The recent article, “Comparison of Telephone Sampling and Area Sampling: Response Rates and Within-Household Coverage” (Am J Epidemiol 2001;153:1119–27), raised a number of issues related to two sampling methodologies that can be used for selecting population-based controls for case-control studies: random digit dialing (RDD) and area probability sampling. Some of these issues are discussed in this commentary in more detail to help in making sample design decisions, including the need to take the analysis plan into account when developing a sample design. Data from the paper are used to illustrate how the choice of sample design can affect analyses. Relative costs associated with the two methodologies as well as variance and bias concerns are also discussed in detail. Sample coverage issues, including those associated with list-assisted RDD, are considered, as are some advantages of the list-assisted approach. A discussion of the use of concurrent screening and sampling with an RDD approach as an alternative to periodically selecting fixed sample sizes is provided. 


bias; case-control studies; cluster sampling; data collection; intraclass correlation; random digit dialing; sample design

The article, “Comparison of Telephone Sampling and Area Sampling: Response Rates and Within-Household Coverage,” by Brogan et al. (1) provides a useful comparison of two potential methods of identifying and selecting population-based controls for case-control studies: random digit dialing and area probability sampling. As statisticians involved in sample design and data collection issues on the study discussed in the paper as well as other case-control studies, we thought it might be helpful to explore further some of the sample design and analytical issues raised. Since the random digit dialing (RDD) methodology generally used now is the list-assisted approach, not the one in use at the time of the Women’s Interview Study of Health (WISH), implications with respect to current applications are noted.

SAMPLE DESIGN FACTORS AND THE NEED TO MAKE TRADE-OFFS—AN ILLUSTRATION

When a sample design is being developed, trade-offs between various factors are typically made. Chief among these factors are survey costs, the variability of sample estimates, the potential for bias in sample estimates, and analytical objectives. Consider a trade-off discussed in the paper by Brogan et al.: weighing real-life costs against potential bias when comparing sample methodologies.

An in-person screening for an area probability sample (APS) survey will virtually always have a higher response rate than a corresponding RDD survey. The higher the response rate, the less potential for biased estimates. Thus, an APS approach generally has less potential for bias than does an RDD; however, an RDD survey is generally considerably cheaper to administer.

For the WISH study in Atlanta, Georgia, the authors report that the difference between the screening rates of 94.9 and 89.4 percent for the APS and RDD, respectively, is statistically significant. However, a 5.5 percent difference in response rates is generally considered well worth the use of RDD methodology to achieve substantial cost savings. Of course, the WISH study took place in the early 1990s. The difference between RDD and APS response rates is usually larger now, and there are other concerns identified by the authors that make RDD surveys less attractive than they once were. Thus, the fact that the WISH RDD screener response rates are high does not diminish the authors’ point that it is important to explore the possibility of improvements in current RDD methodology and that other methods of identifying controls merit consideration.

ANALYTICAL REQUIREMENTS TO CONSIDER IN DESIGNING A SAMPLE FOR CONTROLS

To design a sample appropriately, it is important to consider the types of analyses planned as well as sampling and
operational methods. Typical requirements for selecting a sample of population-based controls include the following:

1. **Match the case distribution.** Sample controls so their sample distribution matches the expected distribution of cases across a set of demographic variables in a particular area over a given time period;
2. **Minimize the overall error.** Use efficient procedures to provide close to the lowest total mean square error (i.e., the combination of sampling error and bias) possible for the available budget;
3. **Hit targeted sample sizes.** Obtain targeted sample sizes of controls almost exactly rather than in expectation;
4. **Permit analysis without special software.** Analyses for case-control studies are frequently done without using sample weights or “special” software (required for handling data from complex sample surveys) for estimating standard errors, establishing confidence intervals, or testing hypotheses. When neither weights nor special software will be used, sample selection should be done with equal probability within strata (demographic groups designated for matching purposes and used as independent variables in regressions). In addition, if sampled units are clustered, the effects on estimates of variability should be negligible. Keeping cluster sizes small helps reduce the effect of the intraclass correlation, a measure of the extent to which units from the same cluster are similar with respect to a data item, on confidence intervals.

