Survey Methodology, December 2001
Vol. 27, No. 2, pp. 121-130
Statistics Canada, Catalogue No. 12-001

Using Propensity Scores to Control Coverages Bias in Telephone Surveys

Kristin Blenk Duncan and Elizabeth A. Stasny [1]

Abstract

Telephone surveys are a convenient and efficient method of data collection. Bias may be introduced into population estimates, however, by the exclusion of nontelephone households from these surveys. Data from the U.S. Federal Communications Commission (FCC) indicates that five and a half to six percent of American households are without phone service at any given time. The bias introduced can be significant since nontelephone households may differ from telephone households in ways that are not adequately handled by poststratification. Many households, called "transients", move in and out of the telephone population during the year, sometimes due to economic reasons or relocation. The transient telephone population may be representative of the nontelephone population in general since its members have recently been in the nontelephone population.

This paper develops a weighting adjustment for transients in an effort to reduce the bias due to noncoverage while controlling the increase in variance due to weighting. We use a logistic regression model to describe each household's propensity for transience, using data collected from a survey of distressed and non-distressed regions of Kentuckiy, Ohio, and West Virginia. Weight adjustments are based on the propensity scores. Estimates of the reduction in bias and the error of estimates are computed for a number of survey statistics of interest, using the propensity based weight adjustments and several alternative weight adjustments. The error in adjusted estimates is compared to the error of the standard estimate to assess the effectiveness of the adjustment.

Key Words: RDD survey; Weight adjustments; Non-sampling error.

1. Introduction

The telephone is a standard mode of communication in today's world, and hence it is extremely useful for conducting surveys. Telephone surveys have come into use more and more as a growing percentage of people have phone connections. Most people who belong to the population that a survey seeks to make inferences about, the survey's target population, can be reached by phone. Therefore, the sample is drawn from the set of all people in households reachable through residential phone numbers. However, this sampling frame excludes all the people without telephone service who may compose a significant portion of some populations. It is currently estimated that in the United States, five and a half to six percent of households are without telephone service an any given time (Belinfante 2000). People without phone service tend to be different from people with service, particularly with regards to economic factors (Smith 1990). Results of the survey will not truly reflect the entire population if these differences are significant on matters of importance to the survey. The coverage bias is particularly troublesome in surveys that examine subgroups of the population with lower telephone penetration rates. These groups include people in lower income households and people who have not obtained a high school degree.

Poststratification on demographic variables associated with telephone coverage is helpful for reducing the coverage bias, but it does not completely solve the problem (Massey and Botman 1988). Another way to account for this coverage bias is to let people who are currently without telephone service be represented by people in the survey who have not had continuous service recently. People whose phone status has changed within the last year are referred to as transients. Transients move in and out of the telephone population, possibly for economic reasons, or service interruptions during relocation. Transients who currently have phone service may be good representatives of the nontelephone population because they are included in the sampling frame, yet they have recently been part of the nontelephone population.

A weighting adjustment suggested by Brick, Waksberg and Keeter (1996) uses transients in the sample to represent the nontelephone population. They use data from the U.S. Current Population Survey (CPS) to estimate unbiased weighting class adjustments for the transient respondents in their survey. Frankel, Ezzati-Rice, Wright and Srinath (1998) also employ this weighting class adjustment, and consider two similar adjustments. Brick, Flores Cervantes, Wang and Hankins (1999) and Frankel, Srinath, Battaglia, Hoaglin, Wright and Smith (1999) evaluate these adjustments using surveys that ask questions about telephone service, but that are not subject to telephone coverage bias. These studies found that employing weight adjustments based on transient status generally led to improved estimates.

This article studies an alternative method for computing a transient weight adjustment. Our method develops a model for predicting transience using demographic variables. The weight adjustment is then based on the

1. Kristin Blenk Duncan and Elizabeth A. Stasny, Department of Statistics, Ohio State University, Columbus, OH 43210-1247.

respondent's propensity for transience. We also compare our propensity method to the method suggested by Brick *et al*. (1996), and to response probability method where the weight adjustment is based on the length of interruption in telephone service.

We use data from the Appalachian Poll, an RDD telephone survey conducted by the Ohio State University's Center for Survey Research during June and July of 1999. The survey was sponsored by *The Columbus Dispatch*, and compared distressed and non-distressed regions of Kentucky, Ohio, and West Virginia. The study gathered information on quality of life issues and perceptions about the Appalachian regions, and also posed a series of standard demographic questions. A stratified sample was used, and just over 400 surveys were completed from each of the six strata (Appalachian and non-Appalachian regions of Ohio, Kentucky, and West Virginia). The poll targeted English speaking adults, 18 years of age or older, residing in the three states. Coverage bias is of particular concern in this survey since telephone coverage rates are lower than usual in the distressed Appalachian regions.

In section 2, we report on the literature describing telephone and transient populations. In this section we also explore differences between these groups in our data, illustrating the concern about coverage bias. Section 2 ends with our proposed model for predicting transience. Section 3 details the various weighting procedures. In section 4 we discuss the trade-off between bias reduction and increased variance from adjusted weights, and compare the weighting schemes. The final section summarizes the findings.

2. Nontelephone and Transient Telephone Populations

The target population for a telephone survey can be categorized by telephone status into four groups: continuous service households, transient households which are currently with service, transient households which are currently without service, and chronic nontelephone households. We need to know something about the size of each of these groups in order to account fro coverage bias in the survey. Data from the FCC is useful for examining long term trends in the size of the nontelephone population. Not as much is known, however, about the short-term changes in phone coverage.

