MAINTAIN AND EXPAND
THE HEALTHCARE COST AND UTILIZATION PROJECT (HCUP)
Contract No. HHSA-290-2006-00009-C

QUARTERLY DATA INFRASTRUCTURE AND RECOMMENDATION REPORT
DELIVERABLE #825.21B

August 21, 2009

Submitted to:
Jenny Schnaier, Project Officer
Agency for Healthcare Research and Quality
540 Gaither Road
Rockville, MD 20850

Submitted by:
Thomson Reuters
5425 Hollister Ave, Suite 140
Santa Barbara, CA 93111

Table of Contents

Executive Summary 1
Introduction 1
Background 2
  Availability of Quarterly Hospital Data 2
Timing Implications 4
  Timing-Related Differences between Quarterly and Annual Data 4
  Availability of Quarterly Data 4
  Quarterly Data or Statistics and NIS Data 7
    Data Available within Five Months of a Quarter's End 7
    Data Available within Seven Months of a Quarter's End 8
HCUP Processing Infrastructure 8
  Current Processing Stream 10
Changes Necessary to Accommodate Quarterly Data 10
  Assumptions 11
  Data Creation Costs 12
    Data Acquisition 12
    Data Purchase Costs 12
    Processing 13
    Data Delivery 14
    HCUP-US Documentation 15
    Data Destruction 15
  Other Cost Considerations 15
    Infrastructure 15
    Productivity 16
  Aggregated Costs of Creating Quarterly Data 17
  Central Distributor 18
    Data Acquisition 19
    Creating Central Distributor Files 19
    Central Distributor Operations (SSS) 19
    Other Central Distributor Considerations 20
    Aggregate Central Distributor Costs 20
Process Modifications 22
  Data Focus 22

Data Acquisition 22
Processing 22
Data Elements 23
Data Reviews 24
File Storage and Distribution 24
Eliminate State Data Returns 25
Central Distributor 25
Using Quarterly Data 25
Raw Statistics 25
Estimates 27
Appendix A – Suggested Data Elements for Quarterly Data Files A-1
Appendix B – Suggested Data Quality Tests B-1

Index of Tables

Table 1. Data and Program Files Created During Annual Data "Processing" for Each State 10
Table 2. Cost Estimates for Quarterly Data by Data Type for One State (2009 Dollars) 11
Table 3. Data Purchase Costs, 2007 Hospital Data (2009 Dollars) 13
Table 4. Processing Teams Needed for Quarterly Data 17
Table 5. Incremental Costs for Processing Quarterly Data (2009 Dollars) 18
Table 6. Estimated Central Distributor Costs for Bi-Annual Releases (2009 Dollars)..... 21
Table 7. Estimated Central Distributor Costs for Quarterly Releases (2009 Dollars) 21
Table 8. Pivot Table Example: Discharges by Age and Diagnosis CCS..... 26
Table 9. Pivot Table Example: Discharges by DRG and Diagnosis CCS 26
Table 10. Pivot Table Example: Discharges by Diagnosis and Procedure 27
Table 11. Suggested Data Elements, Core File A-1
Table 12. Suggested Data Elements, Severity File..... A-8
Table 13. Suggested Data Elements, Charges File A-11

Index of Figures

Figure 1. Availability of Quarterly Data as Reported by HCUP Partners 3
Figure 2. Timeline for Processing Q1 Data 5
Figure 3. Timeline for Processing Q4 Data 5
Figure 4. Hospital Additions and Deletions, 2004-2007 6
Figure 5. Timeline for Data Available within Five Months of a Quarter’s End 7
Figure 6. Timeline for Data Available within Seven Months of a Quarter’s End..... 8
Figure 7. The HCUP Data Flow - Current Process 8
Figure 8. Estimate of Quarterly and Annual Files Created by Month..... 14

Executive Summary

This report summarizes the issues and estimated costs of using quarterly data from Healthcare Cost and Utilization Project (HCUP) Partners. This activity is part of the broader task of creating Rapid-Cycle National estimates from inpatient (IP) and emergency department (ED) data collected as part of the Healthcare Cost and Utilization Project (HCUP), sponsored by the Agency for Healthcare Research and Quality (AHRQ). HCUP provides researchers, policymakers, and stakeholders with the tools they need in order to conduct data-driven analyses on a wide range of health care issues.

With each passing year, health care becomes a larger segment of the U.S. economy and the availability of timely information becomes more important. HCUP provides vital data for informing health policy. The value of HCUP data is unprecedented and continues to grow in popularity.

This report follows an earlier effort that assessed HCUP Partners' ability to provide quality quarterly data. The initial feasibility study evaluated the capacity of HCUP Partners; 34 of 40 states could provide good quality data at more frequent intervals than an annual cycle.

This report describes the infrastructure and organizational impacts necessary for using quarterly data to generate Rapid-Cycle estimates. The quarterly data efforts described are complementary additions to the present work with annual data files. Duplicating the process used for annual data, however, even with modifications will be costly. The estimated cost for processing data from 39 states is more than three million dollars. Such an effort would more than double the current data acquisition budget and triple the data processing budget. The basis for these cost estimates as well as suggestions for activities and methods that can reduce costs follow.

Introduction

This document follows a recent summary by Thomson Reuters that determined the ability of current HCUP Partners to provide more timely data (State Quarterly Data Evaluation Report, Deliverable #825.21A). This report details the issues and costs associated with obtaining and processing quarterly data are examined.

The Healthcare Cost and Utilization Project (HCUP) is a family of health care databases and related software tools and products developed through a Federal-State-Industry partnership and sponsored by the Agency for Healthcare Research and Quality (AHRQ). 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 discharge-level health care data. HCUP includes the largest collection of longitudinal hospital care data in the United States, with all-payer, encounter-level information beginning in 1988. These databases enable research on a broad range of health policy issues, including cost and quality of health services, medical practice patterns, access to health care programs, and outcomes of treatments at the national, state, and local market levels.

As mentioned above, this report presents a summary of potential issues and estimated costs for obtaining, processing, and using quarterly data from existing HCUP Partners. The quarterly data efforts described here are complementary and thus supplement the present work with annual data files. The broader task for the Rapid-Cycle National Estimates Team is to develop more timely nationwide and regional estimates from inpatient (IP) and emergency department (ED)

data. Therefore, this report also describes the infrastructure and organizational impacts of using quarterly data to generate those estimates.

Background

As a follow-up to a previous study regarding the feasibility of "collecting" state quarterly data from HCUP Partners, AHRQ tasked Thomson Reuters to "develop and test" the feasibility of using quarterly data for HCUP processing as a way to produce timelier estimates of care. Due in part to the recent findings from the State Quarterly Data Evaluation Report and discussions with the Rapid Cycle Estimates Team, the context and goals for the current task have evolved with a broader scope which involves using the timelier data to create the HCUP databases.

The following report details the proposed infrastructure required to accomplish the objectives mentioned above. It will focus primarily on proposed changes to current data acquisition and processing efforts, as well as associated costs and staffing, to test the feasibility of processing quarterly data **in addition to** the annual files. Given the complexities involved in this effort, Thomson Reuters will consider four different scenarios developed by AHRQ as a framework for this report:

1. Quarterly state intramural files (fully documented)
2. Quarterly state intramural files (Scenario 1) + quarterly Central Distributor files
3. Quarterly state intramural files (Scenario 1) +semi-annual Central Distributor state files (instead of quarterly)
4. Use quarterly data collected as part of Scenario 1 for special time-sensitive adhoc research projects (use the rotavirus study as the prototype).

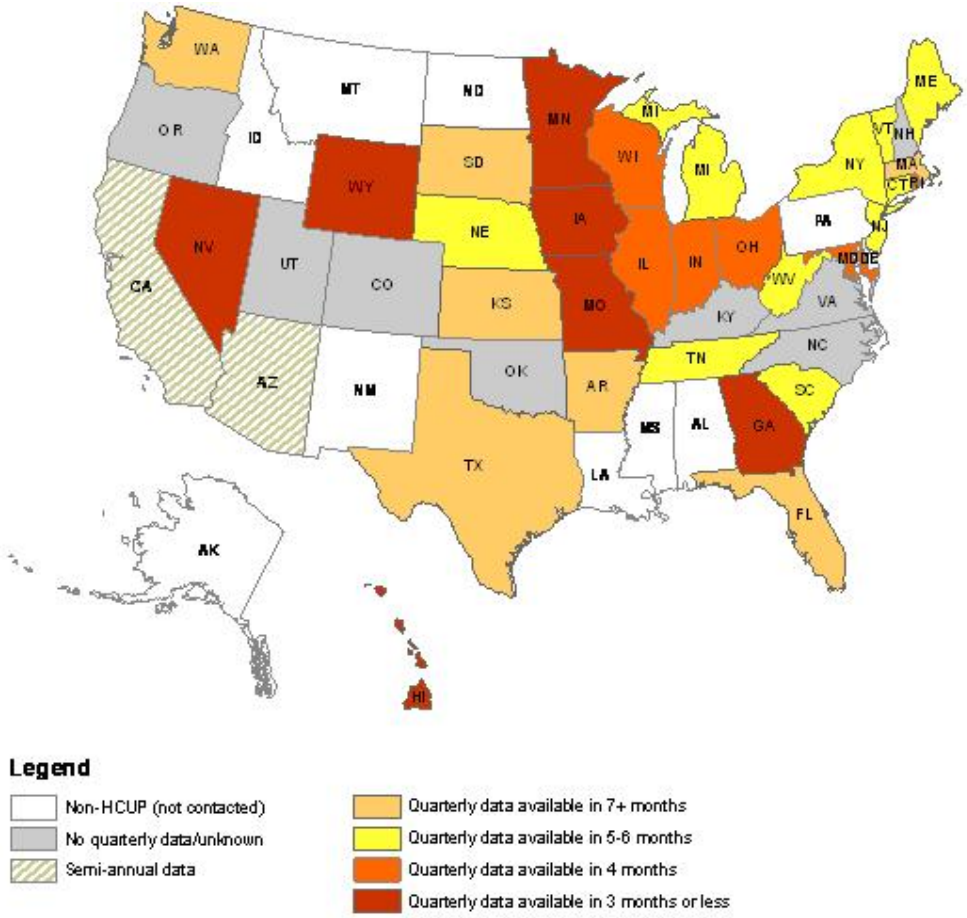
Availability of Quarterly Hospital Data

In an effort to better understand the feasibility of collecting quarterly data from HCUP Partners, Thomson Reuters conducted a previous study to learn more about the completeness, timeliness, and availability of their state quarterly data and the efficacy of collecting those data (HCUP Deliverable 825.21A).

