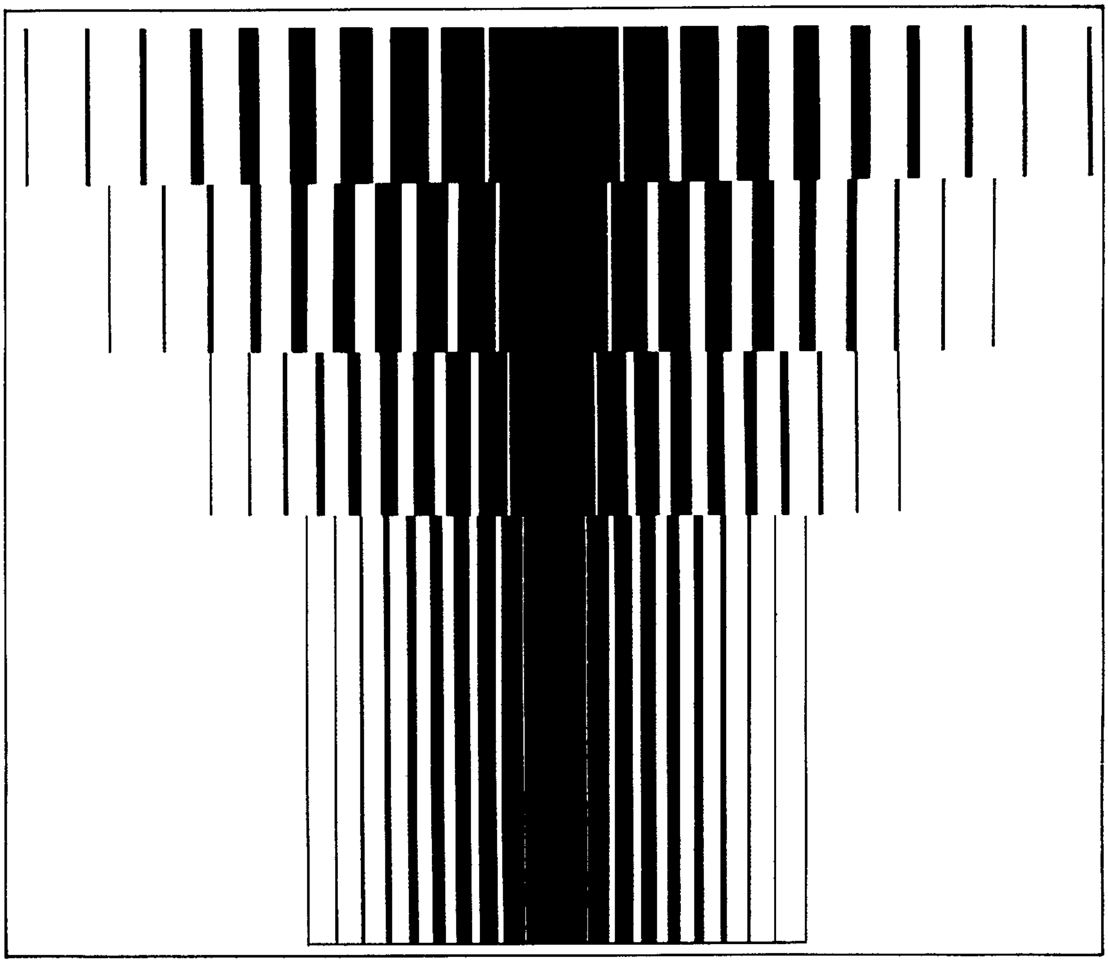
Evaluation of Data Collection Frequency and the Use of a Summary in the National Medical Care Utilization and Expenditure Survey

Series A, Methodological Report No. 2



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Published by
Public Health Service
National Center for Health Statistics

May 1987

National Medical Care Utilization and Expenditure Survey

The National Medical Care Utilization and Expenditure Survey (NMCUES) is a unique source of detailed national estimates on the utilization of and expenditures for various types of medical care. NMCUES is designed to be directly responsive to the continuing need for statistical information on health care expenditures associated with health services utilization for the entire U.S. population.

NMCUES will produce comparable estimates over time for evaluation of the impact of legislation and programs on health status, costs, utilization, and illness-related behavior in the medical care delivery system. In addition to national estimates for the civilian noninstitutionalized population, it will also provide separate estimates for the Medicaid-eligible populations in four States.

The first cycle of NMCUES, which covers calendar year 1980, was designed and conducted as a collaborative effort between the National Center for Health Statistics, Public Health Service, and the Office of Research and Demonstrations, Health Care Financing Administration. Data were obtained from three survey components. The first was a national household survey and the second was a survey of Medicaid enrollees in four States (California, Michigan, Texas, and New York). Both of these components involved five interviews over a period of 15 months to obtain information on medical care utilization and expenditures and other health-related information. The third component was an administrative records survey that verified the eligibility status of respondents for the Medicare and Medicaid programs and supplemented the household data with claims data for the Medicare and Medicaid populations.

Data collection was accomplished by Research Triangle Institute, Research Triangle Park, N.C., and its subcontractors, the National Opinion Research Center of the University of Chicago, Ill., and SysteMetrics, Inc., Berkeley, Calif., under Contract No. 233–79–2032.

Co-Project Officers for the Survey were Robert R. Fuchsberg of the National Center for Health Statistics (NCHS) and Allen Dobson of the Health Care Financing Administration (HCFA). Robert A. Wright of NCHS and Larry Corder of HCFA also had major responsibilities. Daniel G. Horvitz of Research Triangle Institute was the Project Director primarily responsible for data collection, along with Associate Project Directors Esther Fleishman of the National Opinion Research Center, Robert H. Thornton of Research Triangle Institute, and James S. Lubalin of SysteMetrics, Inc. Barbara Moser of Research Triangle Institute was the Project Director primarily responsible for data processing.

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402

Suggested Citation

Tourangeau, R., and Rasinski, K. A.: Evaluation of Data Collection Frequency and the Use of a Summary in the National Medical Care Utilization and Expenditure Survey. *National Medical Care Utilization and Expenditure Survey.* Series A, Methodological Report No. 2. DHHS Pub. No. 87–20002. National Center for Health Statistics, Public Health Service. Washington. U.S. Government Printing Office, May 1987.

Library of Congress Cataloging in Publication Data

Tourangeau, Roger.
Evaluation of data collection frequency and the use of a summary in the National Medical Care Utilization and Expenditure Survey.
(National Medical Care Utilization and Expenditure Survey. Series A, Methodological report ; no. 2) (DHHS publication ; no. 87–20002)
By: Roger Tourangeau and Kenneth A. Rasinski.
Bibliography: p.
Supt. of Docs. no.: HE22.26/2:3
1. National Medical Care Utilization and Expenditure Survey (U.S.) 2. Health Surveys—United States—Statistical methods—Evaluation. I. Rasinski, Kenneth A. II. National Center for Health Statistics (U.S.) III. National Medical Care Utilization and Expenditure Survey (U.S.) IV. Title. V. Title: Data collection frequency and the use of a summary in the National Medical Care Utilization and Expenditure Survey. VI. Series: National Medical Care Utilization and Expenditure Survey (Series). Series A, Methodological report ; no. 2) VII. Series: DHHS publication ; no. 87–20002. [DNLM: 1. Data Collection—methods. 2. Health Surveys—United States. WA 950 T727e]
RA410.53.T67 1987 362.1'0973'021 86–23823
ISBN 0–8406–0358–4

Contents

List of Text Tables

List of Text Figures

Symbols

- - -	Data available for fewer than 75 cases
. . .	Category not applicable
-	Quantity zero
0.0	Quantity more than zero but less than 0.05
*	Test statistic is significant at 0.05 level
**	Test statistic is significant at 0.01 level
***	Test statistic is significant at 0.001 level

Evaluation of Data Collection Frequency and the Use of a Summary in the National Medical Care Utilization and Expenditure Survey

By Roger Tourangeau and Kenneth A. Rasinski of NORC (formerly the National Opinion Research Center)

Executive Summary

This is one of a series of reports that evaluate the methods of the National Medical Care Utilization and Expenditure Survey. The survey was designed as a means to collect data about the U.S. civilian noninstitutionalized population during 1980. During the course of the survey, information was obtained on health, access to and use of medical services, associated charges and sources of payment, and health insurance coverage. The information that was collected covered the entire calendar year.

This report examines two methodological issues raised by the survey. Most of the respondents were interviewed five times, but for a substantial minority, one interview was deliberately skipped. The first part of this report examines how the difference in the number of interviews affected the level of reporting and data quality. A related issue is the use in the survey of a computer-generated summary of key information from previous interviews. Respondents reviewed the summary as part of the interview and could add to, delete, or change information provided in earlier interviews. The second part of the report examines the impact of the summary on the survey data.

The report presents three major findings. First, the number of scheduled interviews had little or no effect on levels of reported events of or charges for medical services. This finding is somewhat surprising, because the mean length of the reference period for respondents scheduled for four interviews was longer than the mean length of the reference period for respondents scheduled for all five; for this reason, one would expect the four-interview respondents to make more memory errors. The second finding was that the length of the time period covered by an interview had small and inconsistent effects on the reported events. Although events do seem to be forgotten over time, the rate of forgetting is apparently slow. Finally, the summary seems to have been an effective means for gathering additional data on charges for medical services. The summary was especially useful because charge figures were unavailable in the initial interview for some of the most expensive services.

NOTE: This report was prepared under contract (Contract No. 282–84–2109) from the National Center for Health Statistics (NCHS). Harrison Greene and Hy Bern of NORC (formerly the National Opinion Research Center) provided programming support for the work reported here, and Sophia Ravin of NORC provided editorial assistance. A number of reviewers commented on earlier versions of the report, including Hal Nisselson of Westat and Margot Brown, Earl Bryant, Robert Casady, Andrew White, and Robert Wright of NCHS. Michele Chyba of the Center provided invaluable help and advice on several issues.

Introduction

The National Medical Care Utilization and Expenditure Survey (NMCUES) was conducted during 1980 and 1981 to provide accurate data from a nationally representative sample of the civilian noninstitutionalized population on the use of and charges for medical services. Approximately 17,000 respondents in the national household sample were contacted for initial interviews; by 1981, when the final interviews were completed, more than 16,000 respondents had provided data for all of 1980.

Like its predecessor, the National Medical Care Expenditure Survey, NMCUES depended on the recollections of panels of respondents. Survey respondents are prone to many kinds of reporting error, the most common of which involves errors of memory. Sometimes respondents completely forget an event. Forgetting reduces the levels of reporting in surveys. Drawing on results from the study of memory, Sudman and Bradburn (1973) suggested an exponential decay function to describe the rates of retention and forgetting over time. Built on the assumption that the rate of forgetting is rapid at first but slows over time, the exponential decay model has a long history in psychology, dating back to classic studies conducted by Ebbinghaus (1885) before the turn of the century. Recently, however, the model has been called into question by Loftus (1982) and others; the more current results suggest that forgetting is a linear function of the passage of time.

Sometimes the event can be remembered but cannot be placed accurately in time. This kind of memory error is called "telescoping" and generally leads to overreporting (Sudman and Bradburn, 1973) because the events tend to be recalled as more recent than they actually were (Brown, Rips, and Shevell, 1985). The telescoping effect is generally thought to be greatest when the event to be recalled is distant in time. Several characteristics of the event—including emotional impact, duration, and frequency—are thought to influence both forgetting and telescoping.

The NMCUES design incorporated several procedures to reduce both kinds of memory error. Respondents were given a calendar-diary to record medical events. After the first round of data collection, the respondents received a summary of key items from previous rounds to review for accuracy and completeness during the last part of the interview. Finally, the reference period for each interview was kept reasonably short; for most respondents data were collected five times, with each interview covering a period of approximately 70 days.

Sudman and Bradburn (1973) suggest that an optimal length might be found in which telescoping errors would more or less offset omissions due to forgetting; however, in practice most surveys are based on the assumption that shorter reference periods produce more accurate results than longer ones. This assumption has been questioned by recent analyses of data from the National Medical Care Expenditure Survey (NMCES), the precursor survey to NMCUES. Cohen and Burt (1984) compared results from respondents who completed all scheduled NMCES interviews with results from a group of respondents who were deliberately skipped (held over) during one round of data collection. The holdovers generally reported higher levels of medical utilization and expenditure than the other NMCES respondents did; in addition, the holdovers' questionnaire data were more consistent with data from medical provider records.

Cohen and Burt (1984) attribute this difference between the two groups of respondents to another type of reporting error—conditioning effects, which refer to the impact of repeated interviewing in a longitudinal study. With repeated interviews, respondents may learn that answers to certain questions may lead to extensive followup questioning, and some respondents may therefore intentionally misreport their answers to shorten the interview. Conditioning is, thus, analogous to fatigue within a single interview, and is thought to produce underreporting.

NMCUES provides another opportunity to examine memory errors and conditioning effects in surveys on medical care. This task is especially important because plans are already under way for a third study on medical care costs, the upcoming National Medical Expenditure Survey.

The NMCUES national household sample included 17,123 "key persons" (members of the sample households at the time of the first interview and certain relatives). Of the key people, 16,207 provided data for the entire reference year. More than 11,000 of those people responding for the entire year were interviewed five times during 1980 and 1981; the remainder were cases who, for one reason or another, came to be interviewed during the last part of the third round field period and were therefore skipped during round 4. Thus, the four-interview group does not constitute a random selection from the entire sample; instead, it includes respondents who are, in effect, self-selected because they were hard to interview or because they presented some other special problem.

