Journal of Official Statistics, Vol. 18, No. 2, 2002, pp 203–215

Two-Phase List-Assisted RDD Sampling

J. Michael Brick, David Judkins, Jill Montaqula, and David Morganstein[1]

Random digit dial (RDD) telephone surveys are cost-effective means of sampling and interviewing households in the U.S. In many RDD telephone surveys, telephone numbers are sampled with equal probability using a list-assisted procedure. We consider an alternative two-phase approach to sampling telephone numbers. A first phase sample is selected and auxiliary data are used to stratify telephone numbers by whether they are listed or whether an address can be associated with the telephone number. The telephone numbers in the second phase strata are sampled at rates that balance data collection costs and the variances of the estimates. Our findings suggest that stratification by mailable status has greater benefits than stratification by listed status in RDD surveys. Methods for estimation and variance estimation for the two-phase design are also considered.

Key words: Optimal allocation; cost models; telephone surveys.

1. Introduction

Telephone sampling has been used for conducting household surveys in the U.S. for several decades. This mode of interviewing is generally more cost-effective and less intrusive than interviewing by personal visit, especially for surveys with relatively short questionnaires. Differences between telephone and nontelephone households are important because of the potential for bias. As a result, telephone coverage has been examined many times since telephone samples became popular in the U.S. Thornberry and Massey (1988) present historical data on telephone coverage from 1963 to 1986 and Anderson et al. (1998) and Ford (1998) update these studies. These studies show the biases from excluding nontelephone households are small, except for subgroups with very low coverage such as low-income persons.

Sampling frames for telephone surveys in the U.S. comprise registers of area code-prefix combinations (the first six digits of the telephone number) that have been designated for the assignment of residential telephone numbers. However, due to the prevalence of business numbers and unassigned numbers in these area code-prefix combinations, about 75 percent of sampled telephone numbers selected completely at random (Casady and Lepkowski 1999) would be expected to be nonresidential.

Various telephone-sampling designs have been developed over the years in efforts to

[1] J. Michael Brick is Senior Statistician and Vice President, Westat, 1650 Research Blvd., Rockville, MD 20850, U.S.A. Email: MikeBrick@Westat.com. David Judkins is Senior Statistician, Westat. Email: DavidJudkins@Westat.com). Jill Montaquila is Senior Statistician, Westat. Email: JillMontaquila@Westat.com. David Morganstein is Vice President and Director, Westat. Email: DavidMorganstein@Westat.com.

reduce the amount of dialing to unproductive (business and nonworking) telephone numbers. Casady and Lepkowski (1999) review several of these approaches. The list-assisted method is now generally used, supplanting the Mitofsky-Waksberg method (Waksberg 1978), which was widely used in the 1980's for random digit dialing (RDD) surveys.

The list-assisted method introduced by Casady and Lepkowski (1993) is a single-stage, unclustered method that produces a self-weighting sample that covers most telephone numbers. In a list-assisted design, a simple random sample or systematic sample of telephone numbers is selected from all telephone numbers in 100-banks (the 100 consecutive telephone numbers with the same first eight digits) in which there is at least one residential telephone number listed in the White Pages directory. The telephone numbers in the sampling frame include both listed and unlisted numbers and both residential and nonresidential numbers. Telephone numbers in 100-banks with no listed telephone numbers are not sampled because Brick et al. (1995) show it has a very low yield and the bias associated with the truncated frame is generally negligible. They estimated that only 1.4 percent of the telephone numbers in these 100-banks were residential and this accounts for less than four percent of all households. Variants of this method of sampling that require more than one listed telephone number in the 100-bank to include the bank in the sampling frame are not discussed here.

Two procedures are available from vendors of telephone numbers that reduce the number of calls to ineligible telephone numbers in household surveys (nonworking and business numbers) even further. First, matches to Yellow and White Pages directories are used to identify business numbers. If a telephone number is listed in a Yellow Pages directory but not in a White Pages directory, it is classified as a business number. If the number is listed in the White Pages directory, it is classified as a listed telephone number. Second, tritone tests are used to identify nonworking numbers. If a tritone sound is detected (typically followed by a recording indicating that the number is out of service), the number is classified as a nonworking number. Since the tritone method is not completely reliable, some survey organizations repeat the tritone test, and the number is classified as nonworking only if a tritone is detected on each of the tests. See Casady and Lepkowski (1999) for more details on these methods.

