CHAPTER 15
Methods

Christopher S. Saigal, MD, MPH
Hangsheng Liu, PhD, MS, BMed
Jan M. Hanley, MS
Rodger A. Madison, MA
Alexandria C. Smith, MSPH
Julie C. Lai, MPH
Andrew W. Dick, PhD

Suggested Citation: Saigal CS, Liu H, Hanley JM, Madison RA, Smith AC, Lai JC, and Dick AW. Chapter 15: Methods. In: Litwin MS, Saigal CS, editors. Urologic Diseases in America. US Department of Health and Human Services, Public Health Service, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases. Washington, DC: US Government Printing Office, 2012; NIH Publication No. 12-7865 [pp. 497-522].

OVERVIEW

The purpose of the Urologic Diseases in America (UDA) project was to assess the burden of illness imposed upon the United States by the major urologic diseases. To accomplish this task, the UDA team at UCLA and RAND reviewed a large number of existing public and private datasets. Component elements of these databases were evaluated to compare their specific characteristics, uses, benefits, and limitations. Criteria for selecting the preliminary set of databases included (a) availability of information on key features of the data collection process, e.g., the unit of observation, reliability of the data, etc.; (b) issues related to the study design, e.g., the target population selected, whether incidence or prevalence data were available, etc.; (c) analytic information, e.g., whether adjustment for sample design characteristics, such as clustering was necessary, etc.; (d) the robustness of the dataset relative to others available to assess the same UDA condition; and (e) an estimate of the time required to procure and analyze the dataset. Ultimately, a complementary set of data sources was selected for this project, in coordination with approval from various experts in the field of urologic illnesses, as well as at the National Institute of Diabetes and Digestive and Kidney Diseases (see Appendix B). Together, these datasets allowed us to paint a broad picture of the burden of urologic diseases in America.

DATABASE SOURCES

Databases selected to study the UDA conditions included in this compendium fall into three categories. The first group describes the Medicare program's experience with the UDA conditions. The datasets were derived from Centers for Medicare and Medicaid Services (CMS) administrative records as either a complete, or a 5% sample (which was then appropriately weighted to represent the national Medicare population). These datasets include the Medicare inpatient (MedPAR) sample, the Medicare carrier file (previously referred to as the Physician/Part B file), and the hospital outpatient file. Finally, the Medicare denominator file, which includes all Medicare beneficiaries enrolled in a given year, was used to supply denominator data for analysis. Medicare data is also linked to information from the Surveillance Epidemiology and End Results (SEER) database.

The second group of datasets allows computation of national estimates of health care utilization, costs, and, for some conditions, prevalence. Data for inpatient utilization measures were obtained from the Healthcare Cost and Utilization Project – Nationwide Inpatient Sample (HCUP-NIS), conducted by the Agency for Healthcare Research and Quality. Data for pediatric inpatient stays were collected using the Kids' Inpatient Database (KID), conducted as part of the HCUP. Data for physician office and hospital outpatient utilization measures were obtained from two surveys conducted by the National Center for Health Statistics: the National Ambulatory Medical Care Survey (NAMCS) and the outpatient and emergency department components of the National Hospital Ambulatory Medical Care Survey (NHAMCS). Data on ambulatory surgery services were obtained from the National Survey of Ambulatory Surgery (NSAS). These databases contain data on national samples of visits to physician offices, outpatient hospital departments, and emergency departments, respectively, and yield a higher number of patients with diagnoses and procedures of interest than do population-based surveys. Finally, we examined the National Health and Nutrition Examination Survey (NHANES), a population-based survey, for items that could be used to create estimates of true nationally representative disease prevalence.

The third type of dataset allowed us to make utilization and cost estimates regarding the commercially insured population. The i3 dataset contains claims for inpatient stays, physician office, and hospital outpatient utilization. Data on patient medication use and lab values are also available in i3.

The combination of databases (Medicare, nationally representative datasets, and i3) allowed us to complete a comprehensive evaluation of the following primary service utilization categories: (1) inpatient stays, (2) hospital outpatient visits, (3) ambulatory surgery center visits, (4) physician office visits, and (5) emergency room visits for the UDA conditions in this compendium. The data also enabled us to derive estimates of disease prevalence for some conditions. The following is a detailed description of the databases analyzed in this compendium, and an in-depth discussion of the analytic approach we used for each data source.

MEDICARE DATA

Description

Medicare enrollment and claims data are available from the Centers for Medicare and Medicaid Services (CMS). Data from 2002 to 2007 claims were used for the tables in this compendium. The enrollment file contains information on all Medicare beneficiaries enrolled, or entitled in the year, and these data were used to generate counts for the denominator when calculating rates. The Medicare claims data consist of three separate files: MedPAR, which contains records for Medicare beneficiaries who used hospital inpatient services during the given year; the carrier file (previously referred to as the Physician (Part B) claims file); and the outpatient claims file (which contains hospital outpatient, laboratory, radiology, rehabilitation, and various other facility charges). For our analyses, we used 5% random samples drawn from these files. Previous work using CMS data has found that this sample size is adequate to detect significant racial and ethnic differences in use of cardiac procedures and tests (7). The carrier and outpatient files contained individual claims for provider services, and the MedPAR sample contained information on hospitalizations incurred by those same Medicare enrollees.

Analytic Approach

Data from the three Medicare files (MedPAR, carrier, and outpatient) were linked to determine inpatient, ambulatory surgery center, hospital outpatient, physician office and emergency room (ER) utilization, as well as to calculate average payments for the various UDA conditions by place of service. The procedure we used is described below.

First, personal identifiers and dates from facility records in the inpatient MedPAR and outpatient files were evaluated to ascertain the number of visits to inpatient hospitals, ERs, hospital outpatient departments, and ambulatory surgery centers. Ambulatory surgery centers were identified in both the outpatient file using revenue center codes (for hospital-based ambulatory surgery centers), and from the carrier file (for free-standing ambulatory surgery centers). Next, person identifiers and dates of service for these visits were linked to the matching line items listing payment for those services recorded in the carrier file. For records that did not have an exact match, an algorithm was developed to assign the remaining carrier file line items and outpatient file records to the appropriate place of service. Utilization of physician office visits was determined by examining line items in the carrier file for appropriate place-of-service and physician-evaluation-and-management billing codes.

Remaining unmatched line items and claims (primarily laboratory charges) from the outpatient file were totaled by disease entity and by place of service (physician office, hospital

outpatient, hospital inpatient, ambulatory surgery, or ER). Total dollars of expenditure associated with these unmatched items were then added to the total expenditure calculation for each place of service, stratified by disease. Average cost-per-service unit was calculated by dividing this total by the number of disease-related visits to the place of service.

