Overview of the HCUP American Hospital Association Annual Survey SAS® Databases

1987-2003

March 7, 2005

TABLE OF CONTENTS

INTRODUCTION **1**

AHA ANNUAL SURVEY DATA **2**

Consistency of Data Elements Across Years 2

Data Features of Note 2

UNIFORM SAS RESEARCH DATABASE CONSTRUCTION **6**

Data Element Naming Conventions 6
- Prefixes 6
- Numeric Roots 8
- Suffixes 9
- Other Data Elements 9

Data Element Labeling Conventions 10

HCUP AHA ANNUAL SURVEY DATABASE DOCUMETATION **10**

Special Reports 11
- AHA Survey Changes 11
- AHA Hospital Changes 11

Description of Data Elements 11
- HCUP AHA Annual Survey Database Construction Index 11
- Description of Data Elements for AHA 11
- AHA Annual Survey Database File Layout 12
- Annotated AHA Annual Survey Instrument 12
- 1970-2003 HCUP AHA SAS Databases Data Element Availability 12

Summary Statistics 12

Processing Programs 12
- AHA SAS Load Program 12
- AHA Summary Statistics Program 13

CONCLUSION **14**

INTRODUCTION

The American Hospital Association (AHA) conducts an annual survey of hospitals in the United States and associated territories and produces an Annual Survey Database™. This database includes data elements containing demographic, utilization, financial, and other hospital characteristics of interest to health care researchers. Examples include:

- Beds
- Inpatient days
- Employees
- Medical staff
- Financial information
- Organizational structure (e.g., ownership and control).

Each year the Healthcare Cost and Utilization Project (HCUP), sponsored by the Agency for Healthcare Research and Quality (AHRQ), licenses the Annual Survey Database from the AHA. HCUP database developers reprocess the AHA Annual Survey Databases into Statistical Analysis System (SAS®) research databases with uniform data element names and coding across years. The HCUP databases contain the AHA Annual Survey data elements and additional hospital characteristics indicators calculated from the AHA data. HCUP uses AHA data for a variety of purposes, including the following:

- Identifying U.S. community, non-rehabilitation hospitals for the NIS and KID (AHA data element Z998 = 1).
- Stratifying, sampling, and weighting hospitals and discharges for the NIS and KID.
- Linking hospital characteristic values to patient data for research.

This overview describes the contents, construction, and documentation of the HCUP AHA Annual Survey SAS Databases. Its purpose is to help researchers understand and utilize the data and documentation.

AHA ANNUAL SURVEY DATA

Consistency of Data Elements Across Years

The 1987-2003 AHA Annual Survey Databases provide information corresponding to hospital fiscal years. Although most data elements are consistently represented across years, data element lists do vary by year. This is because the AHA updates the Annual Survey instrument each year, adding some data elements and deleting others from the survey.

The 2003 AHA database added 71 new data elements and deleted eight previously available data elements. There were 36 new data elements from Survey Section C, Facilities and Services, and 35 new data elements from Survey Section E, Total Facility Beds, Utilization, Finances and Staffing. The AHA moved three data elements, A266A, A266B, and A266C, from Survey Section B, Organizational Structure, to Section E, Total Facility Beds, Utilization, Finances and Staffing. The AHA also moved these three data elements to the end of the file layout and gave them new names with slightly different descriptions.

The 2002 AHA database added 33 new service data elements from Survey Section C, Facilities and Services, and two new data elements from Survey Section D, Community Orientation. Beginning with the 2001 AHA database, Medstat added a new data element (Z998) to identify the universe of U.S. community, non-rehabilitation hospitals, as defined for the Nationwide Inpatient Sample (NIS). Also in 2001, the AHA deleted three data elements from Survey Section B.3.c, Organizational Structure. The 2000 database included 28 new service data elements and dropped five previously available data elements, all from Survey Section B, Organizational Structure. The 1999 database included four new group purchasing data elements. The 1998 database included 40 new data elements, mostly service data elements, and dropped nine previously available data elements, primarily from Survey Section B, Organizational Structure. The 1997 database included a number of new data elements.

