

Methods

Christopher S. Saigal, MD, MPH

*Assistant Professor of Urology
David Geffen School of Medicine
University of California, Los Angeles
RAND Health, Santa Monica, California*

Geoffrey F. Joyce, PhD

*Economist
RAND Health, Santa Monica, California*

Sandy A Geshwind, DrPH

*Natural Scientist
RAND Health, Santa Monica, California*

Mark S. Litwin, MD, MPH

*Professor of Urology and Health Services
David Geffen School of Medicine and School of Public Health
University of California, Los Angeles
RAND Health, Santa Monica, California*

Contents

OVERVIEW	699
DATABASE SOURCES	699
MEDICARE DATA	700
Description	700
Analytic Approach	701
NATIONALLY REPRESENTATIVE HEALTHCARE UTILIZATION AND COST DATA	701
Description	701
Analytic Approach	703
SPECIAL POPULATION DATA	704
Description	704
Analytic Approach	708
SPECIAL NOTES ON THE SEXUALLY TRANSMITTED DISEASES CHAPTER	711
ESTIMATING COSTS ASSOCIATED WITH UROLOGIC DISEASES	714
LIMITATIONS	717
DATA QUALITY	717
APPENDIX A: TECHNICAL PROGRAMMING FOR MEDICARE DATA	719
CREATING THE FILES	719
APPENDIX B: SUMMARY OF DATASETS	722
NOTES	734
GLOSSARY OF SELECTED TERMS	736

Methods

Christopher S. Saigal, MD, MPH
Geoffrey F. Joyce, PhD
Sandy A Geshwind, DrPH
Mark S. Litwin, MD, MPH

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 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 available linked to information from the Surveillance Epidemiology and End Results (SEER) database.

The second group of datasets allows computation of national estimates of healthcare 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 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. We supplemented our analyses of these databases with the household component of the Medical Expenditure Panel Survey (MEPS), a population-based survey. We used the MEPS data to create nationally representative estimates of expenditures on diseases of interest. 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 group of datasets was selected to provide greater depth on special populations and topics of interest. This group included the National Association of Children's Hospitals and Related Institutions (NACHRI) dataset, the Kids' Inpatient Database (KID), the National Nursing Home Survey (NNHS), the Society of Assisted Reproductive Technology (SART) database, the Veterans Health Administration (VA) Outpatient Clinic (OPC) dataset, The Veterans Health Administration National Surgical Quality Improvement Project (NSQIP) database, The Veterans Health Administration Pharmacy Benefits Management version 3.0 database, the urology subset of the MarketScan Health and Productivity Management (HPM) database, private claims data from the Center for Health Care Policy and Evaluation (CHCPE), and the Ingenix claims dataset. Data from CHCPE, KID, and NACHRI were used to enhance analysis of the burden of urologic illnesses on the pediatric population. The NNHS provided information on individuals residing in nursing homes. The SART database allowed assessment of male factor infertility as part of the burden of overall infertility. The VA databases allowed description of veterans' use of outpatient services for urologic care. Data from Ingenix were used to model costs of care for various urologic illnesses. Because some urologic conditions have costs that accrue to employers of

affected individuals, we felt it important to measure indirect costs of illness as well. Data from MarketScan provided unique information on indirect costs, e.g., work absences associated with medical services for urologic conditions.

The combination of databases (Medicare, nationally representative datasets, and special populations and topics) allowed us to complete a comprehensive evaluation of the following primary service utilization categories: (1) inpatient stays, (2) physician office visits, (3) hospital outpatient visits, (4) emergency room visits, and (5) ambulatory surgery center visits for the UDA conditions in this compendium. The data also enabled us to derive estimates of disease prevalence for some conditions. 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 1992, 1995, 1998, and 2001 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, nursing home, 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 and outpatient files were evaluated to ascertain the number of visits to inpatient hospitals, ERs, hospital outpatient departments, and 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. 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 that place of service.

At the completion of the matching process, descriptive tables were generated using appropriate International Classification of Diseases, 9th ed. (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 HEALTHCARE UTILIZATION AND COST DATA

Description

We used six datasets to derive nationally representative estimates of disease-specific service use, disease prevalence, and healthcare payments. These datasets include data for inpatient stays or hospitalizations (HCUP-NIS) and data for outpatient and ambulatory care utilization (NSAS, NAMCS and NHAMCS). In addition, MEPS, a population-based survey, was used to evaluate costs. Finally, NHANES was used to determine the prevalence of urinary incontinence and urinary tract infection.

The databases assessed had different designs, depending on the goals of the surveys they represented. The NAMCS and NHAMCS databases used a nationally representative multistage 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 22,000 to 35,000 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 HCUP database is also a nationally representative probability sample, but rather than using a multistage 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 6 million to 7.5 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 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.

MEPS is a nationally representative survey of healthcare service use and expenditures conducted under the auspices of the Agency for Healthcare Research and Quality (AHRQ). MEPS relies on self-reports and medical record abstraction and describes utilization of all healthcare services, expenditures, sources of payment, and insurance coverage by individuals in the US civilian non-institutionalized population. The data are collected five times per year, the first collection having been made in 1996. The years evaluated for this compendium are 1996, 1997, and 1998. The sample includes approximately 10,000 families, or 24,000 individuals, per year. Medical expenditure data at both the person and the household level are continuously collected for the database, which has an overlapping panel design. Two calendar years of data are collected from each household in a series of five rounds. These data are then linked with additional information collected

from the respondents' medical providers, employers, and insurance providers. The series of data collection activities is repeated each year on a new sample of households, resulting in overlapping panels of survey data from 195 communities across the nation.

The MEPS database is particularly valuable for the purposes of this compendium because it contains detailed information on utilization and payments across treatment settings. In addition, the medical provider component supplements and validates self-reported information in the household component. However, to preserve respondent confidentiality, nearly all of the condition codes in the MEPS file have been collapsed from fully specified (five-digit) ICD-9 codes into three-digit code categories. This limits the ability to examine certain conditions, such as urinary incontinence and urinary tract infection. Also, the sample sizes are relatively small, so unusual urologic conditions are not captured well in the data. Finally, there may be underreporting of some conditions because the data are obtained from self-reports of illness (though these reports are later followed up by abstraction of medical charts and financial data).

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 NHANES III, 1988–1994 and NHANES data from 1999–2000.

The Surveillance, Epidemiology, and End Results (SEER) Program maintains several population-based registries in the United States and Puerto Rico and provide 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 percent of African Americans, 40 percent of Hispanics, 42 percent of American Indians and Alaska Natives, 53 percent of Asians, and 70 percent of Hawaiian/Pacific Islanders. The database contains information on 6 million in situ and invasive cancers diagnosed between 1973 and 2003; 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 US. SEER provides authoritative genitourinary cancer prevalence and incidence data which provide context for trends in expenditures and utilization documented in the other UDA datasets.

Analytic Approach

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 healthcare 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 2002. 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).

NAMCS, NHAMCS, HCUP, KID, MEPS, NHANES, NSAS

The years of NAMCS, NHAMCS and HCUP data analyzed are 1994, 1996, 1998, and 2000. In addition, the 1992 NAMCS data were reviewed. The MEPS data we evaluated were for 1996 through 1998. NSAS data were available from 1994 to 1996. KID data were available for 1997 and 2000.

First, we identified individuals with visits for specific urologic conditions, based on the ICD-9 diagnosis or procedure codes or SEER disease codes that defined each of the conditions and any age and gender specifications necessary to create subpopulations for the analyses (for MEPS data, three-digit ICD-9 diagnosis codes were used). 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. The MEPS database was used to calculate payments for all services, as well as to derive nationally representative estimates of outpatient prescription drug use.

