



Symposium

Explanation in causal inference: developments in mediation and interaction

Tyler J VanderWeele

Departments of Epidemiology and Biostatistics, Harvard T. H. Chan School of Public Health

*Corresponding author. 677 Huntington Avenue, Boston MA 02115. E-mail: tvanderw@hsph.harvard.edu

Accepted 20 July 2016

Epidemiology is sometimes described as the study of the distribution and determinants of disease. Tremendous progress has been made in our understanding of, and in developing methods to study, such distributions and determinants. The questions of distribution and of determinants perhaps find their intersection in the more focused query, ‘For whom?’ For whom, or for what sub-populations, is each particular determinant relevant? We are led, in other words, to questions of interaction or heterogeneity of effect. And then, if we probe the study of the determinants of disease further, we are led to the question of ‘Why?’ Why is this particular determinant a cause of disease? What are the mechanisms or pathways that result in an effect? In other words, we are led to questions of mediation.

Whereas questions of interaction have been present within epidemiology for decades and, in the historical literature, one can find numerous methodological papers addressing such questions, the story within epidemiology is rather different for mediation. Questions of mechanisms have of course often been of interest, but formal methods for mediation have until recently not often been widely discussed or taught within epidemiology. Occasionally tools from the social sciences have been employed, but formal training on methods for mediation has until recently been scant. Methodology for both mediation and interaction have, however, expanded dramatically in the past decade and have been the focus of much of my own research. It is the various developments in methodology on mediation and interaction relevant to epidemiology, and to the social and biomedical science more generally, that I have attempted to describe in my book *Explanation in Causal Inference: Methods for Mediation and Interaction*,¹ published

last year, in 2015, by Oxford University Press. The book attempts to move beyond simple inference for cause and effect to questions of ‘Why?’ and ‘For whom?’, that is to say, to questions of mediation and interaction.

Content

The book is framed by more general conceptual material in Chapters 1 and 16. Chapter 1 considers the nature of explanation, and its relation to causation itself, in broader conceptual terms. Different forms of explanation when reasoning about causation are discussed, along with how the phenomena of mediation and interaction provide different types of explanations for cause-effect relationships. Discussion is also provided as to what might motivate a researcher to investigate these phenomena of mediation and interaction empirically. The first chapter concludes with a brief description of the structure of the remainder book. The final chapter, Chapter 16, after a brief assessment of what future methodological research development might look like, once again turns to broader conceptual questions of explanation and causation and situates some of the discussion in the book within the philosophical discourse on these issues. It is noted that the potential outcomes perspective that forms the basis of the methodologies described in the book, in fact constitutes within the philosophical literature a subset of broader counterfactual queries, which itself is a subset of questions concerning the nature of causality. Discussion is offered as to the relation between causation and explanation, natural laws and human agency, touching upon philosophical ideas concerning causation from Hume, Aristotle, Lewis and Aquinas.^{2–7}

The focus of the book is of course on empirical methodologies rather than philosophy, and the methodological content (Chapters 2–15) is divided into three main parts: I. Mediation (Chapters 2–8); II. Interaction (Chapters 9–13); and III. Synthesis and spillover effects (Chapters 14 and 15). Chapter 2 provides a comprehensive overview of the recent methods in the causal inference literature for assessing mediation. A number of straightforward regression-based methods are described, and discussion is also given to how these relate to methods that have been used for many decades within epidemiology and in the social sciences, such as the Baron and Kenny approach.⁸ Essentially, the new methods in the causal inference literature allow for much greater model flexibility and allow also for effect decomposition of a total effect into a direct and indirect effect even in the presence of interaction between the exposure and the mediator or other non-linearities. When such interaction is absent, the new methods in many cases effectively collapse to the more traditional approaches. The new methods constitute a generalization of the traditional approaches; the conditions under which the traditional approaches fail and when they are appropriate is discussed. Throughout, emphasis is placed on the assumptions needed to interpret estimates from statistical models causally as direct and indirect effects. This second chapter of the book, which forms the book's core mediation chapter, aims both to provide the conceptual foundation for modern approaches to mediation and also to equip the reader with a set of practical tools. The chapter includes descriptions of software available in SAS, Stata, SPSS and R to carry out the types of analyses described. The chapter could be read on its own as an introduction to modern methods for mediation analysis.