At the completion of the matching process, descriptive tables were generated using appropriate International Classification of Diseases, 9th edition (ICD-9) diagnosis codes for the conditions of interest. Hospitalization or facility visit was used as the unit of analysis for the number of claims for each type of service. Denominators were derived using the CMS enrollment file. Because a 5% sample of Medicare records was utilized, national estimates of service use were obtained by multiplying counts by a constant weight of 20 to represent use in the entire Medicare-eligible population. The data were stratified by age, gender, and race variables. Confidence intervals were calculated using standard methods for proportions (1). In Medicare data analyses, 5% samples are considered adequate for meaningful comparisons among different minority, geographic, and age groups (2).

The analytic methodology is described in more detail in Appendix A, Technical Programming for Medicare Data.

NATIONALLY REPRESENTATIVE HEALTH CARE UTILIZATION AND COST DATA

Description

We used six datasets to derive nationally representative estimates of disease-specific service use, disease prevalence, and health care payments. These datasets include data for outpatient and ambulatory care utilization (NAMCS, NHAMCS and NSAS), data for inpatient stays or hospitalizations (HCUP-NIS and KIDs'), and data for outpatient and ambulatory care utilization (NSAS, NAMCS and NHAMCS). Finally, NHANES was used to determine the prevalence of urinary incontinence and kidney stones.

The databases assessed had different designs, depending on the goals of the surveys they represented. The NAMCS and NHAMCS databases used a nationally representative multi-stage probability sample. The sample design consisted of a number of stages that subcategorized the sample into units. First, counties or groups of counties were selected. Next, a probability sample of hospitals and their associated clinics or physicians (depending on the database) was selected within each county. Finally, a systematic sampling of patient visits to those physicians, or clinics was selected within a randomly assigned window of time during the year. The sample size for the years of data evaluated in these two databases ranged from approximately 25,000 to 40,000 patient visits per year, and the sample was used to describe utilization of physician office visit, hospital outpatient, and ER services in the United States. Similarly, the NSAS used a multistage probability sample, with the hospitals, or freestanding ambulatory surgery centers sampled at the first stage, or second stage, and specific surgical procedures sampled at the final stage. The 'hospital' universe includes non-Federal general, short-stay and children's hospitals located in the 50 states and the District of Columbia. The universe of "freestanding ambulatory surgery centers" comprises facilities which are state-licensed or Medicare-certified or which provide ambulatory surgery as the primary business activity, and operate independently as separate businesses.

The Healthcare Cost and Utilization Project, Nationwide Inpatient Sample (HCUP) database is also a nationally representative probability sample, but rather than using a multi-stage approach, the design is based on a sample stratified on five characteristics: geographic area (US

Census Region), location (Metropolitan Statistical Area (MSA), the teaching status of the hospital (teaching or non-teaching), the control of the hospital (public, voluntary, or proprietary), and size, by number of beds (small, medium, or large). This database is much larger than the NAMCS or NHAMCS; it contains from 7.5 million to 8 million discharge records from community hospitals for any given year of our analysis. HCUP data are thus adequate to describe utilization of hospital inpatient services in the United States.

The Kids' Inpatient Database (KID) was created as part of the Healthcare Cost and Utilization Project, sponsored by the Agency for Healthcare Research and Quality. KID is the only all-payer inpatient care database for children in the United States. KID contains a sample of pediatric discharges from approximately 3,000 community hospitals nationwide; it contains data from 2 to 3 million pediatric hospital discharges. For this compendium, data was available for 2003 and 2006. KID provides information on primary and secondary diagnoses, admission and discharge status, patient demographics (e.g., gender, age, race, median income for ZIP code), expected payment source, total charges (regardless of payer), length of stay, and hospital characteristics (e.g., ownership, size, teaching status) related to pediatric inpatient stays. KID samples all pediatric discharges from all hospitals in its sampling frame, stratified as "uncomplicated in-hospital birth", "complicated in-hospital birth," and "other pediatric discharges." After sorting discharges by state, hospital, diagnosis related group, and a random number within the diagnosis related group, systematic random sampling was applied. KIDs' sampling design allows nationally generalizable observations to be made about inpatient utilization for pediatric conditions.

The NHANES, conducted by the National Center for Health Statistics (NCHS), collects data by household interview, supplemented by medical examination and laboratory testing in a mobile center. The sample design is a stratified, multi-stage, probability sample of clusters of persons representing the civilian non-institutionalized population (African-Americans and Mexican-Americans are oversampled). Data include medical histories in which specific queries are made regarding urological symptoms and conditions. These items were selected for analysis. NCHS releases public use data sets from the continuous NHANES in two-year cycles. In our analyses, we present data on information available regarding urologic conditions from NHANES data from 2001-2008.

The benefits of using this combination of data sources are numerous. First, the databases are nationally representative samples that allow for the evaluation of genitourinary conditions even within special subpopulations (e.g., pediatric or ER patients). Demographic information is also available to complement the clinical data provided. However, the datasets have some limitations; for example, they use an inpatient stay or clinic visit, not an individual patient, as the unit of analysis, thus making it impossible to follow patients over time. Also, some of the databases sample a small fraction of total service use, so rare or more-chronic conditions may be missed.

The Surveillance, Epidemiology, and End Results (SEER) Program maintains several population-based registries in the United States and Puerto Rico and provides data on all residents diagnosed with cancer, and follow up information on all previously diagnosed patients. A continuing project of the National Cancer Institute, the SEER Program collects cancer data twice a year from designated population-based cancer registries in various areas of the country. Data are compiled twice a year. SEER is a product of the National Cancer Act of 1971, which mandated the collection, analysis, and dissemination of all data useful in prevention, diagnosis, and treatment of cancer. Trends in cancer incidence, mortality and patient survival in the United

States, as well as many other studies, are derived from this data bank. The geographic areas comprising the SEER Program's database represent an estimated 26% of the US population. SEER coverage includes 23% of African Americans, 40% of Hispanics, 42% of American Indians and Alaska Natives, 53% of Asians, and 70% of Hawaiian/Pacific Islanders. The database contains information on 6 million in situ and invasive cancers diagnosed between 1973 and 2008 (as of this printing); approximately 350,000 new cases are accessioned yearly in 18 geographical areas in the United States including Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, Utah, Los Angeles, San Jose-Monterey, Rural Georgia, the Alaska Native Tumor Registry, Arizona Indians, Greater California, Kentucky, Louisiana, and New Jersey. Cancer mortality data are obtained from vital statistics for the entire United States. SEER provides authoritative genitourinary cancer prevalence and incidence data which provide context for trends in expenditures and utilization documented in the other UDA datasets.

SEER-Medicare data

Through a collaborative effort between the National Cancer Institute and CMS, SEER data have been linked to Medicare claims in order to allow greater specificity when analyzing utilization of resources by older patients with cancers. SEER data, which are replete with clinical detail, are paired with related Medicare claims for covered health care services from the time of a person's Medicare eligibility until death. Linkage is accomplished by matching SEER identifiers with identifiers located in Medicare's master enrollment file. Linkages have been completed for years subsequent to 1991. Linkages are updated every 3 years. Data are currently available through 2008. SEER-Medicare data are requested as a series of files containing data on inpatient stays, outpatient claims, clinical cancer information, etc. Investigators may link individual patients across files using the unique SEER case ID number. Data are available for both subjects with cancer from SEER, and a random sample of Medicare beneficiaries without cancer (for comparison purposes).

