



# H·CUP

HEALTHCARE COST AND UTILIZATION PROJECT

## HCUP Methods Series



Agency for Healthcare  
Research and Quality



U.S. Department of Health and Human Services  
Agency for Healthcare Research and Quality

**Contact Information:**  
**Healthcare Cost and Utilization Project (HCUP)**  
**Agency for Healthcare Research and Quality**  
**5600 Fishers Lane**  
**Room 07W17B**  
**Mail Stop 7W25B**  
**Rockville, MD 20857**  
**<http://www.hcup-us.ahrq.gov>**

**For Technical Assistance with HCUP Products:**

**Email: [hcup@ahrq.gov](mailto:hcup@ahrq.gov)**

**or**

**Phone: 1-866-290-HCUP**

Recommended Citation: Fingar KR, Owens PL, Barrett ML, Steiner CA. *Using the HCUP Databases to Study Incidence and Prevalence*. HCUP Methods Series Report # 2016-06 ONLINE. December 6, 2016. U.S. Agency for Healthcare Research and Quality. Available:  
Available: <http://www.hcup-us.ahrq.gov/reports/methods/methods.jsp>.

## Table of Contents

Introduction .....	1
Incidence .....	1
Prevalence .....	2
Data Considerations .....	2
Population of Interest.....	2
Evaluation Criteria .....	3
Selecting Appropriate HCUP Databases .....	4
HCUP statewide databases. ....	5
HCUP nationwide databases. ....	5
Key HCUP Data Elements.....	6
Diagnosis codes. ....	6
Diagnosis present on admission (POA). ....	6
Discharge disposition.....	7
Patient linkage numbers. ....	8
Examples of Incidence and Prevalence Using HCUP Data .....	8
Incidence.....	8
Table 1. Onset of an event during hospitalization.....	9
Table 2. Development of conditions during pregnancy.....	10
Table 3. Surgeries.....	10
Table 4. Conditions that are likely to result in a hospital encounter .....	12
Prevalence .....	13
Table 5. Comparison of asthma, COPD, and diabetes in BRFSS and HCUP data, 2013 ...	13
Table 6. Examples of prevalence using HCUP data.....	14
Examples in Which HCUP Data Have Been Used Inappropriately.....	15
Table 7. Inappropriate use of HCUP data to measure incidence or prevalence.....	15
Conclusion .....	17

The authors gratefully acknowledge Raynard Washington, Ph.D., M.P.H., Council for Affordable Quality Healthcare, and researchers at the CDC, National Center on Birth Defects and Developmental Disabilities for assistance in developing this report.

## INTRODUCTION

To date, there have been over 1,000 published studies measuring incidence or prevalence using the Healthcare Cost and Utilization Project (HCUP) databases.<sup>1</sup> Although HCUP data can be used to study the incidence or prevalence of hospital utilization for certain conditions or procedures, in some scenarios the use of these measures may be inappropriate or have significant limitations. It is important not to conflate the large sample size of the HCUP databases with the ability to measure incidence and prevalence of a condition itself or of its complications. HCUP data contain information on patients who obtained hospital acute care for their condition and complications. Thus, whereas incidence and prevalence are defined as measures of disease occurrence and frequency, this Methods Series report extrapolates those terms that traditionally have been used to describe disease processes and applies them to health service utilization. In the case of HCUP, these terms can be thought of as measures of event occurrence and frequency, with an event being a hospitalization or visit. After providing background information on incidence and prevalence, this report offers guidance for selecting measures of incidence and prevalence that are appropriate for use with HCUP data and illustrates data considerations through examples.

### Incidence

Incidence is a measure of disease onset that describes the number of events that (1) occurred among a population during a specified time period and (2) can be expressed as an *incidence proportion* or an *incidence rate*. The terms *onset* and *development* imply that the individual did not already have the disease or condition—in other words, that it was a “new” case.

The *incidence proportion*, also referred to as *average risk*, is defined as the number of instances in which the onset of a disease or event occurred during a specified time period divided by the total number of subjects followed for that time period.<sup>2</sup>

In contrast, the *incidence rate* is expressed as a function of person-time. The numerator is the same as that of the *incidence proportion* (i.e., the number of events that developed during a specified time period), but the denominator is the total time, measured in person-days, person-months, or person-years, experienced by all of the subjects at risk who were followed over time. Person-time stops accruing and is not included in the denominator after the occurrence of the event of interest, after a competing event (e.g., a death), or when the subject is lost to follow-up.<sup>3</sup>

There are two methods for determining what constitutes a new case and thus which events should be counted in the numerator. This decision depends on whether the event can occur multiple times. For medical conditions such as hypertension or rheumatoid arthritis, the onset of the disease occurs once, and therefore an individual can be counted only once. Generally, in HCUP one is not able to identify new cases of conditions such as these. Other events, such as

---

<sup>1</sup> Healthcare Cost and Utilization Project (HCUP). Publications Search. Updated February 5, 2010. Rockville, MD: Agency for Healthcare Research and Quality. [www.hcup-us.ahrq.gov/reports/pubsearch/pubsearch.jsp](http://www.hcup-us.ahrq.gov/reports/pubsearch/pubsearch.jsp). Accessed October 24, 2016.