Chapters 3–7 address a number of further topics in mediation analysis. Chapter 3 describes sensitivity analysis techniques for assessing how sensitive one's conclusions concerning mediation are to violations in the assumptions being made. Because the assumptions needed to assess mediation and pathways are quite strong, it is argued in the book that every mediation analysis ought to be accompanied by a sensitivity analysis. This is neglected in much of the research employing methods for mediation in the social and biomedical sciences, but analyses can be very misleading if assumptions are violated, and some dramatic examples of this are presented. If all mediation analyses were accompanied by sensitivity analysis, our inferences about pathways and mechanisms would be on much firmer footing. Several practical approaches to carrying out sensitivity analysis are provided. The remaining chapters on mediation cover more specialized topics. Chapter 4 extends the methods for mediation to time-to-event outcomes. Chapter 5 extends this further to settings in which multiple medi-

ators are of interest. Chapter 6 extends the methodology to longitudinal data with time-varying exposures and mediators. Chapter 7 addresses a number of more specialized topics. Finally, Chapter 8 does not consider mediation *per se*, but rather addresses a number of different concepts and methods that are distinct from, but related to and sometimes confused with, mediation. These include instrumental variables, Mendelian randomization, surrogate outcomes and principal stratification approaches. Each of these topics is given a short description and relations with mediation are discussed. Chapters 2–8, taken together, provide a comprehensive overview of the current state of methodology for mediation, at least as of the book's publication in early 2015. Methods have, of course, continued to develop since, but a reader of these chapters of the book will be well equipped to assimilate and make use of the methodology that has developed in the past year since the book's publication and likely that which will take place in the years to come.

Part II of the book, Chapters 9–13, shifts the focus to the analysis of interaction. Chapter 9 constitutes a reasonably comprehensive overview of conceptual, historical, interpretative and statistical aspects of interaction. The chapter could, like Chapter 2, be read on its own as a general tutorial on interaction analysis.⁹ Chapters 10–13 focus on more specific topics. Chapter 10 discusses so-called sufficient cause interaction, *i.e.* settings in which an outcome occurs if both of two exposures are present but not if only one or the other is present. Such sufficient cause interaction differs from interaction in a statistical model, and the chapter presents modern developments on empirically assessing such sufficient cause interaction, its relation to mechanisms and the limits of such inference. Chapter 11 discusses sensitivity analysis methods relevant for interaction, which are useful in the face of biases such as unmeasured confounding and measurement error. Somewhat surprisingly, in a number of contexts, even if such biases affect estimates of main effects, often interaction estimates are left unbiased. The conditions for this, and how to correct for biases if they are present, are discussed. Chapter 12 discusses selected issues of interaction relevant primarily within, but not exclusively restricted to, a genetic context. Topics include multiple testing, joint tests for main effects and interaction, and case-only estimators of interaction. Chapter 13 provides an overview of sample size and power calculations for interaction analysis, including both additive and multiplicative forms of interaction, from cohort, case-control and case-only designs. In this second part of the book, especially in Chapter 9, some of the material has been well established for decades but many newer methods and approaches are also presented throughout.

Part III of the book offers some synthesis of the ideas of mediation and interaction (in Chapter 14) and also provides an introduction to the notion of spillover effects (in Chapter 15). More specifically, Chapter 14 considers a conceptual and empirical approach to assesses the relative contributions of mediation and interaction simultaneously. It is shown that an overall effect of an exposure on an outcome, in the presence of a mediator with which the exposure may interact, can be decomposed into four components of the portion of the effect due to: (i) neither mediation nor interaction; (ii) just interaction (but not mediation); (iii) just mediation (but not interaction); and (iv) both mediation and interaction. This four-way decomposition, published very recently in the literature,¹⁰ effectively unites methods that attribute effects to interactions and methods that assess mediation. The approach can be implemented using standard statistical models, and software is provided to estimate each of the four components. The four-way decomposition provides the greatest insight into how much of an effect is mediated, how much is due to interaction, how much is due to both mediation and interaction together and how much is due to neither.

