# DATA COLLECTION FREQUENCY EFFECT IN THE NATIONAL MEDICAL CARE EXPENDITURE SURVEY

Steven B. Cohen and Vicki L. Burt, NCHSR

## Introduction

The National Medical Care Expenditure Survey (NMCES) 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 1977 as the reference period. Short recall periods of two to three months in duration were generally implemented to minimize reporting errors of omission. Field conditions, however, did not allow for all interviews to be conducted over the targeted rounds of data collection. 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 and other related predispositional factors. In addition, an analysis was conducted to determine the level of reporting agreement between household and record check data for the two study groups. The research focuses on the implications of a departure from a panel survey characterized by five waves of data collection, with respect to reporting behavior and potential cost savings.

## NMCES Sample Design and Interview Structure

The design of the NMCES 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 14,000 households. The survey was conducted to provide data for a major research effort in the Division of Intramural Research of the National Center for Health Services Research (NCHSR), and was cosponsored with the National Center for Health Statistics (NCHS).

Data collection was applied to the same panel of households in six rounds of interviews during 1977 and early 1978. The first interviews began in mid-January 1977; 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.

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, 1977, through the date of interview. Data collected during the second, third, and fourth rounds covered the period from the immediately preceding interview through the date of the current interview. The fifth interview covered the period from the previous interview through December 31, 1977. (Bonham and Corder, 1981)

A sixth round of interviews were scheduled to obtain responses to a series of supplemental questions covering limitations of activities, status of income tax filing, and the amount of itemized medical deductions. In addition, survey participants were requested to sign a permission form, to allow each physician or facility reported as providing medical care during 1977, to release information about the patient. Consequently, the utilization, expenditure and health insurance coverage data that characterized the survey respondents experience in 1977 was obtained from the initial five rounds of interviewing.

## Characteristics of NMCES Participants Classified by Data Collection Frequency

Of 38,815 sample participants in the NMCES, 14004, or 36.1 percent, did not experience five rounds of data collection for the survey year 1977. As noted, these individuals 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. A distribution of the NMCES participants in terms of data collection frequency is presented in Table 1.

Prior to the round four interview, 3,833 sample participants or 9.9 percent, had missed a scheduled round of interviewing. For these survey respondents, 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 this self-selection classification, this group of holdovers were excluded from subsequent analyses which concentrate on the detection of a data collection frequency effect.

The remainder of the survey participants designated as holdovers, 10,171 or 72.6 percent of the holdovers, only missed the fourth round of data collection. The allocation of round four holdover status was not due to self-selection and primarily a consequence of delays in field procedures. Since the round four interview period extended into December, and another scheduled round of data collection would obtain information for the remaining time interval ending on December 31, 1977, a decision to hold over approximately 29 percent of the NMCES sample was made. These cases were scheduled for interviews early in round five, which started in January, 1978.

In the NMCES, 4,146 or 10.7 percent of all responding survey participants provided data for only part of the time they were eligible to respond. For example, a person could have refused participation after initially cooperating in the first interview by not responding for the remainder of the interviews. Similarly, the inability to re-establish contact with a participant after change of residence would result in this type of nonresponse. Alternative imputation strategies for individuals with partial data were formulated which included a weighted adjustment for partial response, and the use of data exclusively from participants with complete information to characterize the nation (Cohen, 1982). Since the weighted adjustment strategy to partial data was developed for annual estimates of health care utilization and expenditure measures, its consideration in this study was limited. Comparisons of round specific reported health care experiences necessitated consideration of the other adjustment strategy for partial nonresponse. In this study, the 4146 individuals with the partial response profiles were viewed as total nonrespondents and a standard weighted nonresponse adjustment factor was used together with the original NMCES sample weights, which implicitly attributed the characteristics of similar respondents within the same age-race-sex weighting classes to nonrespondents. Consequently, the adjusted sampling weights were directly applied to the data of complete respondents in the derivation of parameter estimates. The distribution of the NMCES participants with complete response profiles, in terms of data collection frequency, is also 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. Consequently, not all individuals with complete response profiles provided data for 365 days. To eliminate the potential effect of differential periods of eligibility when testing for a data collection frequency effect, only those sample participants providing data for the entire survey year were considered in subsequent analyses.

