Medicare Health Outcomes Survey
Data User=s Guide
Cohort II
February 15, 2000

Preface

This data user=s guide is intended to be a general overview of the data cleaning, editing, and scoring processes required for the creation of the HOS Cohort II data set. A detailed document containing data file specifications will be sent to each Peer Review Organization (PRO) along with their electronic data sets in April 2000.

I. Data Receipt

HOS Cohort II data were transmitted to Health Services Advisory Group (HSAG) from the National Committee for Quality Assurance (NCQA). The data were transmitted on a CD-ROM containing individual ACSII flat files for each participating M+CO. The ASCII files contained plan, patient, and survey information as specified in the HEDIS 1999 Medicare Health Outcomes Survey Manual.

II. Data Cleaning and Editing

A. Import Plan Data

The first step in the creation of the HOS Cohort II data set was to import the ASCII files using SAS software. The SAS code utilized to import each file follows the file structure specified in the HOS Manual and saves each plan file as an individual SAS data set. After each file was imported, the results were examined to assure that the file contained the expected number of records and that the file structure was defined properly.

During the file importation process issues related to file structure and record numbers were identified. For each of these issues corrective action was taken in order to retain the records in the HOS Cohort II data set.

B. Check for Out of Range Values

SAS code was written and executed which listed all responses that were outside the expected range for this data set. All responses were examined for those variables with defined ranges, or where a range of reasonable responses was known. Those items that required editing were corrected in the final HOS Cohort II data set.

C. Consistency Check

All variables in the data file were examined for consistency. Contract numbers were evaluated for concordance with plan identification numbers. No discrepancies were noted. Some demographic variables (birth year, race, and gender) are contained in both the member level data provided by HCFA and the respondent provided survey data. Several inconsistencies were found when these two data sources were compared. Although none of the discrepancies resulted in data being altered, indicator variables were created (***C2BDBRTH***, ***C2BDRACE***, and ***C2BDGNDR***) to identify these beneficiaries. It was decided by HCFA that HCFA reported date of birth, gender, and race would be used for all calculations of age or distributions of gender or race presented in the M+CO baseline report. A few beneficiaries responded that they were female (confirmed by HCFA gender information), and that they were currently under treatment for prostate cancer. While these responses were not changed, or set to Amissing@, they should be excluded (where ***C2BDPRST*** = 1) from any further analyses which might examine those who have prostate cancer.

An additional consistency check was performed which examined skip pattern violations. In many records, beneficiaries failed to correctly follow the skip patterns contained within the survey; however, no changes were made to any of the responses. **Caution should be exercised when examining data which utilizes a skip pattern.**

Several inconsistencies pertaining to survey disposition, round number, and survey language were identified. No corrective action was taken regarding these discrepancies since the overall integrity of the data was not affected. **Caution should be exercised when examining survey disposition, round number, or survey language variables during analysis.** Due to these inconsistencies, a variable (***C2CMPSRV***) was generated to indicate whether a survey was complete based on requirements published in the HEDIS 1999 Medicare Health Outcomes Survey Manual (80% of all questions answered). **It is recommended that this variable (*C2CMPSRV*) be used in all calculations involving completed surveys**.

III. SF-36 Scoring

Statistical analysis included the scoring of the SF-36 using the standardized methodology defined by John E. Ware, Jr., Ph.D. and the Health Assessment Lab (HAL) at the New England Medical Center (NEMC); and multivariate analysis used to case mix adjust baseline scores. The SF-36 questions were scored using the standardized methodology. Most of the variables resulting from the scoring algorithm were incorporated into the data file; however, intermediate variables created during the scoring process were not included in the HOS Cohort II data file. During the process of scoring the SF-36, the scoring algorithm, whenever possible, will fill in missing data with an average of completed data. During the processing of the HOS Cohort I data, these estimated values used in the scoring of the SF-36 replaced the original data in the file. This resulted in values of data which were not possible on the original survey. However, during the processing of HOS Cohort II data, the original responses to all questions were maintained.

To validate proper scoring of the SF-36 portion of the HOS, HSAG submitted a random sample of 30,000 unscored records to HAL. HAL independently scored the records and these 30,000 scored records were compared to the same records as scored by HSAG. It was confirmed that all SF-36 scale scores and SF-36 summary scores, as well as minimum, maximum, and standard deviation results were replicated by HAL.

