Adapting Multiple Imputation for Compositional Survey Data

Sana Gupta, Benjamin Stockton, & Ofer Harel

Department of Statistics, University of Connecticut, Storrs, CT

https://doi.org/10.33697/ajur.2025.146

*Students: sana.gupta@uconn.edu**
Mentors: benjamin.stockton@uconn.edu, ofer.harel@uconn.edu

APPENDIX

In this appendix, we have included figures displaying the bias, coverage, and standard error for the MNAR and MCAR simulation data.

MCAR (Missing Completely at Random)

We use an MCAR mechanism that applies a logistic model $\text{logit}(P(R=0)) = \alpha_0$ to apply missingness on ILR_2, where $\alpha_0 = -\log(1/p_{miss} - 1)$. The probability of missingness should be independent of both the response variable Y and other observed data, ensuring that the missingness occurs completely at random. Under MCAR, there should be no bias under CCA or MI.

MNAR (Missing Not at Random)

We use an MNAR mechanism that uses a logistic model,

$$\begin{aligned}\text{logit}(P(R=0)) =& \alpha_0 + 5physical - 2weight + stress - fitness + 2challenge + \\ & 3PIE_1 + PIE_2 + PIE_4,\end{aligned}$$

to induce missingness on ILR_2, such that the probability of missingness depends directly on the unobserved values of ILR_2 and where

$$\begin{aligned}\alpha_0 = -\,(&5\ \text{mean}(physical) - 2\ \text{mean}(weight) + \ \text{mean}(stress) - \\ &\text{mean}(fitness) + 2\ \text{mean}(challenge) + \\ &3\ \text{mean}(PIE_1) + \ \text{mean}(PIE_2) + \ \text{mean}(PIE_4)) - \\ &\log(1/p_{miss} - 1),\end{aligned}$$

and $\text{mean}(x)$ is the sample mean of x. This mechanism causes the missing values to be dependent on the values of ILR_2 as ILR_2 is a function of all of the PIE_j variables, which leads to bias in complete case analysis and MI.