The 1994-1996 databases contained the same set of data elements. The AHA redesigned the survey in 1994 and made major changes to the survey instrument. In the redesign, the AHA dropped many data elements, added or split others into separate categories, and changed the definitions of some data elements. Less than half of the data elements in the 1994 database were consistent with the previous year. From 1987 to 1993, most of the data elements were consistent from year to year.

Data Features of Note

The definitions of most data elements remain constant from one year to the next. There are, however, some exceptions to consider. This section highlights significant changes in data element coding.

1. Beginning with the 1994 HCUP AHA SAS Database, discharges (B005, B005H,

B005N) were imputed to be equal to admissions (B003, B003H, B003N), because the AHA dropped the discharges data elements from the file. Note that the AHA admissions, discharges, and inpatient days totals do not include newborns (B903).

2. Naming and coding of service and raw service availability data elements changed beginning with the 1994 database. Prior to 1994, service availability was coded in a single data element that may have been assigned based on the number of beds for that service. Service and raw service data elements have been split out, so that for each service type there are now four data elements defining service availability at the hospital, health system, network, and joint-venture levels.

   For example, prior to 1994, raw service availability values were coded as single data elements with the following possible values:

   1 = Provided by hospital
   3 = Provided by another hospital or provider
   4 = Service not available.

   The resulting service data element had one of the following values:

   0 = Service not available or provided by another hospital
   1 = Hospital-based service available.

   Beginning with the 1994 database, raw service and service data elements are coded with the following suffixes:

   [Blank] = Hospital or subsidiary
   S = Health system (locally)
   K = Network (locally)
   V = Joint venture (locally).

   Raw service availability values are coded with the following possible values in the 1994 database:

   1 = Provider
   2 = Not provider.

   Beginning with the 1995 database, raw service availability values are coded with the following possible values:

   1 = Provider
   0 = Not provider.

   The resulting service data element has one of the following values:

0 = Service not available
1 = Service available.

Consider the following example of the service data elements for "GENERAL MEDICAL AND SURGICAL CARE (ADULT)":

SAS Data Element Name	AHA Item Name	Description
S011	GENHOS	Hospital or Subsidiary
S011S	GENSYS	Health System (locally)
S011K	GENNET	Network (locally)
S011V	GENVEN	Joint Venture (locally)

Prior to the 1994 database, AHA surveys coded the service simply as S011, with values indicative of the service level.

3. Prior to the 1995 database, most of the raw binary (yes/no) data elements had the following values:

   1 = Yes
   2 = No.

   Beginning with the 1995 database, the values of most of the raw binary (yes/no) data elements changed to the following values:
   1 = Yes
   0 = No.

   The AHA did not document this change in the 1995 Annual Survey Database documentation.

4. The coding of A500, AHA Membership code, changed beginning with the 1997 database to the following values:

   Y = AHA MEMBER
   N = NOT AN AHA MEMBER.

   A500 was coded as follows in the 1995 and 1996 databases:

   01 = YES (REGISTERED A.H.A. MEMBER)
   00 = NO (NONREGISTERED A.H.A. MEMBER).

Prior to the 1995 AHA database, the values of A500 were coded as follows:

01= REGISTERED, SORT-TERM, A.H.A. MEMBER
03 = REGISTERED, LONG-TERM, A.H.A. MEMBER
50 = REGISTERED, NON-A.H.A. MEMBER
53 = NONREGISTERED, NON-A.H.A MEMBER
OM = NONREGISTERED, OSTEOPATHIC MEMBER OF A.O.H.A.
ON = NONREGISTERED OSTEOPATHIC NONMEMBER OF A.O.H.A
25 = NONREGISTERED, PROVISIONAL A.H.A. MEMBER.

5. In the 1997-2001 databases, the values of the following imputation flags are questionable because they are always equal to zero: I_E100H and I_E200H.

6. Prior to the 1998 database, the service data elements were set to zero both when the service was not available, and when the value was missing for non-responding hospitals. Starting with the 1998 database, they are set to missing when the corresponding raw AHA input data element is missing.

