BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-85-21

THE RESULTS OF THE 1984 NHIS/RDD FEASIBILITY STUDY: FINAL REPORT

by

David W. Chapman
Statistical Research Division
Bureau of the Census
Room 3524, F.O.B. #3
Washington, D.C. 20233 U.S.A.

Recommended by: Paul Biemer

Report completed: February 1985

THE RESULTS OF THE 1984 NHIS/RDD FEASIBILITY STUDY: FINAL REPORT

Prepared by the

Bureau of the Census

and the

National Center for Health Statistics

Washington, D.C.

Submitted to the
NCHS-Census Joint Steering Committee on
Telephone Surveys

February 28, 1985

Acknowledgements

The NHIS/RDD Feasibility Study was jointly sponsored by the National Center for Health Statistics (NCHS) and the Census Bureau. The joint agency steering committee, which guided the efforts of the project team, consisted of Earl Bryant and Monroe Sirken of NCHS and Barbara Bailar and William Butz of the Census Bureau.

This analyses report, which includes the findings of ten specific analysis projects, represents the collective efforts of a large number of people. Key members of the NCHS staff included James Massey and Owen Thornberry who were co-directors of the NCHS activities, and Joseph Fitti and Karen Frey who provided assistance on development of the data collection activities. Several people at the Bureau of the Census played active roles in the planning and conducting of the study. Evan Davey and Paul Biemer served as co-directors of the Census Bureau activities. James Avore and Al Giglitto were responsible for the data collection activities. Questionnaires, interviewer training materials, processing specifications and numerous activity reports were developed by Janis Brown, Dale Berglund, Robert Wilson, and Robert Mangold. David W. Chapman, Robert Tegels, and Carma Hogue provided sampling and statistical assistance. Barry Hill, Edna Vega, and Anthony Roman were in charge of the processing, computer programming, and tabulation of the data. The CATI procedures were planned and implemented by William Nicholls, William Mockovak and Warren Besore. The typing and reproduction of a large portion of the manuals, training packages, specifications, memoranda, activity reports, and of this analyses report, all of which were so crucial to the success of this project, were provided by Kristie Spivey, Lillian Stewart and Alice Bell.

David Chapman coordinated the preparation of this analyses report. Although several persons reviewed and commented on all parts of the report, the individuals listed below were responsible for preparing and revising specific sections and appendices. In cases of joint authorship, the analysis project team leader is listed first.

Sections	Authors
1, 2, 3	Paul P. Biemer, Census Bureau David W. Chapman, Census Bureau
4.1, Appendix 1	Anthony M. Roman, Census Bureau
4.2, Appendix 2	Theresa J. De Maio, Census Bureau
4.3, Appendix 3	Gregory E. Sliwa, Census Bureau
4.4, Appendix 4	James T. Massey, NCHS
4.5, Appendix 5	Robert J. Tegels, Census Bureau Paul P. Biemer, Census Bureau
4.6, Appendix 6	David W. Chapman, Census Bureau Anthony M. Roman, Census Bureau
4.7, Appendix 7	Michael C. Tenebaum, Census Bureau
4.8, Appendix 8	Michael E. Haas, Census Bureau
4.9, Appendix 9	William P. Mockovak, Census Bureau Joseph E. Fitti, NCHS Karen Frey, NCHS
4.10, Appendix 10	David W. Chapman, Census Bureau Carma R. Hogue, Census Bureau
Appendix 11	Janis L. Brown, Census Bureau R. Robert Wilson, Census Bureau

TABLE OF CONTENTS

FINAL REPORT
1984 NHIS/RDD FEASIBILITY STUDY

1. INTRODUCTION

In late 1982, the Bureau of the Census and the National Center for Health Statistics formed the Joint Agency Telephone Survey Task Force to plan a three-year program of research and development leading to the implementation of random-digit-dialing (RDD) sampling techniques (via a dual frame design) in the National Health Interview Survey (NHIS). In their final report and three year plan, the Task Force recommended that a feasibility study be conducted during the first quarter of 1984 to investigate a number of major issues involving the use of RDD in the NHIS. Subsequently, the 1984 NHIS/RDD Feasibility Study was conducted. The sample for the study consisted of about 1500 telephone households for each of two questionnaire versions. The data collection phase for this study was late January to May of 1984.

The following nine specific objectives were defined for this study:

1. Test the feasibility of conducting the entire core component of the NHIS questionnaire by telephone.

2. Estimate the response rate for the telephone component of a dual frame NHIS.

3. Estimate the costs for conducting the telephone component as part of a dual frame design.

4. Evaluate alternative questionnaire structures in terms of length and effect on estimates.

5. Identify operational problems associated with administering the NHIS by telephone.

6. Develop and evaluate procedures for identifying and handling special places over the telephone.

7. Conduct preliminary development and testing of estimation procedures, including nonresponse and post-stratification adjustments.

8. Test procedures for the assignment, management, and completion of samples for producing valid estimates.

9. Evaluate the operational feasibility and effect on response rates of using a most knowledgeable respondent rule.

In order to address these objectives and related issues, the following ten analysis projects were carried out:

1. Response Rates
2. Breakoff Analysis
3. Questionnaire Analysis
4. Respondent Rules
5. Interview Period/Sampling Frequency
6. Substitution
7. Special Places
8. Cost Analysis
9. Monitoring
10. Intracluster Correlations

The final reports for these ten projects are attached as Appendices 1-10. Summaries of these ten reports are given in Section 4. Section 3 contains a description of the sample design. The major findings of these projects and some conclusions are provided in the next section.

2. MAJOR FINDINGS

The NHIS/RDD Feasibility Study was highly successful at achieving its objectives. The major findings of the study are:

1. For NHIS, RDD response rates of 85% (for an average interview length of about 50 minutes) are feasible. However, interviewer and field staff experience are critical factors in achieving such rates. Over the 10 weeks of the study, response rates increased

by 12 percentage points and item nonresponse levels were substantially reduced.

2. The length of the NHIS interview and nature of the questions did not appear to cause any operational problems.
3. The Person-by-Section questionnaire version resulted in higher reporting of health events than did the Family/Individual version. Response rates for the two questionnaires were essentially the same.
4. The total time (including all interviewer activities such as dialings, callbacks, etc.) per interviewed household for this study ranged by replicate from an average of 70 minutes (earlier replicates) to 49 minutes (later replicates) for an overall replicate average of 61 minutes.
5. In more than 90% of the completed cases, the most knowledgeable respondent (as identified by the phone asnwerer) was reached on the first household contact. For cases requiring callbacks to reach the most knowledgeable respondent, the refusal rate was three times greater than the average.
6. The automated call scheduler performed efficiently in the later replicates where the number of unresolved cases dropped considerably (though it could still be improved). A longer interview period (say 4 weeks) could have increased the average response rate between 2 to 4 percentage points by reducing the number of unresolved cases.
7. There was a problem with identifying special places (using the NHIS definitions) in the RDD survey. Roughly 30% were classified as nonresidential units and 20% were classified as residential units other than special places.

8. Substitution seems to be a plausible method of accounting for nonrespondents. Compared to an equal-cost, PSU-by-PSU weight-adjustment-based nonresponse procedure, the estimates of variance of the estimated means for the five health characteristics were all lower for the substitution-based procedure.
9. Responses were obtained from about 74% of the substitutes that were generated (about 5 percentage points lower than the original survey cases). This lower response rate for substitutes could indicate the existence of an additional bias relative to the weight-adjustment approach. This potential bias term arises from a more extensive use of early cooperators as compared to the weight-adjustment procedure.
10. Intracluster correlations between units in the same PSU were relatively high for demographic variables (about 0.1) but were relatively low for health variables (.03 or less).

Although much information was obtained from the study, new issues were raised which must be considered for further RDD research. Some of the major questions for future research are:

1. How can cost information obtained from a small-scale feasibility study (such as the present one) be used to project the cost of a larger-scale production survey?
2. How do the data for health characteristics collected in RDD surveys compare with the corresponding data from the NHIS? What is the relative quality of the data?
3. What is an acceptable RDD response rate for NHIS? How can it be achieved?

4. What is the extent of errors in the telephone identification of RDD sample units other than special places? How can all these errors be reduced to acceptable levels?
5. How might special places be redefined for RDD surveys in order to reduce the error in their identification and at the same time be compatible with area/list surveys (for dual frame designs).

3. SAMPLE DESIGN

The telephone households in the sample for the Feasibility Study were selected using the RDD method described by Waksberg (1978). A brief description of how this method was used in this study is given below, followed by more specific details of the design, including the sample size and the use of substitution for nonresponse.

Using the most recent telephone exchange file from AT&T, a list of existing telephone area codes and working three-digit exchanges within each area code was created. To these six-digit combinations, all possible choices of the next two digits were added, forming a frame of the first eight digits of the ten digits in telephone numbers. The eight-digit numbers were the primary sampling units (PSUs). Each PSU contains 100 ten-digit numbers, identified by varying the last two digits. A random selection was made of an eight-digit number (a PSU) and of the last two digits. The number selected was dialed to determine whether or not it was residential. If the number did serve a residence, the PSU was labeled "residential" and was retained for the sample. If this number was not residential, the PSU was labeled "nonresidential" and was excluded from the sample. This procedure, referred to as primary screening, was repeated until a specific number, m, of residential PSUs was selected. For each PSU in the sample, additional last two digits were randomly selected and dialed until a specified number,

k, of residential telephones was identified. An attempt was made to obtain an interview with each of the k residences. The process of selecting and attempting to interview k residences in each PSU is referred to as secondary screening. The total sample size for this design is mk.

The sample for the Feasibility Study was selected in 12 nearly independent replicates.* One replicate was introduced each week for 12 consecutive weeks. Each replicate was interviewed for three weeks. Hence there was some overlap in the data collection phases of adjacent replicates. The total sample size for the study was about 3,000 telephone residences with a sample size per replicate of about 250. Based on the formula for the optimum cluster size given by Waksberg (1978), the optimum cluster size for NHIS was estimated to be 6. Also, it was decided to use the same PSUs for the half of the sample assigned to one questionnaire version as for the half assigned to the other version. Therefore, the total cluster size for each PSU was taken to be k=12 (six for each questionnaire version). With a cluster size of 12, this dictated that m=21 PSUs be selected per replicate to provide about 250 telephone residences. The precise replicate sample size was 252 (i.e., 21 x 12), which provided a total target sample size of 3024 (i.e., 12 x 252).

There were six lists of questions in the survey questionnaires pertaining to different health conditions. Each of the six sample households selected in a PSU for each questionnaire was assigned a different condition list. Since the selection of sample households was completely random, the six condition lists were assigned in a fixed order to the sample cases for each questionnaire (see figure 1).

---

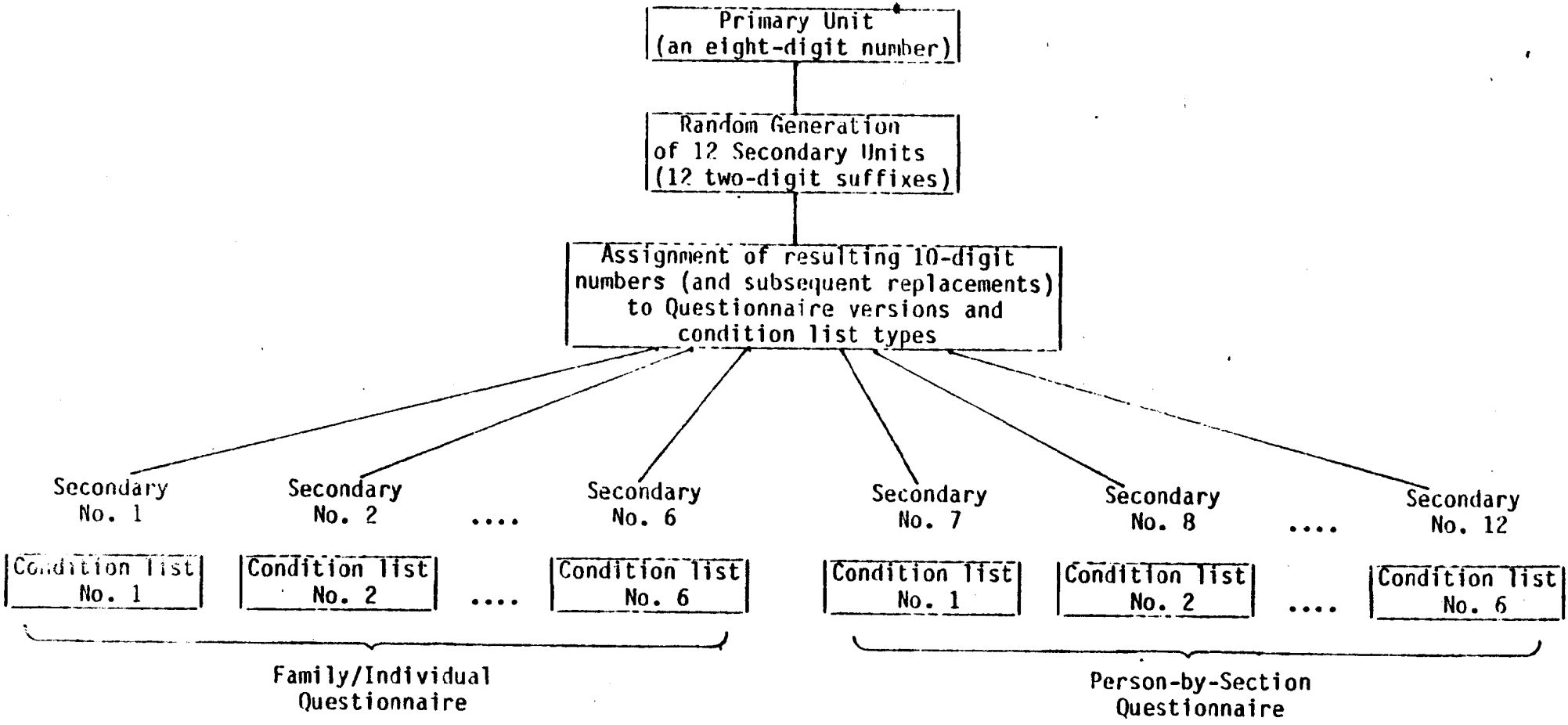
*The replicates were independent except that they were selected without replacement.

Prior to primary screening, the PSUs on the tape file were sorted on the basis of geography, population density, and proximity to urbanized areas. Tegels and Chapman (1984) provide details of this file sort procedure and of the other sampling procedures. For each replicate, a systematic random sample of 135 PSUs was selected from the sorted list and called in a random order until 21 residential PSUs were obtained for the sample. (Based on a residential hit rate of .24, there is better than a 99% chance of finding at least 21 residential PSUs among 135 randomly selected PSUs.)

For the selection within each retained PSU and interviewing of telephone households for the sample (secondary screening), the first step was to select 12 telephone numbers at random from the PSU--six for each questionnaire version. Each number was dialed to determine whether or not it was residential. If it was not a residential number, it was ineligible and was replaced by another number selected at random from the PSU. The new number was dialed and if it also turned out to be nonresidential, it was also replaced. Each number initially selected was replaced repeatedly until a residential number was obtained. In cases for which it was difficult to determine residential status, assistance was requested from the telephone business office. The assignment of selected telephone numbers to the two questionnaire versions and to the six condition lists is illustrated in Figure 1.

For each residential number identified during secondary screening, an attempt was made to obtain an interview. For each residence that refused or could not be interviewed for other reasons, it was decided to try substitution to account for nonresponse, as an alternative to weight adjustment. Specifically, a random substitute not previously selected was generated from the same PSU for any telephone residence that refused to participate or that

Figure 1. Random Assignment of Sampling Units to Questionnaire Versions for One Primary Unit



could not be reached for an interview after a specified number of calls.* The substitutes were called and approached for an interview in the same way as were the original sample cases. In particular, substitute numbers that turned out to be ineligible were replaced repeatedly until a residential number was selected. If a substitute residence refused to participate or could not be contacted, no additional substitute was generated for the original case.

All the interviewing for this study, at both the primary and secondary screening phases, was conducted by newly hired interviewers. Although extensive training was provided and interviewer performance improved substantially as the study progressed, the inexperience of these interviewers could have an important effect on the analyses results.

Additional information on the design and operation of the study, including sampling methods, survey forms, data collection procedures, the case management system, and data processing and editing is given in Appendix 11.

4. SUMMARIES OF ANALYSIS REPORTS

4.1 Response Rates

4.1.1 Background and Purpose

Perhaps the most important objective of the Feasibility Study was to determine the expected response rates for the telephone component of the NHIS. Therefore, response rates were computed for the entire sample, and by replicates to investigate the trend in response rates over the study period.

---

*The specific procedure used to initiate substitutes is described in Section 2 of Appendix 6.

Another objective of the study was to compare two versions of the NHIS questionnaire: Family/Individual (Form THIS-2X) and Person-by-Section (Form THIS-3X).* Thus, response rates were calculated separately for each questionnaire version. The response rates obtained from this study were also compared to those from other RDD health surveys.

4.1.2 Analysis

Since all of the sample cases were selected with equal probability, except for multiple-phone households, the response rates were computed basically as the (unweighted) proportion of eligible sample cases that responded. After completion of the data collection phase, there were a number of unresolved cases (i.e., cases whose eligibility was unknown).** Depending upon the manner in which these unresolved numbers are treated in the calculations, response rates can vary substantially.

Three different ways of treating unresolved cases were included in this analysis. The three methods of calculating response rates associated with these ways of treating the unresolved cases are the following:

(1) R_L = the lower bound on the response rate, calculated as though <u>all</u> unresolved cases were eligible.

(2) R_U = the upper bound on the response rate, calculated as though <u>none</u> of the unresolved cases were eligible.

(3) R_C = the estimator of the response rate calculated under the assumption that a proportion p of the unresolved cases are eligible. The proportion p is simply the observed eligibility rate among resolved cases.

---

*These forms are described in Section 4.3.

**An extended follow-up study was carried out on these unresolved cases to determine the reasons why they were not originally contacted. A report on the results of this investigation is being prepared by Anthony M. Roman of the Census Bureau and will be finalized in March 1985.

These three response rates were calculated for the entire sample and separately by questionnaire version and by replicate group (1-3, 4-6, 7-9, and 10-12). These calculations are given in Tables 1, 2, and 3.

Sixty-seven of the original target sample of 3,024 cases were dropped from the sample. The reasons for these deletions are given in Appendix 1. The 2,957 cases retained for the sample and used for computing response rates received the following final outcome classifications:

Outcome	Number of units
Complete interview	2251
Partial interview	42
Refusal	370
Other noninterview	36
Unresolved	258
Total	2957

As in the continuing NHIS, partial interviews were considered as a form of noninterview for the purpose of response rate computations.

4.1.3 Results and Conclusions

Including all replicates, the overall rates obtained were:*

$R_L = 76.12\%$ $R_C = 78.91\%$ $R_U = 83.40\%$

This indicates that response rates can vary by over 7 percentage points depending upon the manner in which unresolved numbers are treated. Also evident from Tables 1-3 is that no matter which type of rate is considered, a marked improvement in response rates occurred over time, with replicates 10 through 12 always exhibiting the highest rate for any 3-replicate grouping. Finally,

*Substitute interviews were not included as responses in these calculations.

although questionnaire form THIS-3X has slightly higher overall rates than form THIS-2X, these differences were not statistically significant at the .05 level.

The following table compares the values of R_C from the NHIS/RDD Feasibility Study to two other RDD health surveys. Care must be taken in the interpretation of these comparisons due to the differences in target populations, survey procedures, and questionnaire content. The most important differences of these types are noted in the table.

Survey	R_C
(1) NHIS/RDD Feasibility Study	.79
(2) National Telephone Health Interview Survey (THIS) (units for which busy signals were obtained were considered ineligible, rather than unresolved)	.82
(3) National Survey of Personal Health Practices and Health Consequences (NSPHPC)	.69

The response rate from the Feasibility Study compared favorably with the response rates for these other surveys. One additional study which should be used for comparative purposes is the experimental RDD/NHIS study conducted in 1979 by the Survey Research Center (SRC) of the University of Michigan. The SRC reported a response rate of approximately 79%. This rate was obtained using unlimited calls to households and with callbacks allowed a month or more after the initial interviewing period.

Based on the experience of the Feasibility Study, the following three conclusions have been reached:

(1) Response rates can be directly influenced by interviewer experience. The importance of experience can easily be seen by the progress that occurred between the early and later replicates of the Feasibility Study. Values of R_L, R_C, and R_U from replicates 10 through 12 were at least 12 percentage points higher than their corresponding

values from replicates 1 through 3. From replicates 10 through 12, the value of R_L was 83.13 percent, R_C was 84.92 percent, and R_U was 88.09 percent. Thus response rates of 85 percent or higher are within reason for the NHIS using RDD procedures provided that a well-trained and experienced staff of interviewers is maintained.

(2) Improved methods must be developed for quickly identifying ineligible PSUs. Three PSUs (or 36 potential interviews) were lost from the Feasibility Study because they were identified as ineligible too late to generate replacement PSUs. The problem appears related to identifying special places over the telephone (since certain ineligible PSUs contained only special places) and to identifying sparse PSUs (i.e., those with very few eligible residences). Work is proceeding at the Census Bureau in both of these areas.

(3) Even though the Feasibility Study used a 3-week interviewing period, 28 potential interviews were lost because cases were determined to be ineligible too late to generate a replacement. Improvements are needed in this area. Currently being considered are modifications to the automated call scheduler which should assist in contacting and identifying hard-to-reach units more quickly.

Table 1. Values of R_L (lower bound on response rate)

	Form THIS-2X	Form THIS-3X	COMBINED FORMS
Replicates 1-3	.6614	.7082	.6848
Replicates 4-6	.7636	.7108	.7371
Replicates 7-9	.7911	.7961	.7936
Replicates 10-12	.8221	.8404	.8313
Replicates 1-12	.7588	.7637	.7612

Table 2. Values of R_C (compromise response rate)

	Form THIS-2X	Form THIS-3X	COMBINED FORMS
Replicates 1-3	.6874	.7370	.7121
Replicates 4-6	.7929	.7463	.7697
Replicates 7-9	.8258	.8279	.8269
Replicates 10-12	.8425	.8558	.8492
Replicates 1-12	.7864	.7918	.7891

Table 3. Values of R_U (upper bound on response rate)

	Form THIS-2X	Form THIS-3X	COMBINED FORMS
Replicates 1-3	.7310	.7853	.7581
Replicates 4-6	.8413	.8067	.8242
Replicates 7-9	.8738	.8716	.8727
Replicates 10-12	.8790	.8827	.8809
Replicates 1-12	.8309	.8372	.8340

4.2 Breakoff Analysis

4.2.1 Background and Purpose

A breakoff is a discontinuation of an interview with an eligible respondent that occurs, for any reason, prior to the intended completion of the interview. In some cases the interview is completed at a later time. In other cases the interview is never completed.

One vehicle for increasing response rates is to learn as much as possible about interview breakoffs. Accordingly, an analysis of breakoffs in the survey was undertaken.

4.2.2 Analysis

The original aim of this analysis was to produce frequency counts and other descriptive analyses on the location of refusal breakoffs occurring during the Feasibility Study. However, this was not been possible due to problems in the

way the case management data were collected. It is not possible to distinguish between refusal breakoffs and other breakoffs; so the tabulations include information about all breakoff locations that were recorded for refusal cases.

Breakoffs occurring during each of the three distinct parts of the questionnaire (CATI screening, cover booklet, and insert booklet) were collected and stored separately on the case management system. Analysis of these data shows that there is a high level of missing data in the breakoff fields.

It is not possible to determine how many entries are actually missing because of the way the system was set up--some outcome codes required breakoff entries in every case, and some outcome codes required them in some instances but not in others. However, as a rough estimate, only about 21% of the breakoff entries that should have been supplied were actually provided. In light of the magnitude of the missing data, the reader must look at the results with a certain amount of caution.

4.2.3 Results and Conclusions

The results of the investigation are inconclusive. The two major factors contributing to this situation have been described previously: first, an inability to associate specific breakoff locations with the telephone calls on which the breakoffs occurred, and second, a very low response rate for the "location of breakoff" fields. It appears that these problems result from the way the data files were structured and/or programmed; thus, the major conclusion that can be drawn is that the case management system needs to be corrected in order to produce more useful data. Work is currently underway at the Census Bureau in this area [see Nicholls (1984)].

Nevertheless it is useful to present some of the basic results. The table below contains the frequency distribution of breakoff points recorded during refusal cases by the three major questionnaire sections. The number

of total breakoff locations is greater than the number of refusals received since there will generally be more than one breakoff for a refusal case. One breakoff location identifies the initial refusal, one is for the location of the final refusal, and there may be others associated with appointments that were made before the final refusal was received.

Questionnaire Section	Number	Percent
Total Breakoffs in Refusal Cases	2148	---
Location Unknown	1698	---
Location Known	450	100.00
CATI Screen	331	73.6
Cover Booklet	15	3.3
Insert Booklet	104	23.1

Additional analyses contained in Appendix 2 include frequency distributions of breakoff points within each questionnaire section and cross-tabulations of the breakoff points in each section by replicate group and questionnaire version.

In view of the magnitude of the nonresponse to the breakoff items, any other results are subject to a large margin of error. Nevertheless, if breakoffs for cases at unknown locations are assumed to be distributed in the same way as the observed breakoffs, the data suggest that breakoffs at the household roster do not appear to occur as frequently as was indicated by interviewer reports. Since the vast majority of observed breakoffs occurred during the CATI screening section before the household roster was reached, this does not necessarily imply that the household roster is not a problem,

only that other parts of the questionnaire present more serious problems in terms of incurring breakoffs.

4.3 Questionnaire Analysis

4.3.1 Background and Purpose

An objective of the Feasibility Study was to examine the differences between various estimates obtained from the two questionnaire versions used: Family/Individual version and Person-by-Section version. These questionnaires are given in Appendix 11 with a discussion of the differences.

The Family/Individual version closely resembled the questionnaire used in the personal visit NHIS. Based on results of a telephone study conducted by the University of Michigan's Survey Research Center (SRC) in 1979 using a modified NHIS questionnaire, it was hypothesized that the person-by-person style of the SRC questionnaire was responsible for producing higher than expected levels of reporting of certain health characteristics for telephone interviews in comparison to the family style version used in the 1979 NHIS.

The Feasibility Study used the two questionnaire types to test this hypothesis. The two questionnaire versions were very similar, with the main difference being the order in which the questions were asked in a few sections. In the Person-by-Section version, all questions in each section were asked about one family member before proceeding to the next family member. The Family/Individual version had breaks in these sections where interviewers returned to the most recent series of questions and asked these questions about the next family member. Another difference was that some questions in the Family/Individual version obtained individual information through an inquiry concerning the entire family, whereas in the Person-by-Section version, such information was obtained by questions asked directly about each individual family member.

4.3.2 Analysis

An attempt was made to identify whether significant differences existed between the two questionnaire versions. Results were also compared with those obtained in the SRC study and in the regular (personal interview) Health Interview Survey. The two main areas of analysis were the following:

1. Compare the demographic compositions of the interviewed portions of the various questionnaire-version samples, focusing on the following characteristics: sex, age, race, education, income, marital status, usual activity, and veteran status.
2. Compare the reporting of health characteristics for the various questionnaire-version samples. The following health characteristics were examined: two-week bed days, work-loss days, cut-down days, school-loss days, doctor visits, 12-month bed days and doctor visits, 13-month hospital stays, health status, total conditions, conditions by source, and limitations by type. The primary analysis consisted of comparing mean values. In addition, various overall distributional comparisons and comparisons by sex, age, and education were made for these health characteristics.

4.3.3 Results and Conclusions

If reporting a higher number of health events or occurrences is acknowledged to be indicative of better reporting, then consideration should be given to a questionnaire similar to the Person-by-Section version in any future NHIS conducted by telephone. All but one of the 15 significant differences detected for health events or occurrences indicated increased reporting on the Person-by-Section version, and most of the nonsignificant differences tended towards more reporting on the Person-by-Section version. Summary comparisons of the major demographic and health characteristics for the two questionnaire versions are shown in Tables 4 and 5. Detailed comparisons, which identify significant

differences between the health variables for the two versions, are provided in Appendix 3.

Respondent conditioning was perhaps a more important contributing factor with regard to the increased reporting than any methodological differences between the two questionnaire versions. Most of the questionnaire differences were very minor with the exception of the limitation of activities section, which occurred early in both questionnaires. The analysis tended to provide some evidence that the format change in the limitation of activities section may have conditioned respondents to expect similar patterns of questioning in later sections, which led to either increased reporting of the Person-by-Section version or to decreased reporting on the Family/Individual version.

Differences in the demographic compositions of the two interviewed questionnaire-version samples in the Feasibility Study seem to be too small (see Table 4) to explain any observed differences in health characteristics between the questionnaire versions. All sex, age, and education breakdowns generally tended to show higher reporting of health events or occurrences on the Person-by-Section version.

Although greater reporting of health events or occurrences, which might indicate less than excellent health status, was observed for the Person-by-Section version (see Table 5), the percentage of persons reported as being in "excellent" health was also greater on the Person-by-Section version. For three of the demographic breakdowns this percent was significantly greater. Since selfperceived health status is traditionally the most indicative single variable in the NHIS related to other health measures, some concern might be expressed regarding the accuracy of reporting in the Feasibility Study.

Table 4. Percentage Distribution for Demographic Characteristics of Persons Interviewed in the NHSI/RDD Study by Type of Questionnaire

		Type of Questionnaire	
	Characteristic	Family/Individual	Person by Section
Sex	Male	47.4	47.8
	Female	51.5	51.3
	DK/NA/Ref	1.1	1.0
Age	17-24	17.0	15.6
	25-34	23.9	24.3
	35-44	18.6	17.8
	45-55	14.2	13.4
	55-64	11.2	12.7
	65-74	9.2	9.8
	75+	4.7	5.3
	DK/NA/Ref	1.2	1.1
Race	White	88.4	87.6
	Nonwhite	11.6	12.4

Table 5. Percent of Persons Reporting Health Events and Mean Levels of Health Events Reported in the NHIS/RDD Study by Type of Questionnaire

	Percent Reporting One or More Events		Mean Number of Events Reported	
Health Event	Family/ Individual	Person by Section	Family/ Individual	Person by Section
2-wk. bed days	7.5	8.3	0.26	0.29
2-wk. work-loss days	7.4	7.5	0.27	0.22
2-wk. cut-down days	5.3	7.0	0.24	0.36
2-wk. doctor visits	15.5	15.6	0.24	0.26
13-mo. hospital stays	11.8	11.8	0.15	0.15
12-mo. doctor visits	73.4	73.5	3.19	3.42
12-mo. bed days	49.8	52.6	4.32	4.57
2-wk. school loss days	14.2	15.2	0.32	0.42

Except for a few small differences in education, income, and usual activity, the demographic makeup of the feasibility samples resembled that observed for the other health surveys. The comparisons for overall reporting of health characteristics were even more similar. Some differences in health characteristics by sex, age, and education were observed between the other health surveys and each of the Feasibility Study questionnaires, but these differences were most likely due to the small size of the study samples.

4.4 Respondent Rules

4.4.1 Background and Purpose

The NCHS/Census Joint Committee on Telephone Surveys Task Group (1983) devoted a considerable amount of attention to the selection of a respondent rule for the Telephone NHIS. A number of respondent rules were analyzed with respect to cost, sampling error, and nonsampling error. The Task Force recommended that a most knowledgeable respondent (MKR) rule be used for the Feasibility Study. An additional factor considered in the development of the MKR rule used in the Feasibility Study was to select a rule that would most closely approximate the respondent rule used in the face-to-face NHIS. Under the rule developed, the interviewer asked the telephone answerer to identify the MKR for the household and attempted to conduct a household interview with the MKR. The screening questions and procedures used to select and interview the MKR are shown in Appendix 4. Midway through the study the MKR screening questions were revised by shortening the introductions to the questions. The same selection and interviewing procedures were used throughout the study. One of the main objectives of the Feasibility Study was to evaluate the feasibility and effect on response of the MKR rule.

4.4.2 Analysis

The analysis plan developed for the MKR rule is included in Appendix 4. The major analytical issues to be addressed with respect to the MKR rule were:

(1) How well can the telephone answerer identify the MKR?

(2) How often is the MKR home at the time of the first household contact?

(3) How many callbacks are required to reach and interview the MKR when the MKR is not at home on the first household contact?

(4) What is the effect of the MKR rule on the response rate?

(5) What is the quality of information provided by the MKR for other members of the household?

(6) What are the demographic characteristics of the MKRs?

Only part of the analysis for the MKR rule has been completed. NCHS plans to complete the analysis in 1985. The ability of the telephone answerer to identify the MKR was measured by the proportion of time one household member was identified as the MKR, two or more household members were identified as equally knowledgeable, and the proportion of times no one was identified as the MKR. This information will be correlated with the exit information collected about the MKR at the end of the household interview. The exit questions on the MKR are given in Appendix 4.

The effort required to reach and interview the MKR has been partially analyzed. It was hypothesized that the MKR would often be the phone answerer or someone else at home at the time of the initial contact, since the adult most likely to be at home when most of the calls are made is also most likely to have the major responsibility for the health care of any children in the household. The effect of the MKR rule on response rates was measured by the number of times no interview was obtained after repeated callbacks, refusals by the MKR, and refusals by the phone answerer to ask the MKR to come to the phone.

The quality of the information provided by the MKR was partially evaluated using the results of a set of questions asked at the end of the content interview. One of the questions asked the household respondent if he or she still feels he or she is the MKR and if not, who is. Another set of the post survey questions asked the respondent to indicate the accuracy of the information given about each member of the household as follows: very accurate, fairly accurate, or not very accurate. The Telephone Research Task Force also recommended that a number of reinterviews be conducted for proxy respondents to evaluate the quality of the information provided by the MKR for other members of the household. This component of the study was dropped due to its cost and complexity.

4.4.3 Results and Conclusions

The first set of tables shown in Appendix 4 was taken from a response rate progress report prepared by Anthony N. Roman on June 20, 1984. His analysis indicates that approximately 97 percent of the completed interviews were provided by the persons identified by the telephone answerer as the MKR. From these results it appears that there was little difficulty in identifying the MKR and in most cases the MKR was interviewed. From Table 2, however, it appears that when additional callbacks were required to interview the MKR, the refusal and other noninterview rate was much higher than the refusal and noninterview rate for the initial household contact. In fact, when callbacks were required to reach the MKR, there were more about 5 times as many refusals than completed interviews. A different phenomenon occurred when the MKR was not the phone answerer, but was at home at the time of the initial contact. In these cases 62 interviews were completed as compared to only 2 refusals. From the preliminary results one would conclude that the MKR rule works quite well in cases for which the MKR is home at the time of the initial household contact; in cases for which the MKR was not at home at

the time of the initial call, callbacks were not effective in reaching the MKR for an interview. However, a more detailed analysis is needed in order to confirm this conclusion.

An analysis of the exit interview questions about the MKR was done by William Mockovak and is included in Appendix 4. The results of this analysis show that approximately 9 percent of the respondents indicated that they were not the MKR. Of these respondents only slightly more than half identified another member of the household as more knowledgeable (over one-third of the responses to this followup question (3a) were missing). An analysis has not been done to determine how many of the respondents who said they were not the MKR at the end of the interview indicated they were the MKR in the beginning of the interview. Of the respondents who said they were the MKR at the end of the interview, approximately two-thirds said there were other household members equally knowledgeable about the health of other family members.

Further analysis is required to fully address each of the issues listed in Section 3.4.2. The results of the assessed accuracy of information given by the respondent will be evaluated by correlating the respondent assessment with the level of reporting health events. Respondent assessments of information provided for self and proxies will also be studied. Finally, the characteristics of the MKR phone answerers will be contrasted to the characteristics of the MKRs called to the phone and to the respondents who were identified as not being the MKR.

4.5 Interview Period/Sampling Frequency

4.5.1 Background and Purpose

There are many issues that could be studied regarding the optimal length of the interview period and the optimal frequency of introducing replicates (or panels) for an ongoing sample survey. After examining much data it was

decided to address three questions which appear to be most important and interesting and which could be investigated with the data available:

1. How many cases were unresolved after three weeks of interview? Is there any evidence that the distribution of calls over the interview period affects the number of unresolved cases?
2. Were interviewer workloads evenly distributed throughout the survey? That is, was the level of work constant in the facility across weeks?
3. How would response rates be affected if the interview period had been two weeks or four weeks instead of three weeks? Could response rates be improved by stopping the generation of replacements during the last week of the survey?

These issues are discussed in detail in Appendix 5.

4.5.2 Analysis

The analysis for each of these questions consisted of an examination of graphs of data from the case management system. For question 1, the average number of calls made per week to each case by replicate was calculated. An attempt was then made to relate this to the number of unresolved cases in each replicate. In answering the second question, numbers of cases called, contacted and completed by week of the survey were examined. In order to answer the third question, the number of cases completed and the average number of calls made to each case by day of the replicate were enumerated. Also, the generation of replacement cases and how many of them were subsequently resolved were examined.

Some parts of the analysis discuss reps 1-6 and 7-12 separately. This was necessary because survey procedures differed between these two groups. Replicates 1-6 scheduled calls to cases using both the automated call

scheduler and hand scheduling by supervisors. Reps 7-12 relied exclusively on the call scheduler. Therefore, some differences may be expected between these two parts. However, this also enabled us to make some observations about the effectiveness of the automated call scheduler.

The analyses were done completely without the benefit of sampling error estimates. Thus, inferences made about the observed differences in calling patterns, unresolved cases, etc. are limited. However, these preliminary analyses identify areas where fuller investigation is needed using more sophisticated statistical methods.

4.5.3 Results and Conclusions

Using the data examined in this analysis, the following conclusions about the questions posed in the introduction are made:

1. Among replicates the number of unresolved cases (out of the approximately 250 sample cases) varied from eight up to 32. It is difficult to discern which calling patterns were most successful at reducing the number of unresolved cases. However, the data do indicate that replicates whose cases received relatively more attempts in the later weeks than the first week of the interview period had lower numbers of unresolved cases. It appears that this pattern was achieved in this study by using the automated call scheduler and a constant facility staff level.
2. The interviewer workloads were fairly well distributed during the survey. The number of calls, contacts and completions were all stable during the survey period.
3. Response rates could be increased (about 2-4%) if the survey period was increased to four or more weeks. However, the resulting cost increases may not be worthwhile. Further investigation is needed

here. Also, the cutting off of replacements during the last week of the survey would have only a negligible effect on response rates.

The data files from the Feasibility Study are a rich source of information about random digit dialing telephone surveys. The issues that were examined are just a few of the many issues which could have been examined. For example, in answering question 1, tallies were made of the number of calls by week of the survey. This question could also be approached by looking at the number of calls by time of day. Thus, while we believe the conclusions drawn here are valid, there are also data available to support more research.

4.6 Substitution

4.6.1 Background and Purpose

An objective of the Feasibility Study was to conduct a preliminary development of nonresponse adjustment procedures. The procedure that is probably used most often in surveys to account for unit nonresponse is weight adjustment. With this procedure, the entire sample of eligible cases is partitioned into weight adjustment cells. Within each cell, the weights of the respondents are adjusted (upward) so that their sum equals the sum of the unadjusted weights of all sample cases in the cell. Essentially, the characteristics of the respondents in each cell are imputed to the nonrespondents.

Since there is considerable control over the sampling operation with a centralized telephone system, serious consideration has been given to the use of substitution to account for nonresponse in RDD surveys. Consequently, a substitution procedure was used in the Feasibility Study. Whenever a residential unit refused or could not be reached after a specific number of calls, a residential unit not previously selected was drawn at random from the same

PSU to serve as a substitute. An evaluation of the substitution procedure used, including a comparison of substitution and weight adjustment, was carried out as part of the analysis of the Feasibility Study data.

4.6.2 Analysis

The substitution analysis was restricted to ten of the 12 replicates selected because of an error made in the generation of substitutes during replicates 6 and 7. For these ten replicates, substitutes were generated for 668 cases. Four specific analyses were carried out:

(1) Evaluation of the general effectiveness of the substitution procedure

For this analysis, the percentage of original cases which were interviewed after generating a substitute was compared to the percentage of substitutes interviewed. This provided an indication of whether or not substitutes were being generated too early or too late. Also, a comparison of the response rates for substitutes and original sample cases was made.

(2) Costs for substitutes

Although the exact costs attributable to substitution could not be computed from the data available in this study, several cost-related averages in terms of the time and effort associated with substitutes were computed on a PSU basis.

(3) Comparison of substitutes and initial selections

For the 150 cases for which responses were obtained from both the original and substitute residences, comparisons of the responses from originals and their substitutes were made for eight demographic and five health characteristics. These characteristics are listed in Appendix 6.

(4) Variance comparisons for estimates based on substitution and weight adjustment

For the five health variables the variance estimates of the substitution-based estimates of means were compared with the corresponding variance estimates of the weight-adjustment-based estimates. These comparisons were made for substitution and weight-adjustment estimates that are approximately equal in cost.

