on a single statistical model for the observed data, researchers obscure the true objectives of their analysis and the assumptions needed to achieve them, limit the exploration of alternative models that might offer better performance, restrict their ability to accommodate complex relationships and data structures, and hinder the accumulation of scientific knowledge across studies using different modeling strategies to examine the same phenomena (Lundberg et al. 2021).

The final step in our approach to mediation analysis involves a critical examination of the assumptions needed to link the causal estimand to its empirical counterpart and the estimation strategy. Because it is often impossible or impractical to ensure all necessary assumptions are met by design, it is crucial to assess the sensitivity of causal inferences to potential violations of the assumptions on which they are based. In practice, this involves strategically altering the assumptions and observing the impact of these changes on the results, evaluating whether key conclusions remain valid when the assumptions are modified. If minor changes to the assumptions result in significant alterations in the results, the original conclusions may not be very robust. Conversely, if the conclusions hold despite these modifications, researchers can have greater confidence in the reliability of their findings.

To summarize, our approach to causal mediation analysis unfolds in four key steps: define, identify, estimate, and scrutinize. It begins by precisely stating the causal estimand – the central quantity of interest – completely independent of any statistical model. This is followed by linking the causal estimand, which cannot be observed directly, to an empirical quantity based on a set of clearly stated assumptions. The next step

https://doi.org/10.1017/9781009222198.001 Published online by Cambridge University Press

involves estimating this empirical quantity from data and quantifying the uncertainty that afflicts the estimates. It concludes with a critical evaluation of assumptions to determine if the causal inferences are robust to their potential violation. This approach contrasts with conventional approaches to mediation analysis in the social sciences, which often conflate associational and causal notation, rely solely on model-based definitions of effects, obscure the assumptions required to learn about them from empirical data, and consequently, neglect to adequately scrutinize these assumptions.

The approach we adopt offers several advantages for studying mediation. It sharply distinguishes association from causation, ensuring clarity and precision about the objectives of the research. It also focuses on model-independent definitions and identification of target estimands, unlocking the use of a diverse array of competing estimation procedures. This flexibility allows researchers to accommodate many different types of relationships among variables and a variety of different data structures. Moreover, our approach ensures transparency regarding the assumptions underpinning inferences about causal mediation, and it facilitates assessment of how sensitive these inferences are to potential violations of the assumptions on which they are based. Together, these advantages form a solid foundation for generating more credible evidence of causal mediation and for accumulating scientific knowledge about causal processes.

1.5 AN OUTLINE OF THE BOOK

In the ensuing chapters, we provide a comprehensive introduction to causal mediation analysis, with applications to research in the social sciences. We begin by outlining fundamental concepts and relatively simple applications, and then gradually tackle more complex scenarios and the increasingly sophisticated methods they demand. Throughout these chapters, we focus on a running example that examines the impact of college attendance on depression at midlife, and whether this relationship is mediated by unemployment and/or income. This example is revisited in each chapter to ensure continuity and to highlight the subtle differences between methods for analyzing distinct types of mediation. To appeal to an interdisciplinary audience and enrich the narrative, each chapter also concludes with a unique, stand-alone empirical illustration. These additional examples are drawn from a wide range of disciplines across the social sciences.

https://doi.org/10.1017/9781009222198.001 Published online by Cambridge University Press

Chapter 2 lays the foundation for the remainder of the book by introducing concepts and notation that are essential for analyzing causal mediation. We first outline the counterfactual framework, potential outcomes, and graphical causal models, which form the backbone of any causal analysis. Using these tools, we conceptually define the total effect of an exposure on an outcome and explain the "fundamental problem of causal inference" (Holland 1986). We then explain how potential outcomes and causal diagrams can be used to distinguish between a direct effect of the exposure that bypasses a mediator of interest, and an indirect effect that operates specifically through a causal chain involving the mediator. After conceptually defining direct and indirect effects, we introduce the "fundamental problem of causal mediation," a challenge that is related to the fundamental problem of causal inference but is even more difficult to resolve (Imai et al. 2013). This chapter serves as a conceptual primer that illustrates the key ideas, notation, and principles, as well as the challenges, that underlie analyses of causal mediation.