7. Prior to the 1998 database, the following data elements were coded in the AHA Annual Survey Databases and the HCUP SAS databases using numeric AHA state codes with values ranging from 3-95:

SAS Data Element Name	AHA Item Name	Description
A210C	MNGTSTCD	MANAGEMENT ORGANIZATION STATE CODE
A230C	ALL1STCD	FIRST ALLIANCE STATE CODE
A260C	MHSSTCD	HEALTH CARE SYSTEM STATE CODE

Starting with the 1998 database, the coding of the data elements changed to two-character state postal codes in both the AHA survey data and the HCUP SAS files.

8. In the 2000 database, the values of A210A and A210B appeared in upper case only; in contrast, in the 2001 database, the values appear in both upper and lower case.

9. In the 2001 database, the values of the following imputation flags are always missing: I_B012H, I_B015H, I_B022H, and I_B025H.

10. Beginning with the 2002 database, the AHA added the following four new hospital service categories for the A503 and Z210 data elements:

    41: Cancer
    42: Heart
    80: Acute Long-Term Care
    90: Children's Acute Long-Term Care.

11. In the 2002 database, the values of the following data elements are questionable because they are always missing or equal to zero: A313 and A314.

12. In the 2003 database, the values of the following data elements are questionable because they are always missing or equal to zero: A313, A314, I_E100H, I_E200H.

13. In the 2003 database, I_B012H, I_B015H, I_B022H, and I_B025H each only have one non-missing value.

14. In the 2003 database, the values of the following data elements are always missing or equal to one: A504, E126B, E126C, E126D, and E126F. It appears that zero values are missing for these data elements.

UNIFORM SAS RESEARCH DATABASE CONSTRUCTION

HCUP database developers reprocess the AHA Annual Survey Databases into SAS research databases with uniform data element names and coding across years. The HCUP databases contain the AHA Annual Survey data elements and additional hospital characteristics indicators calculated from the AHA data. The uniform data element SAS names and labels adhere to the conventions described below.

Data Element Naming Conventions

The SAS data element names consist of three elements: a prefix, a three-digit numeric root, and a suffix.

<u>Prefixes</u>
Survey data contain nine basic categories of hospital descriptive data elements defined by the data element name prefix, as follows:

A: Affiliations and associations
B: Beds and utilization
E: Employment and personnel

F: Financial data elements
M: Medical staff data elements
N: Health care reform orientation data elements
S: Service available flags, where:

0 = service not available or provided by another hospital, and
1 = hospital-based service available

V: Outpatient services
Z: Core data elements.

In addition, in each survey year there are hospitals that either do not respond to the survey or do not supply data for some questions. To document these occurrences, the HCUP databases include a series of flags for key data elements in the AHA Annual Survey data that define when these data elements are reported by the hospital, estimated by the AHA, or expanded by the AHA. These imputation flag fields are defined by the following prefix:

I_: Imputation flags, where:

0 = Reported by hospital
1 = Estimated by the AHA
2 = Expanded by the AHA.

Finally, service data elements written to the HCUP files are originally read from the AHA Annual Survey data as raw service available flags defined by the following prefix:

X_: Raw service available flags, where

◊ Survey years through 1993 are coded:

1 = Hospital-based service available
2 = Service provided by another hospital or provider
3 = Service not available.

◊ Survey year 1994 is coded:

1 = Provider
2 = Not provider.

◊ Survey years beginning in 1995 are coded:

1 = Provider
0 = Not provider.

Numeric Roots

Each HCUP AHA SAS data element name also contains a numeric root in the last numeric column of the name. Some numeric roots further describe data elements as follows:

0: Describes a total figure
8: Describes miscellaneous data elements
9: Describes "other" data elements.

For example, the data element B200 ends with "0" and refers to a total count for "inpatient days, all hospital units."

Numeric roots of 9 generally describe "other" data elements that are part of a total. For example, B259 is the data element for "inpatient days, other subacute care."

