turning knowledge into practice

Nonresponse Bias Correction in Telephone Surveys Using Census Geocoding: An Evaluation of Error Properties

Paul P. Biemer and Andy Peytchev
RTI International

AAPOR Annual Conference
May 13, 2010

Acknowledgments

- This material is based upon work supported by the National Science Foundation under Grant No. SES-0818931.
- We also thank the Survey Research Center at the University of Michigan for their assistance in compiling the data.

Outline

- Census Geocoding
- Potential error in the CG approach
- Empirical evaluation
  - Overall error
  - Decomposition by source of error
- Conclusions and next steps in ongoing research

Correcting for Nonresponse in RDD

- Very limited information
  - Frame has only telephone number
  - Other identifiers can be matched to gain access to auxiliary data
- Practitioners use this general approach to attempt to correct for unit nonresponse
- Evaluations of such approaches compare estimates with and without the use of the auxiliary data as evidence for the ability to correct for nonresponse

Census Geocoding (CG) Approach

1. Obtain the addresses for as many RDD sample numbers as possible. This usually results in addresses for about 50% to 60% of the sample.
2. For numbers matched to an address, convert addresses to geographic coordinates (i.e., geocode).
   a. Link each sample member with the census block containing its coordinates.
   b. Substitute census block aggregate characteristics.
3. Sample numbers with no (geocodable) addresses can be handled in a number of ways. One method is to substitute with aggregate exchange area characteristics.

Potential Problems with the CG approach

- Correctly matching to auxiliary information
  - Using phone number
  - Using area code and prefix
- Measurement differences with the auxiliary data
- Aggregation

Potential Problems with the CG approach

- Bias: $$Bias(\tilde{y}_r) = E[p_{nr} * (\tilde{y}_{nr} - \bar{y}_{nr})]$$

$$RelBias(\tilde{y}_r) = \frac{Bias(\tilde{y}_r)}{\bar{y}_n}$$

- Decomposing error in the CG approach:

$$Bias(\tilde{y}_r) = \frac{n-r}{n}[p_{cp} * (\tilde{y}_{cp} - \bar{y}_{cp}) + p_{ip} * (\tilde{y}_{ip} - \bar{y}_{ip})$$

$$+ p_{ce} * (\tilde{y}_{ce} - \bar{y}_{ce}) + p_{ie} * (\tilde{y}_{ie} - \bar{y}_{ie})]$$