<sup>2</sup> Rothman KJ. Epidemiology: An Introduction. New York, NY: Oxford University of Press; 2002.

<sup>3</sup> Ibid.

acute myocardial infarctions (AMI), are transient and can recur. In those cases, investigators may choose to count (1) multiple events that recur during a given time period or (2) the first occurrence.<sup>4</sup> More information on these decisions, including considerations for accumulating time in the denominator when counting multiple versus first occurrences of an event, can be found elsewhere.<sup>5</sup>

## Prevalence

Whereas incidence is a measure of disease onset (i.e., the development of a new case), prevalence is a measure of disease status. Prevalence often is used to measure the frequency of diseases, such as diabetes or multiple sclerosis, for which it may be difficult to determine the onset. The prevalence of a condition is defined as the number of individuals with that condition at a specific point in time divided by the total number of individuals in the population.

## DATA CONSIDERATIONS

### Population of Interest

Each investigator will need to identify the population of interest in order to determine what populations contribute to the numerators and denominators of incidence and prevalence measures and to evaluate their appropriateness. For instance, is the goal to assess incidence or prevalence for a geographic population such as all of the United States or a particular State, a subpopulation such as a particular racial or ethnic group, or the hospital population? Measuring incidence or prevalence for a condition among hospitalized patients, for which HCUP is ideal, has a slightly different and less concerning set of limitations than does applying these measures to the general population. However, the latter can be done depending on the extent to which certain criteria are met.

Care should be taken to evaluate whether (1) HCUP data cover all hospital settings in which individuals are treated, (2) the study population of interest is not underrepresented in hospital data, and (3) the condition or procedure under study is well identified and coded in hospital data. For instance, it is important to note that HCUP data generally capture information on care at community hospitals<sup>6</sup> and do not capture diseases treated or procedures conducted in other ambulatory care settings (e.g., physician offices, urgent care centers, and ambulatory surgery centers that are not hospital owned). Furthermore, some populations may be more likely to use hospitals or clinics that are not represented in the HCUP data. For instance, veterans often use Veterans Affairs hospitals, which are federally owned and outside the purview of statewide data organizations that collect hospital data. American Indians/Alaska Natives also are less likely to use community-based hospitals and more likely to receive health care from the Indian Health Service or other tribal clinics. Therefore, many examples of prevalence in HCUP are actually

---

<sup>4</sup> Rothman KJ. *Epidemiology: An Introduction*. New York, NY: Oxford University of Press; 2002.

<sup>5</sup> Rothman KJ, Greenland S. *Modern Epidemiology*. Philadelphia, PA: Lippincott Williams & Wilkins; 1998.

<sup>6</sup> The American Hospital Association (AHA) defines community hospitals as “all nonfederal, short-term general, and special hospitals, including special children’s hospitals, whose facilities and services are available to the public.” Included among community hospitals are specialty hospitals such as obstetrics and gynecology; ear, nose, and throat; orthopedic; and pediatric institutions. Also included are public hospitals and academic medical centers. AHA Data/Viewer. Glossary. <http://www.ahadataviewer.com/glossary>. Accessed October 20, 2016.

measures of *treated* incidence or prevalence at *community hospitals*, and care should be taken when presenting stratified data or when drawing conclusions about certain subpopulations. Finally, caution also should be taken when measuring incidence or prevalence of hospital utilization for conditions often coded as secondary diagnoses, such as hypertension, which may be underreported if they do not contribute to the reason for the hospital stay or visit.

