USING DATA ON INTERRUPTIONS IN TELEPHONE SERVICE TO REDUCE NONTELEPHONE BIAS IN A RANDOM-DIGIT-DIALING SURVEY

Martin R. Frankel, K.P. Srinath, Michael P. Battaglia, David C. Hoaglin, Abt Associates Inc.; Philip J. Smith, Centers for Disease Control and Prevention ; Robert A. Wright, Meena Khare, National Center for Health Statistics

Martin R. Frankel, Abt Associates Inc., 55 Wheeler Street, Cambridge, MA 02138

Key Words: Random-digit dialing, Weighting adjustment, Nontelephone households

1. Introduction

Telephone surveys are subject to coverage bias from noncoverage of nontelephone households. Though the percentage of households not having telephone service is small nationally, it can vary substantially by geographic area and by socioeconomic factors. For example, lack of telephone service is more common among low-income households than in other income groups, so low-income households may not be adequately represented in a telephone survey (Thornberry and Massey 1988). Postsurvey weighting can reduce this bias. Keeter (1995), using a panel constructed from the 1992-93 Current Population Survey (CPS), observed that at any given time telephone households include households that were recently a part of the nontelephone population. Generally, these households have had an interruption in telephone service. By comparing the characteristics of these households with those without telephones, he showed that it is possible to use data from households with interruptions in telephone service to adjust for noncoverage of nontelephone households. Brick et al. (1996) suggested a method of adjusting survey estimates to reduce bias from noncoverage, by using data on interruptions in telephone service.

This paper describes a method to adjust for noncoverage of nontelephone households. This method assumes that the population number of individuals in telephone households is known (or can be estimated from a survey or from alternative sources) and uses the survey results to estimate the weighted proportion of individuals in telephone households with an interruption in service. It then forms a weighted average of the estimates from the non-interruption and interruption parts of the sample, using the interruption estimate for both the interruption and the nontelephone parts of the population.

We use the proposed method to adjust the estimates of vaccination coverage rates for children between 19 and 35 months of age from the National Immunization Survey (NIS), and we compare the adjusted estimates with the current estimates obtained through the usual poststratification methods. Truncation at an upper limit reduces the impact of large weights resulting from this adjustment.

We provide the mathematical theory and empirical results from the 1997 National Immunization Provider Record Check Study (NIPRCS), which indicate that this interruption adjustment method substantially reduces the nontelephone associated bias in estimated childhood vaccination rates. We compare our results to those obtained using poststratified weights without explicit adjustment for noncoverage and to a currently used non-coverage adjustment.

The National Immunization Survey

Since April 1994, the NIS, conducted by the Centers for Disease Control and Prevention, has measured vaccination coverage rates among children 19 to 35 months old, nationally and in each of the 78 Immunization Action Plan (IAP) areas (the 50 states, the District of Columbia and 27 other urban areas). A sample of telephone numbers is drawn quarterly for each of the 78 IAP areas.

The NIS collects data in two phases. First, list-assisted random-digit dialing (RDD) and a screening interview are used to identify households containing a child 19 to 35 months of age. Such households are asked to report the child's vaccinations and also to list the providers of those vaccinations and to give consent to contact them. Second, a mail survey asks those providers for vaccination data from the child's medical record.

The estimates of vaccination coverage discussed in this paper are based on the data from providers.

2. Conceptual Framework for Noncoverage Adjustment

In order to quantify the ability of our proposed estimates to reduce bias associated with noncoverage of nontelephone households, we develop expressions for bias associated with these estimates.

The target population of children in households at the time of the telephone survey can be classified into four groups, as shown in Table 1. Group T/NI contains children coming from households with telephone service (*T*) at the time of the survey and no interruption in service (*NI*) of more than one week during the previous year.

Group T/I contains children coming from households with telephone service at the time of the survey but with an interruption (*I*) in telephone service of more than one week during the year. Group NT/I contains children from households that had no telephone service (*NT*) at the time of the survey but had telephone service at some time during the year, that is, their lack of telephone service was interrupted; and, finally, Group NT/NI contains children from households with no telephone service during the entire year. Let the number of children in each of these groups be as shown in Table 1. (As mnemonic subscripts we use o to denote "no telephone" and *c* to denote "no interruption" [i.e., continuing service or lack of service].)

The population numbers in the cells are generally unknown, though we may know or reliably be able to estimate the numbers of telephone and nontelephone households and children in the population. Let *N* be the size of the total population. Let N_t be the number of children from telephone households. When we use RDD, we obtain a sample of children from only the telephone households. Let this sample size be n_t. Assume that we are interested in estimating a certain population proportion relating to children [e.g., the proportion of children who are 4:3:1 (4 doses of DTP vaccine, 3 doses of Polio vaccine and 1 dose of MMR vaccine) up-to-date]. Let this proportion in the four cells be as shown in Table 2.