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, two conditions must be met for creation of reliable national estimates: (1) there must be at least 30 unweighted counts, and (2) estimates must have a relative standard error (RSE) of less than 30 percent^a. 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 ($N \geq 30$), and therefore no combining or regrouping of stratification variables was necessary.

Population weights were 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 subpopulations of interest. SAS (3) was used to derive the standard errors and 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.

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)^b 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.

SPECIAL POPULATION DATA

Description

The data sources used for special-population analyses target an array of unique populations not completely captured in the databases described above. These include children, the elderly, veterans, and two populations that allow us to combine data to conduct a cost analysis—the privately insured and the employed. Together, these datasets, along with the others described in this compendium, provide a unique picture of the relationship between urologic diseases in America and their impact on healthcare utilization, services, and costs.

NACHRI

The National Association of Children's Hospitals and Related Institutions (NACHRI) maintains a dataset containing information on all inpatient stays at member hospitals, including approximately 2 million pediatric inpatient discharges. Data have been collected annually since 1999. Fifty hospitals located in 30 states participated in 1999, 55 participated in 2000, and 58 participated in 2001. Data include the age, race, sex, and ICD-9 code for the principal diagnosis of each pediatric inpatient cared for at participating facilities. Information on length of stay, total charges, and cost-to-charge ratio is also collected. Because it collects data from children's hospitals, the NACHRI dataset provides a unique opportunity to study the inpatient burden placed on the healthcare system by relatively uncommon pediatric urologic conditions. The dataset is rigorously edited and cleaned to ensure data quality. However, because NACHRI collects data from specialized facilities, its information on such topics as length of stay, patient demographics, and treatment costs may not be representative of the national experience.

NNHS

The National Nursing Home Survey (NNHS) is a series of national sample surveys of nursing homes, the providers of care, and their residents. The dataset contains information on a sample of approximately 1,500 facilities, 8,100 residents, and 6,800 discharge records. The data are collected using a nationally representative, stratified, two-stage probability sampling design. All nursing homes in this sample

have at least three beds and are certified by Medicare or Medicaid or have a state license to operate as a nursing home. Characteristics of each facility, including size, ownership, occupancy rate, number of days of care provided, and expenses, are collected. Additionally, information is provided on the recipients of care, including demographics, health status, and services received. One of the unique aspects of using this database is that it provides information on nursing homes from two perspectives: that of the nursing home facility and that of the residents themselves. However, the survey does not provide detailed information on the health services provided. It indicates only whether the patient received a service from within a general service category. Also, the records for the survey years reviewed for this compendium do not contain facility numbers to allow linkage of the records to the facilities.

KID

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 were available for 1997 and 2000. 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. While the NACHRI dataset may allow description of utilization for relatively uncommon pediatric conditions which may be referred to tertiary care centers, KIDS' sampling design allows nationally

generalizable observations to be made about inpatient utilization for pediatric conditions.

SART database

The Society for Assisted Reproductive Technology (SART) database contains outcome information for approximately 350 fertility clinics nationwide. SART monitors fertility clinic outcomes, in concert with the Centers for Disease Control and Prevention (CDC) in compliance with the Fertility Clinic Success Rate and Certification Act of 1992 (Wyden Act). The database contains information from the large majority of fertility clinics operating nationally, and allows for uniform reporting of outcome variables of interest. Clinical information from each male involved in a fertility procedure is collected. Hence, SART data can be used to examine the burden of male factor infertility in terms of related use of fertility services, the costs of which are often borne by patients. It can also be used to examine the effect of male factor infertility on subsequent outcomes.

MarketScan HPM

The MarketScan Health and Productivity Management Database (HPM) is an integrated inpatient and outpatient medical claims database that provides information on productivity losses associated with medical services. The data contain key information on short-term disability, absence, and worker's compensation resulting from urologic conditions. Absence data are derived from employee time-reporting records collected through employer payroll systems and contain detailed information on when employees were out of work, the number of work hours missed, and the reasons for the absences. Information on work absence can be linked to eligibility files and medical claims data. The linked files allow users to examine medical treatment and its association with work loss and disability. Although the database includes employers from all areas of the country, the data are not nationally representative.

Ingenix

Data for individual-level analyses were obtained from Ingenix, Inc., a healthcare information company that provides cost management and benefit consulting services to employers, health plans, pharmaceutical manufacturers, and others. Data

were obtained on insurance eligibility and medical claims for the employees of 25 large (Fortune 500) companies and their dependents. All individuals had private, employer-sponsored health insurance coverage, including prescription drug benefits. Out of concern that data on healthcare use and costs might be incomplete for the employees' dependents (e.g., in cases of dual coverage), we excluded dependents from the analysis. The sample we used consisted of 278,950 primary beneficiaries 18 to 64 years of age who were continuously enrolled for the entire 1999 calendar year.

The medical claims include detailed financial information, dates of service, diagnosis and procedure codes, types of facility, and providers. Drug claims include prescription fill date, patient and plan costs, and, in most cases, national drug codes (NDC), which were used to examine utilization of specific drugs and therapeutic classes. Claims data contain records only for those who used services. To identify those who may not have used services, enrollment data were also obtained. Enrollment files included each person's age, sex, plan type (FFS, PPO, POS, HMO), zip code of residence, and relationship to employee.

The claims data were linked with information about plan benefits. For each plan, photocopies of the summary of benefits provided by the companies to their employees were obtained, and the benefit information, including the salient features of each plan's covered benefits, was abstracted from its summary-plan design. Drug benefits typically included copayments or coinsurance rates for both retail and mail-order pharmacies; the data on drug benefits included generic substitution rules and a list of drugs or drug classes excluded from coverage. Characteristics of the medical benefit included plan deductibles and patient cost-sharing arrangements for inpatient and ambulatory settings.

Center for Health Care Policy and Evaluation

The Center for Health Care Policy and Evaluation (CHCPE) of UnitedHealth Group maintains a centralized research database (RDB) that contains current and historical medical and pharmaceutical claims and enrollment data for 27 geographically distinct health plans—more than 22 million member-years of data. The majority of UnitedHealthcare members are enrolled through employers (i.e., they are commercially insured), although membership also includes Medicaid and Medicare beneficiaries. The health plans contract with broad networks of physicians, pharmacies, and facilities to provide healthcare to enrollees. Most providers are reimbursed on a fee-for-service basis; pharmacies receive dispensing fees; and most facilities have contracted per diem rates or prospective payments (DRGs). Because information on pediatric urologic diseases in the nationally representative data sources was sometimes limited, data from CHCPE were used to enhance understanding of the burden of urologic disease on the pediatric population. One limitation of this data source is that it is drawn from an insured population, which may differ in important ways from the uninsured population. CHCPE does collect information on children with Medicaid, and these data were reported.

The RDB contains automated utilization data from all types of healthcare sites (e.g., hospital inpatient, hospital outpatient, physician office, emergency department, surgery center) and for all types of services, including care from out-of-network providers. CHCPE assembles enrollment and claims data generated by UnitedHealthcare operations into the RDB as four component files (enrollment, physician, facility, and pharmacy). Table 1 shows the total number of members enrolled in the 15 plans

Table 1. Health plan population and claims count, by year, region, and insurance type

Region	1994		1996		1998		2000	
	Commercial	Medicaid	Commercial	Medicaid	Commercial	Medicaid	Commercial	Medicaid
Midwest	637,093	46,009	887,957	39,192	1,186,702	274	1,350,819	442
Northeast	151,405	816	159,953	34,986	200,994	36,079	164,354	41,604
Southeast	205,934	11,590	471,528	59,600	976,050	46,452	1,099,531	8,266
West	76,084	563	90,689	13,318	109,654	20,599	134,537	29,451
Total	1,070,516	58,978	1,610,127	147,096	2,473,400	103,404	2,749,241	79,763

selected for this project, stratified by year, region, and type of insurance.