Respondents in the four-interview group had reference periods that averaged 20 days longer than the other NMCUES respondents. (In addition, there is considerable within-group and round-by-round variation in the length of reference periods.) The NMCUES data thus afford an opportunity to explore the effects of data collection frequency and length of recall period on reported medical care use and charges.

After the first round of data collection, respondents were mailed a computer-generated summary of their responses to key items from previous rounds. At the end of each followup interview, respondents were asked a series of questions designed to elicit more complete and accurate information regarding events on the summary. The primary objective in using the summary was to allow respondents to update previous responses as additional information became available to them (for example, medical bills); however, the summary also served other functions. It was probably useful as a "bounding" device (Neter and Waksberg, 1964) that reduced telescoping errors, and as a retrieval cue that stimulated recall.

The use of a cumulative data summary and the deliberate manipulation of interview frequency were important features of the NMCUES design. The objective of this report is to examine the effect of those innovations on NMCUES data. The report will focus on five specific questions:

1. Did the respondents who were interviewed four times differ from the other NMCUES respondents in overall levels of reported medical utilization and charges, in relationships among these variables, or in the variability of their data?
2. Did the four-interview respondents differ from other respondents in measures of data quality?
3. Did levels of reporting relate to the length of the reference period?
4. How often did respondents take advantage of the summary review to change data from previous rounds?
5. What was the impact on final survey estimates of changes made during the summary review?

Methods

NMCUES Sample

The NMCUES national household sample included two national samples independently selected by the Research Triangle Institute (RTI) and NORC (formerly the National Opinion Research Center). Both samples were stratified, multistage area probability samples, selected by similar procedures. In both samples, counties and standard metropolitan statistical areas (SMSA's) constituted the primary sampling units (PSU's), and smaller geographical areas were selected in subsequent stages of sampling. Housing units were selected at the final stage. The residents of sample housing units enumerated during round 1 (and selected relatives of the residents) were considered key people to be followed during subsequent rounds. To facilitate the computation of standard errors, the PSU's in both samples were grouped into 69 "pseudostrata," each with a pair of PSU's (or groups of PSU's). Region and size of place were used as stratification criteria to form pseudostrata.

Those joining the households of key people at later rounds (for example, new spouses) were also interviewed; such nonkey respondents are excluded from this analysis. Of the 17,123 key people interviewed, 16,207 provided data for the entire reference year. The current analysis focuses on these respondents. Bonham (1983) gives a more detailed description of the NMCUES procedures.

Sources of the Data

The NMCUES data went through several stages of data processing before a final public use tape was produced. First, a file that combined data from several documents into initial, "uncleaned" data files was produced. Next, minimal cleaning (such as the replacement of blanks with a numerical code) was done, and variables were combined further to produce the "12-month files," on which much of this analysis is based. These 12-month files include event-level files, with records for each hospital stay, medical visit, and so forth, and person-level files with records for each person in each round. The final data on the public use tape differ from the 12-month data in several respects: Data have been cleaned to remove inconsistencies and out-of-range values; values have been imputed for many missing items; data from the initial interview, the summary revision, and administrative records have been combined using a "best-estimate" procedure (in which records data take precedence over summary revisions, and summary revisions take precedence over initial interview responses); and finally, person-level data have been combined across rounds. Given the focus of this analysis, the 12-month files were advantageous for several reasons: Differences between respondents interviewed four times and those interviewed five times are not obscured by imputed values or values based on administrative data; data from different rounds can be easily distinguished; rates of missing and inconsistent values can be examined; and summary revisions can be distinguished from initial responses.

Variables in the Analysis

The variables in the analysis fall into three major classes: First, some variables are treated as independent, such as race, sex, Medicare or Medicaid status, source of the data (self-respondent or proxy), age, and data collection frequency. These variables were drawn primarily from the public use person files and, therefore, are as accurate and free from missing values as possible. The second class of variables involve medical care utilization and charges and are treated as dependent, or covariates. These variables were drawn from the 12-month files, with the utilization variables originating, for the most part, from the round-by-round person files. The charge variables (and the number of medical visits) were drawn from the 12-month event-level files. The final class of variables, the summary revisions, were also drawn from the event-level files. The definition of variables used in this report may differ from those used in other NMCUES reports. The reader is advised to consult Appendix II before making comparisons to other NMCUES reports.

The 12-month data exhibit several problems. Typically, utilization data were gathered in three steps. A filter question was asked to determine whether any relevant events had taken place since the reference date. An initial followup question was then asked to determine how often such events occurred. Finally, a series of followup questions gathered detailed information on each event. For most of the filter items, the questionnaire did not provide an opportunity to note a negative response. If the response was positive (the respondent reported, for example, a hospitalization), the interviewer circled a number; but if the response was negative, both the filter and initial followup items were left blank. Thus, on these items missing data cannot be distinguished readily from a negative answer. The data also exhibit several inconsistencies. Some records have blanks on the filter item, but a positive number on the initial followup; others have the opposite pattern ("yes" on the filter, but zero events); still others have more or fewer event records than responses to the initial followup items would indicate (with only a few exceptions, there should be an event record for each reported event). Such inconsistencies are explored in more detail later in this report.

Standard Errors

Because the NMCUES data are from a complex multistage design, standard errors and inferential statistics based on standard computer packages, such as SAS (1982), are likely to underestimate the variability of survey statistics. For this reason, corrected standard errors and inferential statistics (produced using the balanced repeated replication procedure described in Appendix I) are presented for key results. The inferential statistics reported are all based on standard errors computed from 69 pairs of PSU's; thus, the error terms for these statistics have 69 degrees of freedom.

Effects of Data Collection Frequency on Survey Estimates

This section compares the data from respondents who completed four interviews with data from the NMCUES respondents who completed all five interviews, taking into account the effects of differences in background characteristics of the two groups. The analyses considered seven background variables: sex, race (black respondents compared with all other races), age (18 years and under, 19–64 years, 65 years and over), health status reported during round 1 (excellent, good, fair, or poor), Medicare or Medicaid status (whether or not the person was covered), proxy status (all data from respondent, data partly from proxy, or data completely from proxy), and length of the reference period.

Except where noted, the proxy status variable is a composite reflecting all rounds of data collection. Cases were classified into the "all data from respondent" group only when they provided all of the data for themselves in every interview. Similarly, respondents were classified into the "data completely from proxy" group only when proxies provided all their data in every interview. All other cases were classified into the "data partly from proxy" group.

In addition to background variables, the analysis considered event and charge variables. The five variables related to events are disability days, hospital stays, dental visits, medical visits, and emergency room visits. The three variables related to charges are those charges for hospital stays, those for medical visits, and those for prescribed medicines and other medical expenses. "Hospital stays" refer to episodes in which the respondent was hospitalized for at least 1 night. "Medical visits" is a broad category, covering all visits to medical providers (not necessarily a physician) except those seen while an inpatient. In addition to office visits, this category includes visits to an emergency room and a hospital outpatient department. Variables have been annualized (expressed in terms of an annual period) for each person when data from a particular round were used. For example, a respondent reporting two emergency room visits in a 90-day reference period would be treated as having 8.13 ($2 \times 366/90$) annual visits.

Background Differences

Table A shows the composition of the four- and five-interview groups by the background variables as reported in round 1. The four-interview group seems to include more hard-to-interview respondents, such as black persons and males (Kish, 1965). The difference between the four- and five-interview groups is also reflected in the proxy respondent variable—more of the first round data for the four-interview group were provided partly (30.3 percent compared with 29.0 percent for five interviews) or wholly (41.3 percent compared with 38.3 percent for five interviews) by proxy respondents. Two of the differences between the two groups are statistically significant: The four-interview group included more respondents 18 years of age or under ($z=4.90$) and fewer who provided all of the first round data for themselves ($z=5.89$). It is also possible that respondents interviewed four times began as a somewhat less healthy group; fewer of those respondents rated their health as excellent or good (86.2 percent) than other NMCUES respondents (86.7 percent). This initial difference, although not statistically significant, is further reflected in the round 1 event variables discussed later.

During round 1, respondents provided data for the period from January 1, 1980, through the day of the first interview; in subsequent rounds, the reference period began with the date of the previous interview and ran through the day of the current interview; for the final interview, the reference period ran from the date of the previous interview through December 31, 1980. As Table B shows, the four-interview group had a longer mean reference period than the five-interview group, with the largest differences occurring during round 3 (91.9 days compared with 77.2 days) and round 5 (88.7 compared with 47.6). Table B also shows for each group the mean "lag" in days between the end of the final reference period (that is, December 31, 1980) and the date of the final interview. It should be noted that the differences in length of reference periods are cumulative, so that by round 3 the four-interview respondents were reporting on a somewhat different time of the year than the five-interview respondents; by the third round, the average four-interview respondent had reported on almost an additional month.

Differences in Reported Events

Although the four-interview group had higher means on all five event (disability days and utilization) variables for the year (Table C), the main effect for data collection

Table A

Number and percent distribution of respondents by selected characteristics, according to number of interviews

Characteristic	4 interviews			5 interviews		
		Percent distribution			Percent distribution	
	Number	Raw	Weighted	Number	Raw	Weighted
All persons	5,085	100.0	100.0	11,122	100.0	100.0
Sex						
Male	2,455	48.3	48.7	5,297	47.6	47.8
Female	2,630	51.7	51.3	5,824	52.4	52.2
Race						
Black	681	13.4	13.9	1,132	10.2	10.4
White	4,289	84.4	83.8	9,775	87.9	87.6
Other	115	2.3	2.3	214	1.9	2.0
Respondent						
Self-respondent only	1,440	28.3	29.4	3,632	32.7	34.4
Partly proxy	1,543	30.3	30.2	3,229	29.0	28.8
Proxy only	2,102	41.3	40.4	4,261	38.3	36.8
Public coverage						
Medicare or Medicaid	821	16.2	15.6	1,946	17.5	16.9
Neither program	4,264	83.8	84.4	9,176	82.5	83.1
Perceived health status						
Excellent	2,469	48.6	48.6	5,583	50.2	50.3
Good	1,910	37.6	37.6	4,057	36.5	36.5
Fair	456	9.0	8.9	1,049	9.4	9.4
Poor	193	3.8	3.7	375	3.4	3.4
Unknown	57	1.1	1.1	58	0.5	0.5
Age						
18 years and under	1,819	35.8	34.6	3,439	30.9	29.3
19–64 years	2,807	55.2	56.6	6,413	57.7	59.4
65 years and over	459	9.0	8.8	1,270	11.4	11.2

Table B

Mean length of reference period, by number of interviews and round

Round	Number of interviews	
	4	5
	Average length in days	
Round 1	71.6	66.2
Round 2	114.0	106.3
Round 3	91.9	77.2
Round 4	...	66.8
Round 5	88.7	47.6
Lag[1]	33.3	36.5

[1]The period from the end of the reference period (that is, December 31, 1980) to the date of the final (Round 5) interview.

Table C

Average annual rate per person, by number of interviews, events, and charges

Event and charge	Number of interviews 4	5	F	R^2
Event				
Disability days	5.23	4.34	3.70	0.071
Hospital stays	0.17	0.13	**11.80	0.060
Dental visits	1.24	1.22	<1	0.021
Medical visits	5.55	4.93	<1	0.067
Emergency room visits	0.28	0.25	<1	0.022
Charge				
Hospital stays	$151.41	$114.38	**8.55	0.008
Medical visits	109.77	95.83	<1	0.062
Prescribed medicine and other medical expenses[1]	45.13	45.32	1.23	0.136

[1]Includes prescribed medicines, eyeglasses, orthopedic appliances, hearing aids, diabetic supplies, and ambulance services.

frequency is significant only for hospital visits ($F = 11.80$). The variables were examined using analysis of variance procedures; the analysis of variance model included seven main effect terms (data collection frequency plus the six background variables) and six interactions (the first-order interaction of the data collection frequency variable with each of the background characteristics). The F values for the data collection frequency main effect are shown in Table C. (The F value for the data collection frequency main effect is a "partial" F, which assesses the impact of the data collection frequency main effect after *all* of the other main effects and interactions have been considered.)

This apparent pattern of more disability days and higher utilization by the four-interview group occurred throughout the survey. Table 1 presents means for each group and for each round for the entire year. Almost without exception, the mean number of hospital stays was significantly higher for the four-interview group than for the five-interview group.

The R^2 values in Table 1 indicate that the 13-term analysis of variance model accounts for very little of the variation in the event variables (the median R^2 is less than 0.01). It is also clear that this model does not explain adequately round 1 differences between the four- and five-interview groups. For this reason, additional analyses with extra controls were conducted on the data from rounds 4 and 5. These analyses added terms that reflected the main effect for length of the recall period, the interaction of that variable with data collection frequency, and three covariates—variables from rounds 1, 2, and 3 that corresponded to the dependent variable. These augmented models accounted for considerably more of the variation in the data for rounds 4 and 5 and revealed main effects for the holdover variable only on round 4 hospital visits ($F = 6.07$) and round 4 medical visits ($F = 4.57$).

Thus, the four-interview group initially reported more disability days and higher utilization, a trend that continued throughout the survey, even after statistical controls were introduced to account for background differences. When even more extensive controls were introduced to compensate for the differences between groups on events during rounds 1–3, the round 4 and 5 data showed few significant differences between four- and five-interview respondents. Nonetheless, those differences that remain significant still indicate higher reported use of medical services by the four-interview group.