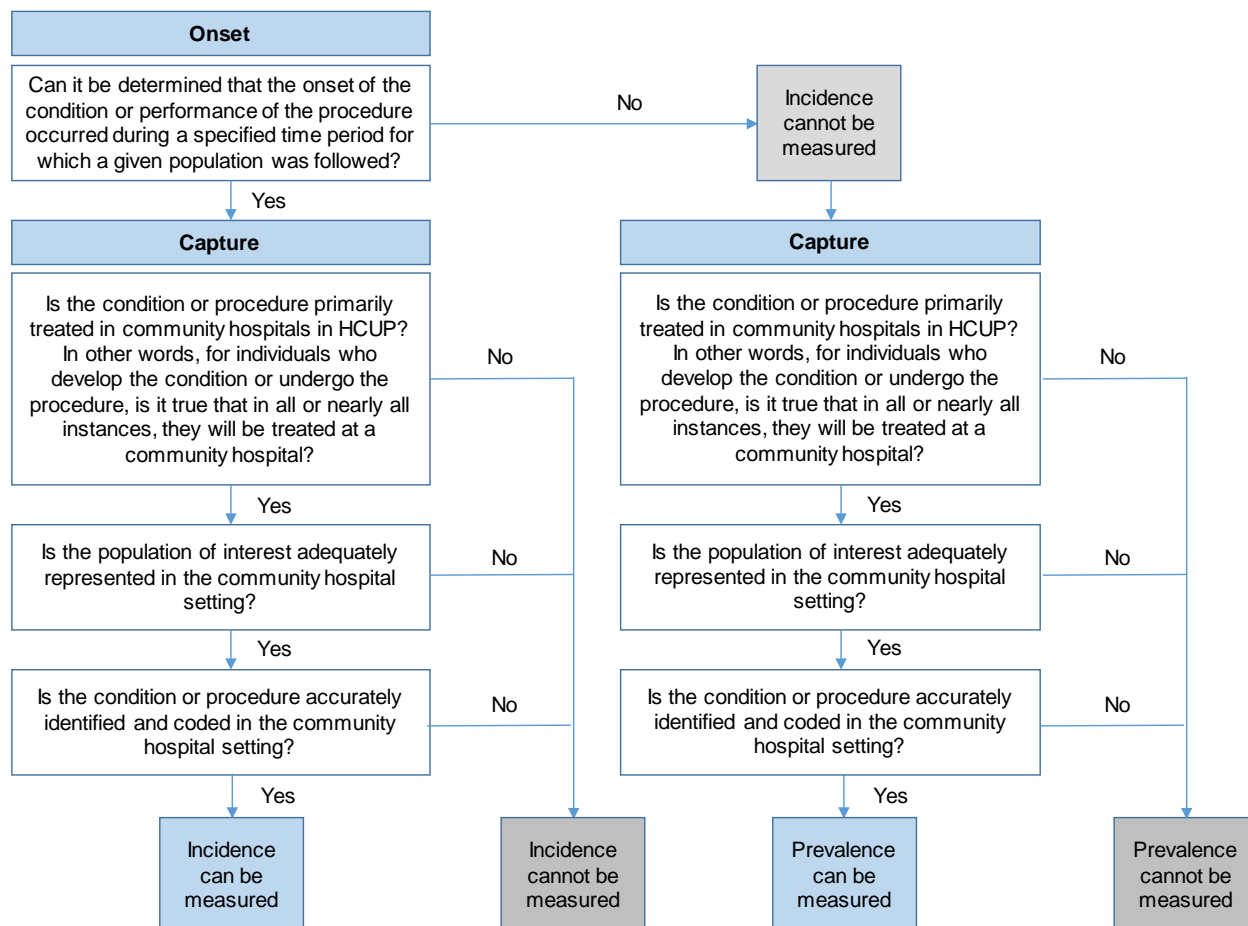
## Evaluation Criteria

This report uses the following criteria—related to *onset* and *capture*—to assess whether the incidence or prevalence of an event can be measured using the HCUP databases.

1. Incidence
  - a. Can it be determined that the *onset* of the event occurred during the specified time period?
  - b. Is the event completely *captured* in HCUP databases? In other words, to what extent does HCUP have complete ascertainment of all cases within a given population? Useful questions include the following: “Is the condition or procedure treated primarily in the inpatient or outpatient hospital setting?” “Is the condition or procedure accurately identified and coded in the hospital setting?” “Is the population of interest adequately represented in the hospital setting?”
2. Prevalence
  - a. Is the event completely captured in HCUP databases? In other words, to what extent does HCUP have complete ascertainment of all cases within a given population? Again, useful questions include the following: “Is the condition or procedure treated primarily in the inpatient or outpatient hospital setting?” “Is the condition or procedure accurately identified and coded in the hospital setting?” “Is the population of interest adequately represented in the hospital setting?”

Figure 1 displays a decision tree highlighting these criteria. This rubric was used to evaluate the examples of incidence and prevalence presented in this report.

**Figure 1. Criteria to Assess Whether Incidence and Prevalence Can Be Measured Using HCUP Data**



Abbreviation: HCUP, Healthcare Cost and Utilization Project

The discussion of the types and features of HCUP data that follows is intended to help researchers make these decisions.

### Selecting Appropriate HCUP Databases

HCUP is a family of databases that contains information on inpatient stays and emergency department (ED) visits, in addition to ambulatory surgery and other hospital-based outpatient services. HCUP databases bring together the data collection efforts of State data organizations, hospital associations, private data organizations, and the Federal government to create a national information resource of encounter-level health care data. Statewide databases include information from most, if not all, community hospitals in a State, translated into a uniform format to facilitate multistate comparisons and analyses. Nationwide databases are a sample of these records weighted to provide national estimates. All HCUP databases include a core set of clinical and nonclinical information on all patients, regardless of payer, including individuals covered by Medicare, Medicaid, and private insurance, as well as those who are uninsured.



*HCUP statewide databases.* The State Inpatient Databases (SID) include discharge-level data on inpatient stays from community hospitals in the State. Together, the SID encompass more than 95 percent of all U.S. community hospital discharges. The SID can be used to investigate questions unique to one State, to compare data from two or more States, to conduct market-area variation analyses, and to identify State-specific trends in inpatient care utilization, access, charges, and outcomes.