**CONSIDERING COSTS AND DATA COLLECTION MODE IN LIGHT OF THE ANALYTICAL REQUIREMENTS**

We focus here on differences in effectiveness between the data collection modes. A convenient way of precisely hitting targeted sample sizes is to carry out screening interviews in several waves over the period of time in which cases are accrued. For an RDD control screener, a wave is a complete set of sampled telephone numbers selected to be fielded over a period of time, say 2–3 months. If assumptions about expected response rates or population distributions (the population distribution of cases, the general population, or both) are not fully met, adjustments in sample sizes or sampling rates can be made in future waves to help ensure that target sample sizes are met almost exactly.

Fielding an APS using multiple waves provides the same sort of flexibility as an RDD in terms of adjusting sampling rates or sizes to meet targets. Thus, waves were used for the APS discussed in the paper. However, it is substantially more expensive compared with an RDD approach, particularly when the interview is a brief screener, as is generally true for case-control studies. The time to implement a screener accounts for only a small proportion of the interviewing costs for an APS, with travel time responsible for most of the cost. Return trips for refusal conversion or when no one is at home can become a resource concern. To help mitigate such travel costs, one could limit the screening for an APS to exactly one wave. However, in doing so, there is no opportunity to make modifications in the event of departures from assumptions, so analytical objectives may be compromised.

The RDD design has both cost and analytical advantages over the APS design described in the paper by avoiding the need for special software. At the time the WISH RDD sample was selected, the cluster sample methodology known as Mitofsky-Waksberg (2) was generally used for RDD sample selection. Clusters of telephone numbers likely to include a relatively large number of residential telephones are selected, with each cluster consisting of 100 consecutive telephone numbers. Residential telephone numbers are sampled within clusters in such a way that each sampled residential number has the same overall probability of selection and the number of sample residential numbers in each cluster is almost identical. The number of clusters sampled for the WISH RDD was determined for achievement of an average of approximately two controls per sampled cluster. Thus, the contribution of the intraclass correlation coefficient to the variance of survey estimates was expected to be negligible. With women sampled to serve as controls with equal probability within strata (age categories), analyses could be undertaken without weights or special software for variance estimation purposes.

For the APS approach discussed by Brogan et al., 180 segments (one or more blocks) were selected in the Atlanta area, with 640 controls ultimately participating, an average of 3.56 controls per segment (cluster). The intraclass correlation of persons within such segments for a given data item can generally be expected to be much higher than the intraclass correlation of persons found in the same group of 100 consecutive telephone numbers, since the telephone numbers within an exchange are spaced over a much broader geographic area. People living in the same area tend to be similar in terms of income, race/ethnicity, educational attainment, and many other factors. Data from table 3 of the paper can be used to illustrate this.

Design effects are standard measures for comparing estimated variances from equal-size samples of different designs, indicating the contribution of variable sample weights and the clustering of a sample to the resulting variance of an estimate. Brogan et al. evaluated the design effects for APS and RDD general population (weighted) estimates they developed and describe those associated with APS as high and those associated with RDD as moderate. The contribution of the sample weights to the variability of the sample estimates should be roughly the same for the APS and RDD estimates since the targeted sample sizes were about the same for each age category and the relative rate of finding women in each such category should be very close in the general household and telephone household populations. For RDD, the sample weights should account for virtually all of an estimate’s design effect. The major difference between the two approaches was in the degree of clustering and the corresponding intraclass correlation of persons within a cluster. A standard approach to assessing the magnitude of the difference in variability between two designs is to examine the ratio of the variance of an estimate, the square of the standard error, from one design to the corresponding variance for the other design.
Consider the ratio of APS estimated variances to RDD estimated variances for the income categories in table 3 of the paper by Brogan et al. for all races combined and for non-Blacks. For all races combined, the ratio of APS to RDD estimated variances for the five income categories ranged from 1.39 to 5.0, with three categories having ratios over 2. For non-Blacks, the corresponding ratios for the five income categories ranged from 0.9 to 7.4, with another category just under seven, an additional category over two, and only the 0.9 ratio under one. The sample sizes for Blacks are substantially lower than those for all races combined and for non-Blacks, and the corresponding estimates of variance are highly unstable and, thus, not reliable for assessment purposes. For the five categories of livebirth in table 3, the differences identified above are also in evidence but are not quite as large.

