Multiple imputation strategies for incomplete data in clustered longitudinal studies

Panteha Hayati Rezvan, Ph.D., phayatirezvan@mednet.ucla.edu
Thomas R. Belin, Ph.D., tbelin@mednet.ucla.edu

Overview

- **Objectives**: assess the robustness of conclusions to proper handling of missing data
- **Methods**: contrast widely-used multiple imputation strategies
- **Results**: the findings from this research can provide *methodological* insights with enormously broad potential application to behavioural and biomedical research settings
- **Future work:** compare MI strategies with a recently developed multi-level imputation technique

Objectives

Missing data: General considerations

- Missing data are frequently encountered in HIV-related research
- Consequences of missing data
  - Can introduce bias
  - Loss of precision/power

Missing data: Modeling and Analysis Process

- Multiple imputation (MI) is a well-recognized technique for handling missing data
- Implementation of MI in standard statistical software typically assumes that data are 'Missing at random' (MAR)

Step 1 - Imputation

- Replace missing values with multiple plausible values

Step 2 - Analysis

- Analyze each imputed dataset separately using statistical methods applicable to the complete data
- Combine results using statistical methods applicable to the complete data, reflecting both between – and within-imputation variability in estimated quantities

Methods

Imputation strategies

- Multivariate Normal Imputation (MVNI)
- Fully Conditional-Specification (FCS)
  - Sequential regression models
  - Variable by variable
    - Linear regression for continuous
    - Logistic regression for binary

Recommendations for building imputation models

- Include all variables in the analysis model
- Include auxiliary variables
  - Predictors of incomplete variables
  - Predictors of missingness

Case study - Background

- Philani study examines a home-visiting prevention program delivered by neighbourhood mentor mothers
- Targeted mothers at risk for hazards alcohol use, HIV, TB and malnutrition as well as their children
- CRCT – 1238 pregnant women recruited
  - Standard Care (SC) - n=594
  - Philani Intervention Program (PIP) - n= 644

Case study - Research question

- Assess the effect of the PIP on children's outcomes through the first five years of life.

Analysis model	Linear mixed-effects model
Outcome	Growth measure (waves 2-6) • Height-for-age z-score (HAZ)
Analysis variables	HIV status, Neighborhood, Time-point, Interaction between time-point & neighborhood
Auxiliary variables	Edinburgh Postnatal Depression Score, Any alcohol use, Married/lived with partner, Income above 2000 RAND, Food insecurity, Any violence, Maternal age*, Education*, Formal housing*

*No missing data

Number (%) of missing values		
Assessments	HIV status	HAZ
Pregnancy	156 (14.50)	-
Birth	159 (14.78)	160 (14.87)
6 month	136 (12.64)	177 (16.45)
18 month	123 (11.43)	259 (24.07)
36 month	226 (21.00)	284 (26.39)
60 month	200 (18.59)	244 (22.68)

- 606/1076 (56.32) incomplete cases
- % of missing data in auxiliaries
  - Min (8.18), Max (20.82), ρ<0.3

Case study - Analysis plan

- Longitudinal Mixed-Effects Models (LMM), include both fixed and random effects
- MI strategies with
  - I. No auxiliaries
  - II. 35 auxiliaries (inclusive strategy)
  - III. 11 baseline auxiliaries
  - IV. 15 baseline (max 10% missing) and post-baseline auxiliaries (max 3% missing)

Results

Imputation model failure

- Inclusive strategy did not converge initially
  - Collinearity among predictors

Strategies for handling lack of convergence

- Collapse HIV status
- Exclude "marital status" at wave 3
- Use augmentation when using MICE
- After modification, inclusion strategy converged
- Inclusive strategy - slightly higher point estimates/SE & wider CI