Differences in Reported Charges

The data on hospital and medical visit charges exhibit a pattern similar to that already demonstrated by the event data. The four-interview group reported higher charges for the year (Table C), a difference that is apparent from the first interview (Table 1). Table 1 also gives F values for the holdover main effect, controlling for the six background factors and six interaction effects. With the use of augmented models that control more adequately for differences between the groups prior to round 4, the data collection frequency main effect is not significant for either variable in round 4 or 5.

Respondents interviewed four times did not show consistently higher means on prescribed medicines and other medical expenses; in fact, the groups differed only by $0.19 in average annual charges of this type. None of the differences between groups regarding this variable is statistically significant.

Interaction Effects

The models used to analyze the event and charge variables also included interaction terms to assess whether the data collection frequency effect differed by subgroup. Several significant interaction effects were observed, most of them involving the age and health status variables. These interactions (which are not tabulated) can be summarized quite simply: The respondents in the four-interview group, regardless of age or round 1 health status, reported more disability days and greater use of medical services and higher charges than did their counterparts who completed five interviews; but the differences were more marked among subgroups that used medical services more frequently (such as the aged and those whose health was described as fair or poor in round 1).

Differences in Item Standard Deviations and Correlations

Beyond an examination of mean differences in disability days and utilization and charges, it is worthwhile to explore whether the four-interview group differed in other ways from the other NMCUES respondents. Because the four-interview group reported over longer periods of time, it is reasonable to assume that the data for this group are "noisier" than those for the five-interview group. Such a difference should be apparent in higher item standard deviations and lower correlations between variables for four-interview respondents. Data in Table D indicate the relationship between round 1 and round 5 utilization and charge variables by subgroup. For seven of the eight variables, the round 1 to round 5 correlation is higher within the four-interview group; only one of the differences is significant (see F values in Table D).

The distributions of all eight event and charge variables are highly skewed, with most respondents reporting no events or charges; the variance for this kind of distribution tends to be related to the mean. Table D gives the standard deviation for each of the event and charge variables, based on the entire year. For six of the eight variables, the standard deviation is higher within the four-interview group. This result is probably an indirect reflection of the higher means for this group.

Table D

Standard deviations (SD), correlations (*r*), and *F* tests, by number of interviews, events, and charges

Event and charge	4 interviews SD	4 interviews *r*	5 interviews SD	5 interviews *r*	*F*
Event					
Disability days	14.5	0.23	12.7	0.21	<1
Hospital stays	0.5	0.16	0.4	0.03	3.63
Dental visits	2.2	0.10	2.1	0.06	1.74
Medical visits	10.8	0.26	9.1	0.29	3.04
Emergency room visits	0.7	0.11	0.7	0.07	<1
Charge					
Hospital stays	$1,028.4	0.02	$1,645.4	0.00	<1
Medical visits	227.8	0.25	202.0	0.11	1.76
Prescribed medicine and other medical expenses[1]	105.8	0.32	98.4	0.14	*4.91

[1]Includes prescribed medicines, eyeglasses, orthopedic appliances, hearing aids, diabetic supplies, and ambulance services.

NOTE: Standard deviations are based on annual data; corrrelations are between annualized round 1 and round 5 data. *F* tests for differences in the regression coefficients relating round 1 and round 5 variables.

Data Quality and Data Collection Frequency

As was noted earlier, the data on events contained missing, inconsistent, and out-of-range values. The comparisons between the four- and five-interview groups on means, standard deviations, and correlations excluded such values. This section examines the number of problematic values by group. For each round, the dependent variable in the analysis is the number of legitimate items (items with values that were not missing, were consistent with other questions, and were within-range) on the five event variables (disability days, hospital stays, medical visits, dental visits, and emergency room visits). Table E lists the mean number of legitimate values for the four- and five-interview groups. The table includes data from each round and averages across all rounds.

There is an insignificant, although consistent, main effect for data collection frequency on the average measure of data quality; across all rounds, respondents who were interviewed four times provided legitimate values for 4.95 of the five event variables compared with 4.97 for the five-interview group. (The F values in Table E are for the data collection frequency main effect, controlling for six background characteristics and their interactions with the data collection frequency variable.) As with those differences between interview groups noted earlier, this trend was already present in round 1 of the survey.

Table E

Mean number of items with legitimate values for five event variables, by number of interviews and round

Round	Number of interviews: 4	Number of interviews: 5	F
All rounds	4.95	4.97	1.35
Round 1	4.93	4.97	<1
Round 2	4.96	4.96	<1
Round 3	4.94	4.97	3.69
Round 4	...	4.98	<1
Round 5	4.97	4.99	<1

NOTE: The event variables include disability days, hospital stays, medical visits, dental visits, and emergency room visits.

Impact of Length of the Recall Period

Round and Length of Reference Period

If forgetting has an appreciable effect on the level of reporting, then respondents with longer reference periods should report fewer events on the average than respondents with shorter reference periods. By contrast, if telescoping has a major impact on reporting, respondents with longer reference periods (those who are more susceptible to telescoping errors) should report more events. Data relevant to these hypotheses are displayed in Table F, which gives the mean level of reporting for five event variables by round and length of reference period. (Round 5 data were tabulated two ways: according to the length of the reference period and according to the length of the period from the beginning of the reference period through the date of the round 5 interview.)

A casual inspection of Table F reveals that the length of the reference period (and, in final interview, the length of the entire recall period) had weak and inconsistent effects on the reported disability days and use of medical services. Respondents with the longest reference periods did not consistently have the lowest reported number of disability days and utilization; likewise, respondents with the shortest reference periods did not consistently report the highest. The data in the table were analyzed using one-way analysis of variance. These analyses support the overall impression given by the means. Of the 30 analyses, only one (involving round 1 disability days) shows a significant effect for the length of the reference period ($F = 9.00$).

This finding has been foreshadowed by the results regarding data collection frequency. If longer reference periods produce lower levels of reporting, then the four-interview group, which had *longer* reference periods on the average, should have reported lower levels of events. In fact, the respondents interviewed four times tended to report higher levels of events (Table C). It is possible that *within* each data collection frequency group, the effects of length of the reference period are more marked; however, the results of two-way analysis that examine the effects of both variables do not differ from those presented already and so are not shown.

A closer examination of Table F suggests that length of the reference period had the greatest (although still generally not significant) impact on round 1 reporting. R^2 values from the analysis of variance (not shown) tend to confirm this impression; except for emergency room visits, the R^2 values are larger for round 1 data than for data from subsequent rounds. Even in round 1, the effects of length of the reference period are inconsistent across the different types of events. For disability days and hospital stays, the round 1 means follow a pattern consistent with the forgetting curve, decreasing as reference periods increase. For dental and medical visits, the pattern suggests that reports were affected by telescoping errors; for these events, the highest round 1 means were reported for the longest reference periods. In round 1, respondents had neither the summary nor the calendar-diary as memory aids. The relatively greater impact of length of the reference period in round 1 may indicate that these devices reduced memory errors in later rounds.

Distribution of Events Within the Reference Period

Both common sense and psychological research on memory argue that longer reference periods should produce more forgetting and, as a result, lower levels of reporting. The analyses presented thus far reveal little evidence of the effects of forgetting. This section takes another look at the issue by examining the distribution of events within the reference period. Because respondents were asked the date of each hospital stay, medical visit, and dental visit reported, it was possible to classify each event relative to the date of interview.

Events were classified into 2-week periods (for hospital stays, which were relatively rare, 4-week periods were used), beginning with the 2 weeks immediately prior to the interview. On the assumption that the events themselves are distributed evenly over time, the distribution of reported events should reflect the shape of the forgetting curve. Because individual respondents were asked to report for different lengths of time, the simple counts of events are somewhat misleading. A respondent with an 8-week reference period could have reported events in the first four 2-week periods, whereas a respondent with a 14-week reference period could have reported events during the first seven 2-week periods. Thus, it made sense to convert the counts for each 2-week

Table F

Mean number of events reported, by length of reference period, events, and round

Event and round	Length of reference period in days: 0–29	30–59	60–89	90–115	120–149	150–179
Annual disability days						
Round 1	---	9.11	8.05	5.93	4.28	---
Round 2	---	5.27	3.78	4.29	4.34	3.53
Round 3	---	2.96	3.68	3.44	3.34	---
Round 4	---	3.72	3.37	4.14	---	---
Round 5	6.99	5.31	5.28	5.38	3.56	---
Round 5 plus Lag[1]	---	11.00	5.32	5.57	5.15	5.82
Annual hospital stays						
Round 1	---	0.19	0.16	0.16	0.04	---
Round 2	---	0.00	0.12	0.13	0.16	0.15
Round 3	---	0.11	0.13	0.14	0.18	---
Round 4	---	0.08	0.13	0.13	---	---
Round 5	0.09	0.14	0.15	0.15	0.13	---
Round 5 plus 1 Lag[1]	---	0.06	0.13	0.14	0.14	0.17
Annual dental visits						
Round 1	---	1.39	1.45	1.45	1.63	---
Round 2	---	1.58	1.35	1.28	1.18	1.17
Round 3	---	1.13	1.19	1.23	0.90	---
Round 4	---	1.23	1.24	0.89	---	---
Round 5	0.87	0.96	1.01	1.13	0.96	---
Round 5 plus Lag[1]	---	0.12	0.94	1.11	0.94	0.91
Annual medical visits						
Round 1	---	5.66	5.44	5.44	6.42	---
Round 2	---	9.07	5.17	5.44	5.38	5.01
Round 3	---	4.67	4.58	5.43	4.43	---
Round 4	---	5.31	4.73	3.77	---	---
Round 5	5.37	4.61	4.74	5.03	4.08	---
Round 5 plus Lag[1]	---	7.43	4.68	4.91	4.67	4.43
Annual emergency room visits						
Round 1	---	0.32	0.29	0.28	0.26	---
Round 2	---	0.26	0.30	0.28	0.30	0.33
Round 3	---	0.24	0.23	0.27	0.22	---
Round 4	---	0.24	0.23	0.25	---	---
Round 5	0.14	0.20	0.20	0.27	0.08	---
Round 5 plus Lag[1]	---	0.32	0.19	0.20	0.22	0.22

[1]The interval from the beginning of the reference period to the round 5 interview date.

NOTE: All means are based on 75 or more events.

period into averages, by dividing the simple count of events for each period by the number of respondents who could have reported events in that period. (The reference period for a respondent did not necessarily divide evenly into 2-week periods. For example, a respondent's reference period might have covered 13 weeks. Such a respondent was considered "eligible" to have reported on six 2-week periods, and any events he or she reported in the 13th week prior to the interview were dropped from this analysis.)

Figures 1, 2, and 3 present the annual averages for hospital stays, medical visits, and dental visits. Figures showing rates per person for each round are included in Appendix IV. (Disability days were not included in this analysis, because respondents were not asked to date them. Emergency room visits are not tabulated separately but are included in the figures for medical visits.) The averages for the entire year were calculated in the same manner as round-by-round averages (that is, the total number of events reported for a period was divided by the total number of respondents eligible to report on that period). All figures plot averages based on 1,000 or more respondents. Because the figures are based on unweighted data, they should be regarded as descriptive.

For hospital stays (Figure 1), the annual curve shows an exponential pattern. The dropoff from the most recent 4-week period to the next period (5 to 8 weeks before the interview) is quite sharp; after that, the curve is more or less level with minor fluctuations. The annual curves for medical and dental visits (Figures 2 and 3)

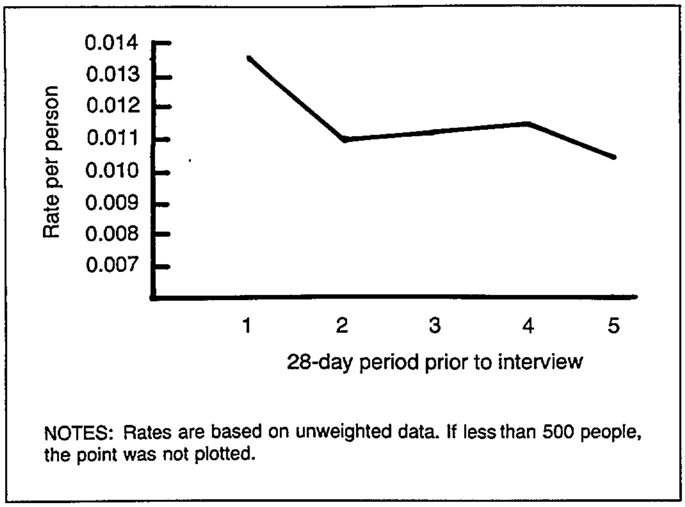


Figure 1

Annual rate of hospital stays per person, by period prior to interview

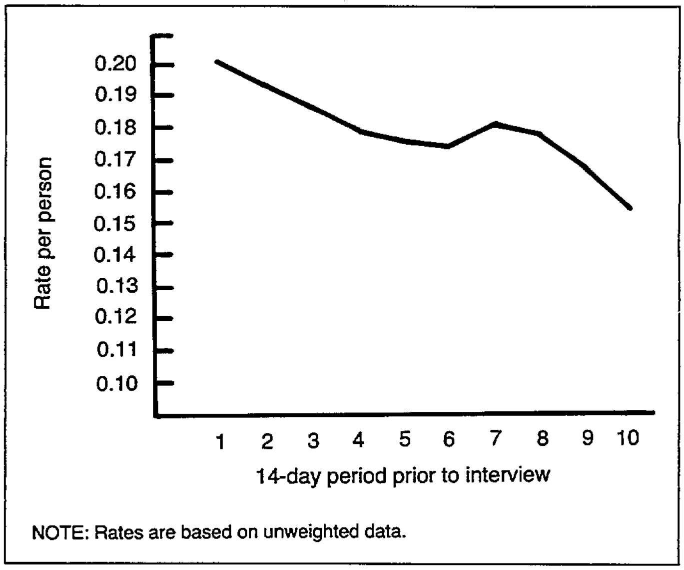


Figure 2

Annual rate of medical visits per person, by period prior to interview

follow a more linear pattern, with fluctuating drops throughout both curves. Although the three annual curves differ in shape, they share an important characteristic—all show a fairly shallow slope, which indicates a slow rate of forgetting. The curves may even exaggerate the rate of forgetting, because the "internal telescoping" effect (Neter and Waksberg, 1964), in which respondents move events forward in time within the reference period, would also produce a curve with higher levels of reporting in the more recent periods than in the more distant periods. The slopes of Figures 1–3 may thus reflect not only the effects of forgetting but also the effects of internal telescoping.

