USER GUIDE: HCUP SUPPLEMENTAL VARIABLES FOR REVISIT ANALYSES

Updated: October 2024

Table of Contents

INTRODUCTION

The Agency for Healthcare Research and Quality (AHRQ) Healthcare Cost and Utilization Project (HCUP) databases capture information on inpatient, emergency department (ED) and ambulatory surgery and other outpatient service encounters in U.S. community hospitals. These databases are often characterized as being "discharge-level" files, meaning that each record in a database represents one discharge abstract from a hospital setting, which can be an inpatient stay, ED visit, or ambulatory surgery or other outpatient service encounter. Thus, if the same individual visited the hospital multiple times in a given year, the HCUP databases would include separate records in the respective HCUP database for each inpatient stay, ED visit or ambulatory surgery or other outpatient service encounter. Many times researchers may be interested in knowing how many visits a distinct patient had rather than simply the number of overall hospital visits. Studying multiple visits is becoming increasingly common as hospital readmission rates are important indicators of the quality of medical care. To facilitate analyses that focus on multiple hospital stays by the same person, AHRQ created a set of supplemental variables that can be linked to the HCUP State databases to track multiple (repeat) patient visits in the hospital setting while adhering to strict privacy regulations. This user guide discusses the methodology used to develop these supplemental variables and how the information can be used with the HCUP State databases. Appendices provide detail on which States, databases, and years are available, in addition to verification statistics. Several SAS coding excerpts are also provided to facilitate the use of these files.

From data year 2003 to 2008, supplemental files called "HCUP Supplemental Files for Revisit Analyses" (herein referred to as the "Revisit Files") were created to include the revisit variables designed to augment the HCUP State databases. Beginning with 2009 data, the variables previously included in the Revisit Files are included in the Core file of the HCUP State databases, when possible. The supplemental variables were labeled "revisit" rather than "readmission" intentionally. The term "revisits" implies multiple health care encounters for a particular patient that are not limited solely to inpatient stays. Most health care research on hospital *readmissions* has focused solely on the inpatient setting – i.e., tracking multiple hospital admissions in the inpatient setting by the same person. The HCUP revisit variables expand on traditional readmission analyses by allowing researchers to study multiple patient visits to the hospital, regardless of the setting of care. In other words, these supplemental variables enable more than identification of hospital readmissions – they also enable tracking of patients admitted to the hospital following an ambulatory surgery or an ED visit and patients who made multiple trips to the ED. Note that revisits may occur for any reason (i.e., they may not be related) and can be separated by days or years. The determining factor in classifying health care events as revisits is that they represent services for the same individual.

In contrast, readmissions are sequential hospital admissions for a related reason, and usually within a specified time frame. Studying readmissions can be difficult as researchers must understand whether patients are admitted for expected follow-up treatment, or, conversely, for unexpected complications. In addition, multiple hospital visits for the same patient may, in fact, be unrelated - and therefore, not considered a "readmission." Identifying readmissions requires specific criteria for the inclusion of events, such as type of condition and appropriate elapsed time. For example, a study of readmissions for congestive heart failure (CHF) may require that the principal reason for the hospitalization, ED visit, or ambulatory surgery is related to CHF and may also require that the time elapsed between events is no longer than a predetermined number of days. The HCUP revisit variables contain key information, such as the days between

multiple visits, that can assist analysts in making informed decisions about whether repeat patient visits qualify as readmissions.

The HCUP revisit variables, used in combination with HCUP State databases, enable analysts to link hospital visits that belong to a unique person, determine the elapsed time between visits, and evaluate valuable clinical information on the HCUP record. These revisit variables afford analysts the flexibility of performing patient-level analyses within and across hospital settings and time periods, without compromising patient confidentiality. Finally, these data elements allow the analyst to determine their definition of a readmission or revisit for the purposes of their study. This User Guide documents the creation of the revisit variables and provides guidance on how to best utilize them in revisit analyses.

HCUP Databases

HCUP develops and maintains a family of health care databases, related software tools, products, and support services. HCUP features the largest collection of multi-year hospital care data in the United States, containing a wealth of all-payer, encounter-level information beginning in 1988. AHRQ relies on vital partnerships among Federal, State, and Industry associations to produce HCUP resources. HCUP databases integrate the data collected by state governments, hospital associations, private data organizations, and the Federal government to create a national health care information resource of inpatient, ED, and ambulatory surgery and other outpatient services data.

The HCUP revisit variables are designed to be used exclusively with the HCUP State databases:

- The State Inpatient Databases (SID), which contain inpatient discharge records from community hospitals in participating States
- The State Emergency Department Databases (SEDD), which contain ED visit records from hospital-owned EDs in participating States
- The State Ambulatory Surgery and Services Databases (SASD), which include data from ambulatory surgery and other outpatient service encounters from hospital-owned and sometimes non-hospital-owned ambulatory facilities.

The revisit variables are unique within state and data year. Prior to 2009 data, users will need to merge the data elements on the HCUP Revisit Files to the corresponding SID, SASD, or SEDD for any analysis (further described in the section on Using HCUP Revisit Variables). Starting with the 2009 data, the revisit variables are on the HCUP Core file.

Note: HCUP revisit variables are available for some, but not all, State databases – SID, SASD, and SEDD – starting in calendar year 2003. Appendix A lists the availability of the revisit variables by State, database, and year.

It should be noted that revisit variables are found on one HCUP Nationwide database, the Nationwide Readmissions Database (NRD), which is sampled from SID with revisit variables. However, the NRD includes further re-identified versions of the revisit variables and therefore, records are unable to be linked back to the SID. The NRD can be used to generate national estimates of readmissions within a single calendar year. Each year of the NRD must be considered as a separate sample as neither patients nor hospitals can be tracked across data years.

