TELEPHONE SCREENING FOR RARE CHARACTERISTICS USING MULTIPLICITY COUNTING RULES

Jane W. Bergsten, Research Triangle Institute
Stephanie A. Pierson, Research Triangle Institute

Screening households for rare characteristics using field interviewers can be both costly and time consuming. With telephone coverage of the household population exceeding 90 percent nationwide, cheaper and faster means of screening are now available to the survey researcher. This paper will describe the procedures used and some of the problems encountered in a recent national random digit dialing telephone survey focusing on the use of sewage sludge on home vegetable gardens.

The problem we were charged with was to locate, by means of a random digit dialing survey, two samples of home vegetable gardeners. The first was to be a sample of gardeners who had treated the area where vegetables were growing with sewage sludge, and the second was to be a sample of vegetable gardeners who had not used sewage sludge on their garden. Subsamples of each of the two groups would then be selected for a personal visit by a field team of interviewers, who would conduct a personal interview with the gardener and take specimens of both soil and vegetable material. The specimens would later be analyzed for the presence of toxic substances.

Because only three or four percent of the households were expected to contain a sludge user home vegetable gardener, even telephone screening would require a massive effort. Some type of multiplicity techniques were called for so that we could concentrate our efforts where success was more likely. Any technique that would be used would have to permit the computation of sample inclusion probabilities. In addition, clustering of the sample into geographic areas would be necessary in order to contain costs during the personal interviewing and specimen taking phase of the project when field interviewers would be sent out from a central location to conduct the field work.

The Survey Design

The random digit dialing survey was conducted during the summer of 1981 in a sample of 100 primary sampling units (PSUs) which had been selected using information about the sewage treatment plants located in the PSUs. (Most sludge is obtained from sewage treatment plants.) Size of the effluent and whether or not the residue was air dried provided us with measure-of-size information related to the potential for sewage sludge distribution in the area. The sample of 100 PSUs was selected giving greater probability to PSUs having the potential to distribute larger amounts of sewage sludge.

Because telephone exchange characteristics do not dove-tail with political boundaries, special rules were developed for linkage. Using a computer tape which was obtained from the Long Lines Department of AT&T and which contained a listing of all working telephone exchanges in the nation together with the central office location, a county identifier was assigned to each exchange. The exchanges were then aggregated to the county level. A PSU was then defined as the aggregation of all 10,000 telephone numbers per exchange for all exchanges linked to the PSU, regardless of whether or not the dwelling unit for any particular telephone number was located inside or outside the politically defined geographical boundaries of the PSU. This definition meant that no geographic location screening was necessary during the telephone interview, and that county level data could nevertheless be used for stratification purposes at the PSU selection stage.

A modification of the Mitofsky-Waksberg cluster sampling approach was used to select clusters containing 8 residential telephone numbers in the 100 PSUs in the sample. Because of the low prevalence of sludge usage, it was decided that multiplicity counting rules would be used to increase the efficiency of the sample and reduce the total number of interviews needed.

Respondents in the sample households would be asked to help identify other persons having the characteristics of interest, namely, having a home vegetable garden and applying sewage sludge as a fertilizer. Several criteria were used in evaluating possible methods of linking other dwellings to the sample dwelling. First, the linkages between the primary respondents of the sample dwellings and the referrals had to be easily and clearly determinable and measurable so that selection probabilities could be calculated. In addition, in order to obtain reliable referral information the screening questions had to tap readily available information.

Three possible options for obtaining referrals were considered. The most obvious option and one which would potentially yield the greatest number of referrals was to ask if respondents knew anyone who had used sewage sludge on a home vegetable garden. This option was immediately discarded, for two reasons. First, it would be almost impossible to accurately determine the selection probabilities for this type of referral. Second, using sewage sludge is not a characteristic that is widely evident to others.