Keeter (1995) used panel surveys to study the dynamics of the transient phone population. In the March 1992 and 1993 CPS, it was found that 94.1% of households in the sample at both times had a phone at both time points, 2.6% at neither point, and 3.4% had a phone at one interview, but not the other. Fifty-seven percent of respondents who reported having no phone at either interview were transient. If the measurements could be taken continuously, rather than at two points in time, even more households would be labeled transient. Keeter concludes that, "a sizable minority of nontelephone households, at the least, have recently been in the telephone population or are soon to join it. Such transient households constitute a measurable segment of telephone households and thus can provide data to characterize the nontelephone population," (Keeter 1995, page 201). The same article asserts that, "Transient telephone households are much more like nonphone households than those with continuous service," (Keeter 1995, page 209). This conclusion is based on formal tests using demographic variables from the CPS. Data from the National Survey of America's Families presented in Brick *et al.* (1999) supports Keeter's findings. Since transients make up a nontrivial proportion of the nontelephone population and transients are more similar to the nontelephone households than they are to continuous service households, it is reasonable to use data from the transients in the sample to attempt to reduce coverage bias.

In the Appalachian Poll, 140 of the 2,463 respondents, or 5.7%, replied positively to the question, "During the last twelve months has your household ever been without telephone service for one week or more?" These respondents are categorized as transients. In the Appalachian regions, the transience rate is 7.4% while the rate is only 3.9% in non-Appalachian regions.

Table 1 compares transient and nontransient households from the sample in regards to selected variables. The large differences between the two populations illustrate the need for bias reduction. People who live in transient households are much younger, have lower incomes, and they are less likely to be employed full time. They also have less access to health insurance and computers.

Table 1
Selected Characteristics of Nontransient and Transient Households

Characteristics	Nonstransient	Transient
Median Age	47.0	37.5
Household income Less than $20K	27.8%	60.0%
Employed full-time or retired	55.0%	34.5%
No health insurance	12.7%	30.0%
Owns or is buying residence	79.4%	61.4%
Computer in home	47.4%	26.4%
Not enough money for food	12.3%	42.9%

Note: Statistics are based on unweighted frequencies in the sample which oversampled from the Appalachian regions, and thus are not representative of population quantities.

A model for transience. Using the Appalachian Poll sample, we develop a logistic regression model to predict transience with demographic variables. The independent variables used to predict transience are age, employment status, race, income, and region. The model is described in the Appendix. Education and tenure are also good predictors of transience, but they are strongly correlated with the other variables in our model, and thus, we chose not to include them. For a comparison of models that predict telephone coverage, see Smith (1990). We will use our model in the propensity weighting adjustment described in the following section.

3. Weight Adjustments

We consider several weighting schemes that attempt to account for the coverage bias inherent in telephone surveys. Each of these schemes is compared to the actual weighting procedure used for the Appalachian Poll. In the standard procedure, a base weight was calculated for each respondent. This adjustment is (# adults in household) / (# voice telephone lines), or the inverse of the respondent's probability of being in the sample. Then weights were raked in each of the six strata to agree with 1990 Census proportions for age group, education level, and gender. Finally, the weights were scaled to the sample sizes within the six strata.

3.1 Length of Disconnect

Respondents to the Appalachian Poll who replied "yes" to the question about an interruption in phone service of one week or longer were then asked how many days they were without service in the last year. A simple approach to the coverage bias problem is to give transients a weight adjustment inversely proportional to the fraction of the year that they were with service. For example, a person who has only had service for six months out of the last twelve receives a weight of two, thus representing himself and one other person in the population with a six-month disconnect who is currently without service.

This naïve approach is included in the analysis for comparison with other schemes. It is referred to as the day scheme (DAY). Weight adjustments are calculated as 365 / (365 – # days without service). This weight adjustment is applied after the base weight described above, and before the weights are raked.

While this approach is logical, it is not practical for controlling variance. It is usually considered undersirable to use weighting factors larger than three. In fact, for many large surveys conducted by the U.S. Census Bureau, if weighting factors are larger than two, respondents are merged into larger groups and a group weight is calculated in order to obtain lower weighting-adjustment factors; see, for example, CPS (1978).

This simple approach becomes more practical when respondents are grouped by the length of their interruption in service. In a scheme called day group (DAYG), transients are grouped into quartiles across the entire sample by length of interruption in phone service. These quartiles correspond to interruptions of one week, more than one week but less than three weeks, three weeks to two months, and more than two months. The weight adjustment for each group is 365 / (365 – avg. # days without service), and it is also applied after the base weight, prior to raking. This grouping procedure is helpful for reducing the variance caused by extremely long interruptions.

