Dual-Frame Weighting Of RDD And Cell Phone Interviews At The Local Level

Michael P. Battaglia[1], Donna Eisenhower[2], Stephen Immerwahr[2]
Kevin Konty[2],
[1]Abt Associates, 55 Wheeler St., Cambridge, MA 02138
[2]New York City DOHMH, Epidemiology Services, 125 Worth St., New York, NY 10013

Abstract

National dual frame telephone surveys use National Health Interview Survey estimates of the size of telephone service groups. How does one go about weighting a local-level dual frame survey? Telephone service control totals were calculated in two ways. A model-based method fit a multinomial regression model to data from the NHIS. The model was applied to New York City data from the American Community Survey to estimate telephone service totals. A second approach used direct estimates from the NYC Housing and Vacancy Survey, an in-person survey. We compared estimates for health risk factors. Differences in overall and subgroup prevalence were small. The results suggest that the model-based weighting approach yields prevalence estimates that are close to the estimates based on weighting to local independent estimates. Implications for local dual frame surveys are discussed.

Key Words: Dual Frame Sample Design, Poststratification

1. Introduction

Dual frame telephone sample designs are now widely used to provide coverage of adults with telephone service in the U.S. (Srinath et al. 2004). In an overlapping dual frame design a list-assisted random-digit-dialing sample is drawn from the landline sampling frame. A random sample of 10-digit telephone numbers is also drawn from dedicated cellular telephone exchanges. For households sampled in the landline sample one adult is typically randomly selected. In the cell phone sample the cell phone is treated as a personal communication device and the adult using the phone is interviewed, or one adult is randomly selected from among the adults in the household that share the cell phone. All adults in the cell phone sample are eligible to be interviewed. A non-overlapping design follows the same approach except that in the cell phone sample only adults who only have a cell phone are eligible to be interviewed.

For both types of dual frame designs the final step in the weighting methodology often involves poststratification to socio-demographic control totals and telephone service control totals. For national surveys the telephone service control totals are generally obtained from the NHIS for three categories: cell-only adults, landline-only adults, and dual service (landline and cell service) adults. Poststratification to control totals in telephone surveys is used to reduce nonresponse bias and noncoverage bias. The ACS or the Current Population Survey can be used to develop control totals for socio-

demographic variables. The telephone service control totals can be obtained from published NHIS reports or by processing the NHIS public-use file (PUF). The NHIS can provide national control totals but those control totals will likely be one to two years out-of-date relative to the field period of the telephone survey. For most poststratification variables this is typically not an issue, but telephone usage patterns are rapidly changing in the U.S. For example, from January-June 2008 to January-June 2009 the percent of adults in the U.S. who only have cell phone service increased by 31 percent, from 16.1% to 21.1% (Blumberg and Luke 2009).

2. Control Totals for Local Surveys

The New York City Community Health Survey (CHS) is an example of a sub-state telephone health survey. The CHS is conducted by the New York City Department of Health and Mental Hygiene. It is an annual random-digit telephone sample of around 10,000 adults. The questionnaire content is similar to the Behavioral Risk Factor Surveillance System. Starting in 2009 a non-overlapping dual frame sample design is being used in order to also provide coverage of cell-only adults in New York City.

For sub-state surveys such as the CHS one can obtain socio-demographic control totals from the 2006-2008 American Community Survey public-use microdata sample (ACS PUMS). The 2005-2009 ACS PUMS will also be released in the near future. However, for sub-state surveys no source of direct telephone service control totals is available. For telephone service poststratification of state and sub-state surveys one has the option of using the national NHIS estimates or NHIS estimates at the Census Region in which state or sub-state area is located, or one can develop indirect model-based estimates of telephone service for the state or sub-state area. We discuss the poststratification of the 2008 and the 2009 NYC CHS and show results from the 2008 NYC CHS.