To accomplish these objectives, Thomson Reuters reviewed information obtained through the Annual Data Assessment sent to HCUP Partners in January 2007, and conducted a survey of HCUP Partners in February 2009 to supplement the information about the availability of quarterly data. In addition, quarterly inpatient data provided by 18 HCUP Partners for a special Rotavirus study was compared with annual data they submitted to HCUP.

Thirty-six of 40 HCUP Partners responded to the survey. While some isolated differences existed related to data types, completeness, and quality, approximately 80% of HCUP Partners (32 states) were identified as collectors of quarterly inpatient data. Two states (Arizona and California) collect bi-annual data.

Figure 1. Availability of Quarterly Data as Reported by HCUP Partners



Thirty-one states provided typical availability dates for the quarterly data, with lag time ranging from 2 months to 14 months after the end of the quarter. Of these 31 states:

- 8 reported they can consistently provide data within three months from the end of the quarter
- 13 reported they can consistently provide data within four months from the end of the quarter (the 8 states within 3 months plus 5 additional states)
- 23 reported they can consistently provide data within six months from the end of the quarter (the 13 states within 4 months plus 10 additional states).

These availability times must be interpreted cautiously as there is often a discrepancy between the date that an HCUP Partner indicates annual files are to be available in a year and the actual time HCUP receives them. On average, HCUP received the 2007 annual data 1.6 months after the scheduled data, with a range of 1 month before to six months after the scheduled date.

The results from the analysis seem promising as they relate to the availability of quarterly data; however questions remain about the quality of those data. Some quality issues reported by HCUP Partners included troublesome data elements, incomplete hospital reporting, and missing hospitals. The primary quality concern was related to data resubmissions from hospitals which inevitably delay the data production schedule.

Timing Implications

The earlier availability of quarterly data impacts the type of data available (Inpatient (IP), Ambulatory Surgery (AS), or Emergency Department (ED)) and may also impact other HCUP tasks such as the NIS.

Timing-Related Differences between Quarterly and Annual Data

Processing quarterly data raises a number of timing issues that do not currently impact annual data. Some of these concerns affect the availability and creation of certain data elements when processing quarterly data. In addition, release of some products using quarterly data may complicate or distract from other HCUP products, such as the NIS.

Availability of Quarterly Data

Processing quarterly data raises a number of timing issues that do not impact annual data. Some coding requirements are defined and announced in advance of use. ICD-9-CM diagnoses and procedures codes are an example. These standards are implemented on October 1st but are communicated several months beforehand. Other coding systems are not available until after a calendar year has begun. The All-Payer Severity Adjusted DRG (APS-DRG) for the calendar year is generally available in March, and the All Patient Refined DRG (APR-DRG) for the first quarter is generally available in May. However, because of the expected time lags for receiving data from HCUP Partners, these are not expected to pose problems. Figure 2 and Figure 3 illustrate significant events in processing quarterly data. For demonstrative purposes, these timelines assume that only data received within seven months of a quarter's end are processed (these are the "yellow," "orange," and "dark orange" as stated in Figure 1. In this case, processing of quarter one data occurs from June through November, with the last data sets available on the Central Distributor in December. Processing of fourth quarter data takes place from March through August of the following calendar year, with the last data sets available on the Central Distributor in September.

Figure 2. Timeline for Processing Q1 Data

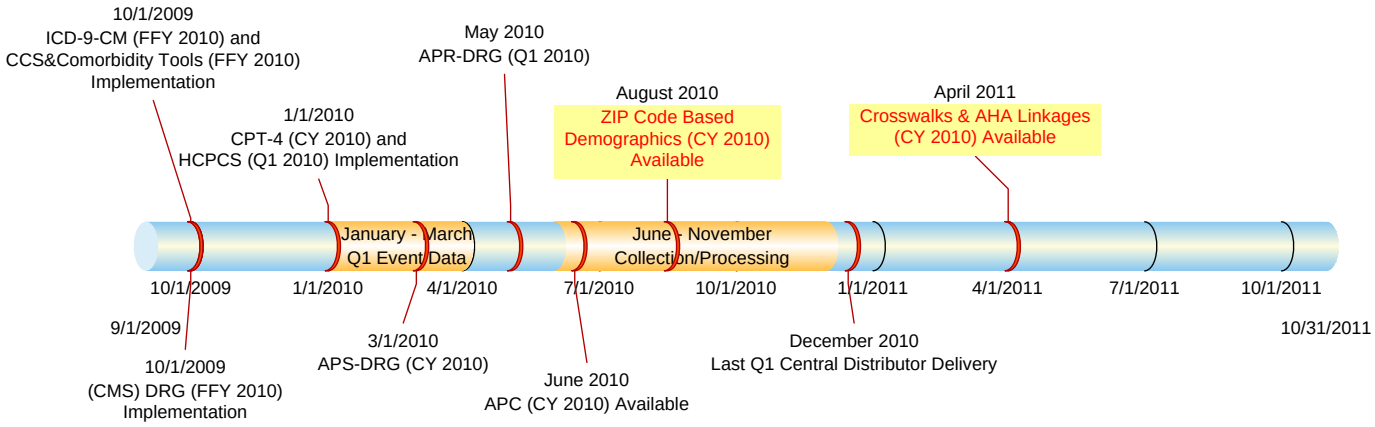
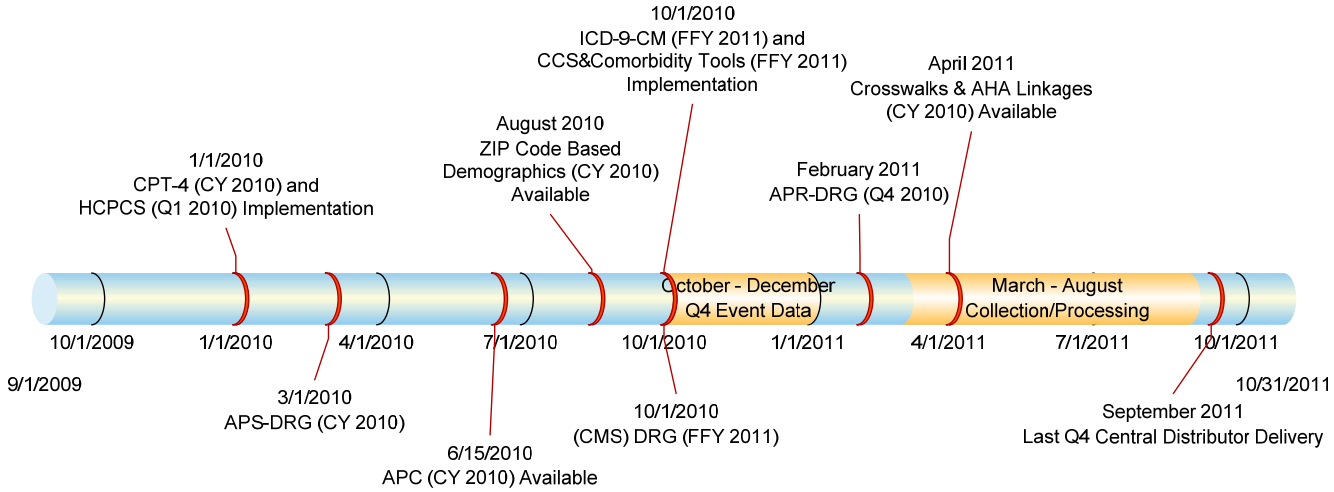


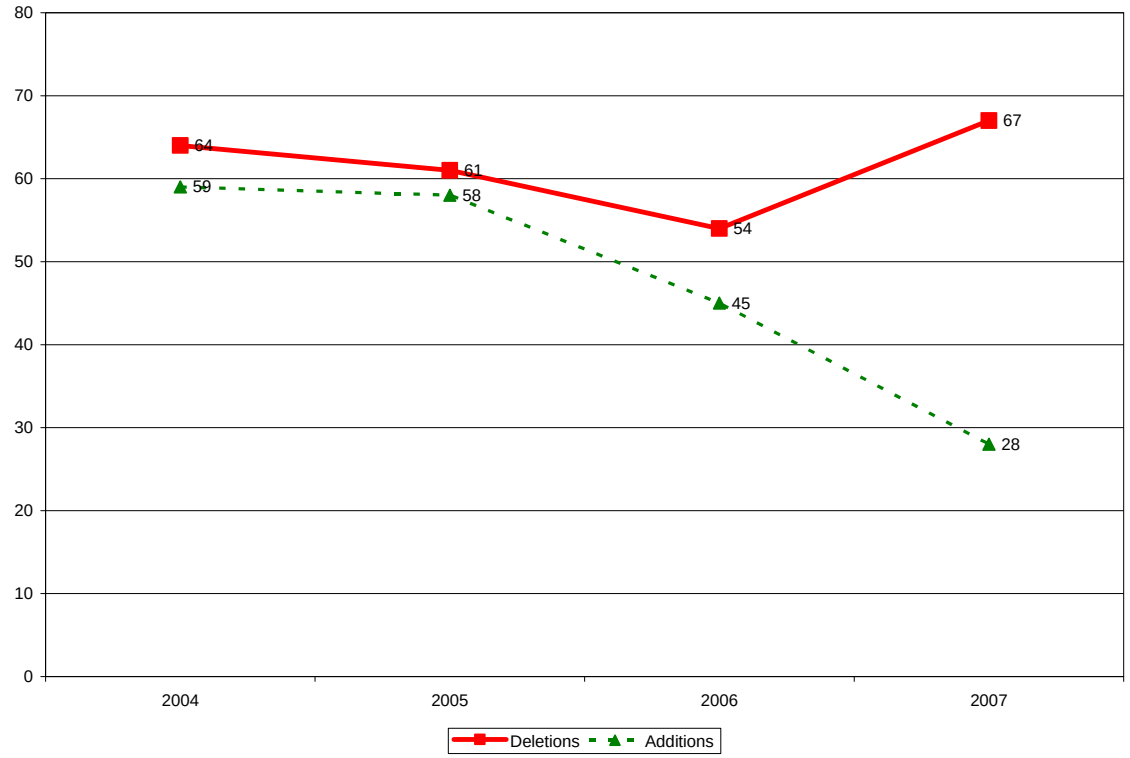
Figure 3. Timeline for Processing Q4 Data



ZIP Code-based demographics could potentially impact delivery of early Q1 data, but should not affect other quarterly data. Related data elements include income quartiles.