3.2 Weighting Class Adjustment Scheme

Brick *et al.* (1996) also implement a response probability adjustment to reduce coverage bias. Under their procedure, they partition the target population into the four components described in section 2: t_1 is the number of persons living in continuous service households; t_2 is the number of persons living in transient households that currently have service; t_3 is the number of persons living in nontelephone households that have not had any service in the last year; and t_4 is the number of persons living in transient households that are currently without service. The response probability model the authors use assumes that $t_3 = 0$. With this assumption, an unbiased weight adjustment is $\mathrm{A} = (t_2 + t_4)/t_2 = 1 + (t_4/t_2)$, the inverse of the proportion of the transient population that currently has service. Unfortunately, these population quantities are unknown and must be estimated. Following the lead of Brick *et al.*, we use CPS data to estimate $t_1 + t_2$, the number of persons who currently have service, and t_4; call these estimates $\hat{t}_1 + \hat{t}_2$ and $\hat{t}_4$, respectively. From the Appalachian Poll, separate estimates of t_1 and t_2 are available; designate these estimates as t_1^* and t_2^*, respectively. Since the estimates come from different surveys, ratios are used in the weight adjustment, and A is estimated by

$$\mathrm{A}' = 1 + \frac{\dfrac{\hat{t}_4}{\hat{t}_1 + \hat{t}_2}}{\dfrac{t_2^*}{t_1^* + t_2^*}}. \tag{1}$$

Some persons are more likely to live in nontelephone households than others, so Brick *et al.* classified transients into cells based on characteristics associated with not having a telephone, and computed the weight adjustment for each cell. Four classification schemes, which categorized respondents by either education or tenure, length of interruption, and race/ethnicity were considered.

Brick *et al.* found schemes that classified respondents as transients if they had an interruption of one week or more to be superior to schemes that used a cut-off of one month, so for the Appalachian Poll data we use the one-week cut-off. Due to the small number of Hispanics in the Appalachian Poll sample, we do not categorize by ethnicity. Thus, for our analyses, the cell classifications for two schemes that use the method described by Brick *et al.* (1996) are defined as follows:

> **BWKE** – households that had a service interruption of one week or more within categories defined by education (less than high school, high school diploma, college diploma or above) and race (black, non-black); and

BWKT – households that had a service interruption of one week or more within categories defined by tenure (own/other, rent) and race.

The disadvantage of using these schemes in our study is that the estimates needed from the CPS are available by state, but not by region since the CPS does not sample from all counties. Persons in Appalachian regions are less likely to have telephones, but we cannot account for this with the available CPS data. Even when we consider statewide data, the sample size of the CPS is not large enough to get reliable values of $\hat{t}_4$ in all of the cells. For example, in 1999 the CPS did not sample any blacks with a college degree or higher who live in Kentucky and do not have telephone service. Thus, the weighting cell adjustments computed for use with the Appalachian Poll are based on CPS data from the three states combined.

3.3 Raking Ratio Adjustment

Lohr (1999) explains the use of raking ratio estimates to adjust for nonresponse in surveys. We propose a similar use of raking to account for coverage bias. We estimate the proportion of the population with continuous telephone service, and then use raking to allow transients in the sample to represent the portion of the population without continuous telephone service.

The percent of households without continuous service is estimated by

$$1 - \left(\frac{\tilde{t}_1 + \tilde{t}_2}{\tilde{t}_1 + \tilde{t}_2 + \tilde{t}_4} \right) \left(\frac{t_1^*}{t_1^* + t_2^*} \right), \tag{2}$$

where $\tilde{t}_i$, $i = 1, 2, 4,$ is obtained from the FCC data. The first fraction estimates the proportion of households that currently have service, and the second fraction estimates the proportion of nontransient households among households with service. Again, we assume that $t_3 = 0$. The FCC gives telephone penetration rates by state, but not by region. Data from the 1990 Census does give penetration rates by county, but rates changed from 1990 to 1999. Therefore, to estimate the 1999 regional penetration rate, we maintained a constant ratio of percent of households without a phone in the non-Appalachian regions to percent of households without a phone in the Appalachian regions and adjusted the 1990 Census regional rates to match the 1999 state rates. Table 2 gives the data we used to compute the 1999 state rates, and the resulting estimates.

In a scheme referred to as transient raking, or TRAK, transient status is included as a control variable for raking along with age, gender, and education level. The totals we used for raking by transient status are given in Table 2.

3.4 A New Propensity Weighting

An estimated propensity score is sometimes used to create a weight adjustment to account for nonresponse in surveys where some variables are known for the non-respondents. For example, in a face-to-face household interview the interviewer knows the address of the nonrespondent and may have information about the person's race, gender, and age. A logistic regression model that describes propensity for response is developed, and respondents are assigned a weight of $1/\hat{p}$, where $\hat{p}$ is the estimated propensity to respond (Little and Rubin 1987). This procedure gives higher weights to sampled households that are more similar to the nonrespondents. Since there is typically no data on the excluded nontelephone population in telephone surveys, a modified approach is taken to using a propensity score. We only adjust the weights for the transients since they will represent the missing part of the sample: weights for nontransients remain unadjusted. The

Table 2
Computation of Transient Status Raking Totals

	Kentucky		Ohio		West Virginia	
	Ap.	Non-Ap.	Ap.	Non-Ap.	Ap.	Non-Ap.
Appalachian Poll Data						
Sample Size	412	407	413	405	411	415
# transients in sample	38	19	18	13	36	16
Percent of sample without cont. service	9.2	4.7	4.4	3.2	8.8	3.9
Census and FCC Data						
1990 State % no phone	10.2	10.2	4.7	4.7	10.3	10.3
1990 Region % no phone	19.1	8.2	11.7	4.5	14.3	8.4
1999 State % no phone	6.7	6.7	5.2	5.2	7.3	7.3
Percent of state pop. living in region	18.6	81.4	2.6	97.4	31.8	68.2
Estimates						
Ratio of Non-Ap to Ap noncoverage	0.429	0.429	0.385	0.385	0.587	0.587
Estimated 1999 region % no phone	12.5	5.4	13.0	5.0	10.1	6.0
Estimated % of pop. without cont. service	20.6	9.8	16.7	8.1	18.0	9.6
Desired # of transients in sample	85	40	69	33	74	40

weight adjustment for transients is $1/(1-\hat{p})$, where $\hat{p}$, the estimated propensity for transience, is described by the model in section 2.1. Households with a higher estimated propensity for transience may be more representative of the nontelephone population and they receive higher weight adjustments. This adjustment is applied to the base weight, and the scheme is called propensity (PROP).

