WEIGHTING FOR NONTELEPHONE HOUSEHOLDS IN THE 2001 CALIFORNIA HEALTH INTERVIEW SURVEY

Ismael Flores Cervantes, J. Michael Brick, Michael E. Jones, Westat; Charles DiSogra and Wei Yen, UCLA
Ismael Flores Cervantes, Westat, 1650 Research Boulevard, Rockville, Maryland 20850

Key Words: Coverage bias, response propensity adjustments, telephone survey

1. Introduction

Since random digit dial (RDD) telephone surveys exclude nontelephone households, the estimates from these surveys may differ from the full population values due to the undercoverage. The size of the undercoverage bias for a particular estimate depends on the relationship between the estimate and telephone status. In this paper, we present our experience when implementing one of the methods for adjusting the weights in RDD surveys to reduce undercoverage bias in the 2001 California Health Interview Survey (CHIS 2001). Adjusting weights is the most appropriate method for general-purpose RDD surveys like CHIS 2001 that collect data on many variables without resorting to surveying nontelephone households.

CHIS was a collaborative project of the UCLA Center for Health Policy Research, the California Department of Health Services, and the Public Health Institute, that focused on public health and access to health care. The survey was the largest state health survey ever undertaken in the United States, producing estimates for the whole state, for the larger counties in the state, and for groups of the smallest counties in the state. The survey also supported the study of the characteristics for the major racial and ethnic groups and a number of smaller ethnic groups within the state. Adults, parents or guardians of children, and adolescents within California households responded. Nearly 58,000 adults were interviewed for CHIS 2001.

2. Adjusting for Nontelephone Households

The obvious solution to remove the undercoverage bias is the inclusion of nontelephone households in the survey. In practice, high costs and short data collection periods are the greatest obstacles in the implementation of in-person surveys. Alternatives with lower costs are dual frame surveys that combine an RDD and area samples (Brick, et. al., 1999a). Although dual-frame surveys have some advantages over telephone-only surveys when dealing with undercoverage bias, they are also too costly to be a viable alternative for most studies with restricted budgets.

In recent years there have been developments in the methodology used to address the problem of undercoverage in telephone surveys. Assuming that a telephone survey is the only feasible mode for data collection, the focus of the problem shifts from data collection to an estimation problem centered on how the sample is weighted. In all these methods, the weights of selected sampling cases are differentially perturbed to account for households without telephones. An additional cost comes from the burden associated with collecting more data items required for the implementation of these methods. Since this cost is much smaller than the cost of in-person surveys, there has been a great interest among survey researchers in both the performance of these methods and their reliable implementation in telephone surveys. A summary of these methods is presented in the following paragraphs.

In most RDD surveys, the only adjustment to the weights that deals with the undercoverage bias is a standard poststratification or raking adjustments to external control totals. If variables correlated to having a telephone are used in this adjustment, then any bias may be reduced by the adjustment. However, the residual undercoverage bias after the standard poststratification adjustment may be relatively large for items that are highly correlated with nontelephone coverage. These adjustments (poststratification or raking) can be seen as "naïve" or implicit adjustments because no special provisions are made to adjust for nontelephone households. This does not mean the weights are not adjusted for telephone undercoverage but rather that the bias is reduced implicitly when the weights are benchmarked to the control totals. The amount of bias reduction depends on the type and number of poststratification cells or raking dimensions and their correlation to the characteristics in nontelephone households. In cases where the data available for forming control totals are not closely related to telephone coverage, the bias reduction may be inadequate. Depending on the subject matter of the survey and the proportion of households without telephone in the population, this approach can be adequate and sufficient in producing estimates that are almost unbiased or with very small biases. In contrast, if the estimates are highly correlated with telephone status, then the associated bias may be too important to be ignored.

