What are Best Practices in Design and Analysis for a Longitudinal Study with Planned Missing Data?

Laura Wray-Lake[1], Wendy M. Rote[1], Aaron Metzger[2], Amy K. Syvertsen[3]
[1] University of Rochester, [2] West Virginia University, [3] Search Institute

ROOTS of ENGAGED CITIZENSHIP PROJECT

Project website: civicroots.org

Introduction

- Multiform planned missing data designs represent a relatively new way of thinking about data collection for developmental research (Enders, 2010; Graham, 2012; Little & Rhemtulla, 2013).
- In survey research with children and adolescents, planned missing data designs are beneficial for:
  - Maximizing the number of questions asked in a fixed time frame.
  - Reducing cognitive demands on participants by reducing survey length, while still measuring all constructs of interest.
  - Increasing the likelihood of survey completion (due to shorter survey length) and thereby minimizing less desirable forms of missing data.

Our Design

- For these reasons, we incorporated a **three-form planned missing data design** into a large multi-site cross-sectional study of elementary, middle, and high school youth.
- We anticipate these data being the first occasion of measurement in a longitudinal study.
- There are few concrete recommendations for effective implementation of longitudinal planned missing designs in published work.
- Thus, our central question for conversation is:
  What are the best practices in design and analysis for a longitudinal study with planned missing data?

Study Content

- The overarching study goal is to understand the developmental roots of civic engagement across childhood and adolescence.
- Constructs were measured in five areas: Civic Engagement, Competencies, Character, Contexts, and Demographics.
- Surveys completed during 45-minute class periods in schools.

Sample Description

- 2505 children and adolescents (*M*age = 13.3 years; Range: 8-20; 55% female)
  - 4th through 12th grade students in 16 schools across three distinct geographic regions.
  - Diverse geographic, socioeconomic, and racial-ethnic background.

Sample demographics by site (%)

Site	N	1st or 2nd Gen. Immigrant	Ethnic Minority	Parent Education HS or below	Parent Education College degree+	Financial Strain	Free/ Reduced Cost Lunch (school-level)
CA	1042	72.2	89.9	18.4	25.7	12.6	26-95
MN	633	49.5	44.4	7.1	45.8	7.8	33-77
WV	838	8.4	6.8	20.4	34.5	6.5	37-55

Note. Financial Strain = % of youth who reported their family "has a hard time buying the things they need."

Planning our Missing Data Design

STEP 1: Choosing Item Sets
- We divided survey items into one core set (X) and three additional item sets (A, B, C).
- CORE SET ("X") → Main dependent variables (civic engagement), demographics, and other central predictors as room allowed.
- A, B, C SETS → Distributed constructs across item sets so that each set would have a mix of competencies, character, and context variables.
  - We tried to group variables together that we had hypotheses about to maximize power to test these hypotheses.
- *After much debate, we decided to keep items in a scale together in the same item set*, but we welcome opinions on this issue going forward.

STEP 2: Creating Three Forms
- We created three survey forms, each of which included the core items (X) and two out of three other items sets.
- Item set X was always included first, as these items were most central to our study goals. Item sets A, B, and C were counterbalanced across versions.

Survey Form	Item Sets
1	X + A + B
2	X + C + A
3	X + B + C

STEP 3: Creating Age-Specific Versions
- Given our large age range (8-20), we needed developmentally sensitive surveys.
- We created age-appropriate versions of the survey forms for elementary, middle, and high school age groups.
- The result was 9 survey versions in total.
- To adjust length across ages, more complex constructs as well as constructs less central to our study were included at older ages only.
- Item wording was the same across ages.

Analysis Plan

- We plan to analyze data primarily in M*plus* version 7.2.
- Missing data will be handled by FIML estimation using principle components as auxiliary variables (Howard, Rhemtulla, & Little, in press; Little et al., 2008).
  - Recommended as an efficient strategy to include information from all variables into the missing data model.