Transience is not that common, and most estimated propensity scores are fairly low. In the PROP scheme, the average weight adjustment for a transient household is 1.167. This adjustment is not large enough for transients to represent themselves and the entire nontelephone population. That is, when the weights are scaled to sum to the population size, the sum of the final weights for transients is less than the size of the transient population. To account for this under-representation, the propensity weight adjustment is applied, and then transient is used as a control variable for raking along with age, education, and gender. The estimated population sizes for transients are computed as in section 3.3. This weighting scheme is called augmented propensity, or AUGP.

4. Findings

The analysis and comparison of the adjustment schemes presented here parallels the analysis performed by Brick *et al.* (1996). We first discuss the change in variance resulting from adjusting the weights to reduce coverage bias and present a statistic for measuring the relative variability. Then, the schemes are evaluated by comparing the variance of adjusted estimates to the mean squared error of the standard estimate.

4.1 Changes in Variability

The goal of the adjustment schemes is to decrease coverage bias while controlling variance. Adjustment of the weights to reduce the bias increases the variability of the weights, hence increasing the variance of the estimates. Kish (1992) gives a formula for measuing the increase in variance due to unequal weights. Brick *et al.* (1996) refer to this expression as the variance inflation factor (VIF). The VIF can be written as

$$\mathrm{VIF} = 1 + [\mathrm{CV(weights)}]^2, \qquad (3)$$

where CV(weights) is the coefficient of variation of the weights. A VIF ratio is computed to compare the VIF of a new weighting scheme to that of the standard weighting scheme. Table 3 gives VIF ratios for the six strata in the Appalachian Poll data under each scheme described in section 3. A VIF ratio if 1.12, for example, indicates an increase in variance of 12 percent over the variance using the standard weighting scheme. The VIF ratio values are reasonable for all schemes except the DAY scheme which sees an average variance increase of 300 percent. The VIF ratio values for our PROP scheme are all very close to one, suggesting that the PROP weight adjustments will not increase the variance of our estimates.

4.2 Coverage Bias Reduction

Estimates of seventeen population proportions using survey variables from the Appalachian Poll were calculated for the standard weighting procedure and for each of the seven adjustment schemes (see Table 4 for a list of the seventeen variables). WesVar software was used to calculate standard errors for these estimates by means of replication. We would like to assess the effectiveness of each scheme for reducing the coverage bias on these seventeen characteristics. Estimates from an independent source that are free of telephone coverage bias would be ideal for such an assessment. Unfortunately, such benchmarks are unavailable and some model assumptions are necessary in order to perform an evaluation. We assume that the weight adjustment procedures reduce the coverage bias. Thus the difference between the standard estimate and the adjusted estimate is considered to be an unbiased estimate of the decrease in coverage bias resulting from the adjustment. The assumption favors the adjusted estimates, considering them to be unbiased.

Table 3
Ratios of Variance Inflation Factor Due to Weight Adjustment

Region	Ratio of scheme's VIF to standard weight's VIF						
	DAY	DAYG	BWKE	BWKT	TRAK	PROP	AUGP
Non-Appalachian Ohio	0.999	0.997	1.004	1.023	1.063	0.999	1.061
Appalachian Ohio	1.480	1.016	1.039	1.091	1.331	0.999	1.336
Non-Appalachian Kentucky	4.151	1.040	1.018	1.054	1.030	0.999	1.029
Appalachian Kentucky	2.433	1.069	1.045	1.042	1.129	1.003	1.145
Non-Appalachian West Virginia	6.331	1.027	1.010	1.029	1.020	0.999	1.024
Appalachian West Virginia	2.935	1.085	1.058	1.053	1.116	1.005	1.119
Scheme Average	3.055	1.039	1.029	1.049	1.115	1.001	1.119