HCUP REVISIT VARIABLES

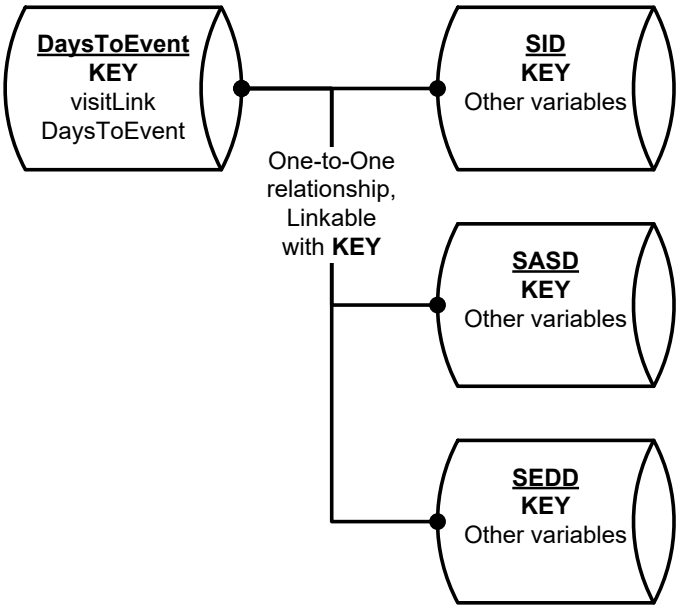
The HCUP revisit variables include only two data elements:

- **visitLink**: linkage variable for all events associated with a unique patient that is assigned during construction of the supplemental revisit variables and based on a unique combination of synthetic patient linkage number, date of birth, and sex
- **DaysToEvent**: the number of days from a randomly chosen "start date" to the admission date for a specific healthcare visit for an individual. The start date is randomly assigned for each unique patient. As a result, **DaysToEvent** will be consistently calculated for all of a patient's linked events, regardless of year (i.e., all visits with the same value of **visitLink**). The **DaysToEvent** variable is assigned during construction of the supplemental revisit variables.

From 2003-2008 the HCUP revisit variables are stored in separate State- and year-specific files (called Revisit Files) that can be linked to the corresponding SID, SASD, and SEDD for that year (Figure 1). For example, for Nebraska, there is one HCUP revisit linkable file for the 2006 data year. Researchers can add the **visitLink** and **DaysToEvent** data elements to the 2006 Nebraska SID, SASD, and/or SEDD files by linking on the **KEY** data element. The Revisit Files for data years 2003-2008 are available through the HCUP Central Distributor.

Starting in data year 2009, the revisit variables (**visitLink** and **DaysToEvent)** are stored in the Core file of the SID, SASD, and/or SEDD, so there are no separate Revisit Files.

Figure 1. Relationship Between HCUP Databases



DEVELOPMENT OF THE HCUP REVISIT VARIABLES

Development of the HCUP revisit variables requires the HCUP State databases files to contain a unique synthetic patient linkage number which enables tracking of unique patients within and across years. Only some of the HCUP statewide data organizations provide this information to HCUP. Each State employs a distinct methodology in producing their synthetic patient linkage number.

Verified Patient Linkage Number (visitLink)

An instrumental part of constructing the HCUP revisit variables is verifying that the synthetic patient linkage numbers accurately represent a unique person in the HCUP State databases. As part of the verification process, the patient's date of birth and sex are used to qualify the synthetic patient linkage number and uniquely identify a person. A new verified patient linkage number (**visitLink**) is assigned for each unique combination of the qualifying information (synthetic patient linkage number, date of birth, and sex). Consider the following example: Five records have the same synthetic patient linkage number, but two records have one date of birth and sex, and the remaining three records have a different, but consistent, date of birth and sex. The two records with identical identifying information have one value of **visitLink,** and the other three records have a different value of **visitLink**. Appendix B contains examples of the assignment of **visitLink** for different scenarios.

No verified patient linkage number is assigned if any of the three pieces of information is missing (i.e., **visitLink** is missing). Additionally, no verified patient linkage number is assigned if there are more than 40 hospital visits in a given calendar year with the same qualifying information. This second qualification excludes less than 0.5 percent of the synthetic patient linkage numbers and aims to eliminate synthetic patient linkage numbers used for multiple people. Appendix B contains examples of the assignment of **visitLink** for different scenarios.

While the term "verified patient linkage number" is used to describe the information in the HCUP data element **visitLink**, the values are not recognizable as specific patient information. **VisitLink** does not include the values of the synthetic patient linkage number, date of birth, or sex.

The **visitLink** variable is created each year for all discharges across all available databases for the particular State. For example, if a State provides SID, SASD, as well as SEDD, the **visitLink** is verified for all records across all three databases for that data year and previous years of data, as appropriate.

Calculating the Days to Event (DaysToEvent)

For a verified patient linkage number (i.e., non-missing **visitLink**) with more than one hospital visit, the elapsed days between visits are calculated as the difference between the two visit dates. This information is often useful for determining readmissions for a specific condition (i.e., 30-day readmissions, 7-day readmissions, etc.) While this information is critical for defining readmissions, the use of admission and discharge dates is highly restricted per Health Insurance Portability and Accountability Act (HIPAA) guidelines.

To comply with HIPAA guidelines and ensure patient confidentiality, no "date" information is released on the HCUP revisit variables. A timing variable (**DaysToEvent)** was calculated consistently for each verified patient linkage number (**visitLink**) based on a randomly assigned "start date." Each verified patient linkage number is assigned a unique start date that is used to calculate **DaysToEvent** for all visits associated with that **visitLink** value. The variable **DaysToEvent** is the difference between the visit's admission date and the start date associated with the **visitLink**.

The calculation of days between visits is the difference of **DaysToEvent** between two selected visits for a unique verified patient linkage number (**visitLink**). For example, consider a patient with CHF that has a hospital admission on 1/10/2008 and an ED visit on 1/25/2008. If the

DaysToEvent value is “9” for the 1/10/2008 admission and the **DaysToEvent** value is “24” for the 1/25/2008 ED visit, then the number of days between the start of the first visit and the start of the second visit is 15 days (24 – 9 = 15). It should be noted that often readmission analyses consider the time between the end of one admission and the start on the next admission. To adjust for the length of the admission, subtract the length of stay from the difference. In the example, above, if the first admission had a length of stay of 2 days then the number of days between the end of the first visit and the start of the second visit is 13 days (24 – 9 – 2 = 13).