4.6.3 Results and Conclusions

There were 668 nonresponse cases for which substitutes were supposed to have been generated, based on the substitution procedures used in this study. Interviews were eventually completed with 216 (32.3%) of these original sample cases. The generation and pursuit of substitutes provided contacts with 543 substitute cases (81.3%) and completed interviews with 435 substitutes. Regarding the general effectiveness of the substitution procedure, these results are inconclusive. The fact that nearly one-third of the cases targeted for substitution were eventually interviewed suggests that perhaps substitutes were generated too early. However, since substitutes were not even contacted for about 19% of the targeted cases, it would not appear advisable to begin generating substitutes any later.

Substitutes were actually generated for only 618 cases because of time constraints. For these 618 cases, the response rate was about 74%. The response rate for the initial sample was about 5 percentage points higher. This higher response rate was apparently due to the fact that less time is generally available to reach and interview a substitute case than there is for an original sample case.

The detailed calculations of the cost-related items listed in the previous section are given in Appendix 6. An important conclusion based on these calculations is that the time and effort expended on pursuing and interviewing substitutes could have been used to increase the PSU sample size by about three units if substitution had not been used. This result was important for the comparison of variance estimates for the weight-adjustment-based estimator with those for the substitution-based estimator.

Regarding the comparative analysis of the 150 late responding original cases and their substitutes, a significant difference between means was obtained at the 10% level of significance for only two variables: age of reference person and average age of all household members. For both of these variables, the average was higher for the substitutes than for the original cases. This is not surprising since it could be anticipated that the difficult-to-reach initial cases would have more mobile young-adult households. For the distributional comparisons made between the initial and substitute samples for four demographic variables--sex, race, education, and marital status of reference person--the only significant difference found was for sex of reference person. There was a significantly higher proportion of female reference persons among the substitutes than among the original sample cases. Although no differences between means for the five health variables were significant, it is interesting to note that for all five comparisons, the average number of illness-related characteristics was higher for the substitutes than for the initial sample cases.

To summarize these comparisons, the reference persons in the substitute households were generally older, had a higher percent female, and tended to report higher numbers of illness-related activities than did their hard-to-interview counterparts. These differences arise because the substitutes

often must be "early cooperators" due to the time constraint. Consequently, the use of substitutes in the Feasibility Study may introduce a bias in the estimates that would not exist for the weight-adjustment-based estimates because of the tendency for substitutes to be early cooperators.

For the variance comparison of the substitution-based and weight-adjustment-based estimators, variance estimates were computed for both estimators of the mean for each of the five health variables. In each case, the variance estimates were based on approximately equal-cost samples for the two methods of accounting for nonresponse. For each of the five health variables, the variance estimate for the substitution-based estimator was less than that for the weight-adjustment-based estimator. Therefore, from a variance standpoint it appears that substitution is a better method of accounting for unit nonresponse than a PSU-by-PSU weight adjustment procedure.

4.7 Special Place Study

4.7.1 Background and Purpose

The RDD surveys conducted to date by the Census Bureau, and by most other organizations, have not attempted to address the problems associated with the telephone enumeration of special places. These are places, such as college dormitory housing or retirement homes, that are different from the usual types of living quarters and where the occupants usually share some common facilities. Special places are believed to house about three percent of the nation's population and omitting them from telephone surveys may result in a coverage bias. An objective of the Feasibility Study was to investigate ways to identify special places over the telephone as a possible prelude to developing procedures for enumerating the occupants of such places.

This special place research was preliminary in nature and its objectives were fairly modest: (1) to determine how well telephone enumerators could

differentiate special place telephone numbers from other types of telephone numbers, and (2) to obtain some empirical evidence as to the feasibility of compiling over the telephone a list of living quarters within special places. The study did not address any issues related to data quality nor was there any attempt to enumerate any persons residing in special places.

Starting with Replicate 2, each replicate was seeded with known special places drawn from two sources: (1) the current survey frame--these were special places that had recently rotated out of a Census Bureau face-to-face survey, and (2) the telephone directory frame--these were special places selected at random from available telephone directories.

Special places identified by the telephone enumerators were referred to the shift supervisor. No interviews were to be conducted at any special places (seeded or unseeded). If an identified special place was part of the current survey frame, the supervisor recontacted it and attempted to compile a list of the living quarters within the place. This listing was then compared to the listing made by the face-to-face enumerator for coverage evaluation. No recontact was made to special places drawn from the telephone directory frame, since these places were used solely to evaluate the ability to identify special places over the telephone. Also, no further contact was made with unseeded special places.

4.7.2 Analysis

The special place research was designed primarily to measure the ability of telephone enumerators to successfully differentiate special place telephone numbers from other types of residential or nonresidential places. The following pertinent measures were obtained:

1. Special place identification success rate, by type of place
2. Special place identification success rate, by replicate group

3. Special place identification success rate, by enumerator

4. Misclassification rate, by type of misclassification

As a secondary consideration, the study produced some rough indications of within-place coverage by comparing the listing of units made in the field to the listing made over the telephone. These data are presented by type of special place.

4.7.3 Results and Conclusions

The RDD interviewers successfully identified only about 39 percent of the seeded special places during replicates 2-12. However, the success rate over the final five replicates improved to about 56 percent. This increased success rate over the final five replicates is attributed to a modification of the screening questions coupled with an intensive refresher training session on special place identification procedures after replicate 7. It is clear, however, that even the 56 percent success rate attained over the final five replicates is not very acceptable for survey work. In 32 of the 100 special places seeded into these five replicates, the enumerators misclassified the telephone numbers as nonresidential. In these places, no telephone interview would have been obtained, with a potential coverage loss to the survey.

From the information obtained in this study, it appears that special places that identify themselves as places of business immediately upon answering the telephone are more often correctly identified as special places. Hotels and motels had the highest identification success rate, while trailer parks were never correctly identified as special places.

In the special places seeded from the current survey frame, the listings made over the telephone were identical to the face-to-face listings in 11 of 15 cases. (Three of the places where the listings differed had undergone some changes since the face-to-face listing was compiled.) While the sample

is much too small to make any generalizations, it appears that telephone coverage of living quarters within special places (once they have been identified) is comparable to face-to-face coverage of units.

This special place research provided some limited empirical evidence on the ability of telephone enumerators to identify special places and to compile a listing of the living quarters within special places. The results suggest that the identification success rate is somewhat lower than desired and that innovative procedures may be needed to improve it. On the other hand, it appears that once enumerators identify a special place, it is possible to compile over the telephone a complete sampling frame of the units within the place. More research into the operational problems associated with telephone enumeration of special places should be given high priority in future RDD surveys.

4.8 Cost Analysis

4.8.1 Background and Purpose

The primary purpose of the Cost Analysis was to estimate the operational cost of conducting the NHIS by phone using an RDD sampling frame. Because of varying salary scales, overheads and other cost allocation methods, dollar amounts could be very misleading. Therefore, the cost-related information is expressed in terms of time components. The data provided in the report were abstracted from both the payroll file and the case management file, with the latter accounting for the "on-line" activities of the interviewing staff. The primary focus of the analysis was the time per case associated with the survey.

4.8.2 Analysis

Before one could interview a sample case in secondary screening, its eligibility status (i.e., residential/nonresidential) had to first be determined. However, during the three-week interviewing period, it was not possible to determine the eligibility of all of the sampled numbers. It was therefore decided to group the cases into three overlapping categories to produce cost-per-case estimates. The first group consisted of the interviewed households. The second group consisted of all potentially eligible cases. This group contained the verified eligible cases as well as those cases whose eligibility status had not yet been determined at the end of the interviewing period. The third group consisted of all sampled telephone numbers. The counts of units in each of these three groups were used as the denominators in the per case averages given in the table in the following section. Also, the average "on-line" time to complete a case for each of 16 final outcome code categories was computed for each of the twelve interviewing replicates, as well as for the entire survey. These times were further partitioned into four time components related to the following interviewing activities:

1) access of case to dialing time
2) screening time
3) interviewing time
4) transcription time

4.8.3 Results and Conclusions

The table below illustrates the operational minutes per case as well as the on-line time per case for each of the three groupings of outcomes.

COMPARISON OF CASE MANAGEMENT DATA TO PAYROLL DATA

CASES	(1) FACILITY ON-LINE MIN/CASE	(2) TOTAL PAYROLL MIN/CASE	(3) SECONDARY SCREENING PAYROLL MIN/CASE
INTERVIEWED HOUSEHOLDS	61	215	104
POTENTIALLY ELIGIBLE HOUSEHOLDS	47	165	80
ALL PHONE NUMBERS	26	90	44

Tables providing minutes per case by outcome code and by replicate can be found in Appendix 8.

One must proceed with caution when applying the data contained in this report to budget estimates for other RDD surveys. The following characteristics of this survey should be taken into account.

1. Interviewers administered two different versions of the questionnaire.
2. Although an automated case management system was utilized, the interview was conducted from and responses recorded on a paper document. Therefore, the interviewer had to switch back and forth between the paper questionnaire and the terminal. This also required a clerical control of partially completed documents so that the appropriate document would be available to the interviewer following up on a case.
3. The skip patterns of the interview had to be applied by the interviewers.
4. There was a manual edit performed on the document instead of an automated one.

5. The total survey period was short and thus the interviewers were still in a learning process when the survey ended. This is evidenced by the fact that the average number of minutes spent per interviewed case dropped from 66 minutes on the first three interviewing replicates to 52 minutes per case by the final three replicates. Also, response rates increased considerably from the beginning to the end of the survey.
6. Substitute phone numbers were introduced into the sample for apparent nonresponse cases. (The procedure used to generate substitutes is described in Section 2 of Appendix 6.)

4.9 Monitoring

4.9.1 Background and Purpose

As part of the Feasibility Study, professionals from both the Census Bureau and NCHS monitored a sample of live interviews to address a variety of questions that would be difficult to answer using objective (response rates, item nonresponse, cost, production, etc.) survey data. A list of these questions follows. Most of these questions reflect concerns about changing the NHIS from a personal, face-to-face interview to a telephone interview.

1. Did the interviewer have difficulty identifying and obtaining an interview with the most knowledgeable respondent?
2. Which sections of the questionnaire, or individual items, were most troublesome to the interviewer, to the respondent, to both? Further, did problems vary by the version (THIS-2X or THIS-3X) of the questionnaire being tested?
3. Did the absence of flashcards cause problems?
4. Was respondent fatigue or frustration a problem?

5. How cooperative was the respondent?
6. How adequate was interviewer performance with respect to knowledge of the questionnaire, probing, answering questions from respondents, and following skip patterns?
7. How did interviewer performance vary during the course of the study?

To structure the monitoring, a special monitoring form was designed that addressed the preceding questions. This form was to be completed for each interview that was monitored.

4.9.2 Analysis

At the conclusion of this study, 151 monitoring forms were available for analysis. However, preliminary analyses and discussions with monitors indicated that conclusions based on the monitoring data would be misleading, rather than informative. Accordingly, although lessons were learned, they were judgmental in nature.

4.9.3 Results and Conclusions

Most of the stated objectives of professional monitoring could not be addressed using the available data. A partial list of the major problems affecting the monitoring data follows.

1. Monitors varied widely in both their knowledge of the Health Interview Survey content and interviewing skills, in general. Therefore, anchor points on the structured rating scales were differentially defined. For example, to one monitor a "cooperative" respondent might have meant someone who completed the interview, but to another monitor, it might have meant someone who was merely nice, even if (s)he refused to be interviewed.

2. Persons varying widely in background knowledge monitored most heavily at different points in the survey. For example, persons most knowledgeable about the content of the survey tended to monitor more heavily in the first half of the survey than in the second half.
3. Persons monitored with widely different objectives. Although a structured monitoring form was used, individuals focused on different aspects of the survey. For example, some monitors were primarily concerned with refusals and why they occurred, some focused on voice quality and style, and others focused on the content of the survey. The result was that relatively few monitoring forms were filled completely.
4. Monitors felt that their standards for judging interviewer performance changed during the course of the study. Initially, some monitors reported that they compared the performance of the telephone interviewers with that of field interviewers. However, this standard was changed when the monitors realized that the telephone interviewers were not in the same "ball park," at least during the first half of the survey. Therefore, raters shifted to comparing telephone interviewers with each other, rather than with field interviewers.

In addition to the problems with monitoring done by professionals, there were also major problems with the quality control (QC) monitoring done by supervisors in the facility. Throughout the 16-17 weeks of interviewing (including practice interviewing of "live" cases), supervisors completed only eight monitoring forms. The limited amount of QC monitoring that occurred can be attributed to a lack of instruction in how to use the monitoring form, lack of supervisor input into the content of the monitoring form, lack of a

sampling plan for conducting monitoring of interviewers, and competing supervisory responsibilities that were viewed as higher priority than monitoring.

In retrospect, the following conclusions were reached.

1. For QC monitoring to be successful, supervisors must view monitoring as an important task. Supervisors must be trained on how to use the monitoring form and how to give feedback to interviewers. Also, the supervisors must be held accountable for completing monitoring sessions. However, requirements to monitor must also be realistic in light of other supervisory tasks.
2. The research objectives of professionals should determine the type of monitoring form used. This project attempted to satisfy a variety of research questions with the same monitoring form, and the results were unproductive. Further, more complex research objectives impose technical constraints that must be addressed to avoid biased results (e.g., acceptable form reliability/validity, interrater reliability, invariant rating standards, and representative sampling plans.)
3. Professionals should be trained on how to monitor, even if an apparently simple question is being researched.
4. Interviewer performance improved during the survey period as indicated by the overall response rate and item response rates in each third of the sample.

4.10 Intracluster Correlations

4.10.1 Background and Purpose

The purpose of the intracluster correlations study was to calculate the intracluster correlations between units in the same PSU for several demographic and health variables and to derive the optimum cluster (PSU)

size for future health surveys based on these correlations and on estimated values of the cost and other parameters needed for the derivations. The formulas used for calculating the correlations and the optimum cluster size are given in Stokes (1983). The general cost model that was presented in Stokes was used but the cost formulas for productive cases and for unproductive cases were slightly modified.

4.10.2 Analysis

Intracluster correlations among households were calculated for 15 different variables: 12 health variables and 3 demographic variables. For each of these variables correlations were derived for each of the 12 replicates as well as for the entire sample. The intracluster correlations for the entire sample for the 12 health variables were all .030 or less. The correlations for the three demographic variables--household income, education level of the respondent, and age of the respondent--were .095, .110, and .153, respectively.

To calculate the optimum cluster size, an estimate of the ratio of the cost of a productive case to the cost of an unproductive case was needed. This estimate was based on the ratio of average number of minutes for productive and unproductive cases obtained from data collected in this study. (A refusal is included as a productive case since it generally provides resolution of an eligible case.) The average time required for productive cases was 70.05 minutes (42.15 minutes on-line and 27.90 minutes off-line). For unproductive cases the average time was 35.11 minutes (9.40 minutes on-line and 25.71 minutes off-line). The optimum cluster sizes which were derived using a cost of productive vs. unproductive cases ratio of 2.00 are given in the table below for intracluster correlations covering the range of the correlations found in the NHIS-RDD study.

Optimum cluster sizes ($\bar{k}^*$)

ρ	$\bar{k}^*$
.150	4
.100	4
.050	6
.030	8
.020	10
.010	14
.005	19
.001	42

4.10.3 Conclusions

From this analysis, it appears that a cluster size of 4 residential telephone numbers is optimum for measuring demographic variables in the NHIS. For estimating the number of activity limitations of household members, total number of household work-loss days, and total number of cut-down days for the household, the optimum cluster size is at least 20 residences. For the other health variables, optimum cluster sizes range from 8 to 13 residential telephone numbers.

It is important to realize that these conclusions on optimum cluster size depend on the productive/unproductive cost ratio of 2.00 that was calculated from the data obtained in this study. In a full-scale NHIS/RDD, the cost ratio and the optimum cluster size would probably differ from those derived here.

REFERENCES

Biemer, Paul (1983). "Final Design of the 1984 NHIS-RDD Feasibility Study," attached to Development Memorandum No. 4, September.

NCHS/Census Joint Committee on Telephone Surveys Task Force (1983). "Research and Development Recommendations Leading to a Dual Frame National Health Interview Survey," an NCHS/Census report to the Joint Agency Steering Committee on CATI Software, March.

Nicholls, William (1984). "Recording the Last Item Answered and Which Interviewer Asked Which Item," internal U.S. Bureau of the Census Memorandum for the Bridge Subcommittee, July.

Stokes, S. Lynne (1983). "Optimization of Waksberg's RDD Sample Design," internal U.S. Bureau of the Census report, May.

Tegels, Robert and Chapman, David W. (1984). "Sampling Specifications for the National Health Interview Survey-Random Digit Dialing (NHIS-RDD) Feasibility Study," internal U.S. Bureau of the Census memorandum, January.

Waksberg, Joseph (1978). "Sampling Methods for Random Digit Dialing," JASA, 73, pp.40-46, March.

Appendix 1. Response Rates

Response rates from random digit dialing (RDD) surveys can vary substantially, depending largely upon the manner in which the researcher handles uncontacted numbers. As an example, in the recently completed NHIS-RDD Feasibility Study, one could determine that the household interview rate was anywhere from 76 percent to 84 percent. This report discusses results from the NHIS-RDD Feasibility Study and methods one can use in estimating overall interview rates.

Introduction

In a 1984 study, the Census Bureau, in cooperation with the National Center for Health Statistics (NCHS), tested the feasibility of conducting the National Health Interview Survey (NHIS) over the telephone using RDD sampling techniques. The design of the study called for 3,024 telephone households to be assigned for interview between February and May 1984. The sample was selected in 12 replicates of equal size (252 households). The Waksberg procedure [1] was used to select the replicates. Each replicate consisted of 21 primary sampling units (PSUs). A PSU was a block of 100 telephone numbers defined by all ten-digit numbers associated with a fixed first eight digits. The 21 PSUs were selected with probabilities proportional to the number of residential phone numbers within the PSU. Twelve (12) telephone numbers were then randomly selected from each PSU. If upon being contacted, a unit was found ineligible for this study, a replacement number was generated from within the PSU. In this manner, an attempt was made to select 12 eligible units from within each PSU. Ineligible units for this study included nonworking telephone numbers, businesses, and special places (e.g., college dormitories, nursing homes, etc.). Interviewing took place over a 3-week period within each replicate with a new replicate being introduced into the study each week. Therefore, interviewing periods of replicates overlapped. Finally, substitutes were generated in this study for potential noninterview cases, but will be excluded from all calculations in this report. An investigationn of substitution as a method of weighting adjustment will be conducted at a later time. More details about the sampling design can be found in [2].

Although the primary goal of the NHIS-RDD Feasibility Study was to determine if the NHIS could be successfully administered over the telephone, a secondary goal was to compare two separate questionnaire forms. These forms were very similar in design, differing only slightly in certain procedures. Form THIS-3X (Person by Section) asked all questions within certain sections of the questionnaire of one person, then asked the same questions of a second person, and proceeded in this manner until the sections were completed for all family members. Form THIS-2X (Family-Individual) broke these sections into segments, asking the questions from one segment of the first person, then proceeding to additional persons before beginning with the next segment. Since these procedures affected only a few sections of the questionnaires, the net difference in questionnaire forms was minimal. An extensive comparison of questionnaire forms is reported in [3].

This report examines issues relating to the response rates obtained in this study. Information concerning the quality of these data (i.e., item nonresponse, etc.), the costs of collecting these data, and additional aspects of the study are discussed in other reports.

Notation

The following notation will be used throughout this report:

C = Number of completed interviews (partial interviews are excluded)

E = Number of units determined upon contact to be eligible

U = Number of units whose outcome status is unresolved (e.g., units where only ring-no-answer call outcomes were obtained, units where busy signals were obtained, etc.)

I = Number of units determined upon contact to be ineligible

The response rates for this study (R_L) were initially calculated in the following manner:

$$R_L = \frac{C}{E+U}$$

This response rate is conservative in that it assumes that all unresolved units are eligible for the study. In fact, it is quite likely that some portion of the unresolved units are ineligible for the survey. Due to this, R_L can be considered as a lower bound on the true response rate obtainable if the eligibility status of all sample units could be determined.

Results

Although the sample design called for 3,024 eligible units to be selected, only 2,957 were used in computing response rates. A total of 36 units were lost when one PSU in replicate 6 and two PSUs in replicate 8 were discovered to be ineligible. These discoveries occurred too late to generate replacement PSUs. An additional 3 units were lost when one PSU in replicate 7 contained only 9 eligible units within its 100 numbers. Finally, 28 units were lost when they were found to be ineligible too late to generate replacements.

The 2,957 units used in computing response rates received the following final outcomes:

Outcome	Number of Units
Complete Interview	2251
Partial Interview	42
Refusal	370
Other Noninterview	36
Unresolved	258

As in the continuing NHIS, partial interviews were considered a form of noninterview in this study. Of the 36 units which received "Other Noninterview" outcomes, 35 were described as language barriers which could not be converted while the remaining unit was not described. This indicates that encountering foreign language households may be a problem in telephone interviewing. Of the 258 unresolved units, 169 had been dialed between 1 and 19 times and were unresolved when the replicate closed out. The other 89 unresolved units were dialed the maximum of 20 times during the replicate and then declared unresolved.

Values of R_L are displayed in Table 1. An overall response rate of 76.12 percent was obtained from the study. Questionnaire form THIS-3X had a slightly higher response rate than form THIS-2X (76.37 percent vs. 75.88 percent), but this difference was not statistically significant at the .05 level. There was an improvement in response rates across time as evidenced by Table 1, which shows a rate of 68.48 percent from replicates 1 through 3 that increases to a rate of 83.13 percent from replicates 10 through 12. The individual rates from each replicate as well as the cumulative rates through each replicate are shown in graphs attached to this report.

An important question in determining response rates for surveys conducted using random digit dialing is how to classify unresolved units. One could consider all unresolved units to be noninterviews (e.g., R_L). If one assumes that all unresolved units were ineligible for the survey (an unlikely event), then an upper bound on the true response rate, R_U, could be computed as:

$$R_U = \frac{C}{E}$$

The values of R_U are displayed in Table 3. These rates are approximately 7 percentage points higher than the corresponding values of R_L.

Another approach is to allocate a proportion of the unresolved units, p, into the eligible unit category. This approach was suggested by Frankel, et al [4], in a 1982 special report to the Council of American Survey Research Organizations (CASRO). Response rates (R_C) are computed in the following manner:

$$R_C = \frac{C}{E+pU}$$

where:

$$p = \frac{E}{E+I}$$

The proportion, p, is computed from the sample using only those units whose eligibility status has been determined. Based upon the sample from the NHIS-RDD Feasibility Study, the estimate of p is 59.54 percent. This was computed as the quotient of the number of eligible units contacted (2,699) and the sum of the number of eligible and ineligible units contacted (2,699 + 1,834). Values of p are displayed in Table 4. It is interesting to note that during an extended followup of 223 unresolved units in the Feasibility Study, 59.00 percent of those that could eventually be resolved were found to be residential. This perhaps lends credence to the possibility that unresolved units have approximately the same proportion residential as the remainder of the sample. The values of R_C obtained from this study are displayed in Table 2.

Comparison to Other Telephone Health Surveys

In a 1983 paper [5], the response rate estimator R_C was computed for five telephone health surveys. All of these surveys were conducted using RDD sampling techniques. Since the specific procedures used in contacting sample units and the specific survey goals and questionnaire content differ from survey to survey, it is difficult to compare the response rates obtained. Still, the attempt will be made for the two surveys most closely resembling the NHIS study.

Survey	R_C
1) National Telephone Health Interview Survey (THIS) (considered units where busy signals were obtained as ineligible, not unresolved)	.82
2) National Survey of Personal Health Practices and Health Consequences (NSPHPC)	.69
3) NHIS-RDD Feasibility Study	.79

The rate from the NHIS-RDD Feasibility Study compared very favorably with the rates from the other RDD surveys. One additional study which should be used for comparative purposes is the experimental RDD NHIS conducted in 1979 by the Survey Research Center (SRC) of the University of Michigan. The SRC reported a response rate of approximately 79%. This rate was obtained using unlimited calls to households with callbacks allowed a month or more after the initial interviewing period.

Observations

Several valuable lessons can be learned from the experience of the NHIS-RDD Feasibility Study. They can best be summed up as follows:

1) Response rates can be directly influenced by interviewer experience and the type of supervision. The importance of experience can easily be seen by the progress that occurred between the early and later replicates of the Feasibility Study. Values of R_L, R_C, and R_U from replicates 10 through 12 were at least 12 percentage points higher than their corresponding values from replicates 1 through 3. From replicates 10 through 12, the value of R_L was 83.13 percent, R_C was 84.92 percent, and R_U was 88.09 percent. Thus, response rates of 85 percent or higher are within reason for the NHIS using RDD procedures provided that a well-trained and experienced staff of interviewers is maintained.

2) Improved methods must be developed for quickly identifying ineligible PSUs. Three PSUs (or 36 potential interviews) were lost from the Feasibility Study because they were identified as ineligible too late to generate a replacement PSU. The problem appears related to identifying special places over the telephone (since certain ineligible PSUs contained only special places) and to identifying sparse PSUs (i.e., those with very few or no eligible residences). Work is proceeding at the Bureau in both of these areas. A report on special places from the Feasibilty Study has been prepared [6] and a method for determining sparse PSUs has been investigated [7,8].

3) Even though the Feasibility Study used a 3-week interviewing period, 28 potential interviews were lost because a unit was determined to be ineligible too late to generate a replacement. Improvements are needed in this area. Currently being considered are modifications to the automated call scheduler which should assist in contacting and identifying hard to reach units more quickly.

References

[1] Waksberg, J. (1978). "Sampling Methods for Random Digit Dialing". Journal of the American Statistical Association, 70, 40-46.

[2] January 4, 1984 Census memorandum from Robert Tegels and David W. Chapman, "Sampling Specifications for the National Health Interview Survey - Random Digit Dialing (NHIS-RDD) Feasibility Study".

[3[September 13, 1984 Census memorandum from Gregory E. Sliwa, "Questionnaire Analysis in the 1984 NHIS-RDD Feasibility Study".

[4] Frankel, L.R. et al (1982). "On the definition of response rates". A special report of the Council of American Survey Research Organizations (CASRO) Task Force on Completion Rates, published by CASRO, New York.

[5] White, Andrew A. (1983). "Response Rate Calculation in RDD Telephone Health Surveys: Current Practices". 1983 Section on Survey Research Methods Proceedings of the American Statistical Association.

[6] July 30, 1984 Census memorandum from Michael Tenebaum, "Results of NHIS-RDD Special Place Study".

[7] May 8, 1984 Census memorandum from David W. Chapman and Carma R. Hogue, "PSU Cutoff Procedures for the NHIS-RDD Feasibility Study".

[8] June 4, 1984 Census memorandum from David W. Chapman, "Technical Documentation of the Procedure to Determine Cutoff Points for the NHIS-RDD".

Table 1: Values of R_L (lower bound on response rate)

	Form THIS-2X	Form THIS-3X	Combined Forms
Replicates 1-3	.6614	.7082	.6848
Replicates 4-6	.7636	.7108	.7371
Replicates 7-9	.7911	.7961	.7936
Replicates 10-12	.8221	.8404	.8313
Replicates 1-12	.7588	.7637	.7612

Table 2: Values of R_C (The CASRO Task Force suggested response rate estimator)

	Form THIS-2X	Form THIS-3X	Combined Forms
Replicates 1-3	.6874	.7370	.7121
Replicates 4-6	.7929	.7463	.7697
Replicates 7-9	.8258	.8279	.8269
Replicates 10-12	.8425	.8558	.8492
Replicates 1-12	.7864	.7918	.7891

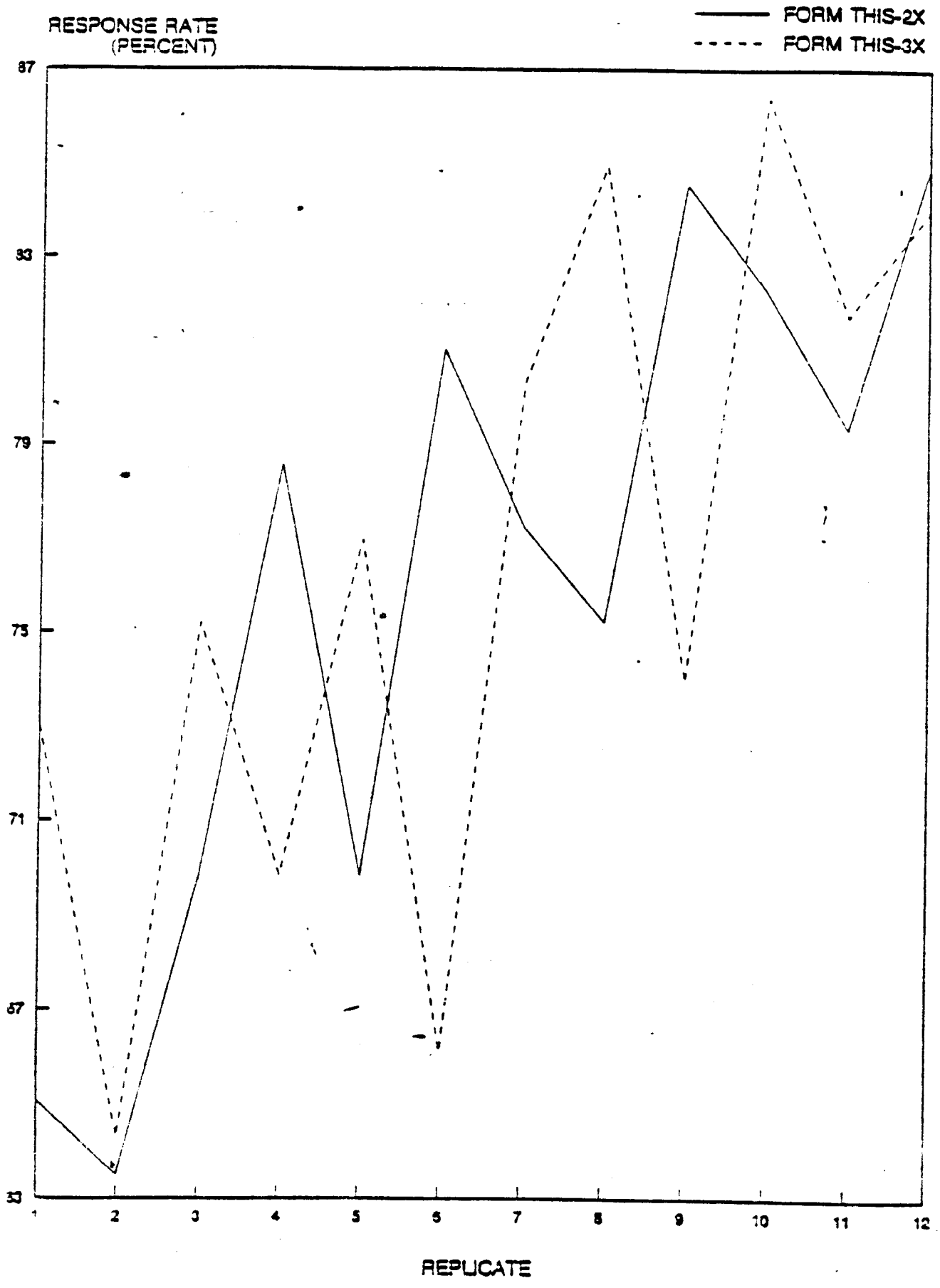
Table 3: Values of R_U (upper bound on response rate)

	Form THIS-2X	Form THIS-3X	Combined Forms
Replicates 1-3	.7310	.7853	.7581
Replicates 4-6	.8413	.8067	.8242
Replicates 7-9	.8738	.8716	.8727
Replicates 10-12	.8790	.8827	.8809
Replicates 1-12	.8309	.8372	.8340

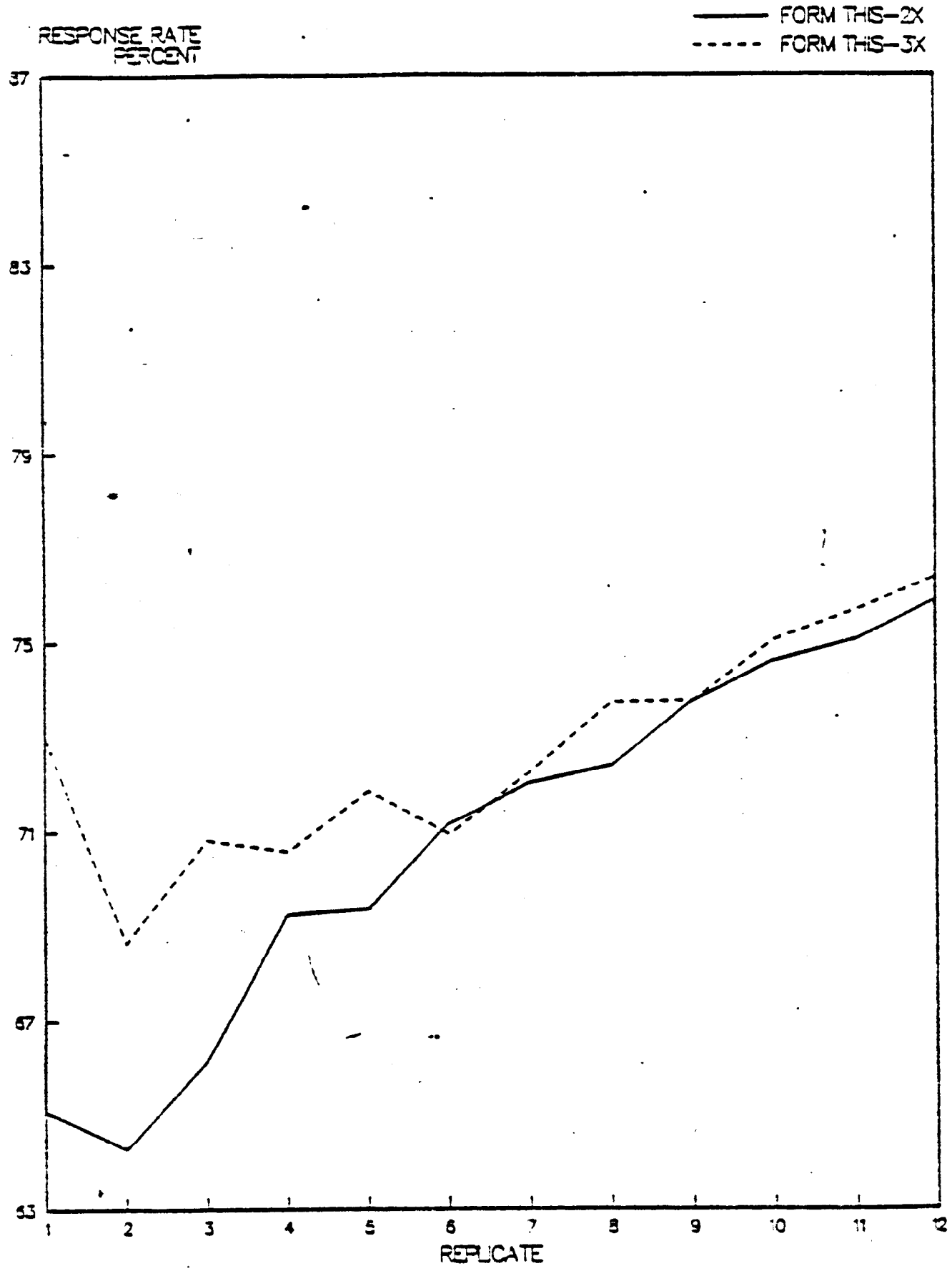
Table 4: Values of p (for use in computing R_C)

	Number of Eligible Units	Number of Ineligible Units	Number of Unresolved Units	p
Replicates 1-3	682	450	73	.6025
Replicates 4-6	660	440	78	.6000
Replicates 7-9	652	521	65	.5558
Replicates 10-12	705	423	42	.6250
Replicates 1-12	2699	1834	258	.5954

IHIS-RDD
.OWER BOUND RESPONSE RATES BY REPLICATE



NHIS–RDD LOWER BOUND RESPONSE RATES
CUMULATIVE THROUGH EACH REPLICATE



Appendix 2. Breakoff Analysis

I. Background

In the last few years, there has been interest within the Census Bureau and other government agencies in exploring the use of telephone interviewing instead of or in combination with face-fo-face interviewing. Telephone interviewing presents a less costly alternative to personal field visits; however, because of its relatively recent development, questions remain concerning comparability of the data collected in terms of response rates and data quality.

Response rates for telephone surveys have typically been lower than those achieved for personal visit surveys. The Census Bureau has been involved in several telephone interviewing projects to date; in each case, one of the objectives has been to see what kind of response rates can be achieved. The latest study, conducted in cooperation with the National Center for Health Statistics (NCHS), tested the feasibility of conducting the National Health Interview Survey (NHIS) using random digit dialing (RDD) techniques.

One vehicle for increasing response rates is to learn as much as possible about breakoffs occurring during the telephone interview. This report presents the results of an investigation of breakoffs in the NHIS-RDD Feasibility Study.

II. Major Findings and Conclusions

The most definitive finding to evolve from this research is that the results of the investigation are inconclusive. Two major factors contrib- to the indefinite nature of the results: first, an inability to associate specific breakoff locations with the telephone calls on which the breakoffs occurred; and second, a very low item response rate (21 percent) for the "location of breakoff" fields. It appears that these problems result from the way the data files were structured and/or programmed; thus, the major conclusion we can draw is that we need to correct the case management system to produce data better able to meet the needs of the research. Work is currently underway in this area.

In view of the magnitude of the nonresponse, any other results are subject to a large margin of error. Nevertheless, if we assume that breakoffs for cases at unknown locations are distributed in the same way as the

observed breakoffs, we can suggest, based on these data, that breakoffs at the household roster do not appear to occur as frequently as was indicated by interviewer reports. Since the vast majority of observed breakoffs occurred during the CATI screening section, before the household roster was reached, this does not necessarily imply that the household roster is not a problem, only that other parts of the questionnaire present more serious problems in terms of incurring breakoffs.

III. Details of the Investigation

The original aim of this analysis was to produce frequency counts of where refusal breakoffs occurred during the NHIS-RDD Feasibility Study. However, this has not been feasible, due to problems in the way the case management data were collected. It is not possible to determine which breakoff points are associated with which outcome codes, so we cannot limit the tabulations to those breakoffs which occurred during refusals.

Although the original plans cannot be implemented exactly, an approximation is made in this report: information is presented about breakoffs occurring in cases in which refusals were received (although we still don't know which of the breakoffs were associated with the refusal itself). This is not as useful as it might have been; however, it does give a general idea of where breakoffs occur.

The NHIS-RDD Feasibility Study questionnaire had three distinct parts:
1) a CATI screening section, containing the interviewer's introduction, privacy act statement, contacting an eligible respondent, etc.;
2) a cover booklet, containing the household roster questions and space for recording information about health conditions, hospitalizations, doctor visits, work history, etc.; and
3) an insert booklet, containing the body of the questions in the National Health Interview Survey.

Breakoffs occurring during each of these parts were collected and stored separately on the case management system. The exact item location of breakoffs occurring during the CATI screening section was stored automatically by the case management system; for breakoffs during the cover booklet and insert booklet, interviewers completed a series of questions (prompted by the case management system) at the conclusion of the interview and recorded the page number, item number, and item suffix (if applicable) of the item at which the breakoff occurred.

IV. Missing Data

There are two sources of missing data in this analysis. First are assignments in which breakoff entries are missing entirely when, according to their outcome code, entries should have been obtained. This could have occurred either during the interviewing or during the programming of the initial CATI screening.

It is not possible to determine how many entries are actually missing because of the way the system was set up--some outcome codes required (or did not require) breakoff entries in every case, and some outcome

codes required them in some instances but not in others. However, a general idea is provided by the following comparison: 2,148 assignments for refusal cases resulted in outcome codes which could have required breakoff entries and 458 breakoff entries were recorded. This may admittedly be an underestimate of the item response rate for these items. Nevertheless, there is a large margin for error in these data, since this comes out to be a 21.3 percent rate of response.

Second, among the fields that did contain information about the point of breakoff for refusal cases, 8 out of 458 assignments (approximately 1.7 percent) contained entries which were nonexistent; that is, the page and item numbers recorded in the data file did not exist on the questionnaire. These could have been the result of typos by the interviewers or perhaps interviewers used a different numbering system to identify the items. For example, in both the cover booklet fields and the insert booklet fields, breakoffs were recorded as occurring in items 1, 2, and 3 on page 1. However, the questions on page 1 of the cover booklet begin with item 3a. and the only question on page 1 of the insert booklet is an interviewer check item (which is also numbered as item 4). These items are excluded from this report.

The combined magnitude of these two sources of missing data, and the size of the first factor in particular, require that the reader look at these results with a certain amount of caution.

V. Results

A. Total Breakoffs

Table 1 presents the distribution of breakoffs received in cases which contained one or more refusal outcome codes broken down according to the three broad sections of the questionnaire. The table shows that almost three-fourths of breakoffs (73 percent) occurred during the CATI section, and the bulk of the remaining breakoffs (23 percent) occurred during the insert booklet part of the questionnaire. Only a very small portion of the breakoffs occurred during the cover booklet.

Another way of viewing the distribution of breakoffs is to look at the final refusal cases and see how much data was obtained before the refusal was finally accepted--that is, how far the interview proceeded before the last "last item answered" was received. Table 2 shows this distribution. Some information is available for 30 percent of the final refusal cases. (This information is the highest entry in any of the last item answered fields; it could have been from the initial refusal if that information was recorded and no subsequent information was recorded for the next refusal.)