The impact of clustering is consistently higher for the APS compared with the RDD approach and is dramatically higher for a number of categories. The APS approach described by the authors for the WISH study clearly does not meet one of the important analytical requirements for the analysis of WISH data: that one should be able to ignore the effect of any clustering in the sample selection of controls when analyzing the collected data. It should be noted that it is not necessary to analyze data ignoring the effects of clustering. The authors cite an article by Graubard et al. (3), in which clustering is taken into account in the analysis of case-control data. (Such clustering should also be reflected in power calculations.) However, this is often not the approach planned. For the WISH study, an important requirement was to ensure a negligible effect of clustering on the variability of sampled persons within the sample strata (age cells). When one undertakes a case-control study, it is important to consider the analytical approaches to be used and to make sure that the sample design will facilitate the planned analyses.

Now consider the issue of cost in conjunction with this analytical requirement. In the article by Brogan et al., the authors state, “...our impression is that the cost of APS sampling would compare favorably to that of RDD sampling in our situation” (1, p. 1125) (although it was pointed out that precise cost comparisons between APS and RDD were not possible). As indicated in the paper, having the area sample limited to three counties, rather than a statewide or national study, does serve to limit the cost of the area sample. However, the concern expressed in the paper that screening out nontargeted counties would add substantially to RDD costs did not occur. Table 1 of the paper shows that only 408 of the 12,033 sampled telephone numbers resulted in the identification of a household’s being outside the targeted geographic area, which is only 3.4 percent of all the telephone numbers selected. Moreover, since the completion of the study described in the paper, databases have been developed that readily identify telephone exchanges that are predominantly outside a targeted local area of interest if the area is defined in terms of ZIP codes or counties. Such exchanges can be excluded from a sample frame. The loss in terms of households within targeted counties by excluding such exchanges can be kept small, reducing the potential for bias, while increasing the chances that a contacted household is in the area of interest.

In fact, concerns about such bias would be eliminated if cases were constrained to the same telephone exchanges as those used to sample controls. Consequently, screening out persons in nontargeted counties is generally not a major cost issue for current RDD control selection efforts.

The other factor identified in the paper as being associated with higher costs of the RDD sample is that the RDD addresses tended to be more geographically dispersed than did the APS. Although this did increase the costs of the RDD approach, the impact of sample clusters on variability was substantially reduced, avoiding the need for special analytical software, as required. If the APS described in the paper were designed to meet the same level of clustering, the APS costs would have been increased substantially.

To illustrate this last point, for the APS sample design to achieve the same level of precision as the RDD design, the APS should have been designed to obtain roughly two controls per cluster, the same as the RDD design. The number of controls per cluster for RDD was approximately 2.3. To achieve 2.3 controls per cluster with 640 controls ultimately participating (the APS number of controls), the number of segments used should have been roughly 278, an increase of more than 50 percent over the 180 segments actually used. In fact, intraclass correlations within a segment are generally higher than the intraclass correlations within a cluster of 100 consecutive telephone numbers, resulting in the need for even more clusters in the APS to achieve the same effect on variability as the RDD approach. APS would thus cost substantially more than RDD to achieve the same degree of precision as the RDD survey (e.g., confidence intervals with the same expected width), particularly since the list-assisted methodology now in use with RDD has no clustering.

