Where, Exactly, are the RDD Non-Respondents: A GIS-Based Approach to Quantifying Potential Bias

Ned English[1], Lisa Schwartz[2], Lisbeth Goble[1]
National Opinion Research Center, 55. E Monroe St, Chicago, IL, 60603[1]
Mathematica Policy Research, Princeton, NJ[2]

Abstract

Non-response is a growing issue in RDD telephone surveys, with the current literature describing how completed interviews are increasingly difficult to obtain. It is usually taken for granted, however, that survey respondents will be similar enough to non-respondents so that non-response bias is small. This paper uses data from 'Poetry in America', a national RDD survey conducted by NORC, to quantify exactly how survey respondents differ from non-respondents. We present a method that employs geographic information systems (or 'GIS') to compare the geographic and demographic characteristics of respondents and non-respondents through geocoding, the appending of census data, as well as spatial visualization. Examples of examined census-derived variables include race/ethnicity, income, and urbanicity. Because, by definition, non-respondents do not disclose useful information, there is usually no means of learning what data are missing due to non-response. Our data demonstrate that increases in response rate would not necessarily change the geographic and attribute distribution of survey data measurably. We argue that geographic examination adds value by enabling some quantification of the differences between respondents and non-respondents.

Keywords: GIS, non-response, bias

1. Introduction

Non-response has become a serious issue in RDD surveys, with a notable decline in overall response rates during the past twenty years (Brehm 1993). To substantiate the merits of their data in the face non-response, survey researchers often conduct non-response bias analyses followed by post-stratification. By definition, however, non-response bias analyses and/or post-stratification require control totals for comparison. At issue, then, is how can one study non-response bias in situations where the population of interest is not well understood or is simply not enumerated. This paper presents a geographic area-based method for studying the effects of non-response in non-enumerated populations where control totals are un available.

2. Background and Problem

'Response rate' is often seen as a general gauge of survey effectiveness, and so it is given considerable importance in survey research (Groves et al. 1988). In the current era of new telephone technology (such as caller-id and privacy manager) one can expect RDD response rates in the 30-40% range, which would have been previously unacceptable (Tuckel and O'Neill 2001).

Consequently, non-response bias analyses are often conducted after data collection to quantify the differences between respondents and non-respondents, by comparing respondents' demographics to control totals. Any differences between the two groups indicate what subpopulations are underrepresented, and therefore implicitly describe the non-respondents. One can then post-stratify and so weight respondents to match population cells of interest.

There are, however, three shortcomings with the usual approach to non-response bias analysis and post-stratification: they require control totals *a priori*; they are "implicit", relying on differences with a known population rather than explicitly describing the non-respondents; they do not take advantage of the geographic relationships between respondents. The first issue, that non-response analyses require control totals, is especially limiting in situations of rare or specialized populations. In such situations one would require a direct means of comparison between respondents and non-respondents, instead of relying on implicit differences.

All three of the above listed drawbacks can be overcome by relying on area-based measures rather than those at the individual level. We demonstrate a method using geographic data quantified within GIS to learn about both respondents and non-respondents.

The analyses discussed herein arose from the *Poetry in America* survey, a national RDD study of poetry usage. *Poetry* collected approximately 1000 completed interviews during 2005, with a final response rate of 37%. While these results are typical for RDD surveys

in the present era, there was still interest in quantifying any non-response bias. This project targeted English-speaking adults who ever "read for pleasure", and so constituted a specialized population with no control totals, precluding traditional non-response analysis.

3. Methodology

Our approach was based on using GIS technology to ascribe area-level measures to both respondents and non-respondents to permit appending Census data. We were able to geocode all 987 respondents that provided addresses. We then utilized a vendor to map telephone exchange to ZIP code for all non-responding cases, whether they were eligible or not. After that, we geocoded each non-respondent to the ZIP-code centroid level to permit the appending of Census tract boundaries. Following geographic processing, we merged salient Census data for both respondents and non-respondents and compared the two groups across a number of measures.

4. Results and Discussion

Table 1 shows differences between the tract-level data for the 987 respondents and three categories of non-respondents, with 5% significant differences indicated by '*' and 1% significant differences indicated by '**'. As shown, the category of 'all non-respondents' tended to reside in denser, metropolitan tracts with fewer white non-Hispanic residents and more foreign-born residents than respondents. Non-respondents, or at least telephone numbers associated with non-response, thus tended to be in generally more urban areas. Note that because the category of 'all non-respondents' includes unresolved numbers, it is believed to include a substantial portion of non-households.

It is therefore more appropriate to make the same comparisons but for known households, as summarized in the 'non-responding known households' category in Table 1. Note that while the cases summarized in this field are known to be households, they are of unknown eligibility for our survey. The only significant differences between respondents and non-respondents for known households was for population density, with this group of non-respondents residing in denser areas than respondents. Otherwise, non-respondents tended to be located in tracts similar to those of non-respondents.

Lastly, Table 1 shows the same results for known eligible refusals. None of the differences between

respondents and known eligible refusals were significant.

As demonstrated in Table 1, as one knows less about a case (e.g., whether they are eligible or even if they are a household), the less like a respondent it becomes. In addition, our results indicate that non-respondents tended to reside in denser, more racially/ethnically diverse areas when compared to respondents. This observation of increasing urbanicity with non-response is shown graphically in Figure 1, comparing the measures of percent White non-Hispanic, percent urban, and population density by category of non-respondent. Our results do, however, demonstrate that tract-level data for known eligible refusals is essentially the same as for respondents in the *Poetry in America* survey. We therefore concluded that our survey was not biased from a spatial perspective.

Our GIS-based approach permits the comparison of respondents and non-respondents without requiring gross control-totals beyond Census area measures. The geographic method demonstrated herein also allows for the geographic examination and comparison of different categories of non-response.

4. Conclusions

This study illustrates a method for conducting a type of non-response bias analysis for non-enumerated populations. Our results are, however, limited because they describe areas, rather than individuals. It would be desirable to use finer levels of geography (e.g., block group), which could potentially reveal more nuanced differences among subpopulations. In addition, one could use commercial database match technology to learn detailed information about non-respondents.

References

Brehm, J. 1993. The Phantom Respondents. Ann Arbor: University of Michigan Press.

Groves, Robert M, P. Biemer, L. Lyberg, J. Massey, W. Nicholls II, J. Waksberg. 198. Telephone Survey Methodology. New York: John Wiley and Sons.

Tuckel, Peter, and Harry O'Neill. 2001. *The Vanishing Respondent in Telephone Surveys.* Paper presented at the annual meeting of the American Association for Public Opinion Research, Montreal, Canada.

Appendix A Tables

Table 1- Respondents vs. Categories of non-Respondents

Tract-Level Measure	All Respondents	Known Eligible non-Respondents	Non-Responding Known Households	All Non-Responding Cases
% White non-Hispanic	**75.9%**	74.1 %	73.7 %	72.1% **
% Foreign Born	**7.9 %**	8.3 %	8.3 %	10.5 % **
% Under 18	**25.1 %**	25.2 %	24.8 %	24.6 %
% Grad Degree	**9.5 %**	8.5 %	8.8 %	9.3 %
% Metropolitan	**79.4 %**	78.2 %	81.4 %	82.8 % *
Median Tract Density (pop/mi2)	**4169**	4245	5351 **	6172 **
Median Household Income	**$ 46,710**	$45,014	$45,758	$ 46,092

Figure 1- Comparison of Measures