PENNSTATE

How Much Do Planned Missing Designs Increase Survey Error in Longitudinal Panel Studies?

Veronica Roth[a]

David R. Johnson[a]

Rebekah Young[b]

[a]Department of Sociology & Demography

The Pennsylvania State University

[b]Department of Biostatistics

University of Washington

This research was partially supported by NICHD Grant R01-HD42241 (David Johnson, PI)

POPULATION RESEARCH INSTITUTE

PENNSTATE

Issues in Longitudinal Studies

- In longitudinal studies, there is a need to keep questions as similar as possible to reduce error (Fowler, 1995; Allen & Yen, 2002)
- Researchers sometimes have to make concessions over long periods of time, where concepts can change (Olsen, 2005)
- The addition or subtraction of questions may be necessary, but complicates designs

PENNSTATE

Issues of Respondent Burden

- There is also a growing awareness of a need to decrease respondent burden (Graham et al., 2001; Enders, 2010)
- Possible complications of item non-response if survey is too long
- Also, there is the possibility of attrition if respondent had a poor experience with survey

PENNSTATE

Planned Missing (PM) Designs

- Planned missing (PM) designs are becoming increasingly popular
  - Currently used in the General Social Survey, National Survey of Family Growth, National Assessment of Educational Progress, (need European surveys!)
- Designs can help decrease respondent burden (Graham et al., 2001; Enders, 2010)
- Haven't heard of a PM design before?
  - Perhaps you're familiar with the terms "split-ballot," "matrix sampling" or "intended skip"

PENNSTATE

Conventional 3-form design

	PM item sets			
Form	X	A	B	C
0	✓	✓	✓	✓
1	✓	--	✓	✓
2	✓	✓	--	✓
3	✓	✓	✓	--

PENNSTATE

Advantages of a PM design

- Reduces the length of the survey for the participants without reducing the total number of items of information assessed
- Can yield scores on scales that retain full reliability and comparability to studies using the complete measures, even when all scale items are not asked of each respondent (Graham et al., 2006; Enders, 2010)

PENNSTATE

Disadvantages of a PM design

- Statistical power will be lost for some comparisons (Graham et al., 2006; Enders, 2010)
- 3-form design is limited in testing higher order interactions (Graham et al., 2006; Schafer et al., 2002)
- Necessary to use statistical methods that permit analysis of incomplete data matrices
  - Imputation
  - Maximum likelihood

PENNSTATE

How should we implement PM designs in panel studies?

- Simulations show that standard errors are lowest when respondents are given full sets of scales in the first and last waves, with PM designs in intervening waves (Graham et al., 2001)
- However, this approach may not be feasible if researchers need to ask many questions in a shorter time frame
- We will focus on how PM designs impact error estimates in panel surveys

PENNSTATE

National Survey of Fertility Barriers (NSFB)

- Nationally representative RDD telephone survey of women age 25-44 and their partners
  - We are focusing on the 3,723 women who were included in the follow-up assessment
- The study examined factors predicting medical help-seeking for infertility and social-psychological outcomes of fertility barriers
- Wave 1 was fielded from 2004-2007 & wave 2 from 2007-2010

POPULATION RESEARCH INSTITUTE

PENNSTATE

PM Design in NSFB

- Reduce interview length and respondent burden without reducing the scales and measures needed to meet research aims
- Include PM design in 21 scales and sets of items to reduce interview length
  - 25 out of 92 scale questions were not asked

PENNSTATE

PM Design in NSFB- Wave 1

- Items in each scale divided into three sets
- Selected random numbers for each scale, each respondent receiving either a 0, 1, 2, or 3
  - Selections into PM design for each scale was independent
- For those with a random number of 1, the first set of items in the scale was dropped
- If 2, second set dropped, etc.
- Those with code 0 were asked all items in that scale

PENNSTATE

Example for the CESD scale

Item	Question Wording	PM Set 1	2	3
1	I was bothered by things that usually don't bother me		x	x
2	I had trouble keeping my mind on what I was doing		x	x
3	I felt depressed		x	x
4	I felt that everything I did was an effort	x		x
5	I felt hopeful about the future	x		x
6	I felt fearful	x		x
7	My sleep was restless	x	x	x
8	I was happy	x	x	
9	I felt lonely	x	x	
10	I could not get going	x	x	

POPULATION RESEARCH INSTITUTE

PENNSTATE

PM Design in NSFB-Wave 2

- Respondents in the follow-up received either the same set of items they previously had or were reassigned to the full set of items
- In Wave 1 13% were asked all items in a scale but in Wave 2 55% were asked all items.
- A small number of respondents during the beginning of the wave 2 interviewing received a different PM set from the set they received in wave 1. The number was too small to analyze so these were excluded from the sample for this study.