The rate of forgetting for these events is apparently slow and may even be somewhat slower than the figures imply. This finding helps explain why extending the reference period has so little effect on levels of reporting. Even the curve for hospital stays, which shows a steep drop at the beginning, quickly stabilizes; after the first two 4-week periods, little additional forgetting seems to occur.

One other phenomenon is illustrated by the figures. For round 1, two of the three curves (for medical and dental visits) show an increase in the average number of reported events for the most distant period plotted. This increase may have resulted from the telescoping of events that happened before the January 1 reference date for round 1, or from the superior recall of events that happened around that date: The New Year's holiday may have provided a useful contextual cue that helped respondents to remember events that occurred around that time.

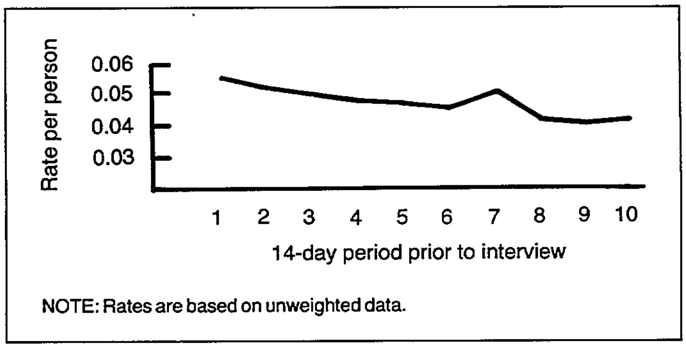


Figure 3

Annual rate of dental visits per person, by period prior to interview

Use of the Summary

One procedure used in NMCUES was the presentation of a summary to respondents as a device to improve the quality of data on medical care utilization and expenditures. The summary was presented to respondents during the interviews at rounds 2 through 5, and consisted of information given during previous rounds about use of medical care services and charges. When presented with the summary, respondents were allowed to update the information given at the previous rounds by changing, adding, or deleting responses.

This section presents an evaluation of the usefulness and effectiveness of the summary. Two strategies are used. The first is a qualitative evaluation of the different kinds of changes made in the summary. This evaluation includes an analysis of changes in information that occurred for each round and for specific event types. For example, the qualitative examination assesses the percent of respondents who replaced data missing in their original report.

The second strategy used to evaluate the effectiveness of the summary is to examine its effect on different types of respondents. Three specific background variables are examined: age, proxy status, and Medicare or Medicaid coverage. An important issue is the impact of the summary on respondents who served as proxies for the medical care user. The accuracy of proxy-generated information has sometimes been questioned in the survey literature (Berk, Wilensky, and Cohen, 1984; Haase and Wilson, 1972). Because it is efficient to use proxies in studies such as NMCUES, it is crucial to know whether the use of a summary can improve the quality of information given by proxies.

Change by Event Type

This section presents a qualitative analysis of the summary's impact on the quality of data reporting. Quality is defined in terms of the type of change made on the summary. The analysis is restricted to four types of reported charges—prescribed medicines and other medical expenses, hospital stays, dental visits, and medical visits—and examines four types of change. "Missing to missing" indicates that the respondent did not give charge information about the event, either at the original interview or during the summary review in the following rounds. If most initially missing charge data remain missing, then the summary was of limited effectiveness. "Missing to legitimate" indicates that an amount was not given at the original round but was reported in the summary review. A high rate of such changes demonstrates the effectiveness of the summary. "Legitimate to legitimate" means that an amount was given both times, although the amount is not necessarily the same. These changes are considered in more detail later in the report. Finally, "legitimate to missing" means that an amount given at the initial round was retracted during the summary review. A large percent of charge reports in this category would suggest that the original household interview data were unreliable. The analyses are organized by individual rounds for all four types of medical charges and by specific type of event for prescribed medicine and other medical expenses, and medical visits. All of the percents are based on weighted data.

Table G shows the percent of each type of change for the four types of charges. Across all types of charges, the percent of events with missing responses both before and after the summary review varies from 16.6 percent for dental visits to 31.9 percent for medical visits. The percent of events for which legitimate responses were given, both initially and at the presentation of the summary, also varies considerably, fluctuating from 41.1 percent for hospital stays to 72.1 percent for prescribed medicines and other medical expenses. Given the initial level of missing charge data, there is ample opportunity for the summary to capture additional data. An examination of the proportions of events that are revised from missing charge data to legitimate values reveals large fluctuations, with percents ranging from 5.4 percent for prescribed medicines and other medical expenses to 33.3 percent for dental visits.

That prescribed medicines and other medical expenses had the lowest percent of change from missing to legitimate charge information is probably a reflection of the high rate of legitimate responses obtained during the initial interview. The high rate of initial legitimate responses for this category may be attributed in turn to the relative ease of obtaining bills for specific items, such as medicine, glasses, hearing aids, and ambulance services. The ease of keeping records about such specific prescribed items contrasts with other types of medical charges, such as hospital stays, which involve multiple

Table G
Percent distribution of charge data revised in the summary by type of revision, according to type of charge

Type of charge	Total	Type of revision: Missing to Missing	Type of revision: Missing to Legitimate	Type of revision: Legitimate to Missing	Type of revision: Legitimate to Legitimate
	Percent distribution				
Prescribed medicine and other medical expenses	100.0	22.0	5.4	0.4	72.1
Medicine	100.0	23.3	4.4	0.4	71.9
Eyeglasses	100.0	6.2	20.3	0.8	72.8
Orthopedic appliances	100.0	11.3	9.3	1.3	78.2
Hearing aids	100.0	7.1	4.3	1.9	86.6
Diabetic supplies	100.0	10.9	2.6	0.4	86.1
Ambulance services	100.0	33.3	12.9	1.1	52.7
Missing[1]	100.0	12.3	20.6	1.6	65.6
Hospital stays	100.0	39.1	19.5	0.3	41.1
Dental visits	100.0	16.6	33.3	0.2	49.9
Medical visits	100.0	31.9	12.6	0.4	55.2
Emergency room visits	100.0	38.3	15.3	0.3	46.1
Hospital outpatient visits	100.0	54.2	15.2	0.3	30.3
Physician visits	100.0	25.9	12.0	0.3	61.8
Other medical provider visits	100.0	36.9	12.4	0.4	50.3

[1]Neither a prescribed medicine nor another medical expense was identified by the respondent.

NOTE: Table entries are weighted percents. Percents may not add to 100.0 because of rounding.

expenses that may be specified less clearly than the charge for individual items. When the different subgroups of prescribed medicine and other medical expenses are examined, it seems that the summary was most effective in obtaining useful information about charges relating to glasses. An additional 20.3 percent of these charges were accounted for after the summary review.

Considerably more of the charge data were missing initially for hospital stays and dental visits than for prescribed medicines and other medical expenses. Nearly 60 percent of the hospital stays and 50 percent of the dental visits had missing charges in the initial interview. After the summary revisions, charges were available for an additional 19.5 percent of the hospital stays and 33.3 percent of the dental visits (Table G). Despite the evident effectiveness of the summary in obtaining additional charge information, 39.1 percent of all hospital stays still have missing charges even after the summary revisions.

Although the overall impact of the summary on charge information concerning medical visits was less dramatic than for hospital stays and dental visits, some variations occurred among the different types of visit (Table G). The most frequent revisions to more complete information concerned charges for emergency room and hospital outpatient visits. The expenses for an approximate additional 15 percent of the visits were reported for these two kinds of medical services. The improvement rate for reporting about visits to physicians and other medical providers was approximately 12 percent.

Among all four types of event, the percent of deletions of charge information in the summary is very low, ranging from less than 0.25 percent in dental visit reports to 1.9 percent for hearing aids. A high percent of changes of this type would suggest that the original reports were unreliable.

Overall, the summary review process produced more improvement in completeness of information for dental visits and for hospital stays than for prescribed medicines and other medical expenses or medical visits. Reports on the prescribed medicine and other medical expense category showed the least improvement in completeness. If only the improvement in completeness rates is considered, the summary seems to have had a negligible effect on reporting about prescribed medicine and other medical expenses. However, the rates for other types of medical charges are encouraging.

It is also useful to examine more closely the events for which legitimate charge information was available at both the original interview and at the summary update. The purpose of this examination is to assess changes in charge estimates produced by the summary. Table H shows the extent to which legitimate charge reports given during the summary were greater than, less than, or the same as charges given at the initial interview for each of the types of charges. What is most apparent from this table is that it was quite rare for a respondent to give a revised charge figure that was greater than the original report. For all types of events, respondents were likely either to leave the original response unchanged or to give a lower figure during the summary review. The least change occurred for prescribed medicine and other medical expenses, perhaps again reflecting the easy availability of this type of charge information; on the other hand, the most change occurred for hospital stays, perhaps reflecting the complexity of this type of medical expense.

Table H

Percent distribution of summary records revised from the original record by result of summary revision, according to type of charge

Type of charge	Total	Result of summary revision: Summary less than original	Summary equal to original	Summary greater than original
		Percent distribution		
Hospital stays	100.0	60.2	38.7	1.1
Dental visits	100.0	50.9	48.1	1.0
Medical visits	100.0	45.3	54.0	0.7
Prescribed medicine and other medical expenses[1]	100.0	28.6	70.8	0.6

[1]Includes prescribed medicines, eyeglasses, orthopedic appliances, hearing aids, diabetic supplies, and ambulance services.

NOTE: Percents are based on weighted data and reflect only those respondents who gave a valid charge on both the original and summary prompted interviews.

Change by Subgroup and Round

Table 2 presents the rates of change in each of the qualitative categories for each round, and Table 3 presents rates for medical care users in different age, insurance coverage, and proxy status groups.

Except for hospital stays, the summary review seemed to have a similar impact on data from all four rounds (Table 2). In the case of hospital stays, the impact of the summary varied with the source of the data (Table 3). When persons responded for themselves, the summary had little impact across the four rounds: self-respondents had puzzlingly high levels of initially missing responses and had relatively little to add (10 to 11 percent) when prompted by the summary. Proxy respondents, on the other hand, supplied new information after the summary about hospital stays in 20.2 percent of the cases.

For events other than hospital stays, there was some tendency for the round 1 data to show higher missing data rates initially, but the differences across rounds are not great (Table 2). Again with the exception of information on hospital stays, the summary may have had a somewhat greater impact on round 1 data; for the other three major charge categories, the highest rates of change from missing to legitimate occurred with round 1 data.

The summary did seem to have a differential impact by age group (Table 3). For each type of medical event, the respondents with the highest percent of missing data after the summary update were under 16 years of age. This group also tended to show the highest rates of missing data initially. Aside from these results, the effect of the summary for different age groups does not seem to follow a consistent pattern.

The effect of the summary also varied depending on whether the care user was covered by Medicare or Medicaid. The charges for all four categories were more likely to be missing initially for persons covered by either program than for those covered by neither program; moreover, the charges were less likely to be provided during the summary review. Thus, the summary was less effective for persons covered by Medicare or Medicaid.

Impact of the Summary on Estimates of Medical Care Charges

An additional question posed by the use of the summary concerns its impact on the medical care charge estimates themselves. Put succinctly, the question is whether the summary affected the estimates.

The impact of the summary on charge estimates can be assessed in three ways. The first involves computing average charge estimates using only initial reports. The same averages can also be computed with the revised values from the summary; then the two sets of averages can be compared. The second method involves examining the means-of-the-difference scores calculated by subtracting the initially reported charges from the charges reported after the summary for those individuals who gave legitimate charge figures in both places. This method is similar to the analysis reported in Table H except that this method examines the amount as well as the direction of change.

A third method of examining the effect of the summary on charge estimates is to compute a "best estimate" score. In the previous section, it was shown that a small number of events had initial charge figures but not summary values. The best estimate score maximizes the amount of nonmissing data by using the summary-generated response when it is available, and the original response when the summary is not available. The number of events and average charge generated by the best estimate will be compared to the values generated using only the initial estimates and only the summary.

The figures in Table J reflect the use of three different methods to derive average annual medical care charge estimates and the number of reported events on which they are based. This table shows explicitly the exact nature of the qualitative changes described in the preceding section.