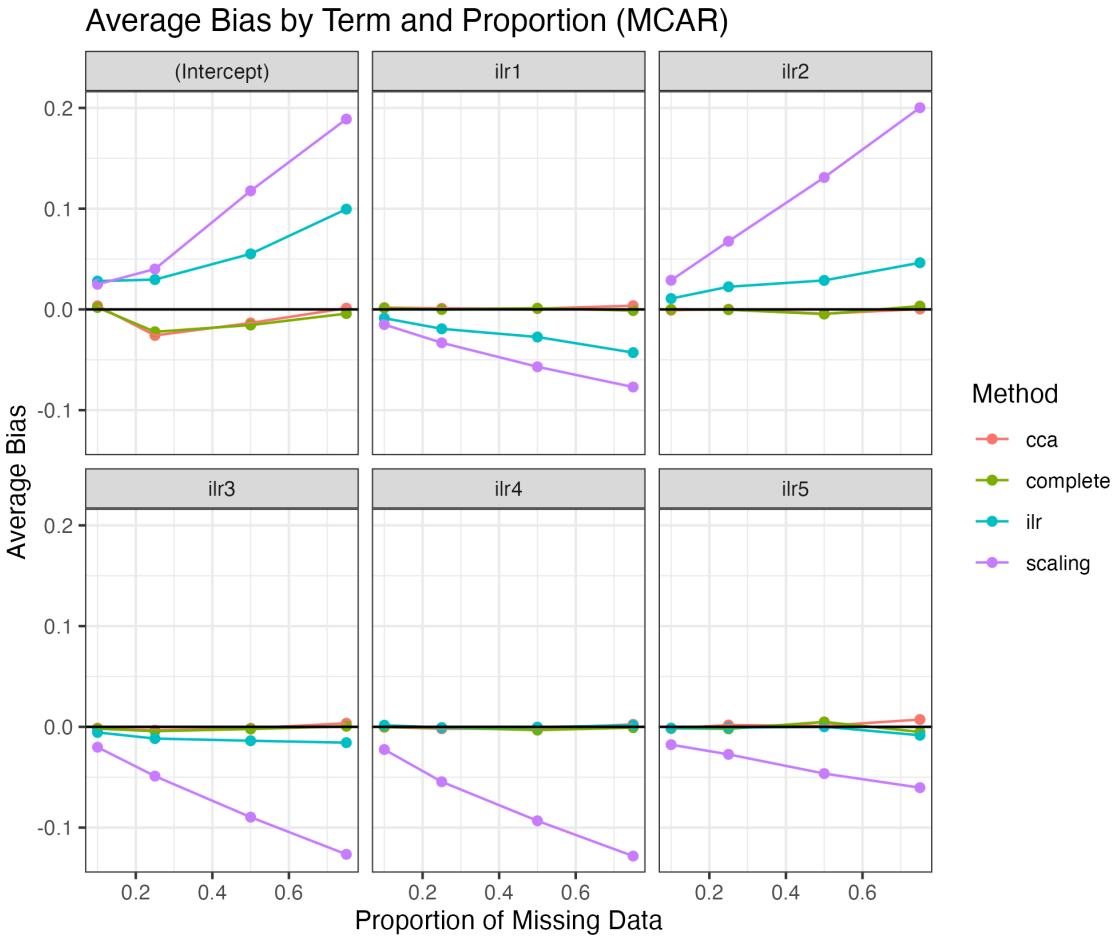

Figure 1. Graph of the average bias for each method at each missingness proportion. The averages were taken after computing the bias for each MCAR simulation iteration.