Chapter 15 provides an introduction to the topic of spillover effects. Like mediation and interaction, this topic is closely related to explanation and mechanism but is discussed somewhat less commonly within epidemiology, with its being a common topic at present perhaps only within infectious disease epidemiology. This topic of spillover effects, also sometimes referred to as ‘interference’, is the phenomenon whereby the exposure of one individual can affect the outcome of another. Such spillover is common whenever an outcome depends upon social interactions between individuals. The chapter provides an introduction to methods for assessing spillover effects from randomized trials and observational studies, and concludes with a brief introduction to causal inference with social network data wherein an entire network of individuals may be related to and influence each other. Many of the methods and approaches for mediation and interaction have direct analogues in the spillover effect context, and these connections are pointed out and discussed. This chapter might be of greatest interest to infectious disease epidemiologists, but the methodology is also likely relevant to researchers in social epidemiology and perhaps well beyond that also. Recent evidence has indicated that phenomena as diverse as smoking cessation and even obesity may in fact travel through social networks.^{11–13} Thinking clearly about such spillover is important in the design and the assessment of the full impact of potential interventions and public health programmes.¹⁴

The book, taken as a whole, moves the reader through basic ideas and foundational concepts to very recent devel-

opments. Although some of the methods presented in the latter parts of the book are relatively advanced, the vast majority of the book was written to be accessible to anyone familiar with linear and logistic regression. Effort was made to relegate as many of the technical details as possible to an appendix at the book’s end. The appendix itself contains more results, definitions and proofs, which may be of interest to statisticians and methodologists, but none of this is essential to the book’s primary content. That primary content is meant to provide a clear, accurate, and relatively accessible introduction to modern methods to assess mediation and interaction so as to empirically address questions of explanation in causal inference.

Motivation

A topic that is addressed in Chapter 1 and recurs throughout the book concerns the motivations for studying questions of mediation and interaction. A number of examples are given to illustrate the various methodologies. The examples are drawn from perinatal, genetic, social, psychiatric and infectious disease epidemiology, as well as various other disciplines. In some of these the applications, the relevant methodology gives important insight into scientific or policy-relevant questions. In other cases, the examples serve more as simple illustrations of the methodology. As noted in the book’s preface, some of the methodology presented is quite new and there were limited options from which to find an ideal example. Nevertheless, emphasis is placed throughout on trying to clarify the motivations for studying mediation and interaction.

Methods for mediation can be useful for a number of purposes. Some potential uses of these ideas and methods include trying to: understand aetiology; provide evidence to confirm and refute theory; assess the impact of intervening on a mediator when it is not possible to alter an exposure; understand why an intervention succeeded or failed. The different motivations are illustrated by various examples.

One prominent example which appears several times in the book involves genetic variants on chromosome 15 that were found to be associated both with smoking behaviour and with lung cancer. A question which arose in this context was whether the variants affected lung cancer only because they affected smoking, and smoking itself causes lung cancer, or whether the variants affected lung cancer through pathways other than, say, number of cigarettes smoked per day.¹⁵ This question is discussed in Chapter 2 of the book, and is revisited in Chapter 14, to clarify the respective roles of mediation and interaction in understanding these relationships. Another example, addressed in

Chapter 5, concerns the repeated finding that low socioeconomic status (SES) during childhood is associated with adverse health outcomes later in life. Debate remains as to whether this is because low SES during childhood affects adult SES which in turn affects adult health (a 'social trajectory' model), or whether childhood SES affects adult health through pathways other than through adult SES (a 'latent effects/sensitive period' model) or both, and to what extent. Empirical methods for mediation are used to evaluate some evidence for these different theories.¹⁶

