Urologic Diseases Methods

OVERVIEW

The purpose of this Urologic Diseases in America (UDA) project was to assess the burden of illness imposed upon the United States by urinary incontinence.

DATABASE SOURCES

Databases selected to study urinary incontinence 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 a 5% sample (which was thenv appropriately weighted to represent the national Medicare population). These datasets include Medicare inpatient (MedPAR file and inpatient claims file) data, 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. The second source we examined was the National Health and Nutrition Examination Survey (NHANES), a population- based survey, for items that could be used to create estimates of population-based prevalence. The third dataset allowed us to make utilization and cost estimates regarding the commercially insured population. The Marketscan Commerical Claims and Encounters dataset contains claims for inpatient stays, physician office, and hospital outpatient utilization. Data on patient medication use and lab values are also available in Truven Marketscan. The combination of databases (Medicare, nationally representative datasets, and Marketscan allowed us to complete a comprehensive evaluation of the following primary service utilization categories: (1) inpatient stays, (2) physician office visits (3) other outpatient visits for urinary incontinence. 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 2011 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 types: the MedPAR and inpatient files, 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. The carrier and outpatient files contained individual claims for provider services, and the MedPAR and inpatient samples contained information on hospitalizations incurred by those same Medicare enrollees. We also used the Medicare part D which records the prescription utilization for those individuals enrolled in the Medicare Part D program.

Analytic Approach

Data from the Medicare files (MedPAR, inpatient, 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. The procedure requires identifying inpatient hospital stays. The MedPAR file is at the stay level and was used for years 2002-2006. The inpatient claims file was used for year 2007-2011 and was processed to create a stay level file similar in structure to the MedPAR. References to MedPAR in the remainder of this document refer to the combined MedPAR/inpatient claims files.

First, personal identifiers and dates from facility records in the 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). At the direction of NIDDK, the Ambulatory, Outpatient and ER services were collapsed into a single place of service labelled "Other". 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. Specific procedures were also associated with various UI types and all claims with those procedures were included regardless if they had a UI diagnosis code (primary or other). 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 denominator for the rates of each individual disease was based on the total population counts for UI. This count included claims which had (1) either a primary UI diagnosis or a UI procedure in the Medpar file, or (2) any UI diagnosis or UI procedure from the carrier or outpatient claims files. The data were stratified by age, gender, and race variables. Confidence intervals were calculated using standard methods for proportions. In

Medicare data analyses, 5% samples are considered adequate for meaningful comparisons among different minority, geographic, and age groups .

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

NATIONALLY REPRESENTATIVE HEALTH CARE UTILIZATION DATA

Description

We used NHANES datasets to derive nationally representative estimates of prevalence of urinary incontinence. 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, multistage, 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 from 2001-2010.

Analytic Approach

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, health status, and reproductive (women only) variables. Using the weights provided by the NCHS, raw counts were weighted to give nationally-representative estimates of disease prevalence.

National estimates of prevalence for the 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, two conditions must be met for creation of reliable national estimates: (1) there must be at least 10 unweighted counts, and (2) estimates must have a relative standard error (RSE) of less than 30 percent. When adequate unweighted counts were unavailable, values corresponding to those counts were suppressed. Population weights were applied to unweighted counts, according to the methodology provided by each organization sponsoring a survey, to obtain national estimates of the prevalence of disease in the entire population and in subpopulations of interest. SAS was used to compute the 95% confidence intervals (CIs) for these estimates. The sample design of the database was taken into account when computing statistics to ensure the proper estimation of variance in each case.

Stratification variables evaluated for databases include age, race/ethnicity, gender, education, income, marital status, health status variables such as body mass index (BMI) and smoking history, and reproductive variables of interest such as number of births and incidence of hysterectomy.

COMMERCIALLY INSURED POPULATION

Marketscan Description

The Marketscan dataset contains claims from 2002-2011 and represents approximately approximately 45-47 million commercially insured individuals from over 100 commerical payers. The Marketscan datasets contain medical and prescription claims, and an enrollment file. The Marketscan 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 Marketscan. In addition to utilization, the Marketscan provides drug claims, which consist of prescription fill date, refills, brand name, therapeutic class and cost. 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 Marketscan dataset, line item claims were designated to inpatient, physician office, and other sites of care. 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 inpatient stays, physician office visits and visits in the emergency room, ambulatory surgery centers and outpatient settings 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

<u>Overview</u>

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. The total costs incurred were summarized for each urologic disease of interest, by place of service, patient population, and calendar year.

<u>Data Sources</u>

The 2002-2011 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.

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 Marketscan claims database of privately insured individuals.

<u>Analytical Approach</u>

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 project, the analysis focused on direct costs only as data on lost work days were not available. Total Payments were calculated as the sum of five type of payments, (1) Net patient payments, (2) coordination of benefits (COB) and Other Savings, (3) Coinsurance, (4) Copayment, and (5) Deductibles.

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 or procedure of a urologic condition. That is, urology-related costs that are secondary to the non-urologic 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.

APPENDIX A: TECHNICAL PROGRAMMING FOR MEDICARE DATA

This appendix describes the process by which data from the Medicare MedPAR, inpatient, 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, inpatient, carrier and outpatient files for a 5% sample of Medicare beneficiaries were used in building the files for this research effort. The MedPAR files contain summary records for all inpatient stays. The inpatient file contains detailed claims information associated with inpatient stays. Before building the analysis file the inpatient file was used to create a "synthetic MedPAR" for years 2007-2011. The process involved grouping claims records, by examination of dates of admission, discharge and inpatient services, and aggregating records identified as belonging to a stay into a single record. This process created a stay record that was functionally identical to the MedPAR and which was used as input to the analysis file in the same manner as the MedPAR (which was used for year 2002-2006). The carrier file contains detailed information at the line-item level, which provided information on payment and place of service by line item. 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.

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. For reporting purpose we summed the number of visits that occurred in the ER, Ambulatory Surgery Center and Outpatient Silos into an Other silo.

CREATING THE FILES

The Inpatient Analysis File

Inpatient stays were identified from MedPAR (the "synthetic MedPAR for years 2007 to 2011) 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, 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 emergency medical screening services
- Emergency room-emergency room beyond EMTALA screening
- Emergency room-urgent care (effective 10/96)
- Emergency room-other

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 90 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 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. 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

Two type of denominators were calculated depending on the level of detail in each table. For the base UI table, which includes all persons having a UI condition, defined as those with a primary UI diagnosis or UI procedure in an inpatient stay or any UI diagnosis or UI procedure in an outpatient setting, the denominator was 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. For individual UI disease tables, which include only persons with a specific UI condition, the denominator was the population of individuals in the base table identified as having any UI condition as described above.

Weighting

The Medicare claims files, MedPAR/inpatient file, 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.

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 inpatient 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 2011.

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: These data sets in combination provide 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 in the Medicare data are 65 years of age and over. This UI analysis excludes those under 65.

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: NHANES is currently a continuing survey, with the latest data release that includes relevant UDA variables covering 2009-2010.

Sample Size: The sample for NHANES data from 2009-2010 includes approximately 10,537 people who were interviewed and 10,253 who were examined. Similar numbers of individuals were interviewed and examined in previous and subsequent cycles.

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
Truvan Marketscan		Commercial Claims Data	Medical claims database providing utilization and cost data for private sector
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

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.