Analytic Approach (NAMCS, NHAMCS, NSAS, HCUP, KID, NHANES)

The years of NAMCS, NHAMCS and HCUP data analyzed are 2002 to 2007. NSAS data were available for 2006 only. KID data were available for 2003 and 2006.

First, we identified individuals with visits for specific urologic conditions based on the ICD-9 diagnosis, or procedure codes that defined each of the conditions and any age and gender specifications necessary to create subpopulations for the analyses. Analytical files for outpatient visits included records of visits with a relevant diagnosis code listed as one of any reasons for the visit. Tables were produced reflecting service use, both when the diagnosis codes in question were listed as any of the reasons for the visit, and when they were listed as the primary reason for the visit. Analytical files for inpatient stays included only those records of inpatient hospitalizations for which a relevant diagnosis code was listed as the *primary diagnosis* during the hospitalization. The raw number of visits in each subset varied by condition and by year. Analyses were conducted at the visit level, or the stay level, depending on which database was being analyzed.

For the NHANES, cases were identified on the basis of answers to specific questions asked in the survey. The frequency of individual "yes" answers and answers regarding the intensity of symptoms were tabulated by gender, age, and other demographic variables. Using

the weights provided by the NCHS, raw counts were weighted to give nationally-representative estimates of disease prevalence.

National estimates of the annual frequency of visits for the demographic groups studied for each of the UDA conditions were calculated when the raw counts were deemed large enough to produce reliable estimates. Under NCHS guidelines there must be at least 30 unweighted counts for creation of reliable national estimates[1]. When insufficient data were available, subgroups (e.g., age categories) were combined to create adequate unweighted counts. In some instances, unweighted corresponding counts for conditions in NHAMCS Outpatient (NHAMCS–OP) and NAMCS were combined to provide reliable estimates of overall outpatient service use. HCUP cell sizes were always large enough to produce reliable estimates (N≥30), and therefore no combining, or regrouping of stratification variables was necessary. For some analyses, a merging of population weights was applied to unweighted counts, according to the methodology provided by each organization sponsoring a survey, to obtain national estimates of the frequency of visits in the entire population and in sub-populations of interest. SAS (3) was used to derive the standard errors and compute the 95% confidence intervals (CIs) for these estimates using the method described below under "Computing Confidence Intervals for Proportions.

To create an estimate of the burden of outpatient visits for urologic conditions in relation to the total burden of illness represented by outpatient visits, national estimates of visits for urologic conditions within various subpopulations were divided by national estimates of the total number of outpatient visits for the demographic groups of interest. This number was multiplied by 100 to generate a percentage. National annual outpatient visit rates were calculated using the US Census non-institutionalized civilian population estimates corresponding to demographic and visit-characteristic groupings for each survey year used. Population estimates were obtained from the Current Population Survey (CPS) for select demographic categories of the US civilian non-institutionalized population.

Stratification variables evaluated for all databases include age, race/ethnicity, gender, region and/or MSA, and other variables selected as appropriate for the database of interest.

COMMERCIALLY INSURED POPULATION

i3 Description

The i3 dataset contains claims from 2002-2007 and represents approximately 30 million insured individuals. The i3 dataset contains three main files: medical claims, prescription claims, and an enrollment file. The i3 medical claims consist of inpatient stays, outpatient and physician utilization. Procedure and diagnosis codes, financial information, dates of service, information regarding the types of facilities and provider are included in the i3. In addition to utilization, the i3 provides drug claims, which consist of prescription fill date, refills, brand name, therapeutic class and cost. For the majority of analysis the national drug codes (NDC), were used to examine utilization of specific drugs and therapeutic classes. The enrollment file contains demographic information, such as the person’s age, sex, plan type (FFS, PPO, POS, HMO), zip code of residence, and relationship to employee.

Analytic approach

Using the place of service variable in the i3 dataset, line item claims were designated to inpatient, hospital outpatient, ambulatory surgery center, physician office, and emergency room utilizations. Inpatient line items were “rolled up” to create a single inpatient stay and any line items that fell into the dates of the stay, regardless of place of service code, were included in the

inpatient costs and utilization. To determine an inpatient stay, line items that had a two day or less gap were matched into one stay episode. Laboratory claims were matched to hospital outpatient, ambulatory surgery center, physician office, and emergency room claims using person identifiers and exact dates of record. Remaining laboratory claims that did not have an exact date match were then matched using a seven day window to the nearest visit claim. If laboratory claims still did not match, these claims were added to the total cost of the disease, but not split out to the different place of service. Utilization of the place of service by disease were determined by aggregating the claims to a person date level. Charges assigned to the place of service by disease were determined by the summation of charges from claims to a person date level.

ESTIMATING COSTS ASSOCIATED WITH UROLOGIC DISEASES

Methods on Estimating Costs Associated With Urologic Diseases

Overview

As one of the goals of the Urologic Diseases in America (UDA) project, we estimated the economic burden of the major urologic diseases in the United States. Two types of analyses were conducted to describe disease burden. First, the total costs incurred were summarized for each urologic disease of interest, by place of service, patient population, and calendar year. Second, incremental costs were estimated through modeling the cost difference between patients with a urologic disease and those without, after controlling for patient socio-demographics, comorbidities, health insurance status, and geographic location. These two sets of analyses provide the UDA compendium users costs estimates that may be used for various purposes.

Data Sources

The 2002-2007 Medicare claims data were used to describe urologic conditions among the population over 65 years old. The datasets contain medical claims and administrative records of a 5% random sample of Medicare beneficiaries, including the Medicare carrier file, inpatient file, hospital outpatient file, and the denominator file.

In order to describe the non-Medicare population, several nationally-representative datasets were used to estimate economic costs of urologic diseases. These include the Healthcare Cost and Utilization Project – Nationwide Inpatient Sample (HCUP-NIS), the National Ambulatory Medical Care Survey (NAMCS), the National Hospital Ambulatory Medical Care Survey (NHAMCS). These databases provide information on a national level about physician office visits, hospital outpatient visits, emergency department visits, and hospital inpatient stays.

Not all of these datasets provided charge data to accompany utilization estimates. To estimate the economic burden of these nationally-representative utilization estimates, we used charge data from the i3 claims database of privately insured individuals.

Analytical Approach

Most cost-of-illness studies distinguish between the direct costs of treating a medical condition and the indirect costs associated with lost work days, reduced quality of life, and premature mortality. Direct costs typically include expenditures for medical treatments, such as hospitalizations, emergency care, ambulatory visits, nursing home and home health care, medical supplies, prescription drugs, and other services provided by medical professionals. Indirect costs usually refer to disability days, work loss, and other labor-market consequences associated with medical illness. For this compendium, the analysis focused on direct costs only as data on lost work days were not available.