In contrast to the implicit methods to remove the bias, explicit or direct methods incorporate special procedures to adjust the sampling weights for telephone coverage. The first explicit method evolved from Keeter's observations of the telephone status of households in different panels of the CPS (Keeter, 1995). Keeter noticed that 42 percent of those not having a phone in 1992 had a phone one year later. Hence, over the course of a year, "transient" telephone households comprise a substantial percentage of nontelephone households. In addition, he reports that transient telephone households bear a closer resemblance in socioeconomic characteristics to nontelephone households than to telephone households that have not experienced an interruption in phone service. Brick, Waksberg and Keeter (1996) took this idea and translated it into a weighting method to account for persons without telephones. The method works by adjusting the weights from respondents who experience telephone interruptions to create a postsurvey weighting adjustment. In particular, the weights of the households with interruptions are adjusted to sum to the total of households with interruptions and households without telephones. The standard poststratification adjustment can be then applied to further reduce the bias. In order to implement this method, questions on the existence and duration of interruptions of telephone service must be asked during the interview.

The Keeter method with some variations has been implemented in several RDD surveys where data on interruptions were collected (Brick et. al., 1999b; Frankel, et al. 1999). Although this approach has shown some improvements in bias reduction, there are some issues to consider. One of the main limitations is that this approach focuses the adjustment on a small subset of all households while the remaining records are not adjusted at all. Consequently, this adjustment could lead to few cases with large weights and subsequent increases in the variability of the weights. Also, the underlying assumption for this adjustment is that every nontelephone household has some probability of having a telephone at a given time even though this is not true. Furthermore, the adjustment is dependent on one variable, which may not be measured very reliably. This last limitation has been overcome to some extent by using additional socioeconomic variables (i.e. rent/own house) in some of the variations of the Keeter method.

A second group of explicit methods to adjust the weights for nontelephone coverage were suggested by Ferraro and Brick (2001). Ferraro and Brick proposed three methods that use models for the telephone propensity of the respondents of the survey. The methods are known as modified poststratification adjustment, calibration (raking) adjustment, and nonignorable model adjustment. These methods rely on auxiliary data to predict telephone status and create an additional poststratification or raking variable that is directly constructed to deal with nontelephone coverage bias. In the particular applications they studied, the modified poststratification approach appeared to have slightly better bias and mean square error properties than the Keeter approach and the two other methods studied. This modified poststratification method is an extension of the calibration method (Deville, Särndal and Sautory, 1993) where the creation of calibration cells is done explicitly for the purpose of adjusting for nontelephone coverage. In this method, logistic regression is used to compute the propensity to be a telephone household using an external survey that includes households with and without telephones. Cells that are homogeneous with respect to the telephone status are formed by grouping cases with similar propensities. Using the same model, these cells are recreated in the sample file to be used as calibration cells. The external survey can also be used to create the control totals since it includes both telephone and nontelephone households.

The modified poststratification approach has several limitations. First, there is an additional burden to the respondents. In order to implement this method, the interview should collect considerably more data for the creation of the model than in the Keeter method. Since many of these questions are sensitive (household income, insurance, participation in welfare programs, etc.) the respondent may refuse to complete the interview. This may not be an issue for surveys where this information is always collected. However, the additional number of questions may be a limiting factor in surveys with short questionnaires or where subject matter is not related to socioeconomic conditions of the respondents (e.g. transportation usage, etc.). The second limitation is that the current and external surveys both must contain all the variables that are used to compute the predicted propensities. The last limitation is related to the consistency of the estimates between the survey file and the control survey. Since the cells are created using common variables, it is important to ensure that the variables are measured consistently. However, this same limitation applies to all poststratification or calibration adjustments to external control totals.

Finally, it is important to note that based on Ferraro and Brick's findings, the Keeter and modified poststratification methods were significantly better than the standard poststratification used in many surveys.

3. Adjusting for Nontelephone Households in CHIS

This section describes the implementation of the adjustment method for CHIS 2001. Since the early stages of the sample design and questionnaire development for CHIS, the inclusion of an adjustment for nontelephone coverage was considered as part of the weighting process. The questionnaire included questions to collect information to be used for a nontelephone adjustment, although the specific form of the adjustment was not defined until after the initial analysis of the data. The weighting process for CHIS 2001 included the standard steps of weighting (i.e., creation of base weight, screener and extended interviews nonresponse). The weights were raked to control totals in the last step. The weighting plan called for the implementation of a nontelephone adjustment as part of the person-level raking procedure in the form of an additional raking dimension. In CHIS separate weights were developed for adults, children and adolescents; therefore, the nontelephone adjustment was done separately for each group.