Table 4
Estimated Reduction in Bias and Bias Ratio for Selected Characteristics

Characteristic	Standard estimate		Estimated reduction in bias							Bias Ratio						
	Estimate	St. error	DAY	DAYG	BWKE	BWKT	TRAK	PROP	AUGP	DAY	DAYG	BWKE	BWKT	TRAK	PROP	AUGP
Owns Home																
Non-Appalachian Ohio	72.2	3.1	0.6	0.5	0.5	1.2	1.4	0.1	1.6	0.2	0.2	0.2	0.4	0.5	0.0	0.5
Appalachian Ohio	75.4	2.8	4.4	0.6	0.6	2.1	3.2	0.3	3.5	1.6	0.2	0.2	0.8	1.1	0.1	1.2
Non-Appalachian Kentucky	68.6	3.1	7.2	0.8	0.9	1.8	1.5	0.2	1.5	2.3	0.3	0.3	0.6	0.5	0.1	0.5
Appalachian Kentucky	80.5	2.2	2.9	0.8	0.3	1.3	0.3	0.0	0.3	1.3	0.3	0.1	0.6	0.1	0.0	0.1
Non-Appalachian West Virginia	80.0	2.3	14.2	1.6	0.9	1.9	1.4	0.2	1.4	6.1	0.7	0.4	0.8	0.6	0.1	0.6
Appalachian West Virginia	81.9	2.2	8.2	0.7	−0.4	0.5	−0.3	0.0	−0.2	3.7	0.3	−0.2	0.2	−0.1	0.0	−0.1
No Health Insurance																
Non-Appalachian Ohio	7.3	1.7	0.0	−0.1	−0.6	−1.4	−1.7	−0.1	−1.8	0.0	−0.1	−0.4	−0.8	−1.0	−0.1	−1.1
Appalachian Ohio	12.6	2.1	0.9	0.1	0.3	0.3	0.5	0.1	0.6	0.4	0.1	0.1	0.2	0.3	0.0	0.3
Non-Appalachian Kentucky	8.8	1.8	1.8	0.4	0.2	0.3	0.0	0.1	0.1	1.0	0.2	0.1	0.2	0.0	0.0	0.0
Appalachian Kentucky	22.2	2.4	3.4	0.1	−0.1	−0.2	−0.8	−0.4	−1.5	1.4	0.0	0.0	−0.1	−0.3	−0.2	−0.6
Non-Appalachian West Virginia	14.2	2.1	−4.8	−0.5	−0.7	−1.0	−1.2	−0.3	−1.4	−2.3	−0.2	−0.3	−0.5	−0.6	−0.1	−0.7
Appalachian West Virginia	24.6	2.5	2.5	−0.8	−1.7	−1.3	−2.7	−0.6	−3.0	1.0	−0.3	−0.7	−0.5	−1.1	−0.2	−1.2
Not enough Money for Food																
Non-Appalachian Ohio	10.8	1.9	−0.7	−0.6	−0.9	−1.6	−2.2	−0.1	−2.1	−0.4	−0.3	−0.5	−0.9	−1.2	0.0	−1.2
Appalachian Ohio	16.2	2.5	−4.7	−0.8	−0.6	−1.3	−3.3	−0.2	−3.4	−1.9	−0.3	−0.3	−0.5	−1.3	−0.1	−1.4
Non-Appalachian Kentucky	11.4	2.4	−3.3	−0.8	−1.3	−1.7	−1.6	−0.4	−1.8	−1.4	−0.3	−0.5	−0.7	−0.7	−0.2	−0.8
Appalachian Kentucky	20.2	2.4	−7.4	−2.3	−2.1	−2.1	−3.8	−0.4	−3.8	−3.1	−1.0	−0.9	−0.9	−1.6	−0.2	−1.6
Non-Appalachian West Virginia	14.0	2.1	4.3	−0.1	−1.0	−1.4	−1.7	−0.3	−1.8	2.1	0.0	−0.5	−0.7	−0.8	−0.2	−0.9
Appalachian West Virginia	16.4	2.0	1.5	−0.7	−1.0	−0.9	−2.2	−0.5	−2.6	0.8	−0.3	−0.5	−0.4	−1.1	−0.3	−1.3
Computer in Home																
Non-Appalachian Ohio	60.1	3.0	0.4	0.3	0.6	1.2	1.3	0.1	1.4	0.1	0.1	0.2	0.4	0.5	0.0	0.5
Appalachian Ohio	40.0	3.0	1.2	0.2	0.3	0.8	1.8	0.1	2.0	0.4	0.1	0.1	0.3	0.6	0.0	0.7
Non-Appalachian Kentucky	44.5	3.0	6.7	0.9	0.8	1.1	0.9	0.2	1.0	2.3	0.3	0.3	0.4	0.3	0.1	0.3
Appalachian Kentucky	29.7	2.3	1.9	1.0	0.9	1.1	2.3	0.0	1.9	0.8	0.4	0.4	0.5	1.0	0.0	0.8
Non-Appalachian West Virginia	46.2	2.6	7.6	0.6	1.1	1.2	1.5	0.3	1.6	2.9	0.2	0.4	0.4	0.6	0.1	0.6
Appalachian West Virginia	36.1	2.7	4.3	1.0	0.3	0.4	0.2	0.3	0.5	1.6	0.4	0.1	0.2	0.1	0.1	0.2
Summary of Seventeen Variables																
Mean absolute value			0.032	0.005	0.006	0.009	0.013	0.002	0.014	1.396	0.235	0.620	0.412	0.885	0.075	0.885
Median absolute value			0.022	0.005	0.006	0.011	0.014	0.001	0.014	0.995	0.240	0.245	0.420	0.605	0.055	0.665

Note: In addition to the four proportions listed in the table, the summary of seventeen variables includes worry about income, better off economically in the 1990's, dissatisfied with own net worth, married, have children, unemployed, college graduate, in good or excellent health, serious illness in household, no family doctor, satisfied with own housing, very safe drinking water, and internet access in home.

Using our assumption, we compare the estimate from each scheme to the standard estimate. The reduction in coverage bias is estimated by the difference between the standard estimate and the adjusted estimate. There are seven different estimates of the bias reduction, one for each scheme. The estimated reduction in bias is given by

$$b_i = \hat{p}_s - \hat{p}_i \qquad (4)$$

where b_i is the estimated bias reduction using scheme i, $\hat{p}_s$ is the standard estimate, and $\hat{p}_i$ is the estimate from adjustment scheme i. Estimated reductions in bias for four characteristics by the six strata are given in Table 4 for each scheme. For the characteristics owns home, not enough money for food, and computer in home, the direction of the bias is fairly consistent across schemes and regions. Reassuringly, the bias is in the expected direction for these characteristics, with fewer people owning homes, more people not having enough money for food, and fewer people having computers in their homes, than is indicated by the estimates using the standard weighting scheme. For health insurance, the direction of the bias is mostly consistent across regions. The standard estimate is biased upward for Appalachian Ohio and non-Appalachian

Kentucky, and generally biased downward in the other regions.