Another option considered was to ask respondents to identify relatives who lived in the area and who had used sewage sludge on their gardens. Although it would have been easier to determine the selection probabilities using this technique, it was again thought that respondents would probably not know whether their relatives had used sewage sludge, and may not even know whether or not relatives had vegetable gardens.

The method that was finally chosen was to ask respondents to identify their immediate neighbors who had vegetable gardens on the property. It was thought that linking to neighboring dwellings using predetermined linkage rules would allow for an easy and accurate determination of selection probabilities. It was also thought that respondents would likely know if their immediate neighbors had vegetable gardens, since most gardens would be visible to the

neighborhood. It was decided not to ask respondents if their neighbors applied sludge to their gardens, however, since that information would be less readily available.

In establishing the specific rules for linking with neighbors, the previously mentioned criteria were used, namely, (1) being able to clearly and easily determine the linkages and (2) tapping readily available and reliable information. "Neighbors", therefore, had to be close enough to have knowledge of gardens and also have clear and measurable linkages to the sample dwelling. The following counting rules were employed. Respondents living in single family dwellings were asked to identify vegetable gardens on the property next door to the right and to the left, but only for next door neighbors living in single family dwellings. Respondents living in structures containing 2 or 3 housing units were asked about the other dwelling(s) in the same structure. Respondents living in structures containing 4 or more units were not asked about their neighbors at all. It was thought that people living in large multifamily structures would be unlikely to have vegetable gardens, therefore there would have been little to gain from such querying. Even if this were not the case, the wide variety of housing configurations possible in large multifamily structures made defining a neighbor close enough to assure acquaintance very difficult.

Respondents were asked only if their neighbors had vegetable gardens on their home property; no inquiries were made about gardens maintained at other locations. Respondents reporting a neighbor's vegetable garden were asked for the name, telephone number, and address of the referral household.

A pretest of this procedure was conducted in order to see if respondents (1) would correctly identify eligible neighbors with gardens, and (2) were able and willing to supply the information needed in order for contact with the referral to be made. The local Raleigh-Durham-Chapel Hill area provided an ideal setting for a rigorous pretest, being abundant with potential problem locations such as heavily wooded or fenced neighborhoods, unusual street and property configurations, rural areas where houses are far apart, and high density urban areas.

In summer, Chapel Hill, for example, looks like a downtown and campus completely surrounded by areas of thick woods and heavy vegetation. It's not until autumn that the houses hidden behind the trees actually become visible from the street. Our concern was whether in this type of neighborhood respondents would know if their neighbors had gardens.

Another potential problem that concerned us was neighborhoods that are not laid out in the normal pattern of rows of houses set in rectangular blocks. In Chapel Hill that street pattern is the exception rather than the rule, for quite often streets are irregular and houses are situated in odd directions. The house that is technically next door might actually be situated around a bend, in the woods, and on the other side of a ravine. We were concerned that in this type of situation respondents would be unable to correctly identify next door neighbors and report their gardens.

For the pretest the local area was scanned and a few of these "worst case" neighborhoods, all of which contained gardens, were selected. Each neighborhood was mapped out and the telephone numbers of selected residents were obtained. Several residents in each neighborhood were then called and administered the questionnaires, and their responses were compared to the known information about the neighborhood.

In every case, the obscure garden was reported, and proper identifying information for the referral household was provided. In one instance a "possible" vegetable garden was reported for a dwelling that actually did not have one. This type of reporting error would have no effect on our coverage because a later interview at that referral dwelling would reveal the true situation. Such errors, if they occurred with high frequency, would of course have cost implications.

Encouraged by the excellent pretest results, we implemented the multiplicity counting rule plan on the nationwide random digit dialing telephone survey.

Problems and Results

We had considerable difficulty with the county linkage at the PSU selection stage. Even though large SMSAs had been set aside as self-representing PSUs, there were nevertheless numerous instances of central-office cities being located on county boundaries. Our rules called for combining counties when this arose, which created some very large PSUs. This type of problem could be handled in the future by linking a city to a second county only if a specified percentage, for example 20 percent or more, of the population of the city was contained in the second county.