The State Emergency Department Databases (SEDD) include encounter-level data on ED visits that do not result in an admission. The SEDD contain information from ED encounters from hospital-based EDs, translated into a uniform format to facilitate multistate comparisons and analyses. Researchers and policymakers use the SEDD to investigate access to health care in a changing health care marketplace; to identify State-specific trends in ED utilization, access, charges, and outcomes; and to conduct market-area research and small-area variation analyses.

The State Ambulatory Surgery and Services Databases (SASD) include encounter-level data for ambulatory surgeries and also may include various types of outpatient services such as observation stays, lithotripsy, radiation therapy, imaging, chemotherapy, and labor and delivery. The specific types of ambulatory surgery and outpatient services included in each SASD vary by State and data year. All SASD include data from hospital-owned ambulatory surgery facilities. In addition, some States include data from facilities not owned by a hospital. The designation of a facility as hospital owned is specific to its financial relationship with a hospital that provides inpatient care and is not related to its physical location. Hospital-owned ambulatory surgery and other outpatient care facilities may be contained within the hospital, physically attached to the hospital, or located in a different geographic area.

*HCUP nationwide databases.* The National (Nationwide) Inpatient Sample (NIS), the Kids' Inpatient Database (KID), and the Nationwide Readmissions Database (NRD) are nationwide samples drawn from the SID.

Beginning in 2012, the NIS is a sample of discharge records from all HCUP-participating hospitals. Prior to 2012, the NIS was a sample of hospitals from which all discharges were retained. The NIS is the largest all-payer inpatient care database in the United States, containing data on more than seven million hospital stays. Researchers and policy analysts use the NIS to identify, track, and analyze trends in health care utilization, access, charges, quality, and outcomes in the United States. Its large sample size is ideal for developing national and regional estimates and enables analyses of hospitalizations associated with rare conditions, uncommon treatments, and special populations.

The KID is a sample of pediatric discharges from all community, nonrehabilitation hospitals in States participating in HCUP. The KID contains data from approximately two million hospital stays for patients aged 20 years and younger. Researchers and policy analysts use the KID to analyze hospital utilization, charges, quality, and outcomes for children across the United States. Its large sample size is ideal for developing national and regional estimates and enables analyses of inpatient care for rare conditions, such as spinal cord tumors and congenital anomalies, as well as uncommon treatments.

The NRD is drawn from the SID with reliable, verified patient linkage numbers and is designed to support various types of analyses of national readmission rates for all payers and those with no insurance. The NRD contains data from approximately 14 million discharges each year. Its

large sample size allows for analysis across hospital types and the study of readmissions for relatively uncommon disorders and procedures. Outcomes of interest include national readmission rates, reasons for returning to the hospital for care, and hospital costs for discharges with and without readmissions. Unlike the NIS, the NRD can be used to identify unique individuals with admissions during a calendar year.

The Nationwide Emergency Department Sample (NEDS) is a sample of hospital-based EDs with data drawn from the SID and SEDD. The SID contain information on patients initially seen in the ED and then admitted to the same hospital. The SEDD capture information on ED visits that do not result in an admission (i.e., treat-and-release visits and transfers to another hospital). The NEDS includes about 30 million records a year and supports health care policy and research on a variety of topics, including the use of and charges for ED services, medical treatment effectiveness, quality of ED care, access to care, and utilization of ED services by special populations.

## Key HCUP Data Elements

There are several HCUP data elements worth noting because they have implications for being able to assess capture and disease onset using the HCUP databases and therefore have implications for being able to accurately measure incidence and prevalence.

*Diagnosis codes.* To determine whether selected codes produce accurate results, the investigator always should evaluate the consistency of diagnosis coding in the International Classification of Diseases, Ninth and Tenth Revisions, Clinical Modification (ICD-9-CM and ICD-10-CM) across hospitals in a State and across States.<sup>7</sup> Equally importantly, the investigator should review evidence of the positive predictive value of ICD-9-CM or ICD-10-CM codes in hospital discharges as ascertained through validation studies using medical chart abstraction. It is important to remember billing codes are not medical diagnoses.

It also is important to note that certain diagnosis codes may enable investigators to determine the onset of an event. For instance, certain codes pertaining to pregnancy indicate whether a condition developed over the course of pregnancy or was a pre-existing condition, such as pre-existing diabetes versus gestational diabetes.

*Diagnosis present on admission (POA).* The POA indicators in HCUP databases in principle distinguish medical conditions that are present before the patient enters the hospital from those that occur during the hospital stay, such as complications or adverse events.<sup>8</sup> These indicators may be helpful in measuring incidence because they can be used to determine whether the onset of an event occurred during a hospital stay. However, the POA indicators may not always be correctly applied. For example, one study found that among medical admissions with hospital

---

<sup>7</sup> Haghghat N, Hu M, Laurent O, et al. Comparison of birth certificates and hospital-based birth data on pregnancy complications in Los Angeles and Orange County, California. *BMC Pregnancy and Childbirth*. 2016;16:93.