The question then is, “When is the higher quality of an APS worth the additional cost?” An APS approach may be a favorable alternative when survey response or coverage is a critical issue. For example, there may be reasons to believe that nonrespondents are very different from respondents, and the reduction of potential bias is important. Studies involving sexually transmitted diseases have such concerns. Circumstances in which coverage may be an issue include those in which a disease or a health issue disproportionately affects poor persons to a high degree. A study of the effects of lead-based paint is an example. Even in such studies, a telephone survey will permit appropriate analyses to be undertaken if the comparisons are restricted to cases who live in households with telephones and have the disease or health problem. However, leaving out an important component of the population with the disease could provide an incomplete and perhaps distorted perception of the risk factors associated with the disease or health problem. Unless there is reason to believe that survey response and/or coverage poses such special concerns, a telephone survey will generally be the preferable option for carrying out the screening for the sample selection of controls.

THE ISSUE OF COVERAGE OF THE TARGETED POPULATION

Undercoverage of a target population can be a potential source of bias in survey estimates, an important concern of the authors. However, although the RDD screening for the
study resulted in a substantially lower proportion of households identified as having members of the target population, this undercoverage does not necessarily drive up the cost of RDD screening, as suggested in the paper. To understand this, it is important to understand the process of determining how many households to screen.

The number of households to be screened for a given survey to identify controls is determined by the cell (e.g., age or age–race category) that is “hardest to fill.” Filling the cell depends on two factors: the number of persons in households with a telephone who fall in the cell and the number of cases expected to be found among persons in that cell over a particular period of time (e.g., a year). Except for studies restricted to narrow age ranges, the cells “harder to fill” are generally those associated with older persons (since they have a higher incidence of disease and are relatively rare in the general population) who are a member of a minority race (and, thus, even rarer in the general population). The number of screeners to complete is based on the number needed to fill the hardest-to-fill cell. All other cells are subsampled.

Thus, unless there is poor coverage for a cell other than the one hardest-to-fill and the subsampling rate for that cell is high (a larger proportion of the cell is taken), undercoverage in one or many cells will not affect the RDD screening rate and, thus, will not affect survey costs. Undercoverage of the cell hardest to fill will, of course, result in additional screening. Consequently, it is necessary to assess which components of the population are likely to be undercovered and the sampling rates associated with those components before one can assess the effect of the undercoverage on survey costs.

There are also bias concerns related to undercoverage. In assessing such concerns, questions to consider include “Who is undercovered?” and “Are they a large part of the target population?” For case-control studies, the coverage of cells from which the largest number of controls is to be selected will have the most important impact on potential bias. As a result, undercoverage of young adults, who are frequently undercovered in surveys, would be relatively important if they happen to represent only a small percentage of cases. Conversely, undercoverage of older persons could be a great concern if they represent a relatively large proportion of cases, even though coverage of the general population might be high. Thus, for assessment of the impact of undercoverage on a study, an evaluation is needed to determine the specific groups likely to be undercovered, the extent of this undercoverage, and the relative importance of the undercovered groups in gaining an understanding of the nature of a target population. As noted by Brogan et al., the lack of coverage of persons without telephones in their households is routinely addressed for case-control studies when an RDD approach is used by eliminating from analysis cases who live in households without telephones.

A NOTE ON SCREENER RESPONSE RATES

Another suggestion made in the paper had to do with the sampling of women for participation in the study. Specifically, it was stated, “Probability sampling of women for interview at the time of RDD screening, rather than 1–3 months later as in our study, could reduce the need for iden-

tifying information such as name and/or address at RDD screening and increase the RDD screening response rate” (1, p. 1123). In the WISH RDD study, the screener was focused on obtaining information needed for sampling purposes, that is, basic demographic data. This is usually the case when sampling controls. Failure to obtain names and addresses represents nonresponse for a completed extended interview. This is largely a bookkeeping issue, as it has no effect on the overall response rate, but the suggested approach for possibly improving RDD screener response rates will not accomplish this objective. Nevertheless, there are other reasons for considering the approach of concurrent screening and sampling in RDD, which are discussed next.