In our presentation we classify the telephone interviews as either screening or extended interviews. The screening interviews provide data to determine if the telephone number is residential, if there are any persons in the household who are eligible for the survey, and any other data needed to sample eligible persons within the household. In the extended interview, persons sampled in the screening interview are asked to respond to the substantive content of the survey. Note that both screening and extended interviews may be conducted at the same time or they may require more than one call attempt. Many RDD surveys only conduct interviews in a subset of households (e.g., households with children) and this has implications for the sampling strategies that are discussed below.

2. Two-Phase Sampling

In many RDD telephone surveys in the U.S., telephone numbers are sampled with equal probability. An alternative approach is to select banks of 100 numbers with probabilities proportional to a measure of size such as the estimated number of listed telephone numbers, households, or individuals in the bank and take a fixed sample size within the selected banks. The weights to account for the unequal sampling rates. This method is attractive because it increases the percentage of telephone numbers that are residential as compared to an equal probability sample. Casady and Lepkowski (1993) found this approach has about the same efficiency as equal probability sampling if both cost and variance are considered. This finding is consistent with the theoretical discussion by Waksberg (1978), who showed that these probability proportional to size designs are generally not efficient under the conditions that exist for list-assisted samples. Thus, these designs are not more efficient than the equal probability list-assisted method.

Another approach is to stratify telephone numbers by auxiliary data available on the telephone frame of all numbers and independently sample at possibly different rates in the strata. This option is not feasible because the frames do not have data that would support this sampling approach currently. The commercial firms that have the sampling frames maintain lists of 100-banks (there were about 2.5 million such banks in 2000) rather than of all telephone numbers in the 100-banks. As a result, a simple stratified sampling approach of telephone numbers rather than of telephone numbers in 100-banks is not possible.

We propose two-phase stratified sampling as a method of dealing with this problem. Judkins (1996) considered a similar approach. A simple random sample of all telephone numbers in the frame (100-banks with at least one listed number) is selected in the first phase. The data on the listed status of the telephone numbers selected in the first phase is obtained by matching of the sample of telephone numbers to White Pages directories as described above. The first phase sample of telephone numbers is stratified by listed status (or some other variable obtained in the first phase) and a differential subsample of the numbers is selected by simple random sampling within these strata. The second phase subsample is then used as the RDD telephone sample for interviewing.

The two-phase approach has some operational and statistical appeal. Listed telephone numbers are more likely to be residential so the cost of finding a residence is lower in the stratum of listed numbers. In addition, households with listed telephone numbers are more likely to cooperate with most surveys. The level of cooperation can be further enhanced if an address can be obtained for the telephone number and used to send an advance letter to inform the household of the legitimacy of the survey or to provide respondents with an incentive. Singer et al. (2000) review the effects on response rates from using these types of procedures.

In a survey Westat conducted in 1999, about 55 percent of all the cooperating households had listed numbers even though only 33 percent of all telephone numbers are listed (Nolin et al. 2000). These findings are consistent with another 1999 study, where the estimated residency rate was 80 percent for listed numbers and only 39 percent for unlisted numbers (Judkins et al. 2001). In the latter study, higher screener response rates were obtained for households with listed numbers than for those with unlisted numbers (80% versus 71%).

Although stratifying at the second phase by listed status has some benefits as shown later, our research led us to consider the alternative of stratifying at the second phase by whether a mailing address can be associated with the telephone number. Several companies provide services of this type in which a telephone number is matched to commercially available files of addresses. We refer to a telephone number that can be identified in this fashion as having a ‘‘mailable’’ address. Because companies use multiple data sources including telephone listings to obtain mailable addresses, the percentage of telephone numbers with mailable addresses may exceed the percentage of listed numbers. In a survey Westat conducted during 2000, about half of all telephone numbers had mailable addresses but this half accounted for 85 percent of all the cooperating households. In this survey, the percentage of mailable telephone numbers that were residential was 68 percent while the percentage residential for those without mailable addresses was only 15 percent.

In the next section, we present a cost model for RDD surveys and use this cost model to examine the effectiveness of the two-phase sampling approach.

3. Cost Models and Optimal Allocation

To assess the usefulness of two-phase sampling for this application, we need a reasonable model of the costs in RDD samples. In some circumstances, sampling listed or mailable telephone numbers at a higher rate might be a cost-effective way of reducing the variance of the estimates, but using highly differential rates might be counter-productive. After presenting a cost model, we apply the theory of optimal allocation (Neyman 1938) to determine the appropriate subsampling rates.