<sup>8</sup> Barrett ML, Owens PL, Bolhack J, et al. Examination of the Coding of Present-on-Admission Indicators in Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID). HCUP Methods Series Report #2015-06. Rockville, MD: Agency for Healthcare Research and Quality; 2015. <http://www.hcup-us.ahrq.gov/reports/methods/methods.jsp>. Accessed October 24, 2016.

discharge codes for deep vein thrombosis or pulmonary embolism, the majority of hospital-associated events were coded as POA=Y.<sup>9</sup>

A standardized definition of POA has been specified by the National Uniform Billing Committee (NUBC) and the Centers for Medicare & Medicaid Services (CMS), which define a condition as POA if it was present when the order for inpatient admission occurs, including those that develop during an outpatient encounter, such as an ED visit or outpatient surgery.<sup>10,11</sup> POA may be indicated with principal and secondary diagnoses, as well as external cause of injury codes. When using the POA indicator, investigators may need to exclude hospitals or diagnoses that are exempt from reporting. For instance, CMS considers critical access and long-term care hospitals, among other types of hospitals, exempt from POA reporting. Examples of exempt diagnoses include old myocardial infarction, normal delivery, and congenital anomalies. More information on these exemptions is available in the HCUP Methods Series Report *Examination of the Coding of Present-on-Admission Indicators in HCUP SID*.<sup>12</sup> It should be noted that HCUP does not assess the sensitivity or specificity of the POA indicator, which would require validation through linked chart reviews.

POA indicators (HCUP data elements DXPOAn) are available only in some SID, SASD, and SEDD. Starting in data year 2014, the HCUP State databases include edit flags (HCUP data elements with “POA” and “edit” in the name) that can be used to verify the coding of POA on individual discharges and by hospital. More information on the POA edit flags and the quality of POA data in the SID is available in the HCUP Methods Series Report *Examination of the Coding of Present-on-Admission Indicators in HCUP SID*.<sup>13</sup>

*Discharge disposition.* The HCUP data element on disposition (DISuniform) indicates whether the patient was discharged routinely or against medical advice, transferred to another health care facility, or died. Certain conditions with a high likelihood of hospitalization, such as a heart attack, are likely to be captured in the SID, except for those events resulting in death before transfer to a hospital. In those instances, HCUP data may need to be combined with another data source, such as death certificates, to completely capture the event. The data element on disposition can be used to exclude in-hospital deaths so that those events are not double counted when both sources are used. Discharge disposition is available on all HCUP databases.

---

<sup>9</sup> Khanna RR, Kim SB, Jenkins I, El-Kareh R, Afsarmanesh N, Amin A, Sand H, Auerbach A, Chia CY, Maynard G, Romano PS, White RH. Predictive value of the present-on-admission indicator for hospital-acquired venous thromboembolism. *Med Care*. 2015 Apr;53(4):e31-6.

<sup>10</sup> The NUBC Data Specifications manual is not publicly available; it is available only by subscription. The CMS Coding Guidelines are available online at [http://www.cdc.gov/nchs/data/icd/icd9cm\\_guidelines\\_2011.pdf](http://www.cdc.gov/nchs/data/icd/icd9cm_guidelines_2011.pdf). Accessed May 11, 2016.

<sup>11</sup> Centers for Medicare & Medicaid Services. Hospital-Acquired Conditions (Present on Admission Indicator): Reporting. Updated February 12, 2014. <http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Reporting.html>. Accessed May 11, 2016.

<sup>12</sup> Barrett ML, Owens PL, Bolhack J, et al. Examination of the Coding of Present-on-Admission Indicators in Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID). HCUP Methods Series Report #2015-06. Rockville, MD: Agency for Healthcare Research and Quality; 2015. <http://www.hcup-us.ahrq.gov/reports/methods/methods.jsp>. Accessed October 24, 2016.

<sup>13</sup> Ibid.

*Patient linkage numbers.* Patient linkage numbers (HCUP data element visitLink) can be used to track sequential visits for a patient within a State and across facilities and hospitals settings (inpatient, ED, and ambulatory surgery) while adhering to strict privacy guidelines. They also allow the exclusion of multiple visits for a patient to obtain an accurate count of individuals treated for a certain disease. Patient linkage numbers are available in some HCUP State databases and the NRD. Information on using the patient linkage numbers and on which State databases contain this information is available on the HCUP User Support (HCUP-US) Web site under HCUP Supplemental Variables for Revisit Analyses (<http://www.hcup-us.ahrq.gov/toolssoftware/revisit/revisit.jsp>).

## **EXAMPLES OF INCIDENCE AND PREVALENCE USING HCUP DATA**

Below we illustrate through a number of examples some considerations to keep in mind when using HCUP data to measure incidence and prevalence. In some examples, hospital encounters may ascertain all or nearly all cases of a condition or procedure in the general population. In those cases, the general population is noted as the denominator. In other cases, the measure of incidence or prevalence is among the population of individuals using hospital services, which is noted in the denominator. Both examples of incidence proportions (i.e., the denominator is a count of individuals) and rates (i.e., the denominator is a measure of person-time) are included. For conditions or procedures that are transient but can recur, the examples allow individuals who experienced past events to contribute to measures of incidence. Investigators would need to decide whether to count multiple events per patient or to count only one event per patient during the specified time period, which may be done using patient linkage numbers.