Methods for mediation can likewise be of policy relevance in the design and refinement of interventions, even in the context of randomized trials. An example discussed in Chapter 3 concerns a randomized trial of a cognitive therapy intervention¹⁷ that was found to have a beneficial effect on depressive symptoms. However, it was also noted that the intervention had an effect on the use of antidepressants: those in the cognitive behavioural therapy arm were more likely to use antidepressants during follow-up. This led to questions concerning whether the cognitive behavioural therapy intervention had a beneficial effect on depressive symptoms simply because it led to greater antidepressant use, or whether the intervention affected depressive symptoms through other pathways, e.g. by changing the thought and behavioural patterns of the participants. If the intervention were only beneficial because of higher use of antidepressants, then the cognitive-behavioural aspects of the intervention could perhaps be abandoned without much loss and a more cost-effective intervention, focusing only on antidepressant adherence, could be developed. Alternatively, it may have been the case that the intervention was effective both because of increased antidepressant use and because of cognitive-behavioural changes. Methods for assessing mediation can again be useful in assessing the relative contribution of these various pathways. However, as described in the book's analysis of this trial, there are also a number of potential problematic pitfalls due the assumptions required to assess mediation in this example, and here sensitivity analysis techniques turn out to be critical in accurately evaluating the evidence.

There are also a number of practical and theoretical considerations, discussed in the book, that motivate the study of interaction. One of the most prominent of these is that, in a number of settings, resources to implement interventions may be limited. It may not be possible to intervene on or treat an entire population. Resources may only be sufficient to treat a small fraction. If this is the case, then it may be important to identify the subgroups of persons for which the intervention is likely to have the largest effect. Chapter 9 discusses how methods for assessing additive interaction can help determine which subgroups would

benefit most from treatment. Even in settings in which resources are not limited and in which it is possible to intervene on everyone, it may be that a particular intervention is beneficial for some individuals and harmful for others. In such cases, it is critical to identify those groups for whom treatment may be harmful and refrain from treating such persons. Techniques for assessing so-called 'qualitative' or 'cross-over' interaction are discussed in Chapter 9 and are useful in this regard. Other more sophisticated methods based on a large number of covariates can also help identify groups of individuals who would or would not benefit, or who would benefit to the greatest extent, from treatment.

Another reason sometimes given for empirically assessing interaction is that it may give insight into the mechanisms for an outcome. Chapter 10 of the book is devoted to how the empirical study of interactions can give insight into the mechanisms for the outcome, and to the limits of such reasoning. Yet another reason sometimes given for studying interaction is that leveraging a potential interaction may in fact help increase power in testing for the overall effect of an exposure on an outcome. In some settings, by jointly testing for a main effect and for an interaction simultaneously, it is possible to detect an overall effect when a test ignoring the interaction would otherwise not be able to detect the effect. These topics are covered in Chapter 12 of the book.

As noted above, one of the motivations for studying interaction is to identify which subgroups would benefit most from an intervention when resources are limited. However, in some settings it may not be possible to intervene directly on the primary exposure of interest, and one might instead be interested in which other covariates could be intervened upon to eliminate much or most of the effect of the primary exposure of interest. In these cases, methods for attributing effects to interactions, discussed in Chapter 9, can be useful in assessing this and identifying the most relevant covariates for intervention. For example, as discussed also in Chapter 9 of the book in the context of the genetic epidemiology example of lung cancer described above, although it is not possible to intervene directly on the genetic variants themselves, if we were able to eliminate smoking this would in fact also eliminate almost all of the effect of the variants. Methods for attributing effects to interaction can help establish such results.

As with mediation, so also with interaction: the motivations for studying the phenomenon are numerous and diverse, and it is important to clarify the motivation and the scientific or policy question of interest, as different motivations in fact correspond to different types of effects and different appropriate methodologies. Interaction can be assessed on different scales; with mediation, different types

of direct and mediated effects can be defined. The different measures correspond to different types of questions concerning policy and science, and it is thus essential to make sure that the motivation, the effect of interest and the relevant methodology are clear.

Concluding remarks

Taken as a whole, the book provides a comprehensive up-to-date treatment of contemporary methods to empirically assess the phenomena of mediation and interaction, from introductory material to very recent developments. Attention is given to both conceptual foundations and statistical tools. Software is provided; methodology is illustrated by example; motivations are discussed; future directions are noted; and although the book focuses on methodology, the content is also placed within a broader philosophical context.