In the two-phase sampling approach for RDD surveys, the second phase subsampling rate tends to be large and is fixed *a priori*. In these circumstances, the variance of the estimated total from the two-phase design is approximately equal to the variance from a simple stratified sample design. Since this approximation simplifies the presentation of the allocation problem, it is discussed briefly.

Särndal, Swensson, and Wretman (1992, Remark 9.4.3. on p. 354) derive an expression for the variance of the expansion estimator for two-phase sampling that can be simplified when the first phase finite population correction is ignored, the population variance is constant, and the subsampling rate is fixed. Under these conditions, the variance of the expansion estimator is

$$V(\hat{t}_{\pi *}) = \frac{N^2}{n_a} S_y^2 \left(1 + \sum_{h=1}^{H} W_h \left(\frac{1}{v_h} - 1\right)\right)$$

where n_a is the first phase sample size, v_h is the fixed second phase subsampling rate in stratum h, and W_h is the proportion of the population in stratum h. Designate the first stratum ($h = 1$) as the listed/mailable stratum and the second stratum the other stratum. Since $v_1 = 1$ when all the listed/mailable numbers are subsampled, the variance simplifies further to

$$V(\hat{t}_{\pi *}) = \frac{N^2}{n_a} S_y^2 \left(W_1 + \frac{W_2}{v_2}\right)$$

This can easily be shown to be equal to the variance of the estimated total from a stratified sample if the finite population corrections are ignored, the stratum-level population variances are constant, and the sample sizes for the strata are set equal to their expected values ($n_h \equiv n_a W_h v_h$). Because of this relationship, we use standard results for stratified sampling in the discussion of the cost model and optimal allocation.

The cost model should account for the major sampling and data collection activities in RDD surveys. As mentioned in the introduction, the current data collection procedures for list-assisted RDD samples include:

- drawing a sample of telephone numbers;
- matching the sampled numbers to the White and Yellow Pages directories (when residential listed numbers can be identified first);
- conducting tritone tests to identify nonworking numbers;
- screening the remaining telephone numbers for households with eligible household members and sampling these individuals for extended interviews; and
- conducting extended interviews with sampled household members.

Typical features of RDD surveys are relatively low initial cooperation rates (often less than 60%) when the household is first contacted and low rates of contacting persons per attempt due to persons not answering the telephone. For those surveys in which high response rates are very important, significant efforts are invested in dialing telephone numbers multiple times and in converting reluctant households. Because these efforts increase data collection costs and these costs vary substantially depending on the nature of the efforts, we include a separate parameter in our cost model to account for this activity.

Our model of the total expected cost, C, of an RDD survey that is stratified by listed or mailable telephone number status ($h = 1$ is the listed or mailable stratum and $h = 2$ is the other stratum) and has a fixed number of completed extended interviews, n, is:

$$C - c_0 = \sum_h n_h\{(t_h - 1)r_h c_{dh} + r_h c_s + c_e\} \tag{1}$$

where

- n_h is the number of completed extended interviews in stratum h, with $n = n_1 + n_2$;
- t_h is the average number of telephone numbers dialed to obtain one completed household screener interview;
- r_h is the average number of residences with completed screeners required to obtain one completed extended interview in stratum h;
- c_0 is the fixed cost of conducting the survey (planning, cost modeling, sample design and selection, training, data processing, etc.);
- c_{dh} is the average cost of data collection for telephone numbers in stratum h that do not result in screening interviews (see decomposition below), where the denominator is the count of sampled telephone numbers that do not complete the screening interview;
- c_s is the average cost of one completed residential screening interview, where the denominator is the count of completed screening interviews; and
- c_e is the average cost of one extended interview.

Some additional explanation of the parameters in the model may be useful. The cost term, c_{dh}, can be represented as a combination of two costs,

$$c_{dh} = \alpha_h c_n + (1 - \alpha_h)c_u \tag{2}$$

where

- c_n is the average cost of eliminating one nonresidential telephone number, including all nonresidential numbers, even those removed by the tritone and Yellow Pages matching;
- c_u is the average cost of unproductive attempts to obtain a completed screener interview with a nonresponding residential telephone number or a telephone number that cannot be classified as residential or nonresidential; and
- α_h is the proportion of unproductive numbers that are nonresidential in stratum h (see below for more details).

The cost of eliminating a nonresidential telephone number, c_n, includes the match to the Yellow Pages and tritone purging and the cost of interviewers dialing into business and nonworking telephone numbers that are not identified in the tritone purge efforts. The cost is averaged over all nonresidential telephone numbers. The average cost of unproductive attempts to obtain a completed screener, c_u, includes unsuccessful attempts to convince households to respond to the screening interview and unsuccessful attempts to determine residential status (mainly repeated ring/no answer and answering machine results). Here the cost is averaged over all nonresponding residential numbers. Clearly, the data collection protocol could have a significant effect on these costs.

