Statistica Sinica **36** (2026), 1-12
doi:https://doi.org/10.5705/ss.202025.0393

RESPONSE TO DISCUSSIONS OF "CAUSAL AND COUNTERFACTUAL VIEWS OF MISSING DATA MODELS"

Razieh Nabi, Rohit Bhattacharya, Ilya Shpitser and James M. Robins

Emory University, Williams College,
Johns Hopkins University and Harvard University

1. Introduction

We are grateful to the discussants—Levis and Kennedy (2025), Luo and Geng (2025), Wang and van der Laan (2025), and Yang and Kim (2025)—for their thoughtful comments on our paper (Nabi et al., 2025). Below we summarize our main contributions before responding to each discussion in turn.

Graphical models have emerged as an important tool for clarifying identifying assumptions made in both causal inference (Pearl and Robins, 1995; Pearl, 2000) and missing data (Robins and Gill, 1997; Bhattacharya et al., 2019; Nabi, Bhattacharya and Shpitser, 2020; Malinsky, Shpitser and Tchetgen Tchetgen, 2021; Mohan and Pearl, 2021). Our paper (Nabi et al., 2025) shows how recent techniques motivated by causal graphical modeling may be fruitfully applied to obtain identification in missing data models. These recent techniques allow us to obtain novel identification results that would not be possible to obtain in standard causal inference problems, except by imposing implausible additional assumptions, such as rank preservation.

Specifically, we formalize how identifiability of the target (i.e., complete) data law can be viewed as identification of a joint distribution over counterfactuals $L^{(1)}$, the variables that would have been observed if all missingness indicators R were set to one (that is, if no data were missing).

This reframing yields a counterfactual analogue of the g-formula. When assumptions encoded in the Markov properties of a missing data DAG (m-DAG) allow us to express the counterfactual g-formula in terms of the factuals, we obtain (nonparametric) identification. That is, noting that the target law $p(l^{(1)})$ satisfies $p(l^{(1)}) = p(l, R = 1)/p(R = 1 \mid l^{(1)})$, it follows that when the missingness selection model $p(R = 1 \mid L^{(1)})$ is identified from the observed law, so is $p(l^{(1)})$.

In our paper, we used the word "nonparametric" in two different ways. The joint distribution of $(L, L^{(1)}, R)$ is Markov to a given m-DAG if it factorizes as the product of the conditional densities of each variable given its parents. In the

*Corresponding author. E-mail: razieh.nabi@emory.edu