



Practice of Epidemiology

Estimating Nonresponse Bias in a Telephone-based Health Surveillance Survey in New York City

Sungwoo Lim*, Stephen Immerwahr, Sunghye Lee, and Tiffany G. Harris

* Correspondence to Sungwoo Lim, Bureau of Epidemiology Services, Division of Epidemiology, New York City Department of Health and Mental Hygiene, Gotham Center, 42-09 28th Street, New York, NY 11101-4132 (e-mail: slim1@health.nyc.gov).

Initially submitted March 12, 2013; accepted for publication May 21, 2013.

Despite concerns about nonresponse bias due to decreasing response rates, telephone surveys remain a viable option for conducting local population-based surveillance. However, this becomes problematic for urban populations, which typically have higher nonresponse rates. Unfortunately, traditional methods of evaluating nonresponse bias pose challenges for public health practitioners due to high costs. In this study, we sought to increase understanding of survey nonresponse at the zip code level in an urban area and to demonstrate the use of a practical tool for assessing nonresponse bias. Data from the 2008 New York City Community Health Survey, a landline telephone survey of residential households in New York, New York, were matched with zip-code-level data from the 2000 US Census. Although response rates varied across zip codes and zip-code-level sociodemographic characteristics, estimated nonresponse bias for the 5 health measures (general health status, current health insurance coverage, asthma, binge drinking, and physical activity) was not substantial (ranging from -3.8% to 2.4%). Findings confirmed previous research that survey participation rates can vary a great deal across small areas and that there is no direct relationship between response rates and nonresponse bias. This study highlights the importance of assessing nonresponse bias for local urban surveys and demonstrates a workable assessment tool.

bias; health surveys; New York City; urban population

Abbreviations: CHS, Community Health Survey; NYC, New York City; SD, standard deviation.

Both low response rates and systematic differences in survey measures between respondents and nonrespondents can lead to nonresponse bias (1–3). The decision to respond to a survey is hypothesized to be a complex stochastic process that is influenced by 1) survey features, such as design features, and survey topic interest and saliency; 2) respondent characteristics, such as sociodemographic attributes; and 3) the interaction between the two, forming response propensity for a given person (1, 4–7). Biased estimates may pose a serious challenge for public health practitioners who use survey data to make resource-allocation decisions and evaluate programs.

Despite recent advancements in the understanding of nonresponse bias, much of the existing research has been at the national or state level, making implications unclear for local-level surveys (2). Moreover, for highly urban local areas, such as New York City (NYC), there is a greater concern about nonresponse bias, since response rates are lower among urban dwellers; they are not only harder to contact but also less likely

to cooperate even after contact (7–9). However, the current literature on nonresponse bias in urban settings is limited, and directly assessing nonresponse bias is often prohibitively expensive because it requires extra data collection with nonrespondents.

Recently, Lee et al. (2) proposed a practical tool for assessing nonresponse bias in a telephone survey by linking survey data with existing US Census data. Using regression models with Census block-level predictors and individual-level health outcomes, they calculated the effect of nonresponse bias on the estimated prevalence of major health conditions in the California Health Interview Survey. Thus, our objective in this study was 2-fold: 1) to evaluate nonresponse in a health telephone survey for small geographic areas within a large urban environment and 2) to demonstrate that the cost-effective approach for nonresponse bias assessment originally proposed by Lee et al. could be applied to a local health survey in a highly urban population.

MATERIALS AND METHODS

Study population

The study included 39,415 landline telephone numbers generated by Marketing Systems Group (Horsham, Pennsylvania) for the 2008 NYC Community Health Survey (CHS). The CHS is an annual, cross-sectional, random-digit-dial telephone survey of the general population of New York, New York, on health conditions and related behaviors that is conducted by the NYC Department of Health and Mental Hygiene. It employs stratified sampling in which random sample selection is performed within 42 NYC neighborhood strata, which are aggregations of adjoining zip code areas as defined by the United Hospital Fund (10, 11). For the contacted households that agree to participate in the survey, 1 adult is randomly selected to complete the interview (10). To generate NYC-representative estimates of health risk behaviors and chronic disease conditions, 2 weighting factors are used: 1) the inverse of the probability of selection and 2) the poststratification adjustment factor based on NYC neighborhood-level distributions of age, race/ethnicity, and sex (10). The CHS included a cell-phone sampling frame starting in 2009, but for this study the sampling frame included only those persons who could be reached by landline telephone.

