

A Comparative Simulation Study: Evaluation and Improvement of Multiple Imputation Method of MIHEml

Mo Li, Xin Shore

Submitted to: JMIR Data
on: February 23, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5
Supplementary Files..... 11
 Multimedia Appendixes 12
 Multimedia Appendix 1..... 12
 Multimedia Appendix 2..... 12



A Comparative Simulation Study: Evaluation and Improvement of Multiple Imputation Method of MIHEml

Mo Li^{1*}; Xin Shore^{2, 2*} PhD

¹ University of Louisiana Lafayette US

² University of New Mexico Albuquerque US

*these authors contributed equally

Corresponding Author:

Xin Shore PhD

University of New Mexico
2400 Tucker Ave NE
Albuquerque
US

Abstract

Background: Missing data imposes significant challenges in clinical epidemiology, particularly when the missingness mechanism is missing not at random (MNAR). Although substantial progress has been made in addressing missing completely at random (MCAR) and missing at random (MAR) mechanisms, methods for handling MNAR data remain underexplored. Galimard et al.[12] (2018) proposed a multiple imputation method using Heckman's One-Step Maximum Likelihood Estimation for binary MNAR outcomes with continuous MAR covariates (MIHEml). A detailed exploration of MIHEml's performance is crucial to assess its potential to produce more reliable inference when the outcome is MNAR and the covariates are MAR.

Objective: This study aims to upgrade the method of MIHEml, assess its performance comparing to a traditional missing data method, complete case analysis (i.e., CCA) and then validate its effectiveness through extensive simulations. We hope to provide researchers with a more robust and reliable tool for addressing missing data challenges, especially when data are missing not at random.

Methods: The simulation study began by generating datasets with no missing values as the full datasets. The datasets included a binary outcome and various types of covariates. Missingness was assigned to the outcomes under MNAR and to the covariates under MAR. Five sample sizes and five missing rates were combined to create 25 scenarios. Therefore, twenty-five different imperfect datasets were generated. CCA and MIHEml were applied to correct the missingness situation in the imperfect datasets. A logistic regression was separately fit to the full datasets, datasets with only observed values under CCA and multiply-imputed datasets under MIHEml. Estimation accuracy was compared in terms of the absolute value of the percent relative bias, root mean squared error, coverage and average confidence interval length. Prediction accuracy was measured by accuracy from a classification table.

Results: Comparing to CCA, MIHEml presents a comprehensively better performance with the less absolute value of the percent relative bias, smaller RMSE, narrower average 95% confidence interval, and higher coverage probability. MIHEml also effectively improves the prediction accuracy when the types of missing variables, sample sizes and missing rates are varying.

Conclusions: MIHEml showed the expected clear-cut superiority over CCA when the binary outcome is assumed MNAR and different types of covariates are assumed MAR.

(JMIR Preprints 23/02/2025:73025)

DOI: <https://doi.org/10.2196/preprints.73025>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to the public.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in a JMIR journal, I will be able to make my accepted manuscript PDF available to anyone.

No. Please do not make my accepted manuscript PDF available to anyone.