Direct costs were quantified as a dollar-denominated measure of resource utilization, by assigning prices to a comprehensive list of utilization and services. For the Medicare population, prices of medical services were estimated based on average payments made by the enrollee (co-payments, deductibles, excluded expenses) and by Medicare as well as other third-party payers. Medicare claims were used to estimate utilization and average reimbursements for the Medicare population. For the non-Medicare population, average prices were approximated using charges derived from the i3 data because the actual payments were not available in the data. National

surveys and inpatient discharge databases were relied upon for deriving estimates of medical service utilization for the non-Medicare population, where the data source depends on the type of service provided. All expenditures for medical and pharmacy services were reported in 2009 dollars, after adjusting inflation using the Consumer Price Index compiled by Bureau of Labor Statistics, U.S. Department of Labor. Finally, all descriptive analyses used appropriate sampling weights to obtain national estimates.

Similar to the analysis of health care utilization, the cost analysis was also based on episodes of care with a primary diagnosis of a urologic condition. That is, urology-related costs that are secondary to the primary diagnosis were excluded, while costs related to non-urologic conditions incurred during a visit, or hospitalization for a urologic illness listed as a primary diagnosis were included. This approach might over-estimate average expenditures by including treatment costs of non-urologic conditions. However, urology-related costs that occur during visits for which a urologic diagnosis is not the primary diagnosis are not included in our estimates.

The incremental medical costs incurred by persons with urologic conditions were estimated using i3 data. Individuals with an inpatient, or outpatient claim for specific urologic conditions were identified. Multivariate regression models were used to predict medical and pharmacy spending during 2003-2006 for persons with and without a particular condition, controlling for differences in patient demographics, health status, and insurance coverage. The primary outcomes of interest included annual medical and pharmacy expenditures for each person. Expenditures consisted of total annual provider charges for medical services, and outpatient prescription drug claims. Covariates included age, sex, region of residence, insurance type (HMO or not), and Charlson co-morbid conditions. Medical claims were used to identify individuals treated for any of 18 chronic conditions based on ICD9 diagnostic codes, including hypertension, diabetes, congestive heart failure, and depression, and included a binary indicator for each condition.

Statistical analyses used generalized linear models that incorporated within-patient correlations based on generalized estimating equations. A log link and Gaussian family were used, and the correlation structure was unstructured. The models did not converge when using a Gamma family or a Poisson family. All patients with at least one urologic condition of interest were included in the disease group, while others without any urologic condition formed the control group. Due to the sheer size of the control group data, a 3% random sample of the control patients was selected for modeling purposes and appropriate weights were used to reflect the sampling method.

Limitations

Our national economic burden estimates reflect charge data, and as such may be overestimating the burden since charges are generally higher than reimbursements. We were unable to estimate medication costs, which are a substantial component of the economic burden for several diseases (e.g. erectile dysfunction).

Data Quality

A systematic approach was developed to evaluate the quality of the data generated for this project. A multi-tiered effort was made to ensure that the data met a high level of accuracy and consistency throughout. Data generated from each database were subjected to multiple levels of examination.

The first level of review required confirmation that the base populations used for each database were correct for each condition being evaluated (e.g., the population at risk for BPH included only males aged 40 years and older, whereas both sexes are at risk for kidney stones). Also, the total frequencies were checked to ensure that they were correctly reported (e.g., that there was no double counting of cases).

Next, individual frequencies were evaluated within patient subgroups to ensure that the counts were appropriate. Any numbers that appeared inconsistent were flagged for a programmer to recheck and review. For example, one would not expect to find a dramatically greater incidence of a particular condition among patients from two different regions of the country, and this inconsistency might be identified for further review.

Third, the rates were compared over all years for which data were available. This allowed for an evaluation of whether any unusual rates were reported for a particular year, or service. We developed a graphical tool to compare the rates across data sets including the results from the previous compendium for each disease. We examined the rate both for overall and individual age groups. This approach was particular valuable in identifying potential data errors or trend changes. Any rates that appeared out of range were flagged for further review. To this end, a comprehensive literature review was performed using the relevant disease search terms. Rates generated from the datasets were compared with published estimates, and clinical experts adjudicated whether discrepancies signaled analysis errors. Also, confidence interval calculations were reviewed to ensure that they were within the appropriate range for all rates reported.

For the next level of verification, a mean-annual-payment summary table was produced to compare payments across years and services. Again, any payments that appeared out of range were flagged for further evaluation. In many cases, a small sample size explained a wide variation in reported payments.

Finally, summary base population tables were generated for all conditions and years. These tables revealed cases where the sum of subpopulations did not total the base population for any given year, or where a base population was mistakenly used for the wrong year.

This systematic approach to reviewing data quality successfully uncovered issues that were later remedied at all levels of evaluation.

APPENDIX A: TECHNICAL PROGRAMMING FOR MEDICARE DATA

This appendix describes the process by which data from the Medicare MedPAR, carrier, and outpatient files were combined to assign number of visits and costs to five separate types of service: inpatient stays, physician office visits, hospital outpatient visits, ambulatory surgery visits, and emergency room (ER) visits.

Claims records from the MedPAR, carrier and outpatient files for a 5% sample of Medicare beneficiaries were used in building the files for this research effort.[2]. The MedPAR files contain summary records for all inpatient stays. The carrier file contains detailed information at the line-item level, which provided information on payment and place of service by line item[3]. Therefore, the carrier records were processed by line item rather than claim for this project. The outpatient file also contains detailed information, but not about payments, or place of service[4].

An iterative process was used to build the analysis files. First, inpatient stays were identified, using MedPAR records. All costs from claims in the outpatient and carrier records with a date of service that occurred during an inpatient stay, as determined by the admission and discharge dates, were added to the inpatient silo. Next, ER, outpatient surgery, and ambulatory surgery visits in the outpatient file were defined, using appropriate revenue center codes. Stand-alone ambulatory center were defined using the place of service code from the carrier files and remaining line items with place of service as office and procedure codes with a range of 99024-99058 or 99199-99999 became the physician office visit core records. Payments from other line items with the same patient identifier, provider, and date of service were added to these physician office visit records; Finally, the line items and outpatient records that were not facility charges were matched to these visits, using the following procedure: (a) person and exact dates of service were matched; (b) unassigned line items and outpatient records were assigned, using place of service and date ranges; (c) payments from any line item or facility records that had not yet been assigned were aggregated by place of service. These “orphan” payments were included only in the calculation of cost per visit.

CREATING THE FILES

The Inpatient Analysis File

Inpatient stays were identified from MedPAR as those stays in which the admitting diagnosis matched one of the diagnoses used to define a UDA condition. This provided the count of inpatient stays for the UDA utilization tables. All other data added to the stay were used to track payments that were occasioned by the stay.

Assigning Payments from Carrier Line Items to Inpatient Stays

Line items were matched to stays, using person identifier and dates of service. Each stay had an admission date and a discharge date. Each line item also had a begin date and an end date (although for most line items they were equivalent). The rules for assigning line-item payments to stays varied by whether the line item matched the admission date, the discharge date, or a date in between (or an interim stay date).

Payments from any line item that matched a person and an admission, or interim stay date were assigned to the stay. Payments from line items that matched a person and discharge date, and had place of service equivalent to *inpatient* or *ambulance* were assigned to the stay. Payments from any line item with a place of service equivalent to *emergency room* that matched

a stay on admission date, or any interim dates were included with the stay. If the line item also matched an emergency room facility, the payments were included with the emergency room visit.

Matching Outpatient Files with Inpatient Stays

Outpatient claims were matched to inpatient stays using HICs,[5] inpatient admission and discharge dates, and outpatient begin and end dates. Outpatient dollars were added to the inpatient stay if at least one of the following rules was met:

- The outpatient claim began and ended between (or including) the inpatient admission and discharge dates.
- The outpatient claim began during an inpatient stay and ended after the stay.
- The outpatient claim began and ended on the inpatient admission date.
- The outpatient claim began and ended on the inpatient discharge date.

An outpatient claim with an ER revenue center "flag" that occurred on the same day as an admission date counted as an ER visit in the ER facility of service.

Facility claims matching the discharge date of one stay and the admission date of a second stay were assigned to the second stay. These were generally ambulance services related to hospital transfers.

The Hospital Outpatient, Ambulatory Surgery, and ER Analysis Files

Each of these files was created using the revenue center codes found on the claims. The reason for the visit to one of these places of service was determined by the UDA condition found at the revenue center, not on the condition shown in data imported from the carrier file.
The revenue centers used to define *hospital outpatient* were:

- Clinic-general classification
- Clinic-chronic pain center
- Clinic-psychiatric
- Clinic-OB-GYN
- Clinic-pediatric
- Clinic-urgent care
- Clinic-family practice
- Clinic-other
- Free standing clinic-general classification
- Free standing clinic-rural health, clinic
- Free standing clinic-rural health, home
- Free standing clinic-family practice
- Free standing clinic-urgent care

The revenue centers used to define an *ambulatory surgery visit* were:

- Ambulatory surgical care-general
- Ambulatory surgical care-other
- Operating room services-general classification
- Operating room services – minor surgery

The revenue centers used to define an *emergency room visit* were:

- Emergency room-general classification

- Emergency room-EMTALA[6] emergency medical screening services
- Emergency room-emergency room beyond EMTALA screening
- Emergency room-urgent care (effective 10/96)
- Emergency room-other

If an individual had two ER visits on the same day, they were counted as separate encounters. Claims were also assigned to the ambulatory surgery silo if the Facility type was "Special Facility or ASC Surgery" and the claim type was Ambulatory surgical center in hospital outpatient department. There could be up to 45 revenue centers on a single outpatient claim record. For some claims, the revenue center fell into more than one facility of service. They were then assigned to the appropriate facility of service based on their HCPCS[7] codes

Physician services were drawn from the line-item file (carrier), and the payments associated with these services were assigned to an emergency room visit, hospital outpatient visit, or ambulatory surgery visit, using place of service, HIC, and exact date matches, as follows.

Payments from line items that matched an ER visit by person and exact date, and had a place of service that included ER, ambulance, or independent laboratory, or had a CPT code ranging from 99281 to 99285, were assigned to the emergency room facility of service. Payments from line items that matched a hospital outpatient visit by person and exact date, and had a place of service that included outpatient hospital, ambulatory surgery center, ambulance, or independent laboratory, were assigned to the hospital outpatient facility of service. Similarly, payments from line items that matched an ambulatory surgery visit by person and exact date, and had a place of service equivalent to outpatient hospital, ambulatory surgery center, ambulance, or independent laboratory, were assigned to the ambulatory surgery facility of service. Claims for free standing ambulatory surgery centers are only in the carrier file and have place of service coded as ambulatory surgery center. These claims were included in the ambulatory surgery center silo and counted as ASC visits.

The remaining line items on the carrier file that had a place of service that included inpatient, ER, outpatient, or ambulatory surgery were examined. The number of days between each line item and each visit for a person were reviewed, and payments for remaining line items (most of which were laboratory services) were matched to the payment total for the type of service encounter that occurred closest in time to the date of the line item[8]. For example, the payment for a line item with a place of service listed as *hospital outpatient* that occurred within seven days of a hospital outpatient visit was added to the grand total of all hospital outpatient payments, but was not assigned to the cost of that particular visit. The mean payment for a hospital outpatient visit would be calculated by dividing the grand total for all hospital outpatient payments by the total number of hospital outpatient visits. If the nearest date for a service encounter was more than seven days from the date of the line item, the cost for the line item was not added to any silo but the cost was added to the total cost for the disease.

The Physician Office Analysis File

After the above steps were performed, the remaining line items, having procedure codes equivalent to 99024–99058 or 99199–99999, formed the core physician office visit file. Payments from any line items from the carrier file or remaining facility records from the outpatient file that matched by patient, provider, and exact date of service were added to this visit file.

Remaining Carrier and Outpatient Payment Items

Remaining facility records that were not matched in the steps outlined above were matched to ER visits, hospital outpatient visits, or ambulatory surgery visits based on exact date of service. Payments from these facility records were added to the payment total for the relevant visit. If a record matched more than one such place of service, its payment amount was split between them. All remaining ambulance service revenue center payments were added to the total payments for ER visits. All radiation therapy revenue center payments were added to the total for hospital outpatient visits.

The remaining facility records were those that did not match a place of service by exact date, and hence were coined "orphan" records. These records' payments were added to the established total payments for physician office visits, ambulatory surgery visits, hospital outpatient visits, and ER visits by HIC to the nearest date of service, using the following rules:

- Facility records were matched to the nearest visit by date of service within seven days.
- Matches were allowed to the ER only by plus, or minus one day.
- Records that matched more than one place of service by the same number of days were assigned in the following order: physician office, hospital outpatient, ER, ambulatory surgery.

Counts—Units of Analysis

Counts presented in the tables of this compendium are claims for each type of service. An individual could be counted more than once in each table if he or she had multiple events during the year. Within each facility of service, group counts, as well as payments, were tabulated for all persons and were stratified by age group, gender, race, and region. Gender and race codes used were those found on the claims record. The age category was derived from the age recorded on the claim record. The region code used was the census region, with claims recoded to region, using the state of residence.

Calculation of Denominators

Denominators for tables were derived from the CMS denominator file. This file includes the entire Medicare-eligible population and contains one record for each individual. Data from the denominator file can be linked to all other CMS files, using a unique identifier (ID) common to all files. In addition to eligibility status, the denominator file contains information about HMO membership. Individuals who were members of an HMO at any time during a year were dropped from the analysis because HMO claim records contain no payment information.