### **Incidence**

The examples of incidence are organized into the following types of hospital encounters:

1. Onset of an event during hospitalization
2. Onset of conditions during pregnancy
3. Surgeries
4. Conditions that are likely to result in a hospital encounter

For each set of examples, we discuss various considerations including (1) the possible HCUP databases, (2) whether there is complete capture of the event in the HCUP databases, (3) whether the onset of the event can be determined, (4) whether incidence can be measured, and (5) how to define the numerator and denominator for the rate.

Table 1 uses the criteria to evaluate studying the incidence of the onset of two types of events during hospitalization: development of sepsis during hospitalization and in-hospital mortality.

**Table 1. Onset of an event during hospitalization**

<b>Development of sepsis during hospitalization</b>	
Possible HCUP databases	NIS, SID
Complete capture in HCUP databases?	Yes, by definition the event must occur during a hospital stay.
Can the onset of the event be determined?	Yes, however, researchers must use POA to restrict to patients who did not have sepsis on admission.
Can incidence be measured?	Yes
Numerator	Number of inpatient stays involving sepsis in which sepsis was not POA
Denominator	Number of total initial inpatient stays (which can be used to calculate a proportion) or the number of person-days in the hospital (which can be used to calculate a rate)
<b>In-hospital mortality</b>	
Possible HCUP databases	NIS, SID
Complete capture in HCUP databases?	Yes, by definition the event must occur during a hospital stay.
Can the onset of the event be determined?	Yes
Can incidence be measured?	Yes
Numerator	Number of inpatient stays in which the disposition equals death
Denominator	Number of total initial inpatient stays (which can be used to calculate a proportion) or the number of person-days in the hospital (which can be used to calculate a rate)

Abbreviations: HCUP, Healthcare Cost and Utilization Study; NIS, National (Nationwide) Inpatient Sample; POA, present on admission; SID, State Inpatient Databases

Table 2 uses the criteria to evaluate studying the incidence of conditions that develop during pregnancy: gestational diabetes, preeclampsia/eclampsia, and peripartum cardiomyopathy.

**Table 2. Development of conditions during pregnancy**

<b>Gestational diabetes, preeclampsia/ eclampsia, peripartum cardiomyopathy</b>	
Possible HCUP databases	NIS, SID
Complete capture in HCUP databases?	Yes. Nearly all births result in inpatient hospitalization in community hospitals, except those that occur at home or at a federal hospital.
Can the onset of the event be determined?	Yes. Certain ICD-9-CM codes, such as those for gestational diabetes, preeclampsia/eclampsia, and peripartum cardiomyopathy indicate that the condition arose during pregnancy. Records should be evaluated for the use of these codes in addition to codes indicating pre-existing conditions (e.g., Type I or II diabetes, hypertension, cardiomyopathy), and those records with pre-existing conditions should be excluded from the numerator and denominator.
Can incidence be measured?	Yes
Numerator	Number of inpatient stays for childbirth with the peripartum condition
Denominator	Number of inpatient stays for childbirth

Abbreviations: HCUP, Healthcare Cost and Utilization Study; NIS, National (Nationwide) Inpatient Sample; SID, State Inpatient Databases

Table 3 uses these criteria to evaluate studying the incidence of three surgeries: kidney transplant, cataract surgery, and hysterectomy.

**Table 3. Surgeries**

<b>Kidney transplant</b>	
Possible HCUP databases	NIS, SID
Complete capture in HCUP databases?	Yes. Kidney transplant would require inpatient hospitalization.
Can the onset of the event be determined?	Yes
Can incidence be measured?	Yes
Numerator	Number of inpatient stays for kidney transplant
Denominator	Number of individuals in the general population

**Table 3. Surgeries (continued)**

<b>Cataract surgery</b>	
Possible HCUP databases	SID, SASD
Complete capture in HCUP databases?	No. Cataract surgery is captured in the SID and SASD, but this procedure often is performed in non-hospital-based outpatient settings.
Can the onset of the event be determined?	Yes, although the investigator may need to consider whether to count multiple cataract surgeries for the same patient during the specified time period, or only the first occurrence.
Can incidence be measured?	No. Because of the movement of these procedures to office-based surgical settings, <sup>a</sup> it is likely that procedures performed in hospital-owned outpatient facilities drastically underestimate the incidence of cataract surgery.
Numerator	Not applicable
Denominator	Not applicable
<b>Hysterectomy</b>	
Possible HCUP databases	SID, SASD
Complete capture in HCUP databases?	Yes, with limitations. The majority (60 percent) of hospital-based hysterectomies are performed in an inpatient setting with 40 percent in hospital-owned ambulatory surgery centers. <sup>b</sup> However, it should be acknowledged that these procedures may be performed in non-hospital-owned facilities, leading to an underestimate of incidence.
Can the onset of the event be determined?	Yes
Can incidence be measured?	Yes, with limitations. It should be acknowledged that procedures performed in hospital-owned facilities may underestimate the incidence of hysterectomy.
Numerator	Number of inpatient stays plus number of ambulatory surgeries at hospital-owned facilities for hysterectomy
Denominator	Number of individuals in the general population

Abbreviations: HCUP, Healthcare Cost and Utilization Study; NIS, National (Nationwide) Inpatient Sample; SID, state inpatient databases; SASD, State Ambulatory Surgery and Services Databases

<sup>a</sup> Ifft D. Moving cataract surgery into the practice setting. *Ophthalmology Management*. 2015;19:6–8.