In CHIS 2001 both the Keeter interruption method and the modified poststratification approach were explored as options to adjust the weights. The Keeter approach was not used due in part to issues consistent with those described in the previous section. Very few records in CHIS 2001 data had interruptions in service (see Table 1). If we had used the Keeter adjustment, the weights of around 1.7 percent of the adults (adults in households with interruption in telephone service for more than one week) would have been increased so much that they might have had undue influence on some estimates.

Instead of the Keeter method, a variation of the modified poststratification method was implemented in CHIS 2001. As described in the previous section, the modified poststratification method uses models for the respondent's propensity of having a telephone (adult, children or adolescent separately) created using data from a different survey. The March 2000 Current Population Survey (CPS) was the survey used for the creation of the propensity models. The March 2000 CPS sample included telephone and nontelephone households and was large enough to produce reliable estimates for California. Table 2 shows the set of variables that are captured in both the CHIS and CPS. Tabulations of these variables were produced to verify that both surveys produced consistent estimates. Only variables that produced similar estimates were considered as predictors.

The main problem encountered when implementing the modified poststratification method was that most of the socioeconomic indicators that were key auxiliary variables in the previous applications did not produce consistent estimates. Some estimates of the CHIS 2001 variables (Table 2) did not mirror those of CPS and could not be used. The inconsistencies were the result of differences in the way the questions were asked. For example, the CHIS 2001 and March 2001 CPS questions for income used to create the poverty variable were asked differently, making the distributions of the variables dissimilar. In particular, CHIS 2001 asked for total household income in the past 12 months, while CPS asked for total household income the previous calendar year. Previous research showed poverty as the most important variable when used to determine telephone propensity. A naïve application of the modified poststratification approach would have actually increased the weights of those above poverty rather than those below poverty, a very undesirable outcome.

Since many socioeconomic variables similar to poverty could not be used for the adjustment, we examined variables related to the structure of the household (e.g., total number of children, adolescents and adult in the household) and demographic variables (e.g., race-ethnicity of the respondent) as alternative candidates for the model. These changes were made to the general approach to make the adjustment more appropriate for this survey.

Before comparing the estimates between CPS and CHIS, missing values in some of the variables in the CHIS sample were imputed. The missing values were due to item nonresponse and were imputed using "hot-deck" imputation. Hot-deck imputation is a technique where cases with missing values for specific variables are filled in with values from other cases. Potential donors (cases that may contribute a value) and recipients (cases with missing values) are classified into cells. The cells are constructed in such a way that characteristics are as homogeneous as possible for potential donors and recipients. Recipients are imputed from donors within the same cell.

Once the imputation was completed, the set of predictors common to the CHIS and CPS files were determined. We then created the calibration cells. The goal was to create cells where the households had a similar propensity of having a telephone. We used the categorical search algorithm CHAID (Kass, 1980) to partition the CPS data, where the dependent variable was the telephone status (i.e., telephone household, nontelephone household). CHAID divides the data into groups so that the propensities between the cells are as different as possible. Given a set of categorical predictors, CHAID divides the data into groups in a stepwise fashion. Through a series of chi-square tests for equality of distributions, CHAID identifies the most important predictor and splits the data set into categories. Each of these categories is further segmented based on other predictors. The merging and splitting continues until no more statistically significant predictors are found or until a user-specified stopping rule is met.

Using CHAID has two advantages over logistic regression as used in Ferraro and Brick. First, the interactions among the predictors are easily identified. Second, there is no need to group records with similar telephone propensities because the cells are created in the CHAID analysis. The final cells were created by collapsing the CHAID cells so there were 100 or more respondents in each cell. Table 3 shows the definition of the cells used for the nontelephone adjustment and the computed telephone propensity rate under the model. The adjustment cells can be classified as respondents with low and high telephone propensities. Table 3 includes the overall raking factors shown in the next to the last column.