Prescribed Medicine and Other Medical Expenses

When the charges for the prescribed medicine and other medical expenses category were first considered, the summary resulted in additional charge data for 2,822 events. The best estimate procedure provided charge data for another 242 events over and above those added by the summary. The more detailed figures in Table 4

Table J

Number of reported events and mean annual charges, by events and sources of interview data

Source of interview data	Event: Prescribed medicine and other medical expenses	Hospital stays	Dental visits	Medical visits
Original		Percent distribution		
Number[1]	40,941	1,067	11,326	46,234
Mean charge[2]	$16.55	$2,121.39	$44.79	$30.57
Summary				
Number[1]	43,763	1,573	18,718	56,287
Mean charge[2]	$16.00	$2,005.23	$62.62	$35.08
Both original and summary				
Number[1]	40,699	1,059	11,277	45,964
Mean difference	$-0.20	$-35.56	$-3.01	$-0.13
Best estimate				
Number[1]	44,005	1,581	18,767	56,557
Mean charge[2]	$16.01	$2,003.03	$64.64	$35.15

[1]Data are unweighted.
[2]Data are weighted.

show that the summary had its greatest effect in round 1. Among individual event types, the summary had the most impact in providing additional information for reports about medicines.

An examination of differences between the initial and summary charges shows that the summary charges were lower for every type of expense except hearing aids. (For hearing aids, the summary values were considerably higher than the initial charge figures.) The overall mean difference for those giving both initial and summary values was 20 cents, with lower charges after the summary revisions. The resulting *t*-ratio of the charge difference in the total sample was -3.45, indicating a significant difference ($p<0.01$). Although small, the difference between the initial report and the summary revision seems to be reliable.

Hospital Stays

Reported hospital stays followed a similar pattern. Over all rounds, charge data were available for an additional 506 visits after the summary revisions (Table J). The best estimate increased the useful data by another eight visits. The detailed data in Table 4 show that, once again, the summary had the most impact on round 1 charge data. Again, the average charge given for the summary was in most cases lower than for the original reports. The variation in the differences between initial and summary charge estimates was much greater for hospital stays than for prescribed medicines and other medical expenses. The *t*-ratio of the mean difference score to its standard error was -1.68—not significant, although the negative difference persisted across three of the four rounds.

Dental Visits

For dental visits, the number of events with complete charges increased by 7,392 after the summary. Forty-nine more cases were added when the best estimate was used. The difference scores for dental visits display a pattern similar to the one observed for prescribed medicine and other medical expenses and hospital stays. The overall mean difference of -3.01 is significantly different from zero ($t = -2.23$). When the average initial estimates are compared with the average summary estimates, however, the initial estimates are lower in all rounds, a pattern that differs from that found in reports about other medical events. This variation can be explained by the relatively large number of individuals (33 percent) who revised a missing response to a legitimate response (Table G). Apparently, the dental visits with initially missing data are costlier than other dental visits are.

Medical Visits

Table J also shows the charge estimates for medical visits. Overall, the summary was responsible for adding charge data for 10,053 events. Use of the best estimate resulted in another 270 nonmissing cases. The biggest difference in the charge data occurred in the second round, in which data were obtained from the summary for 3,069 additional medical visits (Table 4). Among individual types of medical visits, physician visits showed the biggest improvement in reporting.

As for dental visits, the averages based on all summary reports for medical visits were higher than those based on initial reports. Again, this difference must be related to the 12.6 percent of reports revised from initially missing to nonmissing responses, following use of the summary (see Table G). The average mean change for respondents who revised their initial report (-0.13) is not significant.

When the best estimate is considered across all events, two results are found. First, use of the best estimate adds relatively few events with nonmissing charge data over and above those added by the summary. The best estimate yields an increment that ranges from 270 for medical visits to 8 for dental visits. Second, the means for the best estimate vary little at all from the means computed using the summary data alone. It seems that the summary data alone provide estimates at least as good as the estimates that use a combination of other available data.

Discussion and Recommendations

Effect of Data Collection Frequency

Those respondents interviewed four times appear to have more disability days and higher utilization than do those interviewed five times. For the most part, however, the differences between the groups are not significant. The pattern for data on charges is similar, although somewhat less consistent. Because the two groups of respondents do not constitute random subsamples, analysis of variance and covariance were carried out to compensate for any pre-existing differences. Although these analyses indicate that the data collection frequency effect is generally not significant, higher levels of disability days and utilization and charges are evident in the four-interview group. Additional analyses indicate that less frequent interviews do not result in a statistically significant loss in data quality; further, correlations between variables tend to be somewhat stronger in the four-interview group than in the five-interview group (although, again, the differences are not statistically significant).

The results from these analyses are surprising, because the four-interview group reported on longer periods of time in each interview than the five-interview group did; therefore, more forgetting would be expected in this group, and thus, lower reported events and charges. Several analyses to assess the impact of the length of the recall period show that this variable had weak and inconsistent effects on the level of reporting, especially after the first round of data collection. Analyses of the distribution of events within the reference period suggest that the rates of forgetting for those events are fairly slow.

Two hypotheses drawn from the methodological literature (Cohen and Burt, 1984; Neter and Waksberg, 1964) might explain why less frequent interviews do not result in less reporting. Less frequent interviews might produce more telescoping errors (where events that occurred before the reference period are inappropriately reported), thus offsetting the effects of forgetting. There was little evidence, however, for telescoping errors after round 1; the summary review procedure seemed to have reduced or eliminated such errors effectively. The other hypothesis is that the four-interview group was less "conditioned" than the five-interview group and thus provided more complete reports. This "conditioning" hypothesis suggests that differences in reporting will emerge over time, as respondents learn that reporting more events increases the length of the interview. In fact, the differences between the two groups are no more marked in subsequent rounds than they are in round 1. It can be concluded, therefore, that there is little evidence to support this hypothesis.

One interpretation consistent with the results is that the four-interview group tends to report more events simply because it has more to report. The difference in actual use of services more than compensates for any differences in forgetting. The four-interview group includes a number of respondents who were, for one reason or another, more difficult to interview during the first three rounds of the survey. It is possible that a substantial number of these respondents were difficult to interview *because* of their medical problems. Although some of the analyses attempted to take into account the level of reported events and charges in the first three rounds, it is quite possible that these models still failed to correct completely for the differences between the four- and five-interview groups; even the most extensive model generally accounted for less than 10 percent of the variation in the event and charge data from rounds 4 and 5.

Recommendation—The results concerning the effects of the length of the reference period and the distribution of events within the reference period indicate that the reference period might be lengthened without significant reductions in the level of reporting or the quality of of the data. Respondents do seem to forget events over time, but the rate of forgetting seems to be slow. Thus, four interviews may actually be preferable to five, because the cost savings outweigh any reductions in data quality that result from increased forgetting. It should be noted that this recommendation is made in the context of a survey that covers a year, deals with relatively memorable events, and employs numerous memory aids (such as the summary and the calendar-diary). Within this context, a cost-effective course would be to extend the recall period from roughly 70 days on average to roughly 90 days.

Use of the Summary

The effectiveness of the summary received considerable support from the analysis in this report. This conclusion is also supported by previous work (Holt, 1981;

Kasper, 1984). Across four types of medical care events, the summary increased the amount of charge data by about 15 percent. The percent of additional data obtained through the use of the summary ranged from 5 percent for prescribed medicines and other medical expenses to 33 percent for dental visits (Table G). The greatest increase in data generally occurred with events initially reported in round 1.

The summary also resulted in quantitative changes in medical care charge estimates. When the revised charges were compared with the initial charge reported, the summary charges were typically lower. Average charges computed using the summary reports, however, were often higher than averages computed using the initial charges. This seemingly inconsistent finding can be resolved by considering the rate of missing data in the initial reports relative to the amount of missing data after the summary revisions. Charge information for more expensive medical events apparently took longer to obtain, making the summary particularly valuable.

Recommendation—The summary seems to have been a valuable tool for obtaining additional information on medical charges. Furthermore, it may have produced additional, though less tangible benefits as a means for reducing telescoping errors. It is recommended that the summary be used in future longitudinal surveys of medical care. Even with the summary, the rate of missing data for charges was still quite high (for example, 39 percent for hospital stays). Because missing data rates were especially high for such costly events as hospital stays and among such subgroups of special policy relevance as those persons covered by Medicare or Medicaid, additional techniques should be considered to improve the rate of completeness. These techniques might include repeated callbacks and the use of supplementary medical provider or admininstrative records data.

References

Berk, M., Wilensky, G., and Cohen, S.: Methodological issues in health surveys, An evaluation of procedures used in the National Medical Care Expenditure Survey. *Eval. Rev.* 8(3):307–326, June 1984.

Bonham, G.: Procedures and questionnaires of the National Medical Care Utilization and Expenditure Survey. *National Medical Care Utilization and Expenditure Survey*. Series A, Methodological Report No. 1. DHHS Pub. No. 83–20001. Public Health Service. Washington. U.S. Government Printing Office, Mar. 1983.

Brown, N., Rips, L., and Shevell, S.: The subjective dates of natural events in very long-term memory. *Cognitive Psych.* 17(2):139–177, Apr. 1985.

Cohen, S., and Burt, V.: Data collection frequency effect in the National Medical Care Utilization and Expenditure Survey. Paper presented at Annual Meeting of the American Statistical Association. Philadelphia, Aug. 1984.

Deming, W.: On simplification of sampling design through replication with equal probabilities and without stages. *J. Am. Stat. Assoc.* 51:24–53, 1956.

Ebbinghaus, H.: (1885) *Uber das Gedachtnis*. Leipzig: Dunker, H. Ruyer and C. E. Bussenias (trans.), *Memory*. New York. Teachers College Press, 1913.

Frankel, M.: *Inference from Survey Samples: An Empirical Investigation*. Ann Arbor. Institute for Social Research, University of Michigan, 1971.

Haase, K., and Wilson, R.: The study design of an experiment to measure the effects of using proxies resources in the National Health Interview Survey. *Proceedings of the Social Statistics Section, American Statistical Association*, 1972.

Hansen, M., Hurwitz, W., and Madow, W.: *Sample Survey Methods and Theory*, Vol. II. New York. John Wiley, 1953.

Holt, M.: The use of summaries of previously reported interview data in the National Medical Care Expenditure Survey: A comparison of questionnaires and summary data for medical provider visits. in NCHSR Research Proceedings Series, *Health Survey Research Methods*, Third Biennial Conference, DHHS Pub. No. 81–3268, 1981.

Kasper, J.: Comparisons of three data sources from the National Medical Care Expenditure Survey: Household questionnaire, household summary, and medical provider survey. In C. Cannell (ed.) *Proceedings of the Fourth Conference on Health Survey Research Methods*. Rockville, Md. National Center for Health Services Research, 1984.

Kish, L.: *Survey Sampling*. New York. John Wiley, 1965.

Kish, L., and Frankel, M.: Inference from complex samples. *J. Royal Stat. Soc.: Series B (Methodological)*, 36(1):1–37, 1974.

Loftus, E.: Memory and its distortions. In A. Kraut (ed.) *G. Stanley Hall Lectures*. Washington, D.C. American Psychological Association, 1982.

Neter, J., and Waksberg, J.: A study of response errors in expenditures data from household surveys. *J. Am. Stat. Assoc.* 59:18–55, Mar. 1964.

SAS Institute, Inc.: *SAS Users Guide, Basics*, 1982 edition. Cary, N.C. SAS Institute, Inc., 1982.

Sudman, S., and Bradburn, N.: Effects of time and memory factors on response in surveys. *J. Am. Stat. Assoc.* 68:805–815, Dec. 1973.

Tourangeau, R., McWilliams, H., Jones, C., et al.: *High School and Beyond First Follow-Up Sample Design Report*. Chicago. NORC, 1983.

Tukey, J.: *Exploratory Data Analysis*. New York. John Wiley, 1978.

List of Detailed Tables

Table 1

Average annual rate per person, by number of interviews, events, charges, and round

Event, charge, and round	Number of interviews: 4	5	F	R^2
EVENT				
Disability days	Annual rate			
All rounds	5.23	4.34	3.70	0.071
Round 1	8.37	7.89	<1	0.054
Round 2	4.45	4.05	<1	0.034
Round 3	4.00	3.22	1.07	0.029
Round 4	...	[1]3.52	*5.41	0.029
Round 5	[1]5.70	[1]5.26	2.06	0.025
Hospital stays				
All rounds	0.17	0.13	**11.80	0.060
Round 1	0.23	0.15	***12.98	0.023
Round 2	0.18	0.12	<1	0.023
Round 3	0.16	0.12	<1	0.016
Round 4	...	[1]0.12	**8.23	0.015
Round 5	[1]0.16	[1]0.13	3.10	0.013
Dental visits				
All rounds	1.24	1.22	<1	0.021
Round 1	1.48	1.43	<1	0.008
Round 2	1.30	1.26	<1	0.011
Round 3	1.14	1.20	<1	0.008
Round 4	...	[1]1.23	<1	0.008
Round 5	[1]1.08	[1]0.98	1.69	0.006
Medical visits				
All rounds	5.55	4.93	<1	0.067
Round 1	6.28	5.22	<1	0.050
Round 2	5.77	5.23	<1	0.039
Round 3	5.31	4.61	<1	0.041
Round 4	...	[1]4.84	1.39	0.039
Round 5	[1]5.10	[1]4.69	<1	0.033
Emergency room visits				
All rounds	0.28	0.25	<1	0.022
Round 1	0.34	0.28	<1	0.010
Round 2	0.31	0.27	<1	0.009
Round 3	0.26	0.24	<1	0.005
Round 4	...	[1]0.24	**5.28	0.005
Round 5	[1]0.23	[1]0.19	<1	0.004
CHARGE				
Hospital stays				
All rounds	$151.41	$114.38	**8.55	0.008
Round 1	177.17	85.17	**8.81	0.010
Round 2	130.37	131.75	<1	0.003
Round 3	123.72	83.51	*5.68	0.003
Round 4	...	[1]77.72	*6.45	0.007
Round 5	[1]233.71	[1]362.81	<1	0.002
Medical visits				
All rounds	$109.77	$95.83	<1	0.062
Round 1	111.73	92.96	*6.75	0.030
Round 2	112.04	102.24	1.67	0.030
Round 3	99.01	89.49	<1	0.014
Round 4	...	[1]88.64	1.78	0.027
Round 5	[1]109.28	[1]102.38	<1	0.020

See footnotes at end of table.