Other confusion occurred because of multiple cities of the same name located in the same state.

The questions asking about the vegetable garden status of neighboring dwelling units had the desired effect of capturing information for a much larger number of households than was actually selected into the sample. On the average, slightly more than one neighboring dwelling was reported, per sample dwelling interviewed (Table 1). Single family homes averaged 1.51 referrals while two and three family homes averaged one and two referrals respectively, as would be expected from the linking rules employed. For these three types of units combined, the average number of referrals was 1.48.

As you can see in Table 2, the percentage of households with home vegetable gardens varies from a high of 45 percent for single family dwellings to 35 percent for two family dwellings and 11 percent for three family dwellings. Recall that referrals were not asked for when interviewing households in structures containing 4 or more dwellings. Note that only about 5 percent of these latter dwellings reported having a home vegetable garden indicating little loss by omitting these dwellings from the

Table 1

Distribution of Sample Dwellings and Average Number of Referral Dwellings Per Sample Dwelling by Number of Dwellings in Structure of Sample Dwelling

Number of Dwellings in Structure of Sample Dwellings	Sample Cases		Population Estimates			
	f	%	%	%	Avg. # Referrals Per Sample Dwelling	
1	2819	66.1	67.6	75.0	1.51	
2	150	3.5	4.2	4.7	1.00	1.48
3	48	1.1	1.4	1.6	2.00	
4+	663	15.5	16.9	18.8	-	
NA*	312	7.3	9.9			
NI	272	6.4				
Total Excluding NA & NI				100%	1.21	
Total Including NA & NI	4264	100%	100%		1.09	

*NA = Not ascertained. NI = Noninterview

Table 2

Percentage of Dwellings Reporting a Vegetable Garden and Percentage of Referral Dwellings Reported to Have a Vegetable Garden by Number of Dwellings in Structure of Sample Dwelling

Number of Dwellings in Structure of Sample Dwelling	Population Estimates of Percent of Dwellings Reported to Have Vegetable Garden	
	Sample Dwelling	Referral Dwelling as Reported by Sample Dwelling
1	45.4	28.1
2	35.4	21.4
3	10.7	9.7
4+	5.2	0
NA & NI*	-	-
Total including NA & NI	33.2	
Total excluding NA & NI	36.9	
Total excluding 4+, NA, & NI	44.2	27.3

*NA = Not ascertained. NI = Noninterview

Table 3

Result of Asking Sample Dwelling for Information About Neighbors

Report by Sample Dwelling:	Sample Cases f	Sample Cases %	Population Estimates %
1. No vegetable garden at neighbor dwelling	3363	75.3	74.9
2. Vegetable garden at neighbor dwelling	1103	24.7	25.1
Total	4466	100.0%	100.0%
Identification Information Reported			
a. Name, full address, phone number	165	15.0	15.6
b. Name, street name, phone number	128	11.6	13.2
c. Name, no address, phone number	24	2.2	1.0
d. Name, street name, no phone number	401	36.4	36.9
e. Name only	59	5.3	5.6
f. Phone number only	4	0.3	0.0
g. Insufficient or no information	322	29.2	27.6
Total	1103	100.0%	100.0%

referral procedures. The percentage of referrals reported to have a home vegetable garden is considerably smaller than that for the sample dwellings. For single family dwellings only 28 percent of the referrals are reported to have a garden as compared to 45 percent for sample dwellings. A similar relationship is found for duplex dwellings where the corresponding percentages are 21 as compared with 35 percent. One possible explanation of this underreporting is that very small gardens at referral dwellings might not have been noticed and consequently not reported. Somewhat over a third of single family dwelling gardens are under 200 square feet in size and about half of duplex dwelling gardens are that small.

