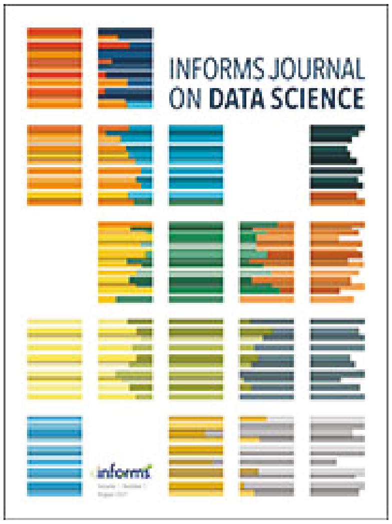

INFORMS Journal on Data Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Commentary on “Causal Decision Making and Causal Effect Estimation Are Not the Same… and Why It Matters”

Edward McFowland III

informs
http://pubsonline.informs.org/journal/ijds

INFORMS JOURNAL ON DATA SCIENCE
Articles in Advance, pp. 1–2
ISSN 2694-4022 (print), ISSN 2694-4030 (online)

Commentary on "Causal Decision Making and Causal Effect Estimation Are Not the Same … and Why It Matters"

Edward McFowland III[a]

[a] Technology and Operations Management, Harvard Business School, Harvard University, Boston, Massachusetts 02163
Contact: emcfowland@hbs.edu, https://orcid.org/0000-0001-5249-7117 (EMF)

Received: November 16, 2021
Accepted: November 25, 2021
Published Online in Articles in Advance: March 9, 2022

https://doi.org/10.1287/ijds.2021.0010

Funding: This work was partially supported by the National Science Foundation Program on Fairness in Artificial Intelligence in Collaboration with Amazon [Grant IIS-2040898].

Keywords: casual inference • treatment effect estimation • treatment assignment policy • fairness • human-in-the-loop

1. Introduction

There has been a substantial discussion in various methodological and applied literatures around causal inference; especially in the use of machine learning and statistical models to understand heterogeneity in treatment effects and to make optimal decision allocations. I commend Fernández-Loría and Provost (2021) for highlighting the important, and in retrospect, intuitive differences between causal estimation and causal decision making. This commentary is aimed at expanding the conversation in fruitful directions, with an eye toward real-world practice in organizations. Specifically, I highlight that future work will need to address how to exploit these theoretical results in practice (Section 2), how to ensure that causal decisions are fair (Section 3), and what are the additional benefits and challenges when there is a human in the loop (Section 4).

2. From Theory to Practice

Fernández-Loría and Provost (2021) provide an intuitive argument about the differences between causal effect estimation (CEE) and causal decision making (CDM), and now the conversation in the literature should progress to identify concrete steps for utilizing these compelling "theoretical" results in practice. Their figure 2 provides three scenarios juxtaposing the effect estimates of a biased model (BM) with an unbiased model, relative to a decision boundary τ. In their final scenario, representing reinforcing bias, BM (based on large confounded data) is more valuable for decision making; however, in practice, one is not generally able to determine whether BM provides desirable reinforcing bias or, alternatively, large and opposite bias. If we also consider their figure 1 (in light of figure 2), we realize that it shows the error for an individual (covariate profile). Therefore, unless Model 1 is particularly fortuitous, it will also underestimate some effects. When considered in conjunction with Model 1's larger variance, it appears that this can actually produce the undesirable consequences of large and opposite bias (figure 1(a)), engendering less confidence from practitioners. This may precisely be why one often opts for the CEE model built on (potentially) smaller but unconfounded data because even though "traditionally 'good' estimates of causal effects are not necessary to make good causal decisions," (Fernández-Loría and Provost 2021, p. 4) these estimates (in the limit) are reliable. Therefore, now that we are convinced that CDM is different from CEE, it must be asked how to reliably exploit this difference in practice.

3. Fairness

Staying with the concept of practice, in the present day, it is difficult (and often unwise) to consider the topic of algorithmic decision making and not consider the question of algorithmic fairness,[1] especially when discussing algorithms making automated decisions. Therefore, given that fairness is beyond the scope of the original work, I want to take this opportunity to highlight its importance in the setting of CDM. Although CDM exploits the important fact that "overestimating (underestimating) the effect has no bearing on decision making when the focal individuals have an effect greater (smaller) than" the decision threshold, the (unintended) consequences of misestimation can be profound and should be an integral part of this conversation. Borrowing the authors' tagline, I contend that "Fair Casual Decision Making and Fair Causal Estimation Are Not the Same … and It Matters."