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

Since the allocation of round four holdover status was not due to self-selection, 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, sex, health status, poverty status, marital status, medicare coverage, medicaid coverage, private health insurance coverage, and health insurance coverage status. 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 2. 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 $\alpha$ 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 (Woodruff, 1971).

Overall, no significant differences in the demographic distributions were evident across data collection frequency classification for region, size of city, race, health status, poverty status and private insurance coverage, and health insurance coverage status. The comparison of age distributions for the two groups revealed a significantly greater representation of individuals 55 years of age or older for the non holdovers, and a greater representation of individuals aged 6-24 for the round four holdovers. There was also a significantly greater representation of individuals covered by Medicare for the nonholders, which was consistent with the differential in age distributions between the two groups. With respect to marital status, nonholdovers were more liklely to be married or widowed than their holdover counterparts, who had a greater probability of classification under 17 years of age. Similarly, the holdovers had a significantly greater representation of men.

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 admissions, dental visits and the number of prescribed medicines. More specifically, physician visits consisted of all outpatient physician contacts including telephone calls. Hospital admissions included admissions of less than 24 hours and those for women giving birth. Newborns were not counted as separate admissions unless they were admitted separately following delivery. 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: physician visit expenditures, and total hospital expenditures for all hospital admissions, with charges excluded for inpatient physician services. Disability days served as the measure of morbidity, which included the number of days illness or injury kept a person in bed, away from job or other work, or usual activity (e.g., work around the house, school).

A comparison of the mean number of outpatient physician contacts for 1977 by data collection frequency indicated a significantly higher annual utilization estimate for individuals with only four rounds of data collection. (Table 3A.) 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 distinquished 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. Although women had a significantly higher utilization experience than men, the pattern of a higher number of physician contacts for round 4 holdovers held across sex. Comparisons across classes of 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 outpatient physician contacts.

Comparisons of the mean number of hospitalizations and the mean number of dental visits (Tables 3B-C) 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 comparison in mean number of prescribed medicines by frequency of data collection did not reveal a significant difference in utilization estimates. (Table 3D) However, when controlling for the demographic differentials between the groups, all statistically significant differences in utilization that were detected were in the same direction, with higher estimates characterizing the round four holdovers.

The comparisons of the mean annual medical expenditures for physician contacts and hospitalizations, by data collection frequency, were generally consistent with the findings for the respective utilization measures (Tables 3 E-F). With respect to the measure of morbidity, a comparison of the mean number of disability days indicated a significantly higher level for the round four holdovers (Table 3G). 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 and women, groups that are typically characterized by higher utilization and medical care expenditure patterns. Consequently, the observation of significantly higher overall utilization estimates for physician contacts, hospitalizations and prescribed medicines 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 disability days observed for the round four holdovers. Given the statistically equivalent health status distributions characterizing the respective study groups, the significant differences observed for this measure of morbidity was 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 comparision of the round specific length of recall period for individuals characterized by four or five rounds of data collection can be observed in Table 4. 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 2.4 days, systematically increasing to 6.3 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 1977 was approximately two months 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.

A distribution of the round specific length of recall periods for individuals further classified by data collection frequency is presented in Table 5. Both study groups, had the same modal recall period for the first three rounds of data collection. The individuals with only four rounds of data collection, however, had a consistently higher percent of

individuals with the longer length of recall periods. In addition, these individuals had a higher model recall period for the fifth round data collection, with 43.2 percent experiencing a length of recall period in excess of 150 days, compared to 0.2 percent for their nonholdover counterparts.

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: (Cohen, Erickson, and Powell (1983)).