In addition, crosswalks based on AHA data are not available until nearly two years after the first quarter's end. This, however, is not an issue with annual file, nor should it be an issue with quarterly data. U.S. community hospitals are a relatively stable population. Historically, the number of hospitals added to or dropping from HCUP have accounted for approximately one percent of the community hospital universe each year (for calendar year 2007, there were 5708 hospitals). Figure 4 illustrates the number of new and discontinued hospitals from 2004 through 2007 (source AHA Annual Survey data).

Figure 4. Hospital Additions and Deletions, 2004-2007



The annual AHA survey is normally available toward the end of the year following the data year. For example, most data for 2008 hospitalizations becomes available during 2009, but the AHA data for 2008 does not become available until November of 2009. Crosswalks can then begin to be created. This process may take weeks depending on the extent of changes and number of hospitals in a state. Because crosswalks are governed by the annual availability of AHA data, they have been decoupled from activities associated with processing annual data. Preparing crosswalks is one of the key steps in creating the NIS, but has no bearing on release of processed state data.

Quarterly Data or Statistics and NIS Data

Depending on the release deadline selected for collecting quarterly data, quarterly data may have timing implications for other HCUP work, primarily NIS data. Below, we highlight two examples of using quarterly data to generate National estimates with data available:

1. Within five months of a quarter's end
2. Within seven months of a quarter's end.

The following timelines demonstrate the potential timing of HCUP events. Both timelines assume:

- A two month lag between processing the quarterly data and generating estimates
- Quarterly data available within the stated timeframe can generate relatively accurate estimates – an as yet unproven assumption
- A NIS release in June.

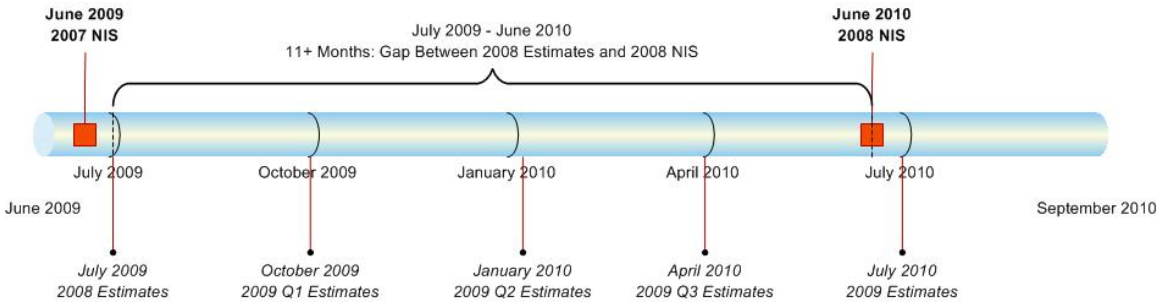
Data Available within Five Months of a Quarter's End

Seventeen Partner states report that their data is available within five months of the quarter's end[1]. In this scenario, the four quarters of 2009 data available from those 17 states by May 2010 can be used to generate full-year 2009 estimates in July 2010. Were this process available right now, 2008 estimates would be available approximately:

- One month after release of the 2007 NIS
- Eleven months before release 2008 NIS.

The impact of announcing 2008 estimates one month after releasing the 2007 NIS is unknown. It could potentially reinforce the importance of the NIS data or alternatively, it may distract from and diminish the importance of the 2007 NIS. Figure 5 depicts the timeline for this scenario.

Figure 5. Timeline for Data Available within Five Months of a Quarter's End



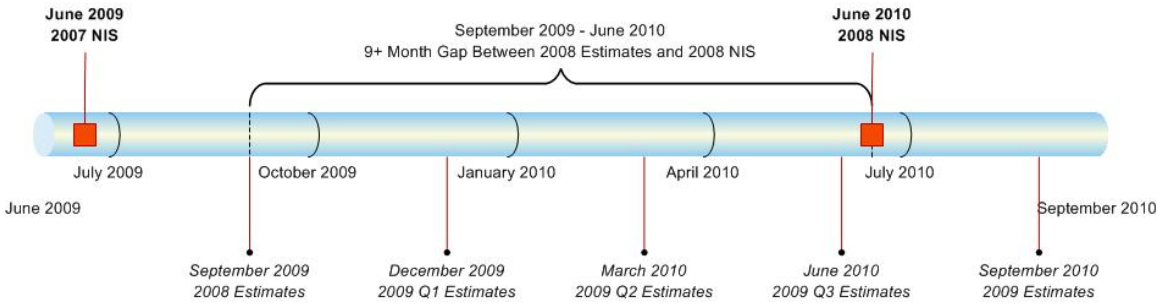
[1]The "dark orange" and "orange" states from Figure 1.

Data Available within Seven Months of a Quarter's End

Extending the timeframe to data available within seven months of a quarter's end increases the number of Partner states to 24[2]. In this scenario, illustrated Figure 6, the four quarters of 2009 data available from the 24 states by July 2010 can be used to generate full-year 2009 estimates in September 2010. Were this process available right now, 2008 estimates would be available approximately:

- Three months after release of the 2007 NIS
- Nine months before release 2008 NIS.

Figure 6. Timeline for Data Available within Seven Months of a Quarter's End

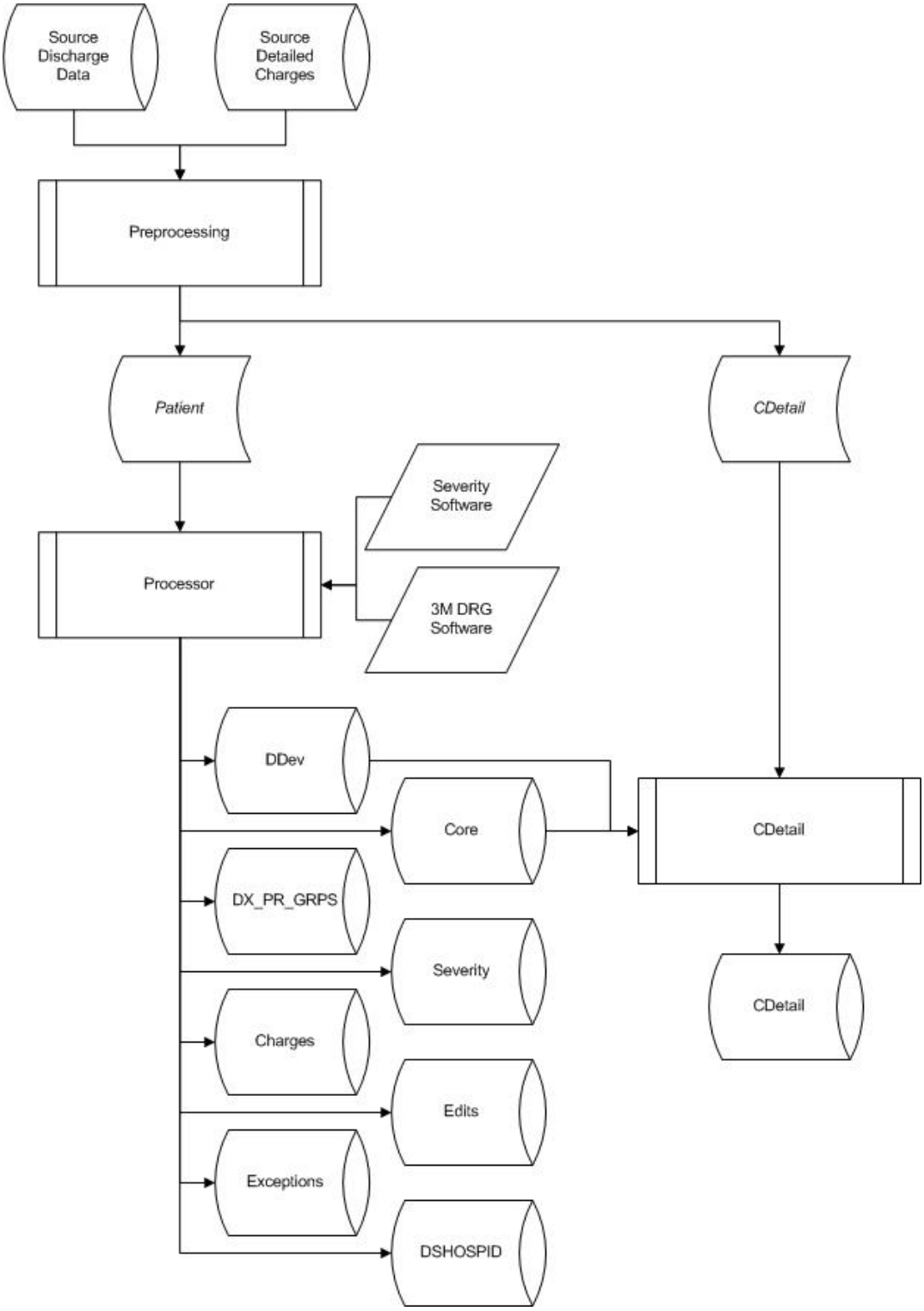


HCUP Processing Infrastructure

Since 1992, Thomson Reuters (nee Medstat) has introduced measures that dramatically increased the efficiency of creating uniform health care research databases (e.g. the SID, SASD, and SEDD). The current protocol for processing source data is an outgrowth of HCUP history. The processor traces its roots to the PL/I processor running on large mainframes – the only computers capable of processing and storing large numbers of hospitalization records – used in building the early State Inpatient Databases (SID). The PL/I processors gave way to stand-alone SAS processors. Mainframe usage was expensive and eventually the capacity of PCs grew into the ability to handle large numbers of records. This is when processing was moved to PCs. Eventually the standalone processor evolved into today's modular system. Together, these modifications essentially slashed the necessary labor to process a state's data from 3 months to 3 weeks to 10 days.

[2]The "dark orange" and "orange" states along with most of the "yellow" states from Figure 1.

Figure 7. The HCUP Data Flow - Current Process



Current Processing Stream

The current processing stream, based on annual data files, involves a large number of intermediate and deliverable files. The number totals nearly 300 files for a state supplying all three data types (inpatient, ambulatory surgery, and emergency department) with detailed charges. This count ignores the hundreds of custom SAS programs utilized for processing.

Table 1. Data and Program Files Created During Annual Data "Processing" for Each State

	Inpatient Data	Ambulatory Surgery Data	Emergency Department Data
SAS Data Sets (files)			
Intramural files	11	10	10
State return files	6	5	5
AHRQ delivery files	6	5	5
SSS files	5	4	4
SAS Programs			
Intramural files	19	18	18
State return creation programs	2	2	2
Central Distributor programs	4	4	4
SAS *CD* load programs	5	5	5
SPSS *CD* load programs	5	5	5
Documentation files			
Intramural files	20	19	19
State return files	6	5	5
Central Distributor files	7	7	7
Other files			
Text (csv) *State Return* files	6	5	5
Text *Central Distributor* files	5	5	5
Total	*107*	*99*	*99*

While responsibilities for most of the file logistics are handled by Thomson Reuters, there is significant burden placed upon staff at AHRQ, SSS, and the Partners. With annual data:

- Each Partner currently receives up to 18 files annually
- SSS currently receives more than 400 files annually which are eventually loaded on to servers
- AHRQ currently receives more than 400 files annually which are eventually loaded on to servers and another 89 data development files managed by the Government Project Officer (GPO).