CONCURRENT SCREENING AND SAMPLING OF CONTROLS WITH AN RDD APPROACH

Concurrently screening and sampling controls eliminates the need to collect names and addresses of nonsampled persons. Since sampling rates must be established in advance, sample sizes for a single wave of data collection may vary somewhat from targets, but sampling rates can be adjusted between waves to help achieve overall targets. Monitoring sample yields to adjust rates appropriately is somewhat labor intensive, adding to screener costs, but may be desirable since sampled controls could be made available for fielding purposes more readily, perhaps every month. Selecting fixed sample sizes for each wave, generally resulting in varying sampling rates as well, requires waiting until a wave is completed, roughly 3 months.

Note that the use of varying sampling rates over time tac-
itely assumes that there are no important differences between sampled controls (or cases) over the full period of data collection. In essence, this assumption permits us to average sampling rates within a single stratum over the full period of data collection, treating them as if they all were selected at the same rate. This is generally a reasonable assumption. If, for some reason, the nature of the underlying populations of cases and/or controls changes substantially over the data collection period, treating samples of either cases or controls as coming from a stable population would be questionable. The nature of the case population might change if a new, more sensitive test becomes available for detecting the presence of the disease in question or if a celebrity develops the disease and speaks out about the need for testing, resulting in substantially more persons being tested than before. The nature of the population from which the controls are sampled might change if, for example, a dramatic economic recession causes the relocation of a relatively large segment of one of the subpopulations of interest. Sample weights, commonly used in the analysis of survey data, could be developed and used in such circumstances to permit appropriate analyses. A text by Korn and Graubard (4) discusses the use of sample weights in the analysis of data for case-control studies.

SAMPLE FRAME FOR LIST-ASSISTED RDD

As mentioned by Brogan et al., the list-assisted methodology has replaced the Mitofsky-Waksberg approach for RDD
surveys. This methodology has the advantage of being less cumbersome (and thus, less costly) to implement and does not incorporate clustering into the sample design, eliminating the contribution of clustering to the variance of sample estimates. Its cost-effectiveness comes from eliminating from the sample frame groups of 100 consecutive telephone numbers that have no listed residential number. Roughly 2–3 percent of residential telephone numbers are thus lost. Brick et al. (5) have examined the effect of excluding such numbers, concluding that little is lost by doing so. Nevertheless, some case-control studies have required the inclusion of all numbers. This increases screening costs considerably. Survey resources are being expended on improving coverage slightly when there is little evidence of concern about bias. Perhaps such resources might be better spent on efforts to improve data quality by increasing response rates (e.g., through mailings, when possible, or by additional follow-up calls) where potential bias concerns are greater.

SUMMARY

Brogan et al. (1) raised many important issues regarding RDD versus APS sampling methods for control selection in population-based case-control studies. We have attempted to address these and related issues in the hope of expanding the dialogue on the assessment of alternative data collection methodologies, so important to epidemiology in general and to control selection in particular. We are unaware of any single procedure that is optimum for all case-control studies. In preparing for such a study, one should consider many issues. These include the extent to which factors influencing the disease will also affect willingness of sample persons to respond, the size of the geographic area and its effect on interview travel costs, the effect on costs or analyses of various administrative decisions (e.g., the number of call backs to be made to achieve high response rates or whether the interviews will be done in waves or as a single data collection effort), whether intraclass correlations within small geographic units such as city blocks are high for some measures of analytical interest, whether analyses are to be done for subclasses of the target population, what confidence levels to use, and the nature and form of both the screening and interviewing questionnaires.

The paper by Brogan et al. contains useful references for readers engaged in case-control studies. Most cited in this commentary are also found there.

REFERENCES