VA datasets

VA-OPC

The Department of Veterans Affairs (VA) delivers healthcare to eligible veterans through the Veterans Health Administration. The VA is the largest healthcare system in the United States, comprising more than 160 hospitals (>45,000 beds), more than 600 community-based outpatient clinics, and more than 100 nursing homes. The VA maintains a centralized data repository reflecting healthcare utilization by the population of veteran users. This repository, known as the Austin Automation Center, contains computerized utilization data on many types of health services, including outpatient visits. The SAS files created from this database allow for file linkages of patients.

The VA outpatient clinic (VA-OPC) files, on which the UDA analyses were performed, include demographics, visits, and clinic stops (i.e., different clinic appointments and services attended in a given visit day) and are available for 1980 to the present. Ambulatory procedures were added to the OPC in 1990, and outpatient diagnoses (ICD-9-CM) were added in FY1997. These datasets provide a rich resource for assessing the prevalence of disease among VA healthcare users. The ability to link files across VA healthcare facilities and across settings within facilities allows a relatively complete portrait of utilization and patterns of care to be obtained. However, the VA datasets do not provide comprehensive information about veterans' healthcare utilization *outside* the VA healthcare system.

The diagnosis codes were derived from outpatient visits from recent physician-patient encounters and thus *do not reflect all existing or historical cases among veteran users*; instead, they reflect the population for whom care was sought during the year being reviewed. Therefore, prevalence based on counts of cases in a given fiscal year of outpatient utilization data is likely to underestimate prevalence in the total population of users.

NSQIP

The VA National Surgical Quality Improvement Program (NSQIP) collects pre-operative risk factors,

intra-operative variables, and 30-day post-operative mortality and morbidity outcomes on most major surgical operations performed in the VA healthcare system. An assessment is done on all patients undergoing a major operation under general, spinal, or epidural anesthesia. Some operations with known low mortality and morbidity are excluded. Also, certain common operations (transurethral prostatectomy, transurethral bladder tumor resection, inguinal hernia) are limited to the first 5 cases in an 8-day cycle. Most VA hospitals collect all of their major cases. A few high-volume hospitals are permitted to take a sample of cases (the first 36 consecutive cases in each 8-day cycle, with each cycle beginning on a different day of the week). NSQIP was initiated in 1991 at the 44 largest VA hospitals that perform both cardiac and non-cardiac surgery. In 1994, the Program was expanded to include all 132 VA medical centers that perform surgery. Each year, risk-adjustment models are created and risk-adjusted 30-day mortality and morbidity surgical outcomes are fed back to the medical centers, so that they can compare their results to other VA medical centers on a blinded basis. A dedicated NSQIP nurse, trained in the data collection procedures, collects the data at each VA medical center. Procedures for the Program are documented in an operations manual, including definitions of all variables. The database contains about 1 million operations, and about 100,000 cases are added each year.

The VA NSQIP is funded by the VA's Office of Patient Care Services in VA Central Office in Washington, D.C. The chairman's office is at the West Roxbury, MA VA medical center. The data coordinating center is located at the Denver VA medical center and the University of Colorado Health Outcomes Program under the direction of Dr. William Henderson, Co-chairman of the NSQIP.

PBM

Disease-specific medication data for veterans were obtained from the Pharmacy Benefits Management version 3.0 database (PBM v3.0). The PBM national database includes medication utilization information based on unique patients for every prescription filled in the VA. The database produces information on each occasion of drug dispensation, which helps to track total prescription usage on a gross or discrete

level. Data originate from the VISTA pharmacy packages at each VA facility and are sent monthly via computer to the PBM office in Hines, IL. Upon receipt, the PBM collects, processes, and translates the information into the national database format. The database includes comparative data for all VA facilities nationwide and data can be aggregated nationally, by veteran integrated service network (VISN), which approximates geographic regions, or by facility. The database contains information from October 1, 1998 to the present. The information in the PBM uses patient-specific social security numbers so that other needed clinical data can be easily linked to the pharmacy data.

Analytic Approach

NACHRI

After obtaining a list of ICD-9 diagnosis codes used to define the pediatric UDA conditions of interest, researchers at NACHRI created an analytic file containing all inpatient discharges reported at member institutions for which a UDA diagnosis code appeared as a primary diagnosis for admission. Information was analyzed for FY 1999–2001. Data were imported into SPSS 4.0 (4), and available stratification variables were examined. Mean values and counts for these variables were compared with those provided by NACHRI to ensure data integrity after importation. One-way ANOVA was used to generate confidence intervals for frequencies and means of desired variables.

NNHS

The years of data evaluated for this dataset were 1995, 1997, and 1999, and the unit of analysis was the individual. The analytic approach follows that used for the nationally representative datasets, with the exception that urolithiasis, sexually transmitted diseases (STDs), and pediatric conditions were excluded due to small sample sizes.

KID

The KID database was evaluated using the approach described for the nationally representative datasets (described above).

CHCPE

Records from component files maintained by CHCPE were linked, using common fields such as member and physician identifiers and dates of service, as described below.

CHCPE enrollment records serve to track plan membership for billing premiums. The enrollment file includes date of birth, gender, and dates of enrollment and disenrollment. The physician file contains information submitted by physicians and other healthcare practitioners, using the CMS-1500 claim form. This file includes the member identifier, unique provider identifier, the service or procedure performed, up to four diagnosis codes, the place of service, billed amounts and payments, and the insurance product under which the service falls. Diagnoses are coded according to the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM), and procedures are coded using the Current Procedural Terminology (CPT) coding system. The facility file contains information submitted by facilities using the CMS-145 claim form. This file includes the member identifier, unique facility identifier, facility type, revenue codes, up to nine diagnosis codes, the place of service (e.g., inpatient, outpatient, emergency room), DRG payments, and the insurance product under which the service falls. As with physician claims, diagnoses are coded using ICD-9-CM codes and procedures are coded using ICD-9-CM procedure codes or CPT-4 procedure codes. Claims from out-of-network facilities are included.

The period of analysis for this compendium included 1994, 1996, 1998, and 2000. Analyses were conducted on members of 15 commercial and Medicaid health plans located in four regions (Midwest, Northeast, Southeast, West) of the United States. Data on commercial and Medicaid health plan members were reported separately, as these populations tend to differ in socioeconomic status.

VA datasets

VA-OPC

The VA-OPC file was used to identify all unique cases of each urologic condition. The event (SE) files of the OPC, which combine the diagnostic and procedural information, were used for generating these data. Three consecutive years of data, 1999 through 2001, were examined. A unique count of

cases was generated by identifying the cohort of veterans with each qualifying ICD-9 code within each UDA urologic condition under review. Redundant cases (i.e., individual patients with more than one qualifying diagnosis code) were then eliminated in order to generate a count of unique cases. Two cohort files were created: prevalence estimates (1) by *first (or primary) diagnosis* and (2) by *all diagnosis* codes (i.e., any mention of eligible diagnostic codes from our case definition).