The part of the fixed cost that is directly a function of the sampling method is assumed to be minor in the development that follows. For organizations that regularly conduct RDD surveys, there is little additional cost associated with applying any of these methods of sampling once the procedure is developed. This may not be the case for organizations that rarely do this work.

Two other parameters of the cost model are the multipliers that account for completion rates. The first is the average number of telephone numbers needed to get one household that completes a screener in stratum h, t_h. We can write this number as $t_h = (\omega_h \gamma_h)^{-1}$, where ω_h is the residency rate in stratum h and γ_h is the screener response rate in stratum h. As noted earlier, these rates differ substantially by stratum. Using this formulation, the parameters in Equation (2) can be written as $\alpha_h = (1 - \omega_h)/(1 - \omega_h \gamma_h)$.

The second parameter is the average number of households with completed screeners required to get one completed extended interview in Stratum h, r_h. We can write this number as $r_h = (\beta_h \xi_h)^{-1}$, where β_h is the proportion of households with eligible members and ξ_h is the extended interview response rate. For example, a survey may only be interested in sampling veterans of the U.S. military and β_h would be the estimated proportion of households in stratum h that have veterans. In this article we assume that in each household with an eligible member, one eligible person is sampled. While this is not always the situation because more than one member may be sampled or some households with eligible members may subsampled, this assumption simplifies the discussion greatly.

Applying standard sampling theory for optimal allocation (e.g., Cochran 1977) when the stratum variances are equal, the optimal ratio of the sampling fraction in the listed/mailable stratum to the other stratum for the number of completed extended interviews is

$$\lambda = \left[\frac{(t_2 - 1) r_2 c_{d2} + r_2 c_s + c_e}{(t_1 - 1) r_1 c_{d1} + r_1 c_s + c_e} \right]^{1/2} \tag{3}$$

The expression is a familiar result of optimal allocation; however, it is in terms of the number of completed extended interviews not the number of sampled telephone numbers. However, the ratio of the sampling rates in the listed to the unlisted stratum for the sampling of telephone numbers is also λ. These statements also apply to stratification by mailable status.

The reduction in the variance of the estimates using the optimal allocation given by (3) relative to the variance obtained under an equal cost design with allocation proportional to the number of telephone numbers in the stratum[2] can be evaluated under this model. The total cost in the design with sampling proportional to the number of telephone numbers is

$$C^* = n_2^{pr}\left\{\left(\frac{M_1 t_2 r_2}{M_2 t_1 r_1}\right)[(t_1 - 1)r_1 c_{d1} + r_1 c_s + c_e] + [(t_2 - 1)r_2 c_{d2} + r_2 c_s + c_e]\right\} \tag{4}$$

where the superscript ‘‘*pr*’’ indicates allocation proportional to the number of telephone numbers and M_h is the number of telephone numbers in stratum h in the population. Using the fixed cost, C^*, the optimal allocation in the other stratum is

$$n_2^{opt} = C^*\left\{\left(\frac{\lambda M_1 t_2 r_2}{M_2 t_1 r_1}\right)[(t_1 - 1)r_1 c_{d1} + r_1 c_s + c_e] + [(t_2 - 1)r_2 c_{d2} + r_2 c_s + c_e]\right\}^{-1} \tag{5}$$

Using Expressions (4) and (5), the ratio of the variance under optimal allocation to the variance under proportional allocation can be expressed in terms either of the number of telephone numbers in stratum (M_h) or of the number of households with eligible persons in stratum (N_h) as

$$\phi = \frac{n_2^{pr}}{n_2^{opt}} = \frac{\dfrac{M_1}{\lambda t_1 r_1} + \dfrac{M_2}{t_2 r_2}}{\dfrac{M_1}{t_1 r_1} + \dfrac{M_2}{t_2 r_2}} = \frac{(N_1 + \lambda N_2)^2}{(N_1 + N_2)(N_1 + \lambda^2 N_2)} \tag{6}$$

The specification of the parameters of the cost model is the remaining, and difficult, task.

The cost model given above is not unique or best, and different cost models could result in very different allocations and reductions in variance. For example, it may be reasonable to have c_s and c_e in the model be stratum specific as opposed to constant across strata. Furthermore, different organizations may conceptualize the costs differently because they use different procedures or for various other reasons.