PENNSTATE

Previous Work on the NSFB

- An analysis of scales from wave 1 evidenced (Johnson et al., 2010):
  - Reliability of scales & covariance of scales with other variables were similar
  - Comparison of PM item sets showed that scale values were generally comparable

PENNSTATE

Focus of this study

- Three health-related scales from NSFB
  - 10-Item version of the CES-D Depression scale (Andresen et al., 1994)
  - 6-Item Medical Locus of Control scale (Wallston et al., 1978)
  - 6-Item Ethics of Assisted Reproductive Technology (ART) Scale (created for this study)

PENNSTATE

Focus of this study

- Does the use of PM in the scales affect the change in scales between waves 1 and 2?
- We first compare scale characteristics at Wave 2 under four conditions:
  - All items asked on the scale (No planned missing)
  - Available items used to estimate scale score
  - Imputation of PM items to create scale score
  - Calibration of imputed PM items

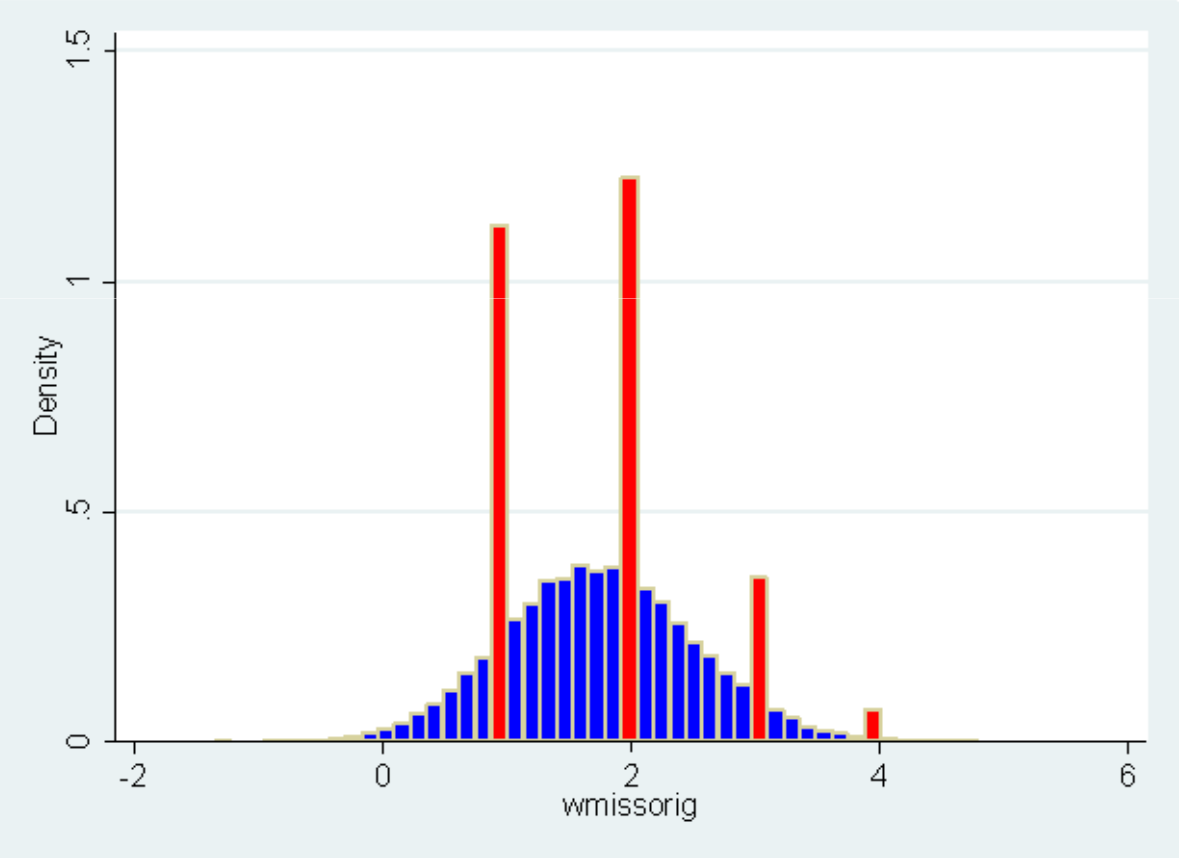
PENNSTATE

Imputation and Calibration Method

- We imputed using chained-equations and the normal model in Stata.
- We generated 25 imputed datasets.
- Imputations were informed by all items in the three scales plus a small number of auxiliary variables
- Imputed values were calibrated using the Yucel method (Yucel, He, & Zaslavsky 2008) to adjust the distributions of the imputed data to match the observed data. This procedure is useful if the distributions of scales is important (e.g. using cutoffs) but may slightly reduce covariances.
- We present some results from both calibrated and uncalibrated variables.