We are interested in estimating the population proportion *P*, which can be written as

$$P=\frac{N_{it}\,P_{it}+N_{ct}\,P_{ct}+N_{io}\,P_{io}+N_{co}\,P_{co}}{N}.$$

P can also be written in terms of P_t and P_o, the proportions for children in telephone and nontelephone households:

$$P=\frac{N_t\,P_t+N_o\,P_o}{N}$$

In terms of the proportions for the four cells defined in Table 2,

$$P_t=\frac{N_{it}\,P_{it}+N_{ct}\,P_{ct}}{N_t}$$

$$P_o=\frac{N_{io}\,P_{io}+N_{co}\,P_{co}}{N_o}.$$

That is, the proportion for children from telephone households is the weighted average of the proportions for children in telephone households with and without interruption. Similarly, the proportion for children in nontelephone households is the weighted average of the proportions for children in nontelephone households with and without interruptions in their lack of telephone service.

Bias in Poststratification

If we obtain a sample of children from the population of households that have a telephone at the time of the survey and if we adjust the base sampling weights to the full known population totals, we obtain the usual telephone-sample estimate. That is, the estimate is based on a telephone sample projected to the total (telephone and nontelephone) population. Let the estimate of the proportion of interest based on the sample of telephone households be p_t. We have

$$E(p_t)=P_t\ .$$

The bias in using p_t as an estimate of *P* is

$$B(p_t)=P_t-P.$$

This can be written as

$$B(p_t)=\frac{N_o}{N}(P_t-P_o)\qquad(1)$$

Thus, the bias of the telephone-sample estimate is a function of the proportion of children in nontelephone households (N_o/N) and the difference (P_t-P_o) in the proportion of interest between children in telephone and nontelephone households .

3. Method of Adjustment

Let the number of children coming from households in the sample with no interruption in telephone service during the year be n_{ct}. Let p_{ct} be the estimated proportion of interest for this group. Let n_{it} be the number of children from households with interruption in telephone service and p_{it} be the corresponding estimate. Let N_t be the total number of children in telephone households and N_o be the total number of children in nontelephone households at the time of the survey. As indicated earlier, these population sizes either are known or can be estimated, either from the survey or from alternative sources.

Let $\hat{N}_{it}$ be the estimate of N_{it}, the number of children in the population coming from households with telephones at the time of the survey but with interruptions in telephone service during the year. To calculate $\hat{N}_{it}$ we use the weighted sample proportion of children with interruption. The weights are denoted by w_{itk} (k= 1,2, 3,..., n_{it}) for children from households with interruption in telephone service and w_{ctk} (k = 1, 2, 3,...,n_{ct}) for children from telephone households without interruption. Applying the weighted proportion of *T/I* children in the sample to N_t yields

$$\hat{N}_{it} = N_t \frac{\sum_{k=1}^{n_{it}} w_{itk}}{\sum_{k=1}^{n_{it}} w_{itk} + \sum_{k=1}^{n_{ct}} w_{ctk}}. \quad (2)$$

We then form the two totals $N_t - \hat{N}_{it}$, the estimated number of children in telephone households without interruption, and $N_o + \hat{N}_{it}$, the estimated number of children in nontelephone households or in telephone households with interruption in service. If we multiply the proportion of interest p_{ct} obtained from the sample of children in telephone households by the first total $N_t - \hat{N}_{it}$,we get an estimate of the number of children in telephone households without interruption and with the characteristic of interest. Then, if we multiply the proportion of interest p_{it} for children in telephone households with interruption by the second total, we get an estimate of the number of children in nontelephone households and telephone households with interruption who have the characteristic of interest. The sum of the two estimates divided by the estimated number of children in the population gives an estimate of the overall proportion of interest in the population. That is,

$$\hat{P} = \frac{(N_t - \hat{N}_{it})\, p_{ct} + (N_o + \hat{N}_{it})\, p_{it}}{N}. \quad (3)$$

The bias in $\hat{P}$ is

$$B(\hat{P}) = E(\hat{P}) - P = E_1 E_2(\hat{P}) - P$$

where $E_2(\hat{P})$ is the conditional expectation over samples in which the two subsample sizes (number of children with and without interruption) are fixed and E_1 is the expectation over all possible sample sizes . Substituting for $E_2(\hat{P})$ and taking the expectation, we get

$$B(\hat{P}) = \frac{N_o}{N} (P_{it} - P_o). \quad (4)$$

Compare (4) with (1). Now the bias is the proportion of children in nontelephone households multiplied by the difference between the proportion of interest for children in telephone households with interruption and the corresponding proportion for children in nontelephone households. $\hat{P}$ has smaller bias than p_t if $|P_t - P_o| > |P_{it} - P_o|$.