3. Model-based Telephone Service Estimates

New York City consists of five boroughs. Our objective was to develop model-based estimates for New York City and for the individual boroughs. The 2008 NHIS PUF was used to classify NHIS households as cell-only, landline-only, dual service, or nontelephone. The nontelephone households were then excluded from a multinomial logistic regression model that used telephone service as the dependent variable (reference group = dual service households) in the NHIS PUF that are also present in the 2006-2008 ACS PUMS:

- Type of living quarter (home or apartment)
- Census Region
- Total number of persons in the household
- Presence of children in the household
- Presence of elderly persons in the household
- Highest education level among all adults in the household
- Tenure status (rent or own)
- Presence of male adults in the household
- Presence of female adults in the household
- Presence of Hispanic adults in the household
- Presence of nonHispanic black adults in the household
- Presence of never married adults in the household

- Presence of currently married adults in the household
- Presence of adults less than or equal to 30 years of age in the household
- Family versus nonfamily household

In Table 1 we show the multinomial model coefficients for the education predictor variable. All household education groups are more likely to be cell only than dual service compared to households where the highest education level is college graduate. The same pattern exists for landline only households versus dual service households with households where the highest education level is less than high school graduate are five times more likely to be landline only than dual service.

Table 1: 2008 NHIS Model Coefficients for Education Predictor Variable

Reference group = College Graduate	*Dependent Variable*	*Odds Ratio Estimate*
Some college	Cell-only versus dual service	1.1
High school graduate	Cell-only versus dual service	1.3
Less than high school graduate	Cell-only versus dual service	1.5
Some college	Landline-only versus dual service	1.5
High school graduate	Landline-only versus dual service	2.8
Less than high school graduate	Landline-only versus dual service	5.0

Following the methodology developed by Battaglia et al. (2008) the 2008 NHIS model was applied to the 2006-2008 ACS PUMS. The first step was to identify the public-use microdata areas (PUMAs) in the ACS PUMS that cover the five boroughs. The ACS households in the New York City PUMAs were extracted from the PUMS and the nontelephone households were set aside. In the second step the NHIS model was used to score the ACS telephone households in New York City. This resulted in three predicted probabilities being assigned to each household (cell-only, landline-only and dual service) with the three probabilities summing to one for each household. In the third step the three predicted probabilities were assigned to all adults in the household. Finally, using the ACS person weight we then estimated the number of adults in each borough falling into the three telephone service categories. As part of the last step we iteratively adjusted the ACS person weights so that at the Census Region level the percent of adults in the three telephone service categories was in close agreement with the NHIS Census Region estimates, and the percent of adults who live in nontelephone households was also in close agreement with the NHIS Census Region estimates.

4. 2008 New York City Housing and Vacancy Survey

The New York City Housing and Vacancy Survey (HVS), conducted by the Census Bureau at the request of New York City every three years, is required by State and City rent regulation laws to determine New York City's overall vacancy rate for rental housing. The survey draws its sample of roughly 21,000 housing units from the 2000 decennial census conducted by the U.S. census Bureau and is updated by the Department of Housing Preservation and Development to include new construction, conversion and alteration. In-persons interviews for the HVS were conducted between February and June 2008. The survey questionnaire was modified to include questions to determine the household telephone service classification. We used the 2008 HVS to develop estimates of the percent of adults by borough in the telephone service groups. In Table 2 we show

the model-based estimates and the direct estimates from the 2008 HVS. For all five boroughs the model-based estimates for the cell-only adult population are lower than the direct HVS estimates, and the HVS estimates for the landline-only adult population are lower than the model-based estimates.

Table 2: Comparison of 2008 NHIS Model-Based Telephone Service Estimates With 2008 HVS Direct Estimates

	Cell-Only		*Landline-Only*		*Dual Service*	
Borough	*Model*	*HVS*	*Model*	*HVS*	*Model*	*HVS*
Bronx	14.3%	15.8%	27.5%	15.4%	58.2%	68.8%
Brooklyn	11.6%	17.3%	25.2%	13.2%	63.2%	69.5%
Manhattan	15.4%	26.8%	24.1%	12.8%	60.4%	60.5%
Queens	9.7%	16.2%	22.3%	10.3%	68.0%	73.5%
Staten Island	6.8%	10.3%	19.8%	7.7%	73.4%	82.0%
New York City	12.0%	18.4%	24.2%	12.3%	63.9%	69.3%

5. New York City CHS Weighting Methodology

We describe the weighting of the 2009 CHS using the same approach as for the 2008 CHS. For the landline sample of adults the design weight took into account the probability of selection of the telephone number, the number of voice-use landline telephone numbers in the household, and the number of adults in the household. The design weight for the cell-only adults in the cell phone sample took into account the selection probability of the telephone number. The poststratification of the combined sample involved raking to three control variables for each borough:

- Neighborhood area by age group by gender
- Neighborhood area by race/ethnicity
- Telephone service group (cell-only, landline-only, dual service)