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

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;
l = 1, 2, 3, ----------7 identifies the round specific length of recall period;
$\bar{Y}_g$ is the overall mean estimate of the unadjusted annual data for individuals in study group g;

and $\hat{Y}_{grl} = \dfrac{\sum_{iegrl} W_i \, 365 \, \dfrac{Y_{ri}}{d_{ri}}}{\sum_{iegrl} W_i}$ is the annualized estimate for individuals in study group g for round r and length of recall period l,
where $Y_{ri}$ is the round specific data for individual i in study group g and length of recall class l,
$d_{rl}$ is the number of days in 1977 that characterize the recall period for round r and
$W_i$ is the ith individual's sampling weight.

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, however, indicated a significantly higher relative level of reported utilization for individuals characterized by four rounds of data collection. For both study groups, however, the congruency ratios were less than unity. The same pattern was observed for the physician visit medical expenditure data (Table 6E). Examination of the congruency ratios for utilization data on dental visits revealed significant differences across the round five estimates (Table 6C). As before, the ratios for individuals with four rounds of data collection were higher, though less than unity. A higher congruency ratio for holdovers was also observed in the comparisons of round one data, experienced by individuals with a length of recall period between 31 to 60 days.

For hospitalizations, the round specific comparisons of congruency ratios also detected significantly higher relative utilization and expenditure estimates for the round four holdovers (Tables 6B, 6F). These differences represented the round three health care experience. No significant differences in round five congruency ratios, however, were noted for the utilization and expenditure data on hospitalizations. Similarly, no significant differences were detected in the round five congruency ratios for the utilization data on prescribed medicines (Table 6D). As before, 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.

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. Since no significant difference in the overall annual estimates of prescribed medicine use was detected, the observation of no differences in congruency ratios for round five was not surprising. Although the annual utilization and expenditure estimates for hospitalizations 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. Contrarily, reporting differences by data collection frequency classification remained evident in round five for physician visit utilization and expenditure data and dental visit utilization data, after controlling for length of recall. Consequently, this level of analysis demonstrates that the presence of a data collection frequency effect was selective in nature.

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 hospitalizations 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 congruence in health care estimates representing the fourth interview for both study groups, supports the notion of a data collection frequency effect. The lower congruency ratios that were observed in round five for the nonholdovers were derived from data collected in a fifth interview for this study group, compared to a fourth interview for the holdovers.

Estimated Data Collection Frequency Effect

A direct test for data collection frequency effect was conducted for those health care measures characterized by round specific congruency ratios that significantly differed for round five. The analysis concentrated on the measures of physician visit utilization and expenditures, and dental visit utilization. A regression strategy was implemented to estimate and test for a data collection frequency effect. Within this framework, the dependent variable was defined as the annualized round five utilization or expenditure estimate for each sample participant. To isolate the hypothesized data collection frequency effect, the model specification in this analysis required the inclusion of predispositional factors associated with reporting differentials. Since the dependent variables of interest were measures of health care demand, the model incorporated a set of explanatory variables consistent with the demand equation specifications in the health economics literature (Newhouse and Phelps, 1976).

The following explanatory variables were included in the model in addition to an indicator for data collection frequency (four rounds, five rounds): length of recall (in days), age, sex, race, education of household head (<12 years, 12 years, >12 years of education), health status (poor, fair, good, excellent), size of city (SMSA, non-SMSA), employment status (employed, unemployed, not in labor force, <16 years of age), region, health insurance coverage status (always insured, sometimes insured, never insured), Medicare coverage (ever covered, never covered), Medicaid coverage (ever covered, never covered), family income, physicians per 100,000 population in county (1975), and hospital beds per 100,000 population in county (1975). The model also included an annualized estimate of the respective health care measures under investigation based on the individual's reported behavior for the first three rounds of data collection, as a method of standardization. Controlling for these explanatory measures, a test for data collection frequency in health care reporting was conducted. The regression analysis considered a weighted least squares methodology appropriate for complex survey data, where variances of estimated model parameters were derived using the Taylor series linearization method (Holt, Smith and Winter, 1980).