Table 1

Average annual rate per person, by number of interviews, events, charges, and round—Con.

Event, charge, and round	Number of interviews: 4	5	F	R^2
Charge—Con.				
Prescribed medicine and other medical expenses[2]	Annual rate—Con.			
All rounds	$45.13	$45.32	1.23	0.136
Round 1	44.95	45.36	<1	0.032
Round 2	46.17	46.52	<1	0.075
Round 3	41.78	42.31	<1	0.069
Round 4	...	[1]44.14	2.05	0.071
Round 5	[1]48.28	[1]48.23	<1	0.059

[1]The rate for round 5 in the 4-interview group is compared with the rate for round 4 in the 5-interview group and for round 5 in the 5-interview group.
[2]Includes prescribed medicines, eyeglasses, orthopedic appliances, hearing aids, diabetic supplies, and ambulance services.

Table 2

Percent distribution of charge data revised in the summary by type of revision, according to charges and round

Charge and round	Total	Type of revision: Missing to Missing	Missing to Legitimate	Legitimate to Missing	Legitimate to Legitimate
Prescribed medicine and other medical expenses[1]	Percent distribution				
All rounds[2]	100.0	22.0	5.4	0.4	72.1
Round 1	100.0	25.6	9.6	0.7	64.0
Round 2	100.0	21.1	5.1	0.5	73.2
Round 3	100.0	21.7	4.5	0.4	73.5
Round 4	100.0	21.1	4.3	0.4	74.3
Hospital stays					
All rounds[2]	100.0	39.1	19.5	0.3	41.1
Round 1	100.0	42.4	21.2	0.3	36.0
Round 2	100.0	35.7	20.4	0.5	43.4
Round 3	100.0	39.6	22.9	0.4	37.1
Round 4	100.0	40.8	25.9	0.3	33.0
Dental visits					
All rounds[2]	100.0	16.6	33.3	0.2	49.9
Round 1	100.0	17.3	37.2	0.2	45.3
Round 2	100.0	15.0	32.7	0.3	52.0
Round 3	100.0	16.6	35.8	0.3	47.3
Round 4	100.0	17.3	33.0	0.2	49.5
Medical visits					
All rounds[2]	100.0	31.9	12.6	0.4	55.2
Round 1	100.0	31.4	14.6	0.8	53.3
Round 2	100.0	31.7	12.3	0.4	55.6
Round 3	100.0	32.5	12.8	0.2	54.5
Round 4	100.0	32.2	13.6	0.2	54.0

[1]Includes prescribed medicines, eyeglasses, orthopedic appliances, hearing aids, diabetic supplies, and ambulance services.
[2]"All rounds" figures include data for round 5.

Table 3

Percent distribution of charge data revised in the summary by type of revision, according to charges and selected user characteristics

Charge and characteristic	Total	Type of revision: Missing to Missing	Missing to Legitimate	Legitimate to Missing	Legitimate to Legitimate
Prescribed medicine and other medical expenses[1]			Percent distribution		
Age:					
Under 16 years	100.0	25.2	5.1	0.3	69.4
16–18 years	100.0	24.9	6.0	0.5	68.6
19–24 years	100.0	21.7	6.1	0.3	71.8
25–54 years	100.0	21.9	5.7	0.4	72.0
55–64 years	100.0	20.1	5.2	0.4	74.4
65 years and over	100.0	21.5	5.2	0.6	72.7
Proxy status:					
Self-respondent only	100.0	21.5	5.1	0.5	72.9
Partly proxy	100.0	21.7	5.7	0.4	72.3
Proxy only	100.0	23.7	5.9	0.4	70.0
Medicare or Medicaid coverage:					
Medicare or Medicaid	100.0	31.9	4.4	0.6	63.2
Neither coverage	100.0	17.2	5.9	0.4	76.6
Hospital stays					
Age:					
Under 16 years	100.0	50.4	18.0	0.4	31.2
16–18 years	100.0	37.2	21.5	0.0	41.3
19–24 years	100.0	34.6	17.2	0.3	47.9
25–54 years	100.0	34.0	19.9	0.3	45.9
55–64 years	100.0	36.7	19.2	0.9	43.2
65 years and over	100.0	44.9	20.7	0.2	34.3
Proxy status:					
Self-respondent only	100.0	76.9	10.6	0.0	12.5
Partly proxy	100.0	44.4	9.8	0.0	45.8
Proxy only	100.0	36.4	20.2	0.4	43.0
Medicare or Medicaid coverage:					
Medicare or Medicaid	100.0	57.6	16.6	0.1	25.7
Neither coverage	100.0	29.9	21.0	0.4	48.6
Dental visits					
Age:					
Under 16 years	100.0	22.1	31.8	0.1	46.0
16–18 years	100.0	15.0	43.6	0.2	41.2
19–24 years	100.0	15.6	30.7	0.2	53.5
25–54 years	100.0	15.4	33.9	0.3	50.5
55–64 years	100.0	12.6	35.4	0.2	51.8
65 years and over	100.0	13.2	25.8	0.3	60.7
Proxy status:					
Self-respondent only	100.0	15.3	32.2	0.3	52.2
Partly proxy	100.0	15.7	34.6	0.3	49.5
Proxy only	100.0	18.6	33.4	0.1	47.9
Medicare or Medicaid coverage:					
Medicare or Medicaid	100.0	34.7	20.7	0.3	44.3
Neither coverage	100.0	14.1	35.0	0.2	50.7

See footnotes at end of table.

Table 3

Percent distribution of charge data revised in the summary by type of revision, according to charges and selected user characteristics—Con.

Charge and characteristic	Total	Type of revision: Missing to Missing	Missing to Legitimate	Legitimate to Missing	Legitimate to Legitimate
Medical visits	Percent distribution				
Age:					
Under 16 years	100.0	36.9	8.6	0.6	53.9
16–18 years	100.0	34.3	13.0	0.1	52.6
19–24 years	100.0	34.6	17.6	0.2	47.6
25–54 years	100.0	28.1	14.7	0.3	57.0
55–64 years	100.0	29.4	11.3	0.3	59.0
65 years and over	100.0	34.5	10.4	0.5	54.6
Proxy status:					
Self-respondent only	100.0	28.0	14.8	0.3	57.0
Partly proxy	100.0	33.0	11.8	0.3	54.9
Proxy only	100.0	36.6	10.0	0.4	52.9
Medicare or Medicaid coverage:					
Medicare or Medicaid	100.0	48.4	7.9	0.5	43.3
Neither coverage	100.0	26.2	14.2	0.3	59.2

[1]Includes prescribed medicines, eyeglasses, orthopedic appliances, hearing aids, diabetic supplies, and ambulance services.

Table 4

Number of reported events and mean annual charges, by sources of interview data, events, round, and charges

Event, round, and charge	Source of interview data: Original Number	Original Mean	Summary Number	Summary Mean	Both original and summary Number	Both original and summary Mean	Best estimate Number	Best estimate Mean
Prescribed medicine and other medical expenses[1,2]								
Round:								
All rounds[3]	40,941	$16.55	43,763	$16.00	40,699	$− 0.20	44,005	$16.01
Round 1	7,723	15.87	8,788	14.81	7,636	− 0.77	8,875	14.81
Round 2	11,291	18.00	11,994	17.41	11,211	− 0.08	12,074	17.40
Round 3	8,611	16.73	9,083	16.39	8,570	− 0.11	9,124	16.42
Round 4	5,582	15.46	5,883	15.09	5,556	− 0.03	5,909	15.14
Charge:								
Medicine	37,219	11.94	39,309	11.76	37,021	− 0.11	39,507	11.73
Eyeglasses	2,272	67.66	2,879	54.07	2,249	− 1.06	2,902	54.31
Orthopedic appliances	479	61.09	523	57.04	471	− 0.23	531	56.87
Hearing aids	227	68.22	233	72.17	222	− 0.33	238	71.01
Diabetic supplies	476	21.72	487	21.56	473	0.14	490	21.66
Ambulance services	182	80.05	222	75.87	179	0.39	225	75.76
Missing[4]	86	90.07	110	57.38	84	− 18.49	112	57.22
Hospital stays[5]								
Round:								
All rounds	1,067	2,121.39	1,573	2,005.23	1,059	− 35.56	1,581	2,003.03
Round 1	196	1,846.76	311	1,651.76	194	− 159.18	313	1,645.09
Round 2	330	1,906.18	484	1,879.94	327	3.20	487	1,860.76
Round 3	206	1,903.78	335	1,927.54	204	− 0.66	337	1,939.81
Round 4	98	1,725.37	176	1,772.72	97	− 78.62	177	1,774.98
Dental visits[6]								
Round:								
All rounds	11,326	44.79	18,718	62.62	11,277	− 3.01	18,767	64.64
Round 1	2,162	54.78	3,912	65.52	2,152	− 13.12	3,922	62.44
Round 2	3,629	42.17	5,862	60.43	3,608	− 0.76	5,883	60.69
Round 3	2,286	41.62	3,962	61.98	2,273	0.09	3,975	61.86
Round 4	1,423	41.10	2,359	65.73	1,419	− 1.72	2,363	65.67
Medical visits[7]								
Round:								
All rounds	46,234	30.57	56,287	35.08	45,964	− 0.13	56,557	35.15
Round 1	8,989	28.65	11,285	31.70	8,870	0.18	11,404	31.86
Round 2	14,522	29.92	17,591	33.90	14,435	− 0.19	17,678	33.89
Round 3	9,742	30.18	11,999	34.13	9,704	− 0.58	12,037	34.18
Round 4	5,382	30.04	6,706	36.81	5,361	0.11	6,727	37.09
Charge:								
Emergency room visit	2,019	76.76	2,677	76.85	2,005	1.65	2,691	77.67
Hospital outpatient visit	2,682	54.59	3,961	61.72	2,654	0.83	3,989	61.71
Physician visit	32,829	24.79	38,911	31.35	32,659	− 0.29	39,081	31.40
Other medical provider visit	8,704	22.90	10,738	28.24	8,646	− 0.23	10,796	28.17

[1]The total number of prescribed medicines is 56,392.
[2]Includes prescribed medicines, eyeglasses, orthopedic appliances, hearing aids, diabetic supplies, and ambulance services.
[3]"All rounds" includes round 5.
[4]Neither a prescribed medicine nor another medical expense was identified by the respondent.
[5]The total number of hospital stays is 2,581.
[6]The total number of dental visits is 22,570.
[7]The total number of medical visits is 83,040.

NOTES: Numbers in the table are unweighted. Means are derived from weighted data. Means given under "Both original and summary" are the mean differences between the initial report and the summary value.

Appendixes

Contents

List of Appendix Figures

Appendix I Method for Estimating Standard Errors

This appendix describes the procedures used to estimate standard errors of statistics derived from the NMCUES data. Most researchers are familiar with the use of standard errors to assess the variability of estimates based on simple random samples. Complex sample designs, such as the NMCUES household sample design, require computational procedures different from those used for a simple random sample to estimate variances. The NMCUES household sample departs from simple random sampling in three respects. First, the NMCUES data are clustered by geographical units. Because the sample was selected in stages (for example, the selection of counties and SMSA's constituted the initial step in sample selection), the respondents are not completely independent. By contrast, in a simple random sample, the selection probabilities for each unit are independent. Second, the NMCUES sample is stratified. For example, the selection of primary sampling units (PSU's) assured proportionate representation of each region in the country. Simple random samples lack such controls. Finally, the NMCUES data are weighted; these weights compensate for differences in the selection probabilities of individual respondents. With a simple random sample, no weights are needed because all respondents have an equal chance of selection.

Each of these departures from the assumptions of simple random sampling affects the variability of sample estimates. The net impact of these departures is that, in general, estimates derived from complex samples vary considerably more than do similar estimates derived from a simple random sample with the same number of cases. Standard statistical packages—such as SPSS and SAS—assume simple random sampling or closely related designs; consequently, the results from such packages can be seriously misleading when estimates are derived from complex sample designs.

The analyses described in this report were based on standard errors estimated by the method of balanced repeated replication (BRR). The remainder of this appendix will first describe BRR and some of the principal alternatives to it and then describe more briefly the program used to carry out BRR computations.

Computational Procedures

In a simple random sample, the mean is estimated as

$$\bar{X}_{srs} = \sum_i x_i / n$$

Only the numerator is subject to sampling error; the denominator (the sample size) is taken as a fixed constant. In more complex sample designs, the mean is estimated as a ratio of estimates; for the NMCUES household sample the ratio is

$$r = \frac{\sum\sum\sum y_{hij}}{\sum\sum x_{hi}} = \frac{y}{x}$$

where

y_{hij} = the weighted value for respondent j from PSU in stratum h, and

x_{hi} = the estimated size of PSU i in stratum h.

