# SAMPLE DESIGN OF THE NMES MEDICAL PROVIDER SURVEY

Steven B. Cohen, Ralph DiGaetano, and Mike Brick

KEY WORDS: NMES, Record Check Survey

## Introduction

The Household Component of the National Medical Expenditure Survey (NMES) was designed to produce unbiased national and regional estimates of the health care utilization, medical expenditures, sources of payment, and health insurance coverage for the U.S. civilian non-institutional population. Data were collected from a panel of approximately 15,000 households, interviewed on four separate occasions over a sixteen month period, to obtain data for calendar year 1987. It was recognized, however, that household respondents were not always the best source for obtaining the required information on medical expenditures. Consequently, a Medical Provider Survey was conducted in order to supplement the household-reported data with medical provider-reported data, for a subset of the medical care events reported by household survey respondents.

More specifically, the NMES Medical Provider Survey (MPS) was primarily designed to serve as a data supplementation strategy to improve the accuracy of medical expenditure estimates derived from the National Medical Expenditure Survey. To meet this objective, all of the providers associated with medical events for which the quality of household reported expenditure data was known to be problematic were targeted for inclusion. In addition, the Medical Provider Survey targeted for selection all of the medical providers associated with a nationally representative sub-sample of households that responded to the Household Survey. This component was included to evaluate reporting differentials between household-reported and provider-reported medical expenditure data.

This paper provides a detailed discussion of the sample design of the NMES Medical Provider Survey and its analytical focus. The sample identification process for the survey is also presented, in addition to a discussion of the planned estimation strategies with MPS data to reduce the bias in survey estimates derived from the National Medical Expenditure Survey.

## Background

The NMES Household Survey (HHS) was designed to produce unbiased national estimates for the general population, for population subgroups of special policy interest, and for the U.S. census regions. The sample is a stratified area probability design with four stages of sample selection : (1) selection of primary sample units (PSUs); (2) selection of segments within PSUs; (3) selection and screening of households within segments; and (4) selection of households based on demographic characteristics (both household and individual) from the set of screened households. The sample of PSUs represents a union of the national sample frames from Westat, Inc. and NORC.

The NMES design required selective oversampling of blacks, Hispanics, the poor and near poor, those 65 years of age and older, and the functionally limited or impaired. A separate screening interview was conducted in the fall of 1986 for a sample of approximately 35,000 addresses to obtain information required to facilitate sample identification of these population subgroups. The screener sample consisted of dwelling units, although the basic analysis units in the NMES are reporting units and persons. The sample dwelling units (DUs) include housing units, group quarters, and other noninstitutional (non-group) living quarters. Dwelling units consist of one or more reporting units. A reporting unit (RU) is defined as a group of individuals related by blood, marriage, or adoption residing in the same housing unit. All members of the civilian population who considered the selected DU as their usual place of residence were included in the interview. Also included in a sample household within the dwelling units were persons considered to be a part of the household but who were temporarily residing elsewhere. Unmarried college students under 22 were selected at their parents' sample address rather than at college, in order to allow for the derivation of health care utilization and expenditure estimates at the family level.

The NMES Round One household sample was selected after the data collection phase for the screening interview was completed and all relevant demographic data necessary for the sample selection of dwelling units were processed (Cohen, DiGaetano, and Waksberg, 1987). The HHS component included four rounds of core interviews, conducted over a 16-month period, so as to provide a full set of annual data on the health care utilization, medical expenditures, health insurance coverage, and related characteristics of survey participants for the reference period January 1, 1987, to December 31, 1987. The interviews were conducted in person for Rounds 1, 2, and 4, and for most respondents by telephone for Round 3. A final Round 5 interview, conducted between May and July, 1988, obtained additional data regarding income tax filing status, day care for children, and pregnancy-related items.

Overall, the NMES household sample consisted of approximately 15,000 reporting units. Based on field results, the overall response rate for the screener round and Rounds 1 through 4 was 79.7 percent.

## The Sample Design of the Medical Provider Survey

In the overall design of the National Medical Expenditure Survey, it was recognized that certain population subgroups in the household survey would not provide "high quality" data on medical expenditures. Furthermore, household-reported charge data for certain medical care events, such as hospitalizations, emergency room visits, and home health care encounters, were expected to be characterized by high levels of item nonresponse and questionable quality. More

specifically, results from the 1977 National Medical Care Expenditure Survey (NMCES), the precursor to NMES, revealed that only half of the population reported the cost of a hospitalization, while only 25 percent of Medicaid beneficiaries and 30 percent of home health users reported charges.

The Medical Provider Survey in NMES was primarily designed to reduce the bias associated with national medical expenditure estimates, derived from household reported data, that was a function of item nonresponse and poor quality data. By selectively targeting those individuals that were most likely to misreport or not possess adequate knowledge about their medical expenditures, and medical care events that were expected to be associated with charge data of questionable quality, optimal use could be made of medical provider reported data to improve the accuracy of national medical expenditure survey estimates. Consequently, the Medical Provider Survey was designed to obtain provider reported charge data for household reported medical care events, and to serve as a data replacement strategy to reduce the level of nonresponse bias in survey estimates due to missing charge data. By also serving to replace household reported charge data of poor quality, the provider survey would also improve the accuracy in estimates derived from the NMES. Furthermore, the data would also be utilized to enhance the imputation strategy that would be employed in NMES to correct for the remaining item nonresponse in expenditure data after the provider survey data replacement strategy was operationalized.

To supplement the data replacement strategy in MPS, and to allow for methodological comparisons on reporting differentials between household and provider reported data at the individual level, the survey included all providers associated with a nationally representative 25 percent sample of the households that completed the Round 1 NMES interview. This component of the Medical Provider Survey would provide a nationally representative pool of provider reported charges for all classes of medical care events identified in the household survey, which would enhance the estimation and imputation strategies employed in NMES. In addition, depending on the response rate that was achieved for the MPS, the survey would yield independent national estimates of medical provider utilization and related expenditures.

It was recognized at the outset of the NMES survey that the survey costs associated with interviewing all of the medical providers associated with the household respondents would be prohibitive. The two complementary design components of the NMES Medical Provider Survey reflect a judicious balance between reductions in survey costs attributable to the 25 percent subsample, while preserving the primary design objective to correct for poor quality household reported charge data.

The definition of a medical provider for the purposes of the Medical Provider Survey includes (a) any Medical Doctor (M.D.) or Doctor of Osteopathy (D.O.) who provides direct patient care; (b) any other medical provider (including inpatient facilities) identified in the household survey providing care under the supervision of an M.D. or D.O.; and (c) any person paid (regardless of the source of payment) to provide home health services as identified in the core questionnaire of the household survey.

Selection of the 25 Percent Sample of Dwelling Units

As noted, a sub-sampling strategy was considered for the Medical Provider Survey, due to the prohibitive costs associated with interviewing all medical providers linked to the NMES household sample. To allow for the derivation of provider reported national utilization and expenditure estimates for population subgroups of particular policy relevance for NMES (the poor, the elderly, the functionally impaired) with acceptable levels of precision, a 25 percent MPS subsample of households was recommended as the minimum required sample size that would not seriously jeopardize analytical objectives. For the MPS, the loss in precision associated with a reduction in sample size due to sub-sampling is calibrated in terms of the increase in the standard errors which characterized survey estimates. Consequently, a 25 percent subsample will increase standard errors of survey estimates by a factor of 2.0, prior to adjusting for further loss in precision due to survey nonresponse.

The design of the 25 percent MPS subsample of NMES households was specified to mirror the NMES Round One sample selection scheme, in order to reflect the oversampling of population subgroups of particular policy relevance. The sample frame for the 25 percent sample of dwelling units (DUs) for the MPS consisted of all DUs with at least one reporting unit (RU) with a completed Round One household interview. Overall, 13,789 DUs comprised the sample frame.

For stratification purposes, and to replicate as closely as possible the sampling scheme that was employed for selection of the NMES household sample, the sample frame was sorted in the following order:

1. Presence in the 10 percent "round 3 exclusion" sample (yes or no). This 10 percent Round 3 exclusion sample was drawn for the purposes of a methodological study. It has been of interest to explore the problems of recall over time in the NMES. Thus, 10 percent of the household sample was randomly selected for exclusion from Round 3 so that responses from this subsample could be compared to those with shorter periods of time between interviews;
2. Within the respective 10 percent sample categories, by the 14 sample domains used in the selection of the NMES household sample (10 collapsed demographic categories plus 4 vacant categories) (DiGaetano, 1987);
3. Within the 14 domains, by the original 25 demographic categories initially defined for selection of the Round One household sample;
4. Within the 25 categories, by a 5 category variable indicating the average number of doctor visits per person in a DU for Round 1 (the categories used were: 0; 1 or 2; 3 or 4; 5, 6, or 7; and 8 or more);
5. Within the average number of doctor visits category, by PSU; and

6. Within PSU, by segment.

A systematic 25 percent sample of DUs was selected after sorting the Round 1 sample by the specified stratification variables. Overall, 3,448 dwelling units were selected for the Medical Provider Survey. All reporting units associated with these DUs were then identified, and all medical providers associated with the "key" sample respondents comprised this component of the MPS. Key sample respondents to the household survey consisted of all civilian non-institutionalized individuals who responded to the Round One interview, in addition to those individuals who joined responding Round One reporting units and did not have an opportunity for selection during the period of time that spanned the Round One field period (new babies, military returning to civilian status, individuals in institutions or outside the country returning to their residence). In the 25 percent sample, medical providers reported as a person's usual source of care were also selected, regardless of whether they provided a service or not (e.g., the patient has a regular doctor but did not see him over the course of 1987).

MPS Certainty Selections: Medicaid Sample

After the 25 percent MPS sample of DUs was drawn, the providers associated with any remaining DUs on the sample frame containing at least one key respondent eligible for Medicaid in Round One were selected for participation. The union of the Medicaid-eligible individuals selected in the 25 percent sample and all remaining Medicaid-eligible individuals represented a certainty sample of all Medicaid-eligible individuals who completed the Round One interview. The sample selection criterion of Medicaid eligibility as of the Round One interview was driven by a MPS design decision to select the MPS dwelling unit based samples prior to fielding the Round 4 household interview. This strategy was adopted to achieve a reduction in survey data collection costs that would be realized by a reduction in the length of the Round 4 interviews for reporting units that were not selected in the 25 percent or Medicaid MPS samples.

With a knowledge of the exact composition of the dwelling unit based samples of the MPS, it was believed that the interviewer burden would be somewhat relieved by the elimination of this sample identification task. Furthermore, the interviewers' progress in obtaining cooperation from household respondents to sign permission forms could be more closely monitored. These permission forms had to be signed to provide authorization for sampled medical providers to release information regarding a NMES respondent's medical utilization, expenditures, and diagnoses. It was recognized that by restricting identification of the Medicaid sample to a NMES respondent's health insurance status as of Round One, this procedure would miss a portion of the NMES sample that was eligible for Medicaid for only part of 1987.

MPS Certainty Selections: Providers Associated With Specific Events

As noted, the MPS was designed to serve primarily as a data replacement strategy to reduce the level of error in household expenditure estimates that was a function of missing or inaccurate data. From the prior 1977 National Medical Care Expenditure Survey, it was determined that the highest levels of missing or inaccurate data were associated with the following medical events: hospitalizations, hospital outpatient visits, emergency room visits, clinic visits (other than visits to a school or company clinic), and home health care events. Consequently, all providers associated with any of these events reported over the four rounds of data collection in the NMES by all household respondents (both in and out of the 25% sample) were selected for participation in the MPS. Furthermore, since the NMES included an institutional population survey which consisted of nursing and personal care homes and facilities for the mentally retarded, all medical providers associated with an admission reported by any household respondent were also selected for participation in the MPS.

MPS Provider Sample

In Round 4 of the household survey, interviewers were responsible for collecting signed permission forms from sample respondents that provided authorization to their medical providers to release information regarding their medical utilization, expenditures, and diagnoses. Computer generated forms were provided for members of dwelling units selected in the 25 percent and Medicaid samples, and for other key household respondents that reported any one of the medical events that comprised the certainty selections during the first three rounds of household data collection. Interviewers were charged with generating additional permission forms in the field during Round 4, for any new reported medical events that comprised the certainty selections, and for obtaining required signatures.

Overall, the joint screener, NMES household survey (4 Rounds of data collection), and permission form signature response rate was 71 percent. Based on the completion of the permission form acquisition activity that occurred in the household survey, the Medical Provider Survey sample was finalized. Table 1 indicates the distribution of providers associated with signed permission forms from key persons identified for inclusion in the MPS.

Table 1. Distribution of providers associated with signed permission forms from key household respondents, by provider type.

| Provider Category Identified for MPS | Number of Providers |
|---|---|
| Home Health Providers | 560 |
| Medium Burden Facilities | 2,048 |
| High Burden Facilities | 321 |
| Low Burden Providers | 11,066 |
| Medium Burden Providers | 223 |
| High Burden Providers | 546 |
| Total | 14,764 |

In the above table, the term "burden" is defined in terms of the number of patients associated with the provider. For facilities, medium burden consisted of 1 to 8 patients while high burden consisted of 9 or more. For physicians, low burden consisted of 1 to 3 patients, medium burden consisted of 4 to 6 patients, and high burden consisted of 7 or more patients. A provider may be an individual office-based physician, a group practice (which may contain several doctors in the sample), a clinic, or an HMO.

Data collection activities for the Medical Provider Survey began in January, 1989, and were targeted for completion in September of 1989. To reduce the potential bias in using MPS data associated with nonresponse, the following response rate goals were specified for the survey. For MPS, the minimum acceptable response rate for the survey was 85 percent, defined as the number of person-provider pairs for which data are collected in the MPS, divided by the number of valid permission forms collected in the household survey. One permission form defined each person-provider pair. In addition, for the 25 percent and Medicaid certainty person-level sample components of the MPS, complete responses to the MPS were required for at least 85 percent of the selected household respondents with only one provider, and at least 70 percent of those with more than one provider.

In the design of the MPS, it was recognized that a significant number of physicians associated with a sampled hospitalization would not bill for their care through the hospital. These physicians were referred to as "separate billing doctors" and were identified through the administration of the MPS hospital questionnaire. Since charge data associated with these separate billing doctors was an essential component of the overall medical expenditures associated with a hospitalization, it was necessary to locate these linked providers to obtain the additional charge information. It was estimated that approximately 5,000 separate billing doctors would be identified in the Medical Provider Survey.

## Estimation in the MPS

### 1. The 25 Percent and Medicaid Sample Components

For the person based component samples of the MPS, as defined by the 25 percent and the Medicaid samples, an estimation strategy will be developed to derive approximately unbiased national estimates of the health care utilization and expenditure parameters that characterize the civilian non-institutionalized population, using the data obtained from their medical providers. The estimation strategy will adjust for person level nonresponse to the household survey, for permission form nonresponse, for medical provider-person pair nonresponse, and for person level nonresponse where not all targeted providers for a given person respond in the MPS.

Using only the provider reported data to make national estimates has certain disadvantages. The first concern is that provider reported estimates are subject to nonresponse from several sources that have a multiplicative effect in reducing the overall response rate. The nonresponse in estimates associated with the inability to obtain some (if any) valid permission forms, and the failure to obtain the required data from the sampled provider, are two additional sources of nonresponse unique to the provider survey. As a result, the level of nonresponse for the provider derived estimates is likely to be substantially greater than the household data, suggesting the presence of a serious component of nonresponse bias.

A second disadvantage of the MPS derived estimates is the increase in the sampling errors associated with using data only from the MPS subsample. Since this component of the MPS sample is based on a 25 percent subsample of the household respondents, the sampling errors for these MPS derived estimates will be at least twice as large as the sampling errors for household derived estimates.

The estimation strategy specified for these person level components of the MPS will consider a straight-forward nonresponse adjustment procedure. More complex adjustments are relatively costly to implement and offer few improvements over the planned strategy. The initial steps of the adjustment strategy will take into account both the permission form and provider level nonresponse. Sampled persons will be categorized based on the number of permission forms required for a sample person (0, 1, 2-3, 4-5, 6+). Adjustments will be made for each of the groups separately, in order to have those individuals for whom all required permission forms are signed and all providers have responded represent the targeted sample.

The rationale for dividing the sample according to the number of required permission forms is related to the likelihood that a full response (permission forms and providers) will be obtained. A person for whom only one permission form is required will be more likely to have a full response in MPS than a person for whom 10 permission forms are required.

The last step in the adjustment procedure is for the purpose of reducing the sampling error associated with the 25 percent subsample in the MPS. The number of persons estimated from the MPS person based samples will be post-stratified to Current Population Survey (Bureau of the Census) totals of the number of individuals in the nation. The adjustment classes will be defined by cross-classifications of age, race/ethnicity, and gender.

### 2. MPS Data Replacement Strategy to Supplement Household Reported Expenditure Estimates

As noted, the MPS was primarily designed to reduce the bias associated with national medical expenditure estimates derived from household reported data. The estimation strategy that was devised to support this data replacement strategy was more comprehensive in nature, making full use of MPS data to correct for missing and poor quality household reported expenditure data. In addition, it includes a recalibration of household reported data, to reflect the reporting differentials observed in expenditure data between households and medical providers.

The foundation on which this estimation strategy rests is the household reported

utilization experience. It is clearly recognized that household reports of medical utilization will be affected by errors of omission and over-reporting that are a consequence of length of recall, lack of knowledge, salience, and proxy response. However, the primary focus of this estimation task will be to correct household expenditure estimates associated with a household reported medical event. At this stage in the MPS estimation strategy, no adjustments to household reported utilization patterns will be made. Separate analyses will be conducted, however, using the person based MPS component samples to assess the level of divergence between household and provider reported national estimates of health care utilization.

For the purposes of this estimation strategy combining the household reported and provider reported expenditure data, the unit of interest is the household reported utilization. A utilization may be a visit to a specific doctor or clinic, or it may be an event involving several providers, such as a hospitalization. Once the data collection phase of the MPS survey is completed, the first stage of this estimation strategy will attempt to match all of the provider reported expenditure data to the household reported utilization.

For a sample person participating in the MPS, there are three distinct outcomes with respect to matching the MPS and the Household survey data. First, the household respondent may report a utilization that matches to the utilization data reported in the MPS. The second possibility is that a utilization is reported in the MPS, but not by the person in the household survey. The third possibility is that a person may report a utilization that does not match any utilization in the MPS. This could happen if the permission form is not signed by the household respondent, if the provider does not respond to the MPS, if there is insufficient information to match their reports, if the provider did not give a complete response, or if the household respondent erroneously reported the event.

A computerized matching algorithm developed at Statistics, Canada, referred to as CANLINK will be used to match household and provider reports of medical care utilization. The matching criteria will include characteristics of the date of the utilization, the type of event (hospitalization, clinic visit, medical provider visit), services rendered, and the household reported condition and provider reported diagnosis that described the purpose of the utilization. The matching rules will be developed to maximize the correct matches while minimizing the false matches and non-matches. The following table illustrates the potential errors in the matching process:

Results of Matching Algorithm to Link Household and Provider Reported Utilizations

| | | Matching Algorithm | |
|---|---|---|---|
| | | Match | Non-Match |
| Truth | Match | correct | false - |
| | Non-Match | false + | correct |

Consider the following steps:

A. For all household and provider reported utilizations that match, and for which MPS reported expenditure data exist, the MPS data will be used as the appropriate value of the expenditure :

Yij = MPS expenditure data for matched utilization j associated with person i.

B. For the subset of household and provider reported utilizations that match and for which both household and provider reported expenditure data exist, the relationship between these alternative sources of expenditure data will be modelled to support a recalibration procedure. More specifically, let Yij be estimated as a model based function of Xij, or

$Yij = f(Xij)$ where

Xij = HHS reported expenditure data for matched utilization j associated with person i.

The purpose of the recalibration procedure is to rescale the person-reported data so that it is comparable to the provider reported data when no match exists. The improvement from recalibration is based on the assumption that the provider's responses are more accurate than the person's expenditure responses. Given this assumption, the recalibration strategy should serve to reduce some of the bias in NMES national expenditure estimates associated with person-level reporting.

Based on the resultant model, all remaining household reported utilizations not included in A. for which a household reported expenditure is present, Xij, will be recalibrated to a predicted provider reported response

$\hat{Yij} = f(Xij)$.

C. The remaining household reported utilizations not characterized in A. and B. for which no household reported expenditure data are present will be corrected by an imputation strategy. The imputation strategy that is implemented to adjust for missing expenditure data, regardless of the technique employed (e.g., whether it is model based or a "hot-deck" approach), should rely (1) wholly on the MPS data, or (2) consider the combination of replacement MPS and recalibrated household data that characterize the household respondents identified in A and B.

## Summary

The complex survey design of the Medical Provider component of the National Medical Expenditure Survey has been described in detail. Particular attention has been given to the sample identification process that facilitated the selection of all medical providers associated with a nationally representative 25 percent subsample of NMES households, a certainty selection of Medicaid eligible households, and explicitly defined health care events. Furthermore, the paper has included a discussion of the analytical focus of the Medical Provider Survey. A discussion is also provided of the planned estimation strategies with MPS data to reduce the bias in survey estimates derived from the household component of the National Medical Expenditure Survey.

References

Cohen, S.B., R.G. DiGaetano, and J. Waksberg (1987). Sample Design of the National Medical Expenditure Survey - Household Component. Proceedings of the American Statistical Association, Section on Survey Research Methods.

Cooley, P.C. (1981) NMCES Matching of MPS and Household Summary Data Methodology Report. Research Triangle Institute, Research Triangle Park, North Carolina.

Cox, B.G., and S.B. Cohen (1985). Methodological Issues for Health Care Surveys. New York, Basel: Marcel Dekker, Inc.

Cox, B.G. and R.E. Folsom (1980) The Sample Design and Weighting Plan for the Medical Provider Survey: An Administrative Record Check Component of the National Medical Care Expenditure Survey. Research Triangle Institute, Research Triangle Park, North Carolina.

DiGaetano, R.G. (1987). Sampling Report for the Selection of Dwelling Units for the Household Survey of the NMES. NMES Report No. 1.00. Westat, Inc., Rockville, Maryland.

Statistics Canada (1985) Generalized Iterative Record Linkage System (CANLINK). Statistics Canada, Ottawa, Ontario.

Williams, R.L. (1979) Medical Provider Survey Imputation Strategy: Expenditure Variables. Research Triangle Institute, Research Triangle Park, North Carolina.