

The Centers for Medicare & Medicaid Services' Office of Research, Development, and Information (ORDI) strives to make information available to all. Nevertheless, portions of our files including charts, tables, and graphics may be difficult to read using assistive technology. Persons with disabilities experiencing problems accessing portions of any file should contact ORDI through e-mail at [ORDI\\_508\\_Compliance@cms.hhs.gov](mailto:ORDI_508_Compliance@cms.hhs.gov).

**Medicare Health Outcomes Survey**  
**Data Users Guide**  
***Cohort III Baseline***  
**March 15, 2001**

**Preface**

This data users guide is intended to be a general overview of the data cleaning, editing, and scoring processes required for the creation of the HOS *Cohort III Baseline* 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 June 2001.

**I. Data Receipt**

HOS *Cohort III Baseline* 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 plan. These plans included all Medicare+Choice Organizations (M+COs) and continuing cost contractors, Program of All-Inclusive Care for the Elderly (PACE) plans, Social HMOs, Medicare Choices, and DOD Subvention Demonstration Plans with contracts in place on or before January 1, 1999. In total, 326 individual files were submitted containing 302,714 records. The ASCII flat files contained plan, patient, and survey information as specified in the *2000 HEDIS, Volume 6, Specifications for the Medicare Health Outcomes Survey* (also referred to as the 2000 HOS Manual) published by NCQA in February 2000.

**II. Data Cleaning and Editing**

**A. Import Plan Data**

The first step in the creation of the HOS *Cohort III Baseline* 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 2000 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 III Baseline* data set.

**B. Check for Out of Range Values**

**Medicare Health Outcomes Survey  
Data User's Guide - Cohort III Baseline  
March 15, 2001**

To verify the presence of unique beneficiaries in the HOS data file, the file was examined for duplicate HIC numbers. Duplicate HIC numbers were found for 133 records, however, subsequent review revealed these records to be members of two PACE plans who directly submitted their sample files to the vendors with HIC numbers equal to 999999999. Several variables were created, such as unique identifiers for each record and each contract market area, as defined by plan identification number. An indicator for the beneficiaries of the PACE plans (**C3PACE**) was added. To be consistent with the variables included in the *Cohort II Baseline* data files, an EverCare indicator (**C3EVER**) was also created; however, no EverCare plans were included in the *Cohort III Baseline* administration.

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. Frequency distributions of all categorical variables as well as cross tabulations by vendor were performed to identify both out of range values and data shifts in value assignment. The cross tabulations were performed utilizing the entire data file and also specified subsets of the data file (M+CO plans only and PACE plans only). Those items that required editing were corrected in the final HOS *Cohort III Baseline* data set. All dates contained within the data file were verified to correspond to the appropriate range. Some records submitted by selected vendors contained values for survey reported year of birth outside of the acceptable range. **Caution should be exercised when examining the survey reported year of birth variable (C3BRTHYR).**

The cross tabulations indicated that selected information in the variables for institutional status, Hospice status, Medicaid status, and reason for entitlement located in the header record row were missing for selected records, although further investigation revealed that this information was predominantly missing from selected PACE plans. No corrective action was required. The records with missing information were primarily limited to plans that had provided their own samples directly to the vendor. **Caution should be exercised when examining these variables for members of PACE plans.**

### **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

**Medicare Health Outcomes Survey  
Data User's Guide - Cohort III Baseline  
March 15, 2001**

the discrepancies resulted in data being altered, indicator variables were created (*C3BDBRTH*, *C3BDRACE*, and *C3BDGNDR*) 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 *C3BDPRST* = 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. When appropriate, corrected data was ascertained and incorporated into the data file. A variable (*C3PCTCMP*) was calculated by HSAG to document the percent completeness of the survey portion of each record, and an indicator variable (*C3CMPSRV*) was generated to indicate whether a survey was complete based on requirements published in the 2000 HOS Manual (greater than 79.5% of all questions answered). **It is recommended that this variable (*C3CMPSRV*) be used in all calculations involving completed surveys.**

### **III. SF-36<sup>®</sup> Scoring**

Statistical analysis included the scoring of the SF-36<sup>®</sup> 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<sup>®</sup> 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 III Baseline* data file.

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.

**Medicare Health Outcomes Survey  
Data User's Guide - Cohort III Baseline  
March 15, 2001**

**Table 1 Covariates Used in the Case Mix Adjustment Model of the HOS SF-36<sup>®</sup> 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

---

**Medicare Health Outcomes Survey  
Data User's Guide - Cohort III Baseline  
March 15, 2001**

Table 2 below lists the SF-36<sup>®</sup> 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<sup>®</sup> data for comparison to the HOS results.

Table 2 SF-36<sup>®</sup> Elderly Population Norms

<b>SF-36<sup>®</sup> Measure</b>	<b>N</b>	<b>Elderly Mean</b>	<b>Standard Deviation</b>
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 III Baseline 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.

**Medicare Health Outcomes Survey  
Data Users Guide - Cohort III Baseline  
March 15, 2001**

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](http://www.hcfa.gov/stats/monthly.htm). The statistics from the April 2000 report were incorporated into the HOS *Cohort III Baseline* 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 III Baseline* sample (all M+COs nation wide, excluding PACE plans) has the following characteristics:

- ! Of the 298,883 records submitted to HSAG, 208,655 were completed surveys, based on the 2000 HOS Manual specifications, (*C3CMPSRV* = 1) resulting in an overall response rate of 71.7%.
- ! Of the 208,655 completed surveys, 195,353 are considered the “Baseline 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 “HOS Total” that is presented on all of the data illustrations. These respondents are identified by the variable *C3ANALYT*.
- ! Due to missing data in the variables included in the adjustment model for some records, the case mix adjustment was computed using 143,123 cases.
- ! All variables in this data set (with the exception of *HICNUM*) begin with the characters “C3”. The C = Cohort (as opposed to Remeasurement), and 3 = *Cohort III Baseline*. Variables for Cohort IV will be named C4... 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 III Baseline* data set should be submitted to Health Services Advisory Group at:**

**azpro.hos@sdps.org**

**or**

**1-888-880-0077**