Case mix adjustment variables included: demographic and socioeconomic characteristics, patient reported comorbidities, and a series of variables unique to the HOS study design (geographic region, data collection vendor, mode of administration and person who completed the survey). See Table 1 for a complete listing of the specific variables included in the case mix adjustment.

Table 1 Covariates Used in the Case Mix Adjustment Model of the HOS SF-36 Measures

Demographics
- Age (Continuous)
- Gender (Male or Female)
- Race (White, Black, Other Minority)
- Education
- Marital Status
- Income

Comorbid Medical Conditions (Beneficiary Reported on Health Outcomes Survey)
- Hypertension or high blood pressure
- Angina pectoris or coronary artery disease
- Congestive heart failure
- Myocardial infarction or heart attack
- Other heart conditions such as problems with heart valves or the rhythm of heartbeat
- Stroke
- Emphysema, or asthma, or COPD (Chronic Obstructive Pulmonary Disease)
- Crohn=s disease, ulcerative colitis, or inflammatory bowel disease
- Arthritis of the hip or knee
- Arthritis of the hand or wrist
- Sciatica
- Diabetes, high blood sugar, or sugar in the urine
- Any cancer other than skin cancer

HOS Study Design Variables
- Who Completed Survey (Self or Other)
- Mode of Survey Administration (Mail or Telephone)
- HCFA Region
- Vendor

Table 2 below lists the SF-36 Elderly Population Norms provided by HAL at NEMC. The scores were limited to those 65 years of age or older, and were provided from unpublished SF-36 data for comparison to the HOS results.

Table 2 SF-36 Elderly Population Norms

SF-36 Measure	N	Elderly Mean	Standard Deviation
Physical Component Summary (PCS) Score	657	41.41	11.40
Mental Component Summary (MCS) Score	657	51.88	10.20
Physical Functioning (PF) Scale	657	41.19	12.41
Role-Physical (RP) Scale	657	43.24	12.53
Bodily Pain (BP) Scale	657	46.05	11.35
General Health (GH) Scale	657	44.48	10.80
Vitality (VT) Scale	657	48.11	11.02
Social Functioning (SF) Scale	657	47.96	11.80
Role-Emotional (RE) Scale	657	48.05	11.77
Mental Health (MH) Scale	657	50.82	10.47

IV. HOS Cohort II Data File Characteristics

After the data file was cleaned and edited, additional variables were added to the file. Plan specific variables included number of ineligible beneficiaries, sample size, total number of completed surveys, number completed by mail, number completed by telephone, overall response rate, mail response rate, and telephone response rate. All date variables contained in the data file were converted to SAS date format (elapsed date variables) to facilitate calculation of the duration of enrollment and age, which were then incorporated into the data file.

In addition to those variables listed above, general plan characteristics information was downloaded from the HCFA monthly report located at the following web site: www.hcfa.gov\stats\monthly.htm.

The statistics from the April 1999 report were incorporated into the HOS Cohort II data file and included the following plan specific variables: type, model, population, description, HCFA region, state, tax status, and contract start date. Duration of plan contract was calculated and incorporated into the data file.

The final HOS Cohort II sample (all M+COs nation wide, excluding those plans participating in the Frail Elderly Pilot, i.e. PACE and EverCare) has the following characteristics:

! Of the 301,184 records submitted to HSAG, 194,378 were completed surveys, based on the HOS Manual specifications, (***C2CMPSRV*** = 1) for an overall completed survey rate of 66.5%.

! Of the 194,378 completed surveys, 180,482 are considered the ABaseline Data Report Sample@ which is limited to respondents age 65 or older whose reason for entitlement is equal to 10 (aged without ESRD). This is the AHOS Total@ which is presented on all of the data illustrations. These respondents are identified by the variable ***C2ANALYT***.

! The case mix adjustment was computed using 127,823 cases.

! All variables in this data set (with the exception of ***HICNUM***) begin with the characters AC2". The C = Cohort (as opposed to Remeasurement), and 2 = Cohort II. Variables for Cohort III will be named C3... Variables for the remeasurement of Cohort I will be named R1..., etc. The variable ***HICNUM*** is the unique identifier for each beneficiary, and does not ever change regardless of which sample is being analyzed.

Any additional questions regarding data quality issues, changes made to the data set, or the structure of the final HOS Cohort II data set should be submitted to:

azpro.hos@sdps.org
or
1-888-880-0077