<sup>b</sup> Carroll CP, Cochran JA, Guse CE, et al. Are we underestimating the burden of traumatic brain injury? Surveillance of severe traumatic brain injury using Centers for Disease Control International Classification of Disease, Ninth Revision, Clinical Modification, traumatic brain injury codes. *Neurosurgery*. 2012;71(6):1064–70; discussion 1070.

Table 4 uses these same criteria to evaluate studying the incidence of ruptured appendix, which is likely to result in a hospital encounter.

**Table 4. Conditions that are likely to result in a hospital encounter**

<b>Ruptured appendix</b>	
Possible HCUP databases	SID
Complete capture in HCUP databases?	Yes. It is likely that a ruptured appendix will result in an inpatient stay.
Can the onset of the event be determined?	Yes
Can incidence be measured?	Yes
Numerator	Number of inpatient stays for ruptured appendix
Denominator	Number of individuals in the general population

Abbreviations: ED, emergency department; HCUP, Healthcare Cost and Utilization Study; SID, State Inpatient databases.



## Prevalence

HCUP data can be used to study patients hospitalized for conditions such as asthma, chronic obstructive pulmonary disease (COPD), and diabetes, but this represents *hospital-treated* prevalence. Other data sources such as the Behavioral Risk Factor Surveillance System (BRFSS) have information on the prevalence of these conditions in the general population. Table 5 compares the prevalence of asthma, COPD, and diabetes among all adults<sup>14</sup> and children<sup>15</sup> living in the United States from BRFSS with the hospital-treated prevalence of these conditions in HCUP data.

**Table 5. Comparison of asthma, COPD, and diabetes in BRFSS and HCUP data, 2013**

<b>BRFSS</b>	<b>Individuals in the U.S. population with condition, %</b>
Lifetime asthma, adults	14.1
Current asthma, adults	9.0
Lifetime asthma, children	14.0
Current asthma, children	9.2
Ever told had COPD, adults	6.3
Ever told had diabetes, adults	9.8
<b>Inpatient stays from the HCUP NIS</b>	<b>Inpatient hospital stays per 100,000 U.S. population</b>
Asthma diagnosis, adults	101.5
Asthma diagnosis, children	136.9
COPD diagnosis, adults	264.8
Diabetes diagnosis, adults	213.7
<b>Treat-and-release ED visits from the HCUP NEDS</b>	<b>Treat-and-release ED visits per 100,000 U.S. population</b>
Asthma diagnosis, adults	426.8
Asthma diagnosis, children	855.6
COPD diagnosis, adults	525.1
Diabetes diagnosis, adults	276.6

Abbreviations: BRFSS, Behavior Risk Factor Surveillance System; COPD, chronic obstructive pulmonary disease; ED, emergency department; HCUP, Healthcare Cost and Utilization Study; NEDS, Nationwide Emergency Department Sample; NIS, National (Nationwide) Inpatient Sample.

Notes: For the HCUP data, the Elixhauser Clinical Classifications Software (CCS) was used to define the following on the basis of the first-listed diagnoses: CCS of 128 for asthma, CCS of 127 for chronic obstructive pulmonary disease and bronchiectasis, and CCS of 49 and 50 for diabetes.

Source: Centers for Disease Control and Prevention, Behavioral Risk Factor Surveillance System, 2013; Agency for Healthcare Research and Quality (AHRQ), Center for Delivery, Organization, and Markets, Healthcare Cost and Utilization Project (HCUP), National Inpatient Sample (NIS) and Nationwide Emergency Department Sample (NEDS), 2013.

<sup>14</sup> Centers for Disease Control and Prevention. BRFSS Prevalence & Trends Data. Updated August 18, 2016. <http://www.cdc.gov/brfss/brfssprevalence/index.html>. Accessed October 24, 2016.

<sup>15</sup> Centers for Disease Control and Prevention. 2013 Asthma Data: BRFSS Prevalence Tables and Maps. Updated March 20, 2015. <http://www.cdc.gov/asthma/brfss/2013/default.htm>. Accessed October 24, 2016.

HCUP data may, however, be used to measure prevalence in more specific scenarios, in which complete capture of both the numerator and the denominator is present. Table 6 provides three examples. The first two examples use only HCUP data: diabetes among inpatient stays at the time of delivery and hypertension among inpatient stays involving a kidney transplant. In the third example (hospitalization among the U.S. population with asthma), the denominator comes from BRFSS.