Numeric roots of 8 are often used with miscellaneous data elements that don't have a clear affiliation. For example, B238 represents "inpatient days, other special care units."

Whenever possible, data element roots were kept consistent within and across classes of data elements. For example, the last digits are the same for:

E163 represents "full-time employment, physical therapists," and

E263 represents "part-time employment, physical therapists."

Similarly:

B114 represents "beds setup, OB unit beds," while

S014 represents "service availability flag, OB unit."

Suffixes

Each data element name suffix further describes the data element:

A-G, J:	These data elements are the components of an overall total, or the components of a sum or union of binary data elements (data elements with G suffixes are reported beginning in 1992; data elements with J suffixes are reported beginning in 1997).
H:	Defines a hospital data element.
K:	Defines service provided by "network (locally)" [reported beginning in 1994].
N:	Defines a nursing home data element
S:	Defines service provided by "health system (locally)" [reported beginning in 1994].
V:	Defines service provided by "joint venture (locally)" [reported beginning in 1994].

Other Data Elements

Note that the following data elements in the HCUP AHA Annual Survey SAS Databases are exceptions to the above naming conventions:

AHAYEAR
HOSPST
HOSPSTCO
HOSP_BEDSIZE
HOSP_LOCTEACH
HOSP_LOCATION
HOSP_CONTROL
HOSP_REGION
HOSP_TEACH
IDNUMBER
LOCTEACH
ST_BEDSZ
ST_OWNER
ST_TEACH
SYSTEM.

AHAYEAR, HOSPST, and HOSPSTCO are HCUP linkage data elements used to link to other HCUP databases. HOSP_BEDSIZE, HOSP_LOCTEACH, HOSP_LOCATION, HOSP_CONTROL, HOSP_REGION, HOSP_TEACH, ST_BEDSZ, ST_OWNER, ST_TEACH, and LOCTEACH are HCUP hospital stratification data elements, derived from AHA survey data, used for NIS sampling and weighting. IDNUMBER and SYSTEM are hospital and system identification numbers used in different data files, and therefore do not follow normal naming conventions. IDNUMBER is a modified AHA hospital identifier, without the leading six.

Data Element Labeling Conventions

The data element SAS labels may include any of the following suffixes:

-T: Total facility
-H: Hospital
-N: Nursing home
-@: Binary (yes/no) data element
! Data element is inconsistent over time (at any point from 1987 to 2003).

The "!" label suffix identifies data elements that have experienced a change in definition during the 1987-2003 period. Such changes usually occur from survey question or definition changes in a given year.

HCUP AHA ANNUAL SURVEY DATABASE DOCUMETATION

The AHA Annual Survey Database documentation on the HCUP Website describes the construction and content of the HCUP AHA databases, providing researchers with a comprehensive reference tool and additional insights into the file conversion process. This will help researchers understand and utilize the HCUP AHA databases. The documentation includes the following sections, described in detail below:

I. GENERAL AHA INFORMATION

- Overview of the HCUP American Hospital Association Annual Survey SAS® Databases (this document)

II. SPECIAL REPORTS

- AHA Survey Changes
- AHA Hospital Changes

III. DESCRIPTION OF DATA ELEMENTS

- AHA Annual Survey SAS Database Construction Index
- AHA Annual Survey Description of Data Elements
- AHA Annual Survey File Layout
- Annotated AHA Annual Survey Instrument
- 1970-2003 HCUP AHA SAS Databases Data Element Availability

IV. SUMMARY STATISTICS

- Summary Statistics (Proc Contents, Means, and Frequencies)

V. PROCESSING PROGRAMS

- HCUP AHA Annual Survey Database: SAS Load Program

- HCUP AHA Annual Survey Database: Summary Statistics Program

VI. SAS DATABASE INFORMATION

- HCUP AHA SAS Database Information.

Special Reports

The year-specific special reports describe the construction of the AHA files, the changes from year to year, and the source data files.

AHA Survey Changes
This document describes AHA Annual Survey Database changes, such as new and revised items, renamed items, and reordered items.

AHA Hospital Changes
This section displays facility changes since the previous year, including openings, closings, mergers, and splits.