The main theses of the book, perhaps more implicit throughout rather than explicitly made central, might be summarized as follows.

- i. We know far more about appropriate empirical methodologies to assess mediation and interaction than we did a decade ago. Methodology has expanded dramatically in the past 10 years and can be used to assess these phenomena in a variety of settings that were not previously possible.
- ii. With mediation especially, but with interaction to a certain extent also, empirical inferences are difficult and subject to numerous pitfalls, even more so than is the case for examining simple cause-and-effect relationships. When we examine these questions of mechanisms, we are making more nuanced inferences requiring additional assumptions, and it is thus essential that we investigate the sensitivity of our conclusions to the assumptions being made; sensitivity analysis can assist with this.
- iii. There are numerous motivations for studying the phenomena of mediation and interaction, and the methods can be used in a variety of contexts. However, clarifying why the relevant methodology is being employed, what the substantive question is and how it maps onto the specific effect of interest is essential, since the methodology can also be employed to generate estimates that will likely have little scientific or policy relevance.

Further methodological developments on these topics are likely to take place in the years ahead, but much has been accomplished already and some of the foundational approaches are perhaps unlikely to change. The methodological toolkit has seen dramatic expansion in these areas; many of the methods are new. What remains to be seen perhaps is the extent to which they will be useful. This will depend in part on adoption and dissemination, in part on correct use, in part on

careful mapping of questions to relevant effects and methodologies, in part on assessing sensitivity to assumption violations and, perhaps more fundamentally, on the extent to which the epidemiological questions of greatest significance truly require assessment of mediation and interaction to address. Careful delineation of when various mechanistic considerations are, or are not, relevant will be an important undertaking in the years to come. It is my hope that the book will help in some small way with this task.

Funding

This research was funded by United States NIH grant R01 ES017876.

Conflict of interest: None declared.

References

1. VanderWeele TJ. *Explanation in Causal Inference: Methods for Mediation and Interaction*. New York, NY: Oxford University Press, 2015.
2. Aristotle. (384-322 BCE). *Physics*.
3. Aristotle. (384-322 BCE). *Metaphysics*.
4. Aquinas T. (1225-1274). *Summa Theologica*.
5. Hume D. (1739). *A Treatise of Human Nature*.
6. Hume D. *An Enquiry Concerning Human Understanding*. Reprint. LaSalle, IL: Open Court Press, 1958.
7. Lewis D. *Counterfactuals*. Cambridge, MA: Harvard University Press, 1973.
8. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J Pers Soc Psychol* 1986;51:1173-82.
9. VanderWeele TJ, Knol MJ. A tutorial on interaction. *Epidemiol Methods* 2014;3:33-72.
10. VanderWeele TJ. A unification of mediation and interaction: a four-way decomposition. *Epidemiology* 2014;25:749-61.
11. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. *N Engl J Med* 2007;357:370-79.
12. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. *N Engl J Med* 2008;358:2249-58.
13. VanderWeele TJ. Sensitivity analysis for contagion effects in social networks. *Sociol Methods Res* 2011;40:240-55.
14. VanderWeele TJ, An W. Social networks and causal inference. In: Morgan SL (ed). *Handbook of Causal Analysis for Social Research*. New York, NY: Springer, 2013.
15. VanderWeele TJ, Asomaning K, Tchetgen Tchetgen EJ *et al*. Genetic variants on 15q25.1, smoking and lung cancer: An assessment of mediation and interaction. *Am J Epidemiol* 2012;175:1013-20.
16. Nandi A, Glymour MM, Kawachi I, VanderWeele TJ. Using marginal structural models to estimate the direct effect of adverse childhood social conditions on onset of heart disease, diabetes and stroke. *Epidemiology* 2012;23:223-32.
17. Strong V, Waters R, Hibberd C *et al*. Management of depression for people with cancer (SMaRT oncology 1): A randomised trial. *Lancet* 2008;372:40-48.