Table 2 shows that, for those refusals in which we have some information about how much data was obtained from the respondent, the vast majority (88 percent) proceeded only as far as the CATI section. Only 2 percent of the refusals were terminated at the cover booklet, which contains the household roster, and the remaining 10 percent got

as far as the insert booklet. (One wonders why these 10 percent of the cases were coded as final refusals [outcome code=25] rather than partial interviews [outcome code=05 or 06], particularly since two of the cases reached as far as page 40 of the questionnaire.)

B. Breakoffs During CATI Screening Section

The interviewer began the interview by reading questions off the CATI screen. These questions included the interviewer's introduction, the information about the purpose of the survey, the voluntary nature of the survey, the fact that a supervisor might be listening in, etc. The flexibility of the CATI system in questionnaire design introduces new items which are asked only in particular, specialized situations. For example, when a callback is made to a household, a CATI screen contains a question stating that "earlier we talked to someone in your household." This item has a separate item number, which is recorded if a breakoff occurs at this point. In a regular paper-and-pencil interview, although the interviewer may have used the same wording in making his/her own introduction at a callback household, the item number that would be recorded for a breakoff would be item number of the general introduction. Thus, more precise information can be obtained with a CATI design.

Table 3 presents a description of the items contained in the CATI screening section. Since the conditions under which the questions were asked are included along with the questions themselves, the precision of the instrument can be readily observed. For example, the basic introduction, "Hello, I'm ... from the U.S. Census Bureau ..." is included in three different breakoff items: one for initial calls to a household, one for introductions to household members other than the phone-answerer, and one for callbacks to a household. To some extent, the outcome of the calls can be surmised from the identity of the breakoff point; for example, breakoffs occurring at Q2, "Have I reached you on your home phone?" may be mostly nonresidential numbers. However, we can't tell that for sure with these data, and residential respondents may also hang up at this point, without hearing what the interview will be about.

During the NHIS-RDD Feasibility Study, changes were made to the introduction on the CATI section during the course of data collection. The effect of the changes was to shorten the introduction by making it less wordy--items were not eliminated totally, but they were pruned to make them less verbose. These changes occurred at the beginning of rep 7; however, the changes were introduced on all the active cases as of a particular date. Any case active on March 19 was interviewed using the new introduction. Thus, all the cases in rep 7 used the second version of the introduction, cases in the second week of rep 6 used the new version, and cases in the third week of rep 5 used the new version as well.

Table 4 presents a frequency distribution of where breakoffs occurred during the CATI section. The largest portion of them occurred at item OTH, when the most knowledgeable respondent was not available and the interviewer asked to speak to some other eligible household

respondent. This item was responsible for over one-quarter of all the breakoffs that were recorded for refusal cases. Other "popular" breakoff points include Q6B, Q6A, and Q2.

Two different questionnaire versions, which differed slightly in certain procedures, were used in the study. In Form THIS-3X (Person by Section), some sections of the questionnaire were divided into groups of questions (about a page in length) which were asked about each family member before proceeding to the next group of questions. In Form THIS-2X (Family/Individual), all questions within those sections were asked about one family member before repeating all the questions for the next family member.

The CATI questions were the same regardless of which questionnaire version was administered. However, Table 5 shows that there are some differences in the location of breakoffs during the CATI section by questionnaire form. Most notably, there were more breakoffs at OTH using Form THIS-3X than with Form THIS-2X; in contrast, there were more breakoffs at Q2 using Form THIS-2X than with Form THIS-3X.

Data in Table 6 are broken down into groups of three replicates each; the data for Reps 4-6 are further divided into those assignments conducted using the original introduction and those conducted using the revised introduction. The new introduction appears to incur more breakoffs at Q6B and Q6A than the old one does. However, the increase in the number of breakoffs at Q6B may actually be an artifact of the changes to the introduction. Changes to the other items mainly consist of pruning words or phrases here and there; Q6B is the only place where there is a substantive change in the content of the item between the old and the new versions. In Q6B, a statement advising respondents to think carefully about answering all questions was replaced by an transitional lead-in statement to the household roster. The end result of the change is that in the new version, for any breakoff at the household roster occurring before the first name is obtained, the last item answered is recorded as Q6B; using the old version, breakoffs occurring at the same place would be recorded as Item 1, Page 2 of the cover booklet. Thus, the new version of the introduction may incur more breakoffs at the household roster than the old one.

C. Breakoffs During Cover Booklet Section

Questions in the cover booklet were not asked in the order that they appear in the booklet. After the CATI section was completed, the interviewer went to page 2 of the cover booklet and asked the household roster questions on pages 2 and 3. After the insert booklet had been completed, the interviewer returned to the cover booklet and asked the questions on page 4 (about the accuracy of answers to the health questions) and on page 1 (about household income and other telephones).

As noted in Table 1, the number of breakoffs recorded during the cover booklet is small. Since these breakoffs are limited to the household roster, as shown in Table 7, this suggests that the

roster is not responsible for any appreciable number of breakoffs during refusals. However, as pointed out in the previous section, some of the breakoffs recorded as occurring at Q6B of the CATI section may actually have occurred at the household roster. Since 17 percent of all the breakoffs occurred at Q6B (see Table 3), the number of breakoffs at the household roster could be considerably higher than indicated by Tables 1 and 7. Even so, the extent of refusals at the household roster as indicated by the data does not appear to reflect the reports of the interviewers, who said that most refusals occurred at the household roster.

Because of the small number of breakoffs recorded at this point, any breakdowns by questionnaire version or replicate group contain very small cell sizes and unstable percentages. Nevertheless, these data are presented in Tables 7-9.

D. Breakoffs During the Insert Booklet

Because of basic differences between Form THIS-2X and Form THIS-3X, the two questionnaire versions have very different numbering schemes in some parts of the interview. The frequency distribution contained in Table 10 presents a single listing of where the breakoffs occurred in this section. It shows that breakoffs occur throughout the insert booklet, from the beginning page through to the end, and that no one item is responsible for any appreciable number of breakoffs. The more noteworthy table is Table 11, which presents these breakoffs by questionnaire version. The cell sizes are really too small to make any generalizations. However, there are several asterisked (*) items which represent item numbers that do not exist on the questionnaire version in question. This suggests that the interviewers in these instances were not using the questionnaire version as instructed by the CATI system at the beginning of the interview.

Table 12 presents the distribution of breakoffs by replicate group. Again, it is not possible to draw any conclusions from these data but they are presented for informational purposes.

Attachments

Table 1. Frequency Distribution of Breakoffs for Refusal Cases by Questionnaire Section

Questionnaire Section	N	%
Total	2148	---
Location Unknown	1698	---
Location Known	450	100.0
CATI Screen	331	73.5
Cover Booklet	15	3.3
Insert Booklet	104	23.1

Table 2. Frequency Distribution of Furthest Point Reached in the Interview for Final Refusals by Questionnaire Section

Questionnaire Section	N	%
Total	457	---
Location Unknown	321	---
Location Known	136	100.0
CATI Screen	120	88.2
Cover Booklet	3	2.2
Insert Booklet	13	9.6

Table 3. Description of Screening Items Contained in CATI Section of the NHIS/RDD Feasibility Study -- Documentation of Changes Made During the Data Collection Period

Description of Items

Item Number	Original Version	Changed on March 19 (beginning of Rep 7)
Q1A	On original call: Hello, I'm ... from the U.S. Census Bureau in Washington, D.C. We are conducting a survey for the U.S. Public Health Service. To make sure I have dialed correctly, is this (tel #)?	On original call: Hello, I'm ... from the U.S. Census Bureau in Washington, D.C. We are conducting a health survey for the U.S. Public Health Service. Is this (tel #)?
Q1B	What number have I reached?	What number have I reached?
Q2	Have I reached you on your home phone?	Have I reached you on your home phone?
Q3	What kind of place does this telephone number serve?	What kind of place does this telephone number serve?
Q4	Does anyone use this number for a home phone?	Does anyone use this number for a home phone?
Q5	This survey is being conducted to collect information on the Nation's health. It is very important to have good answers to the health questions I will be asking. For that reason, I would like to speak to someone in the household who is at least 19 years old and knows the MOST about the health of the people in this family.	This survey collects information on the Nation's health. I would like to speak to someone in the household who is at least 19 years old and knows the MOST about the health of the people in this family.
Q5C	Phone-answerer not eligible respondent; MKR called to phone--breakoff occurred at "Hello. I'm ... from the U.S. Census Bureau in Washington, D.C. We are conducting a survey for the U.S. Public Health Service to collect information on the nation's health. I was told that you would know the MOST about the health of the people in this family."	Phone-answerer not eligible respondent; MKR called to phone--breakoff occurred at "Hello, I'm ... from the U.S. Census Bureau in Washington, D.C. We are conducting a health survey for the U.S. Public Health Service. I was told that you would know the MOST about the health of the people in this family."

Item Number	Original Version	Changed on March 19 (beginning of Rep 7)
Q6A	The survey is authorized by the Public Health Service Act. The results of the survey will be used for statistical research on health problems and all information you give will be kept confidential. Of course, your help on this survey is voluntary, but it is important that you and everyone selected for our survey participate so that we can make accurate estimates of the nation's health. In order to evaluate my performance, my supervisor may listen in.	This survey is authorized by the Public Health Service Act. The results of the survey will be used for statistical research on health problems and all information you give will be kept confidential. Your voluntary participation is extremely important to help us obtain complete and accurate results. In order to evaluate my performance, my supervisor may listen in.
Q6B	Since it is very important to have good answers to the health questions I will be asking, I would like you to think carefully about each question before answering, even those questions which seem unimportant to you.	I will ask you about hospital stays, visits to doctors, illness in the family, and other health related items. Since some questions won't apply to everyone, first I'll need to ask a few questions about the people living in your household.
OTH	MKR is not available; on a callback, "Is there someone else in the household who could answer these questions?	MKR is not available; "I wanted to talk to ... since I was told he/she knew the most about the health of people in the family. But, since he/she is not available, I would like to speak to anyone in this family who is at least 19 years old.
QCB1A	Household already contacted; on callback, person answering phone refused for respondent.	Household already contacted; on callback ("Hello, this is ... from the U.S. Census Bureau in Washington, D.C. May I please speak to ...?"), person answering phone refused for respondent
BRK	Household already contacted and some questions answered; on callback, respondent broke off at "Hello, I'm ... from the U.S. Census Bureau. I'm calling about the National Health Interview Survey. Our records indicate that part of an interview was completed for this household, and I would like to complete that interview now if you have a few minutes."	Household already contacted and some questions answered; on callback, respondent broke off at "Hello, I'm ... from the U.S. Census Bureau. I'm calling about the National Health Interview Survey. Our records indicate that part of an interview was completed for this household, and I would like to complete that interview now if you have a few minutes."

Table 4. Frequency Distribution of Breakoffs in CATI Section by Item Number for Refusal Cases

Item Number	N	as % of known breakoffs during CATI section (N=331)	as % of total breakoffs with known location (N=450)
Q1A	1	.3%	.2%
Q2	35	10.6	7.8
Q3	1	.3	.2
Q5C	31	9.4	6.9
Q6A	43	13.0	9.6
Q6B	77	23.3	17.1
OTH	117	35.3	26.0
BRK	26	7.9	5.8

NOTE: The accuracy of these figures is unknown, due to high item nonresponse rates.

Table 5. Percentage of Breakoffs During CATI Section by Questionnaire Version for Refusal Cases

Item Number	Questionnaire Version	
	Form THIS-2X	Form THIS-3X
N (known breakoffs)	185	146
%	100.0	100.0
Q1A	.5%	0.0%
Q2	15.1	4.8
Q3	.5	0.0
Q5C	1.1	7.5
Q6A	12.4	13.7
Q6B	24.9	21.2
OTH	28.6	43.8
BRK	7.0	8.9

NOTE: The accuracy of these figures is unknown, due to high item nonresponse rates.

Table 6. Percentage of Breakoffs During CATI Section by Replicate Group for Refusal Cases

Item Number	Replicate Group				
	Reps 1 - 3	Reps 4 - 6 old intro	Reps 4 - 6 new intro	Reps 7 - 9	Reps 10 - 12
N (known breakoffs)	13	36	21	79	110
%	100.0	100.0	100.0	100.0	100.0
Q1A	1.2%	0.0%	0.0%	0.0%	0.0%
Q2	2.1	13.9	4.8	1.3	9.1
Q3	1.2	0.0	0.0	0.0	0.0
Q5C	12.9	11.1	9.5	15.2	1.8
Q6A	8.2	13.9	0.0	20.3	13.6
Q6B	10.6	16.7	9.5	30.4	32.7
OTH	29.4	38.9	66.7	27.8	38.2
BRK	15.3	5.5	9.5	5.1	4.5

NOTE: The accuracy of these figures is unknown, due to high item nonresponse rates.

Table 7. Frequency Distribution of Breakoffs During Cover Booklet by Item Number for Refusal Cases

Page No.	Item No.	Description of Item	N	as % of known breakoffs during cover booklet (N=15)	as % of all breakoffs with known location (N=450)
2&3	1	roster, name item	4	26.7%	.9%
2&3	2	roster, relationship item	6	40.0	1.3
2&3	3	roster, date of birth item	5	33.3	1.1

NOTE: The accuracy of these figures is unknown, due to high item nonresponse rates.

Table 8. Percentage of Breakoffs During Cover Booklet by Questionnaire Version for Refusal Cases

Page and Item Numbers	Questionnaire Version	
	Form THIS-2X	Form THIS-3X
N (known breakoffs)	8	7
%	100.0	100.0
Page 2&3, Item 1	37.5%	14.3%
Page 2&3, Item 2	25.0	57.1
Page 2&3, Item 3	37.5	28.6

NOTE: The accuracy of these figures is unknown, due to high item nonresponse rates.

Table 9. Percentage of Breakoffs During Cover Booklet by Replicate Group for Refusal Cases

Page and Item Numbers	Replicate Group			
	Reps 1 - 3	Reps 4 - 6	Reps 7 - 9	Reps 10 - 12
N (known breakoffs)	3	6	3	3
%	100.0	100.0	100.0	100.0
Page 2&3, Item 1	66.7%	33.3%	0.0%	33.3%
Page 2&3, Item 2	0.0	50.0	66.7	33.3
Page 2&3, Item 3	33.3	16.7	33.3	66.7

NOTE: The accuracy of these figures is unknown, due to high item nonresponse rates.

Table 10. Frequency Distribution of Breakoffs During Insert Booklet by Item Number for Refusal Cases

age mber	Item Number	N	as % of known brks during insert booklet (N=104)	as % of total brks with known location (N=450)
2	1	4	3.8%	.9%
	2	1	1.0	.2
4	1	2	1.9	.4
	5	1	1.0	.2
	6	3	2.9	.7
5	1	1	1.0	.2
	2	1	1.0	.2
	12	1	1.0	.2
6	14	4	3.8	.9
7	6	3	2.9	.7
8	14	1	1.0	.2
9	14	1	1.0	.2
10	3	1	1.0	.2
	6	2	1.9	.4
11	1	1	1.0	.2
	6	2	1.9	.4
16	1	1	1.0	.2
	3	4	1.0	.2
18	4	2	1.9	.4
	5	1	1.0	.2
19	4	1	1.0	.2
20	1	4	3.8	.9
	2	1	1.0	.2
	5	1	1.0	.2
21	1	2	1.9	.4
	3	2	1.9	.4
	4	1	1.0	.2
	5	2	1.9	.4
	6	1	1.0	.2
22	6	1	1.0	.2
	15	1	1.0	.2
	21	1	1.0	.2
	26	2	1.9	.4
23	23	1	1.0	.2
24	2	1	1.0	.2
	6	1	1.0	.2
26	3	1	1.0	.2
	4	1	1.0	.2
27	3	1	1.0	.2
28	2	1	1.0	.2
	3	3	2.9	.7
	5	1	1.0	.2
29	3	1	1.0	.2
	4	1	1.0	.2
	5	3	2.9	.7
30	2	1	1.0	.2
	3	4	3.8	.9
31	5	1	1.0	.2
33	9	1	1.0	.2
	17	1	1.0	.2
	23	2	1.9	.4
34	22	2	1.9	.4
36	1	1	1.0	.2
	2	1	1.0	.2
	6	2	1.9	.4
37	6	2	1.9	.4
38	1	2	1.9	.4
	2	1	1.0	.2
	3	1	1.0	.2
40	1	1	1.0	.2
42	4	1	1.0	.2
43	2	1	1.0	.2
	12	2	1.9	.4
44	5	3	2.9	.7
	6	2	1.9	.4
48	5	1	1.0	.2
55	6	1	1.0	.2
56	1	1	1.0	.2

E: The accuracy of these figures is unknown, due to high item nonresponse rate.

Table II. Percentage of Breakoffs During Insert Booklet by Questionnaire Version for Refusal Cases

age mber	Item Number	Questionnaire Version Form THIS-2X (N=46)	Form THIS-3X (N=58)
2	1	0.0%	6.9%
	2	2.2	0.0
4	1	4.3	0.0
	5	0.0	1.7
	6	4.3	1.7
5	1	2.2	--
	2	2.2	--
	12	--	1.7
6	14	2.2*	5.1
7	6	--	5.1
8	14	2.2	--
9	14	0.0	1.7
10	3	2.2	0.0
	6	4.3	0.0
11	1	2.2	--
	6	2.2	--
16	1	2.2	0.0
	3	2.2	0.0
18	4	4.3	--
	5	0.0	1.7*
19	4	2.2	--
20	1	0.0	6.9
	2	2.2	0.0
	5	2.2	0.0
21	1	0.0	3.4
	3	4.3	0.0
	4	2.2	0.0
	5	4.3	0.0
	6	--	1.7
22	6	0.0	1.7
	15	2.2	--
	21	2.2	--
	26	4.3	--
23	23	0.0	1.7*

Page Number	Item Number	Questionnaire Version Form THIS-2X (N=46)	Form THIS-3X (N=58)
24	2	2.2	0.0
	6	0.0	1.7
26	3	2.2	0.0
	4	2.2	--
27	3	0.0	1.7
28	2	0.0	1.7
	3	6.5	0.0
	5	2.2	0.0
29	3	--	1.7
	4	--	1.7
	5	4.3*	1.7
30	2	0.0	1.7
	3	0.0	6.9
31	5	--	1.7
33	9	0.0	1.7
	17	0.0	1.7
	23	--	3.4
34	22	--	3.4
36	1	0.0	1.7
	2	0.0	1.7
	6	--	3.4
37	6	0.0	3.4
38	1	0.0	3.4
	2	0.0	1.7
	3	0.0	1.7
40	1	0.0	1.7
42	4	2.2	0.0
43	2	2.2	0.0
	12	--	3.4
44	5	6.5	0.0
	6	4.3	--
48	5	--	1.7
55	6	--	1.7
56	1	--	1.7

OTES:
- indicates that this item does not appear on this questionnaire version.
* indicates that this item does not appear on this questionnaire version, but breakoff entries were recorded nevertheless.
he accuracy of these figures is unknown, due to high item nonresponse rates.

Table 12. Percentage of Breakoffs During Insert Booklet by Questionnaire Version for Refusal Cases

Page Number	Item Number	Replicate Group			
		Reps 1-3 (N=29)	Reps 4-6 (N=29)	Reps 7-9 (N=18)	Reps 10-12 (N=28)
2	1	10.3%	3.4%	0.0%	0.0%
	2	3.4	0.0	0.0	0.0
4	1	0.0	3.4	0.0	3.6
	5	0.0	0.0	0.0	3.6
	6	6.9	3.4	0.0	0.0
5	1	0.0	0.0	0.0	3.6
	2	0.0	0.0	5.6	0.0
	12	3.4	0.0	0.0	0.0
6	14	0.0	3.4	0.0	10.7
7	6	3.4	3.4	0.0	3.6
8	14	3.4	0.0	0.0	0.0
9	14	0.0	3.4	0.0	0.0
10	3	3.4	0.0	0.0	0.0
	6	0.0	6.9	0.0	0.0
11	1	0.0	3.4	0.0	0.0
	6	0.0	0.0	5.6	0.0
16	1	0.0	3.4	0.0	0.0
	3	0.0	0.0	0.0	3.6
18	4	3.4	3.4	0.0	0.0
	5	0.0	3.4	0.0	0.0
19	4	0.0	3.4	0.0	0.0
20	1	6.9	3.4	5.6	0.0
	2	3.4	0.0	0.0	0.0
	5	0.0	3.4	0.0	0.0
21	1	3.4	3.4	0.0	0.0
	3	3.4	3.4	0.0	0.0
	4	0.0	0.0	0.0	3.6
	5	0.0	0.0	5.6	3.6
	6	0.0	0.0	5.6	0.0
22	6	0.0	0.0	5.6	0.0
	15	0.0	3.4	0.0	0.0
	21	0.0	0.0	0.0	3.6
	26	0.0	3.4	5.6	0.0
23	23	0.0	0.0	0.0	3.6

NOTE: The accuracy of these figures is unknown, due to high item nonresponse rates.

Page Number	Item Number	Replicate Group			
		Reps 1-3	Reps 4-6	Reps 7-9	Reps 10-12
24	2	3.4	0.0	0.0	0.0
	6	0.0	0.0	5.6	0.0
26	3	0.0	3.4	0.0	0.0
	4	3.4	0.0	0.0	0.0
27	3	3.4	0.0	0.0	0.0
28	2	0.0	0.0	5.6	0.0
	3	3.4	0.0	0.0	7.1
	5	3.4	0.0	0.0	0.0
29	3	0.0	0.0	5.6	0.0
	4	0.0	0.0	5.6	0.0
	5	3.4	0.0	0.0	7.1
30	2	0.0	3.4	0.0	0.0
	3	6.9	3.4	5.6	0.0
31	5	3.4	0.0	0.0	0.0
33	9	0.0	0.0	5.6	0.0
	17	0.0	0.0	5.6	0.0
	23	0.0	0.0	0.0	7.1
34	22	0.0	0.0	0.0	7.1
36	1	0.0	0.0	0.0	3.6
	2	0.0	3.4	0.0	0.0
	6	3.4	3.4	0.0	0.0
37	6	0.0	6.9	0.0	0.0
38	1	0.0	3.4	0.0	3.6
	2	0.0	0.0	0.0	3.6
	3	0.0	0.0	5.6	0.0
40	1	0.0	0.0	0.0	3.6
42	4	3.4	0.0	0.0	0.0
43	2	0.0	0.0	0.0	3.6
	12	0.0	0.0	0.0	7.1
44	5	0.0	3.4	11.1	0.0
	6	3.4	0.0	5.6	0.0
48	5	3.4	0.0	0.0	0.0
55	6	0.0	0.0	0.0	3.6
56	1	0.0	3.4	0.0	0.0

Appendix 3. Questionnaire Analysis

Introduction.

This report analyses the differences between various estimates obtained from two questionnaire types used in the 1984 NHIS-RDD Feasibility Study. Estimates for each questionnaire are also compared with similar estimates from both the regular (personal interview) NHIS and from an RDD health survey conducted by the Survey Research Center (SRC) of the University of Michigan.

Section I of this report focuses on the comparisons between the two questionnaires for the feasibility study, and Section II deals with comparisons to the other two health surveys. In each section differences in characteristics for 8 demographic variables and 21 health variables are examined.

Section I. Differences between estimates obtained from the two feasibility study questionnaire forms.

The sample for the feasibility study consisted of approximately 1,500 telephone households selected for each questionnaire. Most of the sections for both questionnaire versions were identical. However, there were some differences in a few sections. The main difference in these sections between the two questionnaires was the order in which the questions were asked. In one questionnaire, designated the Person by Section Version, all questions within a section were asked about one family member before proceeding to the next family member. The other questionnaire, called the Family/Individual Version, had breaks within sections where interviewers were instructed to return to the most recent series of questions and ask these questions about the next family member. Another minor distinction between the two questionnaires was that some questions in the Family/Individual Version attempted to obtain individual information through an inquiry concerning the entire family. In the Person by Section Version such information was obtained by questions directly asked about each individual family member.

The results in Section I of this analysis are divided into three parts. Part 1 compares demographic characteristics of the interviewed samples for each questionnaire. Part 2 examines the differences between questionnaires in percent reporting for 9 health variables and in mean levels for 8 health variables. Part 3 looks at questionnaire differences in percent reporting for 21 health variables by sex, age, and education. In Section I, all tabulations are for the sample of interviewed households, which includes original, replacement, and substitute households.

I.1. Demographic comparisons.

Table 1 displays the demographic characteristics of the interviewed portion of the NHIS-RDD sample by type of questionnaire. The demographic composition of the sample interviewed using the Family/Individual Version is very similar to that of the sample interviewed using the Person by Section Version. For example, only two variable-category breakdowns (Income-DK/NA/Refusal and Usual Activity-Something else) exhibited differences greater than 2 percent.

For the Person by Section Version, income nonresponse (i.e., DK/NA/Refusal) was 4 percentage points lower (23.5 percent versus 27.5 percent), and all income categories had slightly higher percents. A more surprising result not shown in Table 1 was that for each questionnaire a very high percentage (73.3 percent for the Person by Section and 67.4 percent for the Family/Individual) of the interviewed families reported an actual approximate dollar amount for combined family income during the past 12 months. The remaining 3.2 percent of the Person by Section families and 5.1 percent of Family/Individual families went through a series of "splitting" questions which placed each family in a specified income range.

The response, "something else", to the question on usual activity for most of the past 12 months was 2.3 percent higher (12.1 percent versus 9.9 percent) on the Person by Section questionnaire. One explanation for this higher reporting is that the interviewed portion of the Person by Section sample contained more older people. Although the individual age-category breakdowns are very similar between the two questionnaires, the last four age categories (45-54, 55-64, 65-74, and 75+) had 2.6 percent more persons for the Person by Section Version. These last four age categories are more likely to contain retired people or people unable to work, and for such people the category, "something else", would describe their usual activity, as opposed to the other choices of "working", "keeping house", and "going to school".

I.2. Overall comparisons of health characteristics.

Table 2.A lists the frequency distributions for nine selected health characteristics by type of questionnaire. An examination of Table 2.A reveals little differences in the percent breakdowns between versions for the nine variables. For example, only two categories, "no twelve-month bed days" (50.2 percent on the Family/Individual and 47.4 percent on the Person by Section) and "excellent" health status (38.1 percent on the Family/Individual and 40.2 percent on the Person by Section) differed by more than 2 percentage points, and if DK's, NA's, and refusals are eliminated from consideration only" excellent" health status had a difference exceeding 2 percentage points.

Eight of the nine variables in Table 2.A require that the respondent provide a numerical response. Health status was the only variable for which the responses were nonnumeric. For each of the eight variables requiring a numbered response, the percent reporting at least one occurrence was slightly higher on the Person by Section Version. However, only one variable, "two-week cut-down days", had higher reporting on the Person by Section Version for all numerical ranges.

Table 2.B displays the mean levels, differences, and standard errors of the differences for the eight numeric variables mentioned in Table 2.A. Mean levels and variances of the differences were calculated using equations (1) and (2), respectively, which appear in the Appendix.

"Two-week cut-down days" was the lone variable to exhibit a significant difference in mean levels between questionnaire types at the 5 percent significance level. The average number of two-week cut-down days reported on the Person by Section Version was 0.3606 as compared with only 0.2406 on the Family/Individual Version.

Section D, which contained the question on two-week cut-down days, was exactly the same for both questionnaires. However, the most drastic difference between the two questionnaires was in the format of Section B, which dealt with activity limitations and immediately preceeded Section D*. This difference may have conditioned responses to the two-week cut-down days question.

This explanation seems plausible when viewed in terms of the other variables in Table 2.B. The question on 13-month hospital stays preceeded Section B, and the average number of stays on the Family/Individual Version was identical to that on the Person by Section Version. All six other variables in Table 2.B occurred in sections after Section B, and only one, "two-week work-loss days", had a lower mean level on the Person by Section Version. However, the percent reporting at least one "two-week work-loss day" (see Table 2.A) was slightly higher on the Person by Section Version. The question dealing with two-week work-loss days was the first question asked after Section B, and thus any conditioning effects may have been minimal. Data on two-week bed days and two-week school-loss days were obtained later in Section D. The section concerned with two-week doctor visits and the section referring to 12-month doctor visits and bed days had slight format changes and some family-style questions on the Family/Individual Version, which, in addition to any conditioning, may have been responsible for the higher mean levels observed for these questions on the Person by Section Version.

*Note: No content sections of either questionnaire were labelled as "Section C", since "C" was reserved for a roster of hospitalizations, doctor visits, and conditions.

I.3. Comparisons of health characteristics by sex, age, and education.

Tables 3.A, 3.B, and 3.C display the distributions of the interviewed sample in selected response categories for 21 health characteristics for each questionnaire by sex, age, and education, respectively. Percents in each response category and variances of the differences were computed according to the same equations used for mean levels with appropriate changes as described in the Appendix.

For both males and females, a tendency towards higher reporting of health events or occurrences (i.e., all response categories except percent having "excellent" health status) was observed in the Person by Section Version (see Table 3.A.). The tendency was more pronounced among females, where for the Person by Section Version, 17 of the 20 response categories had higher reporting with six of the categories (cut-down days; total, RA, and CL conditions; play limitation, and 12-month bed days) having significantly higher reporting at the 5 percent significance level. Since self respondents usually report more events than proxies, one might conjecture that these differences were probably a result of more female self respondents for the Person by Section Version. However, this was not the case. The Family/Individual Version had 4.4 percent more female self respondents than the Person by Section Version. Increased reporting of health events or occurrences was observed for males on the Person by Section Version in 12 of the 20 response categories, but no differences were statistically significant. The percentage of males reported as having "excellent" health status was significantly higher for the Person by Section Version.

Table 3.B shows that the Person by Section Version had about the same or slightly greater reporting of health events or occurrences as the Family/Individual Version over all age categories. Five differences were statistically significant. For the over 65 age group, "two-week cut-down days" and "RA conditions" were significantly higher on the Person by Section Version. Twelve-month doctor visits were reported for a significantly higher percentage of persons aged 16 and under and aged 25 to 44 on the Person by Section Version. However, for persons 45 to 64 years old, the Family/Individual Version produced a significantly greater percentage of 12-month doctor visits. Excellent health status, the lone nonhealth event or occurrence variable, was reported for a significantly larger percent of people in the 16 and under and in the 45-64 age categories using the person by section format.

Only one health event or occurrence variable for each education level showed significantly higher reporting, and in each instance the Person by Section Version had the higher reporting (see Table 3.C). "Two-week cut-down days" for people with 11 or less years of schooling, "total conditions" for people with 13 or more years of schooling, and "other" limitations in the 12 years of education category accounted for the significant differences. No differences were significant between questionnaires in any of the three education levels for the percent of persons having "excellent" health status.

One surprising result not observed in any of the previous tables did occur in Table 3.C. There was a tendency towards more reporting, though not significant, of health events or occurrences on the Family/Individual Version for persons with only 12 years of schooling. The number of variables exhibiting this tendency was 14 out of the 19 eligible variables. In the 0 to 11 years of education category, 17 of the 20 eligible variables had slightly higher reporting under the person by section format, and in the 13 years-plus education category, higher reporting occurred under the person by section format for 12 of the 17 eligible variables.

Section II. Comparisons with other telephone and personal interview health surveys.

This section compares estimates obtained from each version of the NHIS-RDD questionnaire with those obtained from both the personal interview NHIS and from an RDD health survey conducted by SRC of the University of Michigan. The SRC survey was conducted in the fourth quarter of 1979 using a modified NHIS questionnaire on a national probability RDD telephone sample of persons 17 years or older. The research design for the SRC survey included a number of different treatments, which will be ignored. The regular, ongoing, personal interview NHIS data used for comparison purposes in this report will also be from the fourth quarter of 1979.*

*The fourth quarter NHIS in 1979 also included different treatments in a portion of the sample. Total estimates appearing in this memorandum are for the combined portions. Basically, the test version of the questionnaire in 1979 was similar to the NHIS-RDD Family/Individual Version. The regular questionnaire in 1979 was not similar to anything. It was not sectionalized to the same extent as the test version or either NHIS-RDD versions. Also, it contained many more family-style questions than any other version. The Michigan survey compared various experimental procedures, and estimates over all procedures are used in the memorandum. Also, the SRC version was most similar to the Person by Section Version.

The results in Section II are divided into three parts. Part 1 compares the demographic characteristics of each interviewed sample for the various health surveys. Part 2 examines comparisons of eight health characteristics, and part 3 looks at detailed comparisons for subsets of the eight health characteristiccs. All comparisons in this section are restricted to persons 17 years or older in interviewed households. Included are original, replacement, and substitute households interviewed in the feasibility study. The SRC survey and the regular NHIS yielded data on 8,210 and 19,800 persons 17 years or older, respectively. For the NHIS-RDD Feasibility Study, the Family/Individual and Person by Section questionnaires contained data on 2,770 and 2,795 persons 17 years or older, respectively. In addition, for comparative purposes the regular NHIS data for households with telephones (18,388 of the 19,800 persons 17 years or older) are also examined. These households are referred to as "telephone households" throughout the remainder of this report.

II.1. Demographic comparisons.

Table 4 displays the demographic characteristics of the telephone and personal interview samples for the various health surveys. The distributions of the sex, race, and marital status variables for each of the NHIS-RDD questionnaires were surprisingly similar to those observed in the other health surveys. The age distributions for each NHIS-RDD version more closely resembled that observed in the SRC telephone survey. Differences were more pronounced for the remaining three demographic variables (education, income, and usual activity).

For education, the 0 to 8 year category and the 9 to 12 year category, each comprised smaller portions of the NHIS-RDD samples, and the 13 to 18 year category represented larger percentages than the corresponding percentages observed for the other health surveys. The SRC telephone survey produced these same education differences when compared with the personal interview NHIS, (both all households and telephone only households), but the discrepancies were not as large.

The most drastic differences occurred among the income distributions, where both NHIS-RDD questionnaire versions displayed much higher percentages of families in the upper income bracket ($25,000+) and much lower percentages in the four lesser income categories when compared with the SRC telephone survey or either income distribution for the regular NHIS. Inflation and differences in the income questions are two explanations for these discrepancies. For the SRC telephone survey the percentage of families in the $25,000 and up category was 6 to 7 percentage points less than that of either of the regular NHIS. In the $5,000 to $9,999 category, the SRC percentage was lower (though not as low as the NHIS-RDD percentages) than the regular NHIS percentages. Otherwise, the SRC income distribution was fairly similar to the income distributions of personal interview NHIS.

The "usual activity" breakdowns were not as different as the breakdowns for education and income, but the telephone surveys showed some slight inconsistencies when compared with personal interview results. "Working" was reported for a greater percentage of persons in the SRC telephone survey and a smaller percentage in each NHIS-RDD version than in regular NHIS. Similar results occurred for the percent of persons reported as "keeping house", but the percent was only slightly higher in the telephone SRC. Only 5.7 percent of the Person by Section sample had "going to school" as their usual activity. The Family/Individual Version and SRC telephone survey reported 7.3 percent going to school as compared with 7.4 and 7.5 percent reported in the regular NHIS for all households and telephone households, respectively. For the catchall category "something else", the Family/Individual Version was about the same as the regular NHIS, while the Person by Section Version and the SRC survey were higher and lower, respectively.

One explanation for the distributional differences in the education, income, and usual activity variables might be the difference in survey periods. Both the SRC data and the regular NHIS data were collected during the fourth quarter of 1979, and the NHIS-RDD Feasibility Study was conducted early in 1984. Each survey had 12 or 13-month and 2-week reference periods which essentially covered the time immediately before the date of interview.

Another reason for the differences could be the high item nonresponse (DK's, NA's, and refusals) rates observed for these three variables in the NHIS-RDD Feasibility Study. Nonresponse to the education question averaged about 2 percent in the 1979 surveys and exceeded 6 percent in the 1984 feasibility study. The regular NHIS had an income item nonresponse rate of just under 9 percent. For the SRC survey, this rate jumped to almost 18 percent, and in the feasibility study another jump to over 25 percent occurred. The usual activity question had very low nonresponse (0.5 and 0.3 percent) in the 1979 surveys, but the nonresponse in the feasibilty study was over 3.5 percent. These differences in nonresponse rates are probably a result of differences in the proficiency levels of the various data collection staffs, with the regular NHIS staff being more proficient than the feasibility study staff. Little is known about the quality of the SRC staff.

II.2. Overall comparisons of health characteristics.

Table 5 displays some distributional comparisons for eight health characteristics between the various telephone and personal interview samples. Except for the health status variable, the responses on each NHIS-RDD questionnaire version seemed to be distributed similarly to responses observed by SRC and in the regular NHIS.

Although the nonresponse rates to the health status question were higher for the NHIS-RDD, the main reason for the discrepancies was probably that only four categories (excellent, good, fair, and poor) were available on the 1979 surveys. The 1984 feasibility study allowed an extra category, "very good", and in Table 5 the persons reported as "very good" in the NHIS-RDD are included under the category designation, "good".

II. 3. Comparisons of health characteristics by sex, age, and education.

Tables 6.A, 6.B, and 6.C contain data on the percent reporting selected health characteristics by sex, age, and education, respectively. For the regular personal interview NHIS, data are displayed only for telephone households.

For females on both NHIS-RDD questionnaire versions, the reporting of two-week work-loss days and doctor visits was more comparable to that observed by SRC, while 12-month bed days compared favorably to the regular NHIS (see Table 6.A). The same observation can be made on 12-month bed days reported for males, but the results for the other two variables were somewhat mixed.

Five health characteristics (two-week work-loss and cut-down days and doctor visits, 12-month bed days, and "excellent" health status) are examined in Table 6.B for the age groups 17-24, 25-44, 45-64, and 65+. Except for two-week work-loss days on the Family/Individual Version, reporting for persons aged 17 to 24 on both NHIS-RDD questionnaires more closely resembled that obtained for telephone households in the personal NHIS. The results were not as clearcut for the remaining age groups, especially 45-64 and 65+.

Table 6.C. displays the percent reporting for four health variables (two-week work loss and cut-down days, 12-month doctor visits, and "excellent" health status) by three education levels (0 to 11 years, 12 years, and 13 or more years). Two variables, "excellent" health status and two-week work-loss days, on both feasibility study questionnaires had percentages similar to the SRC telephone survey. With the exception of two-week cut-down days for persons at the lower education level on the Person by Section Version, percents comparable to those in the regular NHIS were observed for the other two variables.

III. Summary

If more reporting of health events or occurrences is acknowledged to be indicative of better reporting, then more consideration should be given towards a questionnaire similar to the Person by Section Version in any future NHIS conducted by telephone. Out of the 15 significant differences detected for the health events or occurrences in Tables 2.B, 3.A, 3.B, and 3.C, 14 showed more reporting on the Person by Section Version than on the Family/Individual Version. In addition, almost 62 percent of the differences that were not significant had increased reporting on the Person by Section Version.

Respondent conditioning was probably a more important contributing factor towards the increased reporting than any methodological differences between the two questionnaire versions. Most of the questionnaire differences were very minor with the exception of the limitation of activities section, which occurred early in both questionnaires. This section contained the limitations questions and was preceeded by a section containing questions on 13-month hospitalizations. All other health occurrences or events in Tables 2.B, 3.A, 3.B, and 3.C were addressed by questions in later sections. None of the significant differences in Tables 2.B, 3.A, 3.B, or 3.C were for 13-month hospitalizations, and only two, each indicating increased reporting on the Person by Section Version, occurred for the different limitations. Of the nonsignificant differences for 13-month hospitalizations and for limitations, only 48 percent involved more reporting on Person by Section Version. This does provide some evidence that the format change in the limitation of ativities section may have conditioned respondents to expect similar patterns of questioning in later sections, which led to either increased reporting on the Person by Section Version or to decreased reporting on the Family/Individual Version.

Differences in the demographic compositions of the two interviewed samples in the feasibility study were too small to explain any observed differences in the questionnaire versions. All sex, age, and education breakdowns generally tended to show higher reporting of health events or occurrences on the Person by Section Version.

One interesting anomaly, not previously mentioned in this report, did occur in the feasibility study. Although greater reporting of health events or occurrences, which might be indicative of less than excellent health status, was observed for the Person by Section Version, the percentage of persons reported as being in "excellent" health was also greater on the Person by Section Version than on the Family/Individual Version. For three of the demographic breakdowns (males, persons aged 16 and under, and persons aged 45 to 64), the percent was significantly greater. Though not impossible, this result might raise some suspicion about the accuracy of reporting in the feasibility study, and it is mentioned here as a caution to temper any enthusiasm for the Person by Section Version, since self-perceived health status is traditionally the most indicative single variable in NHIS related to other health measures.

Except for the few differences noted for education, income, and usual activity, the demographic makeup of the feasibility samples bore a good resemblence to that observed for other health surveys. The comparisons for overall reporting of health characteristics were even more similar. Some differences in the health characcteristics by sex, age, and education did occur between the other health surveys and each of the feasibility surveys, but these differences were most likely due to the small size of the feasibility samples.

