OVERVIEW

The Medicare Current Beneficiary Survey (MCBS) is a continuous, multipurpose survey of a representative sample of the Medicare population. Work on the MCBS is done under the direction of the Centers for Medicare and Medicaid Services (CMS) Office of Research, Development, and Information (ORDI) through its contractor, Westat, Inc. In 2007, the sample included approximately 15,800 beneficiaries either in or joining the continuing sample. Each continuing sample person, or an appropriate proxy respondent, was interviewed three times a year over a four-year period, regardless of whether he or she resided in a community or facility setting. For a description of the MCBS, see the Survey Overview section of our web page.

http://cms.hhs.gov/mcbs

Sample Design

Respondents for the MCBS were sampled from the Medicare enrollment file to be representative of the Medicare population as a whole and by the following age groups: under 45, 45 to 64, 65 to 69, 70 to 74, 75 to 79, 80 to 84, and 85 and over. Because of interest in their special health care needs, the oldest old (85 and over) and the disabled (64 and under) were over sampled to permit more detailed analysis of these sub-populations. The sample was selected by using a stratified, multistage area probability sample design. The first stage in the sampling process was to select 107 primary sampling units (PSUs). The PSUs were selected to be nationally representative and consisted of either counties or groups of counties containing both metropolitan and non-metropolitan areas. Once the PSUs were selected, Zip Code clusters were selected within the PSUs and beneficiaries within those Zip Codes were selected by systematic random sampling to participate in the survey.

Survey Operations

Field work on the MCBS is conducted for CMS's ORDI by Westat, a survey research firm with offices in Rockville, Maryland. Data collection for Round 1 began in September 1991 and was completed in December 1991. Subsequent rounds of data collection, which involve re-interviewing the same sample persons (or their proxies), begin every 4 months. Interviews are conducted regardless of whether the sample person resides in the community or in a long-term care facility, using the version of the questionnaire appropriate for the setting.

In 2007, data were collected from 15,806 beneficiaries for the Access to Care file. The sample included 14,804 persons who lived in the community at the time of their Round 49 (Fall 2007 interview and 1,002 persons who lived in a long-term care facility at the time of their Round 49 (Fall 2007) interview. Interview strategies and survey instruments used to collect data are described below.

Repeat Interviews. The MCBS is a longitudinal panel survey, with sample persons interviewed three times a year over 4 years to form a continuous profile of their health care experience. An initial large sample of 15,411 beneficiaries was fielded in the fall of 1991. Smaller supplemental panels were added in the fall of 1992 and 1993. These supplementary panels were added to adjust for beneficiaries who became entitled to either Part A or Part B benefits during 1991 and 1992 in addition to adjustments required due to death and sample attrition. In 1993, a decision was made to phase out the 1991, 1992, and 1993 panels after no more than six years of interviews and to limit future panels to four years of interviews. The four year rotating panel design was fully implemented for the Round 19 (Fall 1997) interviews.

The Community Interviews. Sample persons in the community are interviewed through computer-assisted personal interviewing (CAPI) survey instruments. The CAPI program automatically guides the interviewer through questions, records the answers, and compares beneficiary's responses to edit specifications for accuracy and relationships to other responses. CAPI improves data collection and lessens the need for after-the-fact editing and corrections. It guides the interviewer through complex skip patterns and inserts follow-up questions where key data are missing from the previous round. When the interview is completed, CAPI allows the interviewer to transmit the data by telephone to the home office computer.

The interviews yield a time series of data on utilization of health services, medical care expenditures, health status and functioning, and beneficiary information such as income, assets, living arrangement, family assistance, and quality of life. To improve the accuracy of the data, respondents are requested to record medical events on calendars provided by the interviewer, and are asked to save Explanation of Benefit forms from Medicare, as well as receipts and statements from private health insurers. To assist in reporting data on prescription medicines, respondents are asked to bring prescription bottles, tubes, and bags (provided by the pharmacy) to the interview.

An effort is made to interview each sample person directly. However, each sample person is asked to designate a potential proxy, usually a family member or close acquaintance, in case he or she is physically or mentally unable to do the interview. On average, about 12 percent of the community interviews in each round are conducted by proxy. The following instruments are used in community interviews:

■ ***The Baseline Questionnaire:*** Collects health insurance, household composition, health status, access to and satisfaction with medical care, and demographic and socioeconomic information for supplemental sample beneficiaries living in household units in the community. Selected information from this questionnaire (primarily health status, and access to and satisfaction with care) is updated annually for continuing sample persons living in the community using *The Community Supplement to the Core Questionnaire.* Additional supplemental questions are added to the core questionnaire in various rounds to gather information about specific topics, including detailed information about the sample person's

income and assets in the spring-summer round of data collection.

■ ***The Community Core Questionnaire:*** Collects detailed health insurance, medical care use, and charge and payment information. This questionnaire is asked in every round, but the initial one.

The Facility Interview. MCBS interviews of persons in long-term care facilities use a similar, but shortened version of the community instrument. A long-term care facility is defined as having three or more beds and providing long-term care services throughout the facility or in a separately identifiable unit. Types of facilities participating in the survey include nursing homes, retirement homes, domiciliary or personal care facilities, distinct long-term care units in hospital complexes, mental health facilities and centers, assisted and foster care homes, and institutions for the mentally retarded and developmentally disabled.

If an institutionalized person returns to the community, a community interview is conducted. If he or she spends part of the reference period in the community and part in an institution, a separate interview is conducted for each period of time. Hence, a beneficiary can be followed in and out of facilities, and a continuous record is maintained regardless of where the person resides.

The initial contact for the institutional interview is always with the facility administrator. Interviews are then conducted with the staff designated by the director as the most appropriate to answer each section of the questionnaire. It was decided early in the design of the study not to attempt interviews with sample persons or family members. The facility interview does not include attitudinal or other subjective items. The facility instruments include:

■ ***The Facility Screener:*** Collects information on facility characteristics such as type of facility, size and ownership. It is used during the initial interview, and in each fall round thereafter.

■ ***The Baseline Questionnaire:*** Collects information on health status, insurance coverage, residence history, and demographics for supplemental sample beneficiaries in facilities and new admissions from the continuing sample. Selected information from this questionnaire (primarily health status) is updated annually for continuing sample persons residing in facilities using an abbreviated version, *The Facility Supplement to the Core Questionnaire.*

■ ***The Facility Core Questionnaire:*** Collects facility use data and charge and payment information. This questionnaire is asked in every round, but the initial one.

MCBS PUBLIC USE FILES

To date, CMS has released public use files (PUFs) on access to care for calendar years 1991 through 2007, and on cost and use for calendar years 1992 through 2006.

Access to Care

The Access to Care Public Use File is designed to provide early release of MCBS data related to Medicare beneficiaries' access to care. Rapid release of access data is achieved by omitting survey reported utilization and expenditure data. The claims information, while limited to program payments for covered services, third party payments for some Medicare secondary payer situations, and potential beneficiary liability, allows significant analysis of the impact of program changes on beneficiaries. This process eliminates the need for imputation of missing cost and payment variables and bypasses the reconciliation of utilization and expenditure data collected in the survey with Medicare claims data.

The content of the Access to Care Public Use File is governed by its central focus. In addition to questions from the access supplement concerning access to care, satisfaction with care, and usual source of care, the file contains demographic and health insurance data as well as data on health status and functioning. To facilitate analysis, the information collected in the survey is augmented with data on the use and program cost of Medicare services from Medicare claims data under Fee-For-Service.

Cost and Use

The MCBS Cost and Use File links Medicare claims to survey-reported events, and provides complete expenditure and source of payment data on all health care services, including those not covered by the Medicare program. Expenditure data were developed through a reconciliation process that combines information from survey respondents and Medicare administrative files. The process produces a comprehensive picture of health services received, amounts paid, and sources of payment. The file can support a broader range of research and policy analyses on the Medicare population, than would be possible using either survey data or administrative claims data alone.

The strength of the Cost and Use file stems from the integration of information that can be obtained only from a beneficiary, and from Medicare claims data on provider services and covered charges. Survey-reported data include information on the cost and use of all types of medical services, as well as information on supplementary health insurance, living arrangements, income, health status, and physical functioning. Medicare claims data include cost and use information on inpatient hospitalizations, outpatient hospital care, physician services, home health care, durable medical equipment, skilled nursing home services, hospice care, and other medical services.

The Sample

The 2007 Access to Care file is compiled on a point-in-time basis. For this reason, special steps were taken to enhance the file to represent an ever-enrolled population. To obtain an ever-enrolled population, it was necessary to account for two groups of beneficiaries. First, sample persons who were enrolled in the Medicare program after January 2007 (called new accretions) and second, sample persons who were alive for part of 2007, but died prior to the Round 49 (Fall 2007) interview. New accretions were included in these statistical tables through a weight adjustment for Round 49 respondents. The sample weights of the respondents who were enrolled for all twelve months were inflated to equal the sum of those enrolled for all twelve months plus the new accretions. This weighting adjustment was performed within each age strata. Beneficiaries who died during 2007 but prior to the Round 49 interview were included in these statistical tables by using their Round 46 survey responses in combination with data from the 2007 Medicare administrative files.

ACCESS TO CARE FILE STATISTICS

The 2007 Access to Care file contains a cross-sectional weight for each of the 15,806 beneficiaries in the data set. These weights reflect the overall selection probability of each sample person. This includes adjustments for survey non-responses and post-stratification to control totals based on accretion status, age, sex, race, region, and metropolitan area status. The weights inflate the sample to an always-enrolled national Medicare population for 2007. The tables in this book are modified so that the population represents a national ever-enrolled 2007 population (see the section entitled The Sample). In general the weights should be used to estimate population totals, percentages, means, and ratios.

Sampling Error

Sampling error refers to the expected squared difference between a population value (a parameter) and an estimate derived from a sample of the population (a statistic). Because the MCBS is a sample of Medicare beneficiaries, statistics derived from the sample are subject to sampling error. The error reflects chance differences between estimates of a population parameter that would be derived from different samples of the Medicare population. Nearly any MCBS estimate of a population parameter (e.g., a percentage, mean, ratio, or count of persons or events) would be affected by the sampling error.

Standard errors have been calculated for all statistics reported in the tables of this book to assess the impact of sampling variability on the accuracy of the estimates. Data from Section 2, Table 2.1 of this book, for example, indicate that 14.25 percent of all Medicare beneficiaries are in excellent health. The standard error of this estimate (0.36 percent) can be used to assess its statistical reliability by constructing a confidence interval that would contain the true value of the population parameter with some given level of confidence.

The confidence interval can be viewed as a measure of the precision of the estimate derived from the sample's data. For example, an approximate 95 percent confidence interval for statistics in this book can be calculated by using the formula

$$\pi = P \pm 1.96 \times \text{(estimated standard error)},$$

where π is the unknown population proportion and P is the calculated (weight) sample proportion. Based on this formula, the approximate 95 percent confidence interval for the estimated proportion of Medicare beneficiaries in excellent health is 14.25 percent plus or minus 0.71 percent. This is a relatively "tight" confidence interval, suggesting that the MCBS data provide a reliable estimate of the true proportion of beneficiaries in excellent health. The chances are about 95 in 100 that the true population proportion falls between 13.54 percent and 14.96 percent.

Another measure of statistical reliability is the relative standard error (RSE) of an estimate. The RSE of an estimate x is calculated by dividing the standard error of the estimate, SE(x), by the estimate, and expressing the quantity as a percent of the estimate, i.e.,

$$RSE = 100\left(\frac{SE(x)}{x}\right).$$

Using data from the previous example, the RSE of the estimated proportion of Medicare beneficiaries in excellent health is 2.52 percent (100 x (.36/14.25)). An RSE of this magnitude would suggest that the estimate is statistically reliable. Statistical reliability of an estimate decreases as the RSE increases.

Many of the statistics in this book are presented by subgroup, some of which are based on relatively small sample sizes. Estimates for these small subgroups can be subject to very large sampling errors. Therefore, it may be desirable in some instances to combine such subgroups with a similar group for analysis purposes. For example, if X_s is an estimated total for the small subgroup, and X_t is the corresponding estimate for the group with which it is combined, then the combined estimate, X_c, is given by $X_c = X_s + X_t$, and the standard error of the combined estimate ($SE(X_c)$) can be approximated as

$$SE(X_c) = \sqrt{[SE(X_s)]^2 + [SE(X_t)]^2}$$

where $SE(X_s)$ and $SE(X_t)$ are the standard errors of X_s and X_t, respectively.

The above approximation applies to estimated totals and should not be used for combining estimates of means of ratios. For the latter types of estimates, the appropriate formulas must include terms representing the proportion of the population that is represented by each of the two component estimates. For example, if Y_s and Y_t are the estimated means for the two subgroups to be combined, then the combined estimate, Y_c, is given by the formula

$$Y_c = P_s Y_s + (1 - P_s) Y_t$$

and the standard error of Y_c can be approximated by

$$SE(Y_c) = \sqrt{[P_{sES}(Y_s)]^2 + [(1 - P_s)SE(Y_t)]^2}$$

where P_s is the proportion of the combined group that is included in the subgroup *s*. It should be noted that both forms of the standard error given above are approximations that may understate the true standard error of the combined estimate.

Confidence intervals and relative standard errors can be calculated for all statistics derived from MCBS data (e.g., totals, percentages, means, ratios, and regression coefficients). The following section provides a brief explanation of the method used to compute the standard errors for the MCBS estimates.

Variance Estimation (Using the Replicate Weights)

The standard errors reported in the tables in this book reflect the complexity of the MCBS sample design. In many statistical packages, the procedure for calculating variances assume that the data were collected in a simple random sample. Procedures of this type are not appropriate for calculating variances for statistics based on a stratified, unequal-probability, multistage sample such as MCBS. They could produce overestimates or, more likely, underestimates of the true sample error. Because the MCBS has a complex design, standard errors in this book's tables were estimated with WesVarPC, a statistical software package that accounts for survey design. Estimates of standard errors from WesVarPC are produced using "replication" methods. The basic idea behind the replication approach is to use variability among selected sub-samples, or replicates, to estimate the variance of the "full-sample" statistics. These methods provide estimates of variance and standard errors for complex sample designs that reflect weighting adjustments such as those implemented in the MCBS. Replication techniques can be used where other methods are not easily applied, and have some advantages even when other methods can be used.

Replicate weights for the MCBS data have been computed using Fay's variant of Balanced Repeated Replication (BRR). BRR is generally used in multistage, stratified sample designs in which two PSUs are sampled within each stratum, possibly with unequal probabilities of selection. The replicate samples are half-samples formed by selecting one of the two PSUs from each stratum. For BRR, the weights for units in the selected PSUs in each half-sample are doubled and the weights for units in the non-selected PSUs are set to zero. Each replicate consists of a different half-sample; however, it is not necessary to form all possible half-sample replicates, since the information from all possible replicates can be captured by using a smaller number of "balanced" half-samples. Fay's method is a variant of BRR, in which the sample weights are adjusted by factors between 0 and 2. With a judicious choice of the perturbation factor, Fay's method provides good estimates of standard errors for a variety of statistics. (For

more information on Fay's method, see D. Judkins, 1990, "Fay's Method for Variance Estimation," *Journal of Official Statistics,* 6: 223-240.)

Replicate weights in the 2007 Access to Care file are named R49CS001...R49CS100. These replicate weights can be used in WesVarPC (the PC version) or WesVar (the mainframe version) to estimate standard errors for MCBS variables. WesVar is available from Westat for a licensing fee. Documentation is provided with the program. A copy of WesVar for IBM PCS (WesVarPC) can be obtained by submitting a request to WESVAR@WESTAT.COM. WesVar is also available for an IBM VMS SAS environment or a VAX VMS SAS environment.

An alternative to WesVar is for the user to write a small custom program using a very simple algorithm. If X_0 is an estimate of a parameter of interest found using the full-sample weights and $X_1,...,X_{100}$ are estimates (calculated by the user) of the same statistic using the corresponding 100 replicate weights, then the estimated variance of X_0 is

$$Var(X_0) = \frac{2.04}{100} \sum_{i=1}^{100} (X_i - X_0)^2$$

A third option is to use another software package such as SUDAAN (Professional Software for SUrvey DAta ANalysis for multistage Sample Designs) to compute population estimates and the associated variance estimates. Two variables, SUDSTRAT and SUDUNIT, have been included in the 2007 Access to Care file for users of SUDAAN.

For information on how to obtain copies of any of the Access to Care Public Use Files or the Cost and Use Public Use Files, send requests to:

Bill Long
Information and Methods Group
Centers for Medicare and Medicaid Services
7500 Security Blvd., Baltimore, Maryland 21244-1850
Telephone (410) 786-7927