

SOME RESULTS OF AN EXPERIMENT WITH TELEPHONE SAMPLING FOR THE U.S. - NATIONAL CRIME SURVEY

Charles Alexander, Janice Sebold and Pamela Pfaff, U.S. Bureau of the Census

I. Introduction

This paper describes the initial results of an experiment with random digit dialing (RDD) telephone sampling for the National Crime Survey (NCS). The experiment was part of the U.S. Census Bureau's development and testing of a centralized computer-assisted telephone interviewing (CATI) facility in Hagerstown, Maryland.

The NCS presently uses a two-stage stratified cluster sample of addresses (from address lists, maps, and building permits) to select households for interview. A rotating panel of approximately 60,000 households (10,000 per month), are interviewed seven times at six month intervals. All persons age 12 years and over are interviewed about crimes which may have occurred to them in the previous six months. Only the first respondent is asked questions about "household crimes", such as burglary. All respondents are asked about "personal crimes", such as assault.

The first interview at an address is in person, although some hard-to-reach persons may get their first interview by telephone.

In the first visit, a household respondent provides a roster of eligible persons, which is updated on subsequent interviews. Each interview provides a "bound" for the next interview. The initial crime interviews presently are not used in regular crime estimates, because they have no such bound. Subsequent interviews are conducted by telephone whenever possible, from the interviewer's home. An exception is the fifth interview at an address, which is conducted in person to re-establish face-to-face contact. Also, if a household's occupants have moved, the initial interview for the new occupants is done in person. For budgetary reasons, the proportion of telephone interviews was increased in 1980 and again in April 1986 with no noticeable effect on the survey estimates. Further details of the NCS design are given in the NCS Survey Documentation.

Random digit dialing has been proposed as a potentially efficient method for obtaining part of the NCS sample; a dual address sample would undoubtedly be maintained, if only to give representation to households which cannot be reached by telephone. A major advantage is that the RDD sample would be less clustered geographically, which would improve the variance of NCS estimates. Groves and Lepkowski (1985) use a cost model to show that, under a wide range of assumptions about relative bias, mean square error could be reduced by such an RDD/address dual frame sample. (The cost model, however, does not reflect the current high proportion of telephone interviewing in the address frame.)

RDD probably would be implemented from a centralized CATI facility, which has other potential advantages in monitoring and quality control of the interview process. These advantages are also possible through "warm-contact CATI", in which the present design would be used, with most telephone interviews being completed from the CATI facility, rather than the interviewers' house.

"Warm-contact" CATI and "cold-contact" RDD CATI share many common developmental tasks. It was decided that initial testing and experimentation with CATI for NCS would use an RDD sample. The experiment discussed in this paper consisted of several parts:

April - December 1985

Test of RDD and CATI procedures, using two versions of the questionnaire.

October - December 1985

Telephone follow-up of uncontacted cases from earlier months.

June 1985 - June 1986

Include Sunday interviewing increasing the interview period from 12 to 14 days.

January - March 1986

Extended interview period: the period for completion of interviewing was extended, using one version of the questionnaire.

April - June 1986

Second interview for households contacted six months earlier. Movers were followed by telephone.

The next section of the paper describes the background and objectives of the experiment. Subsequent sections present results of specific parts of the study. The paper concludes with a discussion of certain operational issues related to RDD sampling, and a summary of the results.

II. Background and Objectives of the Experiment

The experiment was preceded by earlier developmental work for the NCS. From January to April 1984, 863 RDD cases (approximately 216 per month) were interviewed from Census Bureau headquarters in Suitland, Maryland to develop a CATI version of the NCS questionnaire. During this test, lessons were learned about various aspects of the system: selection and training of interviewers, amount and kind of supervision, computer hardware and software, scheduling of calls, and the introduction and organization of the questionnaire. Of all those sampled telephone numbers which may have been residential, 75.9% had complete interviews of at least one resident. Improvements in the questionnaire and procedures were suggested based on the lessons learned.