Figure 1: Regression coefficient for HIV status

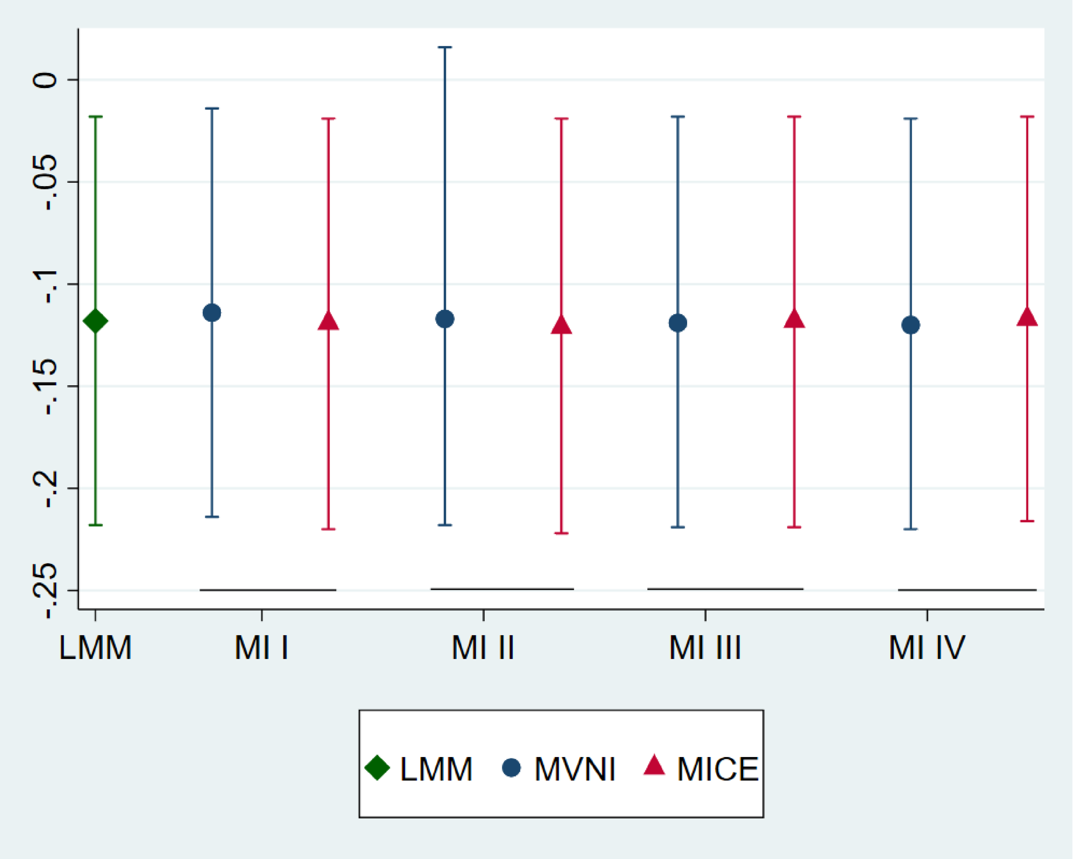
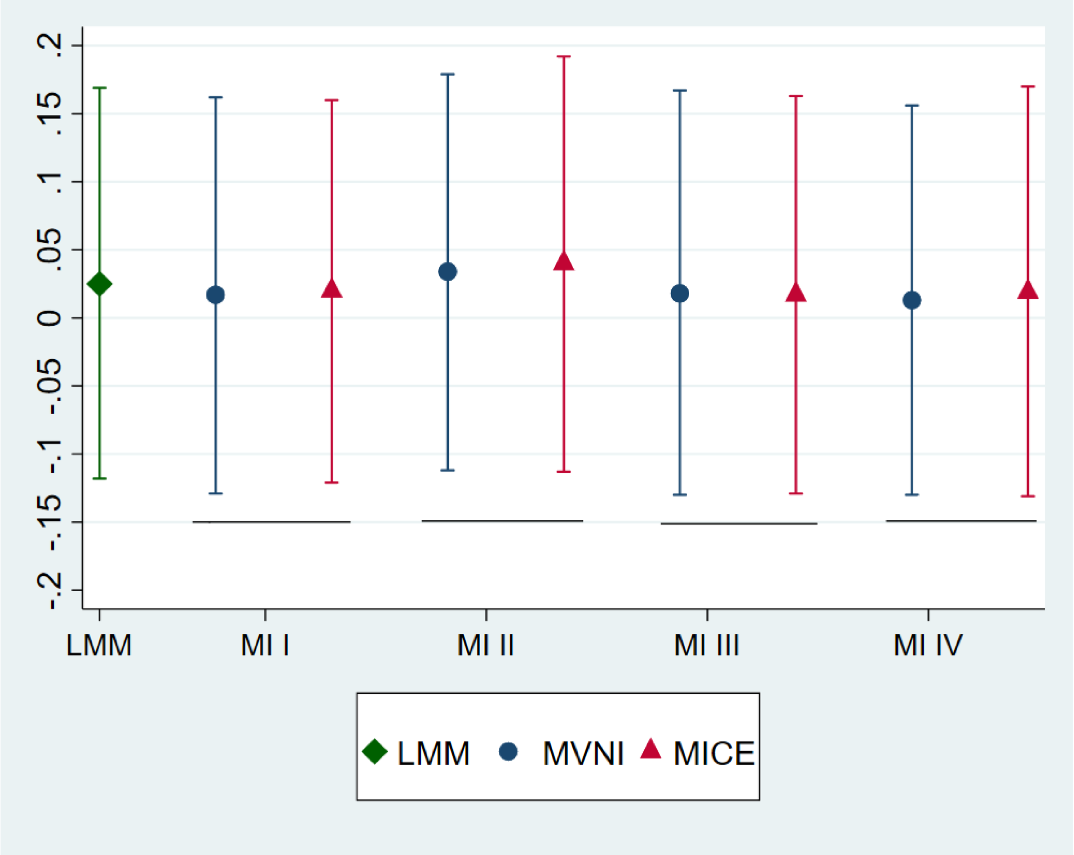


Figure 2: Intervention effect estimate



- Interpretation: Findings robust to choice of imputation strategy

Future work

- Extension to covariates with missing data
- Comparison of the results with the Longitudinal Factor Imputation Method, a recently developed multi-level MI technique
  - Linear mixed model for longitudinal associations
  - Factor-analysis strategy for cross-sectional associations to keep the number of model parameters manageable

Acknowledgements We would like to thank Dr. Mary Jane Rotheram for giving us the opportunity to use the Philani dataset. This work is supported by the T32 Postdoctoral Training Program for HIV Combination Prevention (T32MH109205).