To obtain neighborhood-level information for sampled households to assess survey response patterns (responded, refused, noncontact) and potential nonresponse bias by sociodemographic characteristics, we matched CHS data with Census 2000 data using zip codes (12). Only 30.1% of the 39,415 sampled households provided their residential zip code. For the remaining sample, we determined zip codes by reverse-matching to Telcordia (Piscataway, New Jersey) databases (33.3% of the 39,415 samples) or performing probabilistic assignment based on the zip code distribution of listed households (those listed in telephone directories) sharing the same first 6 digits of the telephone number (36.6% of the 39,415 sampled). Because the CHS is considered health surveillance, this study was not subject to NYC Department of Health and Mental Hygiene institutional review board review.

Study variables

We calculated 3 measures of survey participation for each zip code using the American Association for Public Opinion Research (Deerfield, Illinois) formulas for response rate 1, noncontact rate 1, and refusal rate 1 (13). With a denominator that comprised completed and partial interviews, confirmed noninterviews (refusals and break-offs, noncontacts, and others), and all cases of unknown eligibility (unknown household/occupied housing units, other unknown cases), these 3 measures represented the percentage of completed interviews (response rate 1), the percentage of noncontacts (noncontact rate 1), and the percentage of refusals and break-offs (refusal rate 1) (13).

To evaluate nonresponse bias, we obtained zip-code-level estimates on 1) 5 self-reported survey measures (general health status, current health insurance coverage, asthma, binge drinking, and physical activity) from CHS 2008 data (see Appendix for the exact questions) and 2) 13 sociodemographic characteristics from the Census data known to be highly correlated with response propensities and the selected health

measures (percent male, percent aged ≤ 17 years, percent aged ≥ 65 years, percent non-Hispanic white, percent non-Hispanic black, percent Hispanic, percentage of persons over age 25 years with less than a high school diploma, percentage of households earning less than 200% of the federal poverty level, percentage of people receiving public assistance, percentage who owned their own homes, percentage of female-headed households with children, percentage of single-adult households, and percentage of households with children). We selected these 5 CHS survey measures because they are core surveillance measures for the NYC Department of Health and Mental Hygiene and are used to evaluate the effects of key programs and policies. They also represent major dimensions of population health, including general health perception (general health status), access to care (current health insurance coverage), chronic disease (asthma), and health-related behaviors (binge drinking and physical activity).

Statistical analysis

We first analyzed the data to describe zip-code-level survey participation and sociodemographic characteristics of respondents and nonrespondents. Then we performed multi-level logistic regression analyses to estimate overall nonresponse bias for each of the 5 survey measures (each measure was evaluated in a separate model) by using the 13 zip-code-level sociodemographic variables as predictors in each model. To address the possible violation of independence of observations (i.e., observations in the same neighborhood are more similar than those from different neighborhoods), we allowed the intercept of the regression model to vary randomly across 42 neighborhoods using aggregated zip codes defined by the United Hospital Fund (11).

We then predicted survey estimates (i.e., self-reported health outcomes) for the full sample that included both respondents and nonrespondents. Because the regression models generated predicted values only for respondents, we entered zip-code-level sociodemographic information for the full sample (percent male, percent Hispanic, etc.) into the regression equations, which in turn yielded projected estimates for the full sample. Using this method, we also obtained projected estimates for the nonrespondents and predicted estimates for respondents. For each survey measure, the difference between the projected estimate from the full sample and the observed estimate from respondents was considered the projected bias due to survey nonresponse (2). Additionally, the difference between the predicted estimate from respondents and the observed estimate from respondents was used to evaluate the model fit (2). All statistical analyses were performed using SAS 9.2 software (SAS Institute Inc., Cary, North Carolina). We calculated both unweighted (response rates by neighborhood characteristics) and weighted (nonresponse bias for 5 health measures) estimates. Because of the large sample size, results of statistical significance tests are not reported.

RESULTS

The respondents came from 180 of the 184 residential zip codes in NYC. The average response rate across the 180 zip codes was 19.4% (standard deviation (SD), 6.9%; range,

0%–36.1%). The average rates of noncontact and refusal were 5.5% (SD, 3.1%; range, 0%–20.0%) and 4.2% (SD, 2.6%; range, 0%–20.0%), respectively, and varied widely across zip codes. Response rates also varied across quartiles of each socio-demographic characteristic (Table 1). Average response rates were higher in zip codes with a greater proportion of Hispanics (fourth quartile vs. first: 21.3% vs. 17.4%), single-mother households (21.6% vs. 13.2%), households with children (21.1% vs. 14.7%), persons over age 25 years without a high school diploma (20.5% vs. 16.6%), people receiving public assistance (21.8% vs. 16.4%), and homeowners (22.0% vs. 18.0%). On the other hand, average noncontact rates increased as the proportion of non-Hispanic blacks, single-mother households, and high school dropouts in zip codes increased, while the opposite trend was found for refusal rates and the proportion of homeowners (data not shown).