We ran the raking for each borough twice – using the model-based telephone service estimates and using the HVS direct estimates.
Because the cell-only population has been increasing rapidly in recent years we applied an annual increase adjustment to our 2008 telephone service estimates before using them in the 2009 CHS raking. From the NHIS we determined that at the national level the percent of adults who are cell-only increased by 31 percent from January-June 2008 to January-June 2009. At the same time the percent of adults who are landline-only has been steadily decreasing over time. We also calculated a “milder” increase factor using the January-June 2008 to January-June 2009 increase in the percentage of adults who are cell-only divided by the January-June 2008 percentage of adults who are *not* cell-only. This estimate equals 6 percent. We decided to take the mean of these two adjustment factors and increased the percent of adults who are cell-only in each borough by 19 percent. To force the telephone service percentages to add to 100 percent in each borough we reduced the landline-only percentages. The end result of the weighting process was two set of final weights -- model-based versus HVS.

Table 3 shows health risk factor and health condition estimates from the 2008 CHS for New York City from the two raking approaches. For the nine variables seven of the differences are quite small -- less than one percentage point. The timely mammogram

estimates for females age 40 or over differ by 1.1 percentage points and the timely colonoscopy estimate for adults age 50 and over differ by 2.1 percentage points.

Table 3: Comparison of Health Indicator Variable Prevalence Estimates using 2008 CHS data

Health Indicator Variable	*Estimate using model-based parameters*	*Estimate using HVS parameters*	*Percentage point difference*
Current smoker	16.5%	16.2%	0.3
Binge drinker	14.2%	14.9%	-0.7
Heavy drinker	4.6%	4.8%	0.2
MSM	4.1%	4.1%	0.0
Uninsured	17.0%	16.4%	0.6
Diabetes	9.6%	9.3%	0.3
Obese	22.5%	22.3%	0.2
Timely colonoscopy (age 50+)	64.7%	66.8%	-2.1
Timely mammogram (age 40+)	76.8%	77.9%	-1.1

In Table 4 we show bivariate results (relative risks) from the 2008 CHS for age groups and race/ethnicity by obesity. The relationship between age and obesity and between race/ethnicity and obesity is very similar for the two weighting approaches. This finding holds for all nine health indicator variables.

Table 4: Comparison of Relative Risk Estimates for Obese Variable using 2008 CHS data

Age Group:	*Relative risk using model-based parameters*	*Relative risk using HVS parameters*
18-24 (reference group)	1.00	1.00
25-44	1.60	1.67
45-64	2.10	2.24
65+	1.66	1.75
Race/ethnicity:		
nonHispanic white (reference group)	1.00	1.00
nonHispanic black	1.74	1.72
Hispanic	1.55	1.62
nonHispanic Asian	0.41	0.44
nonHispanic all other races	1.35	1.26

6. Conclusions

The NYC CHS is currently using the 2008 HVS telephone service estimates for poststratification of the combined landline and cell-only samples of adults. The HVS is however only conducted every three years and therefore at some point it might be necessary to use model-based estimates of telephone service for the poststratification. We found that across nine health indicator variables the differences in the prevalence estimates between the two telephone service poststratification control totals are very small. This also held true for subgroup prevalence estimates and bivariate analyses. Most dual frame telephone surveys conducted at the sub-state level will not have access to direct control total estimates from a survey similar to the HVS, making it necessary to develop model-based estimates for the local area. As shown here, however, the model-

based approach is a close approximation to the direct control total method. When models are used it would be useful to conduct a sensitivity analysis by changing the control totals for the telephone service raking margin to reasonable alternative values and examine the impact on the prevalence estimates.

References

Battaglia, M.P., Frankel, M.R., and Mokdad, A.H. 2008. Statistical Challenges Facing Cell Phone Survey. 2008 Proceedings of the Annual Meeting of the American Statistical Association [CD-ROM], Alexandria, VA: American Statistical Association.

Blumberg, S.J. and Luke, J.V. 2009. Wireless substitution: Early release of estimates from the National Health Interview Survey, January-June 2009. National Center for Health Statistics. December 2009. Available from: http://www.cdc.gov/nchs/nhis.htm [Accessed August 31, 2010.]

Srinath, K.P., Battaglia, M.P., and Khare, M. 2004. A dual frame sampling design for an RDD survey that screens for a rare population. 2004 Proceedings of the Annual Meeting of the American Statistical Association [CD-ROM], Alexandria, VA: American Statistical Association.