Our optimism about obtaining names and phone numbers of neighboring vegetable gardeners was not reinforced as much in the regular survey, as it had been in the pretest. We found that in only about 70 percent of the cases was sufficient information provided to make contacting the neighbor possible (Table 3). Though we asked for the name, address, and phone number of each neighbor reported to have a vegetable garden, the complete information was provided only 16 percent of the time. Partial but sufficient information was provided somewhat over half of the time, and insufficient information was provided 28 percent of the time.

The somewhat disappointing nationwide results following a very encouraging pretest made us wonder whether Southerners were more neighborly than people in other regions. We also were concerned about whether the degree of urbanization had any effect on the quality of information provided about neighbors, and also whether people who are gardeners themselves provided better information about neighboring gardeners.

Having a vegetable garden is in fact associated with a somewhat higher percentage usable referral identification. Seventy seven percent of gardners provided good referral information as compared with 66 percent for nongardeners (Figure 1). A similar relationship is found for those living outside of the twenty largest metropolitan areas. Seventy five percent of this group provided usable identification information as compared to only 61 percent for those living inside one of those large SMSAs (Figure 2). Geographic region tabulations show that 8 out of 10 people in the South and North Central states provided usable referral identification as compared to only 6 out of 10 in the combined Northeast and West,* (Figure 3).

When size of community is crossed with geographic region, the Northeast-West combination is seen to be consistently poor at providing usable referral information, both for large urban and other places (Figure 4). In the South a large percentage of people in both community-size groups provided good information. In the North Central States, there was a notable difference between the large metropolitan and other areas, however, with only 63 percent in the former group providing usable information as compared to 81 percent in the latter.

Implications for the Future

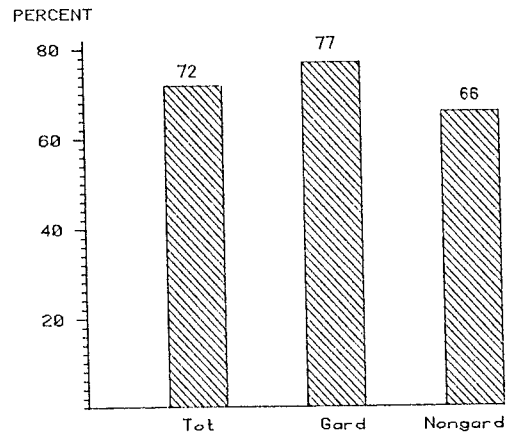
Of those referrals classified as having "sufficient information to contact," what proportion were we actually able to contact? What proportion of those provided an interview, and of those that did, what proportion actually had gardens? Is there any evidence to suggest that it was only the smaller gardens that were missed in the referral process? We have not yet been able to tabulate the data obtained from interviewing the referral households, but doing so should provide some additional information on the quality of this type of referral procedure.

Although this particular investigation focused on the use of a soil enrichment product that is infrequently used on home vegetable gardens, similar referral procedures could be adapted for use on a wide variety of rare characteristics. Some examples are the ownership of outdoor home recreation equipment such as jungle gyms or swimming pools, ownership of campers or house trailers, or even personal characteristics such as persons in certain age or racial groups or those having certain physical handicaps.

Because of the increasing use of telephone surveys to provide fast and comparatively cheap survey information, certainly more exploration of the possibilities for adapting telephone surveys to a wide variety of problems is called for. The use of multiplicity counting rules could conceivable reduce the magnitude of a screening effort considerably, when searching for rare characteristics.

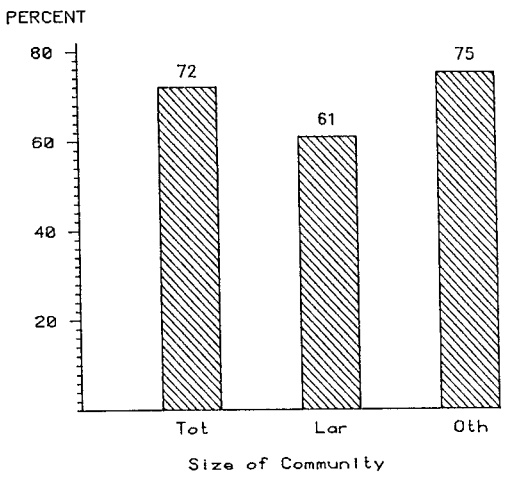
*The Northeast and West were combined because the number of PSUs in the West was considered to be too small to support separate estimates.