Original Manuscript

Figure 1 – Absolute Value of the Percent Relative Bias

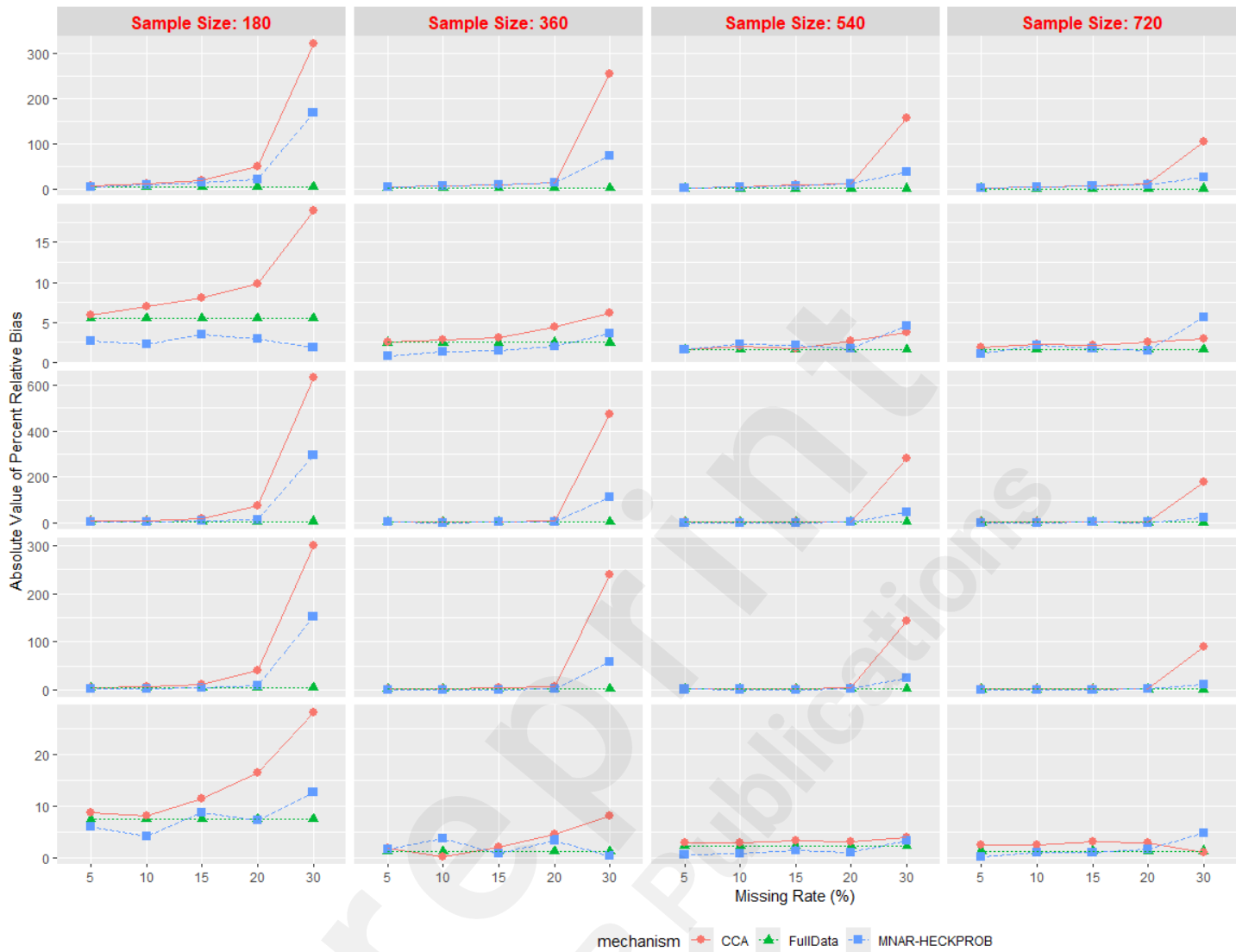


Figure 2 – Root Mean Squared Error (RMSE)