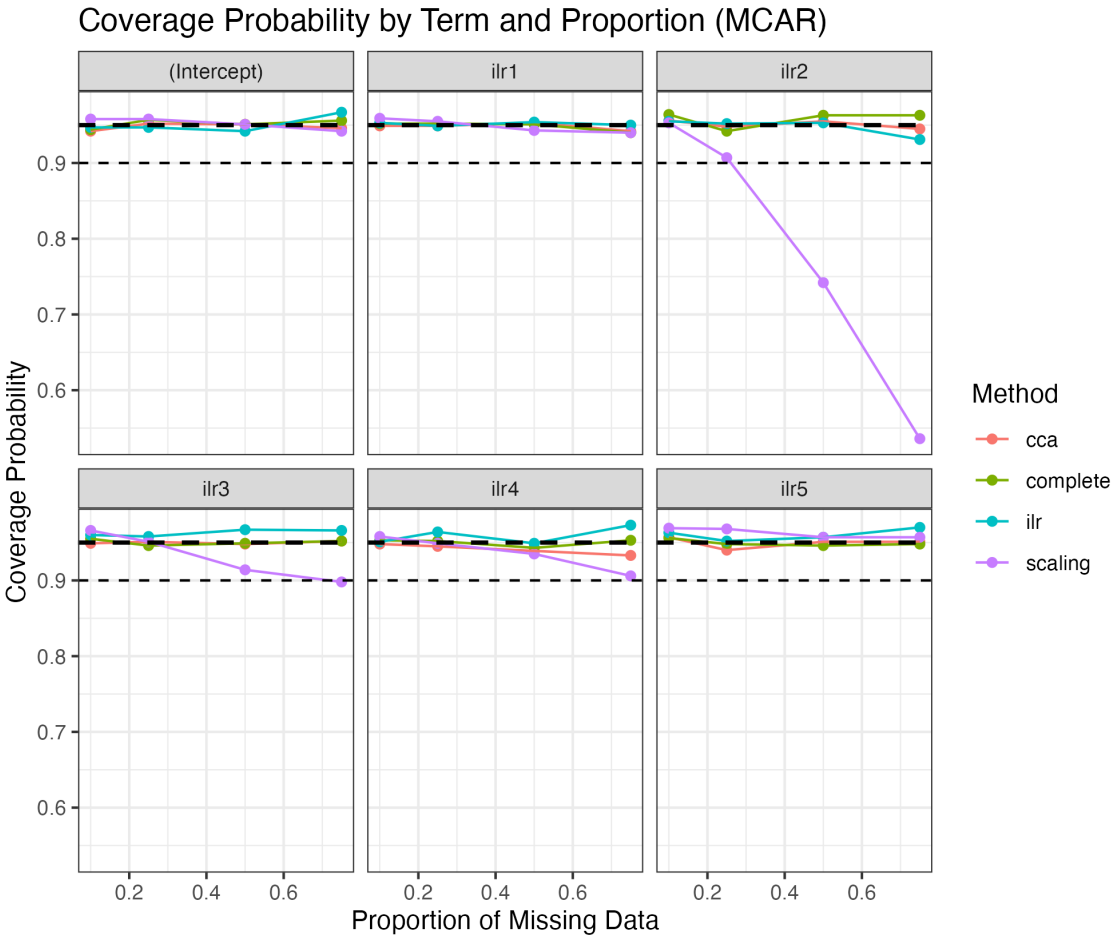

Figure 2. Graph of the coverage probability for each model term at each missingness proportion. The thinner line is at the 0.9 level, and the thicker line is at 0.95.