The predicted prevalence of health outcomes, as opposed to the observed prevalence, differed by –4.0% to 3.0% (data

not shown), indicating that the fit of the models for most outcomes was moderate. Projected bias due to survey nonresponse, which was the difference between the observed estimate for survey respondents and the projected estimate for the full sample, ranged from –3.8% to 2.4% (Table 2). The measure with the largest projected nonresponse bias was current health insurance coverage: the projected citywide prevalence (88.5%) was greater than that in the observed sample (84.7%), indicating that nonresponse bias was –3.8%. Overall, the general direction of bias suggested that nonrespondents were slightly healthier than respondents.

DISCUSSION

The study found that overall survey response rates at the zip code level were negatively associated with several socioeconomic status characteristics but positively associated with

Table 1. Average Response Rates (%) in the New York City Community Health Survey, by Quartile^a of Zip Code Characteristics, New York City, 2008

	Total	Quartile of Zip Code Characteristic			
		First	Second	Third	Fourth
Demographic factors					
Male sex	47.9	19.3	20.3	21.8	15.9
Age, years					
≤17	18.4	12.3	20.7	21.1	20.6
≥65	11.9	16.9	20.7	20.0	20.1
Race/ethnicity					
Non-Hispanic white	41.6	20.3	19.6	18.8	18.7
Non-Hispanic black	21.6	20.3	17.3	19.2	20.7
Hispanic	24.1	17.4	19.8	19.0	21.3
Persons over age 25 years with less than a high school diploma	23.0	16.6	20.6	19.9	20.5
Female-headed households with children	9.7	13.2	20.6	20.5	21.6
Single-adult households	54.0	21.0	20.3	20.1	14.7
Households with children	51.7	14.7	19.6	21.2	21.1
Economic factors					
Households with incomes under 200% of federal poverty level	19.1	18.2	20.2	19.2	20.9
People receiving public assistance	7.1	16.4	19.8	20.0	21.8
Living in a home that one owns	28.0	18.0	16.6	20.1	22.0
Sampling frame^b					
Listed households ^c	53.4	20.9	20.8	17.9	18.1

^a Quartile groups for zip code characteristics were defined using 3 cutpoints (25th percentile, 50th percentile, and 75th percentile). Cutpoints were as follows—percent male: 46.1%, 47.5%, and 48.7%; percent aged ≤17 years: 13.1%, 19.3%, and 23.4%; percent aged ≥65 years: 9.1%, 11.2%, and 14.4%; percent non-Hispanic white: 12.1%, 47.8%, and 69.3%; percent non-Hispanic black: 3.3%, 8.1%, and 31.5%; percent Hispanic: 8.4%, 18.0%, and 30.9%; percentage of persons over age 25 years with less than a high school diploma: 13.4%, 20.6%, and 29.4%; percentage of female-headed households with children: 3.4%, 6.7%, and 15.7%; percentage of households with incomes less than 200% of the federal poverty level: 9.9%, 16.1%, and 26.3%; percentage of people receiving public assistance: 2.5%, 5.4%, and 9.9%; percentage of people living in a home that they owned: 13.1%, 23.8%, and 38.2%; percentage of listed households: 47.0%, 53.0%, and 58.0%; percentage of single-adult households: 46.0%, 54.6%, and 61.2%; percentage of households with children: 43.5%, 51.3%, and 61.6%.

^b “Sampling frame” refers to the source from which a sample is drawn.

^c Listed households are households with telephone numbers that have been published in telephone directories.