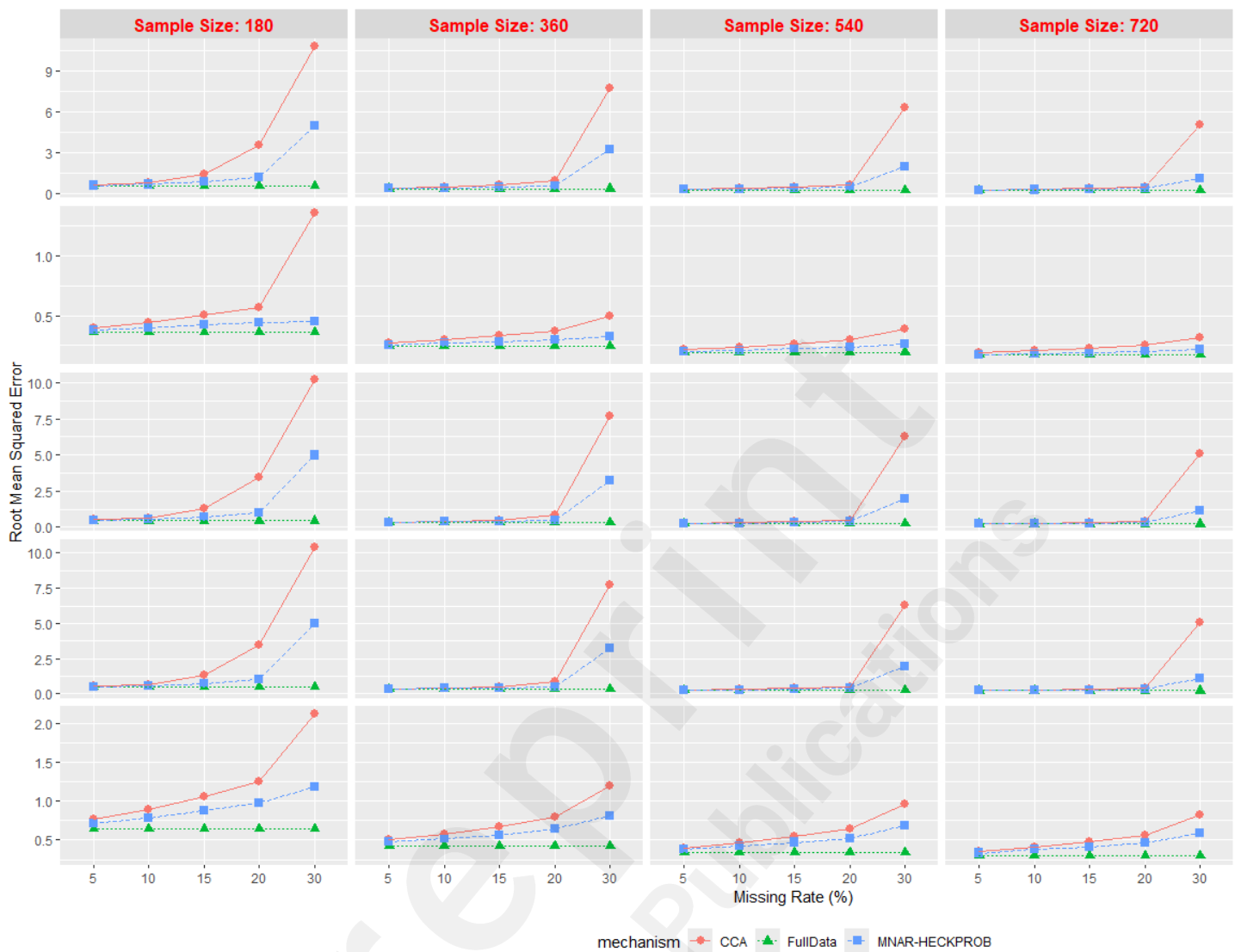


Figure 3 – Confidence Interval Length with the Transformation of $\log_{10}$

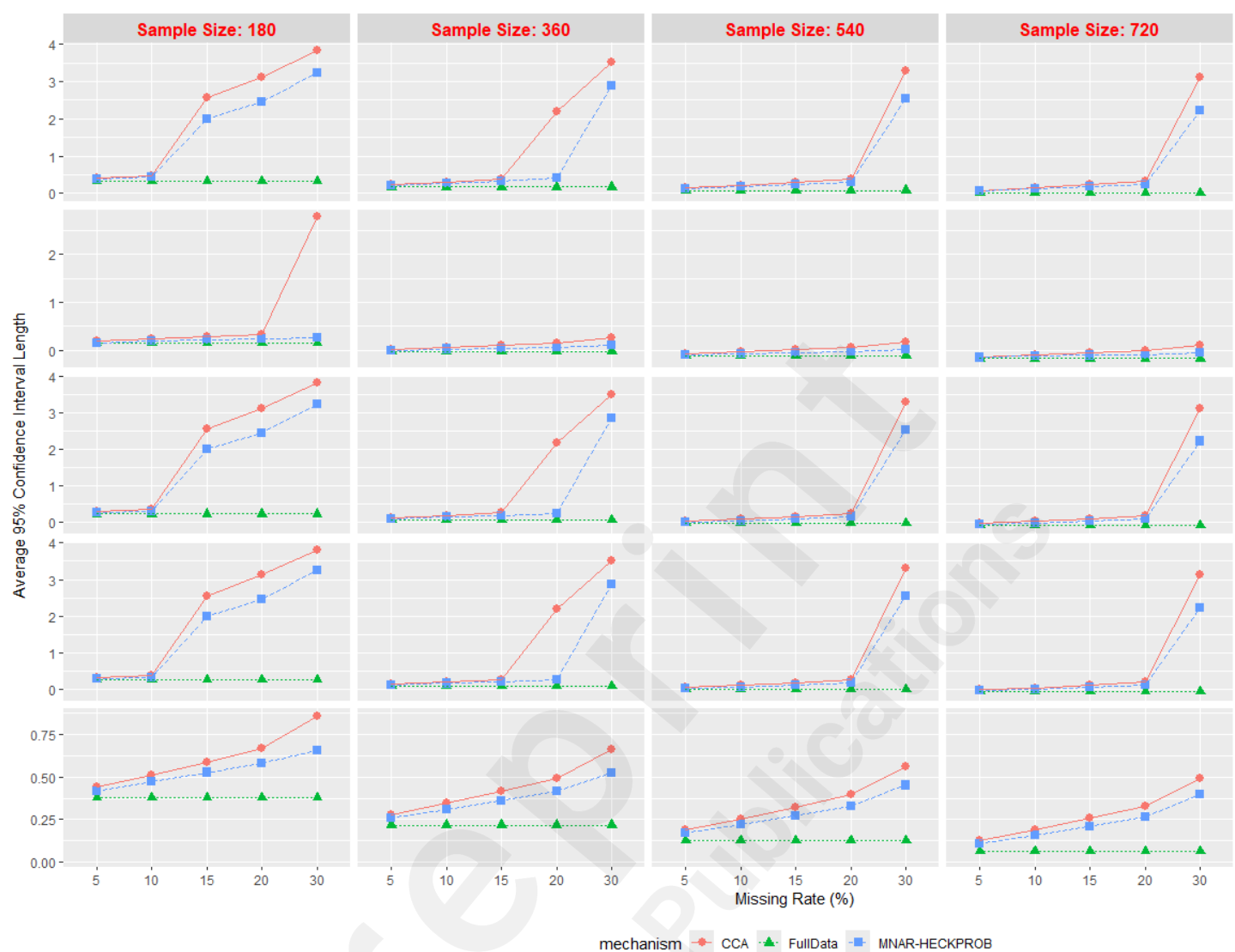


Figure 4 – Coverage