The numerator y represents an estimate of the population total, and the denominator x represents an estimate of the population size. When cluster sizes are unequal, the overall sample size will fluctuate depending on which clusters are selected. For the same reason, the corresponding estimates of the population size will also show sampling fluctuation. Thus, for a ratio estimator, both the numerator and the denominator are subject to sampling error.

In their classic paper, Kish and Frankel (1974) distinguish three major approaches to computing standard errors for statistics based on complex design where ratio estimators must be used: Taylor Series, BRR, and jackknife repeated replication (JRR).

Standard Errors for the NMCUES Analysis

Empirical investigations that compare the three approaches reveal that each yields very similar results (see, for example, Frankel, 1971). For this reason, the decision to use one method over the others must rest primarily on practical considerations such as cost and the availability of software. Although the Taylor Series approach is often preferred because of its computational simplicity for simple estimates (such as means or proportions), the analysis in this report also examined more complex estimates, such as regression coefficients and correlations. As a result, the analysis used a program based on the method of balanced repeated replication.

This program offered a number of practical advantages. First, the authors had used the program before; thus, it was familiar and would minimize startup costs. Second, the program was known to have given accurate results (Tourangeau et al., 1983). Finally, the program was designed to be embedded in SAS, allowing analysts to use results from other SAS procedures while simultaneously calculating accurate standard error estimates.

The program—BRRVAR—was originally developed under contract to the National Center for Education Statistics (NCES). It has been used by NORC and by other researchers to analyze data from the High School and Beyond Survey. NORC's experience with the program indicates that it produces accurate results at reasonable cost. The following section provides an overview of BRR; detailed documentation of the program itself is available through NCES.

Balanced Repeated Replication

The replication approach was originally developed by Deming (1956). The principle underlying replicated sampling is quite simple. If a sample of size n is desired, g independent replicate samples are selected, each of size n/g. The variation among estimates from each replicate can be used to estimate the variance of estimates based on the entire sample. In fact, the NMCUES household sample is such a replicated sample; it consists of two independent national samples, each selected from the general-purpose national samples of the two contractors, the National Opinion Research Center and Research Triangle Institute. The replication approach is limited, however, in that the precision of the standard error estimates depends on the number of replicates in the design; with fewer replicates, the standard errors are less accurate.

Balanced repeated replication, which extends the principle of replication, is usually applied to stratified designs with two primary selections per stratum. When one primary selection is chosen from each stratum, a half sample is created; the unselected primary units form another half sample. In a design with h strata, a total of $2^{(h-1)}$ different pairs of half samples can be formed in this fashion. Each pair is referred to as a replicate. It is customary to form only a portion of the possible replicates using an orthogonal balanced design.

For any given replicate, estimates such as the ratio means, r_1 and r_2, can be computed from each half sample. The sampling variance for the overall statistic (r) can then be estimated in any of several ways (Frankel, 1971). One method compares the estimate from one half sample with the overall estimate:

$$\mathrm{Var}_k(r) = (r_{1k} - r)^2$$

where

$\mathrm{Var}_k(r)$ = the variance estimate based on replicate k,

r = an estimate based on the entire sample, and

r_{1k} = an estimate based on one of the half samples from replicate k.

The final estimate for the variance of r is the average of Var_k across all the replicates. The estimate r need not be a ratio mean; the logic of BRR applies to any type of estimate, giving the method its broad generality.

Appendix II
Definition of Terms

Age—The age of the person as of January 1, 1980.

Annualized charges—The total amount of charges reported for a class of events (such as dental visits) during a single round, projected over the entire year of the survey. The annualized figure was calculated by converting the total reported in one round into a daily rate and then multiplying by 366.

Annualized visits—The number of visits or events in a class (such as hospital stays) reported during a single round, projected over the survey year. The annualized figure was calculated by converting total visits for the round to a daily rate and then multiplying by 366.

Best estimate—A charge estimate, such as a mean or total, based on the maximum number of records. When nonmissing charge figures were available from both the household interview and the summary, the summary figure was used in preference to the initial report from the household interview. Otherwise, either the initial report or the summary value was used, depending on which was available.

Data collection frequency—The number of interviews for which a respondent was scheduled. Except for those respondents who were eligible for only a portion of 1980, all other respondents were scheduled for four or five interviews.

Dental visits—The number of reported visits to dentists, dental surgeons, oral surgeons, orthodontists, dental assistants, or other persons for dental care. This number is based on responses to items 1B and 2B of the provider probe section of the core questionnaire (Appendix III).

Disability days— The number of days in which illness or injury kept a respondent in bed all or most of the day. This figure is based on responses to item 1A of the disability days section of the core questionnaire (Appendix III).

Emergency room visits—The number of reported visits to hospital emergency rooms for medical care, based on responses to item 3B of the provider probe section of the core questionnaire (Appendix III).

Hospital outpatient department visit—The number of reported visits to hospital outpatient departments for medical care, based on responses to item 4B of the provider probe section of the core questionnaire (Appendix III).

Hospital stays—The reported number of times a person spent at least 1 night in a hospital as a patient, based on responses to item 5B of the provider probe section of the core questionnaire (Appendix III).

Housing unit—A group of rooms or a single room occupied or intended for occupancy as separate living quarters: that is, (1) the occupants did not live and eat with any other persons in the structure, and (2) there was either direct access from the outside or through a common hall; or there were complete kitchen facilities for the use of occupants only.

Key person—A key person was (1) an occupant of a national household sample housing unit or group quarters at the time of the first interview; (2) a person related to and living with a State Medicaid household case member at the time of the first interview; (3) an unmarried student 17–22 years of age living away from home and related to a person in one of the first two groups; (4) a related person who had lived with a person in the first two groups between January 1, 1980, and the round 1 interview, but was deceased or had been institutionalized; (5) a baby born to a key person during 1980; or (6) a person who was living outside the United States, was in the Armed Forces, or was in an institution at the time of the round 1 interview, but who had joined a related key person.

Length of reference period—The number of days in the period to be reported on in an interview. In round 1, the reference period began on January 1 and ended on the day of the interview. In rounds 2–4, the reference period extended from the date of the previous interview to the day of the current interview. In the final round, the reference period extended from the date of the previous interview to December 31.

Medical visits—The reported number of ambulatory visits to medical providers, including doctors, nurses, physical therapists, and laboratory technicians. This number is based on responses to the medical provider, emergency room, and hospital outpatient department visit sections of the core questionnaire.

Nonkey person —A person related to a key person who joined him or her after the round 1 interview, but was part of the civilian noninstitutionalized population of the United States at the date of the first interview is considered nonkey.

Perceived health status—The family respondent's judgment of the health of the person compared with others the same age, as reported at the time of the first interview. The categories were "excellent," "good," "fair," or "poor."

Prescribed medicines and other medical expenses—Charges for prescribed medication, and other miscellaneous expenses for items such as prescription eyeglasses, orthopedic appliances, hearing aids, diabetic supplies, and ambulance services.

Primary sampling unit (PSU)—The unit for first-stage sampling, generally consisting of individual counties, groups of adjoining counties, or standard metropolitan statistical areas (SMSA's).

Proxy respondent—A person who provided data on another person. Each person in the survey was classified into one of three groups: those who provided all the data about themselves in every round (self-respondent only); those whose data originated entirely from proxies in every round (proxy only); and all others (partly proxy). This definition differs somewhat from those used in other NMCUES reports.

Pseudostratum—A group of similar PSU's, created after the sample had been selected, to estimate variances. Each pseudostratum contains two PSU's (or a pair of grouped PSU's) that are similar in population and are from the same region of the country.

Race—The race of people 17 years of age and over reported by the family respondent; the race of those under 17 was derived from the race of other family members. If the head of the family was male and had a wife who was living in the household, her race was assigned to any children under 17 years of age. In all other cases, the race of the head of the family (male or female) was assigned to any children under 17 years of age. Race is classified as "white," "black," or "other." The "other" race category includes American Indian, Alaskan Native, Asian, and Pacific Islander.

Round—A round was the administrative term used to designate all interviews that occurred within a given period of time and that used the same instruments and procedures.

Summary of responses (summary)—A computer-generated report sent to the interviewer and reporting unit just prior to a followup interview. It contained summary information of previously reported health care, charges for the care, sources of payment, and health insurance coverage. It was designed for updating information, especially charges and sources of payment which may have not been available to the respondent at the time the health care was originally reported.

DISABILITY DAYS SECTION

Person Name:________________ # ________

1. Since (REF. DATE), did (PERSON) stay in bed because of any illness or injury?

Yes . . . . 01(A)
No. . . . . 02(2)

A. Since (REF. DATE), how many days did illness or injury keep (PERSON) in bed all or most of the day?

☐ Days (B)
None. . . . . 00(2)

B. What conditions caused (PERSON) to stay in bed since (REF. DATE)?

C. Were there any other conditions?

IF MORE THAN ONE CONDITION, AND MORE THAN 1 DAY IN A, ASK D FOR EACH CONDITION.

D. How many of the (NUMBER) days did (CONDITION) keep (PERSON) in bed all or most of the day?

B & C Condition		Cond. #	D # of Days
	CC		
	CC		
	CC		
	CC		

CODE ONE: PERSON IS 14 OR OVER. . . . . 01(2)
PERSON IS UNDER 14. . . . . 02(3)

2. Since (REF. DATE), did illness or injury keep (PERSON) from work, not counting work around the house?

Yes . . . . 01(A)
No/Doesn't work . . . . . . 02(3)

A. Since (REF. DATE), how many days did illness or injury keep (PERSON) from work all or most of the day?

☐ Days (B)
None. . . . . 00(3)

B. What conditions caused (PERSON) to miss work since (REF. DATE)?

C. Were there any other conditions?

IF MORE THAN ONE CONDITION AND MORE THAN 1 DAY IN A, ASK D FOR EACH CONDITION.

D. How many of the (NUMBER) days did (CONDITION) keep (PERSON) from work all or most of the day?

B & C Condition		Cond. #	D # of Days	E # of Days
	CC			
	CC			
	CC			
	CC			

FOR EACH CONDITION LISTED IN BOTH Q. 1 AND Q. 2, ASK E.

E. On how many of the (NUMBER) days that (PERSON) lost from work because of (CONDITION) did (PERSON) stay in bed all or most of the day?

IF ANY DAYS ENTERED IN Q. 2A, ASK F.

F. Was (PERSON) paid in full, in part, or not at all, for the day(s) missed from work?

In full 01
In part 02
Not at all. . . . 03
Self-employed. . 04
Don't know . . . 94

3. (Not counting the days [in bed/lost from work] that you already told me about), since (REF. DATE), did illness or injury cause (PERSON) to cut down on the things (PERSON) usually does?

Yes . . . . 01(A)
No. . . . . 02(NP)

A. Since (REF. DATE) how many days did (PERSON) cut down for as much as a day?

☐ Days (B)
None. . . . . 00(NP)

B. What conditions caused (PERSON) to cut down since (REF. DATE)?

C. Were there any other conditions?

IF MORE THAN ONE CONDITION, AND MORE THAN 1 DAY IN A, ASK D FOR EACH CONDITION.

D. How many of the (NUMBER) days did (CONDITION) keep (PERSON) from the things [he/she] usually does for as much as a day?

B & C Condition		Cond. #	D # of Days
	CC		
	CC		
	CC		
	CC		

NEXT PERSON

DD-3

PROVIDER PROBES

PERSON 1

The next questions deal with visits you (and members of your family) have made to dentists, doctors and other types of medical specialists since (REF. DATE). First, we will talk about dental visits.

1. Since (REF. DATE) did [you/anyone in the family, that is you, (EACH PERSON IN FAMILY)] go to a dentist?

Yes . . . . 01(A)
No. . . . . 02(2)

A. Who was this? CODE "DENTIST" IN PERSON'S COLUMN. Did anyone else go to a dentist since (REF. DATE)? — 1A — Dentist 01

B. Since (REF. DATE), how many times did (PERSON) go to a dentist? RECORD IN PERSON'S COLUMN. — B — ☐ Times

2. (Not counting the visits you just told me about), since (REF. DATE) did [you/anyone in the family] go to a dental surgeon, oral surgeon, orthodontist, dental assistant or any other person for dental care?

Yes . . . . 01(A)
No. . . . . 02(DV)

A. Who was this? CODE "OTHER DENTAL" IN PERSON'S COLUMN. Anyone else? — 2A — Other Dental. 01

B. Since (REF. DATE), how many times did (PERSON) go to such a person for dental care? RECORD IN PERSON'S COLUMN. — B — ☐ Times

DV	ENTER TOTAL OF EACH PERSON'S DENTAL VISITS (Q's 1B & 2B) IN "DV" BOX ON CONTROL CARD.	DV	

3. Since (REF. DATE) did [you/anyone in the family] go to a hospital emergency room for medical care?

Yes . . . . 01(A)
No. . . . . 02(ER)

A. Who was this? CODE "EMERGENCY ROOM" IN PERSON'S COLUMN. Anyone else? — 3A — Emergency Room. 01

B. Since (REF. DATE) how many times did (PERSON) receive treatment in a hospital emergency room? RECORD IN PERSON'S COLUMN. — B — ☐ Times

ER	ENTER TOTAL OF EACH PERSON'S EMERGENCY ROOM VISITS IN "ER" BOX ON CONTROL CARD.	ER	

4. Since (REF. DATE), did [you/anyone in the family] go to a hospital clinic or hospital outpatient department for medical care?