Weighting

The Medicare claims files, MedPAR, carrier file and the outpatient file are simple 5% random samples of the Medicare-eligible population. The sample was drawn using the last two digits of enrollees' SSNs. National estimates presented in the tables were obtained by multiplying counts by a constant weight of 20 to represent the entire Medicare-eligible population.

Computing Confidence Intervals for Proportions

Ninety-five percent confidence intervals were calculated using the normal approximation to the binomial distribution (1). The confidence interval is:

$$(p - 1.96\ sqrt(pq/n),\ p + 1.96\ sqrt(pq/n))$$

where ***p*** is the estimated proportion of interest, ***q***= 1–p, ***n*** is the number of observations, and ***sqrt*** refers to the square-root function.

APPENDIX B: SUMMARY OF DATASETS

Centers for Medicare and Medicaid Services (CMS)

Sponsor:
Robyn Thomas, Director
Division of Quality Coordination and Data Distribution (DQCDD)
OIS/EDG/DQCDD N1-15-03
Centers for Medicare and Medicaid Services (CMS)
7500 Security Blvd.
Baltimore, MD 21244-1850

Design: The Medicare dataset contains a number of files, including the Medicare provider analysis and review (MedPAR) file, the carrier file, the outpatient file, and the denominator file. The *MedPAR* file contains records for Medicare beneficiaries who used hospital inpatient services during the given year. Each record summarizes a stay. The *carrier file* contains final action claims data submitted by non-institutional providers, such as physicians, physician assistants, nurse practitioners, and standalone ambulatory surgical centers. Each observation in this file is at the claim level. The *outpatient file* contains final action claims data submitted by institutional outpatient providers, such as hospital outpatient departments, rural health clinics, and outpatient rehabilitation facilities. The unit of observation is also at the claim level. Finally, the *denominator file* contains demographic and enrollment information about each beneficiary enrolled in Medicare during the calendar year.

Time Frame: Data are available for 2002 to 2007.

Sample Size: The 100% MedPAR dataset contains approximately 11 million records annually. For our analyses, a 5% MedPAR sample was used. The carrier and outpatient dataset samples we used were based on a 5% simple random sample of the HIC numbers from each database. The carrier file contains 30 million records, and the outpatient file contains 5 million records

Use: MedPAR provides in-depth information on all Medicare beneficiaries, including information on their diagnoses and procedures, along with a breakdown of charges for the year.

Benefits: Longitudinal tracking is possible, given the continuous data collection and large sample size. The detailed breakdown of charges allows for calculation of expenditures over a given year. The database also includes multiple diagnosis/procedure codes, thereby allowing for a more detailed level of analysis of charges associated with the urologic conditions under review.

Limitations: These data contain limited demographic information. Most beneficiaries are over 65 years of age.

Healthcare Cost and Utilization Project (HCUP)—Nationwide Inpatient Sample (NIS)

Sponsor:
Healthcare Cost and Utilization Project (HCUP)—Nationwide Inpatient Sample (NIS)
Agency for Healthcare Research and Quality
HCUP Central Distributor
Social and Scientific Systems
8757 Georgia Ave., 12th Floor
Silver Spring, MD 20910
(866) 556-4287

Design: The Nationwide Inpatient Sample (NIS) is a subsample of the State Inpatient Databases (SID). NIS represents a 20% sample of hospital discharges from SID that includes all ages. The database utilizes a nationally representative stratified sample of approximately 6 million to 7.5 million records for the time period analyzed in this study.

Time Frame: The database contains data for 2002 to 2007.

Sample Size: Initially, the database covered only eight states; it has since grown to 28 states. It contains discharge data on approximately 7 million discharges, approximating a 20% stratified sample of US community hospitals. The sample of hospitals comprises about 80% of all hospital discharges in the United States.

Use: Data on hospital inpatient stays can be used to identify, track, and analyze national trends in access, charges, quality, and outcomes. It is the only national hospital database with charge information on all patient stays, regardless of payer.

Benefits: This large, nationally representative sample allows for the evaluation of trends over time. It can also be used to evaluate rare conditions and special populations (e.g., pediatric), and it includes charge information on all patient stays.

Limitations: Only hospitalizations are included, thereby limiting the types of service that can be analyzed. However, it may be possible to document change from inpatient to outpatient care over the years if HCUP is combined appropriately with other databases.

I3

Sponsor:
I3 Innovus
10 Cabot Rd, Ste 304
Medford, MA 02155
www.**i3**global.com

Design: Observational claims database of individuals with private insurance.

Time Frame: 2002-2007.

Sample Size: Contains information on approximately 30 million individuals with private insurance.

Use: Allows characterization of utilization and charges in privately insured individuals with urologic disease.

Benefits: This is a large dataset which allows for adequate numbers to describe less common urologic conditions, for example amongst the pediatric population. Data on medication use and some laboratory values are available.

Limitations: Privately insured individuals differ in socioeconomic status from non-insured or underinsured individuals. The data are not nationally representative. Only charge data are available, as opposed to reimbursement data.

National Survey of Ambulatory Surgery (NSAS)

Sponsor:
National Center for Health Statistics
Centers for Disease Control and Prevention
Division of Data Services
6525 Belcrest Road
Hyattsville, MD, 20782-2003
(301) 458-4636

Design: The NSAS is a multi-stage probability sample, with the hospitals, or freestanding ambulatory surgery centers sampled at the first stage or second stage and specific surgical procedures sampled at the final stage. The “hospital” universe includes non-Federal, general, short-stay, and children’s hospitals located in the 50 states and the District of Columbia. The universe of “freestanding ambulatory surgery centers” is comprised of facilities which are state licensed/ Medicare certified, or which provide ambulatory surgery as the primary business activity and operate independently as a separate business. Facilities specializing in dentistry, podiatry, abortion, family planning, or birthing were also excluded from NSAS.

Time Frame: 2006

Sample Size: The 2006 NSAS abstracted data from 52,000 ambulatory surgery visits to 437 facilities.

Use: As surgical services migrate to outpatient settings, describing ambulatory surgery use is required to present a comprehensive portrait of disease burden. Specific trends within surgical diseases can be examined in relation to subject demographics.

Benefits: This unique dataset allows for national estimation of ambulatory surgery services and trends.

Limitations: ICD-9 procedure codes are used, which are of limited specificity when compared with Common Procedural Terminology procedure codes.

Kids' Inpatient Database (KID)

Sponsor:
Healthcare Cost and Utilization Project (HCUP)—Nationwide Inpatient Sample (NIS)
Agency for Healthcare Research and Quality
HCUP Central Distributor
Social and Scientific Systems
8757 Georgia Ave., 12th Floor
Silver Spring, MD 20910
(866) 556-4287

Design: US community hospitals (defined as short-term, non-Federal, general and specialty hospitals, excluding hospital units of other institutions).

Time Frame: 2003, 2006

Sample Size: KID contains a sample of pediatric discharges from approximately 3,000 community hospitals nationwide; it contains data from 2 to 3 million pediatric hospital discharges.