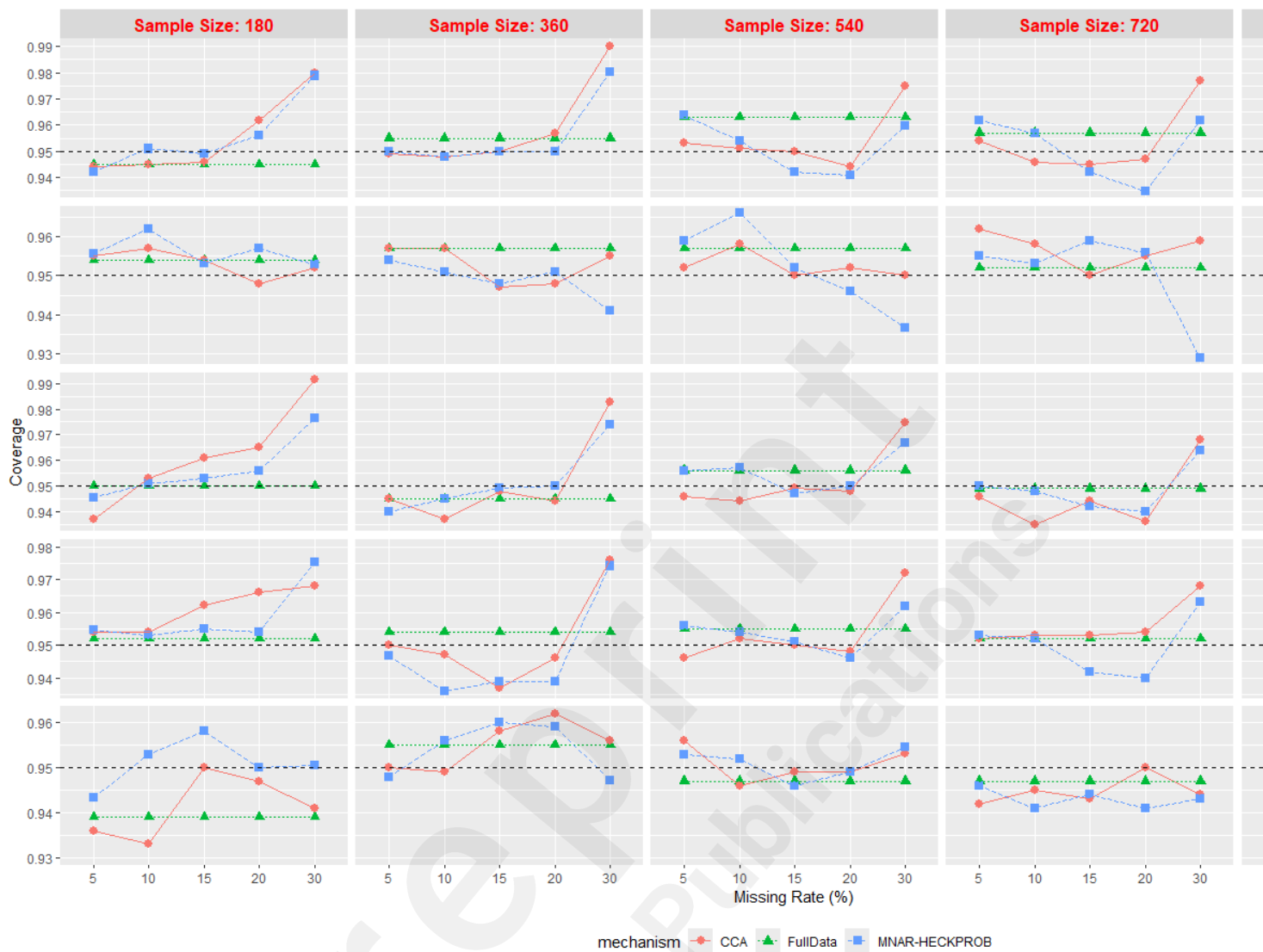
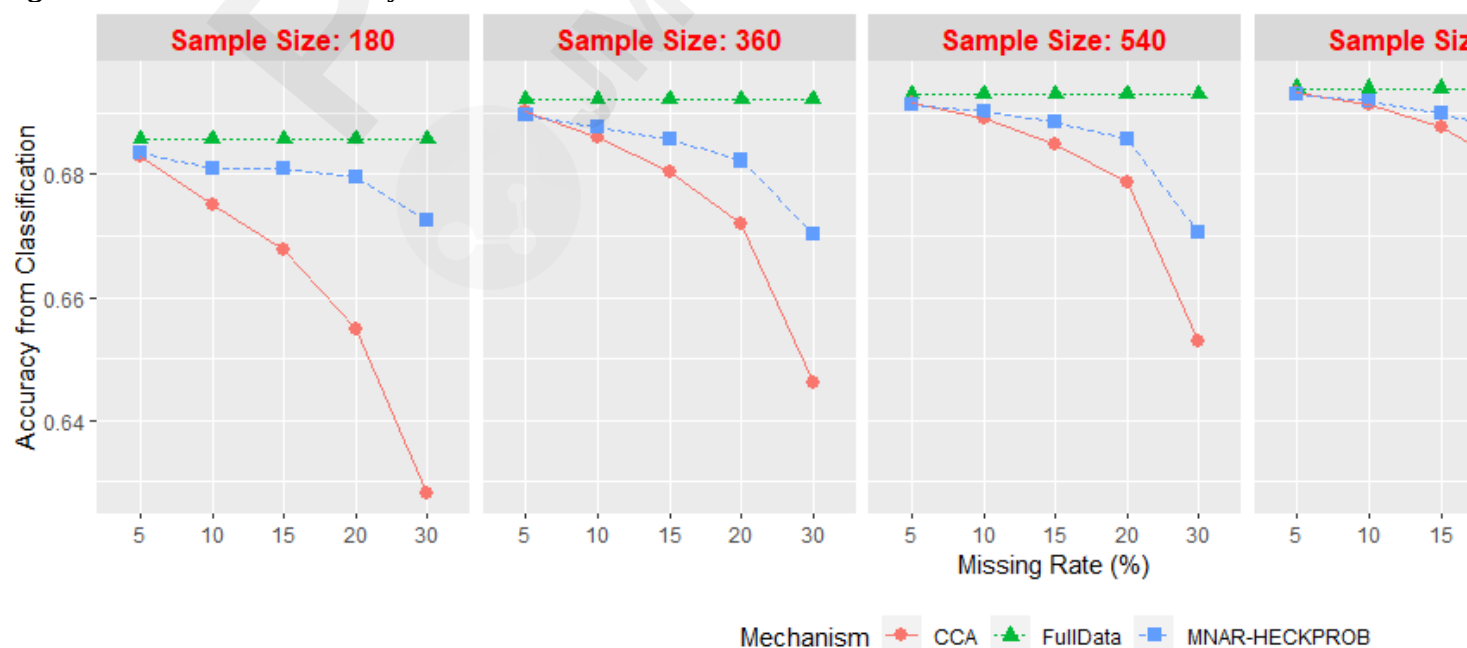


Figure 5 – Prediction Accuracy



Preprint
JMIR Publications

Supplementary Files

Multimedia Appendixes

Supporting Information.

URL: <http://asset.jmir.pub/assets/4f70d05d100ea053a1e020d37913f5a1.docx>

Manuscript.

URL: <http://asset.jmir.pub/assets/403673899b0f495dba9cb12cd87ef1a2.docx>