Further, we can express the difference in (4) as the sum of two differences by using the definition of P_o.

$$B(\hat{P}) = \frac{N_o}{N} \frac{[N_{co}(P_{it} - P_{co}) + N_{io}(P_{it} - P_{io})]}{N_o}. \quad (5)$$

Similarly, we have

$$B(P_t) = \frac{N_o}{N} \frac{[N_{co}(P_t - P_{co}) + N_{io}(P_t - P_{io})]}{N_o}. \quad (6)$$

We expect $(P_{it} - P_{co})$ to be smaller than $(P_t - P_{co})$ and $(P_{it} - P_{io})$ to be smaller than $(P_t - P_{io})$ and therefore, we expect $B(\hat{P})$ to be smaller than $B(p_t)$.

In the 1997 NIPRCS which makes use of in-person interviewing and therefore includes both telephone and nontelephone households, the following estimates of *P* were found for the proportion of children up-to-date on their key vaccinations: *P* (interrupted, telephone)=0.724, *P*(interrupted, no telephone =0.620, *P*(not interrupted, telephone)=0.819 and *P*(not-interrupted, no telephone)=0.649. We see from these data that the proportion for children from households with telephones and interruption is closer to the proportions for children in nontelephone groups than the overall proportion for children from telephone households.

These results provide an empirical confirmation of the conditions under which the interruption procedure results in a decrease in the magnitude of nontelephone associated bias.

4. Applying the Proposed Method to the NIS

We have applied the proposed method to adjust for noncoverage of nontelephone households in the NIS. The process begins with an additional adjustment. The steps described in Section 3 (including poststratification) yield a weight for each child with a household interview, but the estimates of vaccination coverage are based only on the data from providers. Therefore, the household sampling weights of children with provider data are adjusted for nonresponse to the provider data collection, using a method based on response propensities (Smith et al. 2000). The resulting estimates of vaccination coverage on the 4:3:1 series (4 doses of DTP vaccine, 3 doses of polio vaccine, and 1 dose of MMR vaccine) for the U.S., the 50 states and the District of Columbia were produced and are shown for some selected states (with large differences in the estimates) in the INTPWT column of Table 3. (The U.S. estimate and the state estimates are calculated from the estimates for the constituent IAP areas.)

For comparison Table 3 also gives estimates from two methods of adjusting for noncoverage of nontelephone households that do not use data on interruptions in telephone service. Each of these methods produces household weights that are then adjusted for nonresponse to the provider data collection in the same way as the interruption method.

The first method, simple poststratification (SPST), uses the known control totals for cells formed, for each IAP area, by the cross-classification of race/ethnicity of mother, education of mother, and age of the child. Thus within each of these cells simple poststratification assumes that vaccination coverage among children in nontelephone households is the same as that among children in telephone households.

The second method, modified poststratification (MPST), starts with the same cells as simple poststratification. Within each cell it then makes a more refined adjustment by incorporating information from the National Health Interview Survey (NHIS), on the relative levels of vaccination coverage among children in nontelephone and telephone households (Battaglia et al. 1995).

As described earlier, the proposed method first splits the control total of eligible children in an IAP area into those coming from telephone households and those coming from nontelephone households. The split of the total uses the proportion of children, from Census data, that are in telephone households in the IAP area. A weighted estimate of the proportion of children coming from households with interruption in telephone service is obtained from the sample in each IAP area (equation 2). This percentage is used to further split the control total of children coming from telephone households into two groups. The first group consists of children coming from households with telephones and with no interruption in telephone service during the year, and the second group consists of children coming from telephone households but with an interruption in telephone service during the year. This number in the second group is added to the total number of children coming from nontelephone households in the IAP area. The noncoverage adjustment factors for children in households with and without interruption are based on these totals.

In some IAP areas, the noncoverage adjustment factor for children from households with interruption is large, resulting in large weights. After an examination of the ratio of the two adjustment factors, we decided to truncate the ratio to 3.0 for IAP areas in which this factor exceeds 3.0. With this ratio fixed and the adjustment factor for children without interruption also fixed, the adjustment factor for children without interruption was recomputed. The interruption weights are based on the new adjustment factor.