The lowest value of **DaysToEvent** will be on the first or earliest event for a patient. It is important to remember that if patient A has a value of 605 for **DaysToEvent** and patient B has a value of 300 for **DaysToEvent**, patient B’s event did not necessarily take place prior to patient A’s event – in fact, patient B's **DaysToEvent** value has no relation to patient A's **DaysToEvent** value. Because of the use of a random start date in the calculation of **DaysToEvent**, the value of **DaysToEvent** cannot be compared across patients. Appendix B contains examples of the assignment of **DaysToEvent** for different scenarios.

AVAILABILITY OF VERIFIED PATIENT IDENTIFIERS VARIES BY STATE

The availability of verified patient linkage numbers for specific populations and settings varies by State and should be considered prior to any analysis. Table 1 shows the range of the percentage of verified revisit information across 15 States in 2005-2006 for selected patient characteristics, expected payer, and hospital characteristics.

Table 1. Range of Percentage of Records with Verified Revisit Information

Overall		**All Events**			
		Min	**Q1**	**Median**	**Max**
		65.4%	84.7%	90.5%	100.0%
By Patient Characteristics					
Age Group	0	11.0%	26.8%	49.1%	100.0%
	1-17	35.1%	54.1%	73.3%	100.0%
	18-44	69.5%	92.6%	95.9%	100.0%
	45-64	69.2%	96.4%	97.5%	100.0%
	65+	69.8%	97.2%	98.8%	100.0%
Sex	Male	64.3%	82.6%	88.9%	100.0%
	Female	66.2%	86.6%	91.8%	100.0%
Patient Income	Quartile 1 (lowest)	59.3%	87.2%	91.5%	100.0%
	Quartile 2	66.9%	86.7%	91.5%	100.0%
	Quartile 3	70.1%	86.1%	90.9%	100.0%
	Quartile 4 (highest)	70.5%	83.4%	88.9%	100.0%
Expected Payer	Medicare	69.8%	97.2%	99.1%	100.0%
	Medicaid	55.4%	71.7%	85.9%	100.0%
	Private insurance	66.5%	82.2%	89.4%	100.0%
	Self-pay	66.9%	79.8%	88.2%	100.0%
	No Charge	54.2%	79.0%	86.5%	100.0%
	Other	44.0%	87.6%	93.7%	100.0%

Overall		All Events			
		Min	Q1	Median	Max
		65.4%	84.7%	90.5%	100.0%
By Hospital Characteristics					
Hospital Ownership	Government, nonfederal	72.5%	81.7%	90.5%	100.0%
	Private, not-profit	67.9%	84.7%	90.0%	100.0%
	Private, invest-own	57.5%	86.0%	92.5%	100.0%
Hospital Location	Large central metropolitan	78.0%	82.4%	87.8%	100.0%
	Large fringe metropolitan	36.3%	83.1%	89.2%	100.0%
	Medium metropolitan	74.3%	85.4%	92.0%	100.0%
	Small metropolitan	45.6%	87.8%	94.4%	100.0%
	Micropolitan	69.9%	88.7%	95.4%	100.0%
	Noncore	35.1%	83.4%	94.8%	100.0%
Hospital Bed Size	<100	56.7%	87.1%	92.5%	100.0%
	100-299	68.7%	84.3%	91.1%	100.0%
	300-499	70.1%	89.7%	92.2%	100.0%
	500+	72.7%	84.3%	92.5%	100.0%

Source: HCUP State Inpatient Databases, 15 States, 2005-2006

In most cases, verification rates across patient and hospital characteristics and across selected diagnosis and procedure categories were consistent with the overall verification rates. For example, the first and second quartiles of the verification percentage overall were 84.7 percent and 90.5 percent, respectively. The first and second quartiles of the verification percentage for patients from hospitals in large fringe metropolitan areas were 83.1 percent and 89.2 percent, respectively.

Some notable exceptions include:

- Newborns (age 0) – The median of the verification rates across the 15 states was only 49.1 percent.
- Children and adolescents (age 1-17) – The first quartile for verification rates was 54.1 percent and the median was 73.3 percent. A separate analysis examined whether verification rates were better for certain ranges of children, such as adolescents or teens. There was no specific range of pediatric ages between 1 and 17 that were markedly better in terms of the percentage verified patient linkage numbers.
- Expected payer of Medicaid and No Charge – The first quartile for both was less than 80 percent and the median was about 86 percent.

Revisit/readmission analyses for pediatric conditions and certain payers may only be appropriate in selected States.

SELECTING STATES FOR A REVISIT/READMISSION ANALYSIS

When selecting which States to use for a revisit analysis, please reference the following resources:

- Appendix A provides the list of all States, years, and databases with HCUP revisit variables.

- Appendix C lists States that have inconsistent coding across data years of the source synthetic patient linkage numbers in the respective State database provided by the HCUP Partners and should not be used for analyses that span certain years.
- Appendix D provides information on the consistency of **visitLink** in the SID and the SASD and SEDD within a data year.
- Appendix E provides verification rates by State and year that should be used to determine which HCUP States are best for specific types of revisit or readmission analysis.

USING THE HCUP REVISIT VARIABLES

Using the HCUP revisit variables involves four basic steps

1. For a given State and year, merge the HCUP Revisit File with the corresponding SID, SASD, or SEDD by the data element **KEY** to add the revisit data elements **visitLink** and **DaysToEvent**. This step is only needed for data years 2003-2008. Beginning in data year 2009, the data elements **visitLink** and **DaysToEvent** are included on the Core file of the SID, SASD, and SEDD, when possible.
2. Select patients of interest.
3. Use **visitLink** to identify all events for a patient. The same unique value of **visitLink** is coded on all records for an individual patient. Records with missing values for the **visitLink** variable will be a mixture of patients with unknown revisit information. It may be appropriate to exclude these records from the analysis.
4. Use **DaysToEvent** to sequentially order the visits for a patient and to calculate the time between two visits for a patient. If the **DaysToEvent** is 5 on one event and 35 on another, the time between the *start* of each event is 30 (35-5) because **DaysToEvent** is based on the admission date. If you want to consider the time between the end of the first event and the start of the second event, the length of stay for the first event needs to be subtracted. If the length of stay on the first event is two, then the number of days between is 28 (35-5-2 = 28)

Usage Examples

Use of the HCUP revisit variables is relatively straightforward. Below are three examples of applying these variables to research topics.