Given the size of the national VA utilization datasets, all initial diagnosis groups and linkage procedures were pilot-tested on *local regional data* to

examine preliminarily the prevalence of the selected conditions for one year of data. An initial set of tables was created showing the national prevalence of the first four UDA conditions (benign prostatic hyperplasia, urolithiasis, urinary tract infection, and urinary incontinence) *for all diagnoses*. Prevalence rates were presented in total and by selected demographic characteristics (e.g., age, gender, race/ethnicity) and geographic features (e.g., region) as unique cases per 100,000 population of veteran users served by VA in fiscal year 2001 (October 1, 2000, through September 30, 2001). These analyses were then expanded to the other data years. The unit of analysis in the Austin

Table 2. Base population of veteran users of VA healthcare and total veteran population by age, gender, and race/ethnicity

	Veteran Population			
	Veteran Users of VA Healthcare (VA Outpatient Clinic file)			Total Veteran Population (US Census, 2000) ^a
	1999	2000	2001	2000
Total	3,039,688	3,276,298	3,691,533	26,403,703 ^b
Age				
18–24	25,328	23,526	23,033	16,740,194
25–34	150,809	142,082	135,854	
35–44	330,512	312,179	299,820	
45–54	689,196	716,638	753,251	
55–64	501,642	554,117	648,880	
65–74	758,809	825,990	950,660	9,663,506
75–84	536,269	643,466	801,677	
85+	49,123	58,305	78,358	
Gender				
Male	2,898,582	3,125,448	3,526,627	24,810,000
Female	141,106	150,850	164,906	1,593,000
Race/ethnicity				
White	1,366,295	1,472,022	1,610,947	21,888,669 ^c
Black	333,719	342,547	354,807	2,561,159 ^c
Hispanic	114,386	122,469	128,930	1,135,359 ^c
Other	19,356	20,248	21,822	1,240,974 ^c
Unknown	1,205,932	1,319,017	1,575,027	...

...data not available.

^aSOURCE: US Census 2000, Veterans (May 2003). Questions on veteran status asked "Has this person ever served on active duty in the U.S. Armed Forces, military Reserves, or National Guard?" The question was followed by a note "Active duty does not include training for the Reserves or National Guard, but DOES include activation, for example, for the Persian Gulf War." Response categories included the following:

- Yes, now on active duty
- Yes, on active duty in past, but not now
- No, training for Reserves or National Guard only → skip
- No, never served in the military → skip

^bVeterans comprised approximately 12.7% of 208.1 million civilians 18 years and older in the US in 2000.

^cDerived from U.S. Census table as 82.9% White alone (not Hispanic or Latino), 9.7% (Black or African American alone), 4.3% (Hispanic or Latino, of any race), 4.7% other (combined American Indian and Alaskan Native alone, Asian alone, Native Hawaiian and Other Pacific Islander alone, some other race alone, or two or more races) as a percentage of total veteran population in 2000. Note that the percentage for Hispanic includes any race because the available veteran census breakdown did not cross-tabulate race and ethnicity, yielding an overcount in these cells.

Automation Center (AAC) SAS datasets is the patient. A patient who had more than one qualifying diagnosis code was counted as only a single case. Similarly, a patient with one or more of the qualifying diagnosis codes at more than one VA healthcare facility was counted only once. Prevalence rates were stratified by patient characteristics (i.e., age, gender, race/ethnicity, insurance status,^c and region).

Where possible, the VA sociodemographic categories (e.g., 10-year age groups for adults, such as 25–34, 35–44, etc.) were made consistent with those applied to the CMS data. Categories were regrouped as necessary to maintain adequate cell sizes for analysis.

The VA data represent the population of all veteran users of VA healthcare services during the years under study (pediatric cases are excluded from the VA database). Therefore, confidence intervals were included for the calculated rates. Denominator data were obtained for all veteran outpatient users and then refined, based on age, gender, or other restrictions of the UDA conditions, to generate unadjusted prevalence rates for the number of cases per 100,000 population.

The VA data presented are unweighted frequencies representing population prevalence rates among all veteran users of VA healthcare in a given fiscal year. No effort was made to weight veteran users of VA healthcare services to the total US veteran population. Table 2 presents denominator data on this base population analyzed in VA tables that appear in this compendium.

NSQIP

Risk-adjustment models are created each year for all operations combined, and for 9 major subspecialties, including urology. Individual substudies using the database have created risk-adjustment models for some specific operations and for some specific post-operative complications. The post-operative complications that are collected include:

- Post-op wound complications—superficial infection, deep wound infection, dehiscence
- Respiratory complications—pneumonia, unplanned intubation for respiratory/cardiac failure, pulmonary embolism, failure to wean >48 hours

- Urinary tract complications—progressive renal insufficiency, acute renal failure, urinary tract infection
- Central nervous system complications—stroke, coma >24 hours, peripheral nerve injury
- Cardiac complications—cardiac arrest, myocardial infarction
- Other complications—bleeding >4 units, graft/prosthesis failure, deep vein thrombosis/thrombophlebitis, systemic sepsis, other

PBM

The medications extracted from the PBM database for the current project included, LHRH agonists, antiandrogens, erectile dysfunction agents and combined LHRH and antiandrogen therapy (defined as >14 days of overlap). Total prescription data was obtained for each agent. Unique patient counts were then determined for each class and each individual agent within the class for fiscal years 1999 through 2003. Combined therapy was defined as an overlap of at least 14 days or greater of an LHRH agonist and an antiandrogen agent.

Data are presented as total number of unique patients on each class of agents, total unique number of patients on each individual agent nationally and per region. The four regions used are east, west, south, central and other. Data are also presented as total number of unique patients on each class of agent per VISN. A VISN is the unique way in which the Department of Veterans Affairs is divided into regions.

Rates per 100,000 are determined for national and VISN data for each class of agents. The unique rates are based on number of unique veterans in the prescription database nationwide and per VISN (Appendix B). The number of unique veterans per region is not available in the prescription database. Hence, regional data are presented as raw unique numbers per year only.

SPECIAL NOTES ON THE SEXUALLY TRANSMITTED DISEASES CHAPTER

The datasets used extensively for other urologic diseases throughout this compendium (HCUP, CMS, VA, and NAMCS) describe hospital discharge data, inpatient and outpatient medical claims data, and health survey data. For sexually transmitted diseases (STDs), these datasets provide sparse, incomplete, or non-representative data on numbers of cases and patient visits (Table 2) and on patient visit rates (Table 3). Several databases did not capture services rendered to many Medicaid patients, uninsured patients, or patients attending public STD, prenatal, family planning, military, or other clinics that provide a substantial amount of STD care. The inpatient datasets are limited because they enumerate hospitalizations for conditions that rarely require hospitalization. We therefore used additional data from three other databases and from the published literature to obtain more reliable estimates of the burden of the STDs discussed.

National Electronic Telecommunications Surveillance System (NETSS)

The NETSS includes reports of all cases of notifiable diseases, including STDs, sent to CDC by state health departments. After removing personal identifiers, US states and territories report cases of *C. trachomatis* or *N. gonorrhoeae* infection, primary and secondary syphilis, HIV, and hepatitis B virus (HBV). This passive surveillance system has notifiable-disease regulations and is limited by underreporting by clinicians and laboratories. Barriers to reporting include lack of awareness or priority given to public health reporting, patient or clinician concerns about violation of confidentiality for stigmatized diseases, lack of routine reporting systems, insufficient incentives, and administrative costs (5). Reporting by laboratories of cases confirmed by positive STD test results is far more complete than reporting of clinically diagnosed cases by clinicians. However, because of minimal variation in the level of completeness of such reporting from year to year, case reports provided to public health departments have historically been the best source of information about

Table 3. Codes used to identify additional medical visits for genital warts^a in MarketScan data

Any persons having one of the following two ICD-9 Codes and at least one of the following CPT procedure codes:

ICD-9 Codes:

078.10 Viral warts unspecified
078.19 Other specified viral warts
and

CPT Procedure Codes:

00900 Anesthesia for procedures on perineal integumentary system
00902 Anesthesia for anorectal procedure
00910 Anesthesia for transurethral procedures
00920 Anesthesia for procedures on male genitalia
00940 Anesthesia for vaginal procedures
45100 Biopsy of anorectal wall, anal approach
45905 Dilation of anal sphincter under anesthesia other than local
45999 Unlisted procedure, rectum
46030-46999 Surgery of anus
52000-52318 Endoscopy-cystoscopy, urethroscopy, cystourethroscopy, transurethral surgery
53000-53899 Surgery for urethra
54000-55899 Surgery for male genital system
56350-56363 Hysteroscopy
56405-58285 Surgery for female genital system
74400-74485 Radiology, urinary tract
74710-74775 Radiology, gynecological and obstetrical

^aA medical visit was identified as an additional visit for genital warts if the visit was associated with at least one of these ICD-9 codes and at least one of these CPT codes.

the temporal and geographic trends of STDs and the characteristics of infected persons (6, 7). Although a few states have had reporting requirements for herpes simplex infection, pelvic inflammatory disease (PID), and unspecified STDs (8), most states do not have reporting requirements for several STD pathogens, clinical manifestations, and syndromes for which etiologic causes or therapeutic interventions have only relatively recently been identified. These include HSV infections or genital herpes, HPV infections or their manifestations of genital warts and cervical dysplasia, and nonspecific urethritis, epididymitis/orchitis, prostatitis, cervicitis, vaginitis, salpingitis, and STD-related skin disorders. In the NETSS, pathogen-specific codes, not ICD-9 diagnostic codes, are used.

National Disease and Therapeutic Index (NDTI)

The NDTI is a private survey of a random sample of office visits to US physicians in office-based practices, using ICD-9 coding. It allows estimation of the burden and trends of diseases that are not reported by states to the national surveillance systems; these diseases include genital HSV, genital warts, nongonococcal urethritis, and trichomoniasis. However, because the NDTI does not include visits to public health, hospital outpatient, or military medical facilities, it does not permit accurate estimates of the total number of consultations in the United States for these conditions. With ICD-9 codes for abstraction, NDTI data can be used to estimate the diagnosis-specific volumes of office visits to private practitioners for various STDs (including those of reportable diseases such as syphilis, gonorrhea, and chlamydia). Because the NDTI is a national survey based on a sample rather than on the entire universe of patient visits to physician offices, patient visits must be weighted to produce unbiased national estimates.

VA-OPC

Regarding analyses of VA data, medical visits for two syndromic conditions, epididymitis/orchitis and urethritis, were classified according to whether an STD pathogen was specified in any of the four listed ICD-9 codes. For example, the visit was classified as "epididymitis, organism specified" if the first four ICD-9 codes for a given outpatient visit included ICD-9 code 604 for orchitis or epididymitis as well as one of many ICD-9 codes for gonorrhea or chlamydia or

if any of the first four ICD-9 codes were specific to gonococcal epididymitis/orchitis (098.13 or 098.33). The medical visit was classified as "epididymitis, organism unspecified" if the first four ICD-9 codes for a given outpatient visit included ICD-9 code 604 for orchitis or epididymitis but did not include ICD-9 codes for gonorrhea or chlamydia or for gonococcal epididymitis/orchitis (098.13 or 098.33).

In addition, to capture aspects of the burden of various STDs, we applied the ICD-9 codes in Table 3 to data from HCUP, MarketScan, Medicare (MEDPAR, Outpatient, and Physician-Supplier), and the VA (Inpatient and Outpatient). The burden of each STD examined (per 1,000 population) was stratified by demographic variables. The demographic variables considered in HCUP include sex, age, race/ethnicity, region, urban/rural, and expected primary source of payment. The demographic variables considered in Medicare include sex, age, race/ethnicity, region, and setting of care (inpatient, ambulatory). The demographic variables considered in MarketScan include sex, age, region, and urban/rural.

Additional analytic comments specific to each STD or condition evaluated are presented below.

Herpes

With MarketScan data, we evaluated the extent to which medical claims for genital herpes using ICD-9 codes can estimate the burden of the disease. Some care for genital herpes may not result in a diagnostic claim (ICD-9 code) for herpes but may result in a drug claim, e.g., to refill a prescription for a chronic condition. In addition, some clinicians may be reluctant to document a claim coded as genital herpes to protect the confidentiality of patient information. Therefore, we also analyzed claims for drugs (acyclovir, famcyclovir, and valacyclovir) used to treat genital herpes.

Genital Warts

Because MarketScan includes NDCs, we were able to evaluate the use of drugs for treating genital warts. We estimated claims for podofilox and podophyllin, two medications used almost exclusively to treat genital and perianal warts, regardless of ICD-9 codes assigned. We also estimated drug claims for imiquimod regardless of ICD-9 codes assigned, but only if the prescription was obtained from a urologist

or gynecologist, since imiquimod is also commonly used for nongenital warts and other cutaneous lesions commonly managed by other clinical specialties.

A diagnosis of genital warts may be coded with ICD-9 code 078.11 (condyloma acuminata); this was the sole code used in routine analysis of data from HCUP, MarketScan, Medicare, and the VA. However, genital warts may also be coded with ICD-9 code 078.10 (wart—common, digitate, filiform, infectious, viral) or 078.19 (other specified viral warts—genital warts, verruca plana, verruca plantaris), two codes that are also used for nongenital warts. Using MarketScan data, we identified the proportion of those two codes that were likely to represent genital warts. We assumed that medical visits with ICD-9 code 078.10 or 078.19 were for genital warts if the CPT code (listed in Table 3) indicated that the patient had procedures related to destruction or excision of lesions of the anus, penis, vulva, perineum, vagina, or introitus. However, classifying only ICD-9 claims with these procedures as genital warts probably substantially underestimates the number of visits for genital warts; many warts do not require any medication or ablative procedures (and associated NDCs or CPT codes) because they do not create annoying symptoms or cosmetic problems. Restricting our analysis to warts that require ablative procedures may result in minimal estimates of the burden of warts that result in the highest healthcare costs.

Using NAMCS data from 1995–1996, we also estimated the burden of genital warts as follows. We classified medical visits that may or may not have been for genital warts into three groups: definite cases (having ICD-9 code 078.11); probable cases (having ICD-9 code 078.10 or 078.19 and having either NDCs for medication with podofilox, podophyllin, or imiquimod or CPT procedure codes for destruction or excision of lesions of the anus, penis, vulva, perineum, vagina, or introitus); and possible cases (having ICD-9 code 078.10 or 078.19 and no surgical procedure code).

Chlamydia

Using MarketScan data, which include NDCs and CPT codes, we estimated the chlamydia burden using information on drugs for treating chlamydial infection and tests used to identify symptomatic *C. trachomatis* infection or to screen for asymptomatic *C. trachomatis*

infection. First, we selected medical visits with any of the codes listed in Table 17 for nongonococcal urethritis or tests used for *C. trachomatis*. Then, we defined the selected visits as being for *C. trachomatis* infection if the date of drug claims was within 7 days before or 20 days after the date of the medical visit. The drug claims included amoxicillin, azithromycin, doxycycline, erythromycin, and ofloxacin, the drugs recommended by CDC for treatment of uncomplicated, lower genital tract *C. trachomatis* infection in 1998 (9). Note that amoxicillin is a recommended regimen in CDC's guidelines when pregnancy is a consideration and that the recommended drugs are not necessarily the same ones recommended for upper-tract infections such as PID.