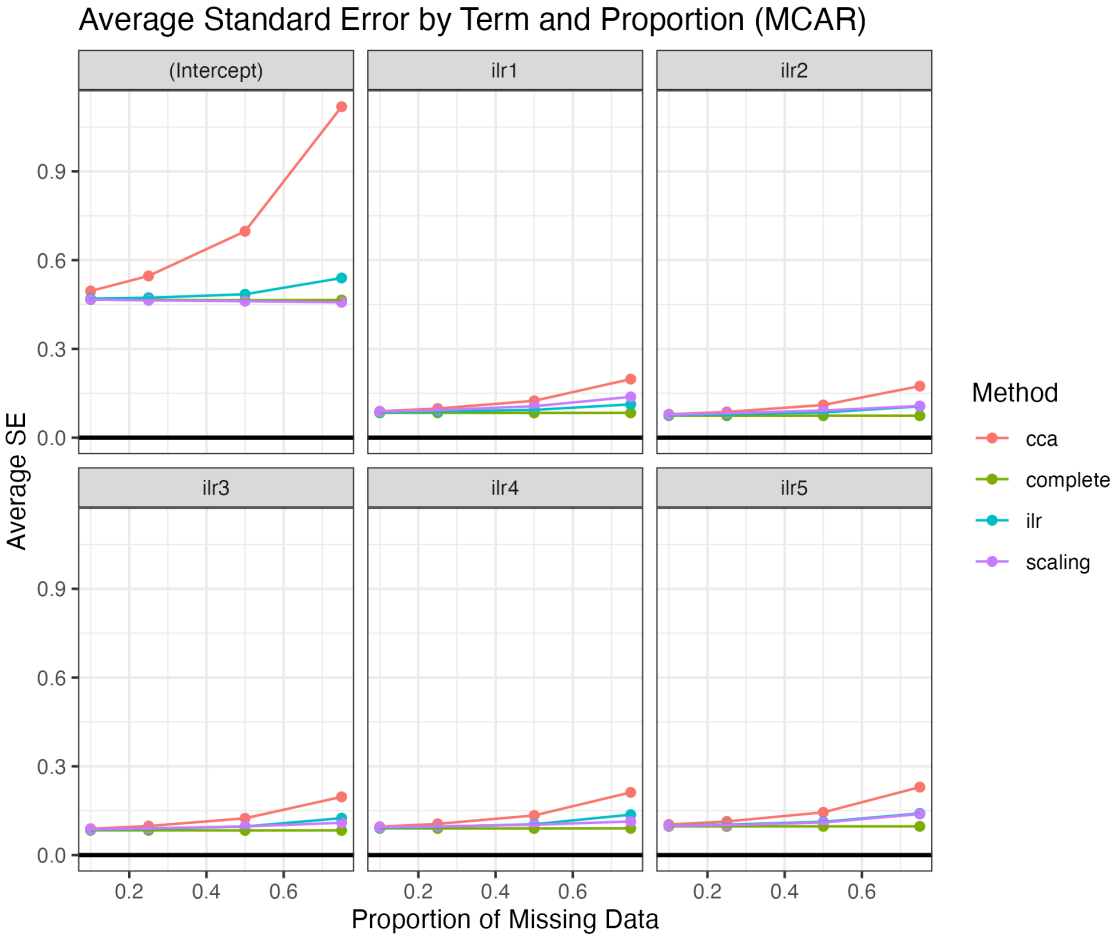

Figure 3. Graph of the standard error for each method at each missingness proportion.

