Statistica Sinica **36** (2026), 1-5
doi:https://doi.org/10.5705/ss.202025.0165

DISCUSSION ON "CAUSAL AND COUNTERFACTUAL VIEWS OF MISSING DATA MODELS"

Shu Yang and Jae Kwang Kim*

North Carolina State University and Iowa State University

1. Introduction

This paper presents a new perspective on missing data by integrating causal and counterfactual frameworks, reinterpreting missing data as a causal inference problem. By drawing an analogy between missing data and unobserved counterfactuals, the paper advances a structured approach to understanding and identifying parameters under different missingness mechanisms, particularly in Missing Not At Random (MNAR) settings. Utilizing Directed Acyclic Graphs (DAGs) and their extensions to missing data DAGs (m-DAGs), this work provides a rigorous framework for characterizing dependencies and assumptions in missing data problems. A key contribution is the extension of the g-formula to accommodate counterfactual distributions in missing data models, offering new insights and methodological advancements for addressing MNAR challenges.

Challenges in Rubin's Hierarchy of Missing Data Mechanisms. Rubin's classification (Rubin, 1976)—*Missing Completely at Random (MCAR)*, *Missing at Random (MAR)*, and *Missing Not at Random (MNAR)*—has long served as the foundational framework for modeling missingness. However, this hierarchy introduces both conceptual and practical limitations, especially under the MAR assumption. Specifically, MAR requires that the missingness indicator R depends only on observed outcomes Y^{obs} and auxiliary covariates X, but not on the missing components Y^{mis}:

$$\mathbb{P}(R \mid Y^{\text{obs}}, Y^{\text{mis}}, X) = \mathbb{P}(R \mid Y^{\text{obs}}, X).$$

Simulating data that strictly conform to MAR is challenging due to the implicit dependencies between Y^{obs} and R. Since the observed data are themselves a function of the missingness mechanism, defining a coherent MAR structure across different patterns is nontrivial. These interdependencies expose foundational limitations in Rubin's taxonomy.

*Corresponding author. E-mail: jkim@iastate.edu