

Nonresponse Bias Studies: Disconnects Between Literature and Practice

Federal Committee on Statistical Methodology
March 8, 2018

Stephanie Coffey, Jim Dahlhamer,
Peter Miller, Joe Schafer, Jeff Gonzalez

Roadmap

- Nonresponse bias studies (NRB Studies) in Federal Surveys
 - Are not widely publicized
 - Findings often lack clear application
- Literature and practice are not integrated
 - Typology based on advances in the literature
 - What is commonly used in some actual work
- What are the disconnects?
 - Institutional
 - Technical
 - Operational
- Suggestions for improving NRB Studies in practice

Indicators for Nonresponse Bias

- Many indicators attempt to measure the risk of nonresponse bias
 - Subgroup response rates, R-indicators, Fraction of Missing Information, Coefficients of Variation, etc.
- Several classifications in the literature
 - Groves et al. (2008)
 - Wagner (2012)
 - OMB Standards and Guidelines for Statistical Surveys (2006)
 - FCSM Working Paper #31 (2001)
 - Brick and Groves JPSM Short Course Material (11/5/2005)

Typology of Indicators

- Groves et al. (2008)
 - Classified by Detail of Estimate
 - Survey-Level vs. Estimate-Level Indicators
- Wagner (2012)
 - Classified by Data Requirements
 - Additive Requirements
 - Response Indicator Only
 - Response Indicator + Frame Data + Paradata
 - Response Indicator + Frame Data + Paradata + Survey Data (the \hat{y} 's)
- Combine these frameworks in a single typology
- Identify which indicators are used where we work

Studies Examined

Survey Name	Sponsor
National Crime Victimization Survey	BJS
American Housing Survey	HUD
National Health Interview Survey	NCHS
National Health and Nutrition Examination Survey	NCHS
National Survey of Family Growth	NCHS
National Survey of College Graduates	NCSES (NSF)
National Compensation Survey	BLS
Survey of Occupational Injuries and Illness	BLS
Monthly Wholesale Trade Survey	Census
Occupational Requirements Survey	SSA

- Nonprobability sample of surveys at our agencies
- Surveys have varied characteristics
 - 6 Demographic Surveys
 - 4 collect household info
 - 3 collect person-level information
 - 4 Establishment Surveys
 - Mix of Longitudinal and X-Sectional
 - 1st Stage RRs Range 65% to 90%
 - 2nd Stage RRs Range 54% to 84%

Indicators by Joint Typology

Survey-Level Indicators

Response Indicator Only

- Overall Response Rate

Response Indicator, Frame Data, Paradata

- Subgroup Response Rates
- CVs of Response Propensities
- R-indicators
- Marginal Imbalance
- Dissimilarity Indicators
(Delta Statistic, Mahalanobis, etc.)
- Propensity Model Goodness of Fit
- CVs of Adjusted Weights
- R's vs. NR's Auxiliary Variables
- Weighting Sensitivity Analysis

Estimate-Level Indicators

Response Indicator, Frame Data, Paradata, Survey Response Data

- Means of \hat{y} 's by Decile of Weights
- Means of \hat{y} 's by Decile of Response Propensity
- Correlation between Auxiliary Variables or Weights and \hat{y} 's
- Estimation of Regression Parameters between Auxiliary Variables and \hat{y} 's
- Fraction of Missing Information
- Prediction of Frame Data Using \hat{y} 's to Identify Ill-Defined Regions
- CVs of Estimates by Phase
- External Data/Estimates Comparison

Indicators by Joint Typology

Survey-Level Indicators

Response Indicator Only

- Overall Response Rate

Response Indicator, Frame Data, Paradata

- Subgroup Response Rates
- CVs of Response Propensities
- R-indicators
- Marginal Imbalance
- Dissimilarity Indicators
(Delta Statistic, Mahalanobis, etc.)
- Propensity Model Goodness of Fit
- CVs of Adjusted Weights
- R's vs. NR's Auxiliary Variables
- Weighting Sensitivity Analysis

Estimate-Level Indicators

Response Indicator, Frame Data, Paradata, Survey Response Data

- Means of \hat{y} 's by Decile of Weights
- Means of \hat{y} 's by Decile of Response Propensity
- Correlation between Auxiliary Variables or Weights and \hat{y} 's
- Estimation of Regression Parameters between Auxiliary Variables and \hat{y} 's
- Fraction of Missing Information
- Prediction of Frame Data Using \hat{y} 's to Identify Ill-Defined Regions
- CVs of Estimates by Phase
- External Data/Estimates Comparison