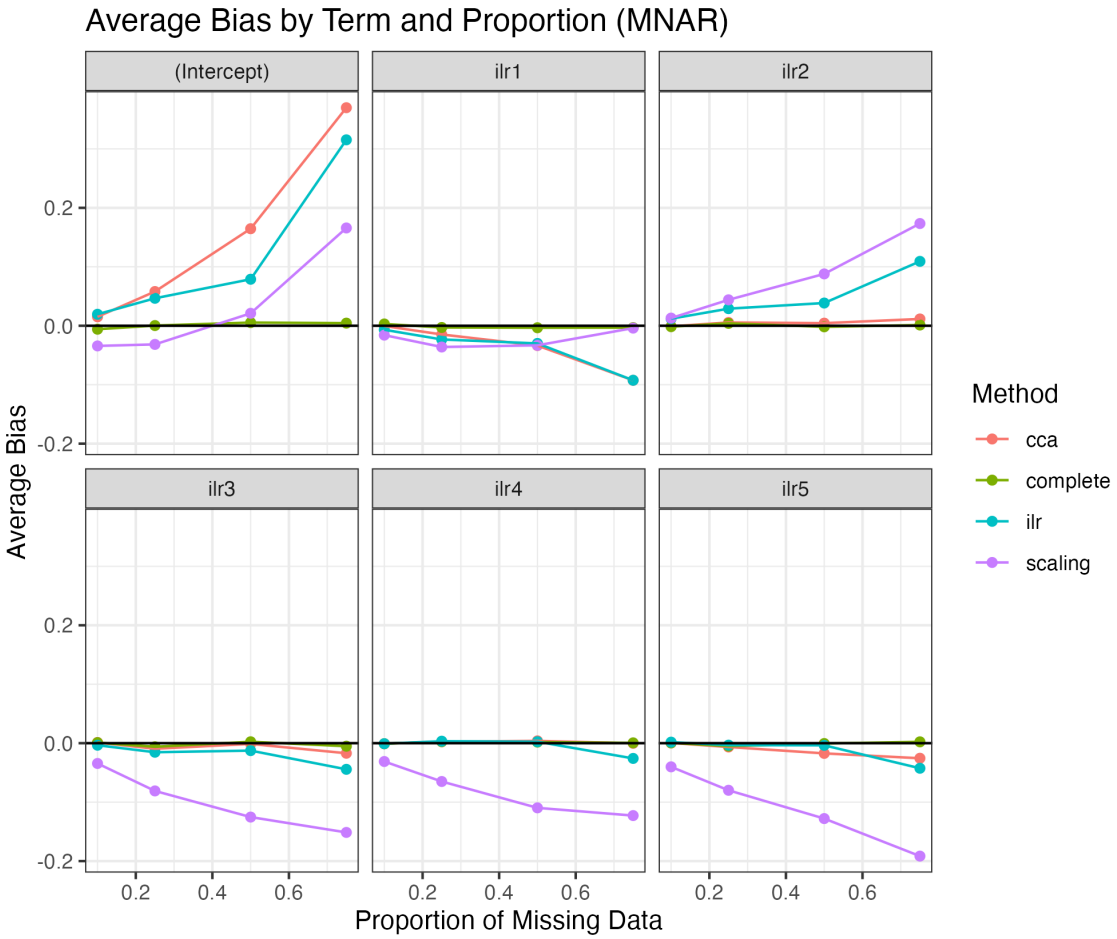

Figure 4. Graph of the average bias for each method at each missingness proportion. The averages were taken after computing the bias for each MCAR simulation iteration.

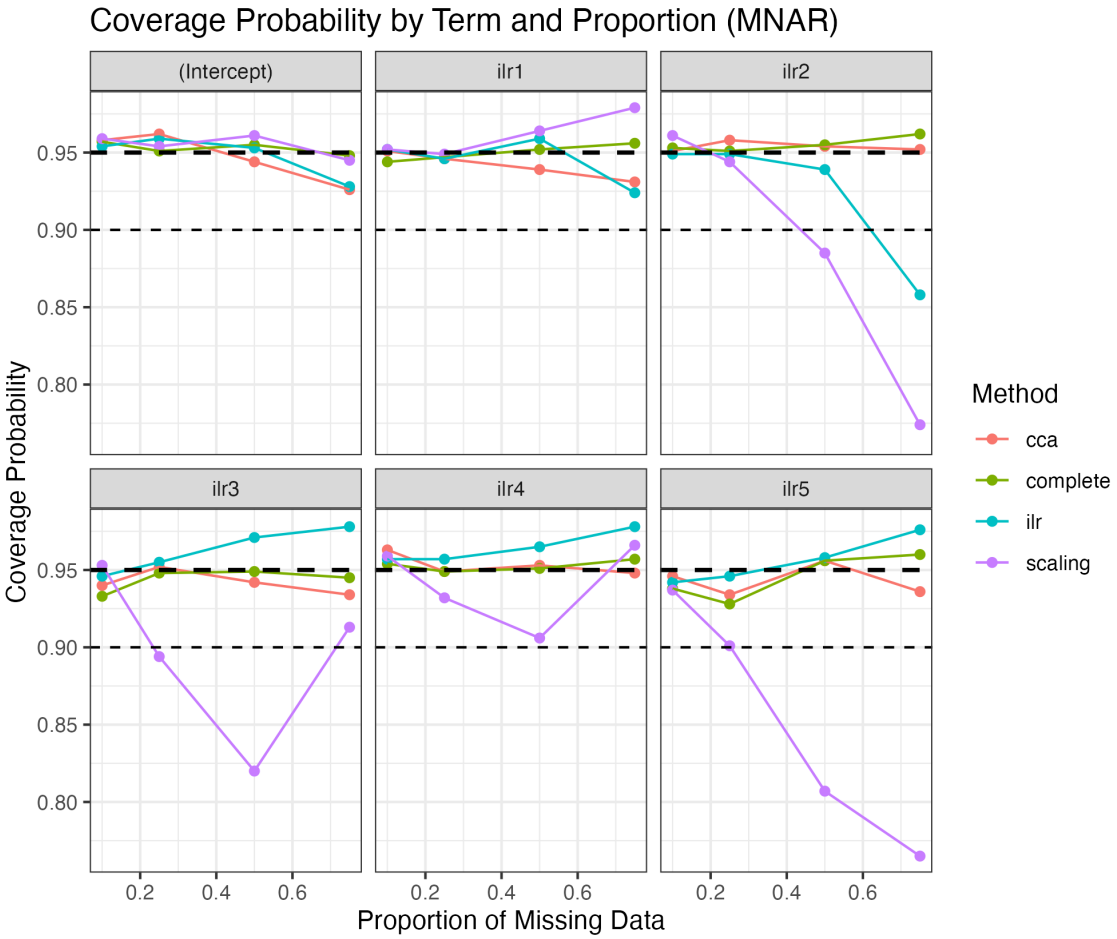

Figure 5. Graph of the coverage probability for each model term at each missingness proportion. The thinner line is at the 0.9 level, and the thicker line is at 0.95.