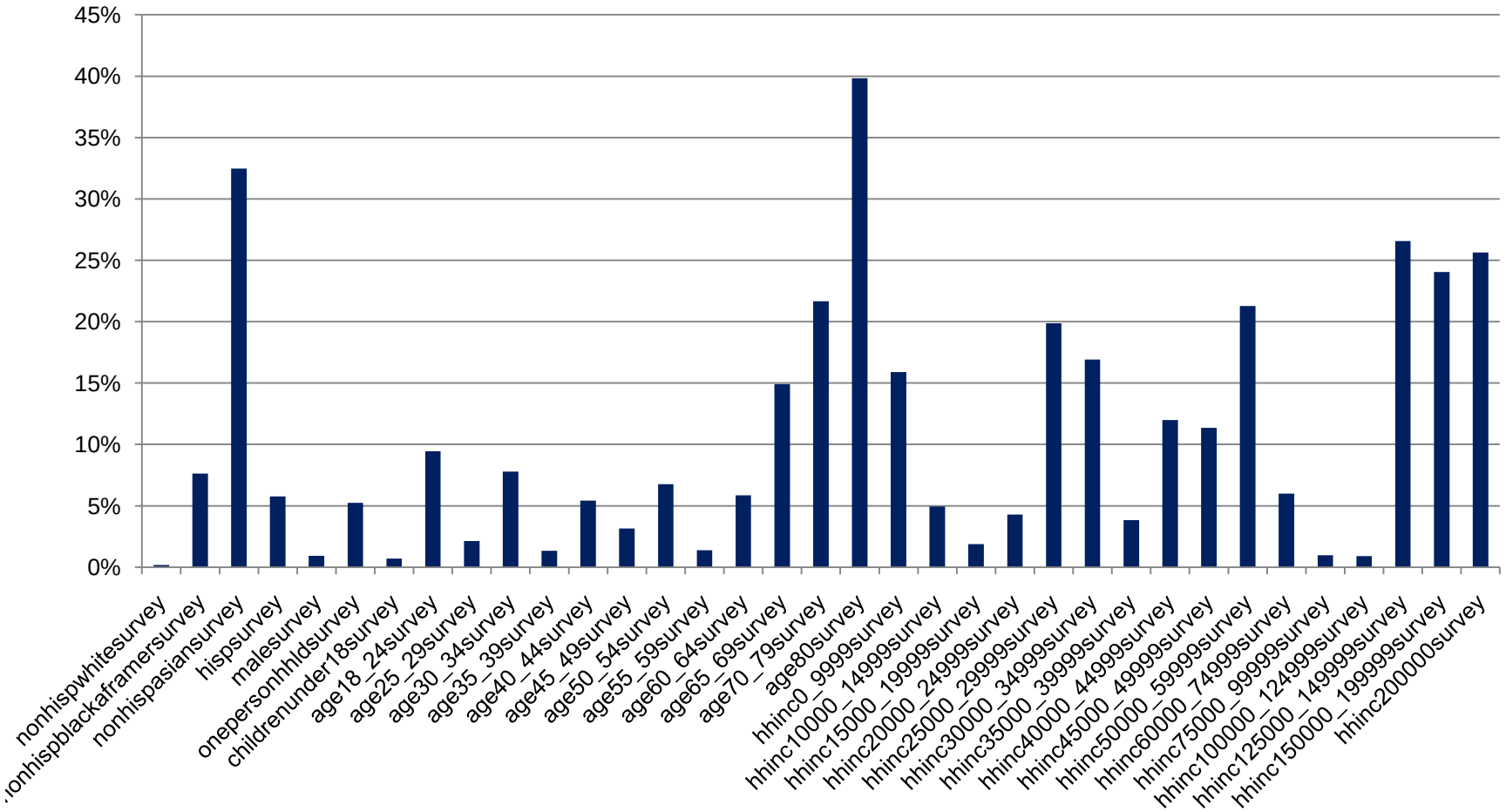
Estimation of the CG Error Components

- Start with a “gold standard” survey that provides responses for almost every sample member and has collected phone numbers
- Perform matching for all available telephone numbers, producing the four types of outcomes found in applying the CG to RDD studies
- Simulate nonresponse under different assumptions, such as number of call attempts or applying model estimated from an RDD survey
- Estimate bias in CG variables and its components

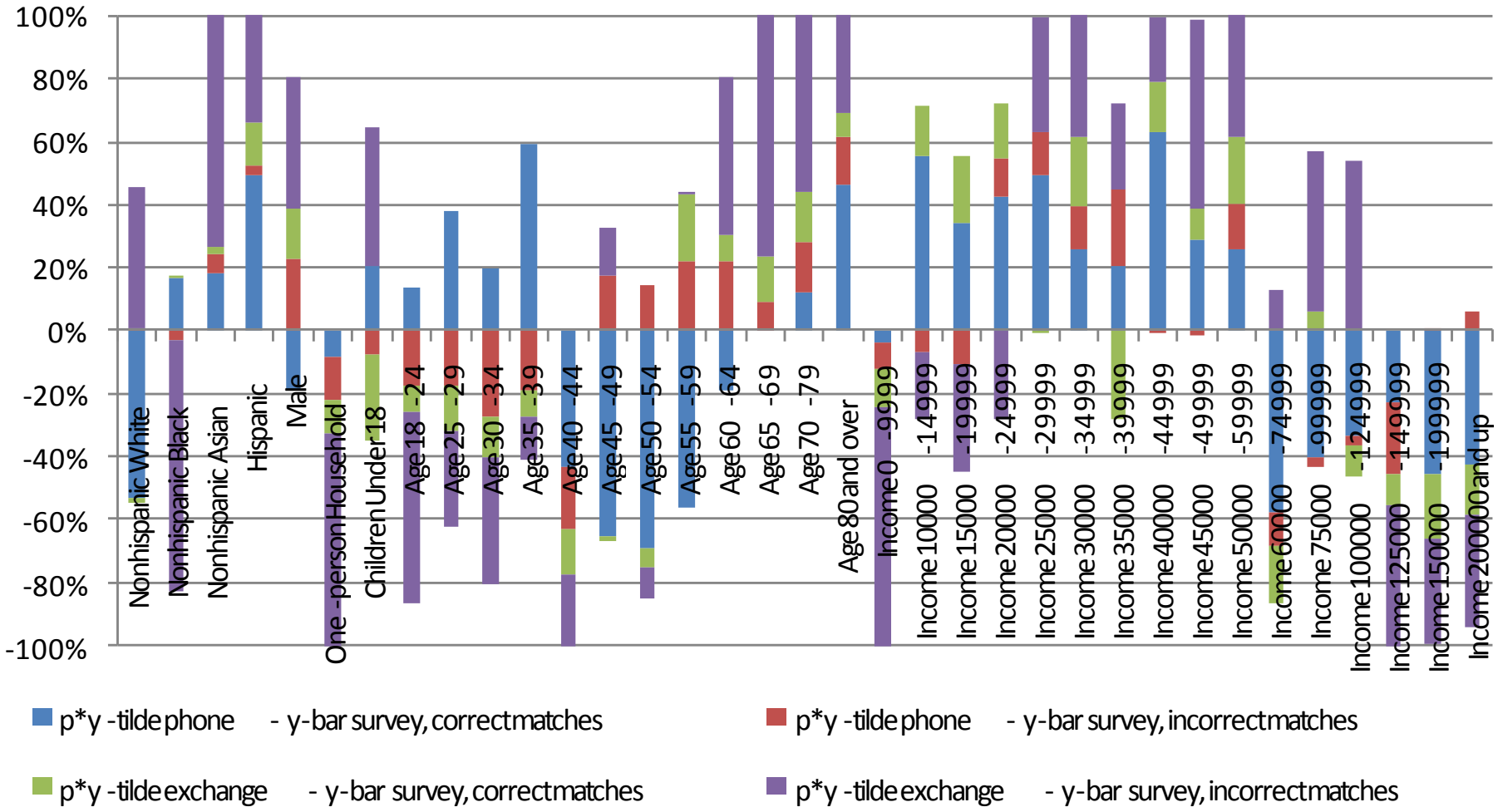
National Comorbidity Survey-Replication

- Area probability sample, face to face survey, n=9285
  - n=8178 who provided a phone number and responses to demographic questions
  - n=4987 who also provided income
- Conducted 2001-2003
- 85.9% response rate

Absolute Relative Bias



Proportion of Bias in CG Process by Source



Conclusions

- Substantial error in the census geocoding approach was found for some demographic estimates and it varied by source
  - Error from cases erroneously mismatched at the exchange level was a major source
  - A large component could also be attributed to cases correctly matched at the census blockgroup level
- Of critical importance is how the CG approach may still be informative of nonresponse bias in survey estimates and can be used to correct for it

Paul Biemer: ppb@rti.org
Andy Peytchev: apeytchev@rti.org

www.rti.org/aapor