In this and the later RDD work, sample telephone numbers were selected using the two-stage method described in Waksberg (6). A cluster of four units was used at the secondary stage of selection. Primary screening for each month's sample was done during the last two weeks of the previous month. The first four units in each cluster had their status resolved before any additional numbers in the cluster were called. If a sample number was determined to be ineligible (i.e., a nonworking or nonresidential number), it was replaced by another number. If a number was residential or potentially residential, it was kept as a final sample unit, even if it was a refusal or "a persistent ring-no-answer which could not be identified as non-residential or non-working" (RNA). Further details on the mechanics of sampling are discussed in the final section.

The small 1984 test was regarded as the first step in a "research cycle", consisting of alternating "development" and "evaluation" stages. The system developed in the first test at Suitland was to be evaluated in the next stage at the telephone facility in Hagerstown, with a total designated sample size of 456 households per month for at least six months. During the evaluation stage, the system was to be left in place, with a minimum of tinkering, long enough to get an accurate measure of its performance. However, since the experiment began shortly after the facility opened, there were some unavoidable operational and personnel changes during the first months of the experiment.

Thus, the main objective of the test at Hagerstown was to evaluate the improved questionnaire and procedures under production conditions. The response rate and percent distribution of various types of nonresponse were of special interest. The effect on crime rates will also be studied, but this analysis has not been completed. In addition, some data for cost models have been obtained.

An important secondary objective was to study the causes and characteristics of nonresponse, so that further improvements could be suggested for subsequent evaluation. Two versions of the questionnaire were tested, with different locations of the household

roster. The "ring-no-answer" cases were contacted several months after the survey closeout to determine whether they were residential and to find out why they could not be reached the first time. When refusals occurred, the time and position in the questionnaire were recorded.

In January 1986, several changes were made to try to improve response rates further. The period for completing interviews was increased from two weeks to four, and changes in question wording suggested by interviewers and observers were adopted. These new questions and procedures were used, with the seemingly preferable roster location, from January through March 1986.

Between April and June 1986, units which had previously been interviewed were contacted for a second interview. If persons had moved, an attempt was made to reach them at their new telephone number, using contact information obtained on the first interview. A questionnaire containing minor revisions was used.

This April-June 1986 part of the experiment was a "development" phase, where the objective was to learn about the problems involved in using a telephone sample as part of a longitudinal panel survey, so that improved recontact procedures could be developed for future testing. The MCS does not presently follow movers, but this feature has been suggested as a possibility for an upcoming survey redesign.

The following sections discuss each part of the experiment in turn.

III. Response and Refusal Rates for the April-December Phase

The average RDD response rate for the April through December 1985 period was 75.7%. The base for this includes units known to be eligible occupied residences plus all those which might have been eligible, making this a lower bound on the "actual" response rate. A unit is counted as a response if at least one person in the household completed the household and personal screen questions and any incident reports. The average refusal rate for the same period was 11.2%. The average "other noninterview" rate was 13.1%.

"Other noninterviews" include cases in which everyone at the sample telephone number was hospitalized, ill or absent for the entire survey period, incomplete callbacks, language barriers, never contacted and never tried cases. Sunday interviewing began in June 1985, increasing the survey period from 12 to 14 days.

One problem in classifying noninterviews is that the classification controls the subsequent handling of the case by the computer system. This can lead to potentially misleading classifications, in order to handle the case properly. In particular, interviewers found that it was necessary to classify a case as an "other noninterview" when the respondent immediately hung up the phone. This allowed the case to stay in the system for subsequent attempts by the interviewer to convert the refusal. (The usual refusal procedure at the Hagerstown facility was that after the initial refusal, the case was removed from the regular flow of work and the shift supervisor called the case and attempted to convert the refusal.) This problem occurred mainly in May. In June, these cases were identified and reclassified as refusals.

In July 1985, the CATI instrument was revised to include an "immediate hang up" outcome that allowed the interviewers to call the case again.

While it was our prior conception that the refusal conversion call made by the shift supervisor would take longer to complete than the initial refusal call made by the interviewer, this was not the case. The first refusal call lasted 6.25 minutes on the average. The second refusal call averaged 3.57 minutes in duration. The combined first and second refusal calls took an average of 5.21 minutes to complete. The shift

supervisors were able to convert 30% of the first refusals taken by interviewers.

Examining all the refusals from both versions of the questionnaire, 86 percent of the refusals occurred during the introduction to the RDD interview (this is before the household roster, even in the roster-first questionnaire). These refusals were concentrated at three main points in the introduction. The initial flow of the interview was for the interviewer to identify him or herself as calling from the Census Bureau, explaining that he or she was calling randomly selected numbers to conduct the National Crime Survey, and verifying the phone number. Sixteen percent of the refusals occurred at this point in the introduction. Next, the respondent was asked if he or she had been reached on his or her home phone. Only a few refusals occurred here. Then, the interviewer asked to speak with the male or female head of the household. Thirty-one percent of the refusals happened at this point. After the interviewer was speaking with the head of the household, he or she offered a detailed explanation of the survey. This explanation read, "We are calling selected telephone numbers throughout the United States to obtain statistics on the kinds and amounts of crime committed against individuals 12 years of age and older." Another 31 percent of the refusals took place at this point in the interview.

The remaining refusals were scattered throughout the rest of the interview. In all, 94% of refusals occurred before the first crime question was asked.

IV. Ring-No-Answer Follow-up

In 1984, both the National Crime Survey and the National Health Interview Survey (NHIS) conducted RDD studies. Both studies experienced a large number of cases for which no contact was made at the sample number. Both surveys conducted a follow-up of the ring-no-answer cases to determine why the cases were not contacted during the original survey period. Both follow-up studies were able to resolve about 80% of the RNA cases and found that about 60% of the previously unresolved RNA cases were residential phone numbers.

To study the nonresponse rate for the NCS RDD experiment at the Hagerstown Facility, the earlier RNA studies were duplicated, demographic questions were added, and the RNA questionnaire was programmed for use on the CATI system.

In October 1985, RNA cases from April through September 1985 were called for a one week period. Then, in November, all the unresolved cases from April through September plus the RNA's from October were called. Finally, in late December, all the unresolved cases from the previous months plus RNA's from November and December were followed up.

Altogether, 237 RNA cases were included in the follow-up. As in the previous studies, the interviewers were able to resolve about 83% of the RNA cases. A breakdown of the resolved and unresolved cases is displayed in the following table. Note that 44% of all the RNA cases are resolved as nonresidential or nonworking and 38% as residential. Of course, some of these could have changed status since the original interview.

Table 1
Resolution of the Ring-No-Answer Cases

	Count	% of Total
Total number of cases	237	100.0%
Resolved Cases	196	82.7%
Residential	91	38.4%
Nonresidential	42	43.9%
Nonworking Numbers	62	
Refusal	1	.4%
Unresolved Cases	41	17.3%
Unreached Numbers	24	10.1%
Partial Refusals	6	2.6%
Tracing Call through TBO	1	.4%
Callback arranged -	10	4.2%
no information retained		

These results showed some inconsistencies with the 1984 studies. Only 46% of the resolved cases were residential as compared to about 60% for the earlier studies. One reason for this inconsistency may be the large number of nonworking telephone numbers included in the resolved cases. This could be produced by differences between the 1984 and the 1985 studies in the lag between the original interview and the follow-up. A more detailed analysis of the 1985 data shows that the months furthest away from the follow-up survey have more nonworking telephone numbers. We have not yet examined the comparable 1984 data to see whether this effect could explain the inconsistencies.

The following table displays the reasons people listed for not answering their phone during the original survey period.

Table 2
Reasons for Noncontact During
Original Survey Period
(Residential Cases)

Reason	Percent of Total
Busy households/Work/Never home	33.4%
On vacation/Absent entire survey period	32.2%
Seasonal residence	12.2%
In the process of moving	6.7%
At school	3.3%
Unplugs phone	2.2%
Don't know/No reason	10.0%

Note that 44% of the response are in categories (on vacation, seasonal residence) where it is unlikely that more persistent calling during the interview period might have reached the person and at least 43% are in categories (busy households, no reason) where more persistent calling would seem to have promise.

Table 3 provides a demographic comparison of the RNA follow-up to the roster-first version of the questionnaire.