4. Applications

The application of the optimal allocation theory depends critically on the estimation of the parameters of the cost model. The development of reasonable cost parameters is always a difficult task and this is true even when some features of the problem are fixed, such as only dealing with surveys done in a centralized, computer-assisted telephone research

[2] This design is the basis for comparison because it is the average design that results when stratification is not used.

facility. It is difficult to provide general guidance on the effectiveness of the design since the cost model has so many parameters that can vary from survey to survey.

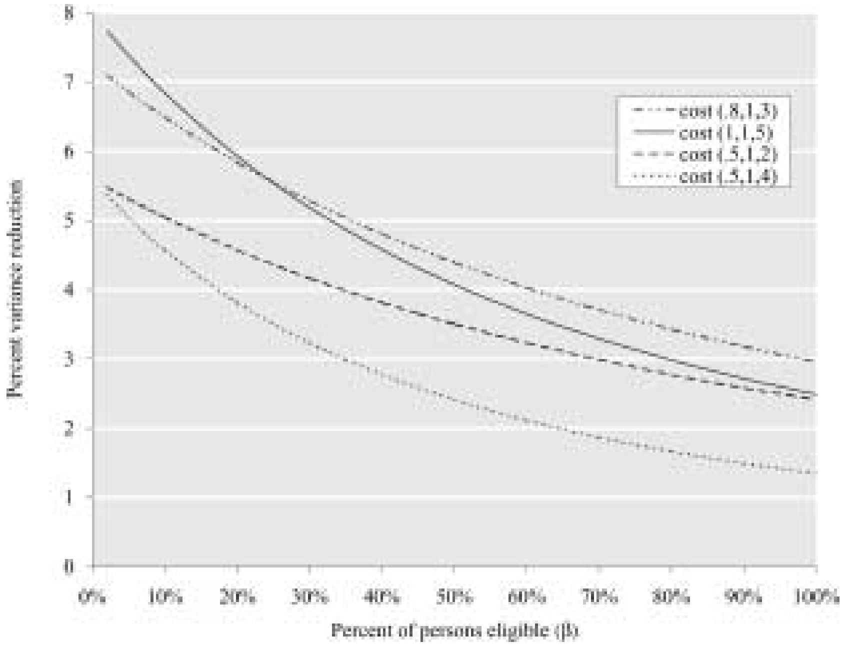
We begin by focusing on the variable costs in Equation (1) and use experiences from our centralized CATI facilities to model the costs. In surveys with high response rates, the relationship $c_{d1} \approx 2c_{d2}$ has been found to be typical and is roughly true for using either listed or mailable status to define the strata. This rule of thumb is based on applying empirical findings that show $\alpha_1 \approx 0.55$, $\alpha_2 \approx 0.85$ and $c_u/c_n \approx 8$ to the cost model given in Equation (2). The large cost ratio is due to the inclusion of the Yellow Pages and tritone purges in the denominator of c_n.

We use the relationship $c_{d1} \approx 2c_{d2}$ to reduce the dimensionality of the problem and express the variable costs as ratios to the cost of screening in a triplet defined as $(c_{d2}/c_s, c_s/c_s, c_e/c_s)$. For example, the ratio (.5,1,3) indicates the average cost per unproductive number in the unlisted stratum is half the cost of completing the screening interview, and the average cost of an extended interview is three times the cost of the screener. We examine various combinations of cost ratios below.

We begin with an examination using the listed status as the second phase stratification variable. From our experiences we estimate $t_1 = 1.6$ and $t_2 = 3.6$, where the factors include both the residency and response rates in the listed and unlisted strata. For national surveys we have found that $M_1 = 0.5M_2$ is a good approximation. The remaining variable to complete the specification is r_h, which is the inverse of the product of the extended interview response rate (ξ_h) and the households having individuals in the subgroup that are eligible for the survey (β_h). To simplify the presentation, we fix $\xi_1 = .80$ and $\xi_2 = .75$ and only allow the percentage of households having individuals in the subgroup to vary. In addition, we assume $\beta = \beta_1 = \beta_2$, which appears to be an entirely reasonable assumption in the absence of other information. For example, in surveys of parents of young children less than ten percent of all households have individuals eligible for the survey, while in surveys of all adults the eligibility rates are nearly 100 percent. We assume these rates are constant across strata once we have completed a screening interview.