Use: KID allows national estimates of pediatric inpatient utilization, including procedure use.

Benefits: This dataset allows national estimates of time trends, quality of care, and charges related to pediatric urologic conditions.

Limitations: Procedures are listed using ICD-9 coding, which limits specificity when compared with CPT coding. Despite the size of the dataset, some urologic conditions are incompletely described by the data. As the trend toward outpatient care in urology continues, this limitation may grow.

National Ambulatory Medical Care Survey (NAMCS)

Sponsor:
National Center for Health Statistics
Centers for Disease Control and Prevention
Division of Data Services
6525 Belcrest Road
Hyattsville, MD, 20782-2003
(301) 458-4636

Design: Data are collected from non-Federally employed physicians engaged in direct patient care (this excludes anesthesiology, radiology, and pathology) during a randomly assigned one-week reporting period. The physicians are selected on the basis of a national probability sample of office-based physicians. During the reporting period, data are gathered on an encounter form that records a systematic random sample of visits per physician. Data collected include patients' symptoms, physicians' diagnoses, and medications either ordered, or provided to the patient.

Time Frame: 2002 to 2007.

Sample Size: The sample size for the years of data evaluated in this compendium ranges from 1,200 to 1,700 physicians and 25,000 to 40,000 patient visits annually.

Use: The data provide information about the provision and use of ambulatory medical care in the United States.

Benefits: This database may be considered nationally representative, since it has a multistage probability design and captures the physician subspecialties that may encounter urologic conditions. Also, this database may identify a number of urologic conditions (e.g. UTI, BPH) that might otherwise go unreported because many of them are identified on the basis of office visits alone.

Limitations: There are no identifiers to track patients longitudinally. Also, some rare pediatric conditions may be missed because of the limited number of visits reported. The number of urologists sampled may be small for specific analyses. There are no cost data, and there may be more than one record per person because the data report number of *patient visits*, not *patients*.

National Hospital Ambulatory Medical Care Survey (NHAMCS)

Sponsor:
National Center for Health Statistics
Centers for Disease Control and Prevention
Division of Data Services
6525 Belcrest Road
Hyattsville, MD, 20782-2003
(301) 458-4636

Design: These data are collected in order to provide a better understanding of the utilization and extent of ambulatory care services in hospital emergency and outpatient departments. Data are collected on a national sample of emergency department and outpatient visits, excluding federal, military, and VA hospitals. The database uses a four-stage probability design. First, a sample of geographic areas is defined. Next, a sample of hospitals is identified within these areas. Third, clinics are selected within these hospitals. Finally, patients are selected on the basis of their visits to these clinics.
A patient record form is completed by hospital staff during a randomly assigned four-week period.

Time Frame: 2002-2007.

Sample Size: The sample size for the years of data evaluated in this compendium is in the range of 25,000 to 40,000 patient visits annually.

Use: The data describe utilization and provision of ambulatory care services in hospital emergency and outpatient departments (excluding federal, military, and VA hospitals).

Benefits: The survey covers a nationally representative multistage probability sample, which includes a pediatric population and contains data on genitourinary care in ERs. Other data reported include demographic characteristics of patients, expected source(s) of payment, diagnoses, medication, and certain characteristics of the hospital, such as type of ownership.

Limitations: There are no cost data and no identifiers to track patients longitudinally. An individual may have more than one record, since the data are based on number of *patient visits*, not *patients*. Because the number of visits is small, rare conditions and those that are chronic in nature may be missed.

National Health and Nutrition Examination Survey (NHANES)

Sponsor:
National Center for Health Statistics
Centers for Disease Control and Prevention
Division of Data Services
3311 Toledo Road
Hyattsville, MD 20782
(301) 458-4636

Design: NHANES is a continuing series of national sample surveys of households and household members in 50 states.

Time Frame: 2001-2008.

Sample Size: The sample for the 2008 NHANES includes approximately 9,762 respondents, age 2 months and older.

Use: The survey allows collection of data regarding urologic diseases and symptoms that can be used to generate true national prevalence for these diseases and symptoms during the time period covered in the survey.

Benefits: The data are unique in that they allow for nationally-representative estimates of the prevalence of certain urologic conditions.

Limitations: Relatively few urologic conditions are asked about in this survey. Subject self-report regarding medical history is subject to error.

Databases Selected for Analysis

DATABASE	ACRONYM	CATEGORY	PURPOSE
Centers for Medicare and Medicaid Services-Medicare Provider Analysis and Review	CMS-MedPAR	Medicare	Records of hospital inpatient services for Medicare beneficiaries
Centers for Medicare and Medicaid Services-Carrier File	CMS-Carrier	Medicare	Claims submitted by non-institutional providers for Medicare beneficiaries
Centers for Medicare and Medicaid Services-Outpatient file	CMS-Outpatient	Medicare	Claims submitted by institutional outpatient providers for Medicare beneficiaries
Centers for Medicare and Medicaid Services-Denominator file	CMS-Denominator	Medicare	Demographic and enrollment information on Medicare beneficiaries
Healthcare Cost and Utilization Project – Nationwide Inpatient Sample	HCUP-NIS	Health care utilization and cost	National sample of inpatient stays and hospitalizations
National Ambulatory Medical Care Survey	NAMCS	Health care utilization and cost	National sample of ambulatory care utilization
National Hospital Ambulatory Medical Care Survey-Outpatient and Emergency Room Components	NHAMCS-OP NHAMCS-ER	Health care utilization and cost	National sample of ambulatory care services in hospital emergency and outpatient departments
I3 database		Cost of disease	Medical claims database providing utilization and cost data for private sector
National Survey of Ambulatory Surgery	NSAS	Health care utilization and cost	National sample of ambulatory surgery performed in hospitals and freestanding ambulatory surgery centers
Kids' Inpatient Database	KID	Health care utilization and cost	Database of hospital inpatient stays for children to examine a broad range of conditions and procedures related to child health issues. Part of Healthcare Cost and Utilization Project
National Health and Nutrition Examination Survey	NHANES	Health care utilization and cost	Continuing series of national sample surveys of households and household members to assess health and nutritional status of adults and children in the US

REFERENCES

1. Modern Elementary Statistics. Fifth ed. Englewood Cliffs, NJ: Prentice Hall, 1979.
2. Escarce JJ, McGuire TG. Methods for using Medicare data to compare procedure rates among Asians, blacks, Hispanics, Native Americans, and whites. Health Serv Res 2003; 38:1303-17.
3. The [output/code/data analysis] for this paper was generated using SAS software, Version 9 of the SAS System for Linux. Copyright © 2002 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.
4. Rothenberg R, Bross DC, Vernon TM. Reporting of gonorrhea by private physicians: A behavioral study. Am J Public Health 1980; 70:983-6.
5. Rothenberg RB. Analysis of routine data describing morbidity from gonorrhea. Sex Transm Dis 1979;6:5-9.
6. Chorba TL. Disease surveillance. In: Thomas JC, Weber DJ, eds. Epidemiologic methods for the study of infectious diseases. New York: Oxford University Press, 2001:138-162.
7. Chorba TL, Berkelman RL, Safford SK, Gibbs NP, Hull HF. Mandatory reporting of infectious diseases by clinicians. JAMA 1989; 262:3018-26.
8. 1998 guidelines for treatment of sexually transmitted diseases. Centers for Disease Control and Prevention. MMWR Recomm Rep 1998;47:1-111.
9. National Drug Data File Plus Tm reference software. First Data Bank Inc; San Bruno, CA.
10. Drug Topics Redbook. 107th ed. Montvale, NJ: Medical Economics Co., 2003.
11. Brown ML, Riley GF, Schussler N, Etzioni R. Estimating health care costs related to cancer treatment from SEER-Medicare data. Med Care 2002; 40:IV-104-17.