Table 3
Demographic Comparison of the Ring-No-Answer
(RNA) and Random Digit Dial-Version A (RDA)

	RDA	RNA
Marital Status of Head of Household		
Married	60.6	44.0
Widowed	12.8	7.7
Divorced	10.3	2.2
Separated	2.1	8.8
Never Married	12.9	23.0
DK/Refused/Blank	1.3	14.3
Sex of Head		
Male	56.1	48.3
Female	43.6	37.4
Blank	.3	14.3
Race of Head		
White	85.6	79.1
Black	9.2	5.5
Other	3.9	2.2
DK/Refused/Blank	1.3	13.2
Age of Head		
Under 20	1.7	1.1
20-34	29.0	23.1
35-49	27.9	23.1
50-59	13.0	16.5
60-64	6.7	11.0
Over 65	19.8	8.7
DK/Blank	1.9	16.5
Average Household Size	2.6	2.0

The RNA cases have proportionally more never-married and fewer married persons, as well as fewer persons over 65 years old. (These results are statistically significant at the .05 significance level, ignoring within-cluster correlation.)

V. Roster First/Roster Last Comparison

An important goal of the experiment was to measure the effect of the location of the household roster on the response rate. In the regular NCS, after the introduction but before any crime questions are asked, a roster of the names, ages, family relationships, and associated demographic variables of all persons in the household is obtained. This

list is used in administering the subsequent interviews of these persons, and for making noninterview adjustments for persons who cannot be interviewed.

For "cold-contact" telephone interviewing, it was suggested that the household roster should be obtained later in the interview, after the first person's crime questions. This might tend to reduce the number of refusals by demonstrating as quickly as possible that the survey has a serious purpose and is indeed about crime. In earlier RDD work, interviewers and observers alike commented that the household roster was received with greater suspicion than the crime questions. The argument against moving the household roster was that it might lead to worse coverage of persons within households. If the crime questions are found to be burdensome by the first respondent, the respondent may deny the existence of other household members to spare them the same burden. Of course, the contrary could also occur: the first respondent might be less suspicious after finishing the crime questions and thus might be more inclined to give an accurate roster.

To investigate these issues, two versions of the questionnaire were used during the April-December 1985 part of the experiment, a "roster first" and "roster last" version. Response rates, refusal rates, average household size, and proportion of one-person households were recorded each month for each version of the questionnaire.

The basic design called for eight interviewers, divided into two groups of four. In the first month (April) Group 1 used the roster-first version and Group 2 used roster-last. In May the two groups switched versions. Group 1 used roster-first in April, June, August, October, December and Group 2 used roster-first in May, July, September, and November. In a given month, the computer system would not permit an interviewer to sign on to the wrong version. Two other alternative designs were considered. One would assign each interviewer at random to a treatment for the entire experiment. The other would have randomly assigned each sample telephone number to a treatment and had interviewers work on cases as they were brought up by the automatic call scheduler. The chosen design was adopted based on several considerations.

1. Because of the small number of interviewers, between-interviewer variance may have had a dominant effect on comparisons between versions if each interviewer worked only on one version.
2. The second alternative would not have controlled the assignment of interviewers to cases. The RDD operation allows several interviewers to work on each case, depending on what hours they work. Some discretion is available to the interviewers and supervisors about which interviewer will take a given case. If interviewers could work on both versions during the same time period, it would be impossible to control the assignments of experimental treatments to the interviewers without major changes in the operation. Since the primary purpose of the whole experiment was to test the RDD operation under production conditions, enforcement of strict experimental controls on call scheduling was ruled out.
3. The interviewers extensively practiced the introductions to different parts of the questionnaire and rehearsed responses to objections the respondents might raise. Being able to concentrate on one version of the questionnaire each month potentially could enhance their proficiency in moving smoothly through the questionnaire, compared to frequent switching of introductory sections.

The chosen design has several potential disadvantages. If interviewers quit during

the experiment, the balance of the "cross-over design" is reduced. If the two groups have different "learning curves," this must be taken into account in the model for analysis. However, since these effects can be recorded and modelled explicitly, they were thought to be preferable to the unmeasurable lack of control of the second alternative.

The facility managers were instructed to treat the two versions of the instruments identically as far as training, monitoring, etc. Observers reported that the training seemed even-handed. However, from the start, the interviewers volunteered the opinion that roster-last seemed preferable. This preference may have been reinforced as the monthly response rates at the beginning of the experiment seemed to favor roster-last. In feedback to the interviewers, emphasis was placed on reducing refusal rates in the first two months of the experiment. Starting with June, further emphasis was placed on high overall response rates. Throughout the experiment the interviewers were aware of their performance on refusal and response rates, and conscious of the relationship of this to the questionnaire version. By contrast, there was no discussion of household size distribution.

The managers took steps to keep the composition of the groups constant. However, the difficulty of controlling staff turnover under production conditions is illustrated in Table 4. The different interviewers are represented by letters. The changes were due to resignations or promotions. In some cases, additional interviewers had to be added to the group of four to replace others who could not work the full month. (Cases in which interviewers made one or two calls on a given instrument have been omitted.) The basic results of the experiment are given in Table 5.

Table 4
Composition of the Interviewer Groups

	Group 1	Group 2
April	ABCE	ZYWX
May	ABCD	ZYWX
June	ABCD	ZYWX
July	ABCD	YWRT
August	ABCDZ	YWRT
September	ABCD	YWRTZ
October	ABCDZ	YWRT
November	ABCDF	YRSQP
December	ABCDFG	YRSQP

Table 5
Response Rates and Refusal Rates
for the Two Experimental Treatments

	Response Rates		Refusal Rates	
	Roster First	Roster Last	Roster First	Roster Last
April	73.2*	75.4	15.8*	13.2
May	68.0	76.3*	14.5	9.2*
June	79.4*	75.9	9.2*	7.3
July	70.2	81.1*	13.6	7.0*
August	71.0*	74.6	9.6*	14.9
September	80.3	80.3*	11.8	7.0*
October	77.6*	75.0	10.1*	13.6
November	73.2	78.5*	15.8	7.0*
December	74.1*	78.1	11.4*	10.5
TOTAL	74.1	77.2	12.4	10.0

*Denotes Group 1

The intended analysis was a two-way analysis of variance (ANOVA) using an additive model including the two groups and two treatments with the response variable being the monthly response or refusal rate for a given group and treatment. (Thus $n = 18$.) This analysis shows that using the roster-last questionnaire reduced the refusal rate by 2.8 percentage points and increased the response rate by 3.4 points compared to the roster first questionnaire. (These differences were statistically significant at the .05 significance level under the usual ANOVA assumptions.)

These results should not be taken as definitive at this stage of the research. As was feared, the analysis is complicated by

changes in composition of the groups. This can be taken into account in two ways:

- changes in the groups can be represented by different levels in the group variable in the ANOVA. For example "Group 2" could be defined to be one group 2A from April to June, group 2B from July to October, and group 2C in November and December. For some such models, the difference due to roster placement is still estimable.
- a measure of individual interviewer performance can be used to define a measure of the monthly group's quality, which can be used as a covariate in a one-way analysis of the roster placement effect.

Several models as described in a. have been tried, with the conclusion that the roster-last refusal rate is 2-3 points lower than roster-first and the response rate is 2.5 - 3.5 points higher. Analyses of type b. have not yet been done. Since the vast majority of refusals occur before the household roster section, this observed difference in refusal rates is probably due to something more complex than a direct respondent resistance to the household roster. One speculation is that interviewers may start out more confidently using a smoother-flowing instrument.

Further complications are:

- The procedures were instituted for the first time in April, with some warm-up in March.
- April, May, and June are affected by the outcome classification issues described in Section III.
- The "response variables" in the ANOVA may not be independent, for numerous reasons.
- The difference between roster-first and roster-last may have been affected by the fact interviewers used both versions.

Deleting the first month, April, did not substantially alter the results of the original ANOVA of response and refusal rates.

Turning to the effect of roster placement on household size, very little difference was found. The average household size for the roster-first instrument was 2.62. The roster-last had an average household size of 2.68. The average household size for the Current Population Survey is 2.6.