Usage Example #1: Assigning Patient Characteristics

Researchers may want to group patients by specific patient characteristics, such as a patient's age or insurance status. When a patient's health care experience includes more than one hospital event, categorizing the patient may be problematic. This difficulty arises because some patient characteristics may change over time. To assign attributes based on when a person began receiving services, consider the following steps:

- Data should first be grouped by patient, in service date order (specifically, the HCUP state-level data file, merged with the HCUP Revisit File if prior to 2009, and then sorted by **visitLink** and **DaysToEvent**).

- All records for a patient are then sequentially examined in order to select the first valid, non-missing value for each patient characteristic (age, sex, race, income quartile, location, expected payer).
- The selected attribute(s) are then applied to all events for the patient.

A SAS coding example of how attributes can be assigned is shown in Appendix F.

Usage Example #2: Revisits for Selected Patients

This example counts the number of related events for selected patients with a specific diagnosis and calculates a number of statistics, including days between the initial event and the first subsequent event, by setting. This example focuses on revisits for diabetes but can easily be adapted to any diagnoses.

The example looks for a "clean period," measured in months, with no hospital events for an individual patient for the specified condition. Use of a "clean period" for counting readmissions is optional. Sometimes when identifying an episode of care, rather than straight utilization, a period of time during which the patient has not been admitted or treated is required. The first event after the "clean period" is considered the index event. Any event in a predetermined period of time after the index event becomes part of the "episode."

For illustrative purposes, we selected "diabetes mellitus with complications" (HCUP Clinical Classification Software (CCS) for ICD-9-CM diagnosis category 50) as the condition and required a clean period of 6 months. The steps are:

1. Combine the event and revisit data
    a. Limit data to linkable patients (a non-missing **visitLink** available)
    b. Select all events with the specified condition (diabetes)
2. Sort the combined events into patient (**visitLink**) and service sequence (**DaysToEvent**) order
3. Find individuals with two or more events
    a. Find patients with a clean period before their first diabetes event
        i. On the first event for a patient, the service must be after clean period, defined as the first six months of the data year
        ii. If the first event was prior to the clean period month, look for a clean period on subsequent events by testing the number of days between the current event and the preceding event
    b. When a clean period is identified
        i. Count the number of events after the clean period
        ii. Determine the settings of the first and second events, and calculate the number of days between the first and second event
    c. Summarize the processing counts
4. Calculate statistics (distribution) for the number of patient events
5. Summarize revisits by the initial and second service settings.

SAS programming code for this example is found in Appendix E.

Usage Example #3: Preceding Visits to any Hospital Setting for Selected Patients

This example identifies patients' hospital events that precede CABG surgery (CCS for ICD-9-CM procedure category 44), regardless of service setting, and summarizes counts by principal diagnosis. The steps are:

1. Combine the event and revisit data
    a. Limit data to linkable patients (a non-missing **visitLink** available)
    b. Identify events with the specified procedure (CABG)
2. Sort the combined events into patient (**visitLink**) and service sequence (**DaysToEvent**) order
3. For patients who received CABG surgery, select all events prior to the surgery
4. Summarize prior events by principal diagnosis and setting.

SAS programming code for this example is found in Appendix E.

Cautionary Note: Transfers and Possible Duplicates

The HCUP revisit variables allow an analyst to identify which records in the SID, SASD, and SEDD belong to the same person, as well as the time between events for that person. An analyst still must decide how to handle the following two types of scenarios:

- Transfers – when a patient is transferred from one acute care hospital to another
- Duplicates – when a record for the same event occurs twice in the HCUP file.

In the SID, there will be two different records if a patient is transferred from one hospital to another. The following can be used to identify the two SID records:

- Same person (**visitLink** is the same on two records)
- Disposition indicating transferred out (**DISPuniform** = 2)
- Admission source indicating transfer in (**ASOURCE** = 2)
- Discharge date of one record is the same as the admission date of another (**DaysToEvent** plus the length of stay of the first record equals the **DaysToEvent** of the second record)
- Different hospital (**DSHOSPID** is different).

Analysts conducting patient-level analyses need to decide how best to use the above information to identify transfers. The coding of admission source and discharge disposition is not always consistent with the timing of events identified by **DaysToEvent** (i.e., **DaysToEvent** may identify two records as two parts of a transfer, but either the disposition or admission source is not coded as such). Table 2 demonstrates the range in the percentage of discharges identified as transfers using different schemes.

Table 2. Range of Percentage of Inpatient Discharges Identified as Transfers

Scheme to identify transfers	Minimum Value Across 15 States	Maximum Value Across 15 States
Percentage of records identified as transfers using one source of information:		
Dates	1.80%	6.41%
Discharged as a transfer to another acute care hospital (DISPuniform=2)	1.36%	3.11%
Admitted as a transfer from an acute care hospital (ASOURCE=2)	0.85%	5.30%
Percentage of records identified as transfers using two sources of information:		
Dates and DISPuniform=2	0.70%	2.07%
Percentage of records identified as transfers using all three sources of information:		
Dates, DISPuniform=2, ASOURCE=2	0.22%	1.50%

Source: HCUP State Inpatient Databases, 15 States, 2006
Note: The Uniform Billing UB-04 Specifications changed coding specifications for Source of Admission to Point of Origin for Admission or Visit starting October 1, 2007.

For some analyses it may be best to combine the two records from a transfer into one by summing the lengths of stay and total charges and combined diagnoses and procedures.