Description of Data Elements

The documents available in the Description of Data Elements section detail the coding of data elements in the AHA files and the availability across years.

HCUP AHA Annual Survey Database Construction Index
The HCUP AHA Annual Survey Database Construction Index cross-references HCUP SAS data element names with the AHA Annual Survey item numbers, item names, and survey page numbers.

For data elements collected on the survey instrument, the index includes a page number reference. For data elements created or substantively changed during HCUP processing, the index includes a reference to the section of the HCUP SAS program that creates or modifies the data element. The program is divided into sections that process specific types of data elements, and the sections are listed in outline format at the beginning of the program. (Refer to the Processing Programs section.) The index also includes SAS data element labels for the HCUP files.

Because the AHA survey instrument changes from year to year, with some data elements added or deleted in any given year, the Database Construction Index is year-specific.

Description of Data Elements for AHA
This table, provided by the AHA, describes the data elements in the AHA Annual Survey Database. It includes descriptions for control codes, service codes, AHA state codes, bed size codes, MSA size codes, city rank codes, and Health System Cluster codes.

AHA Annual Survey Database File Layout
This spreadsheet contains the original AHA Annual Survey Database Public File Layout and Code Descriptions for reference.

Annotated AHA Annual Survey Instrument
This document is an annotated copy of the survey instrument used by the AHA for the Annual Survey. The SAS data element names assigned to the survey responses during HCUP processing are noted beside all survey items that were used for the HCUP files. This offers researchers a cross reference between survey questions and the corresponding SAS data element names.

1970-2003 HCUP AHA SAS Databases Data Element Availability
This table lists all data elements available in any HCUP AHA Annual Survey SAS Database since 1970 and indicates years of availability.

Summary Statistics

This section presents descriptive statistics for all data elements in the HCUP AHA Annual Survey SAS Database. Means, minimum, and maximum values are reported for numeric data elements. Frequency distributions are reported for selected categorical data elements. These statistics are used as initial edit-checks to verify the correct processing of AHA data. They are included on the Website as basic information about AHA data elements.

Processing Programs

This section contains source listings of the programs used to create and document the HCUP AHA Annual Survey SAS Database.

AHA SAS Load Program
This program creates the HCUP AHA Annual Survey SAS Database. It follows an outline format. Part I processes the data into a uniform SAS database. It includes length statements, input statements, creation of new data elements that do not normally change over time, data element changes for given years which generally remain unchanged thereafter, and data element changes that are essentially one-time fixes. Part II generates descriptive statistics used to perform statistical quality review of the data.

The outline structure of the SAS load program is:

I. SAS Database Creation

A. Labels for data elements.

B. Length statements.

C. Input statements.

D. New data elements that do not normally change from year to year:

1. Facility & service data elements.
2. Bed & finance data elements.
3. Personnel & payroll data elements.
4. Medical staff data elements.
5. General hospital-descriptive data elements.
6. Stratification data elements for HCUP sampling.
7. Core linkage data elements.

E. Data element changes for given years which generally remain unchanged thereafter.

F. One-time annual fixes and comments:

1. One-time annual fixes.
2. One-time fixes from edit checks.
3. Notes regarding data elements that were added or dropped.

II. Descriptive Statistics

A. Frequencies of values of state code data elements.

B. Comparison with previous year's survey.

<u>AHA Summary Statistics Program</u>
This SAS program produces the AHA Summary Statistics report. It also adheres to an outline format. It includes a contents listing, means of numeric data elements, and frequencies of values of selected data elements.

The outline structure of the Summary Statistics program consists of:

I. Summary Statistics

A. Proc Contents

B. Means of numeric data elements

C. Frequencies of values of selected data elements.

CONCLUSION

The AHA Annual Survey Databases include data elements containing demographic, utilization, financial, and other hospital characteristics of interest to health care researchers. This overview has described the contents, construction, and documentation of the HCUP AHA Annual Survey SAS Databases with the goal of helping researchers understand and utilize the HCUP AHA databases and documentation.