Figure 1.
Percent of Referral Dwellings
Having Sufficient Contact Information



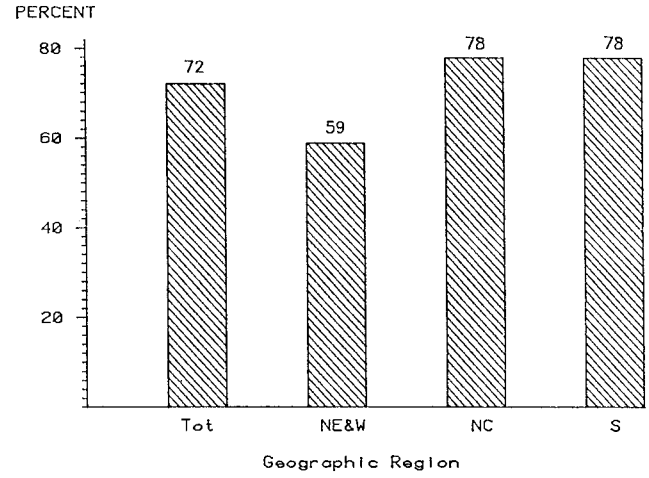
Sample Sizes for Garden Status: Total = 1103
Gardeners = 637 Non-Gardeners = 466
LEGEND: Tot = Total Gard = Gardeners
Nongard = Non-Gardeners

Figure 2.
Percent of Referral Dwellings
Having Sufficient Contact Information



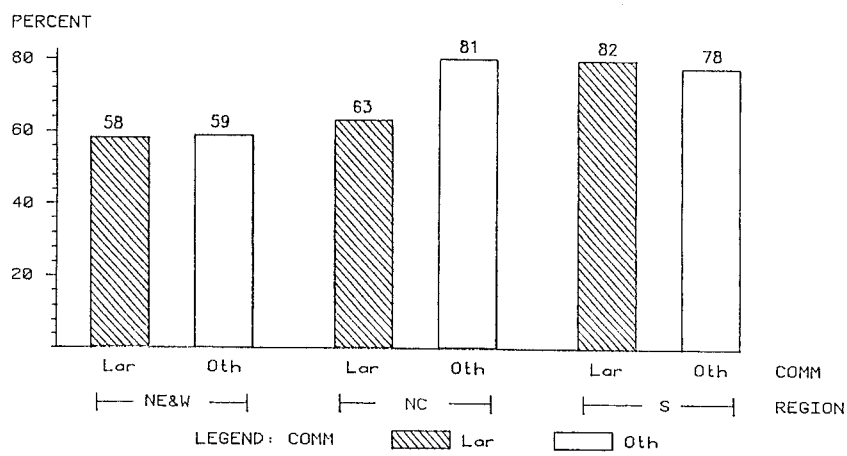
Sample Sizes for Size of Community:
Total = 1103 Largest 20 SMSAs = 277 Other = 826
LEGEND: Tot = Total Lar = Largest 20 SMSAs Oth = Other

Figure 3.
Percent of Referral Dwellings
Having Sufficient Contact Information



Sample Sizes for Geographic Region: Total = 1103
NE&W = 359 NC = 411 S = 333
LEGEND: Tot = Total NE&W = Northeast & West
NC = North Central S = South

Figure 4.
Percent of Referral Dwellings
Having Sufficient Contact Information



LEGEND: Lar = Largest 20 SMSAs Oth = Other NE&W = Northeast & West
NC = North Central S = South
Geographic Region and Size of Community