PENNSTATE

Calibration example: Distribution of observed values (red) and values imputed under the normal model (blue)



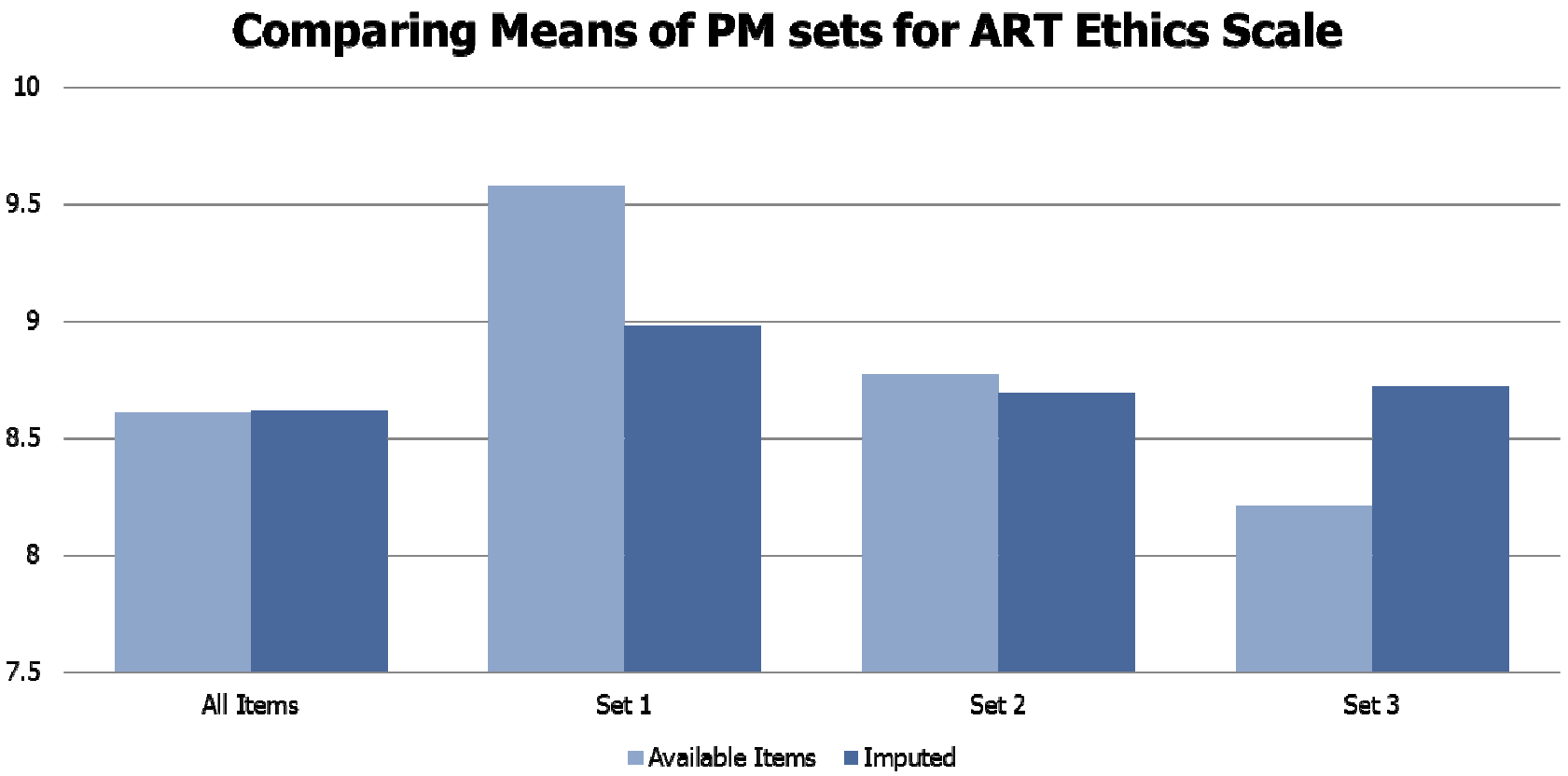
POPULATION RESEARCH INSTITUTE

PENNSTATE

The Available Item Method

- (Mean of available items) * Number of total items in scale
- Commonly used method but can yield a biased estimate of the scale score if there is large variation among items in their mean scores.

PENNSTATE



POPULATION RESEARCH INSTITUTE

PENNSTATE

Comparison of Means, Standard Deviations and Alpha reliability of Wave 2 scale scores under four conditions.