NOTES

[1] 2000 NAMCS Micro-data file documentation, Data Dissemination Branch, National Center for Health Statistics, 6525 Belcrest Road, Room 1064, Hyattsville, MD, 20782.
[2] These files excluded anyone with health maintenance organization (HMO) experience during any years of our analysis.
[3] Line items with place of service other than physician office, inpatient hospital, ER, ambulatory surgery, outpatient hospital, ambulance, or independent laboratory were excluded from the analysis.
[4] Outpatient claims with facility type listed as skilled nursing facilities (SNF) or home health agencies (HHA) were excluded from analysis.
[5] HIC is an acronym for Health Insurance Claim number. It is an 11-digit code made up of a nine-digit claim account number (CAN) (which is actually a social security number (SSN)) and a two-digit beneficiary identification code (BIC), which uniquely identifies multiple people claiming benefits under the same SSN.
[6] The Emergency Medical Treatment and Active Labor Act, a statute that governs when and how a patient may be (1) refused treatment or (2) transferred from one hospital to another when he or she is in an unstable medical condition.
[7] The HCFA Common Procedure Coding System.
[8] If matches of ER and ambulatory surgery were within one day of each other, then half the costs were assigned to each facility of service. Also, when the office visit line item was matched to a place of service, the non-office-visit line items that matched on HIC, provider, and date were also assigned to that place of service.

Glossary

GLOSSARY OF SELECTED TERMS

Race- The concept of race reflects self-identification by people according to the race, or races with which they mostly identify. These categories are socio-political constructs and should not be interpreted as being scientific, or anthropological in nature. Furthermore, the race categories include both racial and national-origin groups. According the Office of Management and Budget (OMB) standards, race is considered a separate concept from Hispanic origin (ethnicity).

White- A person having origins in any of the original peoples of Europe, the Middle East, or North Africa. It includes people who indicated their race as "White", or report entries such as Irish, German, Italian, Lebanese, Near Easterner, Arab, or Polish.

Black or African American- A person having origins in any of the Black racial groups of Africa. It includes people who indicated their race as "Black, African Am", or provide written entries such as African- American, Afro American, Kenyan, Nigerian, or Haitian.

American Indian and Alaska Native (North American Native)- A person having origins in any of the original peoples of North and South America (including Central America) and who maintain tribal affiliation, or community attachment.

Asian- A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent, including, for example, Cambodia, China, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam. It includes "Asian Indian," "Chinese," "Filipino," "Korean," "Japanese," "Vietnamese," and "Other Asian."

Pacific Islander- A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands. It includes people who indicate their race as "Native Hawaiian," "Guamanian or Chamorro," "Samoan," and "Other Pacific Islander."

Other race- Includes all other responses not included in the "White," "Black or African American," "American Indian and Alaska Native," "Asian," "Native Hawaiian," and "Other Pacific Islander" race categories described above. Respondents providing write-in entries, such as multi-racial, mixed, interracial, or a Hispanic/Latino group (for example, Mexican, Puerto Rican, or Cuban) in the "Some other race" category are included here.

Ethnicity- The heritage, nationality group, lineage, or country of birth of the person, or the person's parents, or ancestors before their arrival in the United States.

Hispanic- Persons of Cuban, Mexican, Puerto Rican, South or Central-American, or other Spanish culture, or origin, regardless of race.

Region- The United States is grouped into four regions of states corresponding to those used by the US Census Bureau:

Northeast- Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, and Pennsylvania

Midwest- Michigan, Ohio, Illinois, Indiana, Wisconsin, Minnesota, Iowa, Missouri, North Dakota, South Dakota, Nebraska, and Kansas

South- Delaware, Maryland, District of Colombia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida, Kentucky, Tennessee, Alabama, Mississippi, Arkansas, Louisiana, Oklahoma, and Texas

West- Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada, Washington, Oregon, California, Hawaii, and Alaska

Public Use Microdata Sample (PUMS)- These files are a sub-sample from the American Community Survey (ACS) and show the full range of population and housing unit responses collected on individual ACS questionnaires. These data are used for variables not commonly offered by the US Census bureau. Questionnaire data includes: age, sex, tenure, income, education, language spoken at home, journey to work, occupation, condominium status, shelter costs, vehicles available, and other subjects.

Urban Area- Urban consists of urbanized areas and other urban entities. An urban area consists of densely settled territory with a population of 50,000, or more inhabitants. Other urban areas have from 2,500 to 49,999 populations.

Rural- Territory, population, and housing units not classified as urban.

Source of payment

Medicare- The health insurance program for the aged and disabled administered by the Centers for Medicare and Medicaid Services.

Medicaid- A jointly funded Federal-State health insurance program providing medical care to those unable to afford it.

Private insurance- A private insurance plan not specified as an HMO/PPO. This includes Blue Cross/Blue Shield plans, medical coverage provided by life insurance companies, health insurance companies, and independent plans such as employer/non-sponsored plans and /or self-funded plans (partial or total).

HMO/PPO- Any Health Maintenance Organization (HMO), or Preferred Provider Organization (PPO) sponsored by consumers, communities, physicians, or hospitals.

Self-pay- The majority of the costs for the visits were paid by the patient, spouse, family, or next-of-kin.

Other insurance- Includes any non-profit source of payment (such as church welfare, United Way, or Shriner's Hospitals for Children).

Poverty Income Ratio- This is a calculated variable based on family income and family size using tables published each year by the bureau of the Census in a series "Current Population Reports" on poverty in the United States. The primary reporting categories are:

0.00-0.999 (Below poverty)
1.000 and above (At or above poverty)
Or
0.00-1.850 (Low)
1.851-3.500 (Middle)
3.501 and above (High)

Primary Diagnosis- The condition that is determined during the hospital stay to be the chief reason for causing the hospital admission.

Any Diagnosis- Includes primary diagnosis and additional conditions that coexist at the time of admission, or that develop during the stay, and which have an effect on the treatment, or length of stay in the hospital.

Discharge status- The disposition of a patient at the time of discharge from an inpatient facility.

NIH Publication No. 12-7865