This is a thoughtful and well written book, covering important issues of causal inference in every field of applied data analysis. However, while the book shines in the motivational and conceptual levels, it fades in the mathematical tools that are harnessed to support the conceptual discussions. My review will focus on this weakness, because it is the one factor that prevented me from enjoying the lucid discussion throughout.

The book is about bias or, more specifically, about “identifying and dealing with bias in statistical research on causal effects” (from the back cover). Naturally, readers expect to find methods, criteria, or algorithms that facilitate the “identification” of bias, its assessment or its control. However, with the exception of the illusive and “catch all” assumption of “ignorability,” the book stops short of showing readers what needs to be assumed or what needs to be done to control bias.
This void is a direct consequence of the restricted mathematical language chosen to illuminate examples, concepts and assumptions. This language, which has its roots in Rothman’s “sufficient cause” classification (Rothman, 1976) and Rubin’s “potential outcome” framework (Rubin, 1974) does not recognize modeling notions such as “processes” “omitted factors,” or “causal mechanisms” that guide scientific thoughts, but forces one to articulate knowledge through counterfactual categories such as “doomed,” “causal,” “preventive,” and “immune,” and the proportions of individuals in each category. It is an all-or-nothing framework. If one assumes “ignorability,” bias disappears; if not, bias persists, and one remains at the mercy of the (wrong) assumption that adjusting for as many covariates as one can measure would reduce bias. (Pearl, 2009a,b, 2011a; Rubin, 2009). The question of going from scientific knowledge to bias reduction, as well as the question of defending “ignorability-type” assumptions, remain outside the formal analysis.

The commitment to this mathematical language forces the examples to take the form of numerical tables involving counterfactual variables, rather than depicting the story behind the examples (e.g., through equations or diagrams.) Such tables may convince readers that the phenomenon demonstrated can indeed take place with certain tweaking of parameters, but fail to give readers a sense of the general class of problems where the phenomenon will occur.

Take, for example, Simpson’s paradox, which Weisberg (pp. 164-167) describes as “we might find that conditional effects are very similar within each stratum of the third factor (e.g., man and women), but opposite to the direction of the overall effect.” (Here, the word “effect” means “association naively presumed to represent effect.”) Weisberg rightly continues to the heart of Simpson’s paradox and asks:

“The question then arises of which effect (adjusted or unadjusted) represents a causal effect. Usually, it is assumed that the more “refined” conditional analysis represents the “true” causal effect, reflected in the common effect within strata, whereas the unadjusted effect results from confounding.”

At this point the book does not stop to tell us if this “usual” assumption is valid or not (it is not) or how one can go about deciding when it is valid. Instead, we are instructed
to construct tables involving doomed/causal/preventive/immune categories of individuals, translate them into spreadsheet, from which we can “calculate the empirical effects under any set of assumptions about the various parameters.”

The reader thus gets the impression that, in order to determine the key question: “which effect (adjusted or unadjusted) represents a causal effect” one needs to guess the relative sizes of the doomed/causal/preventive/immune strata in both the male and female population, for both the exposed and unexposed groups, arrange them in a table like those in Tables 7.12, 7.13 and 7.14, go through the arithmetics, and only then conclude which effect is causal?” This is unrealistic, because if we knew the relative sizes of those strata, we would not be facing Simpson’s dilemma in the first place, but calculate the causal effect directly.

Modern treatments of Simpson’s paradox can and should tell us how to make this determination directly from the causal story behind the example (See, for example, Pearl (2009c, p. 383)) without guessing relative sizes of strata and without going through the lengthy arithmetics.

Again, it is not a fault of Weisberg, but of the language of tables and strata which does accept any such notions as “the causal story behind the example.” More generally, this language does not allow for causal assumptions to be articulated in a format that matches the way scientific knowledge is stored and communicated. Weisberg has done an incredibly fine job overcoming this basic limitation of the potential outcome language, but there are limits to what good writing can do when mathematical notation is opaque.

Sailing on good writing, Weisberg manages to walk the reader through an impressive array of concepts and topics, including collapsibility, propensity scores, sensitivity analysis and mediation. Each topic is introduced in a proper technical perspective, starting with its historical roots and ending with its impact on modern causal analysis. It is unfortunate though that the discussion is occasionally marred by myths that once served popular folklore and have since been discarded by analysis.

One such myth is the belief that the use of propensity-score somehow contributes to bias reduction, that it requires no modeling assumptions, and it is ”to some degree, capable of providing warnings that the available data may not support a valid causal estimate.” (Weisberg, pp. 141)
Mathematical analysis has overturned these beliefs. The proper choice of covariates into
the propensity-score is dependent critically on modeling assumptions. (Pearl, 2009a,b, 2011a;
Rubin, 2009). The propensity-score method, like any other model-free analysis, cannot give
us any warning about the invalidity of the causal estimates. Finally, the propensity-score is
merely a powerful estimator, and conditioning on the propensity score would be theoretically
equivalent (asymptotically) to controlling on its covariates, regardless of whether strong
ignorability holds (Pearl, 2009c, p. 349).

Another myth that finds its way to Weisberg’s book concerns causal mediation, sometimes
called direct and indirect effects. According to Weisberg (p. 208), “The theory of principal
stratification has helped to clarify the essential nature of causal mediation.” The hard truth
is that principal stratification has helped circumvent, rather than clarify the essential nature
of causal mediation. Most participants in a public discussion of the usages of principal strata,
including former proponents of this framework now admit that principal strata has nothing
to do with causal mediation. (Joffe, 2011; Pearl, 2011b; Sjölander, 2011; VanderWeele, 2011).

To summarize, this book would be an excellent companion to standard statistics texts,
serving to elucidate the unique problems that data analysts face when challenged to assess
causal-effect relationships. When it comes to solving those problems though, the book should
be supplemented with one that properly demonstrates the mathematics of modern causal
analysis.

References


JOFFE, M. M., YANG, W. P. and FELDMAN, H. I. (2010). Selective ignorability assump-
tions in causal inference. The International Journal of Biostatistics 6. Article 11, DOI:


