DATA COLLECTION FREQUENCY EFFECT IN THE NATIONAL MEDICAL CARE UTILIZATION AND EXPENDITURE SURVEY

Steven B. Cohen, National Center for Health Services Research and Health Care Technology Assessment; Bruce B. Cohen, National Center for Health Statistics

Introduction

The National Medical Care Utilization and Expenditure Survey (NMCUES) was conducted to meet the needs of government agencies and health professionals for more comprehensive data on the utilization, costs and sources of payment associated with medical care in the United States. A longitudinal survey design was adopted for the household component in order to provide accurate measurements of population characteristics which are sensitive to changes in time. Data collection for the core health care measures was to be applied to the same panel of sample households in five rounds of interviewing, with 1980 as the reference period. Short recall periods of two to three months in duration were generally implemented to minimize reporting errors of omission. A subset of sample participants, referred to as holdovers, were not contacted for a particular wave of the survey and data were gathered at the subsequent round of interviewing for the two time intervals that were spanned. This naturally occurring study treatment provided a unique opportunity to investigate the effect of data collection frequency on reporting behavior.

In this paper, national estimates for a representative set of health care utilization, expenditure and morbidity measures were derived from the sample of holdovers and compared with estimates derived from the respondents with five complete rounds of data collection. The analysis controlled for demographic characteristics that distinguished the two study groups. A more detailed round-specific level of analysis was then conducted to test for a data collection frequency effect, further controlling for length of recall period. The research focuses on the implications of a departure from a panel survey characterized by five waves of data collection.

This study replicates the analyses that were conducted to test for a data collection frequency effect in the National Medical Care Expenditure Survey (NMCES) conducted in 1977 and on which the NMCUES was modeled. The NMCES study findings demonstrate the presence of a data collection frequency effect in a panel design similar to the NMCUES. The study results argue for consideration of four rounds of data collection as an alternative to five, in-survey designs which mirror the NMCES (Cohen and Burt, 1984). The findings identify a survey component that would significantly benefit by a redesign strategy to reduce cost without impairing the quality of survey estimates. Given the similarity in survey design and questionnaire wording between the NMCES and the NMCUES, coincident findings of a data collection frequency effect in the NMCUES would provide support for future survey redesign strategies, which considered a reduction in the number of rounds of data collection.

Background

Long reference periods for survey data collection are typically characterized by two distinct types of reporting errors: errors of omission and erroneous inclusion through forward telescoping. Errors of omission are characterized by the respondent forgetting an illness episode or expenditure or inaccurately recalling the event as happening outside the time period of interest. For health care utilization surveys, these omissions are not random, but are usually concentrated among short term illnesses requiring no hospitalization and routine visits to a physician (Sudman and Lannom, 1979). With respect to forward telescoping, the episode is remembered in error in that the episode is viewed as occurring within the time period of interest when in fact it occurred earlier. Telescoping is more of a potential problem with short recall periods and common events. The more frequent the event, the greater the confusion regarding their occurrence in time. Interviewing techniques which include probing, submission of diaries to the respondent, and computer generated summaries have been considered as mechanisms to reduce reporting errors in panel surveys. The use of these summaries, which describe the responses provided by the respondent in previous interviews of a panel design, allow corrections for omissions and telescoping errors. Both the NMCES and the NMCUES surveys made use of these techniques to minimize reporting errors.

A calendar diary was provided to each household at the end of the first interview. Although it was not a data collection instrument, it served as an aid to respondents to record data, improving their capacity to report health problems, health care use, and related costs. Beginning in the second round of interviews and continuing through the fifth, the household respondent was asked to review a computer generated summary of data on health care services received and costs. This review permitted a check for accuracy and completeness and provided the necessary information to check continuity among the interview rounds for data on health insurance coverage and charges for multiple services. This procedure also allowed respondents to provide additional information about previously reported events as well as to report events that had not been mentioned in previous interviews. In addition, short recall periods, of two to three months in duration, were structured into the survey design to also limit the potential for errors of omission.

The effects of time on memory have received considerable attention in the field of survey research. It has been suggested that short term and intermediate memory decays exponentially with time (Wicklegren, 1970).
A negative exponential model has been proposed as a prediction equation for the proportion of events reported in a survey. The model takes form

$$r_o = ae^{-bt}$$

where b determines the rapidity of decay and depends on the events importance to the respondent, conditions of the interview, and respondent characteristics, a is non-time related and measures the social desirability of an event, t measures the length of recall and

$$r_o$$

is the proportion of reported events (Sudman and Bradburn, 1970). Consequently,

$$1 - r_o$$

measures the relative error due to omissions. Using behavioral data with available record check information, Sudman and Bradburn also note that errors of omission are inversely related of the saliency of the question to the respondent. For health care surveys, events that occur with great frequency are more salient to the respondent.

Other record check studies which have focused on interview procedures similar to those adopted in the Health Interview Survey (NCHS) have clearly demonstrated the accuracy of recall of medical events decreases with time (Cannell and Fowler (1965); Balmuth (1965); Cannell, Fisher and Bakker (1965); and Madow (1967)). In these studies, the recall of a hospitalization, an event with great salience to the respondent, was found to remain accurate for several months. Contrarily, a sharp decrease in recall of events of less salience, such as doctor visits, was found to occur within

several weeks after the occurrence of the event (Yaffee and Shapiro, 1981). In another study which concentrated on a determination of optimal recall periods for estimating accidental injuries in the National Health Interview Survey, it was concluded that large memory biases result when long recall periods are used (Massey and Gonzalez, 1976). It was determined that an optimal recall period is either a 2-week or 4-week period depending on the detail of the analysis. The study provided further evidence that the 2-week reference period used to collect information on acute conditions in the HIS survey is close to optimum.

The consistent finding of incremental memory loss with longer lengths of recall provides support for the consideration of repeated interviewing with short recall periods, particularly when the entire survey reference period is too long to accurately recall in one interview. Longitudinal survey designs are also adopted to provide a mechanism to assess change in the behavior of a population over a specified time period. Often referred to as panel designs, they allow for the measurement of seasonal variations in population characteristics. This capacity is of primary importance in major national health surveys, which attempt to measure the health status and morbidity levels of the population. Since these measures are sensitive to seasonal developments (i.e., climatic changes), a point estimate in time would have serious limitations, in its exposure to the risks of seasonal, secular, and catastrophic variation. Repeated interviews over an entire time period, usually a year in duration, may lead to better statistical inferences than a single one-time survey.

The vast majority of studies which have concentrated on the issue of memory loss and period of recall have considered single-period data. As noted, the consistent finding of a strong positive association with memory loss and length of recall period has guided the decision of repeated interviews with short recall periods in panel survey designs. Fewer studies, however, have focused on the effect of repeated interviews on the stability of this relationship.

One methodological experiment from the California Morbidity Project found a tendency for individuals participating in a periodic-visit survey to report a decreasing amount of illness in response to succeeding waves of interviewing (Feldman, 1960). It was noted that since control groups exhibited higher levels of reported morbidity over the complementary time periods, the decrease in reported illness was more likely a consequence of the repeated interviewing over seasonal or extraneous factors. The effect of repeated interviewing over time is often referred to as a conditioning effect. In a study on food purchase behavior, members of a panel design indicated a belief that continued panel membership affected their behavior (Quackenbush and Schaffer, 1960). Repeated interviewing was observed to be associated with a decrease in reported food expenditures in another study of food purchase behavior. (Prais, 1958). In addition, multi-round demographic surveys in developing countries have often been characterized by respondent fatigue (Adlakha, 1980).

A direct test for conditioning effects from repeated interviewing was incorporated in a classic experimental study of response errors in the collection of expenditures data by household interviews. The experiment was a component of the Census Bureau's Survey of Residential Alterations and Repairs. A comparison of data obtained from second and third round household intervals with a one month recall period for the same time period, which were otherwise alike, indicated the presence of a conditioning effect. (Neter and Waksburg, 1965). An aggregate decrease of 9 percent in the reported number of jobs related to residential alterations or repairs was associated with the addition of a third round of interviewing in the panel design. Further, the size of the conditioning effect was inversely related to the size of the job, which was measured in terms of expenditures. In addition, the effect on reported total expenditures was substantially less than on total number of jobs.

The effect of data collection frequency and length of recall on the reporting of medical care utilization, expenditures and source of payment was examined in the Medical Economics Survey - Methods Study. The study was designed to test the feasibility and effectiveness of alternative survey strategies for obtaining accurate health care data in a panel design. The results were used in planning purposes for the National Medical Care Expenditure Survey. Eligible households were randomly divided into four experimental groups which differed on follow-up data collection techniques over a six-month period. Periodicity in terms of monthly versus bi-monthly interviews and type of contact (in person versus telephone) were the two factors that were varied. Using record check data from medical providers and/or third party payers, the reporting accuracy obtained from the alternative survey strategies were examined. No obvious difference in reporting accuracy by data collection frequency was evident. The use of memory aids which included a calendar-diary and a summary of events previously reported, may have reduced the potential for differences due to periodicity. The data collection strategy that considered bi-monthly interviews resulted in substantial cost savings, approximately 30 percent of data collection and processing costs in this study. It was noted, however, that the monthly interviews experienced a significantly higher attrition rate, indicating less desire to remain in the survey with an increased number of interviews (Yaffee and Shapiro, 1981).

The effect of data collection frequency on the reporting of health care utilization and expenditures was also examined in the National Medical Care Expenditure Survey (Cohen and Burt, 1984). The NMCES was a panel survey with five scheduled rounds of data collection to cover calendar year 1977. The survey was conducted by the National Center for Health Services Research and co-sponsored by the National Center for Health Statistics to provide detailed information on the utilization, costs and sources of payment associated with medical care in the United States. In the NMCES, a subsample of participants, referred to as holdovers, were not contacted during a particular wave of the survey and data were gathered at the subsequent round of interviewing for the two time intervals that were spanned.

The analysis focused on a comparison of health care estimates obtained from individuals with five rounds of data collection with those derived from survey participants designated as holdovers for the fourth round of data collection. Controlling for differentials in demographic profiles which characterized the study groups, annual health care utilization, expenditure and morbidity estimates were generally higher for individuals with only four rounds of data collection. Round specific comparisons in the reporting of health care events relative to annual profiles, controlling for length of recall, indicated significantly higher estimates for the holdovers with respect to outpatient physician contacts, related expenditures and dental visits. Further, a comparison to determine the level of reporting concordance between household and medical provider record check data indicated equivalent or superior performance for the individuals with only four rounds of data collection.

The NMCES comparisons between data sources demonstrated equivalent or superior performance in the reporting of health care events by the individuals with four rounds of data collection. The findings argue for consideration of a data collection scheme which follows the schedule of the round four holdovers. The additional interview for the nonholdovers, which was generally characterized by a smaller length of recall period than the complementary fourth interview in round five for the holdovers, appears to have induced a respondent burden. Although shorter lengths of recall are traditionally associated with reductions in reporting errors of omission, the introduction of additional interviews in a panel survey to reduce

periods of recall alters the relationship, and on occasion, may increment errors of omission for the later survey interviews.

NMCUES Sample Design and Interview Structure

The design of the NMCUES is complex; it is best characterized as a stratified multistage probability design from two independently drawn national samples of the Research Triangle Institute (RTI) and the National Opinion Research Center (NORC). The structures of both national sample designs were similar and therefore compatible. Sampling specifications called for the selection of approximately 7,200 households. The survey was co-sponsored by the National Center for Health Statistics (NCHS) and the Health Care Financing Administration (HCFA).

Sampling units in the first three stages of each replicate sample are land areas ranging in size from small groups of contiguous counties in the first stage to small area segments consisting of several dozen housing units. The first stage in both designs consists of primary sampling units (PSUs), which are parts of counties, or groups of contiguous counties. These units were stratified by geographic location, degree of urbanization, and size for RTI, and also by percentage black for NORC. The second stage consists of secondary sampling units (SSUs), which are generally census block groups or enumeration districts in both designs. Smaller area segments constituted the third stage in both designs from each of which a subsample of households was selected in the final stage of sampling. Selection in each of the first three stages was with probabilities proportional to certain size measurements. Combined stage-specific sample size over the two designs was 135 PSUs (covering 108 separate localities), 809 SSUs, and 809 segments. Ultimate sampling units consisted of residential housing units defined as a house, a group of rooms, or a single room occupied as separate living quarters.

Data collection was applied to the same panel of households in five rounds of interviews during 1980 and early 1981. The first interviews began in late January 1980; subsequent rounds of interviews were conducted at intervals of about three months. The first, second, and fifth rounds of interviews were conducted in person, as were about 20 percent of the third and fourth rounds and about half of the sixth round; the remainder were conducted by telephone. Data were obtained for 90 percent of eligible households in the first interview and approximately 95 percent of the individuals in the participating households supplying information for the entire year.

During each of the first five rounds of interviews, information was obtained on use of medical services, charges for services and sources of payment, numbers and types of disability days, and status of health insurance coverage. Data collected during the first interview covered the period from January 1, 1980, through the date of interview. Data collected during the second, third, and fourth rounds covered the period from the immediately preceding interview through the data of the current interview. The fifth interview covered the period from the previous interview through December 31, 1980.

Of 17,123 sample participants in the NMCUES, 16,207, or 94.65 percent, responded to the survey for the entire year (1980). A distribution of the NMCUES participants in terms of survey response status is presented in Table 1. Sample participants who provided data for their entire period of eligibility included individuals who died during the survey year, entered an institution, and newborns. To eliminate the potential effects of differential periods of eligibility and partial response when testing for a data collection frequency effect, only those sample participants responding to the survey for the entire year were considered in subsequent analyses.

For the 16,207 survey participants who responded for the entire survey period, 11,070, or 68.3 percent experienced five rounds of data collection. The remainder were referred to as holdovers, since they were skipped during one or more of the scheduled rounds of interviewing. When contact was re-established in a subsequent round of data collection, the respondents were required to provide information on their health care experience for the entire period between interviews. Prior to the round four interview, there was self-selection in the determination of holdover status for a particular round of data collection. Often, the respondent was away from home at the scheduled time of the interview, infirmed, or difficult to contact. As a consequence of self-selection classification, this group of holdovers were excluded from subsequent analyses which concentrate on the detection of a data collection frequency effect. In the NMCUES, 5,041, or 31.1 percent of the participants who responded for the entire year, only missed the fourth round of data collection (Table 2). These cases were scheduled for interviews early in round five, which started in January 1981.

A Comparison of Demographic and Health Care Measures Between Respondents with Four Versus Five Rounds of Data Collection

Prior to testing for a data collection frequency effect, it was necessary to determine whether individuals with four rounds of data collection exhibited any systematic difference in demographic profiles from their sample counterparts with five rounds of data collection. Those differences that were identified would have to be controlled for in the comparisons of reported health care experiences, to factor out their potential effect on observed differentials in health care estimates.

The demographic variables under investigation included region, size of city, age, race, ethnicity, sex, perceived health status, poverty status, marital status, medicare coverage, medicaid coverage, and private health insurance coverage. Estimates of the national distributions for these demographic measures were derived for the two samples which differed by data collection frequency, and can be observed in Table 3. Large sample two-sided z tests were conducted to determine whether significant differences existed in the demographic configurations of the two respondent groups. All tests considered an α level of .05. Variances of all parameter estimates considered in this paper were derived using the Taylor series linearization method to appropriately account for the effects of clustering and stratification induced by a complex sample design (Shah, 1981).

Overall, no significant differences in the demographic distributions were evident across data collection frequency classification for region, size of city, sex, health status, poverty status and private insurance coverage. The comparison of age distributions for the two groups revealed a significantly greater representation of individuals 25 years of age or older for the respondents with five rounds of data collection, and a greater representation of individuals aged 18 and under for the round four holdovers. There was also a significantly greater representation of individuals covered by Medicare as of December 31, 1980 for the respondents with five rounds of data collection, which was consistent with the differential in age distributions between the two groups. With respect to marital status, individuals with five rounds of data collection were more likely to be married or widowed than their holdover counterparts, who had a greater probability of classification under 17 years of age. In addition, the holdovers had a significantly lower representation of whites but a greater representation of individuals ever having Medicaid coverage.

To provide for a comprehensive investigation, the comparisons for health care measures consisted of a representative set of survey statistics which estimated medical care utilization, expenditures and morbidity. The utilization measures included the

number of physician visits, hospital discharges, dental visits and the number of prescribed medicines. More specifically, physician visits consisted of all medical visits during which a medical doctor was seen. Hospital discharges included the total number of hospital stays for which the hospital was classified as a short stay facility and the discharge date was during 1980. Dental visits included all visits to a dentist, dental surgeon, oral surgeon, orthodontist, other dental specialist, dental hygienist, dental technician or any other person for dental care. Prescribed medicines included any drug or other medical preparation prescribed by a physician, including refills. Expenditure data for selected utilization measures were also considered: total charges for physician visits, and total charges for all hospital stays, with charges included for separately billed doctor charges for visits occurring during these hospital stays. The number of restricted activity days served as the measure of morbidity, which included the number of days illness or injury kept a person away from job or other work, or usual activity (e.g., work around the house, school). This morbidity measure was derived by subtracting the work loss days in bed from the sum of the number of bed days, work loss days, and cutdown days.

A comparison of the mean number of physician visits for 1980 by data collection frequency indicated a significantly higher annual utilization estimate for individuals with only four rounds of data collection. (Table 4A.) Large sample two sided z-tests were conducted to determine whether significant differences existed in the respective health care estimates at the .05 level of significance. When controlling for those demographic characteristics that distinguished the two groups, the same pattern was evident. Comparisons by age revealed the utilization differentials were most prominent for members of the older age categories. Comparisons across classes of race, marital status, medicare and medicaid coverage revealed the same trend. Whenever statistically significant differentials in utilization estimates were detected, individuals with four rounds of data collection had a higher annual mean number of physician visits.

Comparisons of the mean number of hospital discharges (Tables 4B) also indicated a significantly higher annual utilization estimate for individuals with four rounds of data collection. Again, controlling for those demographic characteristics that distinguished the two groups, the same pattern was evident. Whenever statistically significant differentials were detected, individuals with four rounds of data collection were characterized by a higher average utilization measure.

The overall comparisons in mean number of dental visits and prescribed medicines by frequency of data collection did not reveal a significant difference in utilization estimates (Tables 4C-D). However, when controlling for the demographic differentials between the groups, statistically significant differences in prescribed medicine utilization that were detected were predominantly in the same direction, with higher estimates characterizing the round four holdovers.

The comparisons of the mean annual total charges for physician visits and hospital discharges, by data collection frequency, were generally consistent with the findings for the respective utilization measures (Tables 4E-F). With respect to the measure of morbidity, a comparison of the mean number of restricted activity days indicated a significantly higher level for the round four holdovers (Table 4G). The more refined comparisons across demographic classes revealed the same directional differential for those differences in estimates that were statistically significant at the .05 level.

Individuals with five rounds of data collection had a greater representation of the aged, a group that is typically characterized by higher utilization and medical care expenditure patterns. Consequently, the observation of significantly higher overall utilization estimates for physician visits and hospital discharges for the round four holdovers, suggested the presence of a data collection frequency effect. This hypothesis was further supported by the significantly higher mean number of restricted activity days observed for the round four holdovers. Given the statistically equivalent perceived health status distributions characterizing the respective study groups, the significant differences observed for this measure of morbidity were most notable.

Round Specific Comparisons in Health Care Estimates by Length of Recall Period

Although the comparisons of the annual health care estimates are suggestive of a data collection frequency effect, a number of other factors potentially associated with the reporting differentials had to be controlled for, prior to a final determination. The differences in health care estimates that were detected may have been in effect prior to round four, the round of data collection when the study "treatment" of assignment of holdover status went into effect. In addition, differences in the round specific length of recall periods for individuals which comprise the two study groups may have influenced the results. Consequently, a more detailed level of analysis was conducted, which compared round specific estimates of health care measures, controlling for length of recall period.

A comparison of the round specific length of recall period for individuals characterized by four or five rounds of data collection can be observed in Table 5. The mean length of recall period, measured in days, was consistently higher for the round four holdovers over all comparable rounds of data collection. The mean difference was minimal for round one at 5.4 days, systematically increasing to 7.7 days for round two, 2 weeks for round three and over 5 weeks for round five. Further, the round five mean reference period from the previous interview to the end of 1980 was approximately 6 weeks greater for the round four holdovers. Overall, the most dramatic differentials in mean length of recall period occurred after the third round of data collection.

As indicated in the literature, errors of omission are generally associated with longer length of recall periods. Telescoping errors are most evident for short recall periods, and in the NMCES, bounding techniques with repeated interviews and the use of computer generated summaries should have minimized their occurrence. Having established that individuals with four rounds of data collection were characterized by longer length of recall periods, the observations of statistically higher annual health care utilization, expenditure and morbidity estimates for this group was particularly striking.

Controlling for length of recall, round specific health care utilization and expenditure estimates were also compared across the study groups distinguished by data collection frequency. Since it was determined that the two groups often differed in annual health care estimates, the round specific comparisons focused on the detection of relative reporting differences. To facilitate the comparisons, the length of recall period was categorized into seven mutually exclusive classes: 1-30 days, 31-60 days, 61-90 days, 91-120 days, 121-150 days, 151-180 days, and over 180 days. To further control for differential length of recall periods, the round specific health care experience for each individual was annualized. Round specific congruency ratios were then constructed, dividing the mean of the annualized values (based on round specific data) by the overall mean based on the unadjusted annual data. This measure was adopted for the comparisons as a method of standardization. The congruency ratios took the form:

$$CR\ (grm) = \frac{\bar{Y}_{grm}}{\hat{\bar{Y}}_{g}}$$

where g = 1, 2, and (1) denotes four rounds of data collection; and (2) denotes five rounds of data collection. r = 1, 2, 3, 4, 5 identifies the round of data collection; m = 1, 2, 3, ----------7 identifies the round specific length of recall period;

$$\hat{\bar{Y}}_g$$

is the overall mean estimate of the unadjusted annual data for individuals in study group g; and

$$\hat{\bar{Y}}_{grm} = \frac{\sum_{i \varepsilon grm} W_i \, 366 \cdot \frac{Y_{ri}}{d_{ri}}}{\sum_{i \varepsilon grm} W_i}$$

is the annualized estimate for individuals in study group g for round r and length of recall period m, where

$$Y_{ri}$$

is the round specific data for individual i in study group g and length of recall period m,

$$d_{ri}$$

is the number of days in 1980 that characterize the round r recall period for individual i in study group g and length of recall class m, and

$$W_i$$

is the ith individual's sampling weight.

To illustrate this process, consider round five (r = 5) data on physician visits for individuals with only four rounds of data collection (g = 1) and a length of recall period of 91-120 days (m = 4). Each individual i in group g = 1, r = 5, m = 4, has their round specific data

$$Y_{5i}$$

annualized by dividing

$$Y_{5i}$$

by the number of days in 1980

$$(d_{ri})$$

that constitute the recall period (m = 4) to get a rate per day, and multiplying the result by 366. A weighted mean estimate of the annualized data:

$$\hat{\bar{Y}}_{154}$$

is then derived for this subset of respondents. A congruency ratio is produced by dividing the mean of the annualized values based on round specific data by the overall weighted mean for the respondents:

$$\hat{\bar{Y}}_1$$

based on the annual data obtained over all rounds of data collection. If the ratio is larger than unity, then the annualized round specific estimate, controlling for length of recall, is greater than the overall mean based on the reported annual data. Contrarily, when the ratio is less than unity, the overall mean based on the reported annual data is larger.

A comparison of the congruency ratios for data on physician visits revealed no significant reporting differentials for the first three rounds of data collection, after controlling for length of recall (Table 6A). The comparisons of the ratio of the annualized round five estimate to the overall unadjusted annual estimate, were consistent with observation of no significant reporting differentials. For the study group characterized by five rounds of data collection, however, the congruency ratios consistently were less than unity. A similar pattern in the cross-group comparisons was observed for the physician visit medical expenditure data (Table 6E). Examination of the congruency ratios for utilization data on dental visits did not reveal significant differences across the round five estimates (Table 6C). As before, the ratios for individuals with five rounds of data collection were consistently less than unity.

For hospital discharges, the round specific comparisons of congruency ratios also failed to detect significantly different relative utilization and expenditure estimates across study groups. Again, no significant differences in round five congruency ratios, were noted. For the utilization data on prescribed medicines (Table 6D), however, a significant difference was detected in the round five congruency ratios. A higher relative level of reported prescribed medicine utilization was observed for individuals with four rounds of data collection. Large sample two sided z-tests were conducted to determine whether significant differences existed across congruency ratios at the .05 level of significance. Precision requirements restricted comparisons to those classes with a minimum sample size of 100 and a relative standard error of less than 30 percent.

The round specific comparisons in health care estimates, by length of recall period, provided a more sensitive level of analysis in the detection of reporting differentials by data collection frequency classification. Although the annual utilization and expenditure estimates for outpatient physician visits and hospital discharges differed by data collection frequency classification, the round specific comparisons of congruency ratios indicated that when length of recall was controlled for, no evidence of differential reporting for round five was present.

A comparison in health care estimates across data collection periods which constitute the fourth interview of the survey for the respective study groups was also considered. The fourth interview occurred in round four of data collection for the individuals with five rounds of data collection, and in round five for the holdovers. Furthermore, the round five reporting period for the holdovers overlapped with the time period spanned by the round four interview for the nonholdovers. Controlling for length of recall period, the congruency ratios characterizing the fourth interview were statistically equivalent in all but one comparison (Tables 6A-F). A higher relative utilization estimate for dental visits was detected, characterizing the fourth interview for the nonholders. It is important to note that the fourth interview for individuals with five rounds of data collection was usually conducted by telephone.

The inability to detect a statistically significant data collection frequency effect is partially due to the relatively smaller sample size of the NMCUES (17,123) when compared with the NMCES Survey (38,815 individuals). Further, the NMCUES study findings do not indicate that a fifth round of data collection with a shorter length of recall than the complementary fourth interview in round five for the holdovers was associated with a significant differential in reporting behavior. This was noted by comparing Round 5 congruency ratios with longer lengths of recall (91-120, 121-150 days) for the individuals experiencing a fourth interview, to those reflecting a short length of recall (61-90 days) for individuals with a fifth interview. The comparison revealed no significant reporting differentials for the selected utilization and expenditure measures under investigation. Shorter lengths of recall are generally associated with reductions in reporting errors of omission. Although the fifth NMCUES interview was most often characterized by a smaller length of recall period than the complementary fourth interview in round five for the holdovers, the effect of length of recall on reporting behavior was not operational for this later round of data collection. Consequently, the study results provide additional support for the consideration of four rounds of data collection as an alternative to five,

in a panel survey similar in scope to the NMCES and NMCUES.

Summary

In the NMCUES, the departure from five rounds of data collection in a panel survey allowed for an investigation of the effect of data collection frequency on the reporting of health care related events. It was determined that the sample with five rounds of data collection were more likely to be individuals 25 years of age or older, white, medicare recipients, and married or widowed, than their holdover counterparts. Controlling for these demographic differentials, annual health care utilization, expenditure and morbidity estimates were generally higher for the individuals with only four rounds of data collection. However, round specific comparisons in the reporting of health care events relative to the annual profiles, by length of recall, indicated no significant differential in round five congruency ratios across study groups. Consequently, no data collection frequency effect was observed to operational in the NMCUES.

The NMCUES study results do not indicate that a fifth round of data collection with a shorter mean length of recall than the complementary fourth interview in round five for the holdovers was associated with a significant differential in reporting behavior. The findings argue for the consideration of four rounds of data collection as an alternative to five, in a panel survey similar in scope to the National Medical Care Utilization and Expenditure Survey. They identify a survey component which could significantly benefit by a redesign strategy to reduce cost without impairing the quality of survey estimates. This is primarily achieved by a reduction in interviewer costs. Additional savings are to be achieved from reduced data processing costs, which included the generation of round-specific summaries to serve as memory aids. The study, however, does not identify the optimal balance between data collection frequency and round specific length of recall in terms of minimizing reporting errors. Further research in this area is essential, to identify strategies which improve upon the accuracy of data obtained in panel designs.

References

Adlakha, A.L., J.M. Sullivan, and J.R. Abernathy (1980) "Recent Trends in the Methodology of Demographic Surveys in Developing Countries." Proceedings of the American Statistical Association, Section on Survey Research Methods, 60-65.

Balmuth, E. (1965). Health Interview Responses Compared with Medical Records. National Center for Health Statistics: Vital and Health Statistics Series 2, No. 7. Washington: U.S. Government Printing Office.

Bonham, G.S. and L.S. Corder (1981). National Care Expenditure Survey: Household Interview Instruments. National Center for Health Services Research, Instruments and Procedures Series No. 1. DHHS Publication No. (PHS) 80-3280. Hyattsville, Md.