With these specifications, the effectiveness of optimal allocation using listed status can be evaluated using Equation (6). The percentage reduction in variance, $100(1 - \phi)$, was evaluated for cost models with c_{d2}/c_s varying from 0.5 to 1.2, β varying from 0.02 to 1.00, and c_e/c_s going from 2.0 to 5.0. For all the cost models the reductions in variance are small, ranging from slightly above a two percent reduction (when c_{d2} is large, β is small and c_e is small) to 0.1 percent (when c_{d2} is small, β is large, and c_e is large). The optimal value of $\lambda \approx 1.2$ was obtained for most of the cost models. When achieving a high response is important, screening is used to identify a rare domain, and the extended interview is not very long, then listed stratification may be attractive statistically.

The main reason that stratification by listed status produces such small gains in efficiency is that only about 55 percent of all individuals are in households with listed telephone numbers. For oversampling to be effective, a high percentage of all residential numbers needs to be in the oversampled stratum. Recently, the percentage of telephone numbers with mailable addresses has increased substantially and is now larger than the percentage of telephone numbers that are listed. As we noted earlier, nearly 85 percent of all households that completed a recent survey had a mailable address. We speculate about the reasons for this in the concluding section.



Note: Cost triplets given in the legend are (c_{d2}/c_s, 1, c_e/c_s).

Fig. 1. Percent variance reduction with mailable stratification

Only a few changes in parameters of the cost-variance model are needed for evaluating the efficiency of using mailable status as the stratification variable. Based on a national survey conducted in 2000, we estimated $t_1 = 2.0$ and $t_2 = 10.6$ when the strata are mailable and not mailable telephone numbers, respectively. From the same survey, we found that $M_1 \approx M_2$. Otherwise, the parameters in the cost model were allowed to vary over the same ranges as used above.

Figure 1 shows the variance reduction for four cost models that may be appropriate for different types of RDD surveys. Larger reductions in variance are possible using the mailable status. The reductions are largest for rare populations, e.g., when only a small percentage of households contain individuals eligible for the extended interview. When over half the households contain eligible individuals, the variance reductions are relatively small and range from 2.0 percent to 4.5 percent. The variance reduction is largest when high response rates are important so that the percentage of all resources spent on unproductive telephone numbers (the relative value of c_d) is high. For example, the variance reduction for the $(.8, 1, 4)$ cost model is significantly larger than that for the $(.5, 1, 2)$ and $(.5, 1, 4)$ cost models for every value of β in the graph. While a large percent of all resources are expended on unproductive telephone numbers in U.S. federal government surveys in efforts to achieve high response rates, lower values such as $c_{d2} = 0.5$ are more common in public opinion surveys. The relative cost of the extended interview is also important, especially when most individuals are eligible. For example, with $\beta = 100\%$ the variance reduction when $c_e = 2$ is typically about two percent more than when $c_e = 5$.

In addition to the reduction in variance, it is interesting to examine the optimal oversampling rates for the mailable stratum from Equation (3) for these cost ratios. The optimal value of λ ranges from 1.3 to 2.0 for most of the cost models considered. For many surveys we conduct, the optimal value is between 1.5 and 1.7. This is much higher than the value when listed status was the stratification variable. The optimal choice of λ is fairly robust and moderate deviations from the optimal value do not cause large reductions in efficiency. For example, the reduction in variance using the optimal $\lambda = 1.97$ for the $(1, 1, 5)$ cost model is 7.7 percent, but using $\lambda = 1.6$ instead results in a variance reduction of 7.2 percent. Despite this robustness, losses in efficiency are very possible if the optimal allocation is applied and the cost model is misspecified. For example, oversampling the mailable stratum with $\lambda = 2.0$ when the cost model is inappropriate could be worse than using a standard (single-phase) equal probability sample design.

5. Estimation Issues

The procedures for developing the weights to produce estimates from the survey and for estimating the sampling errors of the estimates are discussed in this section. In most circumstances, the procedures are the same as in the standard one-phase design and do not require much additional discussion.

The base weights are the inverse of the first-phase selection probabilities. In most national RDD surveys the base weight is constant for all the telephone numbers. If a stratified sample is selected, then the base weights are constant within a stratum. For simplicity we assume a constant overall base weight, say $w_i = Nn^{-1}$, where N is the total number of telephone numbers in the sampling frame, n is the number sampled, and i refers to a specific sampled telephone number.

The base weight is then adjusted for the second phase of sampling. The adjusted weight for the telephone numbers that are retained in the second-phase sample is

$$w'_{hi} = \frac{\sum_{i \in h} w_{hi}}{\sum_{i \in h} \delta_{hi} w_{hi}} \tag{7}$$

where $\delta_{hi} = 1$ if telephone number i is selected in the second-phase sample in stratum h (where h is defined as above by mailable or listed status). This is a form of the reweighted expansion estimator (REE) described by Kott and Stukel (1997). This estimator has several advantages over alternative two-phase estimators, including stability when the sample size is large in each stratum and simplicity for variance estimation with replication. Notice that the REE weights for the second-phase sample sum to estimated stratum totals from the first-phase sample.