The HCUP SID, SASD, and SEDD occasionally have multiple records for the same person (**visitLink**) with the same **DaysToEvent** and length of stay (**LOS**). These duplicate records may or may not have the similar charge and diagnostic information. Analysts should decide how best to handle such records.

In addition, HCUP made an explicit decision to duplicate records across the SEDD and SASD when a record indicated that the patient received services in both settings. These duplicate records will have the same value for the data element KEY. In this case, the analyst will need to decide how to include and account for these cases. The effect of these duplicated records varies by state from less than 1% in California SASD to about 15% in Tennessee SASD.

Additional considerations for using the HCUP revisit variables for analysis can be found in the HCUP Method Series Report #2011-01, *Methodological Issues When Studying Readmission and Revisits Using Hospital Administrative Data.*

APPENDIX A: AVAILABLE HCUP SUPPLEMENTAL VARIABLES FOR REVISIT ANALYSES

HCUP revisit variables are available for the States, years, and databases listed in Table A of the Excel Appendix file. For data years 2003-2008, the supplemental Revisit Files must be linked using the KEY data element to the corresponding HCUP SID, SASD, or SEDD for any analysis. Starting with 2009 data, the revisit variables are included in the Core file, when available.

Information on some HCUP State databases is to be determined (TBD) after HCUP data processing.

See Excel Appendix for Table A. HCUP Revisit Variables by State and Data Type.

APPENDIX B: EXAMPLE OF ASSIGNMENT OF VISITLINK AND DAYSTOEVENT

The following table lists examples of the assignment of **visitLink** and **DaysToEvent** in different scenarios.

No.	Partner-Provided Synthetic Patient Linkage Number	Date of Birth	Sex	Example of visitLink	Randomly Assigned Base Date for visitLink	Admission Date	DaysToEvent	Notes on visitLink Assignment
1	A	15-Jan-60	M	11111	1/1/1980	1/5/2013	12058	Same patient ID, date of birth, and sex for observations 1 and 2, therefore same VisitLink
2	A	15-Jan-60	M	11111	1/1/1980	3/5/2013	12117	Same patient ID, date of birth, and sex for observations 1 and 2, therefore same VisitLink
3	A	15-Jan-60	F	11112	6/15/1953	11/17/2013	22070	Same patient ID and date of birth as observations 1 and 2, but different sex, therefore different VisitLink
4	B	1-May-40	F	11113	11/11/2011	3/3/2013	478	Same patient ID, date of birth, sex for observations 4 through 6, therefore same VisitLink
5	B	1-May-40	F	11113	11/11/2011	5/5/2013	541	Same patient ID, date of birth, sex for observations 4 through 6, therefore same VisitLink
6	B	1-May-40	F	11113	11/11/2011	11/12/2013	732	Same patient ID, date of birth, sex for observations 4 through 6, therefore same VisitLink
7	B	15-Jun-45	F	11114	5/23/2000	6/1/2013	4757	Same patient ID and sex as observations 4 through 6, but different date of birth, therefore different VisitLink
8	B	15-Jun-45	F	11114	5/23/2000	6/23/2013	4779	Same patient ID and sex as observations 4 through 6, but different date of birth, therefore different VisitLink
9	B	15-Jun-45	F	11114	5/23/2000	7/30/2013	4816	Same patient ID and sex as observations 4 through 6, but different date of birth, therefore different VisitLink
10	C	1-Dec-80	M	11115	12/1/1940	2/3/2013	26362	Same patient ID, date of birth, sex for observations 10 and 11, therefore same VisitLink
11	C	1-Dec-80	M	11115	12/1/1940	6/15/2013	26494	Same patient ID, date of birth, sex for observations 10 and 11, therefore same VisitLink
12	C	Not Available	M	Missing	Not assigned because no VisitLink	8/4/2013	Not assigned because no VisitLink	Same patient ID and sex as observations 10 and 11, but missing date of birth, therefore VisitLink is missing

APPENDIX C: CONSISTENCY OF SYNTHETIC PATIENT LINKAGE NUMBERS ACROSS CONSECUTIVE YEARS

The HCUP data element **visitLink** is derived from synthetic patient linkage number provided by the HCUP Partner. Partners sometimes change their coding scheme between data years, which in turn causes a discontinuity in **visitLink**. Table C of the Excel Appendix file lists the percentage of unique values of **visitLink** that appear in consecutive data years of the SID, SEDD, or SASD. If the percentage is low or different than other pairs of years, it is a good indication that the **visitLink** cannot be used to track patients across those data years. A dash indicates that **visitLink** is not available in one or both years.

To better understand how to interpret the tables below, consider the following examples from the SID.

- In Arkansas, 22 percent of the **visitLink** values in 2004 also appeared in 2005. This is a good indication that **visitLink** can be used to track AR patients between 2004 and 2005. In addition, the percentage of overlap in **visitLink** is 22–24 percent in all pairs of data years in AR from 2004 through 2019. This is a good indication that **visitLink** can be used to track AR patients from 2004 through 2019.
- In Georgia, zero percent of the **visitLink** values in 2011 also appeared in 2012. This indicates that the values of **visitLink** never overlap between the two data years, and **visitLink** should *not* be used to track GA patients from 2011 into 2012.
- In Washington, 100 percent of the **visitLink** values in 2011 also appeared in 2012. The exact same values of **visitLink** were used for different people in these two data years, and **visitLink** should *not* be used to track WA patients from 2011 into 2012.

Information on some HCUP databases is to be determined (TBD) after HCUP data processing.

See Excel Appendix for Table C-1. Percentage of visitLink Values Reported in Consecutive Data Years, State Inpatient Databases; Table C-2. Percentage of visitLink Values Reported in Consecutive Data Years, State Emergency Department Databases; and Table C-3. Percentage of visitLink Values Reported in Consecutive Data Years, State Ambulatory Surgery and Services Databases.