Gonorrhea

Using MarketScan data, which include NDCs and CPT codes, we estimated the gonorrhea burden using information on drugs for treating gonorrhea and tests used to identify symptomatic gonococcal infection or to screen for asymptomatic gonococcal infection. First, we selected medical visits with the following codes: ICD-9 code V02.7 (carrier or suspected carrier of gonorrhea) or CPT codes 87590 (gonorrhea, direct probe technique), 87591 (gonorrhea, amplified probe technique), 87592 (gonorrhea, quantification), or 87850 (*N. gonorrhoeae*). Then, we identified the selected visits having a specific drug claim within 7 days before or 20 days after the date of the medical service claims. The drug claims included cefixime, ceftriaxone, ciprofloxacin, ofloxacin, or levofloxacin, the drugs recommended by CDC for treatment of uncomplicated lower-tract gonococcal infection.

Syphilis

Data from HCUP, MarketScan, NAMCS, Medicare, and the VA were not useful for estimating the incidence of syphilis because the numbers of cases were too low to permit statistically reliable estimates. For example, in the VA data, an ICD-coded syphilis diagnosis was recorded for only 3 per 100,000 unique outpatients. Therefore, we used NETSS data, which include as variables sex, age, race/ethnicity, geographic region, and urban/rural, to estimate the incidence of primary and secondary syphilis. We included data only on individuals with primary and secondary syphilis because they are the most likely

to be seen by urologists for genitourinary symptoms and signs. However, primary and secondary syphilis cases enumerated in national data include many symptomatic cases more commonly seen by family practitioners, internists, obstetrician/gynecologists, and dermatologists, as well as asymptomatic cases identified through routine serologic screening.

Epididymitis/Orchitis

To estimate the incidence of epididymitis/orchitis, we analyzed ICD-9 codes for epididymitis/orchitis not designated as due to chlamydia or gonococcus as well as all ICD-9 codes for epididymitis/orchitis associated with chlamydia or gonococcus. We took this approach because some patients could have both a diagnosis of the syndromic presentation of epididymitis/orchitis and a diagnosis of chlamydial or gonococcal epididymitis/orchitis. We applied the definitions to data from HCUP, MarketScan, and Medicare. Epididymitis/orchitis data from the VA were available in dichotomous categories of organism specified (e.g., chlamydia, gonorrhea, other) and organism unspecified; there is no ICD-9 code to describe orchitis without epididymitis in which the organism is not specified. With HCUP, MarketScan, Medicare, and VA data, the same variables were explored as those considered for genital herpes (see above).

Urethritis

To estimate the incidence of urethritis, we analyzed the available data for urethritis not designated as due to chlamydia or gonococcus as well as aggregate data for all ICD-9 codes for urethritis, including those for chlamydia or gonococcus, listed in Table 1 of the STD chapter. We took this approach because some patients could have both a diagnosis of the syndromic presentation of urethritis and a diagnosis of gonococcal or chlamydial urethritis. We applied the definitions to data from HCUP, MarketScan, and Medicare. Inpatient and outpatient VA data report urethritis in dichotomous categories of organism specified (e.g., gonococcal, chlamydial, other) and organism unspecified. With HCUP, Medicare, VA, and MarketScan data, the same variables were explored as were considered for genital herpes (see above).

ESTIMATING COSTS ASSOCIATED WITH UROLOGIC DISEASES

Marketscan and Ingenix Data

General Methodology for Estimating Costs

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 healthcare, 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.

In this analysis, we were interested in costs as a dollar-denominated measure of resource utilization. Costs are closely linked to other important financial concepts, including charges, out-of-pocket expenses, and payments. These other concepts are more easily measured and can be used to approximate costs, but they are not necessarily equivalent. Since the majority of the databases we examined lacked cost information, direct medical costs were imputed by assigning prices to a comprehensive list of utilization and services. Prices of medical services and pharmacy claims were estimated based on average payments made by the enrollee (copayments, deductibles, excluded expenses) and by all third-party payers (primary and secondary coverage, net of negotiated discounts). The sources of utilization data were national surveys and claims records.

Deriving reliable estimates of indirect costs requires detailed information rarely included in survey data or medical claims. Even when these data are available, converting outcomes such as premature mortality, disability days, and productivity losses into costs requires a set of assumptions about the causal effects of the illness, future wage rates and retirement decisions, and the value of time for heterogeneous patient populations. The available evidence suggests that these underlying assumptions have a considerable effect on the magnitude and reliability of the cost estimates. Because of these limitations, the indirect costs of urologic conditions were not estimated. Rather, administrative data from a large number of

Table 4. Primary Data Source for Medical Care Utilization (non-Medicare population)

Type of Service	Primary Data Source
Hospital Inpatient	HCUP
Hospital Outpatient	NHAMCS
Emergency Room	NHAMCS
Physician Office	NAMCS
Outpatient Prescription Drugs	MEPS
Nursing Home	NNHS

employers were used to impute the average work loss associated with each condition.

A key issue in cost-of-illness studies is the determination of how to attribute costs to a specific condition in an appropriate and consistent way. Ideally, one would like to capture the costs of treating a urologic condition regardless of the primary diagnosis attached to the service. For estimation purposes, this means *excluding* treatment costs of other conditions incurred during a visit or hospitalization for a urologic illness and *including* urology-related costs that are secondary to the primary diagnosis. Medicare claims data permit this level of detail, reporting reimbursements associated with each unit of service, or “line-item.” However, this level of specificity is not available in most claims-based datasets and surveys. Thus for the non-Medicare population, average expenses per unit of service were estimated solely on the basis of hospitalizations or visits with a primary diagnosis of a specific urologic condition. This approach overstates average expenditures by including treatment costs of nonurologic conditions. At the same time, it understates reimbursements by excluding costs of urologic conditions that are secondary to the principal diagnosis. The net effect of this measurement error is uncertain, but it is unlikely to constitute a large fraction of the total economic burden of each condition.

Measuring Direct Costs

Measuring resource costs depends on having appropriate measures of both utilization and unit cost. Medical expenditures were estimated by assigning prices to a comprehensive list of utilization and services. For the non-Medicare population, average prices of a hospitalization, an ER admission,

a hospital outpatient visit, and a physician office visit were imputed, based on average payments reported in the 1996–1998 MEPS. In cases where MEPS lacked adequate statistical power to reliably estimate prices for specific services, average payments from a large administrative database of private employers or Medicare claims were imputed. Average prices for outpatient prescription drugs were based on published compilations from First Data Bank (10) and *RedBook* (11).

Medicare claims from 1992 through 1998 were used to impute average annual growth rates in expenditures over this period. These rates were then applied to 1996–1998 prices derived from MEPS. All expenditures for medical and pharmacy services were reported in nominal dollars.

National surveys and claims-based databases were relied upon for deriving estimates of medical service utilization by the non-Medicare population, where the data source depends on the type of service provided (Table 4).

Medicare claims were used to estimate utilization and average reimbursements for the Medicare population. Medicare does not provide full coverage for all services. Moreover, beneficiaries pay deductibles and coinsurance expenses under Part A and Part B, and these expenses are not included in the Medicare claims. In order to capture these costs, the recommendations of CMS’ Office of the Actuary were followed: Part A payments were inflated by 8%, and Part B payments were inflated by 38% (12).

Computing Work Absences

The MarketScan Health and Productivity Management (HPM) database was used to derive the average work loss associated with each condition. The HPM data are collected through employer payroll systems and include detailed information on when employees are out of work, the number of hours missed, and the reasons for the absences. Reasons for absence include sickness, disability, vacation, and other types of leave. The absence data are linked to eligibility files and medical claims to estimate hours of work loss associated with each condition.