Other than the simple adjustment given by (7), the standard weighting procedures are followed for both the one-phase and two-phase samples. These procedures typically include nonresponse adjustments, adjustments for multiple telephones in the household, and calibration weighting such as poststratification or raking. Calibration estimators introduced by Deville and Särndal (1992) often have important effects in RDD surveys, most importantly adjusting for undercoverage.

The method of adjusting for multiple telephone numbers in a household described by

Massey and Botman (1988, pp. 145–146) has been used in most applications, but there is no theoretical justification for it provided there. However, Kalton and Brick (1995) show that dividing the weight by the number of eligible telephone numbers in the household (where eligible means telephone numbers that could have been used to sample the household) gives an unbiased estimator. Massey and Botman (1988) truncated the adjustment, dividing the weight for all households with more than one eligible telephone number by a factor of two. They argued that the truncated weight would reduce the variability of the weights and offset any bias thus introduced. However, since only about seven percent of households had more than one telephone number in the U.S. in 1999 (Brock-Roth, Montaquila, and Brick 2001), the increase in variance due to the differential weights is very small. Massey and Botman also justify truncation because reports of a large number of reported telephone lines in a household may be due to reporting error. Our experience is consistent with this and as a result we truncate the adjustment at a factor of three rather than two.

To compute sampling errors for RDD surveys we generally prefer replication methods because they provide a standard method for including all the adjustments for nonresponse and calibration. Rust and Rao (1996) discuss creating replicate weights that accomplish this goal. Once the replicate weights have been created, replication software such as WesVar (Westat 2000) can be used to produce estimates of sampling errors that include the adjustments. This facility is particularly important in RDD surveys because the calibration weighting adjustments may have important effects on the variances of the estimates. Binder (1996) presents procedures for estimating variances of estimates from two-phase samples using the linearization method. A drawback of the linearization method is that currently no software supports this method for two-phase samples. In addition, modifications in the procedures would be needed to deal with calibration and other weight adjustment methods.

Replication for the two-phase approach and the REE estimator only requires one additional step corresponding to the full sample adjustment given by (7). Replicate base weights are formed in the usual fashion for balanced repeated replication or jackknife replication. Each of the replicate base weights is then adjusted using (7) but substituting the replicate base weights in that equation for the full sample base weights. Kim (2001) provides more details on replicate variance estimation in two-phase samples. The adjustment for two-phase sampling has exactly the same form as a standard nonresponse adjustment and WesVar can be used to implement this. The other steps in forming the replicate weights are completely identical to those for standard one-phase samples.

6. Discussion

The two-phase sampling method with stratification by mailable status has greater benefits than stratification by listed status, based on the cost model and the current state of the data available for stratification. Neither method produces large reductions in the precision of the estimates for a fixed cost. One of the reasons for this is that technology has reduced the cost of eliminating nonworking and nonresidential numbers. On the other hand, it is becoming more difficult to achieve high response rates in RDD surveys and techniques such as the use of monetary incentives to encourage high response rates may drive up the costs of the screening. Increased screening costs increase the effectiveness of oversampling using these data.

As mentioned earlier, the main reason that stratification by mailable status is more effective than stratification by listed status is that a higher percentage of residential households are identified by mailable status. We suspect this may be due to the procedures used by providers of the data and is subject to change. For example, consider a household with two telephone numbers. Typically only one of the two numbers is listed. However, providers of addresses for telephone numbers may be able to link both telephone numbers to the household. This ability may partially account for the higher percentage of residential telephone numbers identified by mailable status. Clearly, providers of both of these services may use other procedures. In addition, other factors such as new legislation may also affect the availability of the data in the future.

An important aspect of conducting RDD surveys not factored into the cost models is the effect of the pool of telephone numbers on the interviewers. Interviewers have higher job satisfaction and performance when a large proportion of the numbers they dial are residential and when the individuals answering the telephone are more cooperative. Both of these beneficial effects are attained when mailable or listed telephone numbers are oversampled. Thus, even though we may not obtain large gains in precision by oversampling, there are some gains and little reason to avoid the procedure for organizations that regularly conduct RDD surveys.