There is no indication that moving the roster reduces the within-household coverage. There was little difference between the number of people listed in the household in each version of the questionnaire.

In conclusion, although further analysis of these data is desirable, the analysis done so far leads to the conclusion that the roster-last version is preferable. Indications are that the later roster placement provides some improvement in refusal rate, or alternatively overall response rate, although the exact improvement cannot be estimated definitively because of confounding variables. This general conclusion agrees with the subjective impression of interviewers and observers. There is no indication of deterioration in within-household coverage due to the later roster placement.

VI. Extended Closeout

In another effort to increase the RDD response rate, the NCS closeout was extended to give the interviewers four weeks, instead of two, to contact the RDD cases. Only the roster-last version of the instrument was used for the extended closeout for the reasons given in Section V. Several minor wording changes and a reduction in the number of labor force questions were incorporated in the roster-last instrument for the extended closeout that was conducted between January and March 1986.

Table 6 displays the response, refusal and other noninterview rates for the January through March 1986 period after the 2-week interviewing period, for the full extended

closeout period, and the regular April through December 1985 two-week interview period.

Comparing the final extended closeout rates to the regular two-week rates, the response rate went up 2.9 percentage points, the refusal rate went up 3.3 percentage points, and the other noninterview rate went down 6.2 percentage points. An obvious explanation would be that the last two weeks, a number of "other noninterviews" were resolved, some as interviews and some as refusals. However, this explanation is too simplistic, since the first two weeks of the extended closeout are not identical to a two-week interview period.

Table 6
NCS/RDD Response and Refusal Rate Comparisons
by Length of Interview Period

Jan.-Mar. 1986 NCSRDB (roster last)
Using Extended Interview Period

After Regular 2-Week Interview Period:

Average Response Rate - 74.1%
Average Refusal Rate - 11.4%
Average Other-NI Rate - 14.5%

After Extended^{1/} Period

Average Response Rate - 80.1%
Average Refusal Rate - 13.3%
Average Other-NI Rate - 6.6%

Apr.-Dec. 1985 NCSRDB (roster last)

(Two-Week Interview Period)

Average Response Rate - 77.2%
Average Refusal Rate - 10.0%
Average Other-NI Rate - 12.8%

VII. Sunday Interviewing

Census Bureau field procedures traditionally have excluded Sunday as a day to conduct personal interviews for fear of negative respondent reaction. Yet research from the private sector indicates that Sunday is a productive interviewing day. In an effort to improve the RDD response rate and examine the effects of Sunday interviewing, a preliminary test of Sunday interviewing for the RDD experiment was conducted on the second Sunday in June 1985. With a positive reaction from both interviewers and respondents, Sunday interviewing was conducted in subsequent months on both Sundays in the two week interviewing period. Sunday interviewing was conducted between 1:00 pm and 9:00 pm respondent's time. Interviewers were instructed to make an appointment for a callback on another day if there was resistance to calling on Sunday.

Results from Sunday interviewing indicate that Sunday is much like any other day of interviewing with a few minor exceptions. Interviewer productivity increased on Sunday due in part to more ring-no-answer cases and less noneligible phone numbers. These results are to be expected since many businesses are closed on Sundays. Significantly fewer busy signals and fewer nonworking phone numbers were encountered on Sundays.

The Current Population Survey is also conducting Sunday interviewing at the Hagerstown Facility with similar results. The Census Bureau will continue to conduct Sunday interviewing for CPS and NCS at the centralized telephone facility.

VIII. Second Contact

During April, May and June 1986, RDD households interviewed in October, November and December 1985 were called for a six-month, follow-up interview. The purpose of the second contact test was to get an idea of the problems which arise in an RDD follow-up interview. In particular, we wanted to test the alternate contact information we collected in the initial RDD interview. In addition, we wanted to test our ability to follow movers by phone as well as test the CATI system's ability to create "spin-off" cases for people

^{1/} January extension was an additional 13 days. February extension was an additional 10 days. March extension was an additional 12 days.

who move away from the sample household sometime during the six-months between interviews.