The absolute size of the reduction in bias by itself is not fully meaningful, because it does not account for the amount of sampling error associated with the estimate. Therefore, we also calculate the bias ratio, as in Brick *et al.* (1996). The bias ratio for scheme i, r_i, is given by

$$r_i = \frac{b_i}{\mathrm{se}(\hat{p}_s)}, \tag{5}$$

where $\mathrm{se}(\hat{p}_s)$ is the standard error of the standard estimate. Table 4 also gives the bias ratio for the selected estimates. DAY, TRAK, and AUGP give the largest bias ratios; for these adjustment schemes the bias is not negligible when we consider the standard error. DAYG and PROP have low bias ratios, indicating that the bias reduction is small compared to the error of the estimate.

4.3 Mean Square Error

Since the standard estimates are though to be biased, error should be measured with mean square error rather than variance. The MSE of the standard estimate is approximated by

$$\mathrm{mse}_i = \mathrm{var}(\hat{p}_s) + b_i^2 \tag{6}$$

for each adjustment scheme. Recall that we are assuming the adjusted estimates are unbiased, so that the mean square errors of these estimates are equal to their variances. The variance of the these estimates are equal to their variances. The variance of the adjusted estimates can be approximated by two methods. The first approximation is obtained by multiplying the VIF ratio in Table 3 by the variance of the standard estimate. Alternatively, we can us the variance of the adjusted estimate obtained from replication methods.

The error of the adjusted estimate is compared to the error of the standard estimate in the mean square ratio (MSR). Using the VIF variance, the estimated MSR is given by

$$\mathrm{msr}_{\mathrm{VIF}i}(\hat{p}) = \frac{100 \times \mathrm{VIF\ Ratio}_i \times \mathrm{var}(\hat{p}_s)}{\mathrm{mse}_i(\hat{p})}. \tag{7a}$$

For the replication variance, the estimated MSR is given by

$$\mathrm{msr}_{\mathrm{VAR}i}(\hat{p}) = \frac{100 \times \mathrm{var}_i(\hat{p})}{\mathrm{mse}_i(\hat{p})}, \tag{7b}$$

where $\mathrm{var}_i(\hat{p})$ is the estimated variance of the adjusted estimate, obtained through replication. An MSR of 100 indicates that the variance of the adjusted estimate is exactly equal to the mean squared error of the standard estimate. An MSR above 100 means the variance of the adjusted estimate is larger than the MSE of the standard estimate, and the bias/variance trade-off for the scheme is not favorable. An MSR below 100 means that the adjusted estimate is an improvement over the standard estimate in terms of overall error.

Table 5 gives estimated MSR values for selected survey variables from the Appalachian Poll, and a summary of these values for seventeen variables from each adjustment scheme. The MSR estimates vary between regions and between schemes. The msr values computed using the two different variances also differ, but the summary values are similar for both variances. The DAY scheme has the highest msr values, indicating that this weight adjustment is not worthwhile because it increases the variance too much. TRAK and AUGP have the lowest mean and median msr values, though these schemes produced unfavorable estimates for a few characteristics as indicated by the high maximum msr values. The weighting class adjustment schemes BWKE and BWKT performed well and their maximum estimated mean square ratio values are fairly low. All of the msr values for the PROP scheme are near 100, suggesting that the overall error in estimates computed with this scheme is comparable to the error in the standard estimates.

5. Conclusions

While telephone use is commonplace, telephone surveys will always contain some bias since nontelephone households are excluded from the sampling frame, and the nontelephone population has characteristics that differ from those of the telephone population. Coverage bias is alleviated by poststratification on variables such as income and education and may not be a problem in some instances. However, for surveys that target poor or rural areas where telephone penetration rates are lower, the coverage bias is a large concern.

We have proposed a few new methods for reducing the coverage bias by adjusting the weights of respondents in the transient population. We compared the resulting estimates to those from other existing methods. In the analysis of these methods, it was assumed that the adjusted estimates are unbiased. In the absence of unbiased benchmark estimates this assumption cannot be validated. The mean square ratios presented here are likely to be biased downward since the bias of the adjusted estimate is not included. The estimated MSR is still useful for comparing methods, however, and gives a good measure of the effectiveness of the weight adjustments.

As anticipated, the DAY method was found to have too much variability to be useful. The day group (DAYG) method appears to perform better, but most of the mean square ratios for this scheme are close to 100, meaning that we do not see a large improvement over the standard estimate. The advantage of this scheme lies in its simplicity. The weight adjustment is easy to apply and does not require auxiliary data.