APPENDIX D: CONSISTENCY OF SYNTHETIC PATIENT LINKAGE NUMBERS BETWEEN THE SID AND SASD/SEDD WITHIN A DATA YEAR

The year-specific sections in Table D of the Excel Appendix file detail the percentage of verified patient linkage numbers (**visitLink**) in the SID that overlap with the SASD and SEDD. The tables allow the analyst to determine the best possible States for a revisit analysis intended to track sequential visits for patients across settings of care. Across States and available data years beginning data year 2016, the average percent overlap between the SID and SEDD is 43 percent whereas between the SID and SASD, the average percent overlap is 33 percent. The average decreases to 17 percent when examining the percent overlap in the SID with both the SASD and SEDD.

To better understand how to interpret Table D, consider the following example. In Vermont, 46 percent of **visitLink** values in the 2016 SID also appeared in the 2016 SEDD. This is a good indication that **visitLink** can be used to track patients across these settings of care. The percent overlap of **visitLink** values between the 2016 SID and 2016 SASD is 87 percent. Relative to the average, this is a much higher percentage and a possible indication that **visitLink** values may be used for different people between the two settings of care. A percentage overlap of 0 would indicate that **visitLink** values are reassigned with each setting of care and therefore, patients are not able to be tracked.

Some States either do not provide ED or ambulatory surgery and other outpatient services data to HCUP or do not release a SEDD or SASD through the HCUP Central Distributor in given year and are shown as n/a in the table. Information on some HCUP databases is to be determined (TBD) after HCUP data processing.

See Excel Appendix for Table D. Percentage Overlap of visitLink in the SID With the SEDD and SASD.

APPENDIX E: CONSISTENCY OF VERIFIED REVISIT INFORMATION

The consistency of the verified patient linkage numbers is evaluated when the HCUP revisit variables are created for a State. The year-specific sections in Table E of the Excel Appendix file detail the number of total records in the SID, SEDD, or SASD and the percentage of records with a verified patient linkage number (**visitLink**).

The tables allow the analyst to determine the best possible States for a revisit analysis. Researchers should use the HCUP revisit variables with caution when looking at revisits for specific patient populations that have a low percentage of verified patient linkage numbers. If studying pediatric conditions, consider States with a high percentage of verified person identifiers for ages under 18. If a proposed study is specific to other patient or hospital characteristics, generate statistics on the percent verified by the study focus and select States with a high percentage of verified person identifiers.

Information on some HCUP databases is to be determined (TBD) after HCUP data processing.

See Excel Appendix for Table E-1. Consistency of Verified Revisit Information, State Inpatient Databases, Table E-2. Consistency of Verified Revisit Information, State Emergency Department Databases, and Table E-3. Consistency of Verified Revisit Information, State Ambulatory Services and Surgery Databases.

APPENDIX F: SAS CODE FOR USAGE EXAMPLES

Usage Example #1: Assigning Patient Characteristics

```
** Assigning Attributes

** combined event data -- combine all available data types for the years;
** for simplicity, the example uses only 2 attributes: AGE and ZIPINC_QRTL;
data NE_CombinedEvents1;
  keep KEY visitLink DaysToEvent AGE ZIPINC_QRTL
       events   sidEvents   seddEvents   sidSeddEvents   sasdEvents
       verified sidVerified seddVerified sidSeddVerified sasdVerified;
  merge NE_2005_Revisit_Core (in=_inDaysToEvent)
        NE_2006_Revisit_Core (in=_inDaysToEvent)
        NE_SID_2005_Core (in=_inSID_2005)
        NE_SID_2006_Core (in=_inSID_2006)
        NE_SASD_2005_Core (in=_inSASD_2005)
        NE_SASD_2006_Core (in=_inSASD_2006)
        NE_SEDD_2005_Core (in=_inSEDD_2005)
        NE_SEDD_2006_Core (in=_inSEDD_2006)
        end=lastObs;
  by KEY;

  ** indicator for all events;
  events = 1;
  if _inDaysToEvent then verified = 1;
  else                   verified = 0;

  ** indicator for SID events;
  if _inSID then do;
    sidEvents = 1;
    if _inDaysToEvent then sidVerified = 1;
    else                   sidVerified = 0;
  end;

  ** indicator for SEDD events;
  if _inSEDD then do;
    seddEvents = 1;
    if _inDaysToEvent then seddVerified = 1;
    else                   seddVerified = 0;
  end;

  ** indicator for SID/SEDD events;
  sidSeddEvents = max(sidEvents, seddEvents);
  sidSeddVerified = max(sidVerified, seddVerified);

  ** indicator for SASD events;
  if _inSASD then do;
    sasdEvents = 1;
    if _inDaysToEvent then sasdVerified = 1;
    else                   sasdVerified = 0;
  end;
run;


** sort the combined events into link order;
proc sort data=NE_CombinedEvents1;
  by visitLink DaysToEvent;
```

Usage Example #1: Assigning Patient Characteristics (cont'd)

```
** use the first non-missing value for link ID attributes;
data NE_LinkAttribs2;
  keep visitLink foundAGE foundZIP;
  set NE_CombinedEvents1;
  by visitLink DaysToEvent;
  where visitLink;  ** only process events with a visitLink ID;

  ** set attributes to missing for each new verified patient;
  length foundZIP $5;
  retain foundAGE foundZIP;
  if first.visitLink then do;
    foundAGE = .;
    foundZIP = '';
  end;

  ** select the first non-missing attribute;
  if foundAGE le .Z and AGE gt .Z then foundAGE = AGE;
  if foundZIP eq '' and ZIPINC_QRTL ne '' then foundZIP = ZIPINC_QRTL;

  ** create one set of attributes for each verified patient;
  if last.visitLink then output;
run;


** apply the consistent attributes to the event data;
data NE_CombinedEvents2;
  keep KEY visitLink DaysToEvent AGE ZIPINC_QRTL
       events   sidEvents   seddEvents   sidSeddEvents   sasdEvents
       verified sidVerified seddVerified sidSeddVerified sasdVerified;
  merge NE_CombinedEvents1
        NE_LinkAttribs2 (in=_inLinkAttribs)
  by visitLink;

  ** apply uniform attributes to verified patient events;
  if _inLinkAttribs then do;
    AGE = foundAGE;
    ZIPINC_QRTL = foundZIP;
  end;
run;
```