The average response rate for the second contact survey was 80.8% of those units which responded on the first contact. The refusal rate was 5.6% and the other noninterview rate was 13.5%.

Unpublished or incorrect telephone numbers made up a large portion of the second contact nonresponse rate. Although alternate contact information was obtained in the original RDD interview and used to contact those people whose numbers had changed in the six month period between interviews, this information was often incomplete. Greater emphasis on improving the alternate contact information may have improved the second contact response rate. As it stands, however, there is the potential for a serious panel attrition problem over the seven NCS interviews. We were able to create "spin-off" cases on the CATI system for people who moved away from the sample household. This completes the first step in the research cycle for considering longitudinal studies on the CATI system.

IX. Other Sampling Issues

A. Cases encountered in Telephone Sampling

A simple exposition of the two-stage RDD sampling procedure would say that the telephone numbers dialed in both the primary and secondary stage are either eligible ("residential") or ineligible ("nonresidential"). Our experience showed that the situation is more complex. For sample maintenance purposes, at least six different situations may need to be distinguished:

- 1) The telephone number is determined to be residential through contact with the household. (Residential interview, partial interview, or refusal.)
- 2) The number is identified as residential by the telephone business office (TBO) or otherwise, but is never contacted after a specified cutoff number of RNA calls. (Confirmed residential RNA.)
- 3) Identified as residential by the TBO or otherwise, but never reached because there was not time to make the required cutoff number of calls. (Unfinished residential RNA.)
- 4) Not identified as either residential or nonresidential by TBO or otherwise, never contacted after the specified number of RNA calls. (Completed indeterminable-status RNA.)
- 5) Status not determined, case not completed because of lack of time. (Unfinished indeterminable-status RNA.)
- 6) Identified as nonresidential, either by someone at the number, by the TBO, or by a recorded message. (Ineligible number.)

Thought needs to be given as to which of the six situations will be counted as eligible or ineligible during primary and secondary selection. The answer may depend on the nature of the survey. For a one-time survey, only case one might be declared eligible, since little useful information is obtained about the other numbers, which may not be occupied residences.

For a continuing survey, cases two through five might be made eligible nonrespondents in the hope that some information could be obtained on later interviews. Cases two, three, four and five are distinguished from one another because they may have different probabilities of turning out to be occupied residences. (Note that the cases which are eligible for sampling operations need not be treated as eligible in preparing estimates.)

It is advantageous for the same definition of eligible and ineligible to be used at both stages of selection. Otherwise there is a selection bias. (Alexander, 1984). For example, if some nonresidential numbers are eligible at the primary stage but not at the

secondary stage, then there will tend to be over-representation of all the residences in clusters which contain many such nonresidential numbers. In practice, the overall effect of this kind of bias may be small.

B. Operational Complications in Telephone Sampling

The questions of how long to keep calling a telephone number and how long to keep calling in a cluster with few residences have been discussed elsewhere. (For example, Chapman and Hogue (1984).) In our experiment, a limit of 20 calls to a number was used. There was no explicit limit on how many numbers would be contacted in a cluster; the decision to stop with less than four eligible numbers was made on a case-by-case basis when special problems were encountered.

For our experiment, a cluster selected during primary screening could be replaced if the original screened number proved to have been misclassified as eligible when it really was ineligible. This happened in less than 1% of clusters. The original number was recontacted to check its status if secondary screening did not produce any eligible numbers after about ten secondary numbers had been contacted.

In rare cases, clusters may need to be dropped or replaced even if the original number is found to have been eligible. One such instance occurred when a storm eliminated service for several weeks for a cluster between primary and secondary screening. In another instance, it was found that many or all of the numbers in a cluster reached the same residence. The TBO confirmed that there, was a problem. Apparently, the residences actual number was indeed dialed during screening. This cluster was dropped as an act of mercy. Although such cases are rare, a rule for handling problem cases would need to be formulated for an RDD production facility.

During secondary selection, if a number is determined to be ineligible, it is replaced by another number. It can take several days to determine the eligibility of a given number. This could cause problems for surveys with a short interview period and a large number of secondary units chosen from each cluster. For the NCS experiment, with a two-week period for interviewing and only four units per cluster, such problems were rare.