In Chapter 3, we introduce methods for analyzing causal mediation in applications involving a single mediator, where the direct and indirect effects of interest may be confounded by baseline covariates. Baseline confounding occurs when extraneous variables, which are either measured before the exposure or are otherwise unaffected by it, distort the apparent effects of the exposure and/or the mediator, leading to biased inferences. This chapter outlines the conditions necessary for accurately analyzing total, direct, and indirect effects in applications with a single mediator and baseline confounding, and it provides a comprehensive overview of different estimation procedures, including those based on linear models, simulation methods, and inverse probability weighting. To illustrate these methods, we analyze data from a study that assessed the impact of a job training workshop on employment (Vinokur et al. 1995; Vinokur and Schul 1997), focusing specifically on how psychological factors like self-efficacy may mediate these effects.

Chapter 4 addresses a more complex, but still common, scenario in the social sciences: causal mediation analysis focusing on a single mediator in the presence of exposure-induced confounding. This type of confounding occurs when variables affected by the exposure distort the apparent effects of the mediator on the outcome, which can also lead to biased inferences. In this chapter, we introduce an alternative approach for analyzing how a single mediator may account for the effect of an exposure on an outcome when exposure-induced confounders are present. We discuss modified regression, simulation, and weighting procedures as viable

https://doi.org/10.1017/9781009222198.001 Published online by Cambridge University Press

estimation strategies. And to illustrate these methods, we reanalyze cross-national data from a study that examined the economic origins of gender roles (Alesina et al. 2013). Specifically, we analyze whether historical cultivation practices (i.e., the use of plow agriculture) influenced contemporary patterns of female political participation directly, apart from their impact on economic development, all the while accounting for a potentially important exposure-induced confounder – the degree to which a country's current government is authoritarian versus democratic.

Chapter 5 covers analyses of causal mediation with multiple mediators. In this chapter, we outline methods for decomposing the total effect of an exposure on an outcome into an indirect effect that operates through a set of mediators and a direct effect that circumvents them all. We also show how to decompose a total effect into a series of path-specific effects, which capture the unique explanatory role of each mediator in a causal sequence. To estimate these effects, we outline several strategies, including regression-imputation and weighting approaches. We demonstrate these methods by reanalyzing data from an experimental study of public opinion that examined how negative media framing affects opposition to immigration (Brader et al. 2008). Using these data, we attempt to untangle the role of emotions versus beliefs as potential mediating factors that connect discourse in the news to public opinion.

Chapter 6 explores the application of machine learning to enhance causal mediation analysis. Machine learning is a burgeoning field of statistics and computer science that involves training algorithms to inductively identify relationships and patterns in data. When used to analyze causal mediation, this approach combines carefully constructed estimating equations with these data-adaptive algorithms to produce robust estimates of direct and indirect effects. We initiate the discussion by addressing the problem of model misspecification, which often complicates analyses of causal mediation. We then introduce the concept of "multiple robustness" and explain the advantages of machine learning methods for causal inference. Lastly, we present a collection of multiply robust estimation procedures for use in analyses of causal mediation. To demonstrate these methods, we reanalyze data from a study that examined the intergenerational pathways through which exposure to violence shapes descendants' political attitudes (Lupu and Peisakhin 2017). This chapter highlights the challenges of model misspecification, outlines the advantages of multiply robust approaches to estimation, and demonstrates the utility of machine learning for mediation analysis.

https://doi.org/10.1017/9781009222198.001 Published online by Cambridge University Press

In Chapter 7, we explore the design and analysis of experimental and quasi-experimental studies specifically aimed at uncovering causal mediation. The chapter discusses a number of novel experimental designs, which are scarcely applied in the existing literature but hold potential for advancing research on mediation. These include joint, parallel, and multi-arm randomization designs, as well as experiments utilizing randomized encouragements and pre/post designs with repeated measures of the mediator.

Finally, Chapter 8 concludes by reflecting on limitations and omissions. It charts the frontier of the methodological literature on causal mediation and suggests potentially fruitful directions for the future, aiming to inspire ongoing dialogue and investigation in this vital area of social science research.

1.6 STATISTICAL SOFTWARE

The widespread adoption of causal mediation analysis is facilitated by the availability of software for implementation. In support of this, all the data and code used throughout this book, complete with custom packages and detailed documentation for implementing the methods we discuss and illustrate, are accessible at https://causalmedanalysis.github.io/. We provide code in both R and Stata, which are among the most popular statistical software packages in the social sciences. These resources are designed to equip readers with the tools necessary to replicate and explore the examples presented in this book. Additionally, they provide the means for readers to adapt and apply all the methods we cover to their own research. Moving forward, we are committed to continually updating and enhancing the capabilities and functionality of the code, ensuring its ongoing relevance and utility.

https://doi.org/10.1017/9781009222198.001 Published online by Cambridge University Press