	Mean	SD	Alpha
CESD Depression Scale			
All Items (No PM)	16.47	4.58	0.81
Available Items	16.46	4.66	
Imputed and Calibrated	16.43	4.52	0.79
Imputed	16.43	4.54	0.80
Medical Locus of Control Scale			
All Items (No PM)	16.91	2.49	0.79
Available Items	17.07	2.70	
Imputed and Calibrated	17.02	2.53	0.78
Imputed	17.04	2.58	0.80
ART Ethics Scale			
All Items (No PM)	8.61	3.00	0.90
Available Items	8.71	3.02	
Imputed and Calibrated	8.70	2.95	0.89
Imputed	8.72	2.97	0.87

PENNSTATE

Correlation and Regression of Scale over both waves for different methods of handling PM items.

	Correlation	b	SE
CESD Depression Scale			
All Items (No PM)	0.54	0.539	0.026
Available Items	0.49	0.482	0.029
Imputed and Calibrated	0.51	0.508	0.032
Imputed	0.51	0.506	0.031
Medical Locus of Control Scale			
All Items (No PM)	0.47	0.434	0.026
Available Items	0.53	0.520	0.028
Imputed and Calibrated	0.50	0.495	0.032
Imputed	0.52	0.513	0.032
ART Ethics Scale			
All Items (No PM)	0.64	0.637	0.024
Available Items	0.60	0.578	0.026
Imputed and Calibrated	0.62	0.602	0.027
Imputed	0.63	0.606	0.027

PENNSTATE

Test of Differences in Slopes

- Conducted multiple regression analyses of the wave 2 scale on the wave 1 scale.
  - With available item mean scales
  - With imputed and calibrated scales
- Defined 5 conditions:
  - All items both waves
  - PM set 1 both waves
  - PM set 2 both waves
  - PM set 3 both waves
  - All items wave 2 – any PM set wave 1
- Created an interaction term between the scale at wave 1 and the condition to test for differences in slopes.
- For the three scales and there was no significant difference in the slopes or intercepts among the 5 conditions.

PENNSTATE

Conclusions

- Planned Missing designs applied to multiple scale items appear to be a viable method of reducing respondent burden and does so with minimal error in longitudinal analysis of the scales.
- Imputing the deleted items before creating the scale appears to be better than using the mean of available items to estimate the scale score.
- Calibrating the distributions of the imputed values improves the fit between the full scale distributions and the PM scales.
  - Using appropriate categorical data prediction equation seems to yield similar results (Johnson & Young, 2011)

PENNSTATE

Limitations

- Random assignment of PM sets between the waves would have provided a better test of the longitudinal stability of using PM compared to asking all items.
- Using PM with scale items is likely to perform better than PM sets among survey items that are not as highly correlated. Conclusions may not apply as well as more error will be introduced when less predictive survey items are used to inform the imputation.

PENNSTATE

References

Andresen E.M., J.A. Malmgren, W.B. Carter, & D.L. Patrick (1994). Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). *American Journal of Preventative Medicine, 10*(2) 77–84.

Enders, C. K. (2010). *Applied Missing Data Analysis.* New York: Guilford Press.

Graham, J.W. and B.J. Taylor, A.E. Olchowski, & P.E. Cumsille (2006). Planned Missing Data Designs In Psychological Research. *Psychological Methods, 11*(4), 323-343.

Horton, N. J., Lipsitz, S. R., & Parzen, M. (2003). A Potential for Bias When Rounding in Multiple Imputation. *The American Statistician, 57*(4), 229-233.

Johnson, D.R. & R. Young (2011). Toward best practices in analyzing datasets with missing data: Comparisons and recommendations*. Journal of Marriage and Family* (August).

Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data, Second Edition*. New Jersey: John Wiley and Sons.

Schafer, J. L., & Graham, J. W. (2002). Missing Data: Our View of the State of the Art. *Psychological Methods, 7*(2), 147-177.

Wallston, K. A., Wallston, B.S., and Devellis, R. (1978). Development of the multidimensional health locus of control (MHLC) scales. *Health Education Monographs, 6*(1), 160–170.

Yucel, R. M., He, Y., & Zaslavsky, A. M. (2008). Using Calibration to Improve Rounding in Imputation. *The American Statistician, 62*(2), 125-129.

PENNSTATE

Thank You!

Author contact information

Veronica Roth (vlr124@psu.edu)

David R. Johnson (drj10@psu.edu)

POPULATION RESEARCH INSTITUTE

PENNSTATE

National Survey of Fertility Barriers (NSFB)

http://sodapop.pop.psu.edu/data-collections/nsfb

- **PROPER CITATION:** Johnson, David R. and White, Lynn K. National Survey of Fertilty Barriers [Computer File]. Population Research Institute [distributor]. The Pennsylvania State University. University Park, PA.

- Researchers who wish to download the dataset will be requested to leave their names, institutional affiliation, and contact information.

POPULATION RESEARCH INSTITUTE