

DISCUSSION ON “CAUSAL AND COUNTERFACTUAL VIEWS OF MISSING DATA MODELS”

Alexander W. Levis and Edward H. Kennedy*

Carnegie Mellon University

1. Introduction

We congratulate Nabi et al. (2022) on their impressive and insightful paper, which illustrates the benefits of using causal/counterfactual perspectives and tools in missing data problems. This paper represents an important approach to missing-not-at-random (MNAR) problems, exploiting nonparametric independence restrictions for identification, as opposed to parametric/semiparametric models, or resorting to sensitivity analysis. Crucially, the authors represent these restrictions with missing data directed acyclic graphs (m-DAGs), which can be useful to determine identification in complex and interesting MNAR models. In this discussion we consider

- (i) how/whether other tools from causal inference could be useful in missing data problems,
- (ii) problems that combine missing data and causal inference together, and
- (iii) some work on estimation in one of the authors' example MNAR models.

2. Other Causal Identification Tools

The graphical arguments used in Nabi et al. (2022) stem from a powerful analogy to a widely used set of tools in the causal inference literature. In drawing this connection, their work raises the question: *which other tools and methodologies can be borrowed from the causal literature and bear fruit for missing data problems?*

The focus of Nabi et al. (2022) is—for the most part—on characterizing identification of the full data law, $p(L^{(1)}, R)$, under sets of (factual and counterfactual) conditional independence relations implied by a directed acyclic graph and its associated structural model. However, it is often sufficient for practical purposes to identify and estimate certain functions or functionals of this distribution, say the marginal “target” data mean functional $\mathbb{E}(L^{(1)})$, rather

*Corresponding author. E-mail: edward@stat.cmu.edu