Table 2. Projected Values (%) for Survey Health Measures According to Response Status and Estimated Nonresponse Bias in the New York City Community Health Survey, New York City, 2008^a

Health Measure	Observed Respondents (R) ^b	Predicted Respondents	Projected Nonrespondents	Projected Population (PP) ^c	Estimated Bias (R - PP)
General health status is excellent/very good	46.4	43.4	45.7	45.5	0.9
Currently covered by health insurance	84.7	88.8	88.3	88.5	-3.8
Asthma attack in past 12 months	5.3	4.9	4.7	4.7	0.6
Binge drinking in past 30 days	13.6	10.3	10.8	10.7	2.4
Physically active in past 30 days	72.7	72.6	74.1	73.7	-1.0

^a Estimates were weighted using Community Health Survey weights. The following zip-code-level predictors, which varied substantially across New York City zip codes, were included in the regression model: percent male, percent aged ≤ 17 years, percent aged ≥ 65 years, percent non-Hispanic white, percent non-Hispanic black, percent Hispanic, percentage of persons over age 25 years with less than a high school diploma, percentage of households with incomes less than 200% of the federal poverty level, percentage of people receiving public assistance, percentage of people living in a home that they owned, percentage of female-headed households with children, percentage of single-adult households, and percentage of households with children.

^b Observed estimates from respondents.

^c Projected population estimates from the full sample. The projected population consists of respondents and nonrespondents.

home ownership and Hispanic ethnicity. This relationship is not consistent with the finding that Hispanic ethnicity is negatively associated with survey response in the Behavioral Risk Factor Surveillance Survey, which is similar in design and content to the NYC CHS but is conducted at the national level (14). Our findings also differ from those of a similar study conducted using data from the 2005 California Health Interview Survey, where response patterns did not differ according to sociodemographic characteristics (2). These contradictory results could reflect the fact that our target population was highly urban, with unique sociodemographic profiles. For example, households with children in NYC relative to single-adult households have a lower socioeconomic status, which is not what is seen at the national or state level (data not shown). Our finding of high survey response rates associated with Hispanic ethnicity and low socioeconomic status may be due to the fact that these households are more likely to have children and the adults spend more time at home, which makes them easier to contact and more likely to complete interviews than single adults (1).

Despite the observed differences between respondents and nonrespondents, estimated nonresponse bias was not substantial. This may have been due to a null association between nonresponse bias in the health measures and the sociodemographic variables. Such an association was reported previously in a meta-analysis of 23 studies (15). Our finding that it varied by 5 self-reported survey measures was consistent with other findings that bias due to survey nonresponse is item-specific (3, 16). Greater nonresponse bias for the prevalence of current insurance and binge drinking may reflect a lower likelihood of single-adult households to participate in the survey.

This study had several limitations. First, since we used neighborhood characteristics to assess nonresponse bias, we

were not able to account for individual-level variation within each zip code. Second, assignment of zip codes to the sample was imperfect. Third, the population characteristics in the Census 2000 data might have changed by 2008; however, Summary File 3 from Census 2010, containing data on zip-code-level sociodemographic variables, was not available at the time of this study. Fourth, the sampling frame did not include cell phone numbers. Omitting cell-phone-only households, which were estimated to be 19.3% of NYC adults according to the 2008 NYC Housing and Vacancy Survey (17), could reduce the generalizability of our study findings to the NYC population. Lastly, some improvement in the regression models might be possible, since predicted respondent values from the regression models differed somewhat from those of actual respondents. For instance, the premise of this approach is that nonresponse can be explained by the neighborhood sociodemographic variables included in the model. A future analysis expanding the sampling frame to include cell-phone-only households is warranted to address these limitations. Despite these limitations, this study contributed to understanding nonresponse bias in health telephone surveys that may be unique to an urban population, by analyzing a large sample of the NYC adult population.

Because correlates of nonresponse vary across areas and population groups, tools are needed to estimate and adjust for nonresponse bias for individual measures in each survey (18). An approach is needed to go beyond studies that simply describe demographic differences between respondents and nonrespondents rather than possible differences in health estimates (19). Our replication of the method first proposed by Lee et al. (2) suggests that it is a cost-effective and relatively easily applied approach for estimating item-specific bias due to survey nonresponse and that it can be used for surveys conducted in an urban setting.

ACKNOWLEDGMENTS

Author affiliations: Bureau of Epidemiology Services, Division of Epidemiology, New York City Department of Health and Mental Hygiene, New York, New York (Sungwoo Lim, Stephen Immerwahr, Tiffany G. Harris); and Institute for Social Research, University of Michigan, Ann Arbor, Michigan (Sunghee Lee).

This work received no financial support or external funding.

We thank Drs. Donna Eisenhower, Bonnie Kerker, and Mike Battaglia for their helpful comments.

Conflict of interest: none declared.