Changes Necessary to Accommodate Quarterly Data

Overall, the estimated cost of processing quarterly data with no Central Distributor files is $34,100 per state and data type, or $102,300 for a state with inpatient, ambulatory surgery, and emergency department data. Table 2 breaks down this estimate by activity.

Table 2. Cost Estimates for Quarterly Data by Data Type for One State (2009 Dollars)

Activity	Cost Estimate	
	Single Data Type (IP, AS, or ED)	Three Data Types (IP/AS/ED)
Data Acquisition	$5,100	$15,300
Data Purchase Cost	$8,600	$25,800
Production Processing	$19,300	$57,900
Data Delivery	$1,000	$3,000
HCUP-US Documentation	$100	$300
Sub-Total	***$34,100***	***$102,300***

We project costs for processing quarterly data from five states at $530,520. Using the current procedures for 35 states with quarterly data is estimated to cost more than $3 million.

Assumptions

In arriving at these estimates, we made several assumptions.

1. The files and methods employed will be similar to those currently used for processing annual files.
2. Processing quarterly data will supplement, rather than replace processing of annual data files. At some point, if demonstrations and analysis show quarterly data to be accurate and reliable, quarterly data may replace annual data. That point, however, is in the more distant future.
3. Processed quarterly data will differ from processed annual data in that the former will not include severity measures. The severity programs used for HCUP data are provided by a vendor and are not finalized and available to Thomson Reuters until four months after the calendar year ends.
4. Based on experience, we expect some significant data change (e.g., a new variable, or a layout change) from Partners each year. However, we anticipate limited data changes from Partners during a year.
5. Any quarterly data placed in the Central Distributor will exclude AHA-based hospital crosswalks. AHA data is available annually, generally eleven months after the calendar year ends. Older crosswalks to the AHA files, although incomplete, may be used to calculate weights and provide statistics by basic hospital characteristics.

Data Creation Costs

Data Acquisition

Data Acquisition is the first step in the HCUP data process. Quarterly data necessitates repetition of most of these steps and multiplication of the associated costs. Beyond internal HCUP implications, supplying quarterly data increases Partner costs, primarily in terms of time. For this reason, we expect Partners will minimize data changes during a year in an effort to minimize their costs. This is the basis for our assumption/expectation of only one significant change per year.

Processing quarterly data has several *Data Acquisition* implications including ongoing tasks necessary for acquisition of quarterly data. Some modifications may be needed to existing contractual documents including Memorandum of Agreements (MOAs), Data Use Agreements (DUAs), and Applications. Currently, each contractual relationship between HCUP and participating Partners is essentially dictated by both the state and the HCUP Team to create a customized partnership unique for each state. After the data has been obtained and loaded, subsequent communications with the Partners often involves questions around verification of data quality and consistency. For some of these processes mentioned above including follow-up questions and communications to resolve data problems, effort and cost are expected to be straight multiples. However, we expect to find economies in other more involved tasks, such as MOA amendments. The first step necessary in obtaining quarterly data will require a one-time update to the MOA establishing the conditions for using the Partner's quarterly data. Based on recent charges, we project a cost of $1,000 per Partner. Overall, we project quarterly data acquisition activities to require 2.5 times the effort as annual data acquisition work. Based on current charges, this is approximately $5,100 per state and data type per year, or $15,300 per year for a state with inpatient, ambulatory surgery, and emergency department data.

Data Purchase Costs

At AHRQ's request, Thomson Reuters did not query states on the cost of obtaining quarterly data. Partner data organizations may include quarterly data in the purchase of annual data files, or they may increase purchase costs to cover the additional files. For these estimates, data purchase costs were assumed to be the average cost of annual data for the thirteen states that report data availability within four months of a quarter ending[3]. For reference, data purchase costs range from zero up to $24,340 for a single (inpatient) data type. Table 3 summarizes purchase costs for 2008 data. The overall average for all three data types is $8,609[4] per file or slightly more than $25,800 for all three data types.

---

[3]The "orange" and "dark orange" states from Figure 1.

[4]Six Partners report AS data available within four months of a quarter's end; seven Partners report ED data available within four months of a quarter's end.

Table 3. Data Purchase Costs, 2007 Hospital Data (2009 Dollars)

	IP Data	AS Data	ED Data
Average	$9,197	$6,796	$9,069
Minimum	$0	$0	$0
Median	$8,000	$8,375	$8,000
Maximum	$24,340	$11,000	$30,383

These averages are for 2007 data purchases and are expected to increase in future years. For the period spanning the last three data years, the cost of inpatient data has increased at an annual rate of 10.1 percent. Ambulatory surgery and emergency department data costs have demonstrated more stability, growing at annual rates of 1.2 and 2.2 percent, respectively.

Processing

Processing data constitutes the largest component of time and effort required to create discharge data files. HCUP data processing is a complex, labor intensive effort. Separate processors are created for each state and data type. In processing data for a single state and data type, the primary processing program dynamically invokes more than100 different sub-modules which can generate over 10,000 lines of code and create nearly 90 programming files each year. Much of the code generation is accomplished behind the scenes. This process has been optimized for annual data. The HCUP processor has evolved into its current complexity for a number of reasons.

1. It was designed to identify and accommodate more than 300 different data elements that Partners may provide.
2. It was required to accept numerous formats and respond to new, modified, and discontinued data elements.
3. It was expected to ensure all data use restrictions are observed.
4. It was intended to be easy to use so that Junior staff can process data.
5. It performs extensive validity and edit checks on many variables.
6. It adds much derived information, such as severity measures.

The result is a process with complex, internal decision algorithms for creating a uniform, standardized annual data file, but also a process that is difficult to modify.

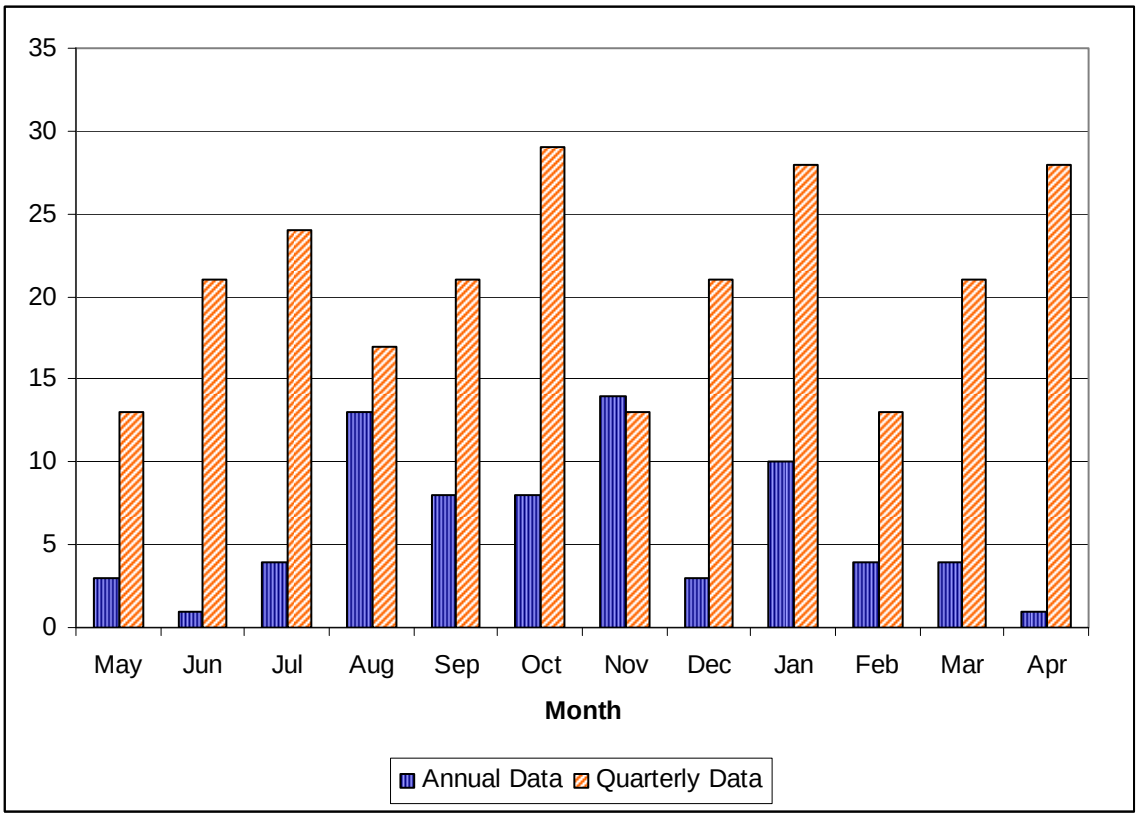
As currently structured, data files are available for use approximately one month (27 days) after received from the Partner. As described in the *Data Acquisition* discussion, we expect some efficiency gain with quarterly data due to the shorter period between data submissions. The more frequent provision of quarterly data may decrease the likelihood that Partners change data formats or coding during the limited span between files. Overall, we project quarterly data processing activities to require 2.75 times the effort as annual data processing work. Based on current charges, this is approximately $19,300 per state and data type per year, or $57,900 per year for a state with inpatient, ambulatory surgery, and emergency department data.

Data Delivery

Data delivery costs are a minor factor in the HCUP budget. The HCUP processing programs automatically create both AHRQ and State Return files. Based on recent charges, state returns including documentation should cost $2000 per state and data type per year, or $6,000 per year for a state with inpatient, ambulatory surgery, and emergency department data. We expect the cost of AHRQ data deliveries at one-half the amount spent for state returns, or $1,000 per state and data type per year, or $3,000 per year for a state with inpatient, ambulatory surgery, and emergency department data.

These cost estimates make no considerations for incremental costs imposed on AHRQ and SSS associated with receiving and loading additional files[5]. Adding quarterly files increases effort required for managing data storage. Based on Partners' claims about their availability of quarterly data, the number of files delivered only to AHRQ would increase by more than three-fold, with the typical (median) monthly delivery increasing from four files to 21 files. Figure 8 compares the number of annual files created and delivered with the number of quarterly files projected based on Partner estimates. Such costs are not reflected in these budget projections.

