Informing Follow-up Strategies to Reduce Nonresponse Bias

CNSTAT Workshop on Improving Consent and Response in Longitudinal Studies of Aging September 27-28, 2021

Andy Peytchev





Acknowledgements

 Part of this work was supported by a grant from the National Institutes of Health (Grant R21 HD063070-01A1)

Outline

- $_{\odot}\,$ Similarities to cross-sectional surveys
- Dissimilarities to cross-sectional surveys
- $_{\circ}$ Areas with greatest potential
- $_{\circ}\,$ Case studies
- A few concluding remarks

Nonresponse Bias in a Longitudinal Study Setting: Similarities to Cross-Sectional Surveys

- $_{\odot}$ Most of the nonresponse usually occurs in the first wave of data collection
- In terms of representation, the first wave of data collection has the greatest potential to induce nonresponse bias
 - When the nonresponse rate is high, relatively small differences between respondents and nonrespondents can result in bias of substantive importance
- Methods to measure and reduce nonresponse bias often share the same constraints as cross-sectional studies
 - Lack of auxiliary information at the sampling unit level (in household sample surveys in the U.S.)

Nonresponse Bias in a Longitudinal Study Setting: Similarities to Cross-Sectional Surveys

- $_{\odot}$ Most of the nonresponse usually occurs in the first wave of data collection
- In terms of representation, the first wave of data collection has the greatest potential to induce nonresponse bias
 - When the nonresponse rate is high, relatively small differences between respondents and nonrespondents can result in bias of substantive importance
- Methods to measure and reduce nonresponse bias often share the same constraints as cross-sectional studies
 - Lack of auxiliary information at the sampling unit level (in household sample surveys in the U.S.)
- It is in this context that we can to dissimilarities as having received less attention

Nonresponse Bias in a Longitudinal Study Setting: Dissimilarities to cross-sectional surveys

A wealth of <u>substantive</u> information on baseline respondents

- 1. To measure nonresponse bias after the initial wave
- 2. To inform data collection efforts after the initial wave
- 3. To inform wave nonresponse adjustments

Nonresponse Bias in a Longitudinal Study Setting: Dissimilarities to cross-sectional surveys

A wealth of substantive information on baseline respondents

- 1. To measure nonresponse bias after the initial wave, *for point estimates that are stable over time*
 - For most longitudinal surveys, these are not the key estimates
 - Estimates of *change* are of essential interest
- 2. To inform data collection efforts after the initial wave
 - **Paradata** to tailor methods based on prior behaviors (e.g., Peytcheva, Kirchner, and Cooney, 2018)
 - **Substantive data** to build models for influence on nonresponse bias ("Anticipatory Survey Design" in Peytchev et al., 2012; "Bias Propensity" in Peytchev, Pratt, and Duprey, 2020)
- 3. To inform wave nonresponse adjustments
 - Use of substantive variables in poststratification weighting adjustments (Peytchev and Presser, 2018)

Case Study 1: Using Prior Wave Paradata

- 2016/17 Baccalaureate and Beyond Longitudinal Study (B&B) field test
 - Survey of individuals who completed their bachelor's degree during the 2014/15 academic year
 - First follow-up for the National Postsecondary Student Aid Study (NPSAS:16)
 - Mixed-mode data collection (web and telephone)
 - 1,130 completes
 - Data collection: July 2016 November 2016
 - 30 min survey

Case Study 1: Treatments

	Relaxed (Reduced)	Default	Aggressive
Early Completion Phase			 Prepaid incentive CATI interviewing (week 2)
Production Phase	• No CATI contact	CATI interviewing (week 5)	 Offer abbreviated interview (week 4)
Nonresponse Conversion Phase	No abbreviated interview	 Offer abbreviated interview 	
Incentives	• \$20 completion	• \$30 promised	 \$10 prepaid \$20 promised

Case Study 1: Groups and Treatments

- Four groups defined by response behavior in NPSAS (prior data collection)
 - Group 1: Early respondents
 - Relaxed (reduced effort and resources) protocol
 - Group 2: Late respondents
 - Default protocol
 - Group 3a and 3b (random assignment): Nonrespondents
 - 3a: Default protocol
 - 3b: Aggressive protocol

Case Study 1: Response Rates



Source: Peytcheva, Kirchner, and Cooney, 2018

Case Study 1: Average Absolute Relative Bias



Source: Peytcheva, Kirchner, and Cooney, 2018

Case Study 2: Using Prior Wave Substantive Variables

Motivation

- Our prior study targeted low propensity cases through an interviewer intervention (Peytchev, Riley, Rosen, Murphy, and Lindblad, 2010)
 - To increase response rates (denominator, below)
 - To decrease the covariance between response propensity and the survey variables of interest (numerator, below)

$$Bias(\overline{y}_r) \approx \frac{\sigma_{y,\rho}}{\overline{\rho}}$$

- ∘ It was not effective
- $_{\circ}\,$ We need a more direct targeting of nonresponse bias

Case Study 2: Using Prior Wave Substantive Variables

o Data

- Wave 8 (2011) of the Community Advantage Panel Survey (CAPS)
- Two samples, home owners and renters
- All interviews conducted by telephone in this wave
- Design overview
 - Start data collection
 - Estimate prioritization models
 - Models for y
 - Models for ρ
 - After a certain period, implement higher respondent incentives for prioritized cases

Case Study 2: Prioritization Models

- $_{\circ}$ Estimate predicted response propensities for all cases, ρ
 - Frame information
 - Paradata during current data collection
 - Prior waves of data collection: Demographic, Substantive variables (i.e., y_{t-1}), Paradata (not being interested, ever hung up), Other known strong predictors (voting)
- $_{\circ}\,$ Conduct factor analysis using the key survey variables, \pmb{y}
- Regress the main factor on the predicted response propensity
- $_{\odot}\,$ Create quintiles based on the predicted factor scores
- Within the quintiles with the lowest mean response propensity, randomly assign cases to treatment

Case Study 2: Response Rates among Nonexperimental Cases



 Models were predictive of the likelihood of response for groups with different Ys.

 Response rates were more different in the owners sample.

Source: Peytchev, Riley, Rosen, Murphy, and Lindblad, 2012

Case Study 2: Response Rates among All Cases



Source: Peytchev, Riley, Rosen, Murphy, and Lindblad, 2012

Case Study 2

- The modeling of response propensities performed well in predicting low propensity groups
- The intervention was more effective (respondent instead of interviewer incentives)
- Yet, still failed at significantly affecting nonresponse bias
 - Models aimed at bias in the prior wave
- $_{\odot}\,$ Case Study 3: Bias propensity, in the session tomorrow on innovation

Concluding Remarks

- Increasing nonresponse calls for more complex study designs, and longitudinal studies offer opportunities for better-informed data collection designs
 - Informed by prior survey behaviors
 - Informed by estimated bias
- A key challenge remains the ability to identify covariates associated with survey variables of interest
 - <u>Covariates of estimates of change</u>