HCUP CODING PRACTICES

Table of Contents

HCUP CODING PRACTICES

The following objectives guided the definition of data elements included in the HCUP databases:

- Make the database as usable as possible without extensive editing by analysts.

- Retain the largest amount of information available from the original sources, while still maintaining consistency among sources.

- Structure the information for efficient storage, manipulation, and analysis.

- Set data element attributes (type and length) to accommodate all expected discharge data. The required characteristics were determined from:
  - The actual characteristics of state and hospital association data tabulated in the HCUP Feasibility Study (*AHCPR Hospital Cost Database Feasibility Study*, Contract No. 282-90-0029).
  - National standards, including the Uniform Hospital Discharge Data Set (UHDDS), Uniform Bill 1982 (UB-82), Uniform Bill 1992 (UB-92), and Uniform Bill 2004 (UB-04).

CODING OF DATA ELEMENTS

Data elements are coded as shown in the following table:

Coding Conventions	
Values have been:	**Examples of data elements:**
Retained in the form provided by the data source	Diagnosis and procedure codes, revenue codes, payer coding
Encrypted into synthetic values	Physician identifiers, patient linkage numbers
Recoded into uniform coding schemes	Sex, race, expected primary pay source
Calculated (when possible)	Age, length of stay, day of principal procedure
Assigned using external algorithms	Medicare-Severity Diagnosis Related Groups (MS-DRGs), Clinical Classifications Software Refined (CCSR)

Some data elements are only retained in the form provided by the data source. Other data elements are retained in both an edited or uniformly recoded scheme and in its original form provided by the data source. In either case, the data element for the information retained in its original form uses the suffix _X (i.e., the naming convention of [data element name]_X). For example, expected primary pay source is uniformly coded as data element PAY1 and retained in its original form as PAY1_X.

Data elements that are encrypted into synthetic values have the naming convention of [data element name]_R for re-identified. For example, medical record number is coded as data element MRN_R.

ATTRIBUTES OF DATA ELEMENTS

Data elements are defined as numeric or character.

- Numeric format is used for data elements that are reasonable to express numerically (e.g., age of patient); and for most categorical data elements (e.g., sex of patient).

    o Categorical data elements are expressed in numeric format because that format:
        - facilitates logical comparisons of indicator data elements and
        - permits flexibility in the creation of summary statistics.

- Character format is used for data elements that contain alphanumeric characters not amenable to recoding. Some data elements are expressed in character format because:
    o the alphanumeric data have a recognized significance that must be preserved (e.g., ICD-10-CM/PCS diagnosis and procedure codes); and
    o there is no reasonable conversion to numeric coding (e.g., encrypted physician identifiers).

- To save storage space, data element lengths are limited to what is necessary to accommodate the expected data.

MISSING VALUES

Special missing values have been used in HCUP data elements to indicate details of data availability and quality. Missing values differ depending on whether you have obtained HCUP data in SAS or ASCII formats.

- **Missing Data**

    When:
    -        the source has defined an explicit value as unknown or unavailable
    -        the source uses a default missing value to indicate missing data
    -        exploratory statistics show an undocumented value with a frequency suggestive of a missing value, and it is a commonly used missing value (e.g., blank, zero, or 9-filled), or when contacted, the source confirms that the value is unknown or unavailable

    The following missing values are assigned:

    SAS
    -        a value of "." for numeric data elements
    -        " " (blank) for character data elements

    ASCII
    -        a negative 9-filled value (-9, -99, -999, etc.) for numeric data elements
    -        " " (blank) for character data elements

- **Invalid Data**

When the source data contain undocumented, out-of-range, or invalid values, e.g., an invalid date, an alpha character in a numeric field, or a value not documented in the Partner-provided source documentation, the following missing values are assigned:

SAS
- a value of ".A" for numeric data elements
- "A" for character data elements

ASCII
- a negative 8-filled value (-8, -88, etc.) for numeric data elements
- "A" for character data elements

- **Data Unavailable from Source**

  In the 1988-1997 HCUP databases, when the data source did not provide a data element, the following missing values were assigned:

  SAS
  - ".B" for numeric data elements

  ASCII
  - a negative 7-filled value (-7, -77, etc.) for numeric data elements

  To conserve space, data elements that were unavailable from the source, i.e., coded as .B for all records in a year, were excluded from the HCUP databases starting in 1998 and some previous years of the publicly released State databases.

- **Inconsistent Data**

  Related data elements within the same record were checked for logical consistency, e.g., a procedure of hysterectomy reported with a sex of male is inconsistent. When such inconsistencies were identified, the following missing values were assigned:

  SAS
  - ".C" for numeric data elements

  ASCII
  - a negative 6-filled (-6, -66, etc.) value for numeric data elements

  See the HCUP Quality Control Procedures section for details on data editing.

- **Not Applicable Data**

  Prior to 2001, when the information is not applicable, e.g., the indication of a HMO or PPO plan for "No Charge" patients, the following missing values are assigned:

  SAS
  - a value of " .N" for numeric data elements

  ASCII
  - a negative 5-filled value (-5, -55, etc.) for numeric data elements

Coding Practices