As the authors do, let us consider the case of deciding whether to advertise to (i.e., allocate to treatment) users and imagine the advertisement is for a higher interest-rate loan. The algorithm is going to be concerned with estimating a quantity of the form $P(x) = P(Y(1) - Y(0) > t \mid X = x)$, the probability that treating an individual with covariate profile x will lead to a "sufficiently higher" outcome for the firm, where the decision threshold t is selected to encode the firm's risk-reward trade-off. Let us now consider two covariate profiles x_m and x_n, where there is a single difference between the two: the feature capturing whether the user belongs to an underrepresented minority group is true for x_m and false for x_n. Finally, assume $P(x_m) = 0.5 + \epsilon$ and $P(x_n) = 0.5 - \epsilon$, for some very small $\epsilon > 0$, representing a slightly larger probability that treating x_m will lead to a sufficiently higher outcome as compared with treating x_n. Further, assume that users are allocated to treatment when $P(x)$ (or a machine learning estimate $\hat{P}(x)$) is greater than the threshold 0.5.[2] Although this may lead to a higher utility (causal lift in revenue), all and only the members of the disadvantaged minority group will be offered the higher-interest advertising, leading to potentially widely disparate impacts. Moreover, these disparate impacts will exacerbate existing societal disparities to further disadvantage an already-disadvantaged group. Informed algorithmic decision making often requires this mapping of a probabilistic measure of benefit into some discrete action space; although it makes sense to focus on maximizing utility, there is a question of if (or at what point) the disparate impact is unjustified (Jung et al. 2018). Intuitively, when ϵ is small, it seems unfair because we believe that similar subpopulations—not just with respect to features but also with respect to utility—should receive similar treatment (Dwork et al. 2012). This is a contrived example; but the core point is critical to the conversation: causal decision making brings with it some benefits but also some challenges. As the literature begins to build upon this work, it seems pertinent to use a lens of how do we maximize utility while minimizing (unjustified) disparate impact.

4. Decisions vs. Recommendations

Finally, there is a large and important question of whether algorithms *should* be making decisions automatically, given the well-documented harms from unintended consequences of these decisions (Rainie and Anderson 2017). I imagine there may not be a single correct answer or even a binary decision as to the appropriateness of algorithmic decision making but rather a continuum from low- to high-stakes settings. Regardless, a curiosity is how the questions and conclusions from this work change if we move from causal decision making to causal *recommendations* provided to a human decision maker, who may consider these recommendations as part of his or her own decision-making process. In such a case, we may not simply consider the estimation of models built on potentially biased data but also (the need to model) the behavior of a likely biased decision maker. Moreover, the causal decision-making framework must also consider how the recommendations (i.e., additional information) will causally impact the decision maker's eventual decision, with the goal of achieving a fairer decision than either the algorithm or the decision maker would achieve on his or her own.

5. Conclusion

In summary, Fernández-Loría and Provost (2021) explore a critical distinction between estimation of causal effects and the causal decisions they intend to drive. Along the way, they articulate the value of this distinction when decision making is the final objective. Armed with this knowledge, the literature must investigate its usability and reliability in practical settings. Doing so could enable organizations to squeeze more value from their data in an era where data are seemingly the key determinate of success.

Endnotes

[1] I refrain from using the term *bias* to not overload its meaning in this context.

[2] I consider $P(x) = P(Y(1) - Y(0) > t \mid X = x)$ because the disparate impact will exist *even if* the machine learning estimates are perfect (i.e., $\hat{P}(x) = P(x)$). The central point will still stand if we consider $E(x) = E[Y(1) - Y(0) \mid X = x]$; but in this case, the disparate impact will exist when the algorithm $\hat{E}(x)$ overestimates (underestimates) the effects such that $\hat{E}(x_m) > t$ while $\hat{E}(x_n) < t$. Therefore, although this is still plausible in practice, that is, in finite samples, it should resolve asymptotically.

References

Dwork C, Hardt M, Pitassi T, Reingold O, Zemel R (2012) Fairness through awareness. *Proc. Third Innovations Theoret. Comput. Sci. Conf.* (Association for Computing Machinery, New York), 214–226.

Fernández-Loría C, Provost F (2021) Causal decision making and causal effect estimation are not the same … and why it matters. Preprint, submitted April 8, https://arxiv.org/abs/2104.04103.

Jung J, Corbett-Davies S, Shroff R, Goel S (2018) Omitted and included variable bias in tests for disparate impact. Preprint, submitted September 15, https://arxiv.org/abs/1809.05651.

Rainie L, Anderson JQ (2017) Code-dependent: Pros and cons of the algorithm age. Accessed December 17, 2022, https://www.pewresearch.org/internet/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/.