Figure 8. Estimate of Quarterly and Annual Files Created by Month



[5]For purposes of costing these activities, we assume that Partners deposit returned files in a large desk drawer and incur no meaningful budgetary burden.

HCUP-US Documentation

Most database documentation is generated from executing the processing programs. Costs associated with HCUP-US updates involve copying processing specification notes and transferring summary statistics and xml documentation files to the HCUP-US server. These involve rather minor costs of approximately $100 per state and data type, or $300 per year for a state with inpatient, ambulatory surgery, and emergency department data.

Data Destruction

The HCUP contract mandates that “at the direction of AHRQ, all source data received by the Contractor from HCUP State Partners shall be destroyed approximately two years after completing the HCUP data files …” It is further specified that “certification of destruction shall be sent to the HCUP Project Officer and the HCUP State Partners …”

In practice, these requirements are fulfilled using the following approaches:

- Thomson Reuters (TR) notifies the Project Officer when two years have elapsed since the last file for a year was produced and asks if AHRQ wishes to proceed with destroying the source data for that year
- When directed to proceed, TR queries the data tracking system to list files for deletion, deletes source files from the servers, erases and recycles or bags for destruction by a certified destruction service all backup tapes of source data, inventories and bags the source media for destruction, and updates the tracking system with data destruction dates
- TR sends the Project Officer a memo certifying the destruction of the source data along with inventory listings of all source media and backup tapes
- TR assists the Project officer in the preparation and mailing of a notification to the State Partners.

The most recent destruction of annual source data required a total of 24 hours at a cost of approximately $2,500. This represents a mix of relatively fixed cost administrative work plus a component that varies by data set. We estimate this cost at $1700 plus $10 per data type.

Other Cost Considerations

Infrastructure

The initial impact of adding quarterly data from one or two states would be minimal. Current server and storage capacity, along with planned enhancements, are more than adequate for the additional files.

In the long run, however, quarterly data would have a substantial impact on storage and costs. Adding all possible quarterly data would nearly double the amount of space and quadruple the number of files managed by the project. Currently, HCUP data is processed on four different state servers. The four state processing servers are already loaded with maximum numbers of disk drives they can physically accommodate, and those servers have relatively little excess capacity. For that reason, we believe that quarterly data will eventually require additional hardware when the number of participating Partners reaches four states.

The cost for a Windows server with an attached RAID array for storage has an estimated cost of $18,000 (based on recent purchases and quotes). This cost does not include SAS subscription fees which are currently covered under the HHS license.

- $4,500 for server hardware
- $1,500 for the operation system
- $12,000 for the RAID array.

The $18,000 is amortized over three years for a cost of $6,000 per year for three years[6]. In addition, each server requires further labor for maintenance: approximately $3,000.

A single server should be able to accommodate the storage and processing requirements for up to 15 states with quarterly data. Managing 16 to 30 states with quarterly data will necessitate a second server, while more than 30 data feeds will most likely require a major redesign of the storage and processing infrastructure based on lessons learned in collecting, creating, managing, and backing up more than 30,000 new separate files each year.

Productivity

As additional servers are added, job management is likely to become a larger issue requiring some form of integrated process administration. Two options are:

1. The SAS product Grid Manager, and
2. Building a custom, HTML-based, job coordinator.

The SAS Grid is not a product covered by the HHS SAS license. While the actual product cost is dependent on the number and type of servers to be managed, the expected range is between $80,000 and $120,000 for the first year with annual renewal fees of $30,000 to $40,000. Building a customized, functionally similar solution involves primarily labor costs of approximately 120 days or $75,000. This is a one-time cost.

Adding quarterly data may decrease programmer productivity if additional processing runs compete for server time and create delays. The smaller size of quarterly files should mitigate this issue to some degree.

Quarterly data will also impact the workload of the four individual processing teams. As currently structured, each team now processes data from nine to eleven states. Given this current workload, we believe that our teams can each accommodate one to three states with quarterly data without affecting processing times. Based on these assumptions, the four available teams can process up to eight Partners with quarterly data before requiring additional staff.

Although we have considered modifications to the team approach, as well as alternatives, we believe that teams are still the most efficient method for processing HCUP data.

- Modifications such as adding additional Programmers or Analysts to teams are problematic. One Analyst managing and reviewing data from two fully utilized Programmers will overwhelm the Analyst, and one Programmer cannot process data from for two fully utilized Analysts in a timely manner.

---

[6]Most likely charged to task C.16 budget: Delivery and Management of Databases, Documentation, and Records.

- An alternative structure, such as staffing pools – where work is assigned to the next available Programmer or Analyst – eliminates the team-based advantage of continuity and familiarity with the Partners and their data.

Our recommendation is to retain a team-structured environment. We find that this structure contributes to quality and building accurate data; it also works well for processing data. Table 4 details the staffing and team requirements.

Table 4. Processing Teams Needed for Quarterly Data

States with Quarterly Data	Teams Required	States per Team
1-8	4	10
9-15	5	8
16-18	6	6-7
19-21	7	5-6
22-32	8	5
33-40	10	4

Aggregated Costs of Creating Quarterly Data

We expect gradual, incremental additions of state quarterly data files over time. At AHRQ's direction, for this budget exercise we assume processing quarterly data will occur at the following pace:

- 5 states in the 2009 data year
- 12 states in the 2010 data year
- 20 states in the 2012 data year
- 30 states in the 2014 data year
- 35 states in the 2015 data year.

To further simplify the exercise, we assume that one half of the quarterly data states provide two data types and the other half of the states provide three data types. Finally, we calculate data destruction costs in the year we process that year's data rather than at a point two or more years in the future. Table 5 summarizes the aggregate cost implications (using constant 2009 dollars).

Table 5. Incremental Costs for Processing Quarterly Data (2009 Dollars)

Data Year	States	Data Acquisition	Data Purchase Costs	Processing	Data Delivery, Documen-tation, and Destruction	Server	Total
2009	5	$66,300	$114,800	$250,900	$16,520	$82,000 *add a fifth processing server add job coordinator software*	$530,520
2010	12 *add a fifth process-ing team*	$153,000	$265,200	$579,000	$35,900	$9,000	$1,042,100
2012	20 *add sixth and seventh process-ing teams*	$255,000	$442,000	$965,000	$58,700	$12,000 *add a sixth processing server*	$1,732,700
2014	30 *add an eighth process-ing team*	$382,500	$663,000	$1,447,500	$87,200	$12,000	$2,592,200
2015	35 *add ninth and tenth process-ing teams*	$448,800	$777,800	$1,698,400	$102,020	$15,000 *add a seventh processing server*	$3,042,020

For perspective, the estimated total cost for five states of 2009 quarterly data ($527,520) is more than 25 percent of the combined Data Acquisition and Processing budgets ($2,087,000). The 35-state estimate for 2015 data ($3,042,020) is 145 percent larger than the combined budgets.

Central Distributor

The preceding estimates exclude Central Distributor costs, at AHRQ's request. These estimates, detailed here, are based on recent charge history for three different HCUP subtasks:

1. Data Acquisition (Central Distributor MOAs) – Task C.7
2. File Creation (Central Distributor files and documentation) – Task C.8

3. Central Distributor (operations and application kits) – Task C.15

Data Acquisition

A Central Distributor release of quarterly data impacts *Data Acquisition* because a Central Distributor MOA amendment is required. Continuing our assumption that Partners will limit major changes to one per year, and based on recent experience, we project a *Data Acquisition* cost of $1,500 per Partner.

Creating Central Distributor Files

Creating Central Distributor files requires extracting a subset of data elements from the Intramural files based on the State's Central Distributor MOA and creating custom documentation for that extracted file. Documentation includes generating custom SAS and SPSS program code for loading the Central Distributor files. Expected costs are relatively small because this process was automated within the past two years; we estimate charges of $700 per state and data type. For bi-annual files this projects to annual costs of:

- $1,400 per state and data type
- $4,200 for a state with inpatient, ambulatory surgery, and emergency department data.

For quarterly files this projects to annual costs of:

- $2,800 per state and data type
- $8,400 for a state with inpatient, ambulatory surgery, and emergency department data.

Central Distributor Operations (SSS)

The final, Central Distributor activity is the operations portion of the *Central Distributor* task. Central Distributor operations include:

- Maintaining the HCUP electronic tracking and distribution system
- Handling Central Distributor inquiries
- Processing Central Distributor applications
- Collecting and managing Central Distributor fees
- Copying CDs and preparing database binders
- Filling and shipping Central Distributor orders
- Preparing Central Distributor application kits (technically, the application kit is separate from the operations task).

Our cost estimates for Central Distributor operations are $2,500 per state data type. For states with inpatient, ambulatory surgery, and emergency department data, this projects to annual costs of:

- $15,000 for bi-annual files
- $30,000 for quarterly files.

This estimate is based on recent charge history and the following assumptions:

1. State data files comprise roughly 55 percent of the Central Distributor operations budget
    - 55 percent of annual Central Distributor operations charges are approximately $263,000
    - Based on recent charges, this amounts to approximately $4,800 per state data type.
2. Demand for quarterly (or bi-annual) Central Distributor files based on quarterly data would be less than demand for annual files. We estimate demand for quarterly data files at 50 percent of demand for annual state files. This reduces the state/data type estimate to $2,400.
   This estimate is based on Central Distributor demand for recently released files:
    - State files (the most recent 2-3 years) account for approximately 61 percent of monthly and year-to-date Central Distributor shipments (ranging from 43 to 88 percent of shipments. We take this as the upper bound of demand for quarterly files.
    - Shipments of the most recent NIS file make up approximately 33 percent of monthly Central Distributor shipments. We take this as the lower bound of demand for quarterly files, although a strong argument may be made that NIS users do not necessarily represent a group interested in timely data.
3. Preparing Central Distributor application kits (this is, technically, separate from the operations task) costs add approximately $100 per state and data type.

While AHRQ incurs some cost in the review and approval of both Application Kits and Central Distributor binders, these are not reflected in our estimates.

Other Central Distributor Considerations

We identified few economies of scale in creating Central Distributor files from quarterly data. In fact, our analysis highlighted data quality issues that could conceivably impose additional costs.

1. Quality concerns and/or incomplete files could increase the number of Central Distributor inquiries fielded by SSS and directly raise costs
2. Quality issues may harm the project's reputation, generating intangible (but very real) costs for both AHRQ and HCUP.