Before oversampling telephone numbers is instituted in a particular survey, it is important to determine whether the cost model given in this article is appropriate for the organization and the survey in particular. As noted earlier, application of the optimal oversampling rates could lower the efficiency of the sample if the cost model is incorrectly specified. One example of a study-specific parameter that could affect the oversampling rate is the proportion of telephone numbers that are mailable. This factor was estimated from a national survey, but in state or local surveys the percentage might be substantially different and it could alter the optimal oversampling rate. Despite these caveats, we believe that two-phase sample design is a valuable method of RDD sampling.

7. References

Anderson, J.E., Nelson, D.E., and Wilson, R.W. (1998). Telephone Coverage and Measurement of Health Risk Indicators: Data from the National Health Interview Survey. American Journal of Public Health, 88, 1392–1395.

Binder, D.A. (1996). Linearization Methods for Single Phase and Two-Phase Samples: A Cookbook Approach. Survey Methodology, 22, 17–22.

Brick, J.M., Waksberg, J., Kulp, D., and Starer, A. (1995). Bias in List-Assisted Telephone Surveys. Public Opinion Quarterly, 59, 218–235.

Brock-Roth, S., Montaquila, J., and Brick, J.M. (2001). Effects of Telephone Technologies and Call Screening Devices on Sampling, Weighting and Cooperation in a Random Digit Dialing Survey. Proceedings of the American Statistical Association, Section on Survey Research Methods [CD-ROM], Alexandria, VA.

Casady, R.J. and Lepkowski, J.M. (1993). Stratified Telephone Survey Designs. Survey Methodology, 19, 103–113.

Casady, R.J. and Lepkowski, J.M. (1999). Telephone Sampling. In Sampling of Populations: Methods and Applications, P.S. Levy and S. Lemeshow (eds.). New York: Wiley.
Cochran, W.G. (1977). Sampling Techniques. New York: Wiley.
Deville, J.C. and Särndal, C.E. (1992). Calibration Estimators in Survey Sampling. Journal of the American Statistical Association, 87, 376–382.
Ford, E.S. (1998). Characteristics of Survey Participants With and Without a Telephone: Findings from the Third National Health and Nutrition Examination Survey. Journal of Clinical Epidemiology, 51, 55–60.
Judkins, D. (1996). Directions for Research on the National Immunization Survey. Final Report, Research Triangle Institute, Research Triangle Park, N.C.
Judkins, D., Brick, J.M., Broene, P., Ferraro, D., and Strickler, T. (2001). 1999 NSAF Sample Design. Report No. 2. Washington, D.C.: Urban Institute.
Kalton, G. and Brick, J.M. (1995). Weighting Schemes for Household Surveys. Survey Methodology, 21, 33–44.
Kim, J.K. (2001). Variance Estimation after Imputation. Survey Methodology, 27, 75–84.
Kott, P.S. and Stukel, D.M. (1997). Can the Jackknife Be Used with a Two-Phase Sample? Survey Methodology, 23, 81–89.
Massey, J.T. and Botman, S.L. (1988). Weighting Adjustments for Random Digit Dialed Surveys. In Telephone Survey Methodology, R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, and J. Waksberg (eds.). New York: Wiley.
Neyman, J. (1938). Contribution to the Theory of Sampling Human Populations. Journal of the American Statistical Association, 33, 101–116.
Nolin, M.J., Montaquila, J., Nicchitta, P., Kim, K., Kleiner, B., Lennon, J., Chapman, C., Creighton, S., and Bielick, S. (2000). National Household Education Survey of 1999: Methodology Report, NCES Technical Report 2000–078, Washington, D.C.: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.
Rust, K. and Rao, J.N.K. (1996). Variance Estimation for Complex Surveys Using Replication Techniques. Statistical Methods in Medical Research, 5, 283–310.
Singer, E., Van Hoewyk, J., and Maher, M. (2000). Experiments with Incentives in Telephone Surveys. Public Opinion Quarterly, 64, 171–188.
Särndal, C.E., Swensson, B., and Wretman, J. (1992). Model Assisted Survey Sampling. New York: Springer-Verlag.
Thornberry, O.T. and Massey, J.T. (1988). Trends in United States Telephone Coverage Across Time and Subgroups. In Telephone Survey Methodology, R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, and J. Waksberg (eds.). New York: Wiley.
Waksberg, J. (1978). Sampling Methods for Random Digit Dialing. Journal of the American Statistical Association, 73, 40–46.
Westat (2000). WesVar 4.0 User's Guide. Rockville, MD: Westat.

Received June 2001
Revised March 2002