After the CHAID analysis, the same cells were created in the CHIS 2001 sample using the model. As mentioned before, the nontelephone adjustment was implemented through an additional raking dimension. In the last step of weighting, the CHIS sample was raked to ten dimensions created by different combinations of variables for geography (selected cities, individual counties and group of counties), race-ethnicity, age groups and gender. The last column in Table 3 shows the overall raking adjustment factors (defined as the ratio of the sum of raked weights to the sum weights before raking). Cells with low telephone propensity tend to have larger raking factors.

4. Evaluation of the Adjustment

As in previous studies, bias reduction and variability of the estimates are the key statistics examined to evaluate the effect of telephone coverage adjustments. However, a direct evaluation of the effect of the adjustment is not possible in CHIS because nontelephone households were not sampled. As an alternative, the effect of the adjustment was evaluated indirectly by comparing estimates created using a second weight. This weight, referred to as *RDD-only weight* $\left(w_{RDD}\right)$, was created following the same steps as in the creation of the nontelephone-adjusted weight except it did not include any explicit adjustment for telephone undercoverage. It does include the standard raking adjustment. As a notational convenience, we refer to the telephone adjusted weight as *NT weight* $\left(w_{NT}\right)$.

To evaluate the bias reduction, a total of 68 estimates of proportions and their standard errors were computed using the same number of variables from the adult extended questionnaire. The variables represented health conditions, health-related behaviors, accessibility to health care (including preventive care), and socioeconomic characteristics (excluding all variables used in raking). At least one question from each section of the extended interview was selected. The variables were coded to "yes/no" categories where the positive answer was linked to measures of good-health, availability or use of preventive care (flu shots, mammograms, etc.), and nonparticipation in welfare programs or receiving benefits. All these conditions were assumed to be uncommon or not readily available for adults in nontelephone households. By recoding the variables in this way, we could evaluate if the direction of differences $\left(\hat{b} = \hat{p}_{NT} - \hat{p}_{RDD}\right)$ was as expected. The difference $\hat{b}$ does not measure the bias with respect to the population because neither estimate is unbaised. Nevertheless, $\hat{b}$ can be used to measure the bias reduction due to the adjustjment.

The average sample size for estimates at the state level was 22,607 and 2,589 for adults under 100 percent poverty level. Table 4 shows the descriptive statistics and the distribution of the estimates (columns 2, 3, 7 and 8) and differences (columns 4 and 9) of the estimates.

The average difference at the state level is negative and very small (less than 0.1 percent). The distribution is skewed to the negative side, indicative of bias reduction in the estimates. However, these differences are very small. To examine the extent of the differences, we computed the same estimates for adults under 100 percent poverty level who have lower telephone ownership. The average difference is -0.08 percent compared to -0.07 percent for estimates at the state level. Still, the difference is less than 1 percent for this group. The largest reduction of bias was for variables such as receiving AFDC, receiving public household subsidies, food stamps, and income less than $20,000, all variables correlated to the nontelephone-raking dimension. On the other hand, the largest positive differences were for mothers never diagnosed with breast cancer, permanent residents with a green card, and adults with insurance for eye exams.

The second part of the analysis examined the variability of the weights. Table 5 shows the coefficient of variation (*cv*) of the two weights and the ratio of *cv*'s. The average ratio shows that there is almost no increase in the variation of the weights after the telephone adjustment. Furthermore, the *cv* for adults under 100 percent poverty is smaller than the *cv* at the state level. The result is a direct consequence of the nontelephone adjustment: the weights were raked to a control total created using variables correlated to poverty (cells created using the variables for participating in AFDC and receiving public housing assistance). In this case when adjusting the weights for nontelephone households, the weights are indirectly benchmarked to patterns of poverty from the external survey. This could potentially transfer biases to the survey if these patterns are not well measured in the external survey.