Yes . . . . 01(A)
No. . . . . 02(OPD)

A. Who was this? CODE "CLINIC OR OPD" IN PERSON'S COLUMN. Anyone else? — 4A — Clinic or OPD 01

B. Since (REF. DATE), how many times did (PERSON) visit a hospital clinic or outpatient department? RECORD IN PERSON'S COLUMN. — B — ☐ Times

IF PERSON WENT TO MORE THAN ONE CLINIC OR OUTPATIENT DEPARTMENT ON A SINGLE TRIP TO THE HOSPITAL, COUNT EACH CLINIC OR DEPARTMENT AS A DIFFERENT VISIT.

OPD	ENTER TOTAL OF EACH PERSON'S CLINIC OR OPD VISITS IN "OPD" BOX ON CONTROL CARD.	OPD	

PROVIDER PROBES

		PERSON 1
5. Since (REF. DATE), [were you/was anyone in the family] a patient in a hospital overnight? (Be sure to include newborn babies.) Yes . . . . 01(A) No. . . . . 02(6)		
A. Who was this? CODE "IN HOSPITAL" IN PERSON'S COLUMN. Anyone else?	5A	In hospital 01
B. Since (REF. DATE), how many different times was (PERSON) a patient in a hospital? RECORD IN PERSON'S COLUMN.	B	☐ Times
6. Since (REF. DATE), [were you/was anyone in the family] admitted as a patient to a hospital and discharged on the same day? Yes . . . . 01(A) No. . . . . 02(7)		
A. Who was this? CODE "IN AND OUT" IN PERSON'S COLUMN. Anyone else?	6A	In and out. 01
B. Since (REF. DATE), how many different times was (PERSON) admitted to and discharged from a hospital on the same day? RECORD IN PERSON'S COLUMN.	B	☐ Times
7. [Were you/was anyone in the family] a patient in a nursing home, convalescent home or similar place since (REF. DATE)? Yes . . . . 01(A) No. . . . . 02(HS)		
A. Who was this? CODE "NURSING HOME" IN PERSON'S COLUMN. Anyone else?	7A	Nursing home. 01
B. Since (REF. DATE), how many different times was (PERSON) a patient in a nursing home or similar place? RECORD IN PERSON'S COLUMN.	B	☐ Times
HS — ENTER TOTAL OF EACH PERSON'S HOSPITAL STAYS (Q's. 5B, 6B & 7B) IN "HS" BOX ON CONTROL CARD.	HS	
8. During this period did [you/anyone in the family] get any medical advice from a doctor over the telephone? Yes . . . . 01(A) No. . . . . 02(9)		
A. Who was the phone call about? CODE "TELEPHONE" IN PERSON'S COLUMN. Anyone else?	8A	Telephone 01
B. How many telephone calls were made to get medical advice about (PERSON)? RECORD IN PERSON'S COLUMN.	B	____# of calls
		DO NOT INCLUDE TELEPHONE CALLS IN V BOX.

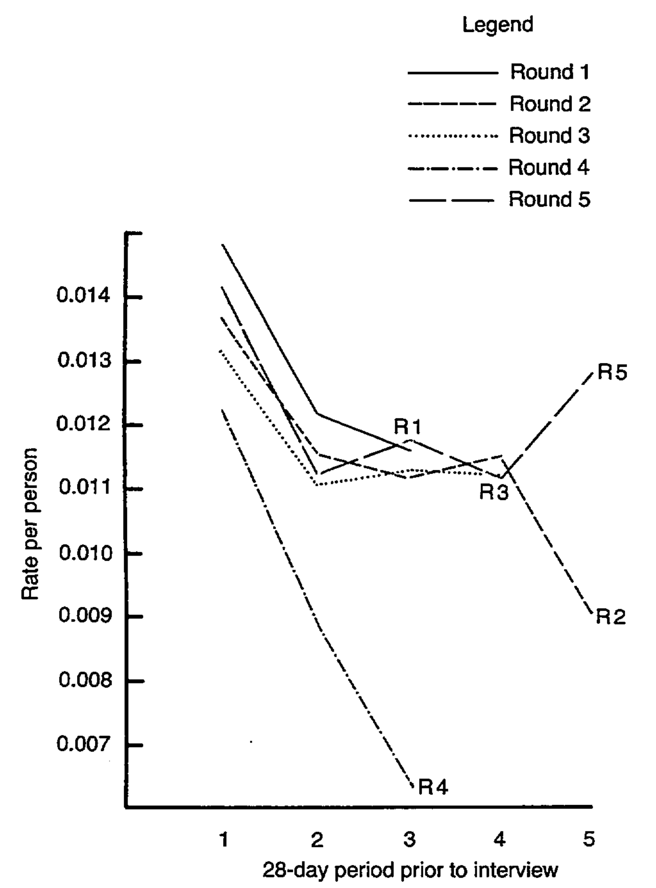
35

PROVIDER PROBES

PERSON 1

		Person 1
9. Since (REF. DATE), how many times did (PERSON) see a medical doctor? (Do not count doctors seen during visits to [an emergency room/hospital clinic or outpatient department/or while a patient in a hospital.]) RECORD IN PERSON'S COLUMN.	9	None seen. . . . . . . 00 Medical Doctor . . . . 01 ☐ Times
10. (Not counting the visits you already told me about) since (REF. DATE), did [you/anyone in the family] see any medical practitioners such as optometrists, foot doctors, chiropractors, or physical therapists? Yes . . . . 01(A) No. . . . . 02(11)		
A. Who was this? CODE "MEDICAL PRACTITIONER" IN PERSON'S COLUMN. Anyone else?	10A	Medical Practitioner . 01
B. Since (REF. DATE), how many times did (PERSON) see such a medical practitioner? RECORD IN PERSON'S COLUMN.	B	☐ Times
11. (Not counting the visits you've already told me about) since (REF. DATE), did [you/anyone in the family] receive treatment from any other medical person such as a nurse, nurse practitioner, paramedic, health aide, physician assistant, or other such medical person? Yes . . . . 01(A) No. . . . . 02(12)		
A. Who was this? CODE "MEDICAL PERSON" IN PERSON'S COLUMN. Anyone else?	11A	Medical Person 01
B. Since (REF. DATE), how many times did (PERSON) see such a medical person? RECORD IN PERSON'S COLUMN.	B	☐ Times
12. (Not counting what you have already told me about) since (REF. DATE), did [you/anyone in the family] see a psychiatrist, a psychologist, a psychiatric social worker or any other mental health person? Yes . . . . 01(A) No. . . . . 02(13)		
A. Who was this? CODE "MENTAL HEALTH PERSON" IN PERSON'S COLUMN. Anyone else?	12A	Mental Health Person . 01
B. Since (REF. DATE), how many times did (PERSON) see such a mental health person? RECORD IN PERSON'S COLUMN.	B	☐ Times
13. (Not counting the visits you've told me about) since (REF. DATE), did [you/anyone in the family] go to a doctor's office, clinic, or laboratory just for an examination, tests, shots, X-rays, or treatments? Yes . . . . 01(A) No. . . . . 02(14)		
A. Who was this? CODE "TESTS, SHOTS" IN PERSON'S COLUMN. Anyone else?	13A	Tests, Shots 01
B. Since (REF. DATE), how many times did (PERSON) go just for examinations, tests, shots, X-rays, or treatments? RECORD IN PERSON'S COLUMN.	B	☐ Times
14. (Besides the visits we've talked about) since (REF. DATE), did [you/anyone in the family] go to a health clinic, company clinic, school clinic, infirmary, neighborhood health center, family planning clinic, mental health clinic or any other medical place? Yes . . . . 01(A) No. . . . . 02(MV)		
A. Who was this? CODE "CLINIC, HEALTH CENTER" IN PERSON'S COLUMN. Anyone else?	14A	Clinic, Health Center. 01
B. How many times since (REF. DATE) did (PERSON) go to one of these places? RECORD IN PERSON'S COLUMN.	B	☐ Times
MV ENTER TOTAL OF EACH PERSON'S VISITS (Q's. 9, 10B, 11B, 12B, 13B AND 14B) IN MV BOX ON CONTROL CARD.	MV	

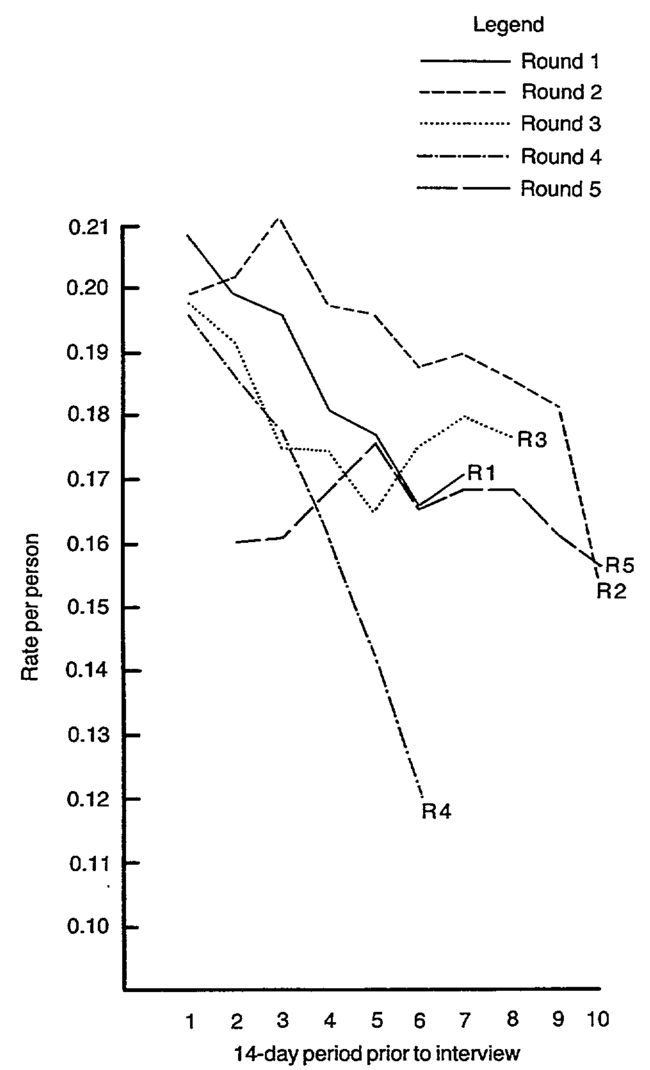
Appendix IV
Rates for Hospital Stays, Medical Visits, and Dental Visits



NOTES: Rates are based on unweighted data. If less than 500 people, the point was not plotted.

Figure I

Rate of hospital stays per person, by period prior to interview and round



NOTE: Rates are based on unweighted data.

Figure II

Rate of medical visits per person, by period prior to interview and round

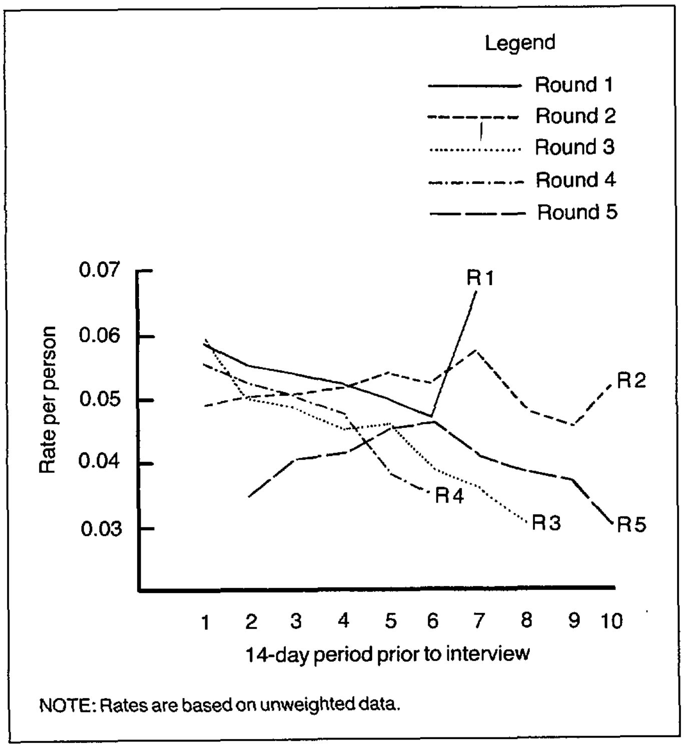


Figure III

Rate of dental visits per person, by period prior to interview and round

Department of Health and Human Services
Otis R. Bowen, M.D., Secretary

Health Care Financing Administration
William L. Roper, M.D., Administrator

Office of Research and Demonstrations
Judith B. Willis, Director

Office of Research
J. Michael Fitzmaurice, Ph.D., Director

Division of Program Studies
Carl Josephson, Director

Survey Studies Branch
Herbert A. Silverman, Ph.D., Chief

Public Health Service
Robert E. Windom, M.D., Assistant
Secretary for Health

National Center for Health Statistics
Manning Feinleib, M.D., Dr.P.H., Director

Office of Vital and Health Statistics Systems
Peter L. Hurley, Associate Director

Division of Health Interview Statistics
Owen T. Thornberry, Jr., Ph.D., Director

Utilization and Expenditure Statistics Branch
Robert A. Wright, Chief