The interruption method yields a lower estimate of the 4:3:1 vaccination rate than the simple poststratifcation estimate in the U.S. and in two-thirds of the states. The interruption estimate is lower than the modified poststratification estimate in the U.S. and in 60% of the states. Among all states and the District of Columbia, the median difference between the INTPWT estimate and the SPST estimate is -0.41 percentage points. In a few states the downward revision is substantial: 2.92 percentage points in Virginia, 2.33 in New Jersey, 1.91 in Mississippi, and 1.86 in South Dakota. The largest upward revision is 1.00 percentage point (in Louisiana). This slight upward revision of vaccination rates in some IAP areas may be due to the small number of households with interruption in telephone service in the sample and to most of the children in these households being 4:3:1 up-to-date. The interruption method puts a higher weight on these children to account for children in nontelephone households. An upward revision is therefore possible with this method. Another reason may be that, in certain IAP areas, low-income households may be targeted for vaccination under special programs. We plan to look at the vaccination rates with this method using data for another set of four quarters to check whether this upward revision is consistent in these IAP areas over time. On the whole, the effect of adjusting for noncoverage of nontelephone households is in the anticipated direction. National data from the NIPRCS show lower vaccination rates among children in nontelephone households. For 16 states and the District of Columbia, however, the revision based on interruption in telephone service is slightly upward.

In contrast, modified poststratification yields estimates of 4:3:1 vaccination rates that are only slightly lower than simple poststratification. The largest difference is -0.37 percentage point (in Wyoming), and only one difference is positive (0.01 in Connecticut).

References

Battaglia, M.P., Malec, D.J., Spencer, B.D., Hoaglin, D.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., and Keeter, S. (1996) "Using Data on Interruptions in Telephone Service as Coverage Adjustments," *Survey Methodology*, 22, 185-197.

Frankel, M., Ezzati-Rice, T., Wright, R., and Srinath, K.P. (1998) "Use of Data on Interruptions in Telephone Service for Noncoverage Adjustment," *1998 Proceedings of the Section on Survey Research Methods,* Alexandria, VA: American Statistical Association, pp. 290-295.

Frankel, M., Srinath, K.P., Battaglia, M., Hoaglin, D.C., Smith, P., and Wright, R. (1999) "Reducing Nontelephone Bias in RDD Surveys," *1999 Proceedings of the Section on Survey Research Methods,* Alexandria, VA: American Statistical Association.

Keeter, S. (1995) "Estimating Noncoverage Bias from a Phone Survey," *Public Opinion Quarterly,* 59, 196-217.

Smith, P.J., Rao, J.N.K., Daniels, D., and Battaglia, M.P. (2000) "Compensating for Vaccination History Nonresponse Bias in the National Immunization Survey Using Response Propensities, To appear in 2000 *Proceedings of the Section on Survey Research Methods,* Alexandria, VA: American Staitstical Association.

Thornberry, O.T. and Massey, J.T. (1988) "Trends in United States Telephone Coverage Across Time and Subgroups," in *Telephone Survey Methodology*, edited by R.M. Groves et al., New York, John Wiley & Sons,

Table 1: Target Population at the Time of the Telephone Survey

Interruption Status	Telephone Status at the Time of the Survey		Total
	Telephone	No Telephone	
Interruption	*(T/I)* N_{it}	*(NT/I)* N_{io}	N_i
No Interruption	*(T/NI)* N_{ct}	*(NT/NI)* N_{co}	N_c
Total	N_t	N_o	N

Table 2: Population Proportion by Telephone Status and Interruption Status

Interruption Status	Telephone Status at the Time of the Survey		Total
	Telephone	No Telephone	
Interruption	P_{it}	P_{io}	P_i
No Interruption	P_{ct}	P_{co}	P_c
Total	P_t	P_o	P

Table 3 4:3:1 Vaccination Estimates (%) for Selected States (Quarter 2, 1998 through Quarter 1, 1999)

State	(1) SPST	(2) MPST	(3) INTPWT	Difference (3)-(1)	Difference (3)-(2)	Difference (2)-(1)
U.S.	80.51	80.43	80.33	-0.18	-0.10	-0.08
CT	89.75	89.76	88.13	-1.63	-1.64	0.01
LA	81.22	81.02	82.22	1.00	1.19	-0.20
MI	79.18	78.96	77.97	-1.21	-0.99	-0.22
MS	82.68	82.59	80.77	-1.91	-1.82	-0.09
NC	84.80	84.70	85.41	0.61	0.71	-0.10
NH	87.49	87.46	86.00	-1.49	-1.46	-0.03
NJ	84.02	83.98	81.68	-2.33	-2.30	-0.04
OH	76.95	76.81	75.83	-1.12	-0.98	-0.14
OK	77.07	76.78	75.61	-1.46	-1.17	-0.29
SD	77.76	77.68	75.90	-1.86	-1.78	-0.08
TX	76.41	76.32	77.30	0.89	0.98	-0.09
VA	80.57	80.55	77.65	-2.92	-2.89	-0.02
WY	77.72	77.35	76.98	-0.74	-0.37	-0.37

SPST: Simple Poststratification Weight, MPST: Modified Poststratification Weight, INTPWT: Interruption Weight