A significant data collection frequency effect was observed for the health care measures of physician visit utilization and expenditures. Even when controlling for the reported health care experience spanning the first three rounds of data collection length of recall, and other related predispositional measures, an individual with only four rounds of data collection was associated with a higher annualized round five utilization and expenditure profile. (Table 7). The estimated data collection organization effect was .832 for the round five utilization data and 20.014 for the related expenditure data. A similar relationship was observed for the dental visit health care utilization measure. Again, the individuals with four rounds of data collection were characterized by a significantly higher annualized round five reported utilization pattern. The estimated data collection frequency effect for the dental visit data was .423.

The same regression analysis strategy was adopted to allow for a more sensitive comparison of the reported health care experience characterizing the fourth interview of the survey for the respective study groups. For the previously specified health care measures, the dependent variable was redefined as the annualized utilization or expenditure estimate derived from data characterizing a sample participant's fourth interview of the survey. Controlling for the set of explanatory measures specified in the test for data collection frequency effect, a direct test for detecting differentials between the two study groups was implemented. As before, an indicator variable for data collection frequency was included to distinguish the two study groups. Once the explanatory measures were controlled for, differences in the annualized health care utilization and expenditure estimates characterizing the fourth interview of the survey were found to be nonsignificant across study groups (Table 7).

## Comparison of Household Reported Data with Record Check Data

A final analysis was conducted to determine the level of reporting concordance between household and record check data for the study groups. Approximately 32 percent of household survey participants were also included in a Medical Provider Survey (MPS). The Medical Provider Survey was a record check procedure to obtain utilization, expenditure and diagnostic data from physicians and hospitals who treated a sample of household respondents during the year. For those sample participants in the respective study groups with MPS questionnaire response from all their eligible physicians and hospitals, the degree of agreement between household reported and record check data was compared. The comparisons focused on the reporting of outpatient physician contacts and related expenditures, which were characterized by a data collection frequency effect, in addition to the number of hospitalizations.

A weighted correlation coefficient was adopted to measure the strength of the association between household reported and Medical Provider Survey data on the number of outpatient physician contacts and related expenditures. A Fisher's z transform was used to test the hypothesis of equivalence in correlation coefficients across the two study groups. For both the utilization and expenditure data, individuals with only four rounds of data collection were characterized by a significantly higher level of agreement between data sources (Table 8).

Due to the departure from normality for the data on the number of hospitalizations, an alternative index of agreement was considered. The index measured the mean absolute relative difference between household reported and record check data for all individuals with an MPS reported hospitalization. For this health care measure, no significant difference in agreement levels for the two study groups were evident (Table 8). This result complemented the observation of no significant data collection frequency effect for the household reported data.

The 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.

## Summary

In the NMCES, 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 55 years of age or older, women, 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. Round specific comparisons in the reporting of health care events relative to the annual profiles, by length of recall, indicated significantly higher round five congruency ratios for the holdovers with respect to outpatient physician contacts, related expenditures and dental visits. No significant reporting differentials were noted for hospitalizations, related expenditures, and prescribed medicine utilization, demonstrating that the presence of data collection frequency effect was selective in nature.

A direct test for data collection frequency effect confirmed the existence of reporting differentials characterizing the fifth round of data collection for select health care indices. Differences in the annualized health care utilization and expenditure estimates characterizing the fourth interview of the survey, however, were found to be nonsignificant across study groups. Finally, the comparison to determine the level of reporting concordance between household and record check data indicated equivalent or superior performance for the holdovers. 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 Expenditure Survey.

The results of the study 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.

## Acknowledgements

The authors wish to thank Jo Crosby, Sandy Widmar, and Melissa Rowe for their skillful typing of this manuscript. The views expressed in this paper are those of the author and no official endorsement by the National Center for Health Services Reserch is intended or should be inferred.

## Note

Tables 1-8 were not presented in this paper, in addition to the references section, due to space limitations. They may be obtained from the author by writing to: Dr. Steven B. Cohen, National Center for Health Services Research, Room 3-50, Park Building., 5600 Fishers Lane, Rockville, Maryland 20857.