Assigning work absences to specific medical treatments required a complex algorithm using all three databases. In general, the dates of an ambulatory visit or an inpatient stay for a specific urologic condition

were matched to the individual's absence data. Absences associated with a hospitalization included any work loss reported between the admission and discharge dates, including days contiguous to those dates. For example, if a person was admitted to the hospital on June 1 with a primary diagnosis of upper tract urolithiasis and was discharged on June 5, any sick time or short-term disability in that period, as well as on contiguous days prior to June 1 or after June 5, was counted. Any work loss on June 7 was not included, however, in the absence of work loss on June 6.

As underscored by this example, appropriately assigning absences to specific treatments is very difficult. Therefore, a set of decision rules and exclusion criteria were established for computing these estimates. The most important criteria were the following:

- To be included, persons had to be fully enrolled in the health plan throughout the year and had to have an inpatient or outpatient medical claim for a specific urologic condition.
- Individuals in the top 0.5% of total absences during the year and persons on long-term disability or COBRA were excluded.
- Work absences were capped at 12 hours if the beginning and end dates of the absence were the same.
- If two outpatient visits occurred in the span of one absence, then hours absent before the first visit counted toward the first visit, and hours absent after the second visit counted toward the second visit. The hours of work lost between the visits counted toward the closest visit. In the event of a tie, the hours were assigned to the first visit. For example, consider an employee absent from work due to sickness from January 1 to January 10, with physician office visits on January 3 and January 7. In this case, work hours missed between January 1 and January 5, were assigned to the first visit and hours missed between January 6 and January 10 were assigned to the second visit.
- Short-term disability hours for persons whose start date coincided with a hospital admission and for whom there was a return-to-work date were included.

- Work absences associated with outpatient visits were calculated in two ways. The first method included absences contiguous to the date of the visit. The second approach excluded contiguous dates. For example, the first method would count a work absence on Wednesday associated with a medical visit on Tuesday. The second approach would not count Wednesday's work loss unless there was also an absence on Tuesday.

Computing Costs at the Individual Level

The Ingenix data were used to estimate the incremental medical costs incurred by persons with urologic conditions. The data link medical and pharmacy claims to health plan benefits for more than 275,000 primary beneficiaries 18 to 64 years of age with employer-provided insurance. Individuals with an inpatient or outpatient claim for specific urologic conditions were identified. Multivariate regression models were used to predict medical and pharmacy spending in 1999 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 payments made by the enrollee (copayments, deductibles, excluded expenses) and by all third-party payers (primary and secondary coverage, net of negotiated discounts) for medical services and outpatient prescription drug claims.

The covariates included a set of variables to describe the medical and drug benefits, including individual plan deductibles, copayments or coinsurance rates, and a binary indicator for plan type (HMO, POS, PPO, FFS). Other covariates included age, sex, work status (active or retired), urban residence, and median household income in the zip code of residence. Observed differences in comorbid conditions were controlled for, based on ICD-9 diagnostic codes from the medical claims files. The medical claims were used to identify individuals treated for any of 26 chronic conditions, including hypertension, diabetes, congestive heart failure, asthma, and depression, and included a binary indicator for each condition.

The statistical analyses used a two-part model. The first part of the model used probit regression to estimate the probability that a member of the study sample had at least one medical or pharmacy claim. The second part of the model used a generalized linear model with a logarithmic link function to estimate the level of spending among members with at least one claim for the outcome of interest.

The two parts of the model were combined to predict average annual spending for persons with and without a urologic condition, controlling for other factors known to affect utilization. Specifically, estimates from the first part of the model were used to predict the probability of nonzero expenditures for persons with and without a specific urologic condition. Similarly, the second part of the model was used to predict expenditures, conditional upon having at least one claim, for each of the two groups. Total expenditures were calculated as the product of the two parts of the model and were averaged over all individuals in the sample, both those with and those without a urologic condition.

LIMITATIONS

We found that for many urologic conditions, population-based datasets contained limited information on true prevalence. Many conditions were not studied in prevalence surveys or were studied in a limited fashion. To buttress our analysis, we turned to published estimates of prevalence and incidence drawn from specific population-based studies focusing on various urologic conditions. For *de novo* analyses, we relied heavily on datasets that use administrative coding systems such as the ICD-9 CM to identify disease burden. Reliance on such administrative codes can result in both underestimation and overestimation of utilization, depending on the sensitivity and specificity of the disease code in question.

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 STDs). 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 greater incidence of a particular condition among divorced persons than among married persons, 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. 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.

Table 5. 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	Healthcare utilization and cost	National sample of inpatient stays and hospitalizations
Surveillance, Epidemiology, and End Results	SEER	Healthcare utilization and cost	Population-based, nationally representative tumor registry data
Surveillance, Epidemiology, and End Results linked with Medicare claims	SEER-Medicare	Healthcare utilization and cost	Claims submitted by inpatient and outpatient providers for Medicare beneficiaries who are tracked in SEER
National Ambulatory Medical Care Survey	NAMCS	Healthcare utilization and cost	National sample of ambulatory care utilization
National Hospital Ambulatory Medical Care Survey–Outpatient and Emergency Room Components	NHAMCS-OP NHAMCS-ER	Healthcare utilization and cost	National sample of ambulatory care services in hospital emergency and outpatient departments
Medical Expenditure Panel Survey	MEPS	Healthcare utilization and cost	National sample of healthcare use, expenditures, and sources of payment
National Association of Children’s Hospitals and Related Institutions	NACHRI	Target populations	Pediatric inpatient stays at member hospitals only
National Nursing Home Survey	NNHS	Target populations	National sample of nursing homes, the providers of care, and their residents
Department of Veterans Affairs-Outpatient Clinic Files	VA-OPC	Target populations	National sample of VA outpatient services
MarketScan Health and Productivity Management Database	MarketScan	Cost of disease	Fortune 500 company inpatient and outpatient medical claims providing productivity and pharmacy data for employees and dependents
Ingenix Database	Ingenix	Cost of disease	Medical claims database providing utilization and cost data for 75 large employers
Society of Assisted Reproductive Technology Database	SART	Target Populations	Data regarding male and female factor infertility procedures from fertility clinics across the US
National Survey of Ambulatory Surgery	NSAS	Healthcare utilization and cost	National sample of ambulatory surgery performed in hospitals and freestanding ambulatory surgery centers
HCUP Kids’ Inpatient Database	KID	Healthcare utilization and cost	Database of hospital inpatient stays for children to examine a broad range of conditions and procedures related to child health issues.
National Health and Nutrition Examination Survey	NHANES	Healthcare utilization and cost	Continuing series of national surveys of households and household members to assess health and nutritional status of adults and children
VA Pharmacy Benefits Management version 3.0 Database	PBM v3.0	Target populations	VA entity responsible for managing the national VA drug formulary process
VA National Surgical Quality Improvement Program	NSQIP	Target populations	Collects pre-operative risk factors, intra-operative variables, and 30-day post-operative mortality and morbidity outcomes on most major surgical operations performed in the VA system

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.

The MEDPAR files contain summary records for all stays. The carrier and outpatient files contain a 5% random sample of the Medicare population. The same 5% sample of stays was used in building the files for this research effort^d. The carrier file contains detailed information at the line-item level, which provided information on payment and place of service by line item^e. 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^f.

An iterative process was used to build the analysis files. First, inpatient stays were identified, using MEDPAR records. Next, ER, outpatient surgery, and ambulatory surgery visits shown in the outpatient file were defined and selected, using appropriate revenue center codes. Finally, the line items and outpatient records that were not facility charges were matched to these visits and inpatient stays, 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) remaining line items with place of service listed 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; and (d) 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 in MEDPAR as those stays in which a UDA diagnosis was the primary diagnosis. This number is 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,^g 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 both as an ER visit in the ER facility of service and also had its associated dollars rolled into the inpatient stay. In other words, it was double-counted.