Aggregate Central Distributor Costs

As stated above, we expect incremental additions of state quarterly data files over time. Extending the budget exercise from above – beginning with five states supplying 2009 quarterly data and building up to 35 states with quarterly data for the 2015 data year – we will assume that approximately 75% of the states participate in the Central Distributor. Again, we will assume that one half of the quarterly data states provide two data types and the other half of the states provide three data types. The following tables summarize the aggregate cost implications (constant 2009 dollars) for, respectively, bi-annual and quarterly Central Distributor releases.

Table 6. Estimated Central Distributor Costs for Bi-Annual Releases (2009 Dollars)

Data Year	States with Quarterly Data	Central Distributor States	Data Acquisition	File Creation	Operations	Total
2009	5	4	$6,000	$14,000	$50,000	$70,000
2010	12	9	$13,500	$32,200	$115,000	$160,700
2012	20	15	$22,500	$53,200	$190,000	$265,700
2014	30	23	$34,500	$81,200	$290,000	$405,700
2015	35	26	$39,000	$91,000	$325,000	$455,000

Table 7. Estimated Central Distributor Costs for Quarterly Releases (2009 Dollars)

Data Year	States with Quarterly Data	Central Distributor States	Data Acquisition	File Creation	Operations	Total
2009	5	4	$6,000	$25,000	$100,000	$131,000
2010	12	9	$13,500	$64,400	$230,000	$307,900
2012	20	15	$22,500	$106,400	$380,000	$508,900
2014	30	23	$34,500	$162,400	$580,000	$776,900
2015	35	26	$39,000	$182,000	$650,000	$871,000

Process Modifications

Data Focus

To contain costs, Thomson Reuters proposes that quarterly data focus on a limited set of essential data elements and a simplified data structure. Following this suggestion would result in uniform data sets for each state, even for cases where the Partner does not supply a particular data element. Focused data sets of this type would minimally impact data acquisition but significantly reduce processing and review efforts.

Data Acquisition

In order to streamline the process of acquiring and processing quarterly data, we suggest requesting custom files containing only the necessary data elements. This reduces time necessary to confirm that files contain information needed. Further, we propose minimal Partner contact:

- One data request per quarter, with the preference that Partners automatically send the data when available
- One reminder per quarter
- Ignore delayed data (data received after a set cut-off date)

Based on our experience, we believe that many Partners would prefer standard data orders that can be calendared. In the past, automatic data orders were eschewed to allow documentation reviews searching for new data elements. Focusing on a core set of variables eliminates that need.

We also suggest seeking replacement data requested only for major issues such as when a large hospital is missing or when there are problems with the primary diagnosis, but not for concerns that have little or no impact on statistics. Replacement data must be received by the normal deadline for that quarter's data and there should be only one replacement request per quarter. These assumptions are built into the cost estimates.

Processing

We believe that a few simplifications and improvements to processing can yield significant efficiencies:

1. Focus on a limited set of data elements
2. Automate most data testing.

These steps could greatly reduce labor and speed processing time, reducing the estimated cost of processing by approximately 30 percent. Transforming this efficiency into dollars, the first year's processing estimate of $267,420 (the "Processing" column from Table 5 plus the "Data Delivery, Documentation, and Destruction" column) would be lowered to roughly $187,000. The total estimated cost for year one would drops from $527,520 to $447,100 (a 15 percent decrease).

Data Elements

We recommend creating quarterly data focused on a limited set of data elements. In comparison to the annual data files, quarterly data would exclude both source value data elements (i.e., _X variables) and edit check variables. To enhance the usability of the quarterly data, we suggest creating all quarterly data with the same set of data elements, even for situations in which the Partner does not supply a particular data element. For example, we would include the data element PNUM_R (re-identified patient number) for all quarterly data, even for states that do not supply patient number.

Much of the work processing annual data focuses on variations in data elements. Transforming source discharge files into standardized research data is an especially complex, labor intensive effort as performed for HCUP because the process must be highly flexible and able to identify and respond to unexpected changes in data composition or format and ensure all data use restrictions are observed. The majority of processing costs reflect the effort required to accommodate varieties of data elements and ensure conformance to data restrictions. As currently structured, processing adjusts to whatever data format the Partner chooses to send. The process must then be able to identify and handle any new or changed variables, verify data content and quality, and conform to all data use limitations.

This process requires a great deal of labor and cost. Alternatively, focusing on a core set of data elements eliminates much of this work while retaining important variables most useful for health care analysis. In addition, a core group of data elements:

1. Simplifies data acquisition and MOA efforts
2. Streamlines processing steps
3. Enables automated reviews with a standard set of statistical comparisons.

A limited group of variables is attractive because it is quite informative, and because information represented is very important to hospitals. To be especially efficient, the lengths of the arrays should be set to some constant number for all data sources. We recommend retaining 15 diagnoses, e-codes, and procedures (ICD-9 and CPT/HCPCS). Generally, after about 15 diagnoses and procedures additional codes don't report important new conditions but instead elaborate on already-indicated conditions, or they provide additional detail about comorbidities. However, restricting the number of codes retained in the data does simplify file creation, documentation, and analysis.

Along with cutting back on the number of data elements retained, we also suggest a simplified file structure with three discharge-, or event-level data sets:

- Core file with a limited set of data development variables (ADATE and DDATE only)
- Severity file with the AHRQ comorbidity variables
- Charges file when charges or detailed charges are available from the Partner. Quarterly data processing would transform any normalized charge detail file into an array of revenue and charge information without losing any detail.

A specialized key (QKEY) would enable linking of all files. Appendix A includes a detailed list of data elements for each data file.

Data Reviews

Limiting the number of data elements simplifies data reviews by reducing the volume of data to evaluate. We would further simplify the process by automating the reviews with statistics-based tests. We see the need for three types of tests, although not all three types will be possible for every data element:

- “Content Tests” (within a data type/year) – for example, ensuring that the percent of discharges with a missing value for age is less than a specified threshold
- “Continuity Tests” (within a data type, across years) – for example, comparing the distribution of discharges by age group compared with the same quarter in the previous year’s annual data using a chi-square distribution test or a nonparametric analog such as the Kolmogorov-Smirnov test.
- ”Context Tests” (comparisons to other data such as NIS or AHA data) – for example, a chi-squared or Kolmogorov-Smirnov distribution test of discharges by age group compared with the most recent NIS (age-adjusted to the most recent Census estimate).

Appendix B illustrates the suggested tests for each of the focus variables. These tests would serve as automated tools to focus attention on specific data elements; we do not expect tests demonstrating significant differences to automatically prove data quality problems.

File Storage and Distribution

As described earlier in this report, quarterly data would greatly increase both the quantity of state data files and the amount of time and effort required to manage those files. One solution might involve replacing the current directory-based structure of individual files with a data warehouse. Such an approach, however, would involve substantial investments in data warehouse software and database administrator staffing for two locations. In addition, this approach requires two copies, one for each of the project’s two principal locations: Thomson Reuters and AHRQ. For this reason, we are not currently recommending a data warehouse approach.

If we continue with directory-based storage of state/year/data type-specific files, there are still several proposals that can lessen the time and storage impact from quarterly data:

1. Focus on a core set of data elements (discussed above). This approach eliminates data development files as well as detailed charge files and tools.
2. Delete processed quarterly data at set intervals such as one year (but not before the completeness and quality of the quarterly data has been assessed against the annual data file).
3. Destroy source data on a regular basis. For example, once every month delete source files for any and all states that completed processing during the previous month. Further, we could eliminate Partner notifications that source data was destroyed – the deletion policy can be specified in the MOA and would eliminate a great deal of paperwork.

Eliminate State Data Returns

Eliminating the creation and delivery of the state return files would save small amounts in terms of processing labor, perhaps $100 per data type per quarter. Data management, however, would greatly benefit by reducing the number of files. Because Partners rarely utilize the State Return files, this would likely benefit Partners as well.

The state return files were originally created so states that did not take part in the Central Distributor might distribute the HCUP version of their files. This has rarely been done and it seems unlikely that a state would want to spend the time and resources to receive and manage quarterly copies of files they seldom use, making quarterly State Return files something of a nuisance to Partners.

Central Distributor

We recommend not releasing quarterly data through the Central Distributor. Costs associated with Central Distributor distribution are substantial (see Table 6 and Table 7) yet offer little benefit to the project. Central Distributor costs are not included in the overall cost estimates shown in Table 5.

Using Quarterly Data

Until quarterly data is shown to be consistently of equal quality to annual data files, use of the quarterly data should remain internal to the AHRQ and Thomson Reuters HCUP teams. This data can and should be used for disseminating estimates and summary statistics on various topics of interest. We highlight a few of these ideas below.

Raw Statistics

An important goal for quarterly data is to deliver results quickly. For that reason, delivering complex, patient-level files may be counter-productive because there will inevitably be delays in analyzing files and generating publicly available results. Instead, we propose expediting this process by creating deliverables in a form that is immediately usable and distributable. Examples might be:

- National estimates
- Counts fed to a query system like HCUPnet
- Counts in Excel pivot tables that allow quick analysis of summary data
- XML-formatted counts available to the public for use in “mashups.”

An important advantage of these approaches is that none release patient-specific data. As such, they are not subject to such stringent confidentiality protections[7]. They could be delivered over the Internet and made widely available. Not only is this a faster way to deliver results, but it also helps address the heavy logistical demands of delivering so much data.

[7]Small cell-size restrictions would still apply.

Pivot tables represent an attractive delivery method because they are informative and easy to use. A single pivot table filled with only summarized, non-sensitive data can generate dozens of different summary reports. The next three tables (Table 8, Table 9, and Table 10) demonstrate a trivial example using 2007 Washington SID data.