Age-Specific Design Details

Number of items from each scale in item sets (age-specific)

Scale	X (Core)	A	B	C
		Item Set		
Social Responsibility Values and Beliefs	7EMH		5MH	4MH
Informal Helping and Volunteering	7EMH			
Extracurricular Activities				6EMH
Environmental Behaviors		3EMH		
Voting	1MH		1E	
Political Behaviors	2MH	4MH		
News/Political Information Consumption				1EMH
Civic Participation Skills			6EMH	
Critical Consciousness				3EMH
Critical Information Analysis		3MH		
Core Concepts of Democracy		4MH		
Empathy	4EMH			
Emotional Competencies			5E, 9MH	
Prosocial Moral Reasoning			4E, 6MH	
Perspective Taking				4MH
Optimism		3EMH		
Parental Support		6EM		
School Support			6EMH	
Peer Support				6MH
Parental Modeling			4E, 6MH	
Peer Modeling		3MH		
Parental Communication				3EMH
School Communication		3EMH		
Peer Communication			3MH	
Quality of Extracurricular Contexts	4E, 6MH			
School Climate				4EMH
Neighborhood Climate		4MH		
Obstacles to Civic Participation			12H	
Discrimination				8H
Personal Responsibility			3EMH	
Perseverance	3MH			3E
Thrifty		3EMH		
Future-Minded				3EMH
Teamwork	3MH	3E		
Humility		3MH		
Leadership			3EMH	
Respect				3EMH
Spirituality			2EMH	
Gratitude		3EMH		
Purpose	3H			3EM
Demographics	16E, 17MH			
ITEM TOTALS	**38E, 50M, 59H**	**29E, 51M, 58H**	**29E, 52M, 56H**	**29E, 48M, 52H**

Note. E = Elementary school, M = Middle school, H = High school

Design Descriptives

- Survey versions were evenly distributed across students:
  - Version 1 *N* = 858, Version 2 *N* = 823, Version 3 *N* = 833.
  - Non-significant one-sample chi square test, indicating equality of group *N*s.
- Versions evenly distributed across age, gender, ethnicity, parental education, and site.
  - All chi square tests *n.s.*

Questions for Discussion

DESIGN
- What are advantages and disadvantages of choosing to add, subtract, or move items across sets in future waves?
  - What would be the implications of distributing scale items across item sets in wave 2 (as opposed to keeping scale items together as we did in wave 1)?
  - Should participants get the same survey form in follow-up waves?

ANALYSIS
- What are the data analytic challenges to capturing intra-individual change over time and time-varying effects of change over time in a planned missing design?
  - Are any longitudinal analyses weakened by the planned missing design?
- What are the pros and cons of FIML versus multiple imputation in data analysis given our planned missing design?
  - To what extent should the analytic approach be consistent across analyses?
- We have encountered multicollinearity problems when conducting single imputation models and principle components analysis.
  - Is it possible that planned missing data causes multicollinearity issues?
  - Can a PCA solution be trusted when the matrix is not positive definite?
  - What are the lowest possible covariance coverage values that would be acceptable to work with?

DEVELOPMENTAL MEASUREMENT
- Is it possible to create measures that are similar across time (for longitudinal analyses) but that are also developmentally sensitive and appropriate for widely different ages?

Contact Information

Laura Wray-Lake: laura.wray-lake@rochester.edu
Wendy Rote: wendy.rote@rochester.edu
Aaron Metzger: aaron.metzger@mail.wvu.edu
Amy Syvertsen: amys@search-institute.org

References

Enders, C. K. (2010). *Applied missing data analysis.* New York: Guilford Press.
Graham, J. W. (2012). *Missing data: Analysis and design.* New York: Springer.
Howard, W., Rhemtulla, M., & Little, T. D. (in press). Using principal components as auxiliary variables in missing data estimation. *Multivariate Behavioral Research.*
Little, T. D., McConnell, E. K., Howard, W. J., & Stump, K. N. (2008). *Missing data in large data projects.* KUant Guide No. KUANT 011.3. Available at www.quant.ku.edu.
Little, T. D., & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives, 7,* 199-204. doi: 10.1111/cdep.12043