Usage Example #2: Revisits for Selected Patients

```
** Example 2, Follow-up Care
** Program code

** diagnosis CCS code to examine and clean period (months);
%let condX = 50;        ** diabetes mellitus with complications;
%let cleanPeriod = 6;   ** number of months for "clean" period with no condX claims;
%let firstYear = 2005;  ** first year of data;
%let lastYear  = 2006;  ** last year of data;


** (1) combine event  and revisit data -- keep only events with the specified condition;
data NE_EventsCondX1;
  keep KEY visitLink DaysToEvent servSetting YEAR DQTR AMONTH LOS;
  merge NE_2005_Revisit_Core (in=_inDaysToEvent)
        NE_2006_Revisit_Core (in=_inDaysToEvent)
        NE_SID_2005_Core (in=_inSID)
        NE_SID_2006_Core (in=_inSID)
        NE_SASD_2005_Core (in=_inSASD)
        NE_SASD_2006_Core (in=_inSASD)
        NE_SEDD_2005_Core (in=_inSEDD)
        NE_SEDD_2006_Core (in=_inSEDD)
        end=lastObs;
  by KEY;

  ** (1.a) limit data to linkable patients;
  if _inDaysToEvent and (_inSID or _inSASD or _inSEDD);

  ** service type indicator;
  select;
    when (_inSID)  servSetting = 'IP';
    when (_inSEDD) servSetting = 'ED';
    when (_inSASD) servSetting = 'AS';
    otherwise;
  end;

  ** (1.b) grab all events with condX;
  if _n_ eq 1 then put "Searched CCS diagnoses codes for values of '&condX'";
  array DXCCS {*} DXCCS1-DXCCS15;
  do i = 1 to NDX;
    if DXCCS{i} eq &condX then output NE_EventsCondX1;
  end;
run;


** (2) sort the combined events into visitLink (patient), service sequence order;
proc sort data=NE_EventsCondX1;
  by visitLink DaysToEvent;
run;
```

Usage Example #2: Revisits for Selected Patients (cont'd)

```
** (3) find people with 2+ condX related events,
** calculate days between the 1st and 2nd event;
data NE_EventsCondX2;
  keep visitLink patientCount patientEvents servSetting1 servSetting2 days_E1toE2;
  set NE_EventsCondX1
      end=lastObs;
  by visitLink DaysToEvent;

  retain patientCount 1;
  label patientCount  = 'count of patients'
        patientEvents = 'number of events for patient'
        servSetting1  = 'type of service for the initial event'
        servSetting2  = 'type of service for the second event'
        days_E1toE2   = 'days between initial and second event';

  retain _condX_events;
  _condX_events + 1;
  _dayLag = lag(DaysToEvent);

  retain _patients _pat_w2Plus _pat_wCleanPeriod _cleanPeriod days_E1toE2 patientEvents 0
         servSetting1 servSetting2;
  if first.visitLink then do;
    _patients + 1;  ** count patients with condX;

    ** reset patient indicators and counters;
    days_E1toE2 = .;
    patientEvents = .;
    servSetting1 = '  ';
    servSetting2 = '  ';

    if not (first.visitLink and last.visitLink) then _pat_w2Plus + 1;

    ** (3.a.i) first claim must be after "cleanPeriod" month of first data year;
    if (YEAR gt &firstYear) or  /* assumes clean period (months) < 12 */
       ( YEAR eq &firstYear and
         DQTR gt &cleanPeriod/3 and
         AMONTH gt &cleanPeriod and
         LOS lt &cleanPeriod*30 ) then do;
      _cleanPeriod = 1;  ** indicator that patient had clean period;
      servSetting1 = servSetting;
      patientEvents = 1;  ** counter for number of patient events;
    end;  /* end-if (clean period) */
    else _cleanPeriod = 0;
  end;  /* end-if (first visitLink) */

  else do;  /* not first visitLink */
    ** (3.b) revisit - clean period already found;
    if _cleanPeriod then do;
      patientEvents + 1;  ** count number of events for this patient;
      if patientEvents eq 2 then do;
        days_E1toE2 = DaysToEvent - _dayLag;
        servSetting2 = servSetting;
        _pat_wCleanPeriod + 1;  ** count patients with a clean period;
      end;  /* end-if (second event) */
    end;  /* end-if (clean period) */

    else do;  /* clean not (yet) period found */
      ** (3.a.ii) no clean pd yet found - check lag days f/ clean pd between events;
      if _dayLag/30 gt &cleanPeriod then do;  /* 99% correct */
        _cleanPeriod = 1;  ** indicator that patient had clean period;
        servSetting1 = servSetting;
        patientEvents = 1;  ** counter for number of patient events;
      end;  ** end-if (lag clean period);
    end;  /* end-else (clean period not found) */
  end;  /* end-else (not first visitLink) */

  ** output one obs per visitLink;
  if last.visitLink and patientEvents ge 2 then output;
```

Usage Example #2: Revisits for Selected Patients (cont'd)

```
  ** (3.c) summarize processing;
  if lastObs then do;
    put '================================================================================';
    put "Processing summary -- events with diagnosis (CCS) category: &condX";
    put '-started with ' _condX_events 'events (total)';
    put '          for ' _patients 'patients.';
    put '-there were ' _pat_w2Plus "patients with 2+ DXCCS '&condX' events";
    put "-clean period, no claim with DXCCS '&condX' for at least &cleanPeriod months";
    put ' before the first claim: ' _pat_wCleanPeriod ;
    put '================================================================================';
    put /;
  end;
run;


proc format;
  picture pctfmt low-high='009 %';
run;


** (4) statistics for the number of condX related events;
title1 "Number of DXCCS &condX Events - Distribution";
title2 "for Patients with Multiple DXCCS &condX Events";
proc means data=NE_EventsCondX2 maxdec=2 mean p25 p50 p75 max;
  var patientEvents;
run;


** (5) summarize revisits by the initial and second service settings;
title "Count and Days for DXCCS '&condX' Patients with Multiple Hospital Events";
proc tabulate data=NE_EventsCondX2 format=comma12.;
  class servSetting1 servSetting2 /descending;
  var patientCount;
  table (servSetting1 all),
        (servSetting2 all)*
        (patientCount*sum*f=comma12.
         days_E1toE2*(mean median)*f=8.1);
run;
```