X. Summary

This section summarizes some conclusions from the earlier sections. The conclusions should be regarded as preliminary, and subject to the qualifications described in earlier sections. Many interesting questions remain to be analyzed.

A. Roster placement

Based on the experiment, it was decided to obtain the household roster at the end of the first respondent's interview. This apparently reduced the refusal rate somewhat, compared to obtaining the roster at the beginning, without any noticeable effect on the coverage of persons within the household. The reduction in the refusal rate is modest, apparently about 2-3%. This estimate of the reduction depends on assumptions about the effect of confounding variables such as interviewer experience. However, the alternative assumptions we have considered so far all lead to the conclusion that there is a modest reduction in the refusal rate.

B. Response rate for a two-week interview period

Using the roster-last questionnaire, the response rate was generally between 75-80% of all numbers which may have been eligible. This value varies with the experience and quality of the interviewers. The rate for a large production facility could of course, differ from the results for this small experiment. About 10% of all the numbers were refusals, and 10-15% were not interviewed due

to other causes. These "other cause" non-responses turned out to include some ineligible telephone numbers, or at least numbers which were non-residential or non-working when contacted several months later. Based on table 1, of the 10-15% "other cause" noninterviews, 44% can be attributed to refusals, 44% can be attributed to ring-no-answers which were ineligible, about 38% were eligible ring-no-answers, and the other 18% still cannot be identified. Eliminating the known ineligible would add on the order of 5% to the overall response rate by reducing the base. Thus, if one could eliminate these ineligible telephone numbers, the response rate would probably be in the range 80-85%. The eligible nonresponse households are somewhat different from the population as a whole, (Table 3).

C. Response rate for a four-week interview period

Table 6 shows that a four-week interview period gives about an 80% response rate, with 13% refusals and 7% other non-responses. We have no estimate of how many of the 7% are eligible households. For comparison, in regular NCS, there are typically about 2-3% refusals, and about 1-2% other noninterviews.

Comparing response rates from samples with different frames is difficult. To these 1-2% "other noninterviews", one may need to add up to about 5% undercoverage of households by the regular survey as compared to the decennial census (NCS data Alexander cited in 1986). These "undercovered" units are never located and not counted as noninterviews. Thus, with a four-week interview period, the "other nonresponse" problem for RDD may be comparable to the combined under-coverage and "other noninterview" problems in the address frame. (This ignores the major difference that the RDD frame does not include nontelephone households.) However, for the NCS, the refusal problem is much more severe for RDD than for the regular "warm-contact" interviewing.

D. Information About Refusals

The majority of refusals occur in the introductory part of the interview, before the crime questions are asked.

E. Sunday Interviewing

Interviewing on Sunday presented no serious problems, and will be done in the future.

Acknowledgements

The authors would like to thank Betty Hanna and William Ireland for their contributions to the random digit dialing research; Census Bureau staff for providing valuable comments, and Donna Brooks and Pat Maier for typing the manuscript.

References

1. Alexander, C.H. (1984). "NCS RDD Experiment: Procedures for Handling Cases with Indeterminable status." U.S. Census Bureau internal memorandum dated July 3, 1984.
2. Alexander, C.H. (1986). "The Present Consumer Expenditure Surveys Weighting Method." Population Controls in Weighting Sample Units, papers from a conference sponsored by the Bureau of Labor Statistics of the Department of Labor, March 18, 1986.
3. Groves, R.M. and Lepkowski, J.M. (1985). "Dual Frame, Mixed Mode Survey Designs." Journal of Official Statistics, 1, 263-286.
4. Hogue, C.R. and Chapman, D.W. (1984). "An Investigation of PSU Cutoff Points for a Random Digit Dialing Survey." Proceedings of the 1984 Survey Research Section of the American Statistical Association.
5. U.S. Bureau of the Census. National Crime Survey, National Sample Survey Documentation.
6. Waksberg, J. (1978). "Sampling Methods for Random Digit Dialing." Journal of the American Statistical Association, 73, 40-46.