Table 5
Mean Square Ratio for Selected Characteristics

Characteristic	VIF Mean Square Ratio							Variance Mean Square Ratio						
	DAY	DAYG	BWKE	BWKT	TRAK	PROP	AUGP	DAY	DAYG	BWKE	BWKT	TRAK	PROP	AUGP
Owns Home														
Non-Appalachian Ohio	96.1	97.2	97.3	88.1	87.5	99.8	84.5	98.6	98.2	98.2	88.4	81.8	99.9	78.7
Appalachian Ohio	42.3	97.4	98.7	68.9	57.9	99.1	52.5	71.7	96.5	89.6	71.1	51.6	99.2	46.6
Non-Appalachian Kentucky	63.9	97.6	94.5	77.7	83.1	99.4	83.5	21.9	98.2	92.8	75.7	81.3	98.8	80.1
Appalachian Kentucky	89.3	96.1	102.3	77.0	110.6	100.3	112.4	116.0	100.7	104.1	88.4	119.2	100.0	118.3
Non-Appalachian West Virginia	16.6	71.2	89.1	62.3	75.8	99.2	75.5	28.6	81.1	94.4	71.4	85.5	100.1	84.2
Appalachian West Virginia	20.2	98.4	103.2	100.3	109.3	100.5	110.6	43.5	106.0	101.1	104.8	108.8	99.1	108.9
No Health Insurance														
Non-Appalachian Ohio	99.9	99.0	88.2	61.4	51.9	99.5	48.5	98.8	100.5	112.1	101.7	82.0	101.9	76.1
Appalachian Ohio	126.8	101.3	102.1	106.5	125.1	99.9	123.5	92.3	98.8	95.6	94.0	105.8	99.0	100.4
Non-Appalachian Kentucky	206.4	99.9	100.9	102.8	103.0	99.8	102.7	39.0	87.9	90.7	86.2	97.8	96.6	95.5
Appalachian Kentucky	82.7	106.7	104.4	103.7	102.1	97.9	84.3	53.5	109.9	104.9	105.3	114.1	100.0	100.5
Non-Appalachian West Virginia	100.2	97.1	90.6	84.0	77.7	97.9	71.3	136.6	99.2	94.0	89.7	90.6	100.8	84.3
Appalachian West Virginia	149.6	99.2	74.1	83.5	52.8	95.7	46.7	107.0	96.5	75.1	84.3	51.5	96.7	45.4
Not enough Money for Food														
Non-Appalachian Ohio	86.5	90.5	80.5	57.9	45.2	99.7	45.6	105.2	100.8	104.3	94.5	66.3	102.0	67.0
Appalachian Ohio	31.9	92.9	97.4	86.1	48.6	99.4	46.4	68.5	98.2	96.8	90.2	69.4	101.1	66.4
Non-Appalachian Kentucky	139.1	94.1	78.2	69.5	69.5	96.7	64.3	320.7	96.8	91.9	85.7	77.6	100.1	68.5
Appalachian Kentucky	22.3	55.8	57.6	58.7	31.0	97.2	31.9	30.5	68.4	68.5	69.5	36.8	100.2	38.5
Non-Appalachian West Virginia	117.3	102.6	82.4	71.0	59.6	97.7	57.2	105.7	101.9	94.7	88.3	71.5	101.6	68.5
Appalachian West Virginia	181.6	97.0	84.1	88.5	50.4	94.1	39.9	92.2	98.8	89.6	92.9	59.0	97.5	48.3
Computer in Home														
Non-Appalachian Ohio	98.1	98.5	96.4	88.2	88.1	99.8	86.1	99.5	99.5	102.0	102.1	106.3	100.6	102.8
Appalachian Ohio	127.2	101.2	103.1	101.2	96.2	99.7	92.5	116.0	99.6	101.2	96.5	94.1	99.1	86.5
Non-Appalachian Kentucky	67.7	94.9	94.4	92.7	93.6	99.5	92.8	27.1	93.7	91.5	89.7	90.3	98.2	88.4
Appalachian Kentucky	147.1	89.0	91.7	85.1	55.7	100.3	68.4	58.9	81.1	85.5	79.5	46.8	100.9	66.6
Non-Appalachian West Virginia	66.8	96.9	86.6	85.8	76.1	98.5	73.5	59.6	95.8	85.1	85.3	72.9	98.6	68.6
Appalachian West Virginia	82.7	95.6	104.4	103.0	111.2	99.6	108.2	41.8	88.1	101.6	99.9	113.3	98.3	107.0
Summary of Seventeen Variables														
Mean	137.6	97.5	94.3	92.2	85.2	99.3	83.8	125.2	99.1	97.1	96.8	96.0	100.2	93.5
Median	107.5	99.0	99.1	97.1	89.8	99.8	86.3	94.8	98.9	98.7	98.5	98.0	100.0	92.4
Minimum	10.9	55.8	0.9	57.9	4.1	94.1	5.7	7.0	68.4	43.1	62.1	7.6	94.6	6.0
Maximum	607.7	108.5	104.8	109.1	133.1	100.5	133.5	695.2	140.8	144.5	147.5	593.8	116.7	545.4
Percent below 100	47.1	60.8	61.8	58.8	65.7	87.3	67.6	63.7	62.7	56.9	58.8	53.9	58.8	58.8

Note: In addition to the four proportions listed in the table, the summary of seventeen variables includes worry about income, better off economically in the 1990's, dissatisfied with own net worth, married, have children, unemployed, college graduate, in good or excellent health, serious illness in household, no family doctor, satisfied with own housing, very safe drinking water, and internet access in home.

The weighting class adjustment schemes have the benefit of giving more weight to respondents in cells where the likelihood of having a phone is lower. For these schemes, greater bias reduction was seen in variables correlated with the classification variables. For example, home ownership and computer ownership are positively correlated, and the BWKT scheme, which classified respondents by home ownership, produced estimates of the percent of households with a home computer that were consistently lower than the standard estimates. Table 5 shows that the BWKE and BWKT schemes produce an improved estimate most of the time. It should also be noted that when these schemes produce an estimate that is not an improvement, the increase in variance remains fairly small. The weighting class adjustment method works well for samples of large populations, such as states or countries, since the outside data needed to compute the adjustments is readily available. The method is more difficult to use for very specific samples such as countries.