Usage Example #2: Revisits for Selected Patients (cont'd)

Output

Number of DXCCS 50 Events - Distribution
for Patients with Multiple DXCCS 50 Events

The MEANS Procedure

Analysis Variable : patientEvents number of events for patient

Mean	25th Pctl	50th Pctl	75th Pctl	Maximum
4.14	2.00	3.00	4.00	97.00

Count and Days for DXCCS '50' Patients with Multiple Hospital Events

	type of service for the second event											
	IP			ED			AS			All		
	count of patients	days between initial and second event		count of patients	days between initial and second event		count of patients	days between initial and second event		count of patients	days between initial and second event	
	Sum	Mean	Median	Sum	Mean	Median	Sum	Mean	Median	Sum	Mean	Median
initial event												
IP	1,850	61.6	6.0	367	102.8	35.0	157	140.3	111.0	2,374	73.2	14.0
ED	254	122.0	75.0	299	93.3	34.0	40	123.0	86.5	593	107.6	52.0
AS	158	151.3	120.0	24	114.0	115.5	466	41.6	14.0	648	71.0	21.0
All	2,262	74.7	16.0	690	99.1	37.0	663	69.9	21.0	3,615	78.5	21.0

Usage Example #3: Preceding Visits to any Hospital Setting for Selected Patients

```
** Example 3, Preceding Events
** Program code

** procedure CCS code to examine;
%let servX = 44;        ** CABG;


** combine event  and revisit data -- flag events with the specified service;
data TN_AllEvents;
  keep KEY visitLink DaysToEvent servX dxCcs1 servSetting events;
  merge TN_2005 _Revisit_Core (in=_inDaysToEvent)
        TN_2006 _Revisit_Core (in=_inDaysToEvent)
        TN_SID_2005_Core (in=_inSID)
        TN_SID_2006_Core (in=_inSID)
        TN_SASD_2005_Core (in=_inSASD)
        TN_SASD_2006_Core (in=_inSASD)
        TN_SEDD_2005_Core (in=_inSEDD)
        TN_SEDD_2006_Core (in=_inSEDD)
        end=lastObs;
  by KEY;
  if _inSID  or _inSASD or _inSEDD;

  ** flag events with the specified service;
  array prCcs {*} prCcs1-prCcs30;
  do i = 1 to dim(prCcs);
    if prCcs{i} eq &servX then do;
      servX = 1;
      leave;
    end;
  end;

  array cptCcs {*} cptCcs1-cptCcs30;
  do i = 1 to dim(cptCcs);
    if cptCcs{i} eq &servX then do;
      servX = 1;
      leave;
    end;
  end;

  ** service type indicator;
  select;
    when (_inSID)  servSetting = 'IP';
    when (_inSEDD) servSetting = 'ED';
    when (_inSASD) servSetting = 'AS';
    otherwise;
  end;

  ** counters;
  retain events 1;
  if _inSID then sidEvents = 1;
  if _inSEDD then seddEvents = 1;
  if _inSASD then sasdEvents = 1;
run;


** sort the combined events into visitLink (patient), service sequence order;
proc sort data=TN_AllEvents out=TN_OrderedEvents (index=(servX));
  by visitLink DaysToEvent;
run;
```

Usage Example #3: Preceding Visits to any Hospital Setting for Selected Patients (cont'd)

```
** find people who received the topic service;
proc sql;
  create table TN_TopicPop as
    select visitLink,
           min(DaysToEvent) as firstService
    from TN_OrderedEvents
    where servX eq 1
    group by visitLink;
quit;


** for people with the service, select all events prior to the service;
data TN_PriorEvents;
  merge TN_OrderedEvents (in=_inEventData)
        TN_TopicPop (in=_inTopicPop);
  by visitLink;
  if _inTopicPop;
run;


** summarize prior events by primary diagnosis and setting;
title 'Count of patients by Primary Diagnosis and Service Setting';
proc tabulate data=TN_PriorEvents format=comma12.;
  class dxCcs1 servSetting;
  format DXCCS1 FDCCSPDX.;
  var events;
  table (all DXCCS1),
        (all servSetting)*(events*sum);
run;
```

Usage Example #3: Preceding Visits to any Hospital Setting for Selected Patients (cont'd)

Output

Count of patients by Primary Diagnosis and Service Setting

		servSetting		
	All	AS	ED	IP
	events	events	events	events
	Sum	Sum	Sum	Sum
All	613,541	87,348	292,336	233,857
CCS: principal diagnosis				
1: Infectious and Parasitic DX	8,351	273	6,726	1,352
2: Neoplasms	27,723	24,110	221	3,392
3: Endocr, Nutri, Metab, Immun DX	7,074	1,626	3,095	2,353
4: Dx of Blood, Blood-Forming Organs	2,689	416	1,510	763
5: Mental Disorders	8,004	216	5,939	1,849
6: Dx of Nervous System, Sense Organs	46,362	9,754	34,891	1,717
7: Dx of Circulatory System	72,483	13,626	18,290	40,567
8: Dx of Respiratory System	59,376	7,848	43,290	8,238
9: Dx of Digestive System	30,309	5,802	18,551	5,956
10: Dx of Genitourinary System	17,253	3,260	11,766	2,227
11: Complic Preg, Birth, Puerperium	15,170	2,244	4,302	8,624
12: Dx of Skin and Subcutaneous Tissue	8,553	826	6,626	1,101
13: Dx of Musculoskel, Connective Tissue	25,937	6,453	16,583	2,901
14: Congenital Anomalies	2,098	917	180	1,001
15: Perinatal Conditions	145,485	810	3,227	141,448
16: Injury and Poisoning	90,679	3,639	79,694	7,346
17: Other Conditions	45,995	5,528	37,445	3,022