Table 1. Demographic Characteristics of the Interviewed Portion of the NHIS/RDD Sample by Type of Questionnaire

Percent	Type of Questionnaire: Family/Individual	Type of Questionnaire: Person by Section
Sex		
Male	47.4	47.8
Female	51.5	51.3
DK/NA/Ref	1.1	1.0
Age		
17-24	17.0	15.6
25-34	23.9	24.3
35-44	18.6	17.8
45-55	14.2	13.4
55-64	11.2	12.7
65-74	9.2	9.8
75+	4.7	5.3
DK/NA/Ref	1.2	1.1
Race		
White	88.4	87.6
Nonwhite	11.6	12.4
Education		
0-8	21.3	21.2
9-12	42.9	41.6
13-18	30.4	30.4
DK/NA/Ref	5.4	6.8
Income		
Less than $5,000	4.7	4.8
$5,000-9,999	8.0	8.2
$10,000-14,999	8.7	8.9
$15,000-24,999	18.6	20.1
$25,000+	32.6	34.6
DK/NA/Ref	27.5	23.5
Marital Status		
Married	63.3	61.6
Widowed	6.0	6.2
Divorced	5.5	6.7
Separated	1.5	1.5
Never Married	23.7	23.9
Usual Activity		
Working	56.3	56.7
Keeping house	23.0	21.8
Going to school	7.3	5.7
Something else	9.9	12.1
DK/NA/Ref	3.6	3.7
Veteran Status		
Veteran	17.7	17.4
Nonveteran	76.2	75.9
DK/NA/Ref	6.1	6.7

Table 2.A. Health Characteristics for the NHIS/RDD Feasibility Study by Type of Questionnaire

Percent	Type of Questionnaire: Family/Individual	Type of Questionnaire: Person by Section
1. Two-week bed days		
None	92.5	91.7
1-3	5.4	6.1
4-7	1.3	1.3
8-10	0.2	0.1
11-14	0.6	0.7
2. Two-week work-loss days		
None	92.6	92.5
1-3	4.8	6.0
4-7	1.4	0.7
8-10	0.9	0.4
11-14	0.3	0.4
3. Two-week cut-down days		
None	94.7	93.0
1-3	3.2	3.8
4-7	1.3	1.8
8-10	0.2	0.3
11-14	0.7	1.2
4. Two-week doctor visits		
None	84.5	84.4
1-3	14.6	14.6
4-7	0.8	0.9
8-10	0.1	0.1
11-14	0.1	0.1
15+	0.0	0.0
5. 13-month hospital stays		
None	88.2	88.2
1	9.7	9.9
2	1.6	1.2
3	0.2	0.5
4	0.1	0.1
5	0.0	0.1
6+	0.1	0.1

Table 2.A. continued

Percent	Type of Questionnaire: Family/Individual	Type of Questionnaire: Person by Section
6. 12-month doctor visits		
None	26.6	26.5
1	24.7	24.0
2-4	29.5	30.8
5-12	15.5	14.7
13-24	2.3	2.4
25-52	1.2	1.3
53+	0.2	0.3
7. 12-month bed days		
None	50.2	47.4
1-7	38.0	38.1
8-30	7.0	7.7
31-180	1.6	1.7
181+	0.3	0.3
DK/NA/Ref	3.0	4.9
8. Two-week school-loss days		
None	85.8	84.8
1-3	11.1	11.5
4-7	3.1	3.3
8-10	0.0	0.1
11-14	0.0	0.3
9. Health status		
Excellent	38.1	40.2
Very good	27.3	27.1
Good	21.1	19.6
Fair	7.2	6.1
Poor	2.4	2.8
DK/NA/Ref	3.8	4.2

Table 2.B. Mean Levels for Health Characteristics in the NHIS-RDD Feasibility Study by Type of Questionnaire

Characteristic	Type of Questionnaire: Family/Individual	Person by Section	Difference	Standard Error
2-wk. bed days	0.2589	0.2892	-0.0303	0.0346
2-wk. work-loss days	0.2673	0.2221	0.0452	0.0436
2-wk. cut-down days	0.2406	0.3606	-0.1200*	0.0387
2-wk. doctor visits	0.2388	0.2551	-0.0163	0.0200
13-mo. hospital stays	0.1481	0.1481	0.0000	0.0141
12-mo. doctor visits	3.1937	3.4184	-0.2247	0.1761
12-mo. bed days	4.3182	4.5682	-0.2500	0.4970
2-wk school loss days	0.3151	0.4194	-0.1043	0.0693

* Significant difference between questionnaire types at the 5 percent significance level.

Table 3.A. Percent of Persons in Selected Response Categories by Questionnaire and Sex

Response category	Males			Females	
	Family/Individual	Person by Section	Family/Individual		Person by Sectio
	Percent with one or more in the past two weeks				
Bed days	6.7	7.4	8.3		9.2
Work-loss days	6.9	6.2	7.8		9.0
Cut-down days	4.7	5.5	5.8	*	8.5
Doctor visits	13.5	13.6	16.9		17.6
School-loss days	13.9	14.9	13.9		15.7
	Percent with one or more				
12-mo.doctor visits	68.8	68.0	77.4		78.5
13-mo. hospitalizations	10.7	9.9	12.8		13.4
Total conditions	38.6	40.3	41.4	*	46.4
LA conditions	15.6	14.0	14.2		15.8
RA conditions	11.2	11.6	12.3	*	14.3
DV conditions	10.4	11.0	12.7		13.2
CL conditions	22.7	23.3	25.3	*	28.5
Work limitation	8.7	6.4	6.5		6.7
Housework limitation	21.2	17.1	14.6		13.7
Other limitation	2.6	3.9	2.2		2.7
ADL limitation	6.5	7.1	9.0		8.7
IADL limitation	8.0	4.4	9.2		12.2
Play limitation	3.5	1.4	0.0	*	2.8
School limitation	5.9	6.2	4.5		3.4
	Percent having				
No 12-mo.bed days	52.4	51.8	51.1	*	48.3
Excellent health status	42.3	46.2	37.4		38.3

* Significant difference between questionnaire types within sex category at the 5 percent significance level.

Table 3.B. Percent of Persons in Selected Response Categories by Questionnaire [1/] and Age

Response category	AGE 0-16	17-24	25-44	45-64	65+
	Percent with one or more in the past two weeks				
Bed days	9.5(10.5)	7.4(7.9)	7.2(7.1)	6.2(7.6)	6.0(8.3)
Work-loss days	NA	6.9(9.2)	7.8(7.9)	7.3(5.8)	2.9(6.9)
Cut-down days	4.9(6.2)	3.5(2.7)	5.1(6.4)	5.8(7.6)	8.8(15.0)*
Doctor visits	15.4(15.6)	12.7(13.2)	15.6(14.3)	15.2(17.3)	19.9(19.2)
School-loss days	14.3(15.6)	12.2(11.9)	NA	NA	NA
	Percent with one or more				
12-mo.doctor visit	77.8(81.7)*	68.0(68.2)	70.4(73.1)*	73.0(66.0)*	78.9(76.1)
13-mo.-hospitalizations	10.4(19.5)	8.2(11.4)	11.3(11.5)	13.3(10.9)	18.9(17.8)
Total conditions	29.3(32.4)	29.4(34.5)	40.2(41.6)	50.6(53.7)	63.9(67.1)
LA condition	7.2(6.9)	7.0(7.0)	11.5(10.6)	25.5(24.6)	37.0(37.8)
RA condition	15.2(14.8)	9.2(9.1)	10.7(11.8)	11.5(12.4)	11.2(17.8)*
DV condition	10.4(12.2)	8.6(9.6)	12.5(10.8)	12.1(14.0)	16.7(15.6)
CL condition	10.1(10.6)	17.1(21.8)	25.0(26.8)	33.8(36.4)	49.7(46.3)
Work limitation	NA	5.1(2.8)	5.7(5.6)	12.9(10.0)	14.3(11.8)
Housework limitation	NA	2.6(0.0)	9.1(7.8)	21.8(19.0)	24.4(31.4)
Other limitation	NA	1.8(3.2)	2.2(2.6)	2.8(4.5)	7.4(4.2)
ADL limitation	NA	NA	NA	NA	9.0(8.5)
IADL limitation	NA	NA	NA	NA	9.4(9.8)
Play limitation	1.9(2.2)	NA	NA	NA	NA
School limitation	5.7(5.5)	0.0(1.7)	NA	NA	NA
	Percent having				
No 12-mo.bed days	57.1(58.2)	47.9(52.9)	52.4(56.0)	37.3(38.8)	31.8(31.9)
Excellent health status	50.5(56.5)*	48.8(48.5)	42.6(43.0)	24.8(31.2)*	17.9(16.6)

1/ Person by Section entries appear in parentheses.

* Significant difference between questionnaire types within age category at the 5 percent significance level.

Table 3.C. Percent of Persons in Selected Response Categories by Questionnaire [1/] and Education

Response category	Education Level 0-11 years	12 years	13 years or more
Percent with one or more in the past two weeks			
Bed days	9.4(10.2)	6.6(6.1)	5.8(7.0)
Work-loss days	7.0(8.5)	6.8(6.5)	7.0(8.1)
Cut-down days	5.7(8.1)*	6.2(6.1)	4.8(5.6)
Doctor visits	16.3(17.0)	14.4(12.6)	15.1(16.8)
School-loss days	12.7(14.8)	20.0(15.2)	NA
Percent with one or more			
12-mo. doctor visits	76.7(74.5)	69.2(67.6)	74.6(75.7)
13-mo. hospitalizations	12.6(13.4)	12.3(11.5)	10.1(9.4)
Total conditions	38.6(41.2)	43.0(43.4)	40.9(46.5)*
LA conditions	15.1(16.7)	16.0(16.0)	12.6(12.3)
RA conditions	13.6(15.6)	11.1(11.0)	9.4(11.7)
DV conditions	12.6(13.2)	11.8(10.0)	11.0(12.2)
CL conditions	20.3(19.9)	28.0(29.7)	27.2(30.9)
Work limitation	11.6(11.7)	7.9(6.5)	5.7(5.5)
Housework limitation	18.5(21.3)	13.3(13.8)	11.8(7.1)
Other limitation	2.1(2.6)	1.6(3.4)*	3.1(3.8)
ADL limitation	6.9(12.6)	9.5(4.0)	5.0(7.7)
IADL limitation	8.9(15.2)	3.5(2.1)	7.9(11.1)
Play limitation	2.2(1.9)	NA	NA
School limitation	5.2(5.5)	0.0(2.9)	NA
Percent having			
No 12-mo. bed days	51.8(49.3)	42.4(44.8)	50.9(54.6)
Excellent health status	39.2(42.5)	32.1(36.3)	47.5(46.8)

1/ Person by Section entries appear in parentheses.

* Significant difference between questionnaire types within education category at the 5 percent significance level.

Table 4. Demographic Characteristics of Telephone and Personal Interview Samples in Various Health Surveys for Persons Aged 17 and Over

Percent of Individuals Reporting	Telephone (SRC) n=8210	Personal (NHIS) Telephone Households n=18388	Personal (NHIS) Total n=19800	NHIS-RDD Family/ Individual n=2770	NHIS-RDD Person by Section n=2795
Sex					
Male	46.7	45.8	46.3	46.5	46.8
Female	53.1	54.2	53.7	52.5	52.4
DK/NA/Ref	0.2	-	-	1.0	0.8
Age					
17-24	18.3	18.5	19.4	17.0	15.6
25-34	22.9	21.7	22.1	23.9	24.3
35-44	16.8	15.9	15.7	18.6	17.8
45-54	14.9	14.6	14.2	14.2	13.4
55-64	13.4	14.0	13.6	11.2	12.7
65-74	8.3	9.8	9.6	9.2	9.8
75+	4.1	5.4	5.3	4.7	5.3
DK/NA/Ref	1.3	-	-	1.2	1.1
Race					
White	87.5	86.7	85.6	88.5	88.1
Nonwhite	12.5	13.3	14.4	11.5	11.9
Education					
0-8	11.0	13.5	14.3	7.2	8.9
9-12	51.7	53.6	53.9	49.1	46.7
13-18	35.2	31.0	29.8	38.4	37.5
DK/NA/Ref	2.1	1.9	2.0	5.3	6.9
Income					
Less than $5,000	8.6	9.1	10.6	4.7	4.8
$5-9,999	11.8	14.6	15.5	8.0	8.2
$10-14,999	15.0	14.9	14.9	8.7	8.9
$15-24,999	26.1	25.1	24.2	18.6	20.1
$25,000+	20.8	27.6	26.0	32.6	34.6
DK/NA/Ref	17.8	8.8	8.9	27.5	23.5
Marital Status					
Married	65.4	65.2	64.4	67.4	65.2
Widowed	6.7	7.8	7.7	6.4	6.6
Divorced	6.1	5.3	5.5	5.9	7.1
Separated	1.8	1.9	2.2	1.6	1.6
Single	20.1	19.8	20.3	18.7	19.5
Usual Activity					
Working	59.5	58.0	57.7	56.3	56.7
Keeping house	24.0	23.8	23.9	23.0	21.8
Going to school	7.3	7.5	7.4	7.3	5.7
Something else	8.7	10.4	10.3	9.9	12.1
DK/NA/Ref	0.5	0.3	0.3	3.6	3.7

Table 5. Health Characteristics of Telephone and Personal Interview Samples in Various Health Surveys for Persons Aged 17 and Over

Percent of Individuals Reporting	Telephone (SRC) n=8210	Personal (NHIS) Telephone Households n=18388	Personal (NHIS) Total n=19800	NHIS-RDD Family/ Individual n=2770	NHIS-RDD Person by Section n=2795
1. Two-week bed days					
None	91.3	92.3	92.2	93.2	92.5
1-3	6.4	5.0	5.0	4.5	5.2
4-7	1.3	1.4	1.4	1.2	1.2
8-10	0.2	0.4	0.4	0.3	0.2
11-14	0.7	1.0	1.0	0.8	0.9
2. Two-week work-loss days					
None	92.4	95.5	95.5	92.6	92.5
1-3	5.3	3.1	3.1	4.8	6.0
4-7	1.0	0.6	0.6	1.4	0.7
8-10	0.2	0.6	0.6	0.9	0.4
11-14	1.1	0.2	0.2	0.3	0.4
3. Two-week cut-down days					
None	90.2	92.9	93.0	94.5	92.7
1-3	6.8	3.4	3.3	2.9	3.5
4-7	1.5	1.9	1.8	1.5	2.0
8-10	0.2	0.4	0.4	0.2	0.3
11-14	1.3	1.5	1.5	0.9	1.4
4. Two week doctor visits (from person section)					
None	84.1	86.5	86.5	84.4	84.4
1-3	15.1	12.8	12.7	14.4	14.5
4-7	0.6	0.3	0.3	1.0	1.0
9-10	0.1	0.1	0.1	0.0	0.1
11-14	0.1	0.0	0.0	0.1	0.1
15+	0.0	0.2	0.2	0.0	0.0
5. Hospital episodes[1/]					
None	87.0	87.7	87.5	87.7	87.8
1	10.9	9.9	10.0	9.7	10.1
2	1.5	1.8	1.8	1.9	1.3
3	0.4	0.5	0.5	0.3	0.6
4	0.1	0.1	0.1	0.2	0.1
5	0.1	0.1	0.1	0.0	0.1
6+	0.0	0.1	0.1	0.1	0.0

Table 5. continued

Percent of Individuals Reporting	Telephone (SRC) n=8210	Personal (NHIS) Telephone Households n=18388	Personal (NHIS) Total n=19800	NHIS-RDD Family/ Individual n=2770	NHIS-RDD Person by Section n=2795
6. 12 Month doctor visits					
None	26.5	26.5	26.8	28.2	29.3
1	17.9	21.3	21.1	24.1	22.0
2-4	34.3	30.2	29.9	28.1	29.8
5-12	17.0	16.7	16.7	15.4	14.5
13-24	3.0	3.7	3.7	2.7	2.6
25-52	1.1	1.4	1.5	1.3	1.5
53+	0.2	0.3	0.3	0.3	0.3
7. 12 Month bed days					
None	46.0	53.9	53.7	53.1	50.0
1-7	38.0	32.7	32.6	34.4	35.4
8-30	10.7	9.5	9.6	6.7	7.3
181+	0.5	0.4	0.5	0.4	0.3
31-180	2.7	2.8	2.9	2.0	2.1
181+	0.5	0.4	0.5	0.4	0.3
DK/NA/Ref	2.1	0.7	0.8	3.3	4.9
8. Health Status					
Excellent	41.5	44.0	43.3	34.3	35.6
Good	41.7	40.0	40.1	50.0 2/	49.5 2/
Fair	11.9	11.7	12.2	8.8	7.4
Poor	3.8	3.5	3.8	3.2	3.6
DK/NA/Ref	1.1	0.7	0.7	3.8	4.0

1/ 12-month period for SRC and Personal (NHIS); 13-month period for NHIS-RDD

2/ Includes the category, "very good".

Table 6.A. Percent Reporting Selected Health Characteristics for Telephone and Personal Interview Samples [1/] by Sex

Health Characteristic	Sex	Personal (NHIS) Telephone Households	Telephone (SRC)	NHIS-RDD Family/Individual	NHIS-RDD Person by Section
At least One 2-wk. Work-Loss Day	Male	5.1	8.1	6.9	6.2
	Female	4.2	7.2	7.8	9.0
At Least One 2-wk. Doctor Visit	Male	11.3	13.6	11.6	12.6
	Female	15.4	17.9	18.6	18.4
At Least One 12-Mo. Bed Day	Male	42.3	50.3	42.8	44.3
	Female	49.4	55.6	47.3	49.9

1/ All samples exclude persons aged 16 or less.

Table 6.B. Percent Reporting Selected Health Characteristics for Telephone and Personal Interview Samples by Age

Health Characteristic	Age	Personal (NHIS) Telephone Households	Telephone (SRC)	NHIS-RDD Family/Individual	NHIS-RDD Person by Section
At Least One 2-Wk. Work-Loss Day	17-24	6.0	9.2	6.9	9.2
	25-44	5.4	8.5	7.8	7.9
	45-64	4.4	7.2	7.3	5.8
	65+	1.0	3.1	2.9	6.9
At Least One 2-Wk. Cut-Down Day	17-24	5.8	8.8	3.5	2.7
	25-44	6.4	10.3	5.1	6.4
	45-64	7.8	9.4	5.8	7.6
	65+	9.0	10.7	8.8	15.0
At Least One 2-Wk. Doctor Visit	17-24	11.8	15.0	12.7	13.2
	25-44	12.0	13.3	15.6	14.3
	45-64	14.1	17.3	15.2	17.3
	65+	17.3	28.2	19.9	19.2
At Least One 12-Mo. Bed-Day	17-24	50.9	62.4	47.9	52.9
	25-44	51.3	59.3	52.4	56.0
	45-64	40.4	45.6	37.3	38.8
	65+	36.6	37.1	31.8	31.9
Excellent Health Status	17-24	51.2	51.8	48.8	48.5
	25-44	51.6	47.0	42.6	43.0
	45-64	37.6	34.9	24.8	31.2
	65+	30.3	29.1	17.9	16.6

Table 6.C. Percent Reporting Selected Health Characteristics for Telephone and Personal Interview Samples 1/ by Education

Health Characteristic	Education in Years	Personal (NHIS) Telephone Households	Telephone (SRC)	NHIS-RDD Family/Individual	NHIS-RDD Person by Section
At Least One 2-Wk. Work-Loss Day	0-11	3.7	8.1	9.7	8.5
	12	4.8	7.6	6.8	6.5
	13+	5.1	7.2	7.0	8.1
At Least One 2-Wk. Cut-Down Day	0-11	8.1	10.7	5.8	11.9
	12	6.3	8.7	6.2	6.1
	13+	6.9	10.2	4.8	5.7
At Least One 12-Mo. Bed Day	0-11	41.3	45.9	40.6	38.3
	12	44.3	53.6	42.4	44.8
	13+	51.3	58.7	50.9	54.6
Excellent Health Status	0-11	30.1	29.1	22.5	20.2
	12	45.4	41.3	32.1	36.4
	13+	57.2	53.6	47.5	46.7

1/ All samples exclude persons aged 16 or less.

APPENDIX

For $i = 1, 2, \ldots, 8; j=2, 3;$ and $k = 1, 2, \ldots, 260$ let

n_{ijk} = total number of persons that had a valid response for question i on questionnaire version j in cluster k, and

x_{ijk} = aggregate value of the valid responses for question i on questionnaire version j in cluster k.

An estimate of the mean value for question i on question version j is given by

$$(1) \quad \hat{R}_{ij} = \frac{\sum_{k=1}^{260} x_{ijk}}{\sum_{k=1}^{260} n_{ijk}} .$$

For each i and j, let

l_{ij} = number of n_{ijk}'s not equal to zero, and let

$l_i = \min (l_{i2}, l_{i3})$

An estimate of the variance of the difference between the mean value on questionnaire version 2 and that on questionnaire version 3 for question i (see Table 2.B.) is given by

$$(2) \quad \text{var } (\hat{R}_{i2} - \hat{R}_{i3}) = \text{var } (\hat{R}_{i2}) + \text{var } (\hat{R}_{i3}) - 2 \text{ cov } (\hat{R}_{i2}, \hat{R}_{i3}),$$

where var $(\hat{R}_{ij}) =$

$$\frac{1}{l_{ij}(l_{ij}-1)\bar{n}_{ij}^2} \left[\sum_{k=1}^{260} x_{ijk} - 2\hat{R}_{ij} \sum_{k=1}^{260} x_{ijk} n_{ijk} + \hat{R}_{ij}^2 \sum_{k=1}^{260} n_{ijk}^2 \right] .$$

$$\operatorname{cov}(\hat{R}_{i2}, \hat{R}_{i3}) = \frac{1}{l_i(l_i-1)\bar{n}_{i2}\bar{n}_{i3}} \left[\sum_{k=1}^{260} x_{i2k}x_{i3k} - R_{i2}\sum_{k=1}^{260} x_{i3k}\, n_{i2k} \right.$$

$$\left. -R_{i3}\sum_{k=1}^{260} x_{i2k}n_{i3k} + R_{i2}R_{i3}\sum_{k=1}^{260} n_{i2k}n_{i3k} \right],$$

and $$\bar{n}_{ij} = \frac{\sum_{k=1}^{260} n_{ijk}}{l_{ij}}.$$

An estimate of the variance of the difference between the percent of persons in selected response categories by questionnaire and one demographic characteristic (see Tables 3.A, 3.B, and 3.C) can be obtained using the same formulas as above with an additional subscript added to each variable to represent the demographic characteristic. Under these circumstances,

n_{ijlk} = total number of persons of characteristic l (l=1 or 2) that had a valid response for question i on questionnaire version j in cluster k, and

x_{ijlk} = total number of persons of characteristic l in the appropriate percent category for question i on questionnaire version j in cluster k.

Appendix 4. Respondent Rules

Attachments

1. Screening Questions for MKR
2. Post Survey Questions About MKR
3. Respondent Assessment of Accuracy of Reported Information
4. Analysis Plan for the MKR Rule
5. Analysis of Screening Questions for MKR
6. Analysis of Post Survey Questions About MKR

Attachment 1

Screening Questions for Most Knowledgeable Respondent

5a. This survey is being conducted to collect information on the nation's health. It is very important to have good answers to the health questions I will be asking. For that reason, I would like to speak to someone in the household who is at least 19 years old and knows the MOST about the health of the people in this family. Are you the most knowledgeable person?

☐ Yes(6) ☐ No(5b) ☐ DK(5b)

---

b. May I speak to someone at least 19 years old and who knows the MOST about the health of people in the family?

☐ Most knowledgeable respondent available(5c)

☐ Most knowledgeable respondent not available (ARRANGE CALLBACK)

---

c. Hello, I'm (name) from the United States Bureau of the Census in Washington, D.C. We are conducting a survey for the U.S. Public Health Service to collect information on the nation's health. I was told that you would know the MOST about the health of the people in the family.

(READ 6a)

5a. This survey collects information on the nation's health. I would like to speak to someone in the household who is at least 19 years old and knows the MOST about the health of the people in this family. Are you the most knowledgeable person?

☐ Yes(6) ☐ No(5b) ☐ DK(5b)

---

b. May I speak to someone at least 19 years old and who knows the MOST about the health of people in the family?

☐ Most knowledgeable respondent available(5c)

☐ Most knowledgeable respondent not available (ARRANGE CALLBACK)

---

c. Hello, I'm (name) from the United States Bureau of the Census in Washington, D.C. We are conducting a health survey for the United States Public Health Service. I was told that you would know the MOST about the health of the people in the family.

(READ 6a)

AA	*Mark first appropriate box.*	PGM 4 ↓ 1 ☐ Only one eligible respondent in family *(Go to question 3 on Household Page of cover booklet)* 2 ☐ Other *(1)*
1.	**Now that you have heard the type of questions we ask in a health study, do you feel YOU ARE the person in your family who knows the most about the health of the family members?**	1 ☐ Yes *(2a)* 2 ☐ No *(3a)*
2a.	**Is there anyone else in the family who would know EQUALLY as much about the health of the family members?**	1 ☐ Yes *(2b)* 2 ☐ No *(Go to question 3 on Household Page of cover booklet)*
b.	**Who would that person be?** *Refer to household composition and enter person number(s) in 2-digit numerals.*	Person number(s) { ☐☐ ☐☐ ☐☐
c.	**Anyone else?**	☐ Yes *(Reask 2b and c)* ☐ No *(Go to question 3 on Household Page of cover booklet)*
3a.	**Is there anyone in the family who would know MORE about the health of the family members?**	1 ☐ Yes *(3b)* 2 ☐ No *(Go to question 3 on Household Page of cover booklet)*
b.	**Who would that person be?** *Refer to household composition and enter person number(s) in 2-digit numerals.*	Person number(s) { ☐☐ ☐☐ ☐☐
c.	**Anyone else?**	☐ Yes *(Reask 3b and c)* ☐ No *(Go to question 3 on Household Page of cover booklet)*

FOOTNOTES

Post Survey Questions
About Most Knowledgeable
Respondent

(Some people provide very accurate health information while others may not be as sure about their answers.)

8. **Overall, about how accurate do you think your answers to the questions about — — are — very accurate, fairly accurate, or not very accurate?**

8.
1 ☐ Very accurate
2 ☐ Fairly accurate
3 ☐ Not very accurate

GO TO AA ON BACK OF COVER BOOKLET

FOOTNOTES

Respondent Assessment of Accuracy of Reported Information

Attachment 3

J. T. Massey
November 14, 1983

Analysis Plan for the Most Knowledgeable Respondent Rule

I. Identification of most knowledgeable respondent

1. Percent of households where one person was identified as most knowledgeable.

2. Percent of households where two or more persons were identified as equally knowledgeable.

3. Percent of households where first adult phone answerer was identified as most knowledgeable.

4. Percent of households where most knowledgeable respondent could not be identified.

These percentages should be computed by replicate and household size.

II. Callbacks required to reach most knowledgeable respondent(MKR) and effect on response rate

1. Percent of households where MKR was <u>not</u> first adult contacted.

The following statistics are subsets of II. 1.

1.1 Percent of households where MKR was at home at time of first contact.

1.2 Percent of households where MKR came to the telephone when at home.

1.3 Percent of households where additional contacts were required to speak to MKR.

1.4 Percent of households where adult phone answerer asked that no callbacks be made to MKR.

1.5 Percent of MKR reached by Xth contact and percent never reached when callbacks were required.

1.6 Percent of households with "proxy" interview when MKR was never reached.

2. Percent of all MKR interviewed.

These statistics should be computed for each replicate.

III. Post survey analysis of MKR

1. Percent of households where MKR still felt he/she was most knowldgeable after interview.

2. Percent of households where another household member was identified as MKR after interview and their relationship to person interviewed.

3. Percent of households with equally knowledgeable respondent(s) after interview and their relationship to person interviewed.

4. Accuracy of reporting by MKR and other household respondents. Complete the following table for MKR's and other household respondents.

Household Member	Accuracy of response for MKR		
	Very	Fairly	Not Very
Self			
Spouse			
Siblings			
Other adults			
Children			

Analysis of Screening Questions for MKR

Attachment D: Most Knowledgeable Respondent Tabulations [1]

Note 1: Substitute cases are omitted from all tabulations.

Note 2: A brief explanation on interpreting Table 3 is required. A most knowledgeable respondent (MKR) indicator is set for each call attempt to a household. This indicator identifies whether the MKR answered the phone, was called to the phone, was not available, etc.. Table 3 is then a crosstabulation of final outcome codes by the lowest category obtained for the MKR indicator during all call attempts. The categories are arranged from lowest at the left of the table to highest at the right. Several anomalies are possible as it is often difficult to determine the most critical MKR indicator for a case. For example, during one call, the MKR may answer the phone and respond to numerous questions before stating he/she cannot continue. A callback is arranged but after several attempts the MKR cannot be contacted and the interview is completed by another respondent. In this instance, although "phone answerer 19+ and MKR" is the lowest category obtained, it may or may not be the most critical in determining the final outcome code.

[1] From June 20, 1984 memorandum prepared by Anthony M. Roman of the Demographic Surveys Division of the Bureau of the Census.

Attachment D:

Table 1: Distribution of Final Outcome Codes (Cumulative through Replicate 12)

Outcome	Number of cases
Fully complete interview with MKR	2140
Fully complete interview with other respondent	68
Complete interview through Section H with MKR	42
Complete interview through Section H with other respondent	1
Partial interview with MKR	35
Partial interview with other respondent	7
Ineligible residence	29
Noninterview	36
Refusal	370
Undetermined	258
Total	2986

Table 2: Final Outcome Codes by Number of Attempts to Reach MKR (Cumulative through Replicate 12)

Final Outcome Code	Number of Callbacks Required in Attempting to Reach MKR				
	0[1/]	1	2	3	4+
Fully Complete Interview with MKR	2123	9	3	2	3
Fully Complete Interview with other respondent	67	0	0	0	1
Complete interview through Section H with MKR	23	10	2	3	4
Complete interview through Section H with other respondent	0	0	0	1	0
Partial interview with MKR	22	5	1	3	4
Partial interview with other respondent	5	1	0	0	1
Ineligible residence	18	3	3	5	0
Noninterview	28	6	0	2	0
Refusal	282	56	14	7	11
Undetermined	192	24	10	7	25
Total	2760	114	33	30	49

1/ In this instance, the MKR was reached on the first household contact resulting in at least item 5 of the survey introduction being answered

Attachment D:

Table 3: Final Outcome Codes by Lowest Category Obtained for MKR Indicator (Cumulative through Replicate 12)

Final Outcome Code	Lowest Category Obtained for MKR Indicator from All Calls to a Household						
	Phone Answerer 19+ and MKR	MKR called to phone	MKR not home or available[1]	No eligible respondents at home[1]	Don't know[1]	Refusal[1]	Blank
Fully Complete Interview with MKR	1995	62	0	0	1	0	82
Fully Complete Interview with other respondent	59	5	0	0	0	0	4
Complete interview through Section H with MKR	41	0	0	0	0	0	1
Complete interview through Section H with other respondent	0	0	0	0	0	0	1
Partial interview with MKR	32	2	0	0	0	0	1
Partial interview with other respondent	6	1	0	0	0	0	0
Ineligible residence	5	0	0	0	0	0	24
Noninterview	3	0	0	0	0	0	33
Refusal	161	2	0	0	1	0	206
Undetermined	13	0	0	0	0	0	245
Total	2315	72	0	0	2	0	597

[1] These outcomes did occur, but not on final outcome dialing except in 1 case.

Attachment 6

Analysis of Post Survey Questions About MKR

Draft

July , 1984

MEMORANDUM FOR The Record

From: William Mockovak (Census Bureau)
Jim Massey (NCHS)

Subject: Respondent Rules for Identifying the Most Knowledgeable Respondent (MKR)

Overview

In the Telephone Health Interview Survey (THIS), an attempt was made to identify the person in a family who knew the most about the health of family members. This person, called the most knowledgeable respondent (MKR), was then asked to provide both health and demographic data about persons in the household. Procedures for locating the most knowledgeable respondent were critical since, unlike the face-to-face interview, the MKR provided proxy information as a routine procedure, rather than when other household members might not be available to join the interview.

Previous analyses 1/ have indicated that out of 2,133 partial or complete interviews, only 65 were completed with someone other than the most knowledgeable respondent. But, these figures were based on entries made by the interviewer at the start of the interview before any health questions had been asked. The data reported in this memorandum came from a series of questions asked toward the end of the interview, after all but the income question, questions about telephones in the residence, and special place questions, had been asked. The results of these questions are summarized in subsequent sections of this report.

1/ See reports prepared by Tony Roman.

Summary of Findings

Following is a brief synopsis of the results. More detailed results are presented in later sections.

1. An estimated 8.6 percent of respondents (in households with more than one eligible respondent) felt that they were not the most knowledgeable respondent.

2. An estimated 31.2 percent of the interviewed households had only one eligible respondent.

3. In those cases where the respondent felt that s(he) did not know the most about the health of family members, 81.2 percent of the respondents reported that there was someone more knowledgeable. This figure translates to 4.5 percent of the 1,823 cases in which there was more than one eligible respondent.

4. An estimated 67.0 percent of respondents who felt that they knew the most about the health of family members, also felt that there were others in the family who knew equally as much.

Did the Respondent Feel That She/He Was the Most Knowledgeable Respondent?

As mentioned previously, the questions that follow were asked at the end of the telephone interview. Table 1 shows that 31.2 percent of the cases had only one eligible respondent in the family. Therefore, no further questions were asked in these cases.

Table 1
Number of Eligible Respondents in the Family

	Absolute Frequency	Relative Frequency	Adjusted Frequency
Only one	826	29.3%	31.2%
More than one	1823	64.7%	68.8%
Not answered	168	6.0%	Missing
Total	2817	100.0%	100.0%

Those respondents with more than one eligible person in the household (1,823 cases) were then asked if they felt they knew the most about the health of family members. Of those responding, about 8.6 percent felt that they were not the most knowledgeable respondent. These results are shown in Table 2. If all interviewed cases, including those with only one eligible respondent are considered, the figure of 8.6 percent drops to 5.5 percent. These estimates are close to the 8.6 percent of the cases 2/ in which monitors reported that the interviewer encountered difficulties identifying the most knowledgeable respondent.

Table 2
Percent of Respondents Reporting That
They Were or Were Not the Most Knowledgeable

Question: Now that you have heard the type of questions we ask in a health study, do you feel you are the person in your family who knows the most about the health of the family members?

	Absolute Frequency	Relative Frequency	Adjusted Frequency
Yes	1637	89.8%	91.3%
No	154	8.4%	8.6%
Don't Know	2	0.1%	0.1%
Missing	30	1.6%	Missing
Total	1823	100.0%	100.0%

In those cases where the respondent felt that s(he) did not know the most about the health of family members, 81.2 percent (82 cases) of these responding reported that there was someone in the household more knowledgeable. These 82 cases were 4.5 percent of the initial 1823 cases in which there was more than one eligible respondent. Table 3 presents these results.

2/ See memorandum "Analysis of Monitoring Data" by Mockovak, Fitti, and Frey.

Table 3
Percent of Respondents Reporting That
There Was or Was Not a More Knowledgeable Respondent

3a

Question: Is there anyone in the family who would know more about the health of the family members?

	Absolute Frequency	Relative Frequency	Adjusted Frequency
Yes	82	53.2%	81.2%
No	19	12.3%	18.8%
Missing	53	34.4%	Missing
Total	154	100.0%	100.0%

Table 4 presents the frequencies with which different person numbers from the questionnaire were entered when the respondent felt another family member know more about the health of the family. A respondent could have mentioned more than one person who knew more about the health of family members.

Table 4
Persons Identified As More
Knowledgeable About the Health of Family Members

Person No. On the Questionnaire	Frequency
1	23
2	59
3	4
4	1
Total	87

When a respondent reported that s(he) was the most knowledgeable, a follow-up question asked if anyone else in the family knew equally as much about the health of family members. As Table 5 shows, about 67 percent of these respondents felt that there were other, equally knowledgeable respondents in the family, Table 6 presents the frequencies with which different person numbers from the questionnaire were entered when the respondent was asked to identify other, equally knowledgeable persons.

Table 5
Percent of Respondents Reporting That There Were Other, Equally Knowledgeable Respondents

	Absolute Frequency	Relative Frequency	Adjusted Frequency
Yes	1084	66.2%	67.0%
No	534	32.6%	33.0%
Missing	19	1.2%	Missing
Total	1637	100.0%	100.0%

Table 6
Persons Identified As Equally Knowledgeable About the Health of Family Members

Person No. On the Questionnaire	Frequency
1	600
2	471
3	65
4	19
5	2
6	4
Missing	17
Total	1178

Appendix 5. Interview Period/Sampling Frequency

I. *Introduction*

The sampling plan for the NHIS-RDD feasibility study is described in NHIS/RDD Development Memorandum No. 4, 9/23/83. The sample was selected in 12 replicates. One replicate was introduced each week for 12 consecutive weeks. Each replicate was then interviewed over a three week period. During any given week of the survey, three replicates were being interviewed. One replicate was in its first week of interview, one in its second, and one in its third.

There are many issues that could be studied regarding this sampling plan. After examining much data regarding the sampling procedures, we narrowed our analysis to three questions which appear to be most important and interesting and which could be addressed with the data available. These are:

1. How many cases were unresolved after three weeks of interview? Is there any evidence that the distribution of calls over the interview period affects the number of unresolved cases?

2. Were interviewer workloads evenly distributed throughout the survey? That is, were we successful at keeping the level of work constant in the facility across weeks.

3. How would response rates be affected if the interview period had been two weeks or four weeks instead of three weeks? Could response rates be improved by stopping the generation of replacements during the last week of the survey?

These issues will be discussed separately in the following sections. Some parts of the analysis discuss reps 1-6 and 7-12 separately. This was necessary because survey procedures differed between these two groups. Replicates 1-6 scheduled calls to cases using both the automated call scheduler and hand scheduling by supervisors. Reps 7-12 relied exclusively on the call scheduler. Therefore, some differences may be expected between these two parts.

Our analyses were done completely without the benefit of sampling error estimates. This somewhat limits what we are able to say about some of the observed differences in calling patterns, unresolved cases, etc. However, these preliminary analyses will identify areas where fuller investigation is needed using more sophisticated statistical methods.

II. *Unresolved Cases and Calling Pattern*

One way of evaluating the survey design is to examine the number of cases which were unresolved[1] at the end of each replicate. The upper portion of Figure 1 shows the number of unresolved cases by replicate. (The last three columns show the average number of Reps 1-6, 7-12 and all Reps.) They range

[1] Unresolved cases are those cases whose residential status has not been determined by the end of the third week of interviewing or by the time the maximum number of calls has been made to the case.

from eight unresolveds in Rep 11 up to 32 in Rep 4. Since there are about 400 cases in each replicate (including replacements but excluding substitutes), having only eight cases (2 percent) unresolved seems like a very good result while having 32 (8 percent) unresolveds is undesirable. Another survey organization has reported averaging 3 to 4 percent unresolved cases.

We would like to determine which features of the survey operations contribute to the final number of unresolved cases. One feature which should have a major effect on this number is when and how often calls are made to a case. The lower half of Figure 1 shows for each replicate, the average number of calls made to each active case during each week of the survey. If there is one calling pattern associated with replicates which have low numbers of unresolved cases or another pattern associated with high numbers of unresolveds, then we know what type of pattern to strive for or avoid in future telephone surveys.

Examining Figure 1 shows that there are two predominant calling patterns. Reps 2, 4, 5, 6 and 7 show a pattern where the fewest number of calls were made to cases during the second week of the replicate. These five replicates averaged 25.6 unresolved cases. Reps 3, 9, 10, 11 and 12 show a pattern where the number of calls per case increases from week one to week two and from week two to week three. These five reps averaged 19.6 unresolved cases. Of the other two replicates, Rep 8 is similar to the second pattern except that the number of calls in week three decreases slightly rather than increasing. This replicate had 16 unresolved cases. Rep 1 had 17 unresolved cases, but it is probably a special case since the survey was first getting started.

These patterns also hold up for the average of Reps 1-6 and 7-12. The first six reps show the pattern of fewest calls during week two of the replicate. They average about 26 unresolved cases. The last six reps show the pattern of increasing calls through the replicate. These replicates average about 18 unresolved cases. Thus it seems that we should strive for a calling pattern which results in increasing numbers of calls to each case through the three weeks rather than a pattern which results in a drop off of calls during the second week. Staffing levels in the facility should be adjusted to allow a sufficient number of calls to be placed to each case in each replicate. The later replicates in Figure 1 show that the desired calling pattern can be attained given adequate staffing levels.

It is likely that other factors such as respondents, interviewers, seasons of the year, and the telephone system also contribute to the problem of unresolved cases. Thus, looking only at the average number of calls per case and staffing may only provide a partial picture of the problem.

III. *Workloads*

One feature of the survey design was the introduction of a new replicate every week. This was done so that the interviewer workloads would be stable across weeks. If, for example, replicates had been introduced every two weeks, we would expect workloads to alternate between heavy loads the week a replicate is introduced and light loads the following week.

In order to study the workloads, we looked at the number of calls, contacts and completions by week of the survey. This data is presented in Figure 2. In this plot, weeks 3, 13 and 14 are outliers. Week 3 contained a holiday and weeks 13 and 14 were the conclusion of the survey where we were no longer interviewing three replicates each week.