REFERENCES

- Groves RM. *Survey Errors and Survey Costs*. New York, NY: John Wiley & Sons, Inc; 1989.
- Lee S, Brown R, Grant D, et al. Exploring nonresponse bias in a health survey using neighborhood characteristics. *Am J Public Health*. 2009;99(10):1811–1817.
- Groves RM. Nonresponse rates and nonresponse bias in household surveys. *Public Opin Q*. 2006;70(5):646–675.
- Groves RM, Cialdini RB, Couper MP. Understanding the decision to participate in a survey. *Public Opin Q*. 1992;56(4):475–495.
- Singer E, Van Hoewyk J, Maher MP. Experiments with incentives in telephone surveys. *Public Opin Q*. 2000;64(2):171–188.
- Groves RM, Presser S, Dipko S. The role of topic interest in survey participation decisions. *Public Opin Q*. 2004;68(1):2–31.
- Groves RM, Couper M. *Nonresponse in Household Interview Surveys*. New York, NY: John Wiley & Sons, Inc; 1998.
- Dimock M, Samaranayake N, Best J. *Are We Under-Counting Blue Counties? Correcting for Disproportional Response Rates at the County Level*. (Paper prepared for the 2005 annual conference of the American Association for Public Opinion Research, Miami Beach, Florida, May 12–15, 2005). Washington, DC: Pew Research Center for the People and the Press; 2005. (<http://www.people-press.org/files/2012/10/Dimock-et-al-pop-density.pdf>). (Accessed August 8, 2013).
- Goyder J, Lock J, McNair T. Urbanization effects on survey nonresponse: a test within and across cities. *Qual Quant*. 1992;26(1):39–48.
- Van Wye G, Kerker BD, Matte T, et al. Obesity and diabetes in New York City, 2002 and 2004. *Prev Chronic Dis*. 2008;5(2):A48.
- New York City Department of Health and Mental Hygiene. *New York City United Hospital Fund (UHF) Neighborhoods and NYC ZIP Code Areas*. New York, NY: New York City Department of Health and Mental Hygiene; 2006. (http://home2.nyc.gov/html/doh/downloads/pdf/survey/uhf_map_100604.pdf). (Accessed March 15, 2012).
- US Census Bureau. *United States Census 2000. Summary File 3 (SF 3)*. Washington, DC: Bureau of the Census, US Department of Commerce; 2003. (<http://www.census.gov/census2000/sumfile3.html>). (Accessed March 15, 2012).
- American Association for Public Opinion Research. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 7th ed. Deerfield, IL: American Association for Public Opinion Research; 2011.
- Link MW, Mokdad AH, Stackhouse HF, et al. Race, ethnicity, and linguistic isolation as determinants of participation in public health surveillance surveys. *Prev Chronic Dis*. 2006;3(1):A09.
- Peytcheva E, Groves RM. Using variation in response rates of demographic subgroups as evidence of nonresponse bias in survey estimates. *J Off Stat*. 2009;25(2):193–201.
- Peytchev A, Baxter RK, Carley-Baxter LP. Not all survey effort is equal: reduction of nonresponse bias and nonresponse error. *Public Opin Q*. 2009;73(4):785–806.
- US Census Bureau. *2008 New York City Housing and Vacancy Survey*. Washington, DC: Bureau of the Census, US Department of Commerce; 2008. (<http://www.census.gov/housing/nychvs/data/2008/nychvs08.html>). (Accessed May 3, 2012).
- Johnson TP, Cho YI, Campbell RT, et al. Using community-level correlates to evaluate nonresponse effects in a telephone survey. *Public Opin Q*. 2006;70(5):704–719.
- Skalland B. An alternative to the response rate for measuring a survey's realization of the target population. *Public Opin Q*. 2010;75(1):89–98.

APPENDIX

Exact Questions Used for 5 Survey Measures in the New York City Community Health Survey, 2008

- General health status: Would you say that in general your health is excellent, very good, good, fair, or poor?
- Current health insurance coverage: Do you have any kind of health insurance coverage, including private health insurance, prepaid plans such as HMOs [health maintenance organizations], or government plans such as Medicare or Medicaid?
Read if necessary: Medicare is a health insurance program for people 65 and older or persons with disabilities.
Read if necessary: Medicaid is a health insurance program for persons whose income and resources cannot cover the costs of health care.
- Current asthma: In the last 12 months, have you had an episode of asthma or an asthma attack?
- Binge drinking: Considering all types of alcoholic beverages, how many times during the past 30 days did you have 5 or more drinks on one occasion?
- Physical activity: During the past 30 days, other than your regular job, did you participate in any physical activities or exercises such as running, calisthenics, golf, gardening, or walking for exercise?