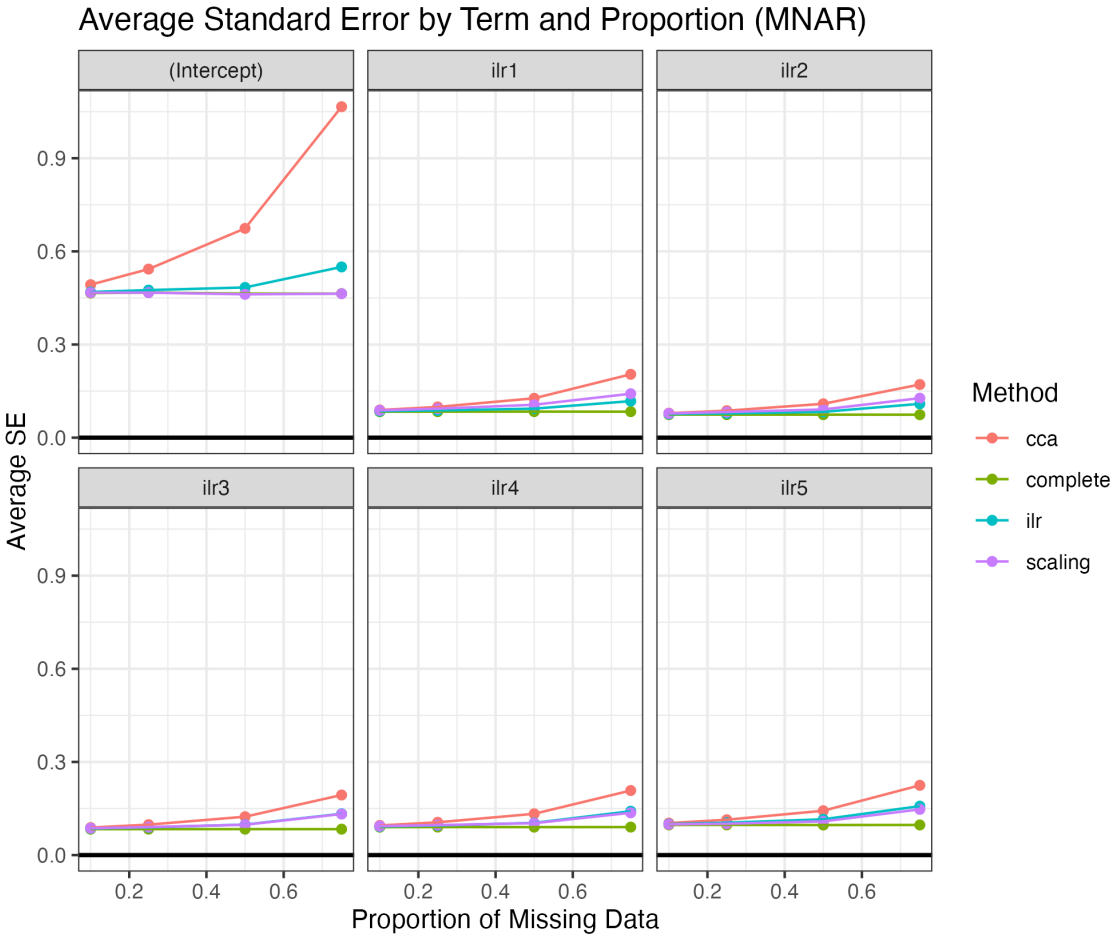

Figure 6. Graph of the standard error for each method at each missingness proportion.