**Table 6. Examples of prevalence using HCUP data**

<b>Diabetes among inpatient stays at the time of delivery</b>	
Possible HCUP databases	NIS, SID
Complete capture in HCUP databases?	Yes. Nearly all births result in inpatient hospitalization in community hospitals, except those that occur at home or at a federal hospital.
Can prevalence be measured?	Yes
Numerator	Number of inpatient stays for childbirth with pre-existing or gestational diabetes
Denominator	Number of inpatient stays for childbirth
<b>Hypertension among inpatient stays involving a kidney transplant</b>	
Possible HCUP databases	NIS, SID
Complete capture in HCUP databases?	Yes, with limitations. Kidney transplant would require inpatient hospitalization, but hypertension may not be coded on the record if it did not complicate care during the stay.
Can prevalence be measured?	Yes
Numerator	Number of inpatient stays for kidney transplant with hypertension
Denominator	Number of inpatient stays for kidney transplant
<b>Hospitalization among a State-specific population with asthma</b>	
Possible HCUP databases	SID
Complete capture in HCUP databases?	Yes, however the count would need to be specific to patients treated in an inpatient setting for asthma and not to inpatient stays related to asthma.
Can prevalence be measured?	Yes
Numerator	Number of patients who were hospitalized with a diagnosis of asthma. To obtain a patient-level count, the first step is to extract all asthma related-stays and then use the HCUP revisit variables to count only one stay per person.
Denominator	Number of individuals in the U.S. population with lifetime or current asthma identified using BRFSS.

Abbreviations: BRFSS, Behavior Risk Factor Surveillance System; HCUP, Healthcare Cost and Utilization Study; NIS, National (Nationwide) Inpatient Sample; SID, State Inpatient Databases.

## Examples in Which HCUP Data Have Been Used Inappropriately

In addition to the examples presented above, in Table 7 we highlight examples from the literature in which HCUP data have been used to measure incidence or prevalence in error.

**Table 7. Inappropriate use of HCUP data to measure incidence or prevalence**

<b>Incidence of hypertension in emergency department visits</b>	
HCUP databases used	NEDS
Complete capture in HCUP databases?	Yes, with limitations. Note that all cases with hypertension do not necessarily result in an ED visit, but if framed as measuring the prevalence (see note below regarding incidence) of hypertension among ED visits as opposed to the general population, that may be appropriate. However, limitations related to underascertainment of hypertension should be noted. Among ED visits, hypertension may not be diagnosed or coded if it was not related to the reason for the visit (e.g., hypertension may not be coded for patient in the ED for a sprained ankle or urinary tract infection).
Can the onset of the event be determined?	No, it cannot be determined whether the ED visit with hypertension was a new case of hypertension or the hypertension was pre-existing.
Can incidence be measured?	No
<b>Prevalence of low back pain disorders</b>	
HCUP databases used	NIS
Complete capture in HCUP databases?	No. All cases of low back pain disorder do not necessarily result in inpatient admission.
Can prevalence be measured?	No
<b>Prevalence of obesity, smoking and alcohol</b>	
HCUP databases used	NIS
Complete capture in HCUP databases?	No. All cases of obesity, smoking, or alcohol use do not necessarily result in inpatient admission.
Can prevalence be measured?	No
<b>Prevalence of a rare condition, such as sickle cell anemia, or cystic fibrosis</b>	
HCUP databases used	NIS
Complete capture in HCUP databases?	No. All cases of rare disorder do not necessarily result in inpatient admission.
Can prevalence be measured?	No.

**Table 7. Inappropriate use of HCUP data to measure incidence or prevalence (continued)**

<b>Prevalence of a comorbid condition among patients with rare disorder</b>	
HCUP databases used	NIS
Complete capture in HCUP databases?	No. All cases of rare disorder do not necessarily result in inpatient admission. Therefore, presence of a comorbid condition is limited to cases of the rare disorder treated in the hospital.
Can prevalence be measured?	No.

Abbreviations: ED, emergency department; HCUP, Healthcare Cost and Utilization Study; NEDS, National Emergency Department Sample; NIS, National (Nationwide) Inpatient Sample.

## **CONCLUSION**

In conclusion, in a number of examples incidence, prevalence, or both can be measured using the HCUP databases. In order to measure incidence and prevalence, researchers should assess whether the event under investigation is completely captured by HCUP data.

Additionally, to measure incidence, researchers will need to decide whether the onset of the event during a specific time period can be determined. However, there are some scenarios in which incidence and prevalence cannot be measured using HCUP data. The HCUP databases have a large sample size which is advantageous for many research studies, particularly for those focused on rare events. Yet, it is important to remember that HCUP data are limited to information on patients who obtained hospital acute care for their condition and not information on the incidence or prevalence of the condition itself is unknown. This report provides a number of tools—including a description of HCUP databases and data elements—that can guide investigators in making these decisions and conducting research.