In examining this plot, it is helpful to put the data into two groups - weeks 1-6 and weeks 7-14. During weeks 1-6 the calls were scheduled using both the automated call scheduler and hand scheduling. From week 7 on, the automated call scheduler was used exclusively.

The second part of the survey showed very consistent numbers of calls except for week 9. (We have no explanation for the high number of calls that week.) The number of calls during the first part of the survey were more variable, but still fairly consistent. The average number of calls per week may be slightly higher during the second part of the survey, but overall, we feel that the number of calls per week is steady across the week of the survey.

Looking at the numbers of contacts and completions by week of the survey, similar patterns are exhibited as for the number of calls by weeks. The main difference is that the number of contacts during the second part of the survey is consistently higher than during the first part of the survey. But the number of contacts and completions are consistent within both parts of the survey. Therefore, these data indicate that our goal of maintaining constant workloads during the survey was met using this sample design.

IV. *Interview Period*

Another question to be answered is how the length of the interview period affects the response rates. Figure 3 shows the average number of cases resolved[2] on each day of the replicate. The number declines steadily through the first week, but then levels off to eight or nine resolutions per day through the last two weeks. There seems to be no decline in the number of resolutions during the later part of the survey period. This indicates that shortening the interview to two weeks may result in substantially fewer completions, while extending the period to four or more week may result in substantially more completions. (This could result in an increase of two to four percent in the response rate.) Of course, it is difficult to assess the real impact of these changes since staffing levels, calling patterns, and call scheduling would accompany any change in interview period length.

The extra completions predicted for a four week period would come only with a substantial increase in effort. Figure 4 shows the average number of calls to each active case by day of the replicate. There is a definite upward trend from day eight to the end of the replicate. If this trend is extrapolated out through a fourth week, there would be a large number of calls to each case each day. This would result in substantial cost increases. Thus the decision on the length of the interview period appears to be primarily a cost consideration. If response rates are our main concern, then the response rates could be increased by extending the period, but this would require more effort and cost.

---

[2] A resolved case can be a completed interview, partial interview or noninterview.

replicates to the last six replicates, we can get an idea of the effectiveness of the automated call scheduler. However, we must be cautious in this comparison. Some differences between the first and second parts of the survey will be due to improved interviewer performance as they learn and an increase in the maximum number of calls from 15 to 20. However, some of the differences are likely due to the procedure change which resulted in exclusive use of the call scheduler.

There were fewer unresolved cases during the second half of the survey. As stated earlier, the first six replications averaged about 26 unresolved cases while the last six averaged 18 cases. This difference could be due just as much to interviewer learning or the increase in the number of calls to a case as to the use of the automated call scheduler.

As noted in section III, the use of the automated call scheduler also coincided with an increase in the number of calls, contacts and completions made by week. It seems to have the largest impact on the number of contacts made. However, we get a different picture if we look at contacts and completions per call. Figure 6 shows, for each week, the number of contacts divided by the number of calls and the number of completions divided by the number of calls. We can see that these contact and completion rates do not increase using the automated call scheduler. Thus, the scheduler helps increase the number of calls and ultimately the number of completions, but does not increase the probability that a given call will result in a contact or completion.

VI. *Conclusion*

The data examined in this analysis has helped to answer the questions posed in the introduction of this report.

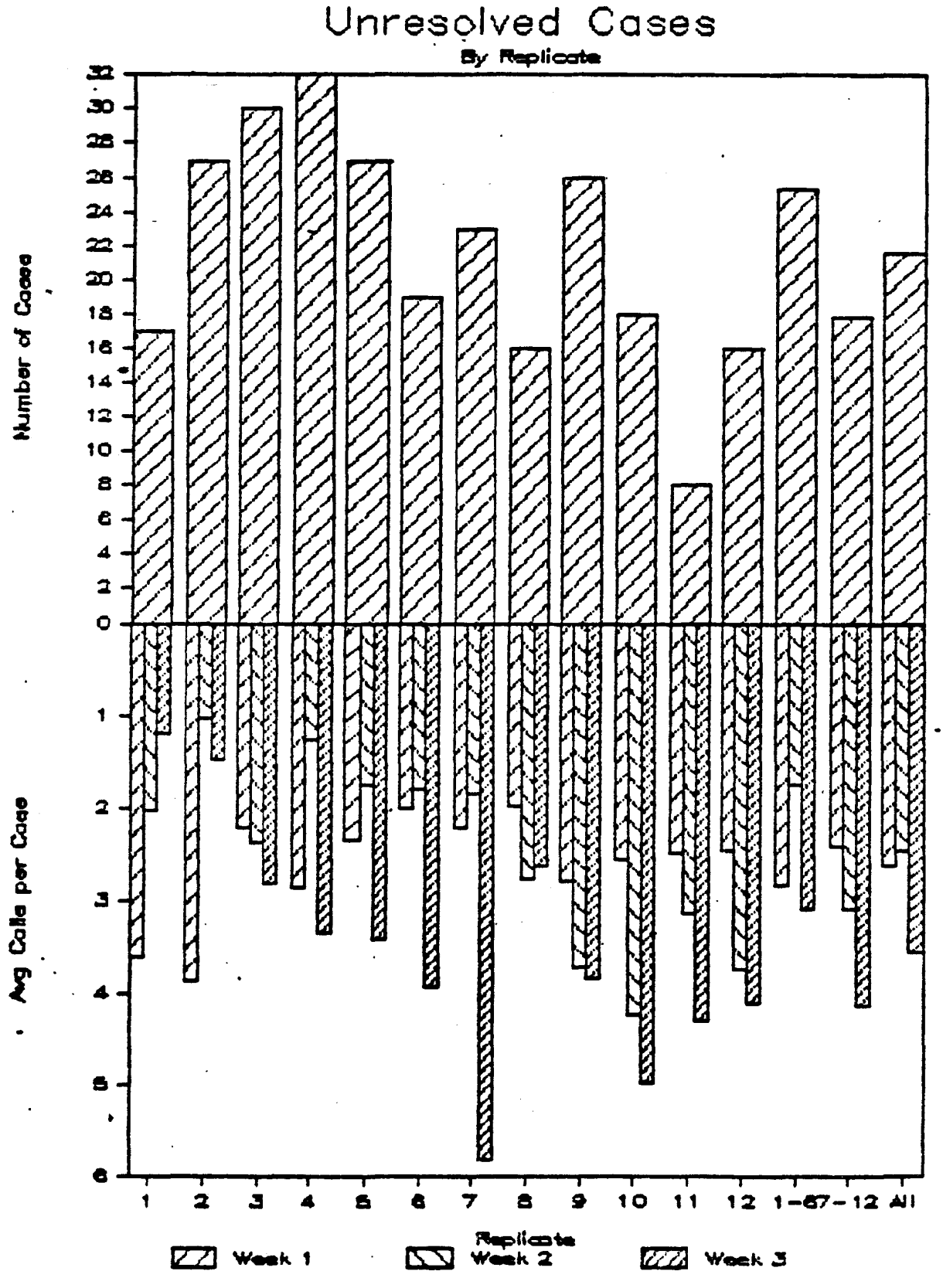
1. The number of unresolved cases varied from eight up to 32. It is difficult to discern which calling patterns are most successful at reducing the number of unresolved cases. However, the data do indicate that a pattern of increasing levels of effort across the weeks of the interview period resulted in low numbers of unresolveds. It appears that this pattern was achieved in this study by using the automated call scheduler and a constant facility staff level. Thus, this pattern is recommended for future surveys to reduce the number of unresolved cases.

2. The interviewer workloads were fairly well distributed during the survey. The number of calls, contacts and completions were all stable during the survey period.

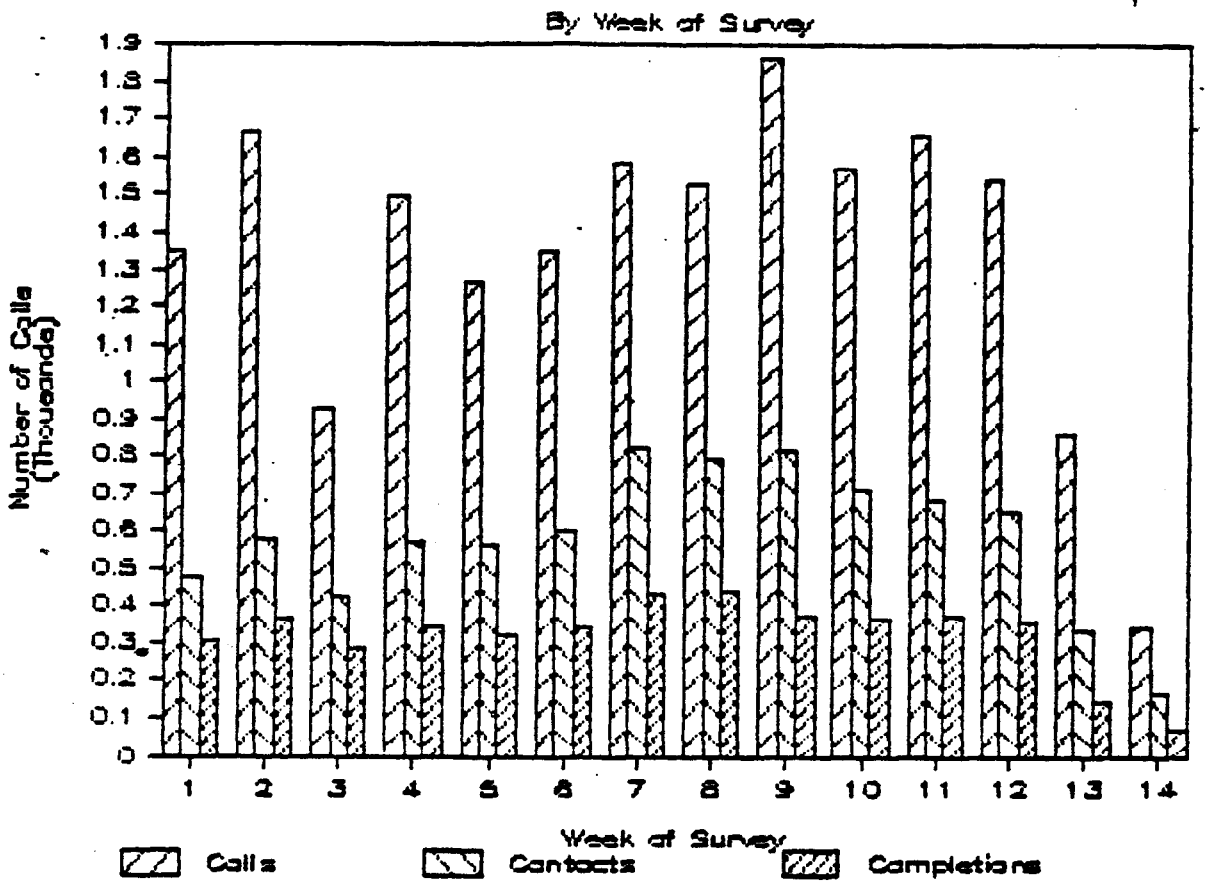
3. Response rates could be increased if the survey period was increased to four or more weeks. However, the resulting cost increases may not be worthwhile. Further investigation is needed here. Also, the cutting off of replacements during the last week of the survey would have a negligible effect on response rates.

The data files from the NHIS-RDD feasibility study are a rich source of information about random digit dialing telephone surveys. The issues we examined are just a few of the many issues which could have been examined. For example, in answering question 1, we looked at the number of calls by week of the survey. This questions could also be approached by looking at the number of calls by time of day. Thus, while we are confident that the conclusions drawn here are valid, there is also data available to support more research.

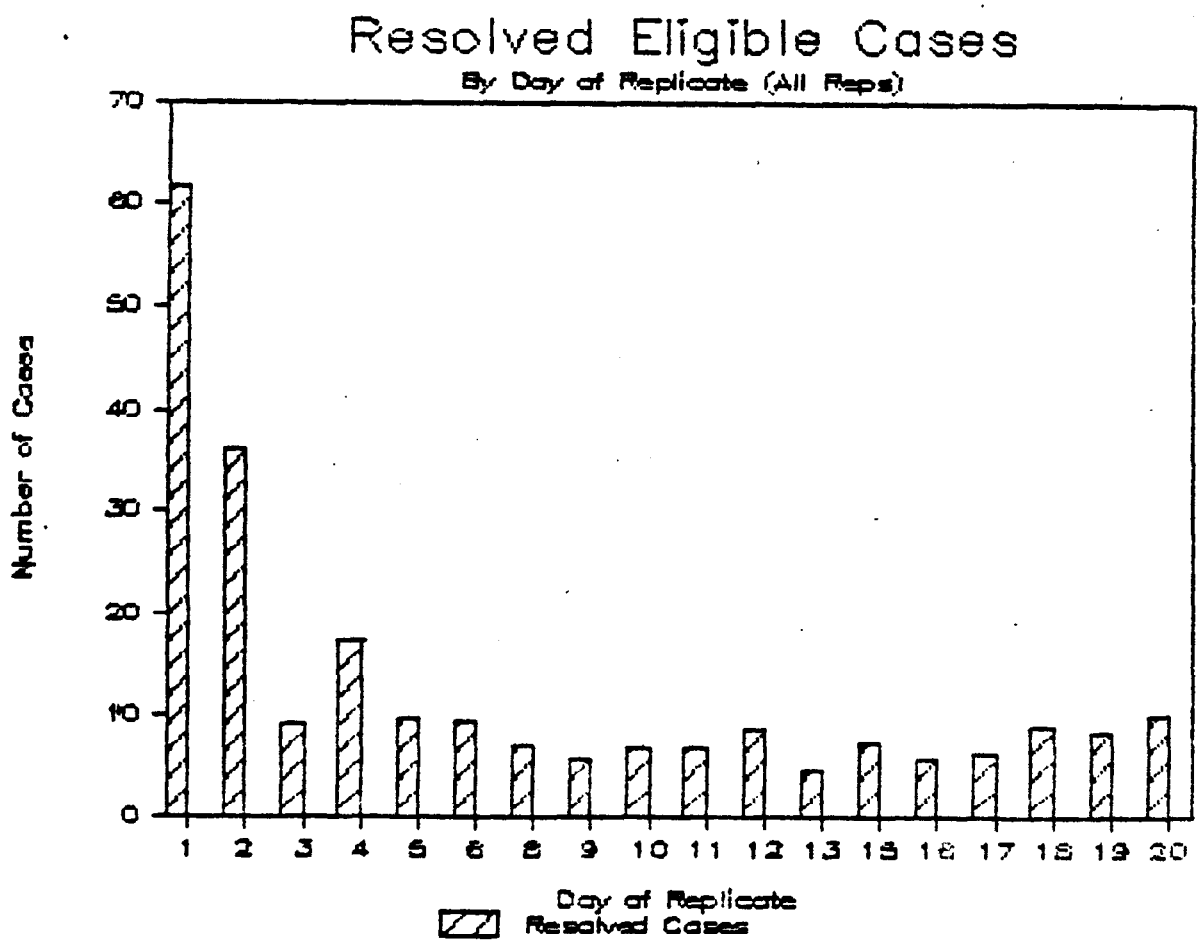
Figure 1


Unresolved Cases
By Replicate
Number of Cases
Avg Calls per Case
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
1
2
3
4
5
6
1
2
3
4
5
6
7
8
9
10
11
12
1-6
7-12
All
Replicate
Week 1
Week 2
Week 3

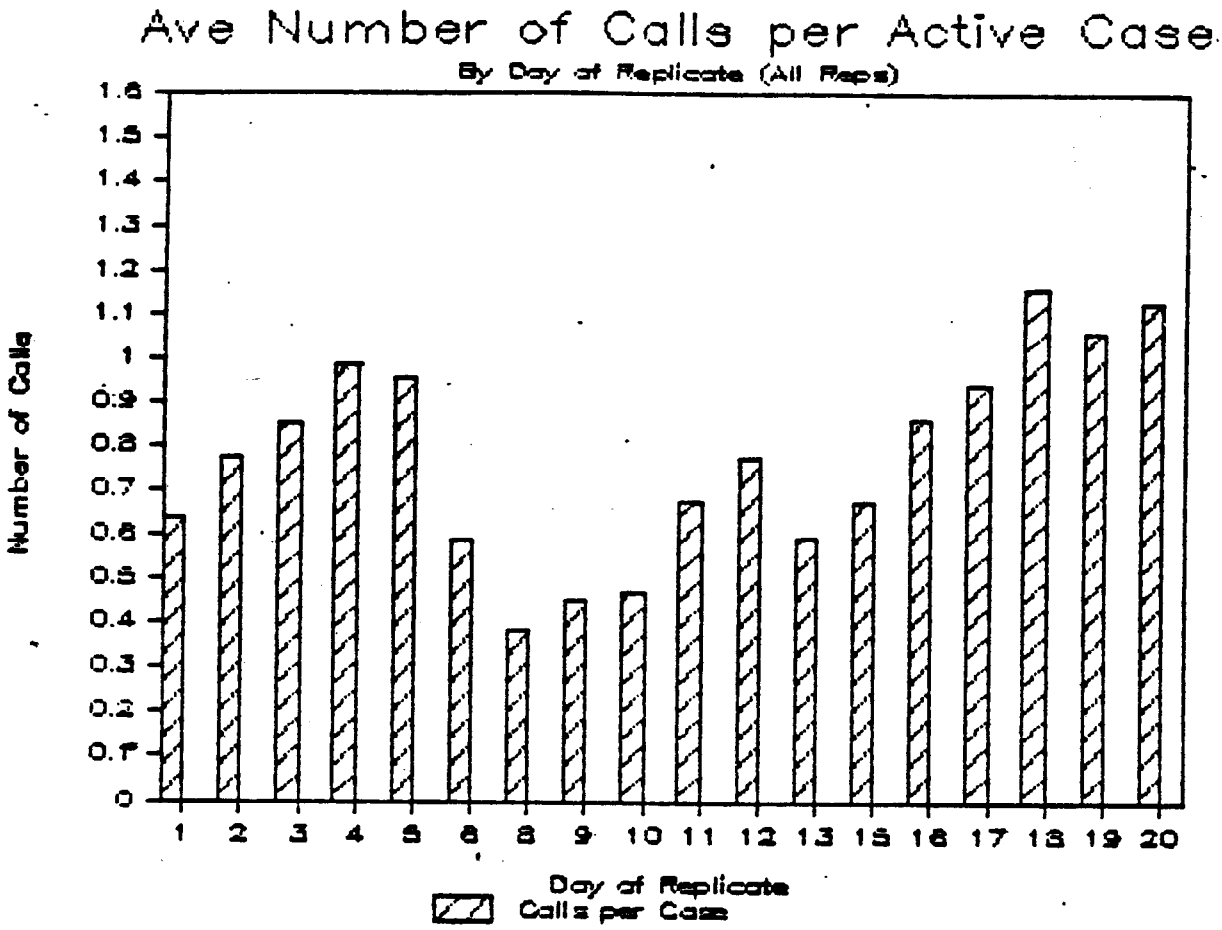
gure 2


Number of Calls, Contacts and Completes
By Week of Survey
Number of Calls (Thousands)
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Week of Survey
Calls
Contacts
Completions

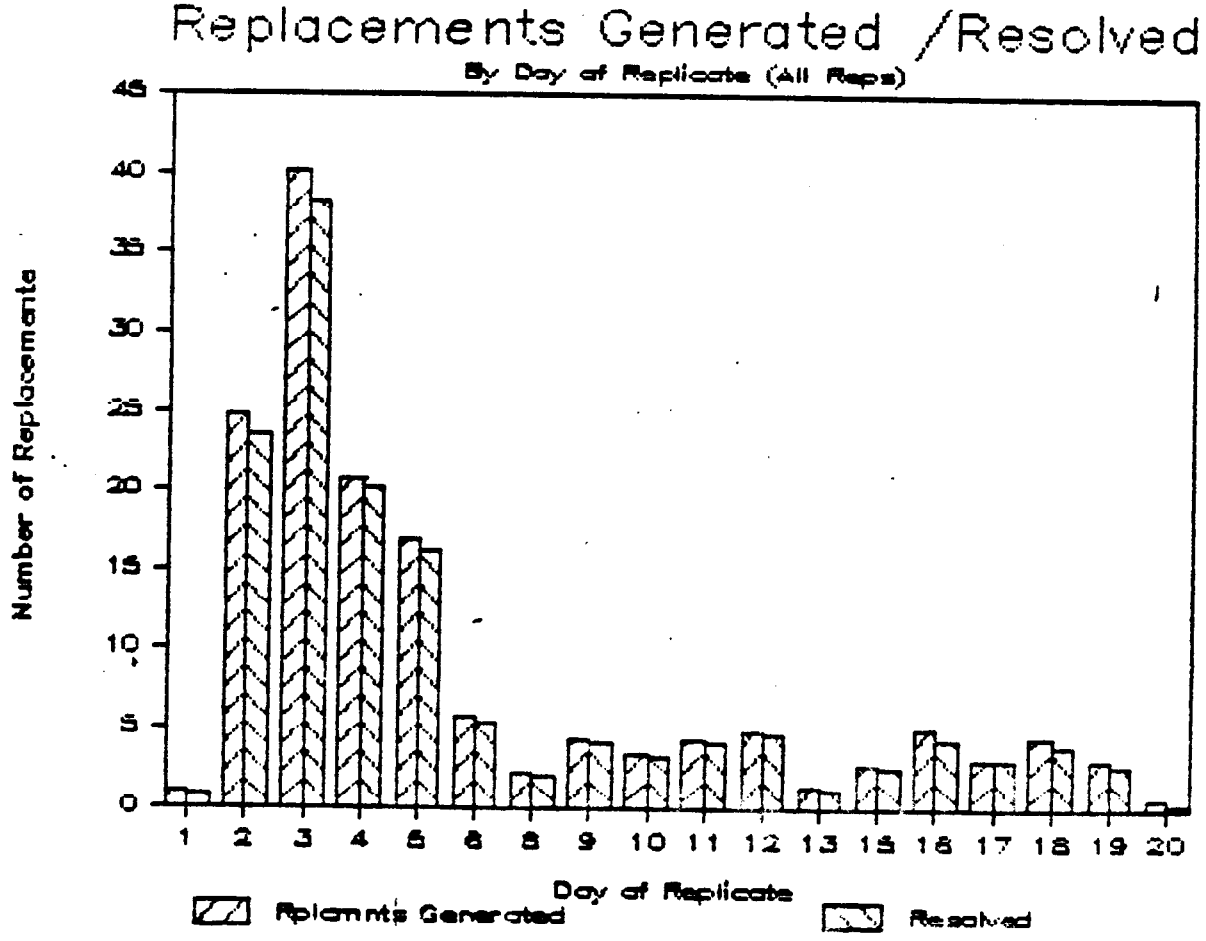
gure 3


Resolved Eligible Cases
By Day of Replicate (All Reps)
Number of Cases
70
60
50
40
30
20
10
0
1 2 3 4 5 6 8 9 10 11 12 13 15 16 17 18 19 20
Day of Replicate
Resolved Cases

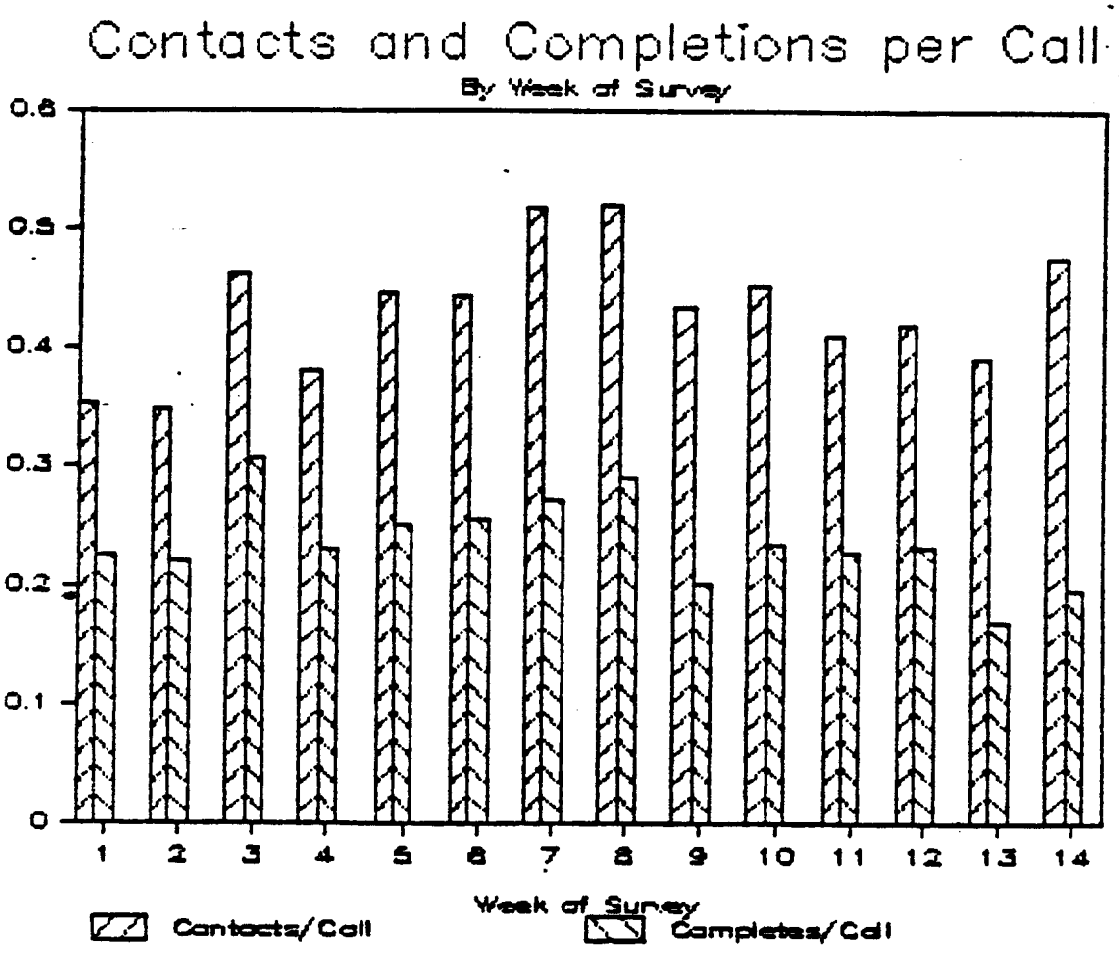
gure 4


Ave Number of Calls per Active Case
By Day of Replicate (All Reps)
Number of Calls
1.6
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 2 3 4 5 6 8 9 10 11 12 13 15 16 17 18 19 20
Day of Replicate
Calls per Case

ure 5


Replacements Generated /Resolved
By Day of Replicate (All Reps)
Number of Replacements
45
40
35
30
25
20
15
10
5
0
1 2 3 4 5 6 8 9 10 11 12 13 15 16 17 18 19 20
Day of Replicate
Rplcmnts Generated
Resolved

ure 6


Contacts and Completions per Call
By Week of Survey
0.6
0.5
0.4
0.3
0.2
0.1
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Week of Survey
Contacts/Call
Completes/Call

Appendix 6. Substitution

1. Introduction

One of the major concerns with the application of RDD for the NHIS is the response rate. If RDD nonresponse rates are 15-20 percent, the method used to account (or impute) for nonresponse in an attempt to reduce nonresponse bias can have an important impact on survey estimates.

The method that is probably used most often to impute for unit nonresponse in surveys is to adjust (upward) the weights of the respondents to account for the nonrespondents. These adjustments are usually made separately within nonresponse weight adjustment classes or cells. Within each cell the weights of the respondents are increased by a factor that causes the sum of the adjusted weights of the respondents to equal the sum of the unadjusted weights of all eligible sample cases in the cell. Effectively, this procedure imputes for the survey items of the nonrespondents in each cell the average values of the survey items of the respondents in the cell. An attempt is made to define weight adjustment cells in such a way that the respondents and nonrespondents in a particular cell will have similar survey characteristics. To the extent that this goal is accomplished, nonresponse bias will be reduced.

Another method of accounting for unit nonresponse is substitution: replacing a nonrespondent with a population unit not originally selected for the sample. The goal in using substitutes is to generate them in such a way that they have characteristics similar to those of the nonrespondents they represent. In the Feasibility Study, a substitute for a nonrespondent was obtained by randomly selecting another telephone residence from the same PSU. With respect to calling and interviewing, a substitute was treated the same as an original selection. Of course, all substitute cases in the respondent

file were identified so that the sample response rate, based on the original selections, could be calculated.

A major criticism of the use of substitution to account for unit nonresponse in surveys has been that a substitute might be viewed by interviewers or other survey personnel as being as good, or nearly as good, as the originally selected unit. Consequently, there is concern that a reduced effort might be extended to obtain a response from the original unit and that substitutes might not be carefully identified in the respondent file. However, with the control over the sampling operation that exists with a centralized RDD-CATI system, these potential problems have been eliminated. The design of the procedures for obtaining responses was not influenced by the fact that a substitution procedure was being used. Also, interviewers did not know whether or not they were dealing with an original sample case or a substitute. Of course, substitute cases were clearly identified in the respondent file.

Also, for the Waksberg RDD method, substitution has two advantages over the use of weight adjustment procedures. First, if substitutes are obtained for all, or nearly all, of the nonrespondents, the PSU sample sizes will be the same, or about the same. This may allow the sample to still be treated as a self-weighting sample for the purpose of survey estimation, a substantial convenience for users. Approximate equality of PSU sample sizes could not realistically be expected if substitution were not used.

Second, if the PSU sample clusters were used for nonresponse weight adjustment classes, there could easily be a large variation of weight adjustments across the sample since the fixed sample size (k) used for each PSU is generally small. This weight variation would tend to increase the variances of survey estimators. To avoid this problem, adjustment classes could be defined by groups of PSU's. However, if this were done, the non-

response bias reduction would suffer relative to the substitution procedure since imputation for nonresponse would then be made over a larger portion of the sample.

Based on the data collected for both the substitutes and original cases, four analyses of the substitution procedure used in this study were made. These analyses are described and the results are given in Section 3. Some conclusions and recommendations are given in Section 4. A description of the substitution procedure used is given in Section 2.

2. Description of the Substitution Procedure

Substitutes were selected within each PSU for those cases which were refusals, other noninterviews, or numbers which could not be contacted but were identified by a telephone business office as working.* For a case selected during the initial interview week of the three-week collection period for a replicate, a substitute was selected after the second refusal within a household or after 10 attempted calls to a working number with no contact. For a case selected during either the second or third interview week of a replicate, a substitute was selected after the first refusal within a household or after 7 attempted calls to a working number with no contact. The additional calls allowed during the initial interview week were due to the fact that two weeks were still available for contacting a substitute.

After a substitute was selected, call attempts were still made to the original sample unit as part of a followup procedure. One or two additional

---

* The generation of substitutes for nonrespondents is completely separate from the operation of replacing ineligible units (e.g., businesses) called during secondary screening. The replacement of ineligible units is part of the routine sampling operations, rather than a nonresponse adjustment procedure.

calls could be made to a refusal household in an effort to convert the refusal. After making a maximum of 20 calls to a working number in an attempt to make contact, the case was classified as a nonresponse.

Beginning with replicate six, it was decided that substitutes could not be selected in the final three days of a replicate. This was due to the experience of the first five replicates in which such cases were observed as not having a realistic chance of being contacted and interviewed. Because of an error made in implementing this modification, no substitutes were selected in replicates six and seven. Therefore, the substitution analysis results cited in this report are based on ten replicates instead of twelve.

3. Project Analyses and Results

There were four specific analysis tasks carried out in this investigation for the data collected in ten replicates. These tasks, which are listed below, are discussed in detail in subsections 3.1-3.4.

(1) Evaluation of the General Effectiveness of the Substitution Procedure.

This analysis included the derivation of the proportion of original cases that provided responses after being targeted for substitutes, the derivation of the proportion of targeted cases for which a substitute was contacted, and a comparison of the response rates of substitutes and of the original sample.

(2) Costs for Substitutes

Exact costs for substitution were not available from this study. However, several items closely related to costs were derived. These items are additional numbers of phone numbers, phone calls, interviews, and minutes associated with generating, pursuing, and interviewing substitutes.

(3) Comparison of Substitutes and Original Selections

This analysis consisted of a comparison of the characteristics of 150 late respondents with those of their substitutes. Comparisons were made for eight demographic and five health characteristics.

(4) Comparison of Variance Estimates Based on Substitution with those Based on Weight Adjustments.

This analysis consisted of a comparison of the two variance estimates for the estimated mean for each of five health characteristics.

3.1 Evaluation of the General Effectiveness of the Substitution Procedure

A total of 668 original sample units met the requirements listed in Section 2 for generating a substitute. Of these original sample units, 216 (32.3%) were eventually interviewed during the followup procedure. Although 668 substitute units were targeted for selection, only 618 were actually selected. Substitute units were not selected for the remaining 50 units because they were not targeted for substitution until replicate closeout had been reached or (in replicates eight through twelve) until the final three days before replicate closeout. Of the 618 substitutes, 543 were contacted resulting in 435 interviews, 84 refusals, 12 partial interviews and 12 other noninterviews. There were 75 substitutes which were determined to be working telephone numbers but could never be contacted. A display of these counts is given in Figure 1.

The response rate for substitutes was computed as $R_C = 74.0\%$.* For the original sample units, the response rate was $R_C = 78.9\%$. The difference of 4.9% can be attributable to the shorter amount of time generally available for contacting substitutes.

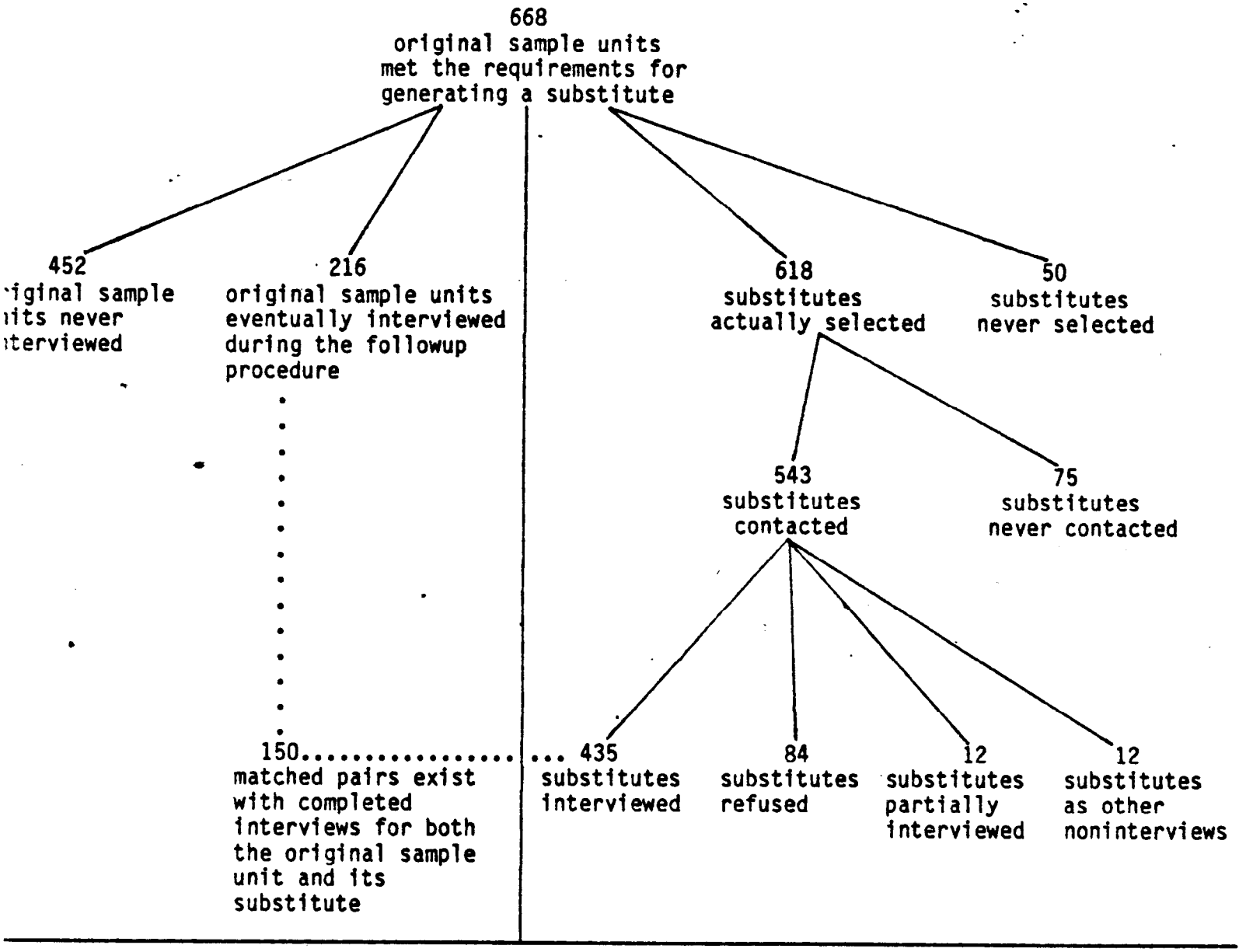
---

* For deriving this response rate, it is assumed that a portion of the noncontacted cases are residential. See Section 3.1 of the main report or Appendix 1 for a precise definition of R_C.

Regarding an evaluation of the rules for selecting subsitutes, an unclear picture is presented. The fact that 32.3% of those sample units which were targeted for substitutes were eventually interviewed suggests that substitutes may have been generated too early. But since 50 substitutes (7.5%) were never selected and since 125 substitutes (18.7%) were either never selected or never contacted, delaying the generation of subsitutes may not be wise.

Finally, as Figure 1 displays, there were 150 instances in which completed interviews were obtained from both the original sample unit and its subsitute. These matched pairs formed the base for the analysis discussed in Section 3.3.

Figure 1. Breakdown of the Basic Counts for Substitution



3.2 Costs for Substitutes

The exact costs incurred due to substitution were not available from this study. Therefore, several items closely related to cost were derived in an attempt to learn how much time and effort was expended in pursuing and interviewing substitutes. Averages were computed on a PSU basis, using the 208 PSU's which were selected for the 10 replicates used to study substitution. Table 1 summarizes these results.

Table 1. Data on the Use of Substitutes

Item	Total from 10 replicates	Average per PSU
Number of times a substitute was supposed to have been generated	668	3.21
Number of substitutes actually selected	618	2.97
Number of additional phone numbers generated due to substitution (including ineligible cases)	1063	5.11
Number of additional phone calls made	3589	17.26
Number of additional complete interviews obtained	435	2.09
Number of minutes of on-line telephone time due to substitution *	26033	125.16

It is interesting to look at the time spent on substitute cases in terms of the equivalent number of original sample cases it represents. From Table 1 it is seen that slightly more than two hours of on-line telephone time was needed, on the average, to pursue an average of 2.97 substitutes per PSU.

---

* Although presented here in minutes, the on-line time was collected in seconds to provide a high degree of accuracy.

In considering only the original sample units, it was found that the average amount of on-line telephone time spent per sample unit was 45.90 minutes. This implies that the time spent on substitutes was equivalent to the time spent on approximately 2.73 (i.e., 125.16/45.9) original sample units per PSU. This average is slightly less than the average number of substitutes selected per PSU (2.73 vs. 2.97) because more time was generally available to pursue original cases than to pursue substitutes.

An important way of interpreting this data is that if substitution was not used, then the original sample size could have been increased by about 3 units per PSU with only minimal additional cost. This interpretation is critical in the development of the variance comparisons presented in Section 3.4.

3.3 Comparison of Substitutes and Original Selections

As was pointed out in Section 3.1 and illustrated in Figure 1, there were 150 matched cases for which interviews were obtained from both the original sample household and its substitute. This provided an opportunity to compare a population of late respondents with one of substitute respondents. This comparison is not the same as the ideal comparison between nonrespondents and substitutes. However, comparing late respondents with substitutes is still useful because the late respondents would have been nonrespondents if follow-up attempts had not been as extensive as they were.

For the 150 pairs of original and substitute cases, a comparative analysis was carried out for eight demographic and five health characteristics. For four of the demographic characteristics and for all five of the health characteristics, a standard large-sample normal test was performed to determine if the sample means for original cases were significantly different from the sample means for the substitutes. Household averages were used as the basic variable of

comparison for each of the five health characteristics. The nine characteristics included in the comparative analysis are listed in Table 2 along with the means for the originals and substitutes, the estimated standard error of the difference between these means, and the Z-score.*

Table 2. Comparisons of Means

Demographic Characteristics	Mean (Originals)	Mean (Substitutes)	Estimated Standard Error of Difference	Z-Score
Household Income	28,109	26,302	2,682	.67
Age (Reference person)	39.84	46.40	1.75	-3.75
Average age of household member	33.87	40.35	2.00	-3.24
Household size	2.39	2.44	.15	-.33
Health Characteristics (Number of)				
Hospital Stays in the Last Year	.105	.138	.034	-.97
Illness Bed Days in the Last Year	3.168	3.601	1.060	-.41
Doctor Visits in the Last Year	2.766	2.810	.466	-.09
Doctor Visits in the Last 2 Weeks	.201	.249	.057	-.84
Work Days Lost in the Last 2 Weeks	.086	.231	.101	-1.44

The other four demographic characteristics are not quantitative; therefore, a comparison of means could not be made. Instead, a standard chi-square

* The standard error of the difference of means was estimated based on the 150 observed differences between late responding originals and their substitutes. The Z-score is simply the difference between means divided by the estimated standard error of the difference.

test was used for each of these characteristics to test the homogeneity of the original and substitute distributions. These four characteristics are listed in Table 3 along with the computed chi-square (test) statistic and the chi-square critical values for the 10% level of significance (i.e., the 90th percentile of the appropriate chi-square distribution).

