Observational Studies 2 (2016) 1-3 Submitted 1/16; Published 1/16

Book Review of *Explanation in Causal Inference: Methods of Mediation and Interaction* (author: T.J. Vanderweele)

Luke J. Keele **ljk20@psu.edu**
Department of Political Science, Penn State University
University Park, PA 16802 USA

Explanation in Causal Inference: Methods of Mediation and Interaction is an introductory text on two widely used methods in statistical analysis: mediation and interaction. The book is both meant to serve as an introduction to these two topics, but also provides considerable mathematical detail in a lengthy appendix. Importantly, the treatment of these two topics is entirely grounded in a counterfactual framework. The counterfactual framework, often referred to as the potential outcomes framework, has been hailed as a revolution in how we think about causality and statistical analysis. I would agree with that sentiment, but the impact of the counterfactual framework is varied. On some topics, the insights have been less revolutionary, but in other areas this framework has I think completely revised how we think. The topics of mediation and interaction analysis are two that I would say have been seriously changed by the counterfactual framework. I think there is already a fairly widespread understanding of how mediation analysis has changed, and this book will only help further spread that awareness. On the topic of interaction analysis, I think there is less appreciation for how the counterfactual framework has changed thinking. This book serves as the remedy.

The book is divided into three parts. Part One of the book is devoted to the topic of mediation. Mediation analysis has long been widely used in some parts of the social sciences like psychology, but there has been I think a renewed interested in this topic in many fields. It was probably inevitable that interest in mediation would increase. Focusing on the identification and estimation of treatment effects in some cases provides us with evidence that a treatment works. It does not tell us why that treatment works. Mediation analysis focuses directly on this why question. At first blush, mediation analysis seems relatively simple. In many cases, multiplying two regression coefficients provides and estimate of a mediation effect.

The assumptions needed to give such estimates a causal interpretation, however, can be rather heroic. Heroic to the point that some reject mediation analysis in almost any form. Vanderweele provides both a brief introduction to the assumptions needed in Chapter 2, which serves as a introduction to basic mediation analysis. In Chapter 2, one can learn both essential ideas, but also understand what best practice should be for a standard mediation analysis. This chapter includes basic code to conduct mediation analysis in either Stata or SAS. Vanderweele returns to the topic of assumptions for mediation in two more parts of the book. First, he devotes Chapter 3 to the topic of sensitivity analysis, which allows analysts

to probe whether their results are likely to change if a key assumption has been violated. Sensitivity analysis is a generally under-utilized tool, and is needed for forms of analysis like mediation that depend on strong assumptions that cannot be justified by experimental design. One strength of the text is the fact that an entire chapter is used to outline and explain sensitivity analysis for mediation. Finally, in Chapter 7 he returns to the topic of assumptions and provides a very useful outline of the controversies that surround mediation, and he covers alternative identification strategies. Chapters 4-6 are devoted to topics that may not concern all readers. Chapter 4 explains the complications that arise when mediation analysis is conducted with survival data. Chapter 5 covers multiple mediators, and Chapter 6 explores the complications created by time-varying treatments.

Part Two of the book is devoted to interaction analysis. This is arguably the more important part of the book. While the topic of mediation is perhaps of greater interest, while there has been a proliferation of journal articles that discuss the intricacies of mediation, there has been considerably less attention to how the counterfactual framework has altered our understanding of interaction analysis. I think many readers might assume they already fully understand interactions since it would seem that such an analysis is simply a matter of including multiplicative terms on the right hand side of an regression model. The chapters in Part Two of this text clearly demonstrate that more subtle issues are often at work.

Chapter 9 serves as a basic primer on causal interactions. Vanderweele defines an interaction as a causal effect that depends on the presence of two treatments. He notes the key distinction between the interaction of two or more treatments as opposed to effect modification, where the effect of one treatment is modified by some typically pretreatment covariate like age or gender. This chapter also lays out the differences between additive and multiplicative interactions, which will be of interest to readers in the biomedical fields. Chapter 9 also includes basic information on mechanistic interaction, which is the case where an outcome is present only when two joint treatments are both present. Chapter 10 then is entirely devoted to the topic of mechanistic interactions. Finally chapter 11 presents sensitivity analysis methods for causal interaction analysis. These chapters deserve to be read by most applied researcher, as Vanderweele clearly explains the nature of interaction analysis when the goal is causal inference.

Part Three concludes with a synthesis of mediation and interactions. Part Three demonstrates how to unify the topics of mediation and interaction in Chapter 14. Chapter 15 then serves as a case study for how unifying the two concepts can be used to study spillover effects. The utility of the synthesis is fully realized in studies of how treatments spillover from treated to nontreated units. The final section of the text concludes with a very interesting discussion of the philosophical aspects of causation more generally. It is a fine and, I think, rare consideration of these topics.

I do have one criticism. The book is meant to be a primer on these topics for an applied audience. However, many applied topics are given short shrift. While the book is nearly 700 pages and includes over 180 pages of mathematical derivations and proofs, there is considerably less time devoted to the actual mechanics needed for an applied analysis. For

example, the text includes SAS and Stata code for additive interactions, but the code is never used in an empirical example. Moreover, while the SAS and Stata code in the text is available online, the data used in the examples is not provided. Therefore, the reader cannot use the code to replicate any of the analyses in the book. I think for many applied analysts the ability to replicate the applications from the text would have been useful.

While the book is designed to appeal to the social and biomedical sciences more broadly, there is a heavy emphasis on examples and methods used mostly in epidemiology and biostatistics. For example, many of the examples and methods are outlined in the context of the odds and risk ratio. Both of which rarely see any use in the social sciences. Readers from outside the biomedical sciences will have to become used to some of the vocabulary, but the adjustment should be quick.