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.

Inpatient payments were inflated by 8% (12) to account for deductible expenses.

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^h
- Operating room services-minor surgery^h
- Operating room services-other operating room services^h

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

- Emergency room-general classification
- Emergency room-EMTALAⁱ 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.

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^j codes.

Physician services were next 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.

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^k. 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 line item was dropped from further analysis.

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:

- Any facility records more than seven days from an existing date of service were deleted.
- 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 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. Individuals were selected from the 100% MEDPAR and denominator files, using the same criteria. 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 1991 through 2000, except in the denominator file, which contains data for 1984 through 2000. The years of data used for the conditions evaluated in this compendium were 1992, 1995, and 1998.

Sample Size: The MEDPAR dataset contains 100% of the Medicare beneficiaries and 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 at least 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 1988 through 2000.

Sample Size: Initially, the database covered only eight states; it has since grown to 28 states. It contains discharge data from 994 hospitals, approximating a 20% stratified sample of US community hospitals. The 2000 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 and 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.

Ingenix

Sponsor:

Ingenix Health Intelligence
Corporate Headquarters
2525 Lake Park Blvd.
Salt Lake City, UT 84120

Design:

This database contains a subset of claims, utilization, and cost data from 75 large employers.

Time Frame: The available data are for 1997 through 1999.

Sample Size: The dataset includes approximately 1.8 million enrolled employees and their dependents.

Use: The medical claims data provide detailed financial information, as well as dates of service, diagnosis and procedure codes, types of facility, and providers. Drug claims include prescription fill date, patient and plan costs, and, in most cases, national drug codes (NDCs). Claims data contain records for only those who used services.

Benefits: This claims-based dataset captures all healthcare claims and encounters for employees and their dependents and includes detailed information on both medical and prescription drug costs.

Limitations: The longitudinal data are available for only a subset of firms.

MarketScan

Sponsor:

Medstat
777 E. Eisenhower Parkway
Ann Arbor, MI 48108
(734) 913-3000

Design: The MarketScan dataset is a collection of integrated inpatient and outpatient medical claims data and encounters; prescription drug, enrollment, and eligibility information; and productivity data. Claims are collected from employers who record corresponding employee absenteeism data and disability claims. Age, gender, and regional distribution of patients are available.

Time Frame: Only one year of data, 1999, is presently available for analysis.

Sample Size: This is a proprietary dataset of claims data from 100 health plans serving Fortune 500 employers. It includes data on 800,000 covered lives and approximately 340,000 employees.

Use: This dataset enables the evaluation of productivity and pharmacy data and associated medical claims information.

Benefits: MarketScan is a unique source of information on the indirect costs of a variety of urologic illnesses. It contains productivity and pharmacy data as well, and cases may be followed longitudinally.

Limitations: MarketScan data are not nationally representative. The database covers a working population, which is not necessarily similar to other patient populations. Issues related to the “healthy worker effect” might also be present (i.e., a healthier subset of people in the general population are more likely to work).

Medical Expenditure Panel Survey (MEPS)— Household Component

Co-Sponsors:

Agency for Healthcare Research and Quality (AHRQ)
and National Center for Health Statistics (NCHS):

Agency for Healthcare Research and Quality
8757 Georgia Ave.
Silver Spring, MD 20910
(866) 556-4287

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

Design: MEPS is a nationally representative survey of healthcare use, expenditures, sources of payment, and insurance coverage for the US civilian non-institutionalized population. It is designed to yield comprehensive data for estimating the level and distribution of healthcare use and expenditures, monitoring the dynamics of healthcare delivery and insurance systems, and assessing healthcare policy implications. The database continuously collects medical expenditure data at both the person and the household level, using an overlapping panel design. Two calendar years of data are collected from each household in a series of five rounds. These data are then linked with additional information collected from the respondents’ medical providers, employers, and insurance providers. The series of data collection activities is repeated each year on a new sample of households, resulting in overlapping panels of survey data in 195 communities across the nation.

Time Frame: Data have been collected five times a year from 1996 to the present.

Sample Size: 10,000 families, or approximately 24,000 individuals.

Use: This national probability survey provides information on the financing and utilization of medical care in the United States. The household

component collects information on demographics, health conditions, health status, payments, access to care, satisfaction with care, insurance, income, and employment. These data are collected at the person and the household level over two calendar years and are then linked with additional information collected from the respondents' medical providers, employers, and insurance providers.

Benefits: The database contains longitudinal data for the core survey components. The medical provider component supplements and validates self-reported information in the household component.

Limitations: Because it is a household sample, MEPS may include data on only a few urologic illnesses. In addition, conditions may be underreported if one household member responds for others in the household and is unaware of some illnesses.

National Association of Children's Hospitals and Related Institutions (NACHRI)

Sponsor:

National Association of Children's Hospitals and Related Institutions
401 Wythe Street
Alexandria VA 22314
(703) 684-1355

Design: This dataset records information on all pediatric inpatient stays at member hospitals.

Time Frame: Data have been collected annually since 1999.

Sample Size: The dataset contains information on approximately 2 million pediatric inpatient discharges. Fifty hospitals located in 30 states participated in 1999, 55 participated in 2000, and 58 participated in 2001.

Use: Data are collected on the age, race, sex, and ICD-9 principal diagnosis of each pediatric inpatient cared for at participating facilities. Additionally, information is collected on length of stay, total charges, and cost-to-charge ratio.

Benefits: Because it collects data from children's hospitals, the NACHRI dataset provides a unique opportunity to study the inpatient burden placed on the healthcare system by relatively uncommon pediatric urologic conditions. The dataset is rigorously edited and cleaned to ensure data quality.

Limitations: Because NACHRI collects data from specialized facilities, information regarding such topics as length of stay, patient demographics, treatment, and costs may not be representative of the national experience.

Society of Assisted Reproductive Technology (SART) database

Co-Sponsors:

Society of Assisted Reproductive Technology
1209 Montgomery Highway
Birmingham, AL 35216
(205) 978-5000

Centers for Disease Control and Prevention
1600 Clifton Rd.
Atlanta, GA 30333
(404) 639-3311

Design: Prospective ongoing data collection from fertility clinics across the US as mandated by Federal law.

Time Frame: 1999

Sample Size: 347 fertility clinics throughout the US.

Use: This dataset allows description of impact of male infertility on use of fertility services, and can be used to examine the influence of male factor infertility on reproductive outcomes.

Benefits: Utilization for infertility services related to male factor infertility are usually born by the patient; this dataset provides a window into utilization that is not easily captured in claims-based datasets.

Limitations: Access to fertility clinics is not uniform and varies by socioeconomic status; these data are not representative of the national burden of male factor infertility. Low income men with male factor infertility may not have access to these services.

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 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” is comprised of facilities which are state licensed or 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: 1994–1996

Sample Size: The 1996 NSAS abstracted data from 125,000 ambulatory surgery visits to 488 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 over a 3-year period.

Limitations: ICD-9 procedure codes are used, which are of limited specificity when compared with Common Procedural Terminology procedure codes. At the time of these analyses, only data from 1994–1996 were available. The National Center for Health Statistics is planning a NSAS data collection for 2006.

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: 1997, 2000

Sample Size: The 2000 KID contains data drawn from 27 State Inpatient Databases on children 20 years of age and younger. The 1997 KID contains data drawn from 22 State Inpatient Databases on children 18 years of age and younger. KID includes a sample of pediatric discharges from over 2,500 to 3,500 US community hospitals (defined as short-term, non-Federal, general and specialty hospitals, excluding hospital units of other institutions