Table 3. Distribution Comparisons

Characteristic (of Reference Person)	Computed Chi-square Statistic	Ninetieth Percentile of Chi-square Distribution
Marital Status	5.21	7.78
Sex	3.22	2.72
Race	.31	4.61
Education	.76	7.78

For both the comparisons of means and the comparisons of distributions, simple random sampling was assumed. Of course, the full sample was selected in clusters of 12 units. However, the 150 pairs of late responding originals and their substitutes are not nearly as clustered as was the full sample. Of the 102 clusters that contain at least one pair, 66 clusters contain exactly 1, 28 clusters contain 2 pairs, and 8 clusters contain 3 or more pairs. Consequently, the assumption of simple random sampling should not cause serious problems in this comparative analysis.

From inspection of Table 2 it is noted that a significant difference between the means at the 10% level was observed for only two of the nine variables: age of reference person and average age of household members. In both cases the mean age of the substitutes was significantly higher than the mean age for the original cases. This implies that the ages of the persons in substitute households are generally higher than the ages of the person in

the late respondent households for which the substitutes were selected. This is not surprising since it could be anticipated that difficult-to-reach original sample households would contain more younger and thus more mobile persons than would their easier to reach substitute households. Although no significant differences were observed between means for health characteristics, it is interesting to note that for all five comparisons the average number of illness-related characteristics was higher for the substitutes than for the original sample cases. This may also be due to the age differences.

For the four distribution comparisons summarized in Table 3, only the distributions of sex of reference persons differed significantly between originals and substitutes at the 10% level. This significant difference arises because the percent of female reference persons in the original sample (32) is significantly less than in the substitute sample (42). This suggests that substitute households contain disproportionately more female reference persons than do the late responding original households. This is not surprising since a higher proportion of men are in the labor force and consequently would be harder to contact than women.

3.4. Comparisons of Variance Estimates Based on Substitution with Those Based on Weight Adjustments

For each of the five health characteristics included in this analysis, a comparison was made between the variance estimate of the estimated mean based on the original sample plus substitutes and the variance estimate of the estimated mean based on an equal-cost sample that utilized weight adjustments, rather than substitutes, to account for nonresponse. The cost-related data given in Section 3.2 was used to develop an equal-cost sample that did not use substitutes. It was demonstrated there that if substitution were not used, three more telephone residences could have been selected per PSU with

only a slight increase in survey costs. Therefore, the weight-adjustment sample that was taken to be equal in cost to the full substitution-based sample was one consisting of the original sample of 12 residential units, plus three additional residential units, per PSU.

An equal-cost weight-adjustment sample could have been defined by retaining the fixed PSU sample size of 12 residences, but increasing the number of PSUs. However, if this were done, an empirical comparison of variance estimates could not have been made without reducing the substitution-based sample, since no additional PSUs had been selected. Although it was not done as part of this analysis, a model-based comparison of these two variance estimators could be developed which would allow for selecting a higher number of PSUs for the weight-adjustment sample than for the substitution-based sample.

The equal-cost weight-adjustment sample was created by adding three "pseudo cases" to each of the 208 PSUs included in the substitution analysis. First, the response rate was calculated for each PSU based on the sample of 12 residences selected. This rate was multiplied by 3 to obtain the "expected number" of additional interviews that would have been obtained in the PSU if 15 residences, rather than 12, had been selected initially. This expected number was rounded to the nearest integer to determine the number of additional interviews (i.e., pseudo interviews) to obtain from the PSU for the weight-adjustment sample. A constraint was included in the procedure for determining the number of additional interviews so that the overall sample response rate would not be altered.

The additional interviews for a PSU were obtained in two ways. First, any substitutes that had been interviewed for the PSU were used as a source of additional interviews. This is appropriate since substitutes are simply additional random selections from the PSU. In cases for which the number of

substitute interviews exceeded the number of additional interviews needed, the number needed was selected randomly from the available substitutes. For 107 of the 208 PSUs included in the substitution analysis, there were enough substitute interviews available to provide the pseudo interviews needed for the weight-adjustment sample. Second, for each of the remaining 101 PSUs, one or more pseudo interviews were provided, as needed, by selecting cases randomly from the completed interviews obtained from the original selections. That is, when necessary, entire interviews were "hot decked" (or replicated) to obtain the required number of additional interviews to complete the weight-adjustment sample. The maximum number of hot deck cases needed per PSU was three; this number was needed for 16 of the 101 PSUs. For the other 85 PSU's, either one or two hot deck interviews were selected.

For the weight-adjustment sample, adjustment classes were taken to be individual PSUs. This choice was made since the substitution classes were also the individual PSUs. With the adjustment classes and substitution classes being the same, the nonresponse bias for the substitution-based estimator of the mean should be about the same as that for the weight-adjustment-based estimator of the mean.* For this choice of adjustment classes, the nonresponse weight adjustment, w_i, assigned to each respondent selected from the i-th PSU is

$$w_i = k_i/k_i' , \qquad (1)$$

---

* Actually, since there is generaly less time available to obtain an interview from a substitute residence than from an original sample residence, there could be a subtle bias that exists for the substitution-based estimator that does not exist for the weight-adjustment-based estimator, even though the weight adjustment classes and the substitution classes are the same. This is discussed further in Section 4.

where

k_i = the PSU sample size (i.e., 15),

k_i' = the total number of completed interviews, including pseudo interviews, for the i-th PSU.

Since completed interviews were not obtained for all substitute cases, the weight adjustment given in equation (1) also had to be used for the substitution-based estimator. In this case, $k_i = 12$ and k_i' = the number of completed interviews in the PSU, including substitutes.

The variance estimates to be compared for the two types of estimators were computed using the same variance formula. To develop this formula, some notation is needed. First, the weighted sum, x_i', for a characteristic, X, for the i-th PSU is equal to

$$x_i' = w_i \sum_{j=1}^{k_i'} x_{ij} \quad ,$$

where x_{ij} = the sum of the values of X for all persons in the j-th respondent household in the i-th PSU.

Similarly, the sum, n_i', of the weights for the i-th PSU is equal to

$$n_i' = w_i \sum_{j=1}^{k_i'} n_{ij} \quad ,$$

where n_{ij} = the number of persons in the j-th respondent household in the i-th PSU.

The estimator of a population mean, $\bar{x}$, can be written as follows:

$$\bar{x} = x'/n' \quad ,$$

where

$$x' = \sum_{i=1}^{208} x_i' \quad ,$$

$$n' = \sum_{i=1}^{208} n_i' \quad .$$

The variance estimate of the population mean, $\bar{x}$, was computed using the standard Taylor Series approximation to the variance of a ratio:

$$\hat{\sigma}_{\bar{x}}^2 \doteq \bar{x}^2 \left[\frac{\hat{\sigma}_{x'}^2}{(x')^2} + \frac{\hat{\sigma}_{n'}^2}{(n')^2} - 2 \frac{\hat{\sigma}_{x'n'}}{x'n'} \right] . \qquad (2)$$

All terms in equation (2) have been defined previously except the two variance estimates, $\hat{\sigma}_{x'}^2$ and $\hat{\sigma}_{n'}^2$, and the covariance estimate, $\hat{\sigma}_{x'n'}$. Each of these three estimates was derived using an ultimate cluster variance estimate. For example,

$$\hat{\sigma}_{x'}^2 = \frac{208}{207} \sum_{i=1}^{208} (x_i' - x'/208)^2 . \qquad (3)$$

The other variance estimate and the covariance estimate were computed in a way analogous to the variance estimator in equation (3).

For both the substitution-based estimator and the weight-adjustment-based estimator, the variance formula given in equation (2) was applied to the estimated means for all five health characteristics. The ten variance estimates, along with the estimated means, are given in Table 4.

Table 4. Variance Estimates for the Substitution-Based and Weight-Adjustment-Based Estimates

Health Characteristics (Number of)	Substitution		Weight Adjustment	
	Estimated Mean	Estimated Variance	Estimated Mean	Estimated Variance
Hospital Stays in the Last Year	.148	.000068	.152	.000077
Illness Bed Days in the Last Year	4.484	.107	4.553	.143
Doctor Visits in the Last Year	3.338	.0184	3.383	.0206
Doctor Visits in the Last 2 Weeks	.248	.00018	.247	.00020
Work Days Lost in the Last 2 Weeks	.247	.00074	.255	.00095

It can be observed from Table 4 that, for the estimated mean for all five health characteristics, the variance estimate for the substitution-based estimator was less than the variance estimate for the weight-adjustment-based estimator. Consequently, with regard to variance estimates, substitution appears to be superior to a PSU-by-PSU weight adjustment procedure as a method of accounting for unit nonresponse.

It should be noted that the variance estimates for the weight-adjustment procedure probably would have been less if groups of PSUs, rather than individual PSUs, had been used for adjustment classes since weight adjustment factors would not have varied as much. However, if this were done, the potential of the weight-adjustment-based estimator to reduce nonresponse bias would likely be diminished as was discussed in the first section of this appendix.

4. Conclusions and Recommendations

The general success of any substitution procedure will depend heavily on the substitution rules used and on the call scheduling applied to substitutes. The rules used in the NHIS-RDD Feasibility Study were discussed in Section 2 of this report. This experimental procedure was chosen primarily on an intuitive basis without a substantial amount of preliminary investigation. It turned out that the procedure used was not particularly successful. A fairly high portion (32%) of the cases targeted for substitution were eventually interviewed, which represents some unnecessary expenditure. Also, for 7.5% of these targeted cases, substitutes were never generated. Finally, for the cases for which substitutes were generated, the response rate was about 5% lower than for the original sample. The rules for initiating substitutes should be carefully considered. Perhaps there were certain types of cases for which substitutes were generated too early. Also, the data collection period might have to be increased or the call scheduling modified

to improve the generation rate and response rate for substitutes. Furthermore, consideration should be given to the possibility of generating additional substitutes for a case when the first substitute turns out to be a nonrespondent, if there is enough time to contact additional substitutes.

The comparison of hard-to-interview original sample cases and their substitutes, discussed in Section 3.3, investigates the potential for nonresponse bias in the use of sustitution to account for unit nonresponse. The reference persons in the substitute respondent households were older, had a higher percent female, and indicated a tendency to report higher numbers of illness-related activities than did their hard-to-interview counterparts. These differences indicate that there is the potential for biases in the survey estimates due to the use of substitutes.

How would such biases compare to those associated with nonresponse weight adjustments in the case where weight adjustment classes are taken to be the same as the substitution classes (i.e., the individual PSUs)? In designing this research it was assumed that the biases associated with these two procedures would be the same since substitutes are additional respondents from the same PSU and weight adjustments within a PSU impute characteristics of the respondents in the PSU to the nonrespondents in the PSU. However, since less time is generally available to pursue substitutes than original sample cases, these biases may not be equal. Since substitute respondents must generally be "early cooperators" because of the time constraint, there may be a bias component associated with the use of substitution that may not exist for the corresponding weight adjustment procedure. To minimize this differential effect, the rules for initiating substitutes, the interview period, and call scheduling procedure should be designed in such a way that adequate time will be available to pursue substitutes. The response rate for substitutes would provide an indication of

whether there was adequate time to pursue substitutes. If it were about the same as the response rate for the original sample, then there probably was adequate time provided to pursue substitutes.

With regard to variance estimation, discussed in Section 3.4, the substitution-based estimates were superior to the weight-adjustment-based estimates for all five health characteristics included in the analysis. However, the weight adjustment classes used were the indivudual PSUs. The variance comparison may have been more conclusive if the weight adjustment classes had been larger--perhaps groups of PSUs. However, if this had been done, the bias associated with the weight-adjustment-based estimates may have increased, which would have complicated the comparison of the two types of nonresponse procedures.

The results from the comparison of the two approaches of accounting for unit nonresponse--substitution and PSU-by-PSU weight-adjustment--were inconclusive. They suggest that there may be a bias-variance tradeoff involved: The weight-adjustment procedure may provide estimates that are less biased, while the substitution procedure seems to provide estimates with a lower variance. A mean square error comparison was not possible for this study. Such a comparison would be difficult to do, but would be very useful.

Due to the more constant sample size per PSU and the lower variance estimates as compared to a PSU-by-PSU weight adjustment, it does appear that substitution is a slightly better method of accounting for nonresponse in an RDD survey, provided that there is generally enough time allowed in the data collection phase to adequately pursue substitute cases.

Appendix 7. Special Places

I. Background

For the past several years the Bureau of the Census has been exploring the use of the Random Digit Dialing (RDD) telephone survey methodology as a cost-saving alternative to its current personal visit demographic surveys. As a relatively new survey methodology, the RDD surveys face unresolved design, operational and analytical issues. One such issue that has recently attracted attention at the Bureau is the need for developing procedures for enumerating group quarters and other unusual types of living arrangements on the telephone.

The RDD surveys conducted to date by the Bureau, and by most other survey organizations, have not addressed the problems associated with enumerating living arrangements found in places the Bureau classifies as "special places." These are places where people live that are different than the usual types of living quarters and where the occupants usually share some common facilities -- for example, college dormitory housing for students and retirement homes for the elderly are two types of special places. Although special places are believed to house about three percent of the nation's population, there has been no serious attempt to enumerate at these places in RDD surveys, primarily because of perceived operational problems. These types of living quarters are usually considered out-of-scope for telephone surveys. It is important to note that special places are an integral part of the sampling frame for the Bureau's personal visit surveys, and the omission of these places in RDD surveys produces an undesirable bias in the sample (unless they are included as part of an area sample for a dual-frame survey.) As such, the Bureau has decided to explore ways to successfully enumerate special places as part of its on-going study of the RDD methodology.

The potential problems associated with enumerating the occupants of special places using an RDD methodology revolve around three basic issues: (1) the ability of enumerators to successfully screen for special places on the telephone, that is, successfully differentiate special places from other telephone numbers, (2) the feasibility of enumerators obtaining an accurate list of the beds, rooms or persons within the place that are eligible for inclusion in the survey, so that a representative sample of units can be selected for interview, and (3) the ability to conduct interviews with designated sample units with an assurance of high quality data.

This paper describes a special place study that was conducted in conjunction with the Bureau's research into the feasibility of collecting health information on the telephone. The Bureau conducts the National Health Interview Survey (NHIS) under the sponsorship of the National Center for Health Statistics. At the request of NCHS, the Bureau began investigating the use of the RDD methodology as a cost-saving alternative to the personal visit method currently used for collecting data on the NHIS, investigated the cost and response rate concerns regarding an RDD survey and, at the same time, looked into operational problems and estimation/weighting problems associated with RDD surveys.

One of the stated goals of the NHIS-RDD study was to develop and evaluate procedures for identifying and handling special places over the telephone.

The special place research associated with the NHIS-RDD study as discussed in this paper was a first step toward more sophisticated research in the future. As such, its objectives were fairly conservative: (1) to provide a reading on the ability of enumerators to screen for special places on the telephone, that is, successfully differentiate special place telephone numbers from other residential and nonresidential numbers and (2) provide a preliminary indication of how well the eligible sample units within special places could be identified and listed. Due to budget constraints, the research did not address the issue of data quality nor did it examine the potential problems associated with actually enumerating persons within a special place. These issues may be studied as part of future RDD programs.

II. Methodology

The sample for the basic NHIS-RDD study was selected using a two stage procedure proposed by Waksberg[1]. Twenty-one primary sampling units were selected and twelve secondary units were selected within each primary, yielding 252 units. The sample was replicated 12 times for a total sample size of 3024 households.

Each replicate could be expected to yield about five special place sampling units. However, these special places were not used in the analysis because not enough was known about these places to have a controlled experiment. A success rate for identifying special places could not be calculated because there was no way of knowing the number of special places which were in sample but not correctly identified as special places. Also, there would be no easy way to verify any listing of living arrangements obtained for these special places. Therefore, the special places selected through the regular sample design were treated as "out-of-scope" for the survey (whenever they were properly identified).

---

[1] Waksberg, Joseph (1978), "Sampling Methods for Random Digit Dialing," Journal of the American Statistical Association, 70,40-46.

In order to measure the success rate for identifying special places and to verify living arrangements, the sample was seeded with known special places. Then, the number of special places correctly identified could be compared to the total number seeded to calculate a success rate. Also, living arrangements could be more easily verified. Each replicate, starting with Replicate 02, was seeded with special places drawn from two sources:

1. Current Survey Special Places

The Bureau's clerical operations unit identified 96 special places that rotated out of prior Bureau surveys since December 1982. These special places were randomly assigned to replicates 02-12. The definitions for special places used for the face-to-face surveys were also used for the NHIS-RDD even though it is recognized that the concept of sample unit is different for telephone surveys. The types of places included in this frame included the full gamut of special places, ranging from student housing to correctional and long term care institutions. The frame included a sizeable number of fairly small places with fewer than 20 eligible units, since these are the types of places most likely to rotate out of the current surveys after their eligible units are enumerated.

The intent of using special places from other surveys was to provide a means to compare the listing of units made by the RDD staff with the field listing made by a Census Bureau enumerator who visited the place in person for the other survey. This provided a rough indication of how well the eligible units within special places could be identified over the telephone. There was no attempt to actually conduct telephone interviews at these places, because of respondent burden considerations and the operational problems associated with special place enumeration.

2. Telephone Directories

In addition to the special places derived from other Bureau surveys, each replicate was also seeded with 20 telephone numbers known to be for special places, drawn from 1983 telephone directories. Three hundred special places were identified by clerks who were provided with randomized lists of types of special places. Each clerk was assigned several commercial telephone directories and instructed to search for listings under the special place types assigned. For example, the clerk whose list showed "Rooming Houses" was instructed to list a specified number of rooming houses from the Yellow Pages assigned to him/her. Each special place was listed on an index card. The card deck was then shuffled, and twenty special places were randomly assigned to each replicate of the study. This process assured that each replicate would include a variety of special place telephone numbers.

The intent of seeding the sample with telephone directory special places was to provide a reading of the ability of RDD enumerators to distinguish special places from regular housing units and from commercial (nonresidential) telephone numbers.

III. Results

Table 1 shows that the RDD enumerators successfully identified only about 39 percent of the special place telephone numbers seeded into the sample during replicates 02 through 12. The low rate of success in identifying telephone numbers for special places suggests that these telephone numbers are difficult for enumerators to distinguish from nonresidential or other residential telephone numbers.

It is interesting that the success rate over the final five replicates increased from 39 percent to 56 percent. This increase followed an intensive refresher training session in which all supervisors and enumerators had the opportunity to review and practice the special place identification procedures. The increased success rate was noticed in the replicate immediately following the retraining and was sustained for the duration of the study. This suggests that special place identification is a difficult task for enumerators and that intensive training followed by periodic refresher training might result in improved performance.

The initial enumerator training on identifying special places was secondary to the training on interview techniques for the regular survey document, and so some enumerators received very little formal training on special place procedures. In addition, the regular survey questionnaire was not set up to easily lead the enumerator through a series of special place screening questions. Thus, in many cases the initial determination of whether a telephone number served a special place was based upon the respondent's reaction to the screening question "Have I reached you on your home phone?" The screening questions were modified after replicate 07 to include the probe "Does this number serve a place where people can live...?" in hopes of identifying special places where the respondent initially indicated that the number was for someplace other than his/her "home". This modified screening procedure, coupled with the formal retraining after replicate 07, led to the improved success rate for the final five replicates of the special place study.

Table 1 also displays the distribution of success rates by type of special place. Fraternity and sorority houses had the highest identification rate (86 percent), and convents, dormitories, motels, rectories, missions, jails and group homes were identified with greater than fifty percent accuracy. There is no readily apparent rationale for why these types of places were more easily identified on the telephone than other similar types of living arrangements. For instance, the success rate for motels was 55 percent, but it dropped to 30 percent for hotels. While hotels may have more permanent guests who consider the place their "home", this should not result in the much lower identification rate. Similarly, it is difficult to explain the difference in rates for missions (62 percent success) compared to halfway houses and YWCA's (zero success rate). It may be that the respondents answered the probe questions differently in the special places with low success rates, although the present study did not collect detailed data on respondent comments to open-ended screening questions.

Table 2 shows the results for replicates 08-12, when the special place identification success rate increased to fifty-six percent. It is interesting that the success rate for trailer parks remained at zero for the final five replicates. These places may present unique problems for telephone interviewers since the living arrangements are very similar to regular housing units. Trailer parks were not treated as special places for the 1980 Census, and perhaps should not be considered special places for telephone surveys.

Table 3 shows the distribution of successfully identified special places by enumerator. Six of the 23 enumerators who contacted any seeded special places identified more than fifty percent of the special places assigned to them. Note also that interviewers I, K, M and Q (with a total of 94 eligible special places assigned to them) had very poor success rates ranging from 43 percent to seven percent. Five enumerators were unsuccessful in identifying any of the special places in their workload.

The results of this preliminary study suggest that some enumerators are more adept than others in distinguishing special place telephone numbers from other types of telephone numbers, although the study did not collect any profiles of the enumerators with the high success rates. There did not seem to be any relation between performance on the special place study and performance on other aspects of the overall RDD study. For example, the interviewers with high success rates on special place identification did not have higher than average production rates on the regular enumeration. In addition, three of the six interviewers with high success rates worked during the night shift, when the level of supervision was somewhat lower than the day shift. Experience on Census Bureau survey work did not correlate with success rate, since three of the interviewers with superior success rates were newly hired and had no previous survey experience.

The present study was undertaken with two goals: (1) to provide a reading on the ability of telephone enumerators to successfully differentiate special place telephone numbers from other residential and nonresidential numbers, and (2) provide a preliminary indication of how well the eligible sample units within special places could be identified and listed. The latter goal was considered secondary to the measurement of the ability of enumerators to identify special places over the telephone, but the results provide some interesting preliminary findings.

In the seeded special places which were derived from other Census Bureau face-to-face surveys where interviewers visited special places and list-enumerated them, we compared the listing made in the face-to face interview situation with the listing made over the telephone. Table 4 provides the results for 15 places where the two independent listings were made. The small size of this sample is related to several factors, including refusals by three special places to provide a listing of units over the telephone, and clerical problems with contacting several places within a reasonable period of time because of higher priority work on the regular enumeration aspects of the overall RDD study.

The Table 4 results show that the telephone listing of special place units was identical to the face-to face listing in 11 of the 15 special places. In one special place the RDD lister erroneously listed over 300 ineligible units because of a misunderstanding of the eligibility rules for the survey. The listings in the remaining three special places differed from the face-to-face listing because of changes that occurred in the special place after the face-to-face listing. Although the sample of 15 places is much too small to make generalizations, the results suggest that telephone enumerators can make accurate and complete listings of units within special places provided they have structured questions to ask and formal procedures to follow. The special place probe questions for listing units are illustrated in Exhibit 1.

The present study was preliminary in nature, and provided some limited empirical evidence on the ability of telephone interviewers to successfully identify special places and to compile a list of eligible units within the identified special places. The results suggest that intensive training is important for adequate success rates, and that special place identification is more difficult on the telephone than face-to-face. The results also showed that the telephone enumerators can successfully compile a sampling frame of eligible units within special places.

Future studies of the viability of identifying and sampling special places using the RDD methodology are currently in the planning stage at the Bureau. These studies will look at such variables as interviewer characteristics and how they relate to successful performance on the special place operations, modified screening and probe procedures and their effect on the ability of enumerators to successfully screen for special places, and an in-depth investigation of some of the operational problems associated with identifying the units which are eligible for inclusion in the various demographic surveys conducted by the Bureau.

cc:

J. Massey	(NCHS)		T. Demaio	(CSMR)
O. Thornberry	"		P. Biemer	(SRD)
S. Rice	"		D. Chapman	"
J. Fitti	"		C. Hogue	"
K. Frey	"		T. Walsh	(DSD)
W. Mockovak	(FLD)		E. Davey	"
M. Haas	"		R. Mangold	"
R. Tegels	(SMD)		R. Wilson	"
L. Love	"		T. Roman	"
L. Cahoon	"		L. Brown	"
J. Paletta	"			
F. Nash	"			
G. Shapiro	"			
L. Baer	"			

SMD:MTenebaum:rw:7/30/84

Table 1. Special Place Identification Success Rate, By Type of Special Place

Type	No. Seeded	No. Coded as Non-working	No. Eligible	No. Correctly Identified	Success Rate (%)	No. Coded as Regular Living Qtrs.	No. Coded as Nonresidenti Telephone Numbe
Group homes	3	2	1	1	(100)	-	-
Fraternity/Sorority	11	3	8	7	(88)	1	-
Jails	4	1	3	2	(66)	-	1
Missions	11	3	8	5	(62)	-	3
Rectories	5	0	5	3	(60)	-	2
Motels	24	2	22	12	(55)	4	6
Dormitories	25	6	19	10	(53)	-	9
Convents	18	1	17	9	(53)	2	6
Rest Homes	33	4	29	14	(48)	3	12
Military Housing	8	2	6	2	(33)	-	4
Hospital Facilities	18	2	16	5	(31)	1	10
Hotels	11	1	10	3	(30)	2	5
Rooming/Boarding Houses	32	18	14	4	(29)	7	3
Tourist Homes	9	1	8	2	(25)	3	3
Children's Homes	10	2	8	2	(25)	1	5
Rehabilitation Centers	12	2	10	2	(20)	2	6
Sanatariums	14	4	10	1	(10)	1	8
Trailer Parks	23	6	17	0	(0)	6	11
Halfway Houses	2	0	2	0	(0)	1	1
YWCAs	2	0	2	0	(0)	0	2
TOTALS	275	60	215	84	(39)	34	97

Table 2. Special Place Identification Success Rate, by Type of Special Place, For Replicates 08-12

Type	No. Seeded	No. Coded as Non-working	No. Eligible	No. Correctly Identified	Success Rate (%)	No. Coded as Regular Living Qtrs.	No. Coded as Nonresidentia Numbers
Group Homes	1	0	1	1	(100)	-	-
Rectories	2	0	2	2	(100)	-	-
Hotels	3	0	3	3	(100)	-	-
Jails	3	1	2	2	(100)	-	-
Motels	12	-	12	10	(83)	-	2
Convents	8	1	7	5	(71)	1	1
Rest Homes	17	2	15	10	(67)	1	4
Tourist Homes	3	0	3	2	(67)	-	1
Sanatariums	6	3	3	2	(67)	0	1
Fraternity/Sorority	6	0	6	4	(67)	1	1
Dormitories	10	2	8	4	(50)	0	4
Missions	4	0	4	2	(50)	0	2
Rehabilitation Centers	8	1	7	3	(43)	2	2
Hospital Facilities	5	0	5	2	(40)	1	2
Rooming/boarding Houses	18	8	10	4	(40)	3	3
Trailer Parks	12	4	8	0	(0)	3	5
Military Housing	3	1	2	0	(0)	1	1
Children's Homes	4	2	2	0	(0)	0	2
TOTAL	125	25	100	56		13	31

Table 3. Special Place Identification Success Rate, By Interviewer

Interviewer Code	No. Cases	No. Coded as Non-working	No. Eligible	No. Correctly Identified	Success Rate (%)	No. Coded as Regular Unit	No. Coded as Nonresidential
A	1	0	1	1	1.000	0	0
B	11	3	8	6	0.750	0	2
C	23	3	20	14	.700	2	4
D	17	4	13	7	.538	3	3
E	18	5	13	7	.538	2	4
F	2	0	2	1	.500	0	1
G	12	3	9	4	.444	2	3
H	21	3	18	8	.444	1	9
I	38	8	30	13	.433	4	13
J	11	5	6	2	.333	2	3
K	48	11	37	11	.297	4	2
L	10	3	7	2	.285	3	2
M	17	3	14	4	.285	3	7
N	5	1	4	1	.250	2	1
O	5	1	4	1	.250	2	1
P	10	3	7	1	.142	1	5
Q	15	2	13	1	.076	1	11
R	1	0	0	0	.000	0	1
S	5	0	0	0	.000	1	4
T	1	0	1	0	.000	1	0
U	3	2	1	0	.000	0	1
V	1	1	0	0	.000	0	0

Table 4. Special Place Units Listed by RDD Compared to Units Listed Face-to-Face

Case ID	Type of Place	Units Listed By RDD	Units Listed by Face-to-Face Survey	Difference 2/
A	Convent	7	4	3
B	Group home	0	0	0
C	Hotel	0	0	0
D	Hotel	0	0	0
E	Motel	0	0	0
F	Rectory	0	4	4
G	Rest home	5	5	0
H	Motel	1	1	0
I	Motel	1	1	0
J	Tourist home	18	18	0
K	Fraternity House	18	16	2
L	Dormitory	60	60	0
M	Retirement Home	1	1	0
N	Retirement Home 1/	330	0	330
O	Rectory	5	5	0

1/ RDD enumerator listed ineligible units due to misunderstanding of the eligibility rules.

2/ Places with zero units listed had no units eligible for the survey, e.g., hotels with no rooms for permanent guests or employees.

FORM THIS-104(X) (1-7-83)

U.S. DEPARTMENT OF COMMERCE
BUREAU OF THE CENSUS

TELEPHONE HEALTH INTERVIEW SURVEY

SUPERVISOR PROBE QUESTIONS FOR SPECIAL PLACES

PRETEST VERSION

a. Case ID	Replicate number
9999999	2

b. Telephone number: 301 684 3951

c. Date of contact	d. Interviewer code
2/9/84	B6

e. Supervisor name:

f. Respondent name and title	Telephone number
BOB JOHNS	301 684 3951

INTRODUCTION — I am *(Your name)* **from the U.S. Bureau of the Census. I am following up on a recent call made by one of our interviewers. May I speak to** *(read name in f or say:* **someone knowledgeable about the living quarters there)?**

Section I - IDENTIFICATION OF SPECIAL PLACE TYPE

Type of place *(Dorm, convent, etc.)*
Mark if known from Q9 on the THIS-2a or 3a; otherwise ASK: **What type of place is this?**

- ☐ Dorm
- ☐ Hotel/Motel
- ☐ Rooming house
- ☐ Boarding house
- ☐ Military
- ☐ Halfway house
- ☐ Home for aged
- ☐ Trailer park
- ☐ Convent
- ☐ Mission
- ☐ Other — *Specify* _____

What is the name of this . . . *(read type from 1a)?*

CHECK ITEM A ▶ **Is this an institution as defined on the Special Place flashcard?**

- ☐ Yes — *Ask 2*
- ☐ No — *Go to Check Item B*

Does this institution provide living quarters for staff members?

- ☐ Yes — Only these staff units are eligible for the survey — *Go to section II*
- ☐ No — This institution is not eligible for the survey — *Terminate the call*

CHECK ITEM B ▶ **Is this Special Place on a military post or military installation?**

- ☐ Yes — *Ask 3*
- ☐ No — *Go to Check Item C*

Does this place house any civilians?

- ☐ Yes — Only these civilian living quarters are eligible for the survey — *Go to section II*
- ☐ No — This military housing is not eligible for the survey — *Terminate the call*

CHECK ITEM C ▶ **Is this place a hotel or motel?**

- ☐ Yes — *Ask 4*
- ☐ No — *Go to Check Item D*

Does this place have any rooms for permanent guests or resident employees?

- ☐ Yes — Only these rooms for permanent guests or resident employees are eligible for the survey — *Go to section II*
- ☐ No — This place is not eligible for the survey — *Terminate the call*

CHECK ITEM D ▶ **Is this place a halfway house?**

- ☐ Yes — *Ask 5a*
- ☐ No — *Go to Check Item E*

CONTINUE ON REVERSE SIDE

Section I — IDENTIFICATION OF SPECIAL PLACE TYPE — Continued

1. Are the residents staying there voluntarily or involuntarily?	☐ Voluntarily — The residents are eligible for the survey — *Ask 5b* ☐ Involuntarily — These residents are not eligible for the survey — *Ask 5b* ☐ Both — Only voluntary residents are eligible — *Ask 5b*
2. Does this . . . *(type of place)* provide living quarters for staff members?	☐ Yes — These staff units and any voluntary residents are eligible for the survey — *Ask section II for these staff units and voluntary residents only.* ☐ No — *Ask section II for voluntary residents. If none, terminate the call.*
CHECK ITEM E ▶ If this Special Place is none of the above, ask section II for all units in the place. All units are eligible for listing.	*Go to section II*

Section II — LIVING ARRANGEMENTS

INTRODUCTION — For these next few questions, I am asking about (all units/staff units/civilian units/mobile homes, etc.) at this . . . *(read type of place)*.

1. Does this . . . have any *(such units)* occupied or intended to be occupied by five or more unrelated (persons/staff/civilians, etc.)?	☐ Yes — *Go to 3* ☐ No — *Ask 2*
2. How many *(. . . units)* are there in this . . .?	_____ Consider each as a separate unit when listing the units — *Go to section III*
3. What is the maximum number of persons that could live in the . . . *(units)* at this . . . *(read type of place)*?	_____ Consider each person or each bed a separate unit when listing the units — *Go to section III*

Section III — LISTING OF UNITS

- Compile on Form 11-213 a list of all eligible units at the special place. Refer to section I and table A to determine which units to list. Refer also to section II to determine what to consider as a separate unit.
- Complete section IV after listing the units.

Section IV — TELEPHONE ARRANGEMENTS

INSTRUCTION — When asking these questions, insert the appropriate phrase i.e. "staff unit," "civilian unit," " unit," etc. Also, be sure the respondent understands what constitutes a "unit" in this Special Place.

1. Does EACH . . . in this place have an individual telephone number?	☐ Yes — *Ask 2* ☐ No — *Go to 3*
2. Is the telephone number *(read number in heading item b)* used by one of the . . .?	☐ Yes ☐ No — *END INTERVIEW*
3. Does ANY . . . have its own telephone number?	☐ Yes ☐ No — *Ask 4*
4. Is the telephone number *(read number in heading item b)* used by any . . .?	☐ Yes — *Ask;* How many? _____ } *END INTERVIEW* ☐ No

Remarks

Section I — IDENTIFICATION OF SPECIAL PLACE TYPE — Continued

5a. Are the residents staying there voluntarily or involuntarily?

- ☐ Voluntarily — The residents are eligible for the survey — *Ask 5b*
- ☐ Involuntarily — These residents are not eligible for the survey — *Ask 5b*
- ☐ Both — Only voluntary residents are eligible — *Ask 5b*

b. Does this . . . *(type of place)* provide living quarters for staff members?

- ☐ Yes — These staff units and any voluntary residents are eligible for the survey — *Ask section II for these staff units and voluntary residents only.*
- ☐ No — *Ask section II for voluntary residents. If none, terminate the call.*

CHECK ITEM E ▶ If this Special Place is none of the above, ask section II for all units in the place. All units are eligible for listing.

Go to section II

Section II — LIVING ARRANGEMENTS

INTRODUCTION — For these next few questions, I am asking about (all units/staff units/civilian units/mobile homes, etc.) at this . . . *(read type of place)*.

1. Does this . . . have any *(such units)* occupied or intended to be occupied by five or more unrelated (persons/staff/civilians, etc.)?

- ☐ Yes — *Go to 3*
- ☐ No — *Ask 2*

2. How many *(. . . units)* are there in this . . .?

_____ Consider each as a separate unit when listing the units — *Go to section III*

3. What is the maximum number of persons that could live in the . . . *(units)* at this . . . *(read type of place)*?

_____ Consider each person or each bed a separate unit when listing the units — *Go to section III*

Section III — LISTING OF UNITS

- Compile on Form 11-213 a list of all eligible units at the special place. Refer to section I and table A to determine which units to list. Refer also to section II to determine what to consider as a separate unit.
- Complete section IV after listing the units.

Section IV — TELEPHONE ARRANGEMENTS

INSTRUCTION — When asking these questions, insert the appropriate phrase i.e. "staff unit," "civilian unit," " unit," etc. Also, be sure the respondent understands what constitutes a "unit" in this Special Place.

1. Does EACH . . . in this place have an individual telephone number?

- ☐ Yes — *Ask 2*
- ☐ No — *Go to 3*

2. Is the telephone number *(read number in heading item b)* used by one of the . . .?

- ☐ Yes
- ☐ No — *END INTERVIEW*

3. Does ANY . . . have its own telephone number?

- ☐ Yes
- ☐ No — *Ask 4*

4. Is the telephone number *(read number in heading item b)* used by any . . .?

- ☐ Yes — *Ask;* **How many?** _____
- ☐ No

} *END INTERVIEW*

Remarks

Appendix 8. Cost Analysis

I. Overview

The NHIS/RDD feasibility study took place from January to May of 1984. The study consisted of a two-stage random digit dialing sample, an automated case management and call scheduling system and a paper questionnaire. The survey was partitioned into 12 samples (replicates). Each replicate consisted of approximately 23 clusters of phone numbers from which 12 sample cases were selected for interview. Six cases in a cluster were to be interviewed using one version of the questionnaire and the other six were to be interviewed using an alternate version (Note: approximately 85 percent of the questions were the same on both versions and for this analysis, they were treated as the same). Beginning with the last week of January and for the next 11 weeks, a new replicate was introduced and remained in sample for three consecutive weeks. Therefore, for the majority of the survey, three replicates were active at any given time.

The primary focus of this analysis will be the secondary screening or interviewing phase of the data collection operation as this is the most sensitive component of cost in regards to sample size, length of interview, etc. There were two major sources of data for the analysis. The first was the case management file maintained throughout the survey. This accounted for interviewer 'on-line' time (i.e. the time the interviewer was logged on the system). The second source of data was the the interviewer payroll forms on which the interviewers allocated their time by activity. Because of varying salary scales, overheads, cost allocation methods, as well as the confounding of research activities, dollar amounts would be very misleading even if they were recoverable. Therefore, the cost related information is expressed in terms of time components.

II. Analysis

The data from the case management file can be analyzed by replicate. However, the data from the payroll files cannot be aggregated by replicate as in any given payroll week, cases from three replicates may have been active. Therefore comparisons made between on-line time and payroll time are only for the entire survey.

Table 1 illustrates the comparisons of average minutes per case between the two data files. The numerator of the rate (total minutes) is the same down a column. The numerator for the

1. column (1) rate is the facility's total on-line minutes for all cases.

2. column (2) rate is the total payroll minutes charged by the facility staff on all activities.

3. column (3) rate is the total payroll minutes charged to secondary screening only.

The denominator of the rate is the same across a row. The denominator for the

1. first row rate , INTERVIEWED HOUSEHOLDS, is the total number of cases that had a final outcome code of 1,2,3,5 or 6 (Note: TABLE 3 gives a description of each outcome code)

2. second row rate, POTENTIALLY ELIGIBLE CASES, is the total number of cases that had a final outcome code of 1,2,3,5,6,21,25,26 or 27.

3. third row rate, ALL PHONE NUMBERS, is all phone numbers attempted.

TABLE 1. COMPARISON OF CASE MANAGEMENT DATA TO PAYROLL DATA

CASES	(1) FACILITY ON - LINE MIN/CASE	(2) TOTAL PAYROLL MIN/CASE	(3) SECONDARY SCREENING PAYROLL MIN/CASE
INTERVIEWED HOUSEHOLDS	61	215	104
POTENTIALLY ELIGIBLE HOUSEHOLDS	47	165	80
ALL PHONE NUMBERS	26	90	44

Table 2 illustrates the average on-line minutes per case by replicate. (Note: column 1 of table 1 is the survey average for the rates in table 2). One may note a significant decrease in the minutes per interviewed households in later replicates whereas the average minutes per all cases remained relatively stable. As interviewers became more efficient at their job, the time per interview decreased. However, a larger percentage of cases in the later replicates resulted in completed interviews which take longer than non-interviews and thus the overall time is increased. Also, in earlier replicates, a case was never attempted more than 15 times before being retired as a non-interview. In later replicates, cases were attempted up to 20 times before being retired. This also would increase the amount of time per case. It appears, too, that more time was spent on refusal conversions in the later replicates.