The amount of variation introduced in the estimates by adjusting the weights for nontelephone households is also evaluated by computing the ratios of the standard errors defined as $ser = se\left(\hat{p}_{NT}\right)/ se\left(\hat{p}_{RDD}\right)$ for the same 68 proportions (columns 5 and 10 in Table 4). The average ratio at the state level and for under 100 percent poverty are very close to 1. Furthermore, the average ratios show the same effect described above. The average and median ratio for 100 percent poverty is smaller than the average and median ratio at the state level. For this set of estimates, the standard errors computed using the nontelephone-adjusted weights are slightly smaller than the standard errors from weights without the adjustment. The largest increase in standard error is for those currently not receiving general assistance (10% increase) at the state level and those not covered by Indian Heath service, Tribal Health Program or Urban Indian Clinic (7%) at under 100 percent poverty.

Measuring the size of the difference between estimates is of limited value because it does not reflect the magnitude of the sampling error. A statistic that measures both the bias and sampling error is the bias ratio. A bias ratio of 0.4 is large enough to reduce a nominal confidence interval from 95 percent to about 93 percent. Although we cannot compute a bias ratio because neither estimate is unbiased, we can observe the bias reduction with respect to the size of the standard error. We computed this "bias" ratio as $\hat{b}r = \left(\hat{p}_{NT} - \hat{p}_{RDD}\right)/ se\left(\hat{p}_{RDD}\right)$ shown in Table 4 columns 6 and 11. There are 32 estimates (47%) with bias ratios greater than 0.4 at the state level and 34 estimates (50%) for adults under 100 percent poverty.

In the last step of the evaluation, we determined the number of differences that are statistically different from zero. There were 46 differences (68%) that were significant at the state level and 56 (82%) for adults under 100 percent poverty. Although the estimates and standard errors are very similar, they are statistically different because the estimates are computed using the same data and the weights are highly correlated as shown in Table 5.

5. Conclusions

There is a "cost" associated with explicitly adjusting the weights for telephone coverage in RDD surveys. Some of factors to be considered are: the survey's subject matter, the additional burden to respondents, the impact on questionnaire development, the possibility of higher nonresponse from adding more sensitive items, the availability of data from similar surveys. Even in cases where these issues have been addressed, it is difficult to anticipate how well the adjustment and/or method will perform. For CHIS 2001, the modified poststratification method was used. Most of the variables shown to be important in previous studies when modeling the telephone propensity could not be used because of lack of consistency between the estimates.

When evaluating the effect of the nontelephone adjustment, we observed very small differences between estimates computed using telephone adjusted and unadjusted weights. The bias in most estimates in this study is reduced as expected; however, these reductions are very small. At the same time, there are only marginal increases in the variability of the estimates. Although the differences are small, more than half of the estimates are statistically different because the precision is very high (large sample sizes). The small differences between the estimates may be the result of the small percentage of households without telephones (2% in

California based on the Census 2000), in combination with the fact that differences between adults in households with and without a telephone in terms of the variables being studied are not large. Our findings tend to confirm previous results (Anderson et al., 1988), that is differences in health–related variables between respondents with and without telephone are small. Although we found small differences in the adjusted and unadjusted estimate, we cannot determine if the estimates are unbiased or if the adjustment was able to remove only a small part of the bias.

6. References

Anderson, J. E., Nelson, D. E., and Wilson, R. W (1998). Telephone coverage and measurement of health risk indicators: data from the National Heath Interview Survey, *American Journal of Public Health*, 88(9), 1392-1395.

Brick, J.M., Waksberg, J., and Keeter, S. (1996). Using data on interruptions in telephone service as coverage adjustments, *Survey Methodology*, 22(2), 185-197.

Brick, J.M., Shapiro, G., Flores-Cervantes, I., Ferraro, D., Strickler, T. (1999a). *1997 NSAF snapshot survey weights* (Report No. 3). Washington, DC: Urban Institute.

Brick, J.M., Flores-Cervantes, I., Wang, K., and Hankins, T. (1999b). Evaluation of the use of data on interruptions in telephone service. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 376-381.

Deville, C., Särndal, C. and Sautory, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, 88, 1013-1020.

Ferraro, D. and Brick, J.M. (2001). Weighting for nontelephone households in RDD surveys. *2001 Proceedings of the American Statistical Association, Survey Research Methods Section* [CD-ROM], Alexandria, VA. American Statistical Association.