The ranking ratio adjustment, TRAK, produced a number of very favorable estimated MSR values. With this scheme we were able to account for the differences in telephone penetration rates by region, but not the differences across other demographic characteristics. Variability was introduced when we estimated the regional rates from the state rates, thus, as with the weighting class adjustment, the scheme works better for samples of larger populations. While the mean and median estimated MSR values were low for this scheme, the scheme also produced some high mean square ratios. The higher ratios occurred in Ohio where the percent of transients in the sample was low compared to the estimated percent without continuous service.

The propensity adjustment alone, PROP, provided too little reduction in bias to be worthwhile. The propensity adjustment is advantageous, however, because it allows us to account for differences in the likelihood of having telephone service without using outside data. When used in conjunction with raking, the propensity based scheme AUGP produced good results.

There are many issues to consider when determining which adjustment scheme is preferred. As mentioned previously, the weighting class adjustment schemes BWKE and BWKT are difficult to implement if you have a very specific target population. These schemes are fairly conservative, however, in that they typically reduce the bias without increasing the variance. The schemes that employed raking usually performed better than the weighting class adjustment schemes, but the larger weight adjustments sometimes led to increased variances. It may be advisable to compute estimates using several schemes and then determine which scheme offers the best bias-variance trade-off.

Brick *et al*. (1996) note that these weight adjustments for telephone coverage should be more beneficial in reducing mean squared error when the sample size of the survey is large. As the sample size increases, the bias ratio increases since the bias is unaffected but the standard error of the estimate, which is in the denominator, decreases.

The findings suggested by this study and others indicate that the adjustments could be useful for many estimates from telephone surveys and should be seriously considered. The benefits of adjustment appear to outweigh the penalties in the weighting class adjustment schemes, the raking scheme, and the augmented propensity scheme. In light of the smaller sample size and special target population of the Appalachian Poll, generalizations of these findings should not be made until the methods receive further evaluation. These weight adjustments still need to be tested using a survey that is free of coverage bias, one that includes nontelephone households in the sampling frame and collects information on telephone status, in order to assess the validity of the assumptions. Data from the National Survey of America's Families, or the National Health Interview Survey may be appropriate for evaluating the adjustment methods and the assumptions.

Acknowledgements

This work was supported in part by a fellowship from the Center for Survey Research at the Ohio State University. We thank Dr. Paul Lavrakas and the Center for Survey Research for allowing us to use the Appalachian Poll data. We would also like to thank the referees for their helpful comments.

Appendix

Logistic Regression of Transient Status

Below is our model for predicting transient status. Most of the variables in the model relate to socioeconomic status. The coefficients indicate that young people, those with low income, those who are not employed full-time, American Indians and African Americans, and residents of distressed counties have higher propensities for transience. The high significance level of the Hosmer and Lemeshow test indicates a very good fit of the model. The large area under the ROC curve tells us that the model discriminates well.

Variable Coding

Age
- 0 – "Refused" (Count = 9)
- 1 – 18 to 29 years
- 2 – 30 to 44 years
- 3 – 45 to 59 years
- 4 – over 60

Low Income
- 0 – Household income over $20,000 or refused
- 1 – Household income under $20,000

Employment Status
- 0 – Employed full-time or retired
- 1 – Other (refused, part-time, housekeeper, student, unemployed, other)

Race
- 0 – Caucasian, Alaskan Native, Hispanic, or Asian
- 1 – American Indian, African-American, Black, or other

Appalachian
- 0 – Does not live in a distressed county of KY, OH, or WV

Kentucky/West Virginia
- 0 – Ohio
- 1 – Kentucky or West Virginia

Results

Variables in the Equation

Variable	B	S.E.
Age (Refused)	– 2.107	12.160
Age (18 – 29)	2.006	0.357
Age (30 – 44)	1.664	0.347
Age (45 – 59)	1.064	0.364
Low Income	1.358	0.189
Employment Status	0.397	0.187
Race	1.136	0.292
Appalachian	0.531	0.196
KY/WV	0.567	0.216
Constant	– 5.712	0.401
Hosmer and Lemeshow Goodness of Fit Test		
Chi-Square		3.568
Degrees of Freedom		8
p – value		0.894
ROC Curve		
Area under the Curve		0.782

References

Belinfante, A. (2000). Telephone Subscribership in the United States. Industry Analysis Division, Common Carrier Bureau, Federal Communications Commission, Washington, D.C. 20554.

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

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

Current Population Survey (1978). Current Population Survey: Design and Methodology. Technical Paper 40. Department of Commerce, Bureau of the Census, Washington, D.C.

Frankel, M.R., Ezzati-Rice, T., Wright, R.A. and Srinath, K.P. (1998). Use of data in interruptions in telephone service for noncoverage adjustment. *Proceedings of the American Statistical Association Section on Survey Research Methods*, 290-295.

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 American Statistical Association Section on Survey Research Methods*, 934-939.

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

Kish, L. (1992). Weighting for unequal Pi. *Journal of Official Statistics*, 8, 183-200.

Little, R., and Rubin, D. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons, Inc. 55-60.

Lohr, S. (1999). *Sampling*: *Design and Analysis*. New York: Duxbury Press, 255-287.

Massey, J., and Botman, S. (1988). Weighting adjustments for random digit dialed surveys. In *Telephone Survey Methodology*, (Eds. R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls and J. Waksberg). New York: John Wiley & Sons, Inc. 143-160.

Smith, T. (1990). Phone home? An analysis of household telephone ownership. *International Journal of Public Opinion Research*, 2, 369-390.