TABLE 2. ON-LINE MINUTES PER CASE BY REPLICATE

	REPLICATE NUMBER											
CASES	1	2	3	4	5	6	7	8	9	10	11	12
* INTERVIEWED HOUSEHOLDS	68	66	67	69	68	70	59	59	58	49	54	52
** POTENTIALLY ELIGIBLE HOUSEHOLDS	46	45	48	53	51	52	47	47	48	40	44	43
ALL PHONE NUMBERS	31	26	25	28	28	26	24	23	23	26	24	26

* Cases with final outcomes of 1,2 3,5 or 6

** Cases with final outcomes of 1,2,3,5,6,21,25,26,27 or 29

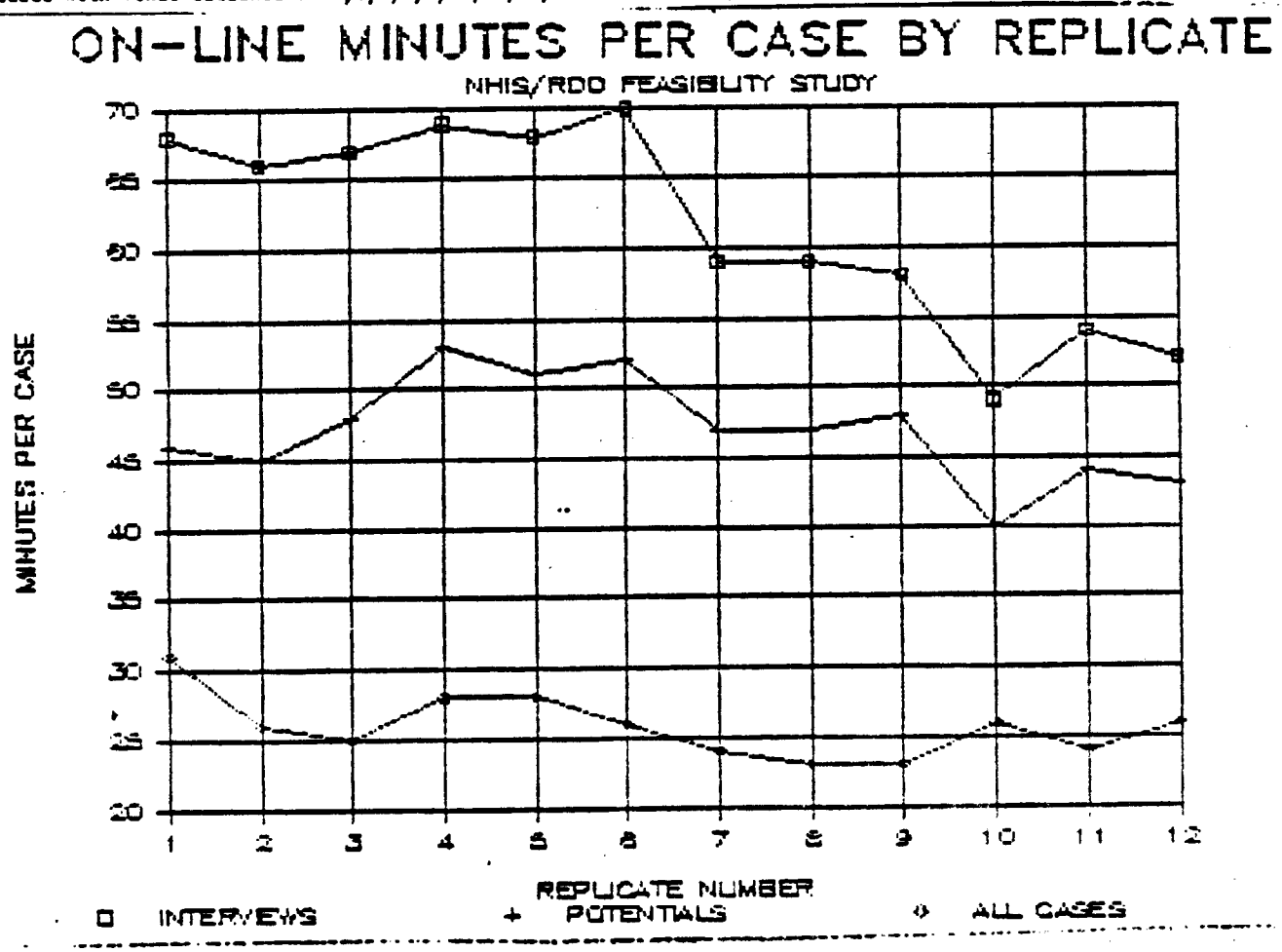


Table 3 examines the same data as in table 2 but further partitions the minutes per case by final outcome code assigned to the cases. The times represent the aggregated minutes for all dialings to a case.

TABLE 3 TOTAL MINUTES PER CASE BY FINAL OUTCOME CODE BY REPLICATE

FINAL OUTCOME CODE	DESCRIPTION OF FINAL OUTCOME CODE	REPLICATE # 1	2	3	4	5	6	7	8	9	10	11	12	AVG OF 1 - 12
1	COMPLETE INT W MKR	52	49	49	51	49	49	45	43	40	37	38	39	45
2	COMPLETE INT W/O MKR	62	36	52	61	76	37	46	44	43	54	49	37	53
3	COMPLETE THRU SEC W W MKR	78	75	84	60	101	46	62		59	65	61	174	73
5	PARTIAL W MKR	39	35	36	45	55	52	91	41	56	56	46	58	49
6	PARTIAL W/O MKR	48	41	23	19				28					33
7	SAMPLE REDUCTION DELETE				2	5				37				4
8	BANK DELETION			11	9	10	10		12	8		16		10
10	NON-WORKING/ # CHANGED	5	4	5	6	5	4	3	5	5	4	4	3	4
11	NONRESIDENTIAL PHONE	8	5	6	8	8	7	5	4	6	8	5	7	6
12	INELIG. RESIDENCE TYPE 1	6	28	30	33	22	21	16	9	11	16	15	6	18
13	INELIG. RESIDENCE TYPE 2	19	20	22	41	22	23	29	20	22	17	23	23	24
21	UNCONVERT. LANG. BARRIER				15	10	24	19	18	13	20	13	11	17
23	REFUSAL CUTOFF REACHED	19	19	18	24	23	31	20	32	36	33	33	33	25
26	SEARCH CUTOFF REACHED			18	33	31	37	48	30	29	22	33	23	30
27	TOTAL CUTOFF REACHED	38	30	36	33	46								33
29	CLOSEOUT CUTOFF REACHED	13	10	16	16	13	34	36	26	38	34	41	45	27
ALL CODES		31	26	25	28	28	26	24	23	23	26	24	26	26

Tables 4 and 5 show the distribution of cases by final outcome code first by number and then by percent for each replicate and for all replicates combined.

TABLE 4 TOTAL CASES BY FINAL OUTCOME CODE BY REPLICATE

FINAL OUTCOME CODE	DESCRIPTION OF FINAL OUTCOME CODE	REPLICATE # 1	2	3	4	5	6	7	8	9	10	11	12	SUM OF 1 - 12	% OF CASES
1	COMPLETE INT W MKR	204	206	207	212	208	173	185	199	233	241	245	265	2578	39.2
2	COMPLETE INT W/O MKR	11	6	5	11	9	5	7	8	4	2	5	3	76	1.2
3	COMPLETE THRU SEC H W MKR	3	10	5	6	5	1	3	0	7	5	4	1	50	0.8
5	PARTIAL W MKR	5	7	3	9	3	3	3	2	3	2	3	5	48	0.7
6	PARTIAL W/O MKR	2	2	1	2	0	0	0	1	0	0	0	0	8	0.1
7	SAMPLE REDUCTION DELETE	0	0	0	17	11	0	0	0	1	0	0	0	29	0.4
8	BANK DELETION	0	0	44	43	62	60	0	63	111	0	40	0	423	6.4
10	NON-WORKING/ # CHANGED	101	172	129	119	105	122	170	98	119	90	136	128	1489	22.6
11	NONRESIDENTIAL PHONE	54	59	86	80	50	47	51	89	68	44	62	73	763	11.6
12	INELIG. RESIDENCE TYPE 1	1	10	14	15	12	12	13	14	18	18	19	12	158	2.4
13	INELIG. RESIDENCE TYPE 2	9	12	12	10	14	5	12	3	13	10	11	12	123	1.9
21	UNCONVERT. LANG. BARRIER	0	0	0	6	5	11	4	2	5	9	6	2	50	0.8
25	REFUSAL CUTOFF REACHED	75	79	48	24	36	33	22	30	22	21	42	25	457	6.9
26	SEARCH CUTOFF REACHED	0	0	1	3	1	2	3	5	5	5	2	3	30	0.5
27	TOTAL CUTOFF REACHED	17	24	11	17	8	0	0	0	0	0	0	0	77	1.2
29	CLOSEOUT CUTOFF REACHED	12	8	28	20	24	19	20	16	24	18	9	22	220	3.3
ALL CODES		494	595	594	594	553	493	493	530	633	465	584	551	6579	100.0

TABLE 5 PERCENT DISTRIBUTION OF CASES BY FINAL OUTCOME CODE BY REPLICATE

FINAL OUTCOME CODE	DESCRIPTION OF FINAL OUTCOME CODE	REPLICATE # 1	2	3	4	5	6	7	8	9	10	11	12	AVG OF 1 - 12
1	COMPLETE INT W MKR	41.3	34.6	34.8	35.7	37.6	35.1	37.5	37.5	36.8	51.8	42.0	48.1	39.2
2	COMPLETE INT W/O MKR	2.2	1.0	0.8	1.9	1.6	1.0	1.4	1.5	0.6	0.4	0.9	0.5	1.2
3	COMPLETE THRU SEC H W MKR	0.6	1.7	0.8	1.0	0.9	0.2	0.6	0.0	1.1	1.1	0.7	0.2	0.8
5	PARTIAL W MKR	1.0	1.2	0.5	1.5	0.5	0.6	0.6	0.4	0.5	0.4	0.5	0.9	0.7
6	PARTIAL W/O MKR	0.4	0.3	0.2	0.3	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.1
7	SAMPLE REDUCTION DELETE	0.0	0.0	0.0	2.9	2.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.4
8	BANK DELETION	0.0	0.0	7.4	7.2	11.2	12.2	0.0	11.9	17.5	0.0	6.8	0.0	6.4
10	NON-WORKING/ # CHANGED	20.4	28.9	21.7	20.0	19.0	24.7	34.5	18.5	18.8	19.4	23.3	23.2	22.6
11	NONRESIDENTIAL PHONE	10.9	9.9	14.5	13.5	9.0	9.5	10.3	16.8	10.7	9.5	10.6	13.2	11.6
12	INELIG. RESIDENCE TYPE 1	0.2	1.7	2.4	2.5	2.2	2.4	2.6	2.6	2.8	3.9	3.3	2.2	2.4
13	INELIG. RESIDENCE TYPE 2	1.8	2.0	2.0	1.7	2.5	1.0	2.4	0.6	2.1	2.2	1.9	2.2	1.9
21	UNCONVERT. LANG. BARRIER	0.0	0.0	0.0	1.0	0.9	2.2	0.8	0.4	0.8	1.9	1.0	0.4	0.8
25	REFUSAL CUTOFF REACHED	15.2	13.3	8.1	4.0	6.5	6.7	4.5	5.7	3.5	4.5	7.2	4.5	6.9
26	SEARCH CUTOFF REACHED	0.0	0.0	0.2	0.5	0.2	0.4	0.6	0.9	0.8	1.1	0.3	0.5	0.5
27	TOTAL CUTOFF REACHED	3.4	4.0	1.9	2.9	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.2
29	CLOSEOUT CUTOFF REACHED	2.4	1.3	4.7	3.4	4.3	3.9	4.1	3.0	3.8	3.9	1.5	4.0	3.3
ALL CODES		100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 6 attempts to partition the on-line time into four phases. Time marks were set at critical locations in the CATI instrument so that time estimates could be constructed for various activities. The first time mark was activated when an interviewer requested a case. The second time mark was activated when an interviewer had completed reviewing any notes from prior attempts to the case obtained the partially completed questionnaire for the case (if one existed) and dialed the telephone number for the case. The third time mark was set when an interviewer departed from the monitor to begin the paper questionnaire. The fourth time mark was recorded when the interviewer returned to the monitor after working on the paper document. The fifth and final time mark was set when the interviewer completed the outcome coding and transcription items. The difference in time between each of these time marks were aggregated together for all dialings to a case and four component times were generated as follows:

1. The time between the first and second time marks is referred to as "Access to Dialing Time."

2. The time between the second and third time marks is referred to as "Screening Time."

3. The time between the third and fourth time marks is referred to as "Interviewing Time."

4. The time between the fourth and fifth time marks is referred to as "Transcription and Outcome Coding Time."

The time the interviewer spends recording any notes about the case is not included in any of the time categories but is included in the total time for the case as indicated in Table 3. This may explain why the sum of the four categories is less than the total time for the case.

Appendix 9. Monitoring

BACKGROUND/PURPOSE

As part of the NHIS/RDD Feasibility Study, professionals from both the Bureau and NCHS monitored a sample of live interviews to address a variety of questions that would be difficult to answer using objective (response rates, item nonresponse, cost, production, etc.) survey data. A list of these questions follows. Most of these questions reflect concerns about changing the HIS from a personal, face-to-face interview to a telephone interview.

1. Did the interviewer have difficulty identifying and obtaining an interview with the most knowledgeable respondent?

2. Which sections of the questionnaire, or individual items, were most troublesome to the interviewer, to the respondent, to both? Further, did problems vary by the version (2 or 3) of the questionnaire being tested?

3. Did the absence of flashcards cause problems?

4. Was respondent fatigue or frustration a problem?

5. How cooperative was the respondent?

6. How adequate was interviewer performance with respect to knowledge of the questionnaire, probing, answering questions from respondents, and following skip patterns?

7. How did interviewer performance vary during the course of the study?

To structure the monitoring, a special monitoring form was designed that addressed the preceding questions. This form was to be completed for each interview that was monitored. At the conclusion of this study, 151 forms were available for analysis. Unfortunately, after careful review of these forms, and discussions with persons who monitored, it was decided that the data were not usable. The reasons for reaching this conclusion are described in the next section of this report.

PROBLEMS WITH THE MONITORING DATA

Prior to the implementation of the monitoring project, plenary sessions were held with representatives from the Bureau and NCHS to discuss the content and structure of the monitoring form. As a result of these meetings, general guidelines were developed for completing the form, but no formal training of monitors was done. This informal approach apparently caused a variety of problems. For example, a partial list follows.

1. Monitors varied widely in both their knowledge of the health interview survey content and interviewing skills, in general. Therefore, anchor points on the structured rating scales were differentially defined. For example, to one monitor a "cooperative" respondent might have meant someone who completed the interview, but to another monitor, it might have meant someone who was merely nice, even if he refused to be interviewed.

2. Persons varying widely in background knowledge monitored most heavily at different points in the survey. For example, persons most knowledgeable about the content of the survey tended to monitor more heavily in the first half of the survey, than in the second half.

3. Persons monitored with widely different objectives. Although a structured monitoring form was used, individuals focused on different aspects of the survey. For example, some monitors were primarily concerned with refusals and why they occurred, some focused on respondent rules, others stressed general interviewing skills such as voice quality and style, and others focused on the content of the survey. The result was that relatively few monitoring forms were filled completely.

4. Monitors felt that their standards for judging interviewer performance changed during the course of the survey. Initially, some monitors reported that they compared the performance of the telephone interviewers with that of field interviewers. However, this standard was changed when the monitors realized that the telephone interviewers were not in the same "ball park," at least during the first half of the survey. Therefore, ratings shifted to compare telephone interviewers with each other, rather than with field interviewers.

Due to the preceding reasons, it was decided that any conclusions drawn from the available monitoring data would be misleading, rather than informative. Nevertheless, the experience was not a total loss. Some monitors felt that they benefited from the experience, even if their observations could not be summarized and compared statistically.

WHAT LESSONS WERE LEARNED?

A lesson which was learned in this study, and which is relevant to similar efforts, is that the objectives of the monitoring and statistical analyses planned must be clearly thought through before a monitoring effort is implemented. These objectives will define the technical constraints that must be satisfied to draw certain types of conclusions from the data. For example, in the NHIS/RDD, one of the most demanding objectives dealt with measuring interviewer change over time on a variety of dimensions. However, valid assessments of change over time require at least the following four conditions when raters and rating scales are used:

1. A rating instrument (or scale) with acceptable reliability and validity.

2. High interrater reliability, initially, and across time.

3. Invariant standards (criteria) for making ratings across time.

4. A sampling plan that is representative of interviewers, time slots, and intervals in the survey.

The difference between points two and three above should be clarified. For example, it is possible to have a rating scale with high reliability and validity. However, if the raters using it change their standards for assessing performance across time, the reliability of the instrument may remain high (i.e., interrater agreement will be high), but the ratings themselves will not be comparable across time because the standards have changed. This problem was identified in paragraph #4 of the "Problems With The Monitoring Data" section.

Another lesson learned was that monitoring done for research purposes, such as those reasons described earlier, should not be expanded to include quality control of interviewer performance. This did occur to a certain extent in the NHIS/RDD study with unwelcome consequences, such as interviewer resentment of the monitors. Quality control (QC) of interviewer performance is a primary task of supervisory staff and should remain totally their responsibility. However, in the feasibility study, a structured QC program was not implemented. Throughout the 16-17 weeks of interviewing (including practice interviewing of "live" cases), supervisors completed only eight monitoring forms. The limited amount of QC monitoring that occurred can be attributed to a lack of instruction in how to use the monitoring form, lack of supervisor input into the content of the monitoring form, lack of a sampling plan for conducting monitoring of interviewers, and competing supervisory responsibilities that were viewed as higher priority than monitoring.

Finally, another lesson was that any monitoring system should be totally unobtrusive. That is, the interviewer should not be aware that she or he is being monitored.

The next section of this report presents data that show major improvements in interviewer performance over time. These data support one point of consensus among the monitors - interviewer performance improved noticeably during the course of the survey. However, the improvements in data quality might also have been influenced by other changes made during the course of the survey. These changes are also described in the next section.

VARIATIONS IN INTERVIEWER PERFORMANCE DURING THE COURSE OF THE SURVEY

At the start of this survey, the quality of interviewing, as judged by response rates and the observation reports of monitors, was considered to be far below the quality of personal-visit interviewing for the HIS. To remedy this situation, several steps were taken. These included the following:

1. Retraining of interviewers on important HIS concepts, procedures, and techniques for introducing the survey.

2. Changes in management approaches, including the use of formal (quality circle) and more informal group meetings, and increased supervisory training.

3. Increased emphasis by supervisors on refusal-conversions, plus increased efforts to communicate to interviewers the importance of the survey and uses of the data.

4. Dropping one interviewer who was not meeting performance standards.

5. Changes in the introductory statements in response to suggestions from interviewers.

Although the individual effect of these changes cannot be assessed, the cumulative effect, in combination with improved interviewing skills as a result of job experience, obviously led to dramatic improvements in response rates and item nonresponse. Anthony Roman has presented detailed results about response rates in another memorandum, but Figure 1 on the next page 2/ graphically shows the improvement in response rates averaged over each third of the survey period.

Item nonresponse for a series of critical health questions also showed sizeable decreases. These changes are shown in Table 1.

2/ Appreciation is expressed to Al Lago (Field) for his assistance in preparing all the graphics shown in this report.

Figure 1
Response Rates for the Questionnaire Variants
Averaged Over Each Third of the Survey Period

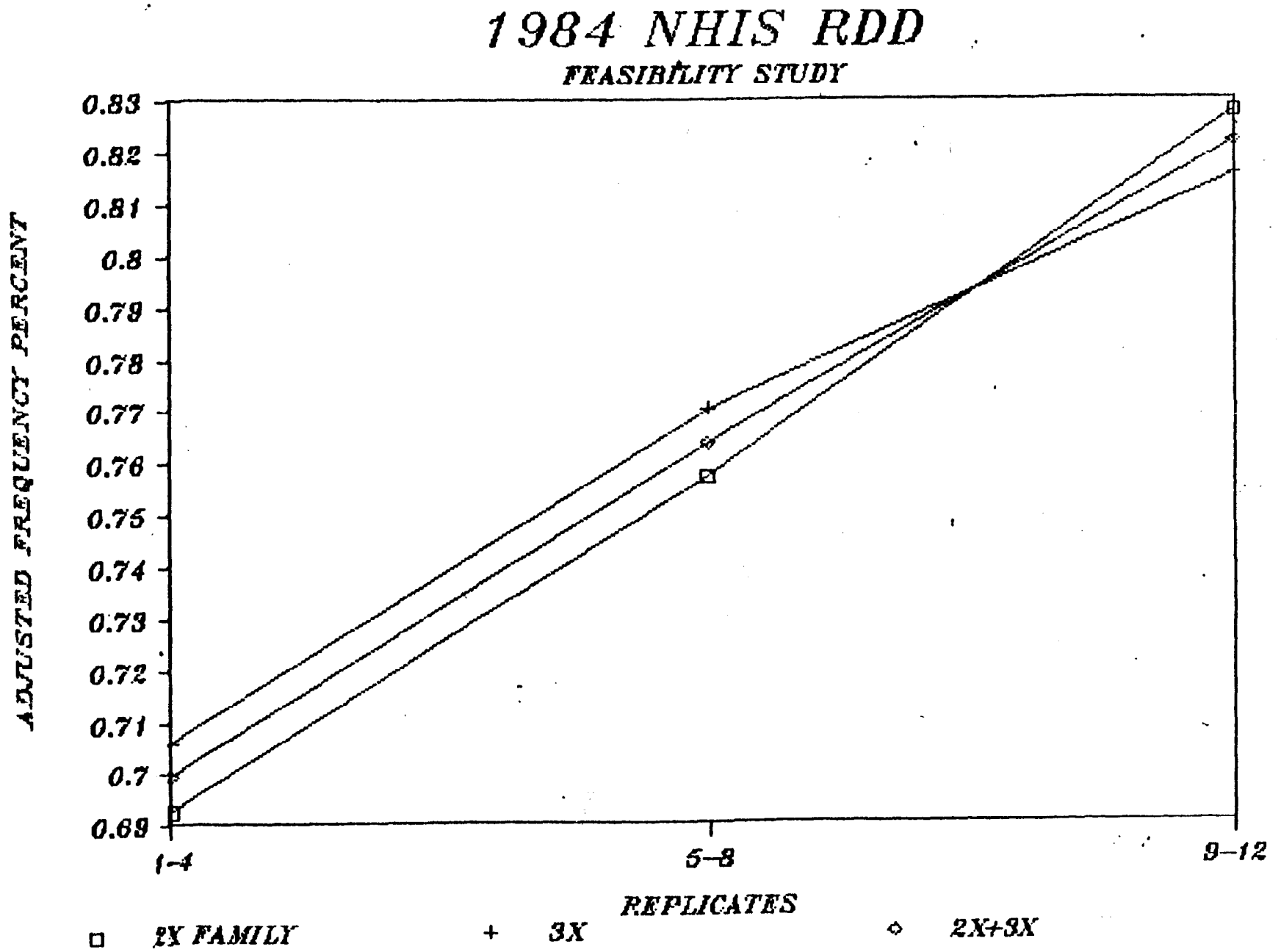


Table 1
Changes in Item Nonresponse (percent) For Major
Health Questions for the Combined Family Style and Person Style Questionnaires

Variable	Reps 1-4	Reps 5-8	Reps 9-12
Race	9.6	6.1	5.6
Age	1.6	1.0	0.9
Usual Activity	4.8	3.0	3.2
Education	8.4	5.6	4.3
Marital Status	7.5	5.0	3.9
2-week bed days	5.4	2.9	2.7
2-week cut-down days	6.1	4.0	2.9
12-month doctor visits	7.9	5.9	4.8
12-month bed days	5.0	3.9	2.7
2-week doctor visits	5.7	5.0	3.0
Health status	6.4	3.5	1.9
Total Conditions	4.4	1.4	0.7

Obviously an important question, but one that could not be addressed by the design of this study, deals with the effect of nonsampling error on the distributions of data obtained. It appears reasonable to conjecture that the interviewer component of nonsampling error varied during the course of this survey. It also seems reasonable to assume that in the initial replicates of this survey, the interviewer component of nonsampling error would be more random in nature, and possibly contribute to more variability, since interviewer inexperience should have its greatest effects on statistical estimates. However, as Table 2 shows below, the standard deviations of some critical estimates tended to increase during the course of the survey, and in no case were they lower than the average of the first four replicates.

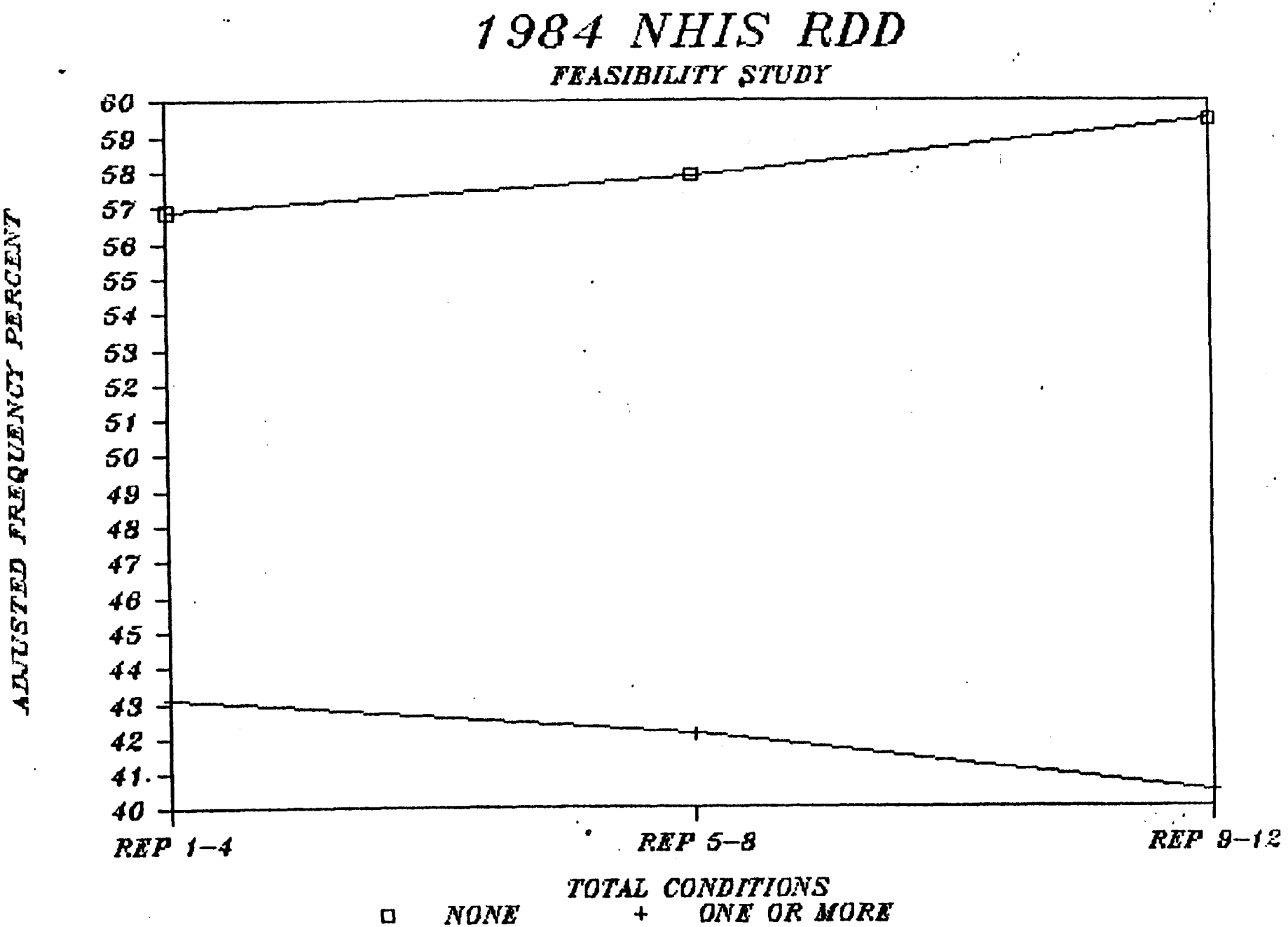
Table 2
Changes in Mean Estimates And Standard Deviations
For the Combined Family Style and Person Style
Questionnaires for Selected Variables

	Reps 1-4		Reps 5-8		Reps 9-12	
Variable	$\bar{X}$	S.D.	$\bar{X}$	S.D	$\bar{X}$	S.D
2-week bed days	.278	1.222	.322	1.628	.231	1.300
2-week cut-down days	.267	1.362	.320	1.648	.315	1.693
12-month doctor visits	3.137	5.575	3.519	7.718	3.294	7.791
12-month bed days	3.925	15.949	5.281	25.299	4.241	19.579
2-week doctor visits	.248	.719	.249	.862	.245	.732

A possible explanation for the increased variability shown in Table 2 is that response rates were significantly higher in later replicates. One could assume that more experienced interviewers would succeed in interviewing more difficult and disparate households. If these households also differed in other ways, for example, in their types of health problems, then their inclusion in later replicates would result in greater variability in statistical estimates.

Distributions of data for some major health variables are shown in Tables 1A to 1L in the Attachment to this report. Of these, one of the more interesting is Table 1L. These data seem to indicate that the number of interviews in which "none" total conditions were reported increased as a function of increased job experience. This trend is illustrated in Figure 2 (on the next page), and could indicate a reluctance of interviewers to record health conditions (which must then be reported upon in great detail).

Figure 2
Adjusted Frequency Percent of the Number
of Total Health Conditions Reported in the NHIS/RDD



Attachment 1 [3]

Response Distributions for the Combined
Family Style and Person Style Questionnaires

Table 1A
Race

	Reps 1-4	Reps 5-8	Reps 9-12	Family-Style 1-12 [4]
White	80.8 (89.4)	75.8 (83.9)	85.0 (90.1)	83.1 (88.4)
Non-White	9.6 (10.6)	15.1 (16.1)	9.3 (9.9)	10.9 (11.6)
Nonresponse	9.6	6.1	5.6	5.9

Table 1B
Age

	Reps 1-4	Reps 5-8	Reps 9-12	Family-Style 1-12
17-24	14.8	17.4	16.8	17.0
25-34	25.2	20.9	25.7	23.9
35-44	18.2	19.0	17.5	18.6
45-54	12.8	15.0	13.8	14.2
55-64	12.4	12.8	10.9	11.2
65-74	9.4	9.4	9.8	9.2
75 +	5.7	4.5	4.7	4.7
Nonresponse	1.6	1.0	0.9	1.2

Table 1C
Usual Activity

	Reps 1-4	Reps 5-8	Reps 9-12	Family-Style 1-12
Working	54.4	57.8	57.4	56.3
Keeping house	23.9	21.6	21.7	23.0
Going to school	6.1	6.2	7.1	7.3
Something else	10.8	11.4	10.7	9.9
Nonresponse	4.8	3.0	3.2	3.6

3/ Adjusted frequency percent (nonresponse cases excluded) is shown in parentheses.

4/ Data for the combined family- and person-style questionnaires over all 12 replicates were not available.

Table 1D
Education

	Reps 1-4	Reps 5-8	Reps 9-12	Family-Style 1-12
Under 12	20.6	22.8	20.4	21.3
12	41.4	42.9	42.5	42.9
Over 12	29.5	28.7	32.8	30.4
Nonresponse	8.4	5.6	4.3	5.4

Table 1E 5/
Marital Status

	Reps 1-4	Reps 5-8	Reps 9-12	Family-Style 1-12
Married	(62.9)	(61.0)	(63.4)	(63.3)
Widowed	(7.0)	(5.4)	(5.8)	(6.0)
Divorced	(6.1)	(5.8)	(6.4)	(5.5)
Separated	(1.3)	(1.9)	(1.4)	(1.5)
Never Married	(22.8)	(25.8)	(23.0)	(23.7)
Nonresponse	7.5	5.0	3.9	5.0

Table 1F
Two-Week Bed Days

	Reps 1-4	Reps 5-8	Reps 9-12	Family-Style 1-12
None	85.4 (90.3)	89.2 (91.9)	91.4 (93.9)	89.7 (92.5)
1-3	7.2 (7.6)	5.3 (5.5)	4.2 (4.3)	5.3 (5.4)
4-7	1.5 (1.6)	1.4 (1.4)	0.9 (1.0)	1.3 (1.3)
8-10	0.2 (0.2)	0.1 (0.1)	0.3 (0.3)	0.2 (0.2)
11-14	0.4 (0.4)	1.1 (1.1)	0.5 (0.5)	0.6 (0.6)
Nonresponse	5.4	2.9	2.7	
Mean	.278 ± 1.222	.322 ± 1.628	.231 ± 1.300	.259 ± 1.342

5/ Marital status was not asked of everyone. Frequencies shown exclude ineligible and nonresponse cases.

Table 1G
Two-Week Cut-Down Days

	Reps 1-4	Reps 5-8	Reps 9-12	Family-Style 1-12
None	87.5 (93.2)	90.1 (93.9)	91.7 (94.4)	91.1 (94.7)
1-3	4.2 (4.5)	2.9 (3.1)	2.8 (2.9)	3.1 (3.2)
4-7	1.5 (1.6)	1.7 (1.7)	1.3 (1.3)	1.2 (1.3)
8-10	0.1 (0.1)	0.4 (0.4)	0.2 (0.2)	0.2 (0.2)
11-14	0.6 (0.6)	0.9 (0.9)	1.2 (1.2)	0.7 (0.7)
Nonresponse	6.1	4.0	2.9	3.8
Mean	.267 ± 1.362	.320 ± 1.648	.315 ± 1.693	.241 ± 1.376

Table 1H
Twelve-Month Doctor Visits

	Reps 1-4	Reps 5-8	Reps 9-12	Family-Style 1-12
None	24.0 (26.1)	26.2 (27.9)	24.5 (25.8)	25.0 (26.6)
1	22.7 (24.7)	22.2 (23.7)	23.3 (24.5)	23.3 (24.7)
2-4	28.3 (30.8)	27.3 (29.1)	28.9 (30.5)	27.8 (29.5)
5-12	13.6 (14.8)	13.9 (14.8)	14.8 (15.6)	14.6 (15.5)
13-24	2.0 (2.2)	2.3 (2.4)	2.4 (2.5)	2.2 (2.3)
25-52	1.1 (1.2)	1.7 (1.8)	0.7 (0.8)	1.1 (1.2)
53 +	0.1 (0.1)	0.3 (0.3)	0.3 (0.3)	0.2 (0.2)
Overnight in Hospital	0.3	0.2	0.4	0.3
Nonresponse	7.9	5.9	4.8	5.4
Mean	3.137 ± 5.575	3.519 ± 7.718	3.294 ± 7.791	3.194 ± 6.088

Table 1I
Twelve-Month Bed Days

	Reps 1-4	Reps 5-8	Reps 9-12	Family-Style 1-12
None	47.2 (50.8)	47.7 (50.4)	49.0 (51.3)	49.3 (50.2)
1-7	37.2 (39.9)	37.3 (39.4)	37.5 (39.3)	37.4 (38.0)
8-30	6.9 (7.4)	7.4 (7.8)	7.4 (7.7)	6.9 (7.0)
31-180	1.6 (1.7)	1.9 (2.0)	1.3 (1.4)	1.5 (1.6)
180 +	0.1 (0.1)	0.5 (0.5)	0.3 (0.3)	0.3 (0.3)
Nonresponse	5.0	3.9	2.7	3.0
Mean	3.925 ± 15.949	5.281 ± 25.299	4.241 ± 19.579	4.318 ± 19.981

Table 1J
Two-Week Doctor Visits

	Reps 1-4	Reps 5-8	Reps 9-12	Family-Style 1-12
None	79.2 (84.0)	80.5 (84.8)	81.9 (84.5)	81.3 (84.5)
1-3	14.3 (15.2)	13.4 (14.1)	14.1 (14.5)	14.1 (14.6)
4-7	0.7 (0.8)	0.8 (0.9)	0.9 (0.9)	0.8 (0.8)
8-10	0.1 (0.1)	- -	0.1 (0.1)	0.1 (0.1)
11-14	- -	0.2 (0.2)	- -	0.1 (0.1)
15 +	- -	- -	- -	- 0.0
Nonresponse	5.7	5.0	3.0	3.7
Mean	.248 ± .719	.249 ± .862	.245 ± .732	.239 ± .751

Table 1K
Health Status

	Reps 1-4	Reps 5-8	Reps 9-12	Family-Style 1-12
Excellent	36.9 (40.2)	37.1 (39.0)	41.1 (42.6)	37.5 (39.6)
Very Good	27.6 (30.2)	26.0 (27.3)	26.6 (27.6)	26.9 (28.4)
Good	19.3 (21.0)	21.3 (22.4)	19.7 (20.4)	20.8 (22.0)
Fair	6.0 (6.6)	7.5 (7.9)	6.3 (6.5)	7.1 (7.5)
Poor	1.8 (2.0)	3.2 (3.3)	2.8 (2.9)	2.4 (2.5)
Nonresponse	6.4	3.5	1.9	3.7

Table 1L
Total Conditions

	Reps 1-4	Reps 5-8	Reps 9-12
None	53.2 (56.9)	56.3 (57.9)	58.0 (59.5)
One or More	40.4 (43.1)	40.9 (42.1)	39.6 (40.5)
Nonresponse	4.4	1.4	0.7

Appendix 10. Intracluster Correlations

1. Introduction

For several years, the Census Bureau has examined random digit dialing (RDD) to determine its feasibility as a method of surveying populations of interest. For most of the Census Bureau surveys to which RDD could be applied, RDD would be a component of a dual-frame system in which an area sample would be used to cover non-telephone households. The purpose of this analysis is to estimate for clustered RDD designs the intracluster correlations which would be used along with cost estimates to compute optimum telephone cluster sizes either for the dual-frame or for a single-frame RDD survey. Intracluster correlation estimates and their estimated variances from the National Health Interview Survey/Random Digit Dialing (NHIS-RDD) Feasibility Study are presented in this report. The calculated correlations were used to obtain optimum cluster sizes that are recommended for use in any future applications of RDD to the NHIS survey. These optimum cluster sizes are given in Section 5.

2. Background

The NHIS-RDD feasibility study consisted of the selection of about 1500 households for each of two versions of the questionnaire. Twelve replicates, each consisting of about 126 households per questionnaire version, were selected and interviewed over a three-week period. Beginning in late January of 1984, a new replicate was introduced each week. The random digit dialing procedure used for this study was the one introduced by Waksberg (1978). The details of the sample design are given in Biemer (1983). Briefly, a six-digit area code-exchange number was selected from an AT&T list of working telephone

exchanges. A two-digit random number was attached to this number to form a PSU or bank of 100 numbers. PSUs were screened by selecting at random one of its 100 numbers and calling the number to determine whether or not it was residential. If the number called was residential, the PSU was labeled "residential" and was retained for the sample. For each residential PSU, 12 telephone numbers from the 100-bank were called, six for each questionnaire type.

Intracluster correlations were calculated for twelve health information variables and three demographic variables for each of the twelve replicates and two questionnaire versions. Correlations were also calculated (1) for each variable and each replicate for the two questionnaire versions combined, (2) for each variable and each questionnaire version for the twelve replicates combined, and (3) for each variable and all households in the survey. The three demographic variables of interest follow:

Household income
Age of the reference person
Education level of the reference person

Values for each of the following twelve health variables were totaled for the entire household:

Number of hospital stays
Number of 2-week doctor visits
Total number of conditions reported
Total number of reported conditions on the condition list
Number of limitations
Number of work-loss days
Number of school-loss days
Number of bed days
Number of cut-down days
Number of 12-month bed days
Number of 12-month doctor visits
Total number of nights in the hospital

3. Computation of the Intracluster Correlations

The equation that was used for calculation of intracluster correlations was derived by S. Lynne Stokes (1983) for the Random Digit Dialing Employment

and Health Survey (RDD I). The estimate of ρ is

$$\hat{\rho} = \frac{(m\, v(\hat{Y})/N^2) - s_2^2/\bar{k}}{(m\, v(\hat{Y})/N^2) + s_2^2\,[\bar{k}(N+M') - N]/(N\bar{k})} \qquad (1)$$

where

m = the number of PSUs selected for the sample (which was 250 for the entire sample and about 21 PSUs for each replicate),

N = the number of residential phone numbers in the universe (which is approximately 78.6 million),

M' = the number of PSUs in the universe that have at least one residential telephone (which is about 1.94 million),

$\bar{k}$ = the average number of telephone residences interviewed per PSU = $\sum_{i=1}^{m} k_i/m$,

k_i = the number of telephone residences interviewed in the i-th PSU,

$$v(\hat{Y}) = N^2 \sum_{i=1}^{m} (\bar{y}_i - \bar{\bar{y}})^2/m(m-1)$$

$$s_2^2 = \sum_{i=1}^{m} \sum_{j=1}^{k_i} (y_{ij} - \bar{y}_i)^2/m(\bar{k} - 1),$$

$$\bar{y}_i = \sum_{j=1}^{k_i} y_{ij}/k_i,$$

y_{ij} = the value of the Y-characteristic for the j-th respondent selected from PSU i, and

$$\bar{\bar{y}} = \sum_{i=1}^{m} \bar{y}_i/m.$$

Correlations were calculated for each replicate and questionnaire version combination for the 15 variables given in Section 2. For each of the 15 variables, correlations for all replicates combined by questionnaire version and for both questionnaire versions combined by replicate as well as correlations for the entire sample were also calculated.

The variances of the overall correlations for each variable and the correlations by questionnaire type were estimated using the random groups method. The 12 replicates were used as the random groups. The random groups estimate of $Var(\hat{\rho})$ is given below.

$$v(\hat{\rho}) = \sum_{h=1}^{12} (\rho_h - \rho)/(12 \cdot 11) \qquad (2)$$

where $\hat{\rho}_h$ = the estimated intracluster correlation from replicate h.

$\hat{\rho}$ = the estimated intracluster correlation from the entire sample.