Table 8. Pivot Table Example: Discharges by Age and Diagnosis CCS

2007 Washington Inpatient Summary

Discharges	**AGE**									
DXCCS1	0	01-17	18-34	35-44	45-54	55-64	65-74	75 plus	missing	**Total**
101			40	427	1,915	3,452	3,534	3,427		*12,795*
106	25	78	314	422	1,091	1,829	2,446	4,912	1	*11,118*
108	8	10	175	351	969	1,699	2,427	7,491		*13,130*
122	509	1,318	613	782	1,493	2,180	2,702	6,841	1	*16,439*
193		623	14,960	2,068	19					*17,670*
195		459	10,171	3,344	45	2				*14,021*
203		3	62	435	2,456	5,491	5,902	4,706		*19,055*
205	1	33	850	1,924	3,061	2,781	2,177	1,924		*12,751*
218	83,734									*83,734*
237	59	380	706	874	1,736	2,367	2,281	2,685	2	*11,090*
OTH	10,027	24,832	87,493	47,743	56,805	58,519	54,418	93,202	60	*433,099*
Total	94,363	27,736	115,384	58,370	69,590	78,320	75,887	125,188	64	*644,902*

Table 9. Pivot Table Example: Discharges by DRG and Diagnosis CCS

2007 Washington Inpatient Summary

Discharges	**DXCCS1**											
DRG	**101**	**106**	**108**	**122**	**193**	**195**	**203**	**205**	**218**	**237**	**OTH**	***Total***
89				9,076							141	*9,217*
127			8,359								419	*8,778*
371					4	2,066					12,227	*14,297*
373					11,794	5,418					19,685	*36,897*
390									8,339		951	*9,290*
391									47,835		938	*48,773*
430											9,853	*9,853*
544							12,305	1		76	2,397	*14,779*
775					3,812	1,672					6,213	*11,697*
795									15,453		195	*15,648*
OTH	12,795	11,118	4,771	7,363	2,060	4,865	6,750	12,750	12,107	11,014	380,080	*465,673*
Total	12,795	11,118	13,130	16,439	17,670	14,021	19,055	12,751	83,734	11,090	433,099	*644,902*

Table 10. Pivot Table Example: Discharges by Diagnosis and Procedure

2007 Washington Inpatient Summary

Discharges	**PRCCS1**											
DXCCS1	45	124	134	137	140	152	153	216	222	228	OTH	**Total**
101	5,654							13	46	7	7,075	*12,795*
106	36						1	67	116	13	10,885	*11,118*
108	68						1	348	511	24	12,178	*13,130*
122	5							548	611	63	15,212	*16,439*
193			7	7,751	8,427				1	3	1,481	*17,670*
195		14	3,685	4,389	1,231			4	21	1	4,676	*14,021*
203	1					11,433	6,195	3	7		1,416	*19,055*
205		1					1	6	23	4	12,716	*12,751*
218								1,435	18	10,729	71,552	*83,734*
237	817					857	926	55	130	1	8,304	*11,090*
OTH	4,439	12,125	20,954	19,289	4,210	670	2,893	10,814	9,244	501	347,960	*433,099*
Total	11,020	12,140	24,646	31,429	13,868	12,960	10,017	13,293	10,728	11,346	493,455	*644,902*

Estimates

In earlier efforts, Thomson Reuters investigated methods for creating early estimates (*Rapid Cycle Estimates Test Methods Report – Deliverable #825.11B*). The two methods for early estimates tested in that study proved somewhat disappointing using annual discharge data. Therefore, rather than expend more effort on improving estimates using annual data, AHRQ and Thomson Reuters researchers concluded that we should shift our attention to developing estimates using early quarterly data. One advantage of early quarterly data is that it is potentially available from many more states than those for which we can obtain early annual data. We anticipate that a broader array of states will be more representative of the overall NIS trends and therefore lead to better early estimates across the entire panel of outcomes and conditions.

We recommend that AHRQ evaluate the feasibility of generating early national estimates from quarterly data. One approach is to investigate methods similar to those tested with early annual data. Another method is to construct an econometric model that incorporates previous NIS estimates, demographic estimates from the U.S. Census Bureau, as well as cost and usage changes observed in early quarterly data.

Testing of various methods can be accomplished using quarterly data previously collected as part of the Rotavirus task, supplemented with annual SID data divided into quarters (to simulate quarterly data).

APPENDICES

Appendix A – Suggested Data Elements for Quarterly Data Files

Table 11. Suggested Data Elements, Core File

Variable Names	Description	Def	Value	Value label	Comments
ADATE	Admission date		YYMMDD . .A .C	Date of admission Missing Invalid Inconsistent	Confidential data development file – **this data element must be removed from the data if quarterly data is released.**
AGE	Age in years at admission	3	0-124 .	Age in years Missing or invalid	Prefer to calculate from dates. Use reported if cannot calculate.
AMONTH	Admission month	3	1-12 . .A .C	Admit month Missing Invalid Inconsistent	
ASOURCEUB92	Admission source, UB92 standard coding	3	Non-newborn admissions (ATYPE ≠ 4): 1 2 3 4 5 6 7 8 A Blank Newborn admissions (ATYPE = 4): 1 2 3 4 Blank	 Physician referral Clinic referral HMO referral Transfer from hospital Transfer from a skilled nursing facility Transfer from another health care facility Emergency room Court/Law enforcement Transfer from a Critical Access hospital Missing or invalid Normal newborn Premature delivery Sick baby Extramural birth Missing or invalid	

Variable Names	Description	Def	Value	Value label	Comments
ATYPE	Admission type	3	1 2 3 4 6 .	Emergency Urgent Elective Newborn Other Missing or invalid	Code deliveries under Urgent (2)
CPTn	CPT-4/HCPCS procedures	$5	5(a) Blank	CPT or HCPCS Procedure Codes 1-15 Missing	n=15 Possibly from line item detail.
CPTCCSn	Clinical Classifications Software (CCS): CPT/HCPCS procedure classification	3	1 - 231 .	CCS Proc Class Missing or invalid	n=15
DDATE	Discharge date		YYMMDD . .A .C	Date of discharge Missing Invalid Inconsistent	Confidential data development file – **this data element must be removed from the data if quarterly data is released.**
DIED	Died during hospitalization	3	0 1 . .A .C	Did not die Died Missing Invalid Inconsistent	Recoded from DISPuniform.
DISPUniform	Disposition of patient, uniform	3	1 2 5 6 7 20 99 .	Routine Transfer to short-term hospital Transfer other: includes Skilled Nursing Facility (SNF), Intermediate Care Facility (ICF), and another type of facility Home Health Care (HHC) Against medical advice (AMA) Died in hospital Discharged alive, destination unknown Missing or invalid	
DRG	DRG in use on discharge date	3	nnn	DRG value	
DSHOSPID	Data source hospital number	$13	13(a)	Data source hospital number	

Variable Names	Description	Def	Value	Value label	Comments
DXCCSn	Clinical Classifications Software (CCS): diagnosis classification	4	1 - 259 2601-2621 .	CCS Diagnosis Class CCS E-code Class No diagnosis code	n=15
DXPOAn	Diagnosis n, present on admission indicator	$1	Y N U W E Blank	Yes No No information on record Clinically undetermined Exempt from POA reporting Missing or invalid	n=15 Includes source values of ‘1’ Includes all other codes not listed
DXn	Diagnosis	$5	annnn Blank	Diagnosis codes 1-15 Missing	n=15
E_CCSn	Clinical Classifications Software (CCS): E code classification	3	2601-2621 .	CCS E-code Class No diagnosis	n=15
ECODEn	External cause of injury code	$5	E code Blank	Ecodes 1-6 Missing	n=15
E_POAn	E Code n, present on admission indicator	$1	Y N U W E Blank	Yes No No information on record Clinically undetermined Exempt from POA reporting Missing or invalid	n=15 Includes source values of ‘1’ Includes all other codes not listed
FEMALE	Indicator of sex	3	0 1 .	Male Female Missing	
HISPANIC	Hispanic ethnicity	3	0 1 .	Non-Hispanic Hispanic Missing	

Variable Names	Description	Def	Value	Value label	Comments
HCUP_ED	HCUP emergency department indicator	3	0 1 2 3 4 .	Record does not meet any HCUP Emergency Department criteria Emergency Department revenue code on record Positive Emergency Department charge (when revenue center codes are not available) Emergency Department CPT procedure code on record Emergency Department admission source Missing	
HOSPST	Hospital state		aa	Hospital state postal code	Coded for all observations
LOS	Length of stay, cleaned	4	0 – 365 .	Days Missing	Same day stays coded as 0. .
MDC	MDC in use on discharge date	3	nn	MDC value	Coded for all observations.
MEDINCSTQ	Median household income state quartile for patient ZIP Code	3	1 2 3 4 .	First quartile Second quartile Third quartile Fourth quartile Missing	Quartiles are individually defined for each state. Median Income Quartile for ZIP Codes.
NCPT	Number of CPT/HCPCS procedures for this discharge	3	0-25	Number of CPT procedures	
NDX	Number of diagnoses for this discharge	3	0 - 15	Number of diagnoses	
NECODE	Number of E codes on this record	3	0-6	Number of E codes	

Variable Names	Description	Def	Value	Value label	Comments
NEOMAT	Neonatal/maternal discharge	3	0 1 2 3	No neonatal or maternal diagnosis or procedure on record Maternal diagnosis or procedure on record Neonatal diagnosis on record Neonatal & maternal diagnoses and procedures on the same record	Based on diagnosis and procedure screens.
NPR	Number of procedures for this discharge	3	0 - 15	Number of procedures	
PAY1	Expected primary payer, uniform	3	1 2 3 4 5 6 .	Medicare Medicaid Private insurance Self-pay No charge Other Missing	
PAY2	Expected secondary payer, uniform	3	1 2 3 4 5 6 .	Medicare Medicaid Private insurance Self-pay No charge Other Missing	
PL_NCHS2006	NCHS Urban-Rural Code, 2006	3	1 2 3 4 5 6 .	Large Central Metro Large Fringe Metro Medium Metro Small Metro Micropolitan Noncore Missing	
PNUM_R	Person number (re-identified)	5	9(n) .	Person number Missing	

Variable Names	Description	Def	Value	Value label	Comments
PointOfOriginUB04	Point of origin for admission or visit, UB-04 standard coding		1 2 4 5 6 7 8 B C D E F 5 6 blank	Non-health care facility point of origin Clinic Transfer from a hospital (different facility) Transfer from a Skilled Nursing Facility (SNF) or Intermediate Care Facility (ICF) Transfer from another health care facility Emergency Room Court/law enforcement Transfer from another Home Health Agency Readmission to Same Home Health Agency Transfer from one distinct unit of the hospital to another distinct unit of the same hospital resulting in a separate claim to the payer Transfer from ambulatory surgery center Transfer from hospice and is under a hospice plan of care or enrolled in a hospice program Born inside this hospital Born outside of this Missing, Invalid, or Information not available	Values when ATYPE not equal to 4
			5 6 blank	Born inside this hospital Born outside this hospital Missing, Invalid, or Information not available	Values when ATYPE equal to 4 (Newborn)
PRCCSn	Clinical Classifications Software (CCS): procedure classification	3	1 – 25 .	CCS Proc Class No Proc code	n=15
PRn	Procedure	$4	nnnn Blank	Procedure codes 1-15 Missing	n=15
PSTCO2	Patient state/county FIPS code	4	nnnnn .	State/county FIPS code Missing	From provided PSTCO if available; otherwise derived from ZIP Code