Disconnects Between Literature and Practice

- Why don't NRB studies explore more of these indicators ... or why don't we know about ones that do?
- Anecdotally discussed three areas of disconnect
 - Institutional, in the Federal Statistical System
 - Technical, in Survey Organizations
 - Operational, in Survey Organizations
- Goal is to have NRB Studies that:
 - Enrich the understanding of nonresponse bias in general
 - Help inform data users about data quality
 - Inform improved practice of survey design and implementation

Institutional Disconnects

- Organizations often see NRB Studies as a formality
 - OMB guideline calls for one if unit RR < 80%
 - Surveys may not actively plan for a NRB Study
 - Metrics used vary; function of other planned analyses
 - There is no central repository of studies
 - Not widely disseminated, even within survey organizations
- Result can be a low-resource attempt to meet an OMB requirement

Institutional Disconnects

- Less focus on downstream data quality issues
 - Non-rigorous analysis could mislead data users
 - Focusing on means ignores other key estimands (e.g. measures of change)
 - Focusing on bias overlooks impacts of NR on efficiency and measures of uncertainty
 - Implications for scientific conclusions drawn from data (e.g. hypothesis tests)
 - Implications for decisions and policies based on those conclusions

Technical Disconnects

- Need infrastructure and resources
 - Data that are readily available are not sufficient to evaluate more indicators
 - Difficult for surveys to justify resources for detailed studies
 - Potential benefits vs institutional/operational roadblocks
 - Initial developmental/retrospective research
 - Same staff is often responsible for production
 - Production infrastructure requires even more resources
 - Some indicators rely on processing – challenging to prioritize
 - Organizations may be waiting for others to show it is “worth it”
- Lack of clarity around best practices with respect to indicators
 - How do they relate to nonresponse bias? What do they tell us?
 - How many should be used?

Technical Disconnects

- Lack of a true evaluation of the risk of NRB
 - Gold standards are very rare
 - If they exist, they are not always timely
 - Administrative records have barriers to use
 - Can be expensive and hard to work with
 - May not be timely
 - Do not exist for every purpose
 - Low topical overlap limits cross-survey comparisons
 - Difficult to benchmark or confirm estimates via other surveys
 - Even if we expend a lot of effort, we don't know if our findings are “correct”

Operational Disconnects

- If risk of nonresponse bias is found, what is the plan of action?
 - Make a change to the survey design?
 - Difficult for ongoing surveys
 - Make a change to the questionnaire?
 - Difficult for surveys that create time series
 - Make a change to post-collection adjustments?
 - Difficult for surveys with little auxiliary information
- Not clear how a survey organization should *respond* to finding a risk of nonresponse bias

Suggestions for Improvement

Address institutional and technical issues together

- More detailed guidance is needed
 - Information on different approaches and indicators
 - Data requirements, pros and cons of indicators
 - Suggestions on how to proceed if risk of NRB is discovered
 - Requirement needs to be more than just to check a box
- Survey organizations should be encouraged to regularly evaluate and monitor data collection
 - Plan NRB studies into schedule/budget
 - Work towards near-real-time processing, automation of indicator creation
 - Proactively store paradata and other data that may be useful for later evaluation

Suggestions for Improvement

- Collect and share knowledge about nonresponse bias:
Agencies should be encouraged to:
 - Conduct and share NRB Studies regardless of survey response rate
 - Use indicators appropriate for the survey
 - Communicate strengths and weaknesses of survey data
- Explore alternative data sources in combination with surveys
 - All potential data sources should be evaluated for data quality
 - May encourage data blending to improve estimates
 - Combine multiple surveys
 - Combine surveys + administrative data or big data

References

- Wagner, J. (2012). A Comparison of Alternative Indicators for the Risk of Nonresponse Bias. *Public Opinion Quarterly*, 76(3): 555-575.
- Groves, et al. (2008). Issues Facing the Field: Alternative Practical Measures of Representativeness of Survey Respondent Pools. *Survey Practice*, 1(3).
- Office of Management and Budget (2006). Standards and Guidelines for Statistical Surveys. fcsm.sites.usa.gov/policies
- Groves, R.; Brick, M. (2005). Practical Tools for Nonresponse Bias Studies. JPSM Short Course Material. <http://projects.isr.umich.edu/jpsm/materials/2005-1110.html>
- FCSM Statistical Policy Office (2001). Measuring and Reporting Sources of Error in Surveys. FCSM WP #31. fcsm.sites.usa.gov/reports/policy-wp

Contacts

- Census Bureau
 - Stephanie Coffey, stephanie.coffey@census.gov
 - Peter Miller, peter.miller@census.gov
 - Joe Schafer, joseph.l.schafer@census.gov
- National Center for Health Statistics
 - Jim Dahlhamer, fzd2@cdc.gov
- Bureau of Labor Statistics
 - Jeffrey Gonzalez, gonzalez.jeffrey@bls.gov