# Table of Contents

HCUP CODING PRACTICES ................................. 1
CODING OF DATA ELEMENTS ............................... 1
ATTRIBUTES OF DATA ELEMENTS ........................... 1
MISSING VALUES .............................................. 2
  Missing Data .............................................. 2
  Invalid Data .............................................. 2
  Data Unavailable from Source .......................... 2
  Inconsistent Data ......................................... 3
  Not Applicable Data ..................................... 3
DIAGNOSIS AND PROCEDURE DATA ELEMENTS ............ 3
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), and Uniform Bill 1992 (UB-92).

CODING OF DATA ELEMENTS

Data elements are coded as shown in the following table:

<table>
<thead>
<tr>
<th>Coding Conventions</th>
<th>Examples of data elements:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Values have been: Retained in the form provided by the data source</td>
<td>Diagnosis and procedure codes</td>
</tr>
<tr>
<td>Encrypted into synthetic values</td>
<td>Physician identifiers, person identifiers</td>
</tr>
<tr>
<td>Recoded into uniform coding schemes</td>
<td>Sex, race, expected primary pay source</td>
</tr>
<tr>
<td>Calculated (when possible)</td>
<td>Age, length of stay, day of principal procedure</td>
</tr>
<tr>
<td>Assigned using external algorithms</td>
<td>Diagnosis Related Groups (DRGs), Clinical Classifications Software (CCS)</td>
</tr>
</tbody>
</table>

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).

  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:
- the alphanumeric data have a recognized significance that must be preserved (e.g., ICD-9-CM diagnosis and procedure codes); and
- 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 EBCDIC/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

EBCDIC/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, or an alpha character in a numeric field, the following missing values are assigned:

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

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

• **Data Unavailable from Source**

In the 1998-1997 HCUP databases, when the data source did not provide a data element, the following missing values were assigned:
SAS
- "B" for numeric data elements

EBCDIC/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

EBCDIC/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**

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

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

**DIAGNOSIS AND PROCEDURE DATA ELEMENTS**

The coding of the diagnosis/procedure-specific data elements is interdependent and changes over time. These data elements are:

• Diagnoses (DXn) and procedures (PRn),
• Validity flags (DXVn and PRVn) in the 1988-1997 data only, and
• Clinical Classifications Software (CCS) codes (DXCCSn and PRCCSn beginning in 1998 and DCCHPRn and PCCHPRn in 1988-1997). CCS was formerly known as Clinical Classification for Health Policy Research (CCHPR).

Starting in the 1998 data, invalid or inconsistent diagnoses and procedures were masked directly instead of setting corresponding validity flags.

The following table demonstrates the relationship between these variables.
<table>
<thead>
<tr>
<th>Diagnosis (DXn)/Procedure Code (PRn)</th>
<th>Year of Data</th>
<th>Validity Check</th>
<th>CCS Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing (Blank)</td>
<td>1988-1997</td>
<td>DXVn/PRVn = missing (.) or (-9)</td>
<td>DCCHPRn/PCCHPRn = missing (.) or (-999)</td>
</tr>
<tr>
<td></td>
<td>Starting in 1998</td>
<td>No action</td>
<td>DXCCSn/PRCCSn = missing (.) or (-999)</td>
</tr>
<tr>
<td>Valid Code</td>
<td>1988-1997</td>
<td>DXVn/PRVn = 0 if consistent with age and sex; DXVn/PRVn = .C or -6 if inconsistent</td>
<td>DCCHPRn = 1-260, PCCHPRn = 1-231</td>
</tr>
<tr>
<td></td>
<td>Starting in 1998</td>
<td>No action</td>
<td>DXCCSn = 1-259, 2601-2620, PRCCSn = 1-231</td>
</tr>
<tr>
<td>Invalid Code</td>
<td>1988-1997</td>
<td>DXn/PRn = “invl”</td>
<td>DCCHPRn/PCCHPRn = .A or -888</td>
</tr>
<tr>
<td></td>
<td>Starting in 1998</td>
<td>DXn/PRn = “invl”</td>
<td>DXCCSn/PRCCSn = .A or -888</td>
</tr>
<tr>
<td>Inconsistent Code (DXn/PRn inconsistent with age or sex)</td>
<td>1988-1997</td>
<td>DXVn/PRVn = .C or -6</td>
<td>DCCHPRn = 1-260, PCCHPRn = 1-231</td>
</tr>
<tr>
<td></td>
<td>Starting in 1998</td>
<td>DXn/PRn = “incn”</td>
<td>DXCCSn/PRCCSn = .C or -666</td>
</tr>
</tbody>
</table>

See the *HCUP Quality Control Procedures* section for details on diagnosis and procedure edits.