Cannell, C. and F. Fowler (1965). Comparison of Hospitalization Reporting in Three Survey Procedures. National Center for Health Statistics: Vital and Health Statistics, Series 2, No. 8. Washington: U.S. Government Printing Office.

Cannell, C., G. Fisher and T. Bakker (1965). Reporting of Hospitalization in the Health Interview Survey. National Center for Health Statistics: Vital and Health Statistics. Washington: U.S. Government Printing Office.

Cohen, S.B. and W.D. Kalsbeek (1981). National Medical Care Expenditure Survey: Estimation and Sampling Variances in the Household Survey, Instruments and Procedures Series, No. 2, DHHS Publication No. (PHS) 81-3281. Hyattsville, Md.

Cohen, S.B. and Burt, V.L. (1984) "Data Collection Frequency Effect in the National Medical Care Expenditure Survey." Proceedings of the American Statistical Association, Section on Survey Research Methods, in press.

Feldman, J.J. (1960) "The Household Interview Survey as a Technique for the Collection of Morbidity Data." Journal of Chronic Diseases. Vol II, No. 5, 535-557.

Holt, D., T.M. Smith, and P.D. Winter (1980). "Regression Analysis of Data from Complex Surveys" Journal of the Royal Statistical Society. 143(4), 474-487.

Madow, W.G. (1967). Interview Data on Chronic Conditions Compared with Information Derived from Medical Records. National Center for Health Statistics: Vital and Health Statistics, Series 2, No. 23. U.S. Government Printing Office, Washington, D.C.

Massey, J.T. and J.F. Gonzalez, Jr. (1976). Optimum Recall Periods For Estimating Accidental Injuries in the National Health Interview Survey. Proceedings of the American Statistical Association, Social Statistics Section, 584-588.

Neter, J. and J. Waksberg (1965). Response Errors in Collection of Expenditures Data by Household Interviews: An Experimental Study. U.S. Bureau of the Census. Technical Paper No. 11. Washington: U.S. Government Printing Office.

Newhouse, J.P. and C.E. Phelps (1976). New Estimates of Price and Income Elasticities of Medical Care Services, in The Role of Health Insurance in the Health Services Sector, R.N. Rossett, editor. New York: National Bureau of Economic Research, 261-313.

Prais, S.J. (1958). "Some Problems in the Measurement of Price Changes with Special Reference to the Cost of Living." Journal of the Royal Statistical Society, Series A, Part 3, 312-332.

Shah, B.V. (1981). "SESUDAAN: Standard Errors Program for Computing of Standardized Rates from Sample Survey Data." Research Triangle Institute, Research Triangle Park, North Carolina.

Sudman, S. and N.M. Bradburn. (1973). Effects of Time and Memory Factors on Response in Surveys. Journal of the American Statistical Association, 68, 805-815.

Sudman, S. and L.B. Lannon. (1979). A Comparison of Alternative Panel Procedures for Obtaining Health Data. Paper available from the Survey Research Laboratory, Urbana: University of Illinois.

Wicklegren, W.A. (1970). "Multitrace Strengths Theory" in Models of Human Memory, D. A. Norman, editor, New York: Academic Press, 65-102.

Yaffe, R. and S. Shapiro. (1981). Medical Economics Survey - Methods Study: Cost-Effectiveness of Alternative Survey Strategies. NCHSR Research Proceedings Series: Health Survey Research Methods Third Biennial Conference. Department of Health and Human Services Publication No. (PHS) 81-3268. Washington: Government Printing Office.

Notes

1. The views expressed in this paper are those of the authors and no official endorsement by the Department of Health and Human Services, The National Center for Health Services Research, or the National Center for Health Statistics is intended or should be inferred.

2. Tables 1 - 6F were not presented in this paper due to space limitations. They may be obtained from the author by writing to: Dr. Steven B. Cohen, National Center for Health Services Research and Health Care Technology Assessment, Room 3-50 Park Building, 5600 Fishers Lane, Rockville, Maryland 20857.

Acknowledgments

The authors wish to thank Bill Ashby, Jo Crosby, and Melissa Rowe for their skillful typing of this manuscript.