Variable Names	Description	Def	Value	Value label	Comments
QKEY	Record identifier	8	15(n)	Record identifier (coded as SSyyyyQTnnnnnnn, where SS = State FIPs code, yyyy = data year, Q = quarter, T = data type 1 (inpatient), 2 (ambulatory surgery), 3 (emergency department), and nnnnnnn = record counter)	Assigned during HCUP processing. KEY does not include a hospital identifier. The easiest way to manage quarterly keys would be to include quarter in their definition, but then they would not match the annual file. This probably would not be an issue though because it is unlikely that the same quarterly and annual keys would be assigned to the same records.
RACE	Race	3	1 2 3 4 5 6 .	White Black Hispanic Asian or Pacific Islander Native American Other Missing	Generally, Hispanic is coded as a separate variable so that it can be combined with various races. Perhaps we should follow this convention and replace the “Hispanic” category with a separate Hispanic variable.
TOTCHG	Total charges, cleaned	6	25 - 1 million .	Total Charge (rounded) Missing	
ZIP_S	Patient ZIP Code (synthetic)	$5	nnnnn C M F blank	ZIP Codes Canada Mexico Other or unspecified foreign Missing	ZIP Codes are primarily used to indicate geographical location so set foreign codes or identifications and homeless to missing.
ZIPINC_QRTL	Median household income national quartile for patient ZIP Code	3	1 2 3 4 .	First quartile Second quartile Third quartile Fourth quartile Missing	

Table 12. Suggested Data Elements, Severity File

Variable Names	Description	Def	Value	Value label	Comments
CM_HTN_C	AHRQ comorbidity measure: Hypertension (combine uncomplicated and complicated)	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_DEPRESS	AHRQ comorbidity measure: Depression	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_PSYCH	AHRQ comorbidity measure: Psychoses	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_DRUG	AHRQ comorbidity measure: Drug abuse	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_ALCOHOL	AHRQ comorbidity measure: Alcohol abuse	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_ANEMDEF	AHRQ comorbidity measure: Deficiency anemias	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_BLDLOSS	AHRQ comorbidity measure: Chronic blood loss anemia	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_LYTES	AHRQ comorbidity measure: Fluid and electrolyte disorders	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_ARTH	AHRQ comorbidity measure: Rheumatoid arthritis/collagen vascular diseases	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software

Variable Names	Description	Def	Value	Value label	Comments
CM_TUMOR	AHRQ comorbidity measure: Solid tumor without metastasis	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_METS	AHRQ comorbidity measure: Metastatic cancer	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_LYMPH	AHRQ comorbidity measure: Lymphoma	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_ULCER	AHRQ comorbidity measure: Peptic ulcer disease excluding bleeding	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_LIVER	AHRQ comorbidity measure: Liver disease	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_AIDS	AHRQ comorbidity measure: Acquired immune deficiency syndrome	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_ WGHTLOSS	AHRQ comorbidity measure: Weight loss	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_OBESE	AHRQ comorbidity measure: Obesity	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_COAG	AHRQ comorbidity measure: Coagulopathy	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_RENLFAIL	AHRQ comorbidity measure: Renal failure	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_HYPOTHY	AHRQ comorbidity measure: Hypothyroidism	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software

Variable Names	Description	Def	Value	Value label	Comments
CM_DMCX	AHRQ comorbidity measure: Diabetes with chronic complications	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_DM	AHRQ comorbidity measure: Diabetes, uncomplicated	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_ CHRNLUNG	AHRQ comorbidity measure: Chronic pulmonary disease	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_NEURO	AHRQ comorbidity measure: Other neurological disorders	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_PARA	AHRQ comorbidity measure: Paralysis	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_PERIVASC	AHRQ comorbidity measure: Peripheral vascular disorders	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_PULMCIRC	AHRQ comorbidity measure: Pulmonary circulation disorders	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_VALVE	AHRQ comorbidity measure: Valvular disease	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_ARYTHM	AHRQ comorbidity measure: Cardiac arrhythmias	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software
CM_CHF	AHRQ comorbidity measure: Congestive heart failure	3	0 1	Comorbidity is not present Comorbidity is present	Assigned using AHRQ Comorbidity Software

Variable Names	Description	Def	Value	Value label	Comments
QKEY	Record identifier	8	15(n)	Record identifier (coded as SSyyyyQTnnnnnnn, where SS = State FIPs code, yyyy = data year, Q = quarter, T = data type 1 (inpatient), 2 (ambulatory surgery), 3 (emergency department), and nnnnnnn = record counter)	Assigned during HCUP processing. KEY does not include a hospital identifier.

Table 13. Suggested Data Elements, Charges File

Variable Names	Description	Def	Value	Value label	Comments
NREVCD	Number of revenue codes for this discharge	3	0-25	Number of revenue codes	
QKEY	Record identifier	8	15(n)	Record identifier (coded as SSyyyyQTnnnnnnn, where SS = State FIPs code, yyyy = data year, Q = quarter, T = data type 1 (inpatient), 2 (ambulatory surgery), 3 (emergency department), and nnnnnnn = record counter)	Assigned during HCUP processing. KEY does not include a hospital identifier.
REVCHGn	Detailed charges for revenue code n (as received from source)	6	7(n).nn . .A	Charges 1-15 Missing Invalid	From line item detail.
REVCODEn	Line item revenue code	$4	4(a) Blank	Revenue Codes 1-15 Missing or Invalid	From line item detail.

Appendix B – Suggested Data Quality Tests

Data Element	Content Tests	Continuity Tests	Context Tests
(discharges/events) A count of observations	n/a	Percent change from one year prior *within specified range* Distribution by facility compared to one year prior *Chi-square distribution test*	Count as a percent of most recent **AHA** Survey *within specified range* Distribution by facility compared to most recent **AHA** Survey *Chi-square distribution test*
NDX	n/a	Distribution compared to one year prior *Chi-square distribution test*	Distribution compared to most recent **NIS/NEDS** *Chi-square distribution test*
DX1-DX15	n/a	n/a	n/a
DXCCS1	n/a	Distribution compared to one year prior *Chi-square distribution test*	Distribution compared to most recent **NIS/NEDS** *Chi-square distribution test*
NECODE	n/a	Distribution compared to one year prior *Chi-square distribution test*	Distribution compared to most recent **NIS/NEDS** *Chi-square distribution test*
ECODE1-ECODE15	n/a	n/a	n/a
E_CCS1	n/a	Distribution compared to one year prior *Chi-square distribution test*	Distribution compared to most recent **NIS/NEDS** *Chi-square distribution test*

Data Element	Content Tests	Continuity Tests	Context Tests
DRG	n/a	Distribution compared to one year prior *Chi-square distribution test*	Distribution compared to most recent **NIS/NEDS** *Chi-square distribution test*
NCPT (test only for OP data)	n/a	Distribution compared to one year prior *Chi-square distribution test*	Distribution compared to most recent **NEDS** *Chi-square distribution test*
CPT1-CPT15	n/a	n/a	n/a
CPTCCS1 (test only for OP data)	n/a	Distribution compared to one year prior *Chi-square distribution test*	Distribution compared to most recent **NEDS** *Chi-square distribution test*
NPR	n/a	Distribution compared to one year prior *Chi-square distribution test*	Distribution compared to most recent **NIS/NEDS** *Chi-square distribution test*
PR1-PR15	n/a	n/a	n/a
PRCCS1	n/a	Distribution compared to one year prior *Chi-square distribution test*	Distribution compared to most recent **NIS/NEDS** *Chi-square distribution test*
ADATE	Percent missing *within specified range*	Distribution by month compared to one year prior *Chi-square distribution test*	n/a
AGE	Percent missing *within specified range*	Distribution by age group compared to one year prior *Chi-square distribution test*	Distribution by age group compared to most recent **NIS** (age adjusted to the most recent **Census** estimate) *Chi-square distribution test*

Data Element	Content Tests	Continuity Tests	Context Tests
ATYPE	n/a	Distribution compared to one year prior *Chi-square distribution test*	Distribution compared to most recent **NIS/NEDS** *Chi-square distribution test*
PointOfOrigin	n/a	Distribution (non-newborns) compared to one year prior *Chi-square distribution test* Distribution (newborns) compared to one year prior *Chi-square distribution test*	Distribution (non-newborns) compared to most recent **NIS/NEDS** *Chi-square distribution test* Distribution (newborns) compared to most recent **NIS/NEDS** *Chi-square distribution test*
DDATE	Percent missing *within specified range*	Distribution by month compared to one year prior *Chi-square distribution test*	n/a
DISPUB04	n/a	Distribution compared to one year prior *Chi-square distribution test*	Distribution compared to most recent **NIS/NEDS** *Chi-square distribution test*
FEMALE	Percent missing *within specified range*	Distribution compared to one year prior *Chi-square distribution test*	Distribution compared to most recent **NIS/NEDS** *Chi-square distribution test*
LOS	Percent missing *within specified range*	Distribution compared to one year prior *Chi-square distribution test*	Distribution compared to most recent **NIS/NEDS** *Chi-square distribution test*
PAY1	Percent missing *within specified range*	Distribution compared to one year prior *Chi-square distribution test*	Distribution compared to most recent **NIS/NEDS** *Chi-square distribution test*

Data Element	Content Tests	Continuity Tests	Context Tests
RACE	Percent missing *within specified range*	Distribution compared to one year prior *Chi-square distribution test*	Distribution compared to most recent **NIS/NEDS** *Chi-square distribution test*
TOTCHG Note: "trended" mean is the prior year's mean multiplied by one plus the increase observed for that year. For example, if the means found in 2007 and 2008 were, respectively, 100 and 110 (for a trend rate of 10%), the trended mean for 2009 would be 121 (110 X 1.1).	Percent missing *within specified range*	Mean compared to trended mean from one year prior *t-test* Mean by facility compared to tended means from one year prior *Chi square test* Mean by DXCCS1 compared to tended means from one year prior *Chi square test* Mean by PRCCS1 compared to tended means from one year prior *Chi square test* Mean by DRG compared to tended means from one year prior *Chi square test* Mean by PAY1 compared to tended means from one year prior *Chi square test*	Mean compared to trended mean from most recent **NIS/NEDS** *t-test* Mean by facility compared to trended mean from most recent **NIS/NEDS** *Chi square test* Mean by DXCCS1 compared to trended mean from most recent **NIS/NEDS** *Chi square test* Mean by PRCCS1 compared to trended mean from most recent **NIS/NEDS** *Chi square test* Mean by DRG compared to trended mean from most recent **NIS/NEDS** *Chi square test* Mean by PAY1 compared to trended mean from most recent **NIS/NEDS** *Chi square test*