Frankel, M.R., Srinath, K.P., Battaglia, M.P., Hoaglin, D.C., Wright, R.A., and Smith, P.J. (1999). Reducing nontelephone bias in RDD surveys. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 934-939.

Kass, G. 1980. An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics* 29, 119-127.

Keeter, S. (1995). Estimating noncoverage bias from a phone survey. *Public Opinion Quarterly*, 59, 196-217.

Lundström, S. and Särndal, C-E. (1999). Calibration as a standard method for treatment of nonresponse. *Journal of Official Statistics*, 15, 305-327.

Table 1. Number of adult completed interviews with interruptions in telephone service

Interruption of service	Count	Percentage
No interruption	53,332	96.22
1 day	226	0.41
More than 1 day to 1 week	891	1.61
More than 1 week to 1 month	589	1.06
More than 1 month	355	0.64
Unknown	37	0.07
Total	55,430	100.00

Table 2. Common variables between the CHIS 2001 and the March 2001 CPS

Included in form the model?	Variables used in cell creation?	Variable
✓	✓	Self-reported race and ethnicity
		Adult education level in household, based on high school education
		Poverty level, less than or greater than 100
		Insurance indicator, if anyone in household is insured
✓	✓	Household receiving aid from the AFDC program
✓	✓	Household receiving public housing subsidies
		Household participates in MEDICAL program
		Household participates in MEDICAID program
✓		Number of persons in the household
✓	✓	Number of adults in the household
✓	✓	Number of children in the household
✓		Number of teens in the household

Table 3. Nontelephone adjustment cell definition for CHIS 2001

Person type	Telephone propensity	AFDC participant or receiving public housing assistance	Number of children	Number of adults	Race/ethnicity	Raking factor
Adult	Low	Yes	0 or 1			1.22
Adult	Low	Yes	2 or more			1.30
Adult	High	No			Latino or Black non-Latino	1.09
Adult	High	No			Other non-Latino	1.01
Child	Low	Yes	0 to 2			0.96
Child	Low	Yes	3 or more			1.33
Child	High	No			Latino	0.91
Child	High	No			Non-Latino	1.09
Teen	Low	Yes				1.31
Teen	High	No		0 or 1		0.72
Teen	High	No		2 or more		1.04

Table 4. Statistics for 68 estimates at the state level and 100 percent poverty

	State level					Under 100 percent poverty level				
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Statistic	$\hat{p}_{RDD}$	$\hat{p}_{NT}$	$\hat{b}$	$\hat{r}_{se}$	$\hat{b}_r$	$\hat{p}_{RDD}$	$\hat{p}_{NT}$	$\hat{b}$	$\hat{r}_{se}$	$\hat{b}_r$
Mean	62.455	62.374	-0.081	0.996	0.485	54.100	54.035	-0.065	0.993	0.401
Standard deviation	27.700	27.655	0.151	0.040	0.866	30.458	30.328	0.436	0.032	0.594
Minimum	3.398	3.400	-0.654	0.712	0.003	2.607	2.608	-1.548	0.850	0.002
5th percentile	8.877	8.884	-0.404	0.983	0.008	6.848	6.825	-1.260	0.933	0.020
25th percentile	42.530	42.555	-0.105	0.995	0.096	26.641	26.789	-0.207	0.983	0.061
50th percentile)	67.842	67.756	-0.045	0.999	0.210	59.953	59.693	-0.032	0.997	0.240
75th percentile	86.632	86.554	0.002	1.005	0.458	81.815	81.557	0.132	1.007	0.478
95th percentile	96.336	95.717	0.114	1.015	2.696	95.352	95.278	0.509	1.038	1.751
Maximum	98.687	98.592	0.132	1.103	5.061	98.596	98.585	0.844	1.065	3.227

Table 5. Coefficient of variation of adult weights

	Coefficient of variation			Correlation between
	w_{RDD}	w_{NT}	Ratio	weights
State total	96.56	96.64	1.00	0.998
Less than 100 percent poverty level	105.98	104.87	1.01	0.994