4. Results

The estimated intracluster correlations by replicate for both questionnaire types combined and the random group estimates of variances are given in Table 1. Due to nonresponse and occasional misunderstandings regarding the assignment of questionnaire versions to households, cluster sizes by questionnaire version were often too small to give reliable estimates of the correlations. Therefore, correlations by questionnaire version were eliminated from the analysis and are not given in Table 1.

The estimated correlations on the demographic variables for all replicates combined were .095 for income, .110 for education, and .153 for age. Estimated correlations for health variables for all replicates combined were all between 0.00 and 0.03. The relatively high intraclass correlation for age may be a result of the way telephone numbers are assigned. In areas other than large metropolitan areas, it is often true that a person can move several times in a rather large geographic area and still keep the same telephone number. If a person does not move out of his/her county, he/she could keep the same telephone number for a lifetime. New 100-banks of telephone numbers are filled largely with young adults getting their first

Table 1: Intracluster Correlations and Variances by Replicate for the Variables of Interest

Variable	Replicate 1	2	3	4	5	6	7	8	9	10	11	12	All Reps	Var. (x.001)
Income	.004	.160	.104	.143	.047	.018	.115	.102	.136	.032	.240	.143	.095	.398
Age	.135	.163	.259	.225	.041	.202	.123	.164	.174	.029	.136	.112	.152	.383
Hosp. stays	.012	.002	.013	.032	.018	.024	.023	-.018	.026	.012	.010	.041	.017	.020
Doctor visits	-.012	.063	.042	.089	.088	.004	.007	.048	.021	.009	.011	-.009	.033	.105
Total cond.	-.045	.032	-.023	.086	-.010	.051	.023	.023	-.008	-.032	.012	.014	.015	.114
Cond. list	-.043	-.008	-.033	.019	.007	.049	.090	.002	-.008	-.045	-.012	.053	.009	.142
Limitations	-.017	.016	.004	-.014	-.027	.092	-.018	-.001	-.006	-.033	.050	.015	.000	.107
Work-loss days	-.011	.121	-.037	.065	-.010	-.026	.023	.024	-.012	-.020	.008	-.025	.003	.175
School-loss	.045	-.033	.151	.090	-.003	.015	.003	.004	-.019	.035	.012	-.014	.019	.226
Bed days	-.014	.058	.004	.026	.016	-.037	-.053	.062	-.028	-.005	.071	-.025	.011	.141
Cut-down days	-.019	.034	.024	.017	.005	-.013	-.048	-.018	-.006	.009	.029	-.007	-.001	.047
12-mo. bed days	.003	.050	-.006	.135	.027	.005	-.045	-.018	.006	.046	-.018	.013	.019	.176
12-mo. dr. visits	.055	.014	.015	.072	.019	.020	.004	.040	-.047	-.028	-.003	.023	.017	.089
Nights in hosp.	.042	.064	-.005	.052	.028	.055	-.004	-.017	.056	-.002	.043	-.006	.030	.077
Educ. level	.135	.081	.101	.108	.107	.101	.102	.171	.111	.132	.120	.090	.110	.048

telephones. If movement out of the geographic area covered by the exchange is rare, a majority of the people in a PSU or 100-bank of telephone numbers may be about the same age. To a lesser extent, education level and income would be affected by the same phenomenon.

5. Calculation of Optimum Cluster Sizes

The formula used to compute the optimum cluster size, $\bar{k}^*$, is that given in equation 2.6 of Stokes (1983). The cost model that she developed in her analysis treats all primary screening calls as unproductive. The formula for $\bar{k}^*$ is repeated below:

$$\bar{k}^* = \left[\frac{(1-\rho)}{\rho}\right]^{1/2} [\pi\, C_p/C_u + (1-t-\pi)]^{-1/2} \qquad (3)$$

where

π = the proportion of residential telephone numbers in all 100-banks combined,

C_p/C_u = the ratio of the cost of a productive call to that of an unproductive call,

t = the proportion of 100-banks that have no residential telephone numbers,

ρ = the intracluster correlation between two households in a PSU.

The cost of a productive call, C_p, was estimated by the sum of the average on-line and the average off-line costs of a productive case. The cost of an unproductive call, C_u, was estimated by the same sum for unproductive cases. The ratio, C_p/C_u, as estimated from the NHIS-RDD follows:

$$\frac{C_p}{C_u} = \frac{42.15 + 27.90}{9.40 + 25.71} = 2.00$$

The value of π was estimated to be .24 from the NHIS-RDD data. In RDD I, t was estimated to be between .57 and .65. The optimum cluster sizes, $\bar{k}^*$, are given in Table 2.

Table 2--Optimum Cluster Sizes

ρ	t .57	t .65
.15	3	4
.10	4	4
.05	6	6
.03	7	8
.02	9	10
.01	13	13
.005	18	19
.001	39	42

6. Conclusion

Table 2 shows that a cluster size of 4 residential telephone numbers is approximately optimum for measuring demographic variables since the estimated intraclass correlations for those variables were all at least .10. The approximate optimum cluster size is at least 20 residences for measuring the number of limitations of household members, the total number of household work-loss days, and the total number of cut-down days for the household. For all other health variables, cluster sizes of 8 to 13 residential telephone numbers are optimum.

Acknowledgement

We acknowledge the programming support of Chris Lashman of the Statistical Research Division of the Census Bureau.

Bibliography

Biemer, Paul (1983). "Final Design of the 1984 NHIS-RDD Feasibility Study," attached to a memorandum from Owen Thornberry and James Massey dated September 20, 1983, National Center for Health Statistics, Washington, D.C.

Stokes, S. Lynne (1983). "Optimization of Waksberg's RDD Sample Design," memorandum for Kent Marquis, Chief, Center for Survey Methods Research, Bureau of the Census, Washington, D.C.

Waksberg, Joseph (1978). "Sampling Methods for Random Digit Dialing," Journal of the American Statistical Association, 73, 40-46.

Appendix 11. 1984 NHIS/RDD Feasibility Study Survey Design Description

The following describes in detail the survey design and operational activities associated with the Feasibility Study. This description was extracted from a longer report (intended for internal circulation only) written in October 1984 by Janis L. Brown and R. Robert Wilson, entitled "Summary Report--1984 NHIS/RDD Feasibility Study."

Page

I. INTRODUCTION

The costs of conducting sample surveys using traditional methods are rising and continue to place a heavy burden on sponsor's budgets. Random Digit Dialing (RDD) techniques hold potential for substantial cost savings over current list/area frame sampling and face-to-face interviewing, but questions arise about the quality of estimates based on RDD because of potentially large coverage biases, higher noninterview rates, and possible overall lower measurement quality due to respondent, interviewer, questionnaire, and procedure differences for the face-to-face and telephone interviewing modes.

The National Center for Health Statistics (NCHS) and the Bureau of the Census agreed to undertake a research and development program leading to a dual frame National Health Interview Survey (NHIS) in 1986. A Steering Committee was established to guide development and implementation and a Task Force was created to recommend the research and development activities needed to achieve the objectives of the program. Steering Committee and Task Force members are listed below:

Steering Committee Members

National Center for Health Statistics:
- Earl Bryant
- Monroe Sirken

Bureau of the Census:
- Barbara Bailar
- William Butz

Task Force Committee Members

National Center for Health Statistics:
- Robert Fuchsberg, co-chair
- Robert Casady
- James Massey
- Owen Thornberry

Bureau of the Census:
- Kent Marquis, co-chair
- Paul Biemer
- Richard Blass
- Charles Jones
- Robert Mangold
- William Nicholls

The approach recommended for resolving many of the methodological and operational issues was to perform developmental work through a feasibility study in 1984. This document describes many aspects of the survey design used for the Feasibility Study.

F. Develop and evaluate procedures for identifying and handling special places over the telephone.

G. Preliminary development and testing of estimation procedures, including nonresponse and post-stratification adjustments.

H. Test procedures for the assignment, management, and completion of samples for producing valid estimates.

I. Evaluate the operational feasibility and effect on response rates of using a "most knowledgeable respondent" rule.

B. Interviewing Schedule - The interview periods for REP 0 and each of the 12 production REPs are noted below:

1/	REP 0	January 16-21
	Down Week	January 23-28
2/	REP 1	January 30 - February 4
	REP 1	February 6-25
	2	February 13 - March 3
	3	February 21 - March 10
	4	February 27 - March 17
	5	March 5-24
	6	March 12-31
	7	March 19 - April 7
	8	March 26 - April 14
	9	April 2-21
	10	April 9-28
	11	April 16 - May 5
	12	April 23 - May 12

1/ REP 0 was a shakedown/throwaway REP which was used to train interviewers using live cases. No data from this REP were used in the final analysis reports for the study.

2/ REP 1 which began January 30 was thrown out when it was discovered that many of the clusters never received primary screening. No data were used in the final analysis reports for cases completed from January 30 through February 4. REP 1 was re-introduced February 6 and included all new sample cases. This caused the Feasibility Study interviewing to be extended to May 12 rather than May 5 as planned in the schedule of operations (section III, paragraph A on the previous page).

IV. SAMPLE DESIGN

This section provides a brief summary of the sample design for the Feasibility Study. Additional details are given in Section 3 of the main report.

The sample was selected using the Waksberg procedure. The sample design called for 3,024 households to be assigned for interview during the survey period. The total case load was divided into 12 replicates (REPs), each consisting of 252 cases. Three weeks were allowed for interviewing each REP and a new REP was introduced into the study each week, therefore, interviewing periods for REPs overlapped.

The original and revised screening questions and introductions follow:

ORIGINAL VERSION	REVISED VERSION
1a. Hello. I'm (name) from the United States Bureau of the Census in Washington, D.C. We are conducting a survey for the U.S. Public Health Service to collect information about the nation's health. To make sure I have dialed correctly, is this (telephone number) in area code (area code)? ☐ Yes(2) ☐ No(1b) ☐ DK(1b)	1a. Hello, I'm (name) from the United States Bureau of the Census in Washington, D.C. We are conducting a health survey for the United States Public Health Service. Is this (telephone number) in area code (area code)? ☐ Yes(2) ☐ No(1b) ☐ DK(1b)
b. What number have I reached? ☐ Same as selected number (2) ☐ Refused } (TERMINATE CALL) ___ _____ Area Code Number	b. What number have I reached? ☐ Same as selected number (2) ☐ Refused } (TERMINATE CALL) ___ _____ Area Code Number
2. Have I reached you on your home phone? ☐ Yes(5) ☐ No(3)	2. Have I reached you on your home phone? ☐ Yes(5) ☐ No(3)
3. Have I reached you on a business phone or something else? ☐ Business(4) ☐ Other - Specify→ _____(4)	3. Does this telephone number serve a business or does it serve some place where people could live, such as a dormitory, hotel, and so forth? ☐ Business(4) ☐ Other - Specify→ _____(4)

6a. This survey is authorized by the Public Health Service Act. The results of the survey will be used for statistical research on health problems and all information you give will be kept confidential. Of course, your help on this survey is voluntary, but it is important that you and everyone selected for our survey participate so that we can make accurate estimates on the nation's health. In order to evaluate my performance, my supervisor may listen in.

(READ 6b)

6a. This survey is authorized by the Public Health Service Act. The results of the survey will be used for statistical research on health problems and all information you give will be kept confidential. Your voluntary participation is extremely important to help us obtain complete and accurate results.

(PAUSE)

In order to evaluate my performance, my supervisor may listen in.

(READ 6b)

---

b. Since it is very important to have good answers to the health questions I will be asking, I would like you to think carefully about each question before answering, even those questions which may seem unimportant to you.

(GO TO HOUSEHOLD COMPOSITION PAGE OF COVER BOOKLET)

b. I will ask about hospital stays, visits to doctors, illness in the family, and other health related items. Since some questions won't apply to everyone, first I'll need to ask a few questions about the people living in your household.

Household Composition Page:

In order to ask you the appropriate health questions, first I'll need to ask a few questions about the people living in your household.

(CONTINUE WITH NHIS CONTENT INTERVIEW)

NOTE: Interviewers were instructed to omit the explanation on the Household Composition Page since the revised version of item 6b contained this.

(CONTINUE WITH NHIS CONTENT INTERVIEW)

B. SURVEY FORMS--An explanation of the questionnaires and other forms used in the study follows. The Census division responsible for each form is noted in parentheses after the description.

1. Questionnaire Forms--In order to expedite printing of the questionnaires and to simplify data processing in Jeffersonville, a cover booklet which was separate from the insert booklet was used. Two versions of the cover booklet were used which differed only by the wording of the income question. The different methods of asking income--question 3 on page 1 of the cover booklet--were tested to determine if response/nonresponse to this item differed between the methods. One approach asked directly for income (version THIS-2a) and, if there was not an acceptable response to this question, then asked if the amount was more or less than $20,000. The second approach (version THIS-3a) was basically the reverse--the first question asked if the income was more or less than $20,000, then the next question asked for the approximate amount.

Two versions of the insert booklet were also tested--the Family/Individual version and the Person by Section version. The Family/Individual version closely resembled the HIS-1 questionnaire used in the personal visit NHIS. Based on results of a telephone study conducted by the University of Michigan's Survey Research Center (SRC) in 1979 using a modified NHIS questionnaire, it was hypothesized that the person-by-person style of the SRC questionnaire was responsible for producing higher than expected levels of reporting of certain health characteristics for telephone interviews in comparison to the family style version used in the 1979 NHIS. The Feasibility Study used two insert booklet versions to test this hypothesis.

- Health Indicator Page - Question 1:

THIS-2 version	THIS-3 version
1a. During the 2-week period from Monday, (date) to Sunday, (date), has anyone in the family had an injury from an accident or other cause that you have not yet told me about?	1a. During the 2-week period from Monday, (date) to Sunday, (date), has -- had any injury from an accident or other cause that you have not yet told me about?
b. Who was this?	b. What was the injury?
c. What was -- injury?	c. Did -- have any other injuries during that period?
d. Did anyone have any other injuries during that period?	d. As a result of the (injury in 1b) did [--/anyone] see or talk to a medical doctor or assistant (about --) or did -- cut down on -- usual activities for more than half of a day?
e. As a result of the (injury in 1c) did [--/anyone] see or talk to a medical doctor or assistant (about --) or did -- cut down on -- usual activities for more than half of a day?	

THIS-2 version	THIS-3 version
• 2-Week Doctor Visits Probe Page:	
- Read introduction.	- Read introduction.
- Complete check item E1 and ask question 1 for each family member before proceeding to the next family member.	- Complete check item E1 through item E2 for each family member before proceeding to the next family member.
- Ask questions 2 and 3 for the family.	
- Complete item E2 for all family members.	
• Health Indicator Page:	
- Ask question 1 for the family.	- Complete check item G1 through question 5, as appropriate, for each family member before proceeding to the next family member.
- Ask questions 2 and 3 for each family member before proceeding to the next family member.	
- Ask questions 4 and 5 for each appropriate family member before proceeding to the next family member.	
• Demographic Background Page:	
- Read introduction.	- Read introduction.
- Complete check item L1 through question 2, as appropriate, for each family member before proceeding to the next family member.	- Complete check item L1 through item R, as appropriate, for each family member before proceeding to the next family member.
- Ask question 3 for each family member.	
- Ask question 4 for each family member.	

b. THIS-100(X) Interviewer's Manual--This contained detailed procedures about conducting the interview (NOTE: Part B of the manual, "Field Procedures" was never prepared, therefore, the manual contains only Parts A, C, D, and E.) (DSD)

c. THIS-104(X) Supervisor Probe Questions for Special Places; Form 11-213(R) Special Place Listing Sheet--Used for the Special Place Research Project for REPs 2-12. Appendix 7 of the main report discusses the Special Place Research Project. (SMD)

d. THIS-501(X) Interviewer's Flashcard Booklet--This consisted of a group of cards used for reference during the interview and afterwards for questionnaire editing. (DSD)

e. THIS-501B(X)--Reference Period Card which showed the three reference periods used during the interview. (DSD)

f. THIS-705A, THIS-705B, THIS-706A, and THIS-706B--These were diagnostic and nondiagnostic error forms used for editing the questionnaire forms. (DSD)

g. "Dear Friend" Letter (no form number)--Form letter which was sent to persons who would not participate in the interview but who would give us their address so that we could provide verification/explanation of the survey. (This letter was produced on the word processor, as required.) See paragraph E of this section for a discussion of the use of "Dear Friend" letters. (DSD)

h. THIS-601(L) Thank You Letter--This was sent to persons who completed the interview but also requested written confirmation from the Census Bureau. (DSD)

i. THIS-899 Self Study for New Interviewers
THIS-899.1 Training Cover Booklet (blank)
THIS-899.2 Training Insert Booklet (blank)
THIS-899.3 Training Cover Booklet (with entries)
THIS-899.4 Training Insert Booklet (with entries)
THIS-899.5 Tape Recorded Interview

These were self-study materials given to interviewers prior to classroom training. These items were originally prepared by Field Division's Training Branch to use for training new personal visit NHIS interviewers. DSD made minor revisions to adapt the material for the telephone interviewing mode.

C. NHIS-RDD Extended Follow-Up of Unresolved Cases--Based on a review of costs through the end of March, it was determined that Field Division would have a sufficient surplus in their budget to conduct an additional research project. Two possibilities were suggested: (1) adding a REP 13 to attempt to maximize the response rate by releasing the poorer performing interviewers and retaining only the better interviewers; or (2) conducting an extended follow-up of uncontacted cases from production REPs 1 through 11 to try to determine the residential/nonresidential status of each. Since it was estimated that adding a REP 13 would only show marginal improvement, the decision was made to conduct the extended follow-up. The procedures used to conduct this follow-up are specified below:

- The follow-up was conducted from May 3 through May 19, during the period when interviewers workloads were winding down.

- Only cases with a final outcome of "unable to contact-status unknown," and "unable to contact-confirmed residential from another source" were included in the follow-up.

- CSMR identified and printed out a listing of the unresolved cases from each REP.

- DSD prepared a paper questionnaire similiar to the Primary Screening instrument which appeared on the CATI system. No health data was collected for cases determined to be residential during this follow-up. A copy of the extended follow-up form appears in Exhibit 1 on page 21 and 22.

- Interviewers were instructed to "blitz" the numbers assigned to them; that is, to attempt calls as frequently as possible until the status was resolved or until the follow-up was discontinued. Periodic calls were also made to the appropriate telephone business office.

- To assure that cases included in this follow-up would have an adequate opportunity for contact, a minimum of five attempts was to be made in each of the weekday time slots with at least two attempts on each Saturday.

- Interviewers attempted to reconcile with the respondent the reason the number was not contacted during the appropriate interviewing period.

- A total of 223 cases from REPs 1-11 were identified for follow-up. Unresolved cases from REP 12 were not available soon enough to be included in the follow-up.

EXHIBIT 1

The form used for the follow-up appears below.

HHS-ADD EXTENDED FOLLOW-UP FORM

Case ID: _____ Tele. No. ( ) -

Name of Place: _____ Time Zone: _____

Final Outcome Code: _____ TSO No. ( ) -

GENERAL INSTRUCTIONS

- Call the sample telephone number as frequently as possible during your work shift.
- Tally each attempt in the appropriate time slot box for the time zone called.
- If contact is made, complete PART A below.
- If contact is not made after 5 attempts, call the Telephone Business Office during normal business hours for the time zone and complete PART B on the back.
- Enter appropriate notes concerning this case in the notes space on the back.

TALLY OF ATTEMPTED CALLS (times refer to residence time zone)

9:00 - 2:00	2:00 - 4:30	4:30 - 7:00	7:00 - 9:00	Saturday	Sunday

PART A - SAMPLE NUMBER CONTACT

(INTRO) Hello. I'm _____ from the United States Bureau of the Census in Washington, D.C. Is this (number) in area code (area code)? (If Yes, continue with 1. If No, apologize and redial)

1. Have I reached you on your HOME phone?

   ☐ YES - (Go to 4)

   ☐ NO - (Ask 2)

   ☐ SPECIAL PLACE - (Go to 5)

2. Does this telephone number serve a business or does it serve some place where people COULD live, such as a dormitory, hotel, or so forth?

   ☐ BUSINESS - (Ask 3)

   ☐ OTHER (SPECIFY) _____ - (Ask 3)

   ☐ SPECIAL PLACE - (Go to 5)

3. Does ANYONE use this number for a HOME phone?

   ☐ YES - (Ask 4)

   ☐ NO - (Go to 5) -

4. Was this your telephone number during the period _____ through _____?

   ☐ YES - (Ask 5)

   ☐ NO - (Go to CLOSING)

5. We tried to contact this number several times during the period _____ through _____, but no one answered. Could you tell me why no one answered?

_____

_____

(CLOSING) Thank you for your help. For now, all we are doing is identifying residential phone exchanges for a survey we will be doing in the future. (END CALL - CIRCLE TIME SLOT ABOVE)

(Continued on next page)

D. Response Rates

1. Reports--Two series of reports were produced during the course of the survey to monitor response.

   a. One series included weekly progress reports by REP computed from a manual review of all call records for each case in a REP. Because the cases could not be identified adequately on the management reports and call records as original sample cases, replacement cases, or substitute cases, these response rates included all types of cases. The purpose of the weekly progress reports was to produce preliminary figures which were available shortly after closeout for each REP. Rates for response, refusal, and "other noninterview" appear in Exhibit 2 on the next page.

The "other noninterview" rate may be somewhat overstated because, for each RDD, a few cases were not included in the call record, and since the exact status of these cases could not be determined, they were included in the "other noninterview" rate. In addition, some cases which were never contacted and for which the status was unknown, were included which probably would have been classified as out-of-scope if interviewers had been able to obtain confirming information from the telephone business office, operator assistance, or some other acceptable source. To determine the extent of this problem, an extended follow-up was done (see paragraph C above).

During the course of the survey, some questions were raised about the adequacy of the method used to compute response rates. Various computation methods have been noted through informal discussions with others involved in telephone surveys. For example:

$$\text{INTERVIEW RATE} = \frac{\text{Interviews}}{\text{Total Cases}} \times 100$$

OR

$$\text{INTERVIEW RATE} = \frac{\text{Interviews}}{(\text{Interviews}) + (\text{Other Confirmed Residential Cases})} \times 100$$

OR

$$\text{INTERVIEW RATE} = \frac{\text{Interviews}}{(\text{Total Cases}) - (\text{Out-of-Scope}) - \left[\frac{\text{Out-of-scope}}{\text{Total}} \times \text{Unresolved}\right]} \times 100$$

The first presents response as a proportion of the total sample, the second as a proportion of residential cases only, and the third as a proportion of in-scope cases after an adjustment was made to delete the estimated out-of-scope cases. To get a better handle on the NHIS-RDD response rate adequacy, research was proposed to determine the response rates achieved on other RDD surveys, the methods used to compute those rates, and similarity/dissimilarity between these surveys and the NHIS-RDD (for example, length of interview, respondent rules, interviewing period, staff composition, etc.). The NHIS-RDD rates would then be compared to those of other surveys. The results of this research are included in Appendix 1 of the main report.

REP	BEFORE SUBSTITUTION BASED ON FINAL CMS FILE RESPONSE RATE
1	69.1%
2	63.9%
3	72.5%
4	74.2%
5	73.4%
6	73.5%
7	78.8%
8	80.1%
9	79.3%
10	84.3%
11	80.6%
12	84.3%

Additional information on response rates by questionnaire version and on cumulative rates by replication are included in Appendix 1 of the main report. Some additional discussion and analysis of the response rates are also provided in Appendix 1.

When reviewing the above response rates, one point should be mentioned. Generally, one expects a certain amount of improvement during the start-up REPs as interviewers become more proficient in explaining and conducting the survey. During REP 2, the response rate fell below REP 1 figures. This may be explained in some part by the unexpected death of the original facility manager during this time, which caused confusion and interruption in the facility's operations.

2. Proposals to Increase Response Rates--Several suggestions were proposed during the course of the survey to increase response and/or efficiency:

   a. Use personal visit NHIS supervisory field representatives or experienced interviewers to convert refusals. This proposal was rejected because it did not appear to be cost effective since most, if not all, of these persons would have to be working here on subsistance, reimbursed for travel costs, and probably could not be committed for more than one week at a time. Also, since virtually all field NHIS interviewing is done in person, they may not be particularly adept at telephone interviewing.

Based on a review of the next week's interviewer performance reports, one interviewer was released due to a high refusal rate. The other three were retained. Because of this, no sample reduction was necessary. However, the earlier work schedule for some interviewers was implemented. No Sunday interviewing was performed.

e. Stop introducing substitute cases late in the interview period for a REP since there was not adequate time allowed to contact these cases. The decision was made to stop substitution for noninterviews and potential noninterviews after Thursday of the third week for a REP. (NOTE: Because of an error made when revising the CMS to account for this procedure, almost all substitution was stopped for REP 7. Once detected, the error was corrected and substitutes were generated according to the recommended procedure.)

f. There was discussion regarding whether it would be more efficient to close-down a 100-block cluster if we did not get enough contacts after trying 50-60 sample numbers rather than continuing to make unproductive attempts for up to 100 numbers. Due to possible complications involved in weighting the data, it was decided that the entire list of 100 numbers must be attempted, if necessary, for any cluster determined to be residential from the primary screening. Additional research into developing cut-off rules was initiated but the final recommendations came too late for implementation during the Feasibility Study.

E. Use of "Dear Friend" Letters--These letters were used in an effort to further explain the purpose of the survey to reluctant respondents. Of course, this method of refusal conversion depended upon several factors: (1) that the respondent would give us their correct address, and (2) that enough time remained in the interviewing period for the REP to allow receipt of the letter, which was sent by first-class mail, and recontact by the telephone interviewer.

During the course of the survey, 24 letters were sent. One of these did not have a control number assigned to it, therefore, final disposition of this case is not known. Of the remaining 23 cases, 18 were completed interviews,. one was an ineligible residence (i.e., all military), and four were refusal cases.

Nineteen of the 23 interviewers originally hired attended training. Seventeen of these persons finished training. This left an interviewing staff of 14 since three persons were chosen from this staff to become clerks. It was estimated that 15 interviewers would be needed to staff the project, therefore, four additional interviewers were hired, which allowed for interviewer attrition during the course of the survey. Field Division received approval to hire these persons based on referrals from the Charlotte Regional Office.

3. Other Staff--Three persons were selected as clerks from the staff of interviewers. Clerks were chosen based on their potential to perform edit and transcription activities and their lack of potential to become good quality interviewers.

   Three shift supervisors were hired at the GS-5 level. All of these persons had previous Census telephone interviewing experience.

B. Training--Due to time constraints, the self-studies and initial training package used in the field for personal visit interviewers were revised and used to train the telephone interviewers. Since many of our personal visit interviewers have previous interviewing experience, the training package focuses on the survey content rather than interviewing techniques. Although interviewers received some training on telephone interviewing techniques, this was inadequate since many of the staff had no previous interviewing experience. Many telephone interviewers had difficulty grasping the detailed content procedures and appeared to be overwhelmed. This problem may also exist in the field to some extent, however, it was probably compounded for the telephone survey since Field Division was forced to hire minimally qualified interviewers (personal visit interviewers are often chosen to work on the NHIS because of their excellent performance on other surveys).

A dry run of the initial training package was held in mid-December 1983. Two of the supervisors were trained at this time. The other supervisor attended one of the interviewer training sessions given in January 1984.

Prior to classroom training, interviewers completed the Pre-Classroom Self Study, which dealt mostly with NHIS conventions, and listened to the audio cassette tapes, which discussed asking questions, probing, and reluctant respondents.

Several weaknesses in our current training methods were identified and are noted below:

- Since it is extremely important to develop a good "sales pitch" in a telephone survey, this area of training needs to be expanded. Although methods to sell the survey were presented during the initial narrative training, practice interview training, and retraining phases, most interviewers were never able to answer respondent's questions or adequately explain the purpose of the survey. Some method of training should be developed to teach interviewers to "think on their feet." (Perhaps this is a factor which should also be considered during the ranking and selection of interviewers stage.)

- The importance of the interviewers role and the value of the survey should be stressed. It's very difficult to sell a survey to respondents if the interviewer believes it's useless. These topics are covered in the current training package but, apparently, interviewers are not convinced that their role in the survey makes any difference or that the survey has any real value. Most interviewers with prior Census interviewing experience looked upon this survey as a demotion rather than a challenge. This attitude appeared to be accepted by persons new to interviewing as well.

- Interviewers should be trained in "listening" techniques. Often, when asked a question by the respondent, inappropriate answers would be given. This is essential to gain the respondent's cooperation as well as to record correct survey data.

- The amount of content training received at one time is overwhelming to new interviewers. A centralized telephone location resolves many problems associated with using different training techniques since interviewers could be trained using self-paced techniques, such as self-studies, rather than by using the traditional lecture method technique. The lecture method currently used does little to stimulate the interviewer's interest in the survey. Other methods of presenting the material should be developed (for example, more extensive use of self-studies, and more practice interviews between content training).

2. Professional Staff Monitoring--The purpose of this was to help assess the feasibility of conducting the NHIS by telephone. The observations were not primarily intended to be used for quality control or an evaluation of interviewer performance. However, this activity was used to provide feedback to interviewers due to the failure of the supervisors to perform quality control monitoring.

DSD and NCHS provided "technical experts" to monitor interviewers and to answer NHIS content and procedure questions during the evening hours for the first four weeks of production interviewing. After this time, persons were not assigned to cover evening hours since few, if any, technical questions were asked by the supervisors.

We experienced several problems in the monitoring program. First, because of the recent break up of AT&T, it was difficult to secure all of the equipment necessary for effective monitoring. The telephone silencers for the monitoring equipment were not installed until the survey was nearing its end. Without the silencing equipment, an audible click was noticeable to both the interviewer and the respondent when monitoring began and ended; a decrease in the sound volume from the interviewer, respondent, or both often occurred during monitoring; and, occassionally, extraneous noise was picked up. During the course of the survey, the interviewers were told that monitors were instructed to discontinue the activity if either the respondent or the interviewer experienced difficulty hearing. This complaint may have been abused since some interviewers saw this as a way to avoid being monitored. Future studies should avoid such comments to the interviewing staff. Of course, the preferred solution would be to secure all necessary monitoring equipment before the survey begins.

Second, although we were told that supervisory monitoring was being or would be performed, there was concern that supervisors would not be able to identify interviewer errors since they had no more experience with the NHIS program than did the interviewers. Because of this, DSD and Field Division agreed to let the professional staff discuss any problem areas noticed during the monitoring session directly with the shift supervisor. The supervisors were then to explain the correct procedures to the interviewer. DSD and Field agreed that the completed monitoring forms should not be shown directly to the interviewer. Unfortunately, this procedure was ignored by the supervisors. Rather than explaining correct procedures, the supervisors often handed the monitoring forms directly to the interviewers. Since these forms sometimes contained indiscrete comments which should not have been seen by the interviewer, this generally created a negative attitude about

These meetings allowed the staff to air their grievances and opinions. Issues raised during the meeting were distributed to appropriate Bureau and NCHS personnel for comments and/or action. Overall, the QC meetings covered some valid concerns and generated responses to explain the rationale behind our procedures.

Specific issues mentioned by the interviewers during the QC meetings are listed below:

- Interviewers had trouble throughout the course of the survey distinguishing the concept of major activity in the past 12 months from work during the past two weeks, particularly for retired persons.

- Interviewers felt many questions were repetitive which caused respondents to become irritated or hostile.

- Interviewers felt that respondents were confused by the term "regular school" and tended to believe that this included high school but not college. Interviewers also had difficulty completing the education item for persons who attended school in foreign countries with grade levels different from our own.

- The question to determine the accuracy of health information (question 8 on the Demographic Background Page of the insert booklet) caused respondent confusion or resentment since it questioned the respondent's credibility.

- The length of the interview is a drawback. Most respondents don't want to spend more than a few minutes on the phone.

- Interviewers questioned the importance of some of the conditions which appear on the Condition Lists.

- Many interviewers felt that most refusals occurred just before or during the household rostering question.

- Interviewers stated during the first QC meeting that the survey introduction was too lengthy and repetitive. After revising the introduction, interviewers felt it was a great improvement since it allowed them to start the health interview more quickly.

- Some interviewers reported respondent resistance to the income question.

- Interviewers felt that their workspace was inadequate, particularly if additional questionnaires were needed because of large households.

- Interviewers said they often had difficulty deciding how to handle respondents who were "hard of hearing."

F. Performance Standards--The study plan stated that performance rates would be calculated for refusal/nonresponse, production, and error rates. The methods used to compute refusal/nonresponse and production rates for personal visit interviewers are relatively easy to compute since each interviewer is assigned specific cases for which (s)he is held accountable. Because RDD surveys assign cases as they become available, there is no direct accountability and we lose the ability to rate interviewer performance using this measure.

During the Feasibility Study, interviewer refusal rates were calculated by dividing a count of the first occurrences of any sort of refusal in the call history by the number of all contact calls. Production rates were calculated based on the number of calls per login hour (e.g., the amount of time per hour the interviewer was "logged on" the CATI system) and based on the number of completed interviews per login hour. During the course of the Feasibility Study, one interviewer was released from the survey due to a high Type A noninterview rate compared to other NHIS-RDD interviewers.

Error rates were calculated beginning in the fourth week of the survey. (This activity was overlooked for the first three weeks.) The clerical staff hand computed rates based on simple formulas printed on the error tally sheets. Although the same basic formulas are used to compute error rates for personal visit interviewers, these were not be used as a standard for this study for several reasons: (1) the experience levels of the two interviewing staffs are not comparable, and (2) the clerical staff had no more experience than the interviewing staff, therefore, they sometimes overlooked errors which should have been charged or charged errors when they shouldn't have.

The Feasibility Study interviewers were not provided with any type of positive incentives to increase their performance. Field Division requested approval from Personnel Division to implement some type of cash awards system, but this suggestion was rejected, for one reason, because performance standards had not been established for telephone surveys.

Generally, the methods used to compute the rates mentioned above cannot be compared with personal visit rates but could be used as a general indication of how the interviewers compared with their peers. Exactly what measures should be included when calculating performance standards requires additional research. We also need to determine in advance the types of feedback that interviewers should receive about their performance (e.g., results of monitoring, noninterview rates, error rates, etc.), establish standard procedures for this feedback, and make sure these procedures are carried out.

4. Cases were sometimes reassigned immediately after a previous call attempt, although they should have been placed in a queue for a different time slot. This often caused interviewers to enter inappropriate outcome codes in an effort to avoid these cases. For example, after receiving and coding a first refusal, the case was immediately reassigned to the same interviewer (who, as it happened, was being monitored at the time). The interviewer again coded it as a refusal without attempting the call, thus resulting in a final noninterview. Other similar instances were observed and the interviewers were then advised to use the "quit-out" code when an assigned case was not attempted rather than assigning a valid outcome code to such cases.

5. Early in the survey period, telephone business office (TBO) calls were often scheduled during non-business hours. This caused a call attempt to be tallied against the 15-call limit although no call was made. This was corrected during the course of the survey; however, another problem then arose which may have resulted from programming revisions to count only one TBO call against the call maximum. During REP 8, very few TBO calls were scheduled at all. As a result, no 'confirmed residential' noninterviews were identified and many of the "status unknown" cases received 14-20 "ring-no-answer" attempts with no TBO calls. This problem was corrected once identified.

6. A number of inappropriate call outcomes were tallied against the 15-call limit which caused cases to be classified as noninterviews without adequate call attempts. For example, "Can't reach TBO/Directory Assistance" and "New TBO number" were tallied against the call limit.

7. The CMS only considered a case as a final refusal if the refusal outcome code was entered two times. This understated the true refusal rate and undermined efforts to control refusals since all refusals were not easily identifiable from the CMS. The final refusal coding became a manual operation rather than an automated one since system modifications could not be made. (This would have created inconsistencies in the way cases were handled during the survey period.)

8. Some interviewers may have entered false outcome codes to avoid difficult cases. During one quality circle meeting, interviewers mentioned that some of their co-workers were coding refusals as soft appointments in order to lower their refusal rate. Several professional staff monitors also mentioned that some interviewers assigned incorrect outcome codes to cases or assigned codes without dialing the number. This illustrates the importance of a supervisory monitoring program for surveys in which interviewers are not directly accountable for specific cases.

15. After reviewing the call records for REP 8, two invalid sampling clusters (i.e., not classified as residential in the primary screening) were discovered late in the interviewing period. Replacement clusters for these were then introduced with only three days or less remaining in the interviewing period, which did not provide adequate time for call attempts. It was jointly decided by Field and DSD to delete the unresolved cases from the REP 8 sample for these clusters. There was at least one other time during the survey period when secondary screening numbers were assigned which never received primary screening.

16. The last outcome code ("lastout") entered for a case sometimes determined its final status rather than using appropriate intermediate codes ("lowout"). For example, a partial interview requiring a callback was coded as "closeout cutoff reached" when the last attempt was a "ring, no answer." We could not be sure that all cases were properly classified due to the merging of the manual and automated systems.

17. Interviewers provided insufficient information to adequately analyze break off points.

18. Interviews for unrelated household members had to be controlled manually since the system could not handle situations which required separate interviews within a household.

19. Problems were experienced in the preparation and conversion of the CMS output files. The principal problem was an inability to read the last case on any CATI tape using generally accepted and tested software. It was almost the end of March before the output file was sucessfully converted to a UNIVAC file in order to conduct the response rate analysis. DSD manipulated the CMS output file to resolve this problem when reading the file.

20. The merging of the CATI CMS and the paper instrument also showed some inconsistency of case status between the CMS file and the paper questionnaires. For example, questionnaires for some cases which were classified as interviews on the CMS were not filled/keyed and vice versa.

Overall, the CMS improved toward the end of the study due to revisions in the programming and the avoidance of manual intervention (i.e., supervisors occasionally assigned cases by hand rather than allowing the system to function as designed). A CATI Software Subcommittee was formed to develop specifications for the CMS so that future problems could be avoided. However, persons planning to use the automated CMS in future surveys would be wise to verify that these problems have been resolved so that they are assured all expected data appears.

IX. CLERICAL EDITING/TRANSCRIPTION

The three clerks assigned to the project were responsible for all editing and transcription activities.

A. Editing--Clerks performed a 100 percent edit for all questionnaire cover booklets and insert booklets filled during the survey. DSD performed a 100 percent re-edit for all cases in REP 1. A sample of each clerk's work was re-edited at several other times throughout the survey period by DSD. Based on a sample of REP 6 interviews which were re-edited, only a few consistent interviewer problems were identified; such as, incomplete or inadequate condition entries on the Condition Page; making careless entries, for example, using one digit when two were required, not marking check items, and missing skip instructions. Although the edit clerks occasionally omitted some recodes or transcription items, they identified and corrected most errors, including diagnostic errors. Overall, based on re-edit, both the interviewers and clerks appeared to perform adequately with respect to filling the questionnaires.

B. Transcription--Due to budget contraints, the entire questionnaire could not be data keyed. Therefore, NCHS identified data items which were needed to compare the telephone and personal visit data. All items which needed to be keyed from the insert booklets were transcribed to pages 2 and 3 of the cover booklets. This was done so that only the cover booklets needed to be transmitted to Jeffersonville for keying.

The data items which were transcribed from the insert booklet to the cover booklet for each person are listed below:

-A2 (Household membership status)
-C2 TOTAL (Total number of conditions recorded in C2)
-C2 LA (Total number of conditions recorded in C2 with Limitation of Activities Page as a source)
-C2 RA (Total number of conditions recorded in C2 with Restricted Activity Page as a source)
-C2 DV (Total number of conditions recorded in C2 with 2-Week Doctor Visits Page as a source)
-C2 CL (Total number of conditions recorded in C2 with Condition List as a source)
-B.1 (Major activity during past 12 months for persons 18-69)
-B.2 (Limited in job or business for persons 18-69)
-B.3 (Limited in housework for persons 18-69)
-B.5 (Would the person be limited in work for persons 18-69)
-B.6a (Limited in any way in any activities for persons 18-69)

The following information was keyed for each family record:

- Control identification
- Condition List asked
- Insert Booklet version used
- Interviewer's code
- Month and date of interview
- Respondent's use of a calendar to recall events during 2-week reference period
- Income
- Other telephone numbers which serve the home
- How many different telephone numbers serve the home
- Are any of the telephone numbers used only for business
- How many are used only for business
- Code to determine if exact address was provided
- Code to determine if mailing address was provided
- Description of type of residence
- Does telephone number serve other residences
- Number of residences served by telephone number
- Is telephone located inside or outside of residence
- Code to determine if name and/or telephone number of someone who could provide a listing of persons living in the unit was provided
- Number of eligible respondents in family
- Was respondent the most knowledgeable respondent
- Would anyone know equally as much about the health of family members
- Person number of "equally" knowledgeable person
- Would anyone know more about the health of family members
- Person number of "more" knowledgeable person

X. DATA KEYING AND PROCESSING

This section refers to the keying and processing of questionnaires in Jeffersonville. Refer to section VII of this report for information regarding the automated case management system used in the study.

A. DATA KEYING--All clerical editing and coding activities were performed by the clerks prior to transmittal of the questionnaires to Jeffersonville for keying. Although two versions of the cover booklets were used, the data fields were exactly the same on both forms. All data were keyed in numerics only. Entries were range checked and keying was verified 100 percent. Refer to section IX, paragraph B for a list of items which were data keyed.