Care Management for High-Cost Beneficiaries (CMHCB) Demonstration Sample Data

Data Description

The Centers for Medicare and Medicaid Services (CMS) has created a sample data file of high-cost beneficiaries to help applicants prepare proposals for the CMHCB demonstration. CMS analyzed claims and risk scores to define selection criteria that would yield a sufficiently large sample of beneficiaries with a substantial potential for savings. The final data file includes those beneficiaries who had claims greater than or equal to $6,000 or a Hierarchical Coexisting Condition (HCC) based risk score greater than or equal to 1.7 in 1999.

The data file contains data for a representative 5 percent sample of Medicare beneficiaries who were continuously enrolled in the traditional Fee for Service (FFS) program throughout 1999, 2000 and 2001, and who met the selection criteria. It includes beneficiaries who were enrolled at the beginning of 2001, but who may have died during the year. The file does not include beneficiaries who died during 1999 or 2000.

The data file is approximately 151 MB in size and contains 265 columns of data for 282,521 beneficiaries. It includes summary claims data, HCC risk data, and Diagnosis Related Group (DRG) data for 1999, 2000, and 2001. The HCC assignment process allocates beneficiaries to one or more of 70 HCC categories based on ICD-9 codes. The HCC categories are used to calculate HCC risk scores. A risk score for one year is broadly predictive of spending in the next. More information on HCCs can be found on the CMS website following the links below:

1. CMS website with background information regarding HCCs.
   http://www.cms.hhs.gov/healthplans/riskadj/PilotReporttoplansfn.pdf

2. CMS website with current and historical information on the use of HCC and Risk Score for rate setting purposes. The middle of the web page has information on the use of the HCC model's development and background reference material.
   http://www.cms.hhs.gov/healthplans/rates/

3. CMS website with FFS statistical data for Part A and B claims, some tables include state data.
   http://www.cms.hhs.gov/statistics/feeforservice/default.asp

Files on this CD

This CD includes four files. They are:

1. CMHCB_sample (152 MB) - this file is the fixed width (column delimited) raw sample data file. It includes claims data, HCC data, DRG data for 3 years (1999-2001), see description above.
2. data_dictionary.xls (52 KB) - this is an MS Excel file that provides variable names, descriptions and layout information (variable position and length).
3. input_statement.sas (20 KB) - this is a Statistical Analysis System (SAS) file that provides a SAS input statement that will put the raw data file (CMHCB_sample) into a SAS data set.
4. readme.rtf (10 KB) - this rich text format file includes descriptive information about the

data and the files on the CD.

Data Confidentiality

To preserve confidentiality for Medicare beneficiaries, the Medicare beneficiary identification numbers on the sample file have been encrypted.

Additional Data Notes

The variable 'Total Part A and B claims payments' is not necessarily the sum of inpatient, outpatient, physician and DME claims payment variables. This is due to other claims data not included in the selection set.

Beneficiaries generally do not have a single HCC and may have multiple HCCs. There is no rank order or priority assigned to an HCC group based on co-morbid conditions or expense. The one exception is diabetes (HCC 15-19). There should only be one of the five diabetic HCCs per individual; however, the beneficiary may have other HCCs.

The data in the sample file are for calendar years, not fiscal years, and the DRG variables are the first 10 DRGs occurring in the calendar year.

Website

For additional information please visit the CMHCB website:
http://www.cms.hhs.gov/researchers/demos/cmhcb.asp