Compensating for Noncoverage of Nontelephone Households Using the NHIS

Karen E. Davis and Meena Khare
National Center for Health Statistics, U.S. Centers for Disease Control and Prevention,
3311 Toledo Road, Room 3213, Hyattsville, MD 20782
(kedavis@cdc.gov), (mkhare@cdc.gov)

Abstract

Telephone surveys are generally the quickest and most cost-effective way to investigate emerging public health issues. However, telephone survey estimates are subject to bias from noncoverage of nontelephone households. To compensate for such bias, sampling weights are typically adjusted using ratio adjustments to match the demographics of the population or model-based weight adjustments of telephone households based on logistic regression estimates of the propensity for each household to have been without service. Although only about 5% of households in the U.S. do not have landline telephones, rates of telephone coverage show substantial variation by geography, demography, and socioeconomic factors. In particular, lack of telephone service is more common among households that contain young adults or persons with lower socioeconomic status. Valid and practical methods are needed to adjust survey estimates for potential bias due to noncoverage.

Keywords: noncoverage, telephone survey, propensity score model

1. Introduction

Based on Keeter's weight adjustment method (1995), Brick, *et al.* (1996), Frankel, *et al.* (2003) and Srinath, *et al.* (2002) showed that socioeconomic characteristics of persons who live in households with interruptions of >1 week in telephone services within the past 12 months are similar to those who live in nontelephone households. Therefore, persons with "interruption in telephone service" can be used to represent persons without phone service in telephone surveys. Researchers have been investigating alternative methods, based on data obtained from telephone households that had interruptions in their telephone service during the preceding year of the survey, to adjust for noncoverage of nontelephone households.

The target population of persons in households, at the time of the telephone survey, can be classified into four groups (Frankel, *et al.,* 2003). Group **T/NI** contains persons from households with telephone service (T) at the time of the survey and no interruption in telephone service (NI) during the entire year. Group **T/I** contains persons from households with telephone service at the time of the survey but interruptions in telephone service (I) during the previous year. Group **NT/I** consists of persons from households that did not have telephone service (NT) during the entire year but had telephone service at some time during the year (that is, their lack of telephone service was interrupted). Group **NT/NI** contains persons from households with no telephone service during the entire year. This last group is considered as permanent nontelephone households, with no chance of being covered in a telephone survey. However, some of the households in the group NT/I would have had a chance of selection at some point in time.

This paper describes an evaluation of three weighting methods used to adjust for noncoverage of nontelephone households in the U.S., based on interruptions in telephone service, to reduce noncoverage bias.

2. Data

For this research, we used data from two years (2001-2002) of the National Health Interview Survey (NHIS). The 2003 NHIS was used for evaluation. The NHIS is a multi-purpose health survey conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), and is the principal source of information on the health of the civilian, noninstitutionalized, household population of the United States. It is an annual survey that covers both telephone and nontelephone households, and collects data on interruptions in landline telephone service during the previous 12 months. Data are collected through in-person interviews (http://www.cdc.gov/nchs/nhis.htm).

In general, approximately 2% of households experience interruptions of >1 day in telephone

service in the U.S. We defined five interruption groups, using the length of interruption reported in the NHIS (2001-2002). The corresponding percentages of households with interruptions are as follows:

With any interruption of >1day   1.86%
With any interruption of >1 week   1.61%
With any interruption of >2 weeks   1.48%
With any interruption of >4 weeks   1.02%
With any interruption of >6 weeks   0.79%

In order to select an appropriate interruption group, based on length of interruption, to compare to the nontelephone group, a logistic regression model was developed for each interruption group as well as for the nontelephone group. Significant covariates for each interruption group model included age group, sex, race/ethnicity, marital status, education, household income, house tenure, poverty ratio, MSA status, U.S. born, and Census region. We then plotted the normal predicted probability from each interruption group model. Figure 1 shows the distribution of households by the interrupted groups and compares with the nontelephone group. The distribution of the nontelephone group appears to be somewhat closer to the group with interruption of >2 weeks as compared to the other four interruption groups. We also compared selected characteristics of interrupted households and assessed the impact on the final estimates by these interruption groups (Table 3). In this paper, we define the interrupted group as telephone households with interruptions of >2 weeks.

3. Methods

Using data from the 2001-2002 NHIS, we divided the household sample into telephone and nontelephone households based on the reported telephone status. ("Is there at least one telephone inside your home that is currently working?")
The 2003 NHIS was used for evaluation, to compare the weighting procedures, and to assess coverage bias in the estimates. In order to simulate a sample from a telephone survey, we selected only those households that reported having landline telephones, regardless of interruptions in service, as our evaluation sample. Thus, the 2003 NHIS evaluation sample did not contain any nontelephone households.

Table 1 compares the demographic and health characteristics by telephone interruption status of U.S. households, and shows the magnitude of the

bias in terms of the under- or over-representation in percentage points by the difference between interruption households and nontelephone households. Nontelephone status appears to be related to young adults 18-34 years old, and to the socioeconomic status of the household members (e.g., having low income, less than high school education, having no health insurance at all, or no private health insurance). Similar results were observed in the data from the 2003 NHIS. Table 2 shows the difference in age distribution of the telephone and nontelephone households.

Weighting class adjustment methods are typically used in complex surveys to adjust for unit nonresponse, and poststratification is used to adjust for noncoverage. An alternative approach is to use logistic regression to model the response propensity of living in a nontelephone household, and then use the inverse of the estimated propensity to adjust sampling weights for noncoverage.

We used the following three weight adjustment methods:

SP1: Simple poststratification by age, gender, race-ethnicity within 32 demographic cells; **NT2:** Model-based adjustments using the parameter estimates from the 2001-2002 NHIS logistic regression no-phone model to obtain response propensities, then apply the inverse of the predicted propensities to the 2003 telephone only evaluation sample; **INT3:** Direct ratio-adjustments by poststratifying for interruption in telephone services using the >2 weeks interruption group.

In the simple poststratification method (SP1), the 2003 NHIS population control totals were used to poststratify the nonresponse-adjusted base weights of adults reported having landline telephones, to compensate for nontelephone households. The NHIS uses population controls from the Current Population Survey (CPS), classified by Census region, age group, sex, and race/ethnicity.

For logistic regression models, a binary response indicator variable is regressed upon potential covariates that are available for both with and without phone households to obtain predicted probabilities of having no phone. The parameter estimates from the model are then used to obtain response propensities in the evaluation sample.

To develop the propensity model (NT2), we identified a set of variables that were common to

the 2001 - 2002 NHIS and the 2003 NHIS data and that were associated with nontelephone status. The model sample (2001-2002 NHIS) was used to develop a logistic regression model to predict telephone status. Since telephone status at the individual case level is dichotomous, a binomial logit model was used (Battaglia, et al., 1995). The logit model was defined as

$$\ln\{\tau_i/(1-\tau_i)\} = x_i\beta + e \qquad (1)$$

where τ_i is the probability that person i lives in a household that did not have telephone service. The beta coefficients from the final model included age group, sex, race/ethnicity, marital status, education, household income, house tenure, and Census region. These covariates were used to estimate the propensity of telephone status in the 2003 NHIS evaluation sample. We used the inverse of the predicted propensity scores, as an adjustment factor $[=1/\exp(\ln\{\tau i/(1-\tau i)\})]$, to the nonresponse-adjusted 2003 NHIS sampling weights. This step adjusted the weights of persons in telephone households to account for similar persons from nontelephone households.

The third method (i.e., INT3) uses the direct ratio-adjustment procedure, currently used for weight adjustment in the National Immunization Survey, to compensate for noncoverage of nontelephone households (Smith et al., 2005). Based on our results in Figure 1, we used >2 weeks of interruption to define the interruption group for the INT3 method. In the direct ratio-adjustment approach, the weights for persons in the **T/NI** sample group are first poststratified using the telephone population controls, excluding the population controls for households with interruptions in service. Then, the weights for persons in the **T/I** sample group are poststratified by adding the no-phone population controls to the population controls for households with interruption in service. Table 4 shows the adjustment factors that resulted from this approach. Since the ratio of the adjustment factors were large (e.g., > 3.0) for several poststratification cells, we trimmed the with interruption group adjustment factors to reduce variability in the sampling weights.

Finally, we added a final poststratification adjustment to methods NT2 and INT3 to adjust the total cell counts to the total U.S. population (based on the population controls from the 2003 CPS). The final post-stratified weights were then used to compute weighted estimates for several key characteristics, and to compare the three weighting methods.

To assess the impact of the three noncoverage adjustment methods, weighted estimates of several health and sociodemographic characteristics were compared. For this comparison, we used the following variables: self-reported health status, no private health insurance, no health insurance at all, income < $20,000 has activity limitation, poverty status, and education. Estimates of bias were computed by taking the difference between the NHIS estimates and the estimates from methods SP1, NT2, and INT3. Standard errors were calculated using SUDAAN (Shah et al., 1997). The mean square error (MSE = bias^2 + standard error^2) for the estimates were calculated for each method. Ratios of the NHIS MSE to the MSEs from the three methods were compared.

4. Results

Table 1 shows that there are substantial differences between the socio-demographic characteristics of the telephone and the nontelephone households as well as between households with and without interruptions in phone service. Figure 1 indicates that households with >2 weeks of interruption can be used to represent nontelephone households in telephone surveys.

After applying method SP1 to adjust the nonresponse-adjusted base sampling weights in the 2003 NHIS evaluation sample, minor differences are observed in the point estimates and in the ratio of standard errors when compared to the corresponding 2003 NHIS estimates (see Tables 5 and 6). Method NT2 produced consistently lower estimates from the 2003 NHIS, except for the self-reported excellent or good health estimate. Method INT3 appears to produce estimates much closer to the estimates from the 2003 NHIS (or method SP1), as compared to method NT2. Table 5 indicates that the MSE associated with the estimates from method INT3 are the smallest. All MSE ratios from Method INT3 to SP1 are less than one. This shows that the direct-ratio method of adjustment using interruptions in service performed better than both the model-based response propensity method (NT2), and the simple poststratification method (SP1).

5. Conclusion

Adjustments based on interruption in telephone service seem to reduce noncoverage bias,

especially for those variables that are highly correlated with the presence or absence of telephone service (e.g. lower socioeconomic status). This adjustment may lead to increased variability in sampling weights, and may require some trimming of extreme weights. An increase in the standard errors of the estimates may offset the decrease in bias. Therefore, it is important to evaluate the mean square error of these estimates and compare them with the mean square error of the simple poststratification estimates, since method SP1 does not make an explicit adjustment for noncoverage of nontelephone households.

References

Battaglia, Michael P., Malec, Donald J., Spencer, Bruce D., Hoaglin, David C., and Sedransk, J. 1995. *Adjusting for noncoverage of nontelephone households in the National Immunization Survey.* 1995 Proceedings of the Section on Survey Research Methods, Alexandria, VA: American Statistical Association, pp. 678-683.

Brick J, Waksberg J, Keeter S. *Using data on interruptions in telephone service as coverage adjustments.* Survey Methodology 1996; 22:185–197.

Frankel, Martin R., Srinath, K.P., Hoaglin, David C., Battaglia, Michael P., Smith, Philip J., Wright, Robert A., and Khare, Meena. 2003. *Adjustments for non-telephone bias in random-digit-dialling surveys.* Statistics in Medicine, Volume 22, pp. 1611-1626.

Keeter S. *Estimating non-coverage bias from a phone survey*. Public Opinion Quarterly 1995; 59:196–217.

Shah BV, Barnwell BG, Bieler GS. *SUDAAN User's Manual, Release 7.5.* Research Triangle Institute: Research Triangle Park, NC, 1997.

Smith, P.J., Hoaglin, D.C., Battaglia, M.P., Khare, M., and Barker, L.E. (2005), *Statistical Methodology of the National Immunization Survey: 1994-2002*. National Center for Health Statistics. Vital Health Stat 2(138).

Srinath, K.P., Battaglia, Michael P., Frankel, Martin R., Khare, Meena, and Zha, Wenxing. 2002. *Evaluation of a procedure based on interruptions in telephone service for reducing coverage bias in RDD surveys.* 2002 Proceedings of the Annual Meeting of the American Statistical Association [CD-ROM], Alexandria, VA: American Statistical Association, pp. 3323-3329.

Table 1: Comparison of Selected Characteristics of Adults by Telephone Interruption Status, US Households, NHIS 2001-02

Characteristics of adults 18+ years	Total (%) (T)	Nontelephone (%) (A)	Telephone with Interruption (B)	Telephone without Interruption (C)	Difference (%) (B - A)	Difference (%) (B - C)
Age:18-34 years	30.2	47.2	50.1	30.2	2.90	19.90
Age:35-54 years	40.6	38.0	36.0	40.6	-2.00	-4.60
Age:55-64 years	12.5	7.6	7.3	12.5	-0.30	-5.20
Age:65+ years	16.7	7.3	6.6	16.7	-0.70	-10.10
Male	47.97	53.66	47.32	47.74	-6.34	-0.42
Female	52.03	46.34	52.68	52.26	6.34	0.42
Non-Hispanic Whites and Others	77.74	62.54	65.96	78.61	3.42	-12.65
Hispanics	10.90	17.31	15.06	10.55	-2.25	4.51
Non-Hispanic Blacks	11.37	20.15	18.98	10.84	-1.17	8.14
Income >= $20,000	81.10	48.86	62.84	82.69	13.98	-19.85
Income < $20,000	18.90	51.14	37.16	17.31	-13.98	19.85
Education: Less Than High School	47.50	69.38	56.56	46.49	-12.82	10.07
Education: Some College	44.51	28.56	39.16	45.22	10.60	-6.06
Education: Graduate School	7.99	2.06	4.28	8.29	2.22	-4.01
Owner	71.45	31.53	45.99	73.54	14.46	-27.55
Renter	28.55	68.47	54.01	26.46	-14.46	27.55
Health Status: Excellent, Very Good	62.75	51.86	55.06	63.35	3.20	-8.29
Health Status: Good	25.57	31.34	28.59	25.27	-2.75	3.32
Health Status: Fair, Poor	11.69	16.80	16.35	11.38	-0.45	4.97
Has Private Insurance	70.23	35.19	49.11	72.12	13.92	-23.01
No Private Insurance	28.61	61.18	49.58	26.82	-11.60	22.76
No Health Insurance	15.63	40.56	30.92	14.28	-9.64	16.64
Has Health Insurance	83.20	55.81	67.77	84.66	11.96	-16.89
Has Limitation of Activity	14.35	17.34	19.25	14.12	1.91	5.13
No Limitation of Activity	85.31	81.65	80.53	85.56	-1.12	-5.03
Poverty Ratio < 100%	10.57	34.37	27.19	9.29	-7.18	17.90
Poverty Ratio 100-199%	16.86	29.35	26.11	16.18	-3.24	9.93
Poverty Ratio 200-399%	31.96	24.93	28.96	32.29	4.03	-3.33
Not Born in US	14.56	15.53	14.84	14.51	-0.69	0.33
Born in US	85.44	84.47	85.16	85.49	0.69	-0.33
MSA	79.89	69.93	76.36	80.38	6.43	-4.02
Not MSA	20.11	30.07	23.64	19.62	-6.43	4.02
Urban	72.96	68.75	74.89	73.09	6.14	1.80
Rural	27.04	31.25	25.11	26.91	-6.14	-1.80

Table 2: Distribution of Adults With and Without Telephone Service by Age, NHIS 2003

Age Group (years)	Phone Inside Household			Phone Not Inside Household
	Total (%)	Without Interruption	With Interruption	Total (%)
18-34	30.2	29.7	54.2	53.1
35-54	39.9	40.1	31.9	33.5
55-64	13.3	13.4	8.0	7.7
65+	16.6	16.7	5.9	5.7

Table 3: Estimated Percentage of Adults with any Interruption, NHIS 2001-2002

Characteristic	More Than 1 Day		More Than 1 Week		More Than 2 Weeks		More Than 4 Weeks	
	Estimate	S.E	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Total	1.86	0.07	1.61	0.06	1.48	0.06	1.02	0.05
18-34	2.83	0.14	2.49	0.14	2.33	0.13	1.67	0.12
35-54	1.68	0.07	1.45	0.07	1.32	0.06	0.88	0.05
Hispanic	2.39	0.16	2.13	0.15	2.05	0.15	1.46	0.13
Black	2.93	0.18	2.55	0.17	2.40	0.16	1.68	0.15
White/Other	1.62	0.07	1.40	0.07	1.26	0.07	0.86	0.05
High School or Less	2.19	0.09	1.93	0.08	1.82	0.08	1.34	0.07
Some College	1.69	0.09	1.44	0.09	1.28	0.08	0.82	0.07
Less than $20,000	3.14	0.19	2.84	0.19	2.74	0.18	2.04	0.16
$20,000 or more	1.53	0.06	1.30	0.06	1.16	0.06	0.76	0.04
Owner	1.29	0.06	1.10	0.06	0.96	0.05	0.61	0.04
Renter	3.31	0.15	2.94	0.15	2.81	0.14	2.06	0.13
Uninsured	3.41	0.19	3.12	0.18	3.00	0.18	2.22	0.15
Insured	1.56	0.06	1.33	0.05	1.18	0.05	0.79	0.04

Table 4: Age by Race/ethnicity Adjustment Factors for Method INT3 (NHIS 2003)

Age x Race/ethnicity	Without Interruption Group	With Interruption Group	Ratio
18-34 Hispanic	1.29642	4.7405	3.6566082
18-34 NH Black	1.3506	4.6441	3.4385458
18-34 NH White/ Other	1.27012	5.4513	4.2919567
35-54 Hispanic	1.2746	5.7853	4.5389142
35-54 NH Black	1.20304	5.1785	4.3045119
35-54 NH White/ Other	1.17138	5.5225	4.7145247
55-64 Hispanic	1.27851	5.3345	4.1724351
55-64 Black	1.1414	3.4708	3.0408271
55-64 NH White/ Other	1.16057	5.6327	4.853391
65+ Hispanic	1.32266	15.3096	11.574857
65+ Black	1.17133	3.9428	3.3660881
65+ NH White/ Other	1.18959	5.0561	4.2502879

Table 5: Comparison of Estimates Using Three Weighting Methods, NHIS 2003, Telephone Sample

Characteristics	NHIS 2003		SP1		NT2		INT3	
	Est	SE	Est	SE	Est	SE	Est	SE
Health status: (1)								
Excellent to good	88.20	0.18	88.40	0.19	89.09	0.19	88.08	0.20
Health status: (3)								
Fair/poor	11.80	0.18	11.60	0.19	10.91	0.19	11.92	0.20
PRIVINS: (2) No private ins	30.03	0.33	28.85	0.33	27.83	0.32	29.58	0.35
Health insurance coverage: (1) Uninsured	16.32	0.24	15.39	0.24	15.35	0.23	15.83	0.25
Activity LIMITED: (1)								
Has limitation	14.49	0.20	14.26	0.20	13.64	0.19	14.55	0.21
INCOME: (1) LT $20K	18.83	0.33	17.52	0.33	16.74	0.33	18.29	0.34
EDUCAT: (1)								
LE High School	47.13	0.38	46.42	0.38	45.39	0.38	47.00	0.38
Poverty status: (1)								
<100%	11.33	0.31	10.46	0.32	10.00	0.32	11.08	0.33

Table 6: Comparison of MSE* Associated with Three Weight Adjustment Methods, NHIS 2003

Characteristics		MSE		MSE Ratio
	INT3	NT2	SP1	INT3/SP1
Health status: (1)				
Excellent to good	0.054	0.828	0.077	0.6989138
Health status: (3)				
Fair/poor	0.054	0.828	0.077	0.6989138
PRIVINS: (2) No private ins	0.325	4.959	1.504	0.2159324
Health insurance coverage: (1) Uninsured	0.309	1.003	0.935	0.3311454
Activity LIMITED: (1)				
Has limitation	0.049	0.761	0.088	0.5525022
INCOME: (1)				
LT $20K	0.408	4.481	1.824	0.2236581
EDUCAT: (1)				
LE High School	0.164	3.152	0.649	0.2529777
Poverty status: (1)				
<100%	0.171	1.867	0.845	0.202687

Figure 1: Normal Predicted Probability Plots for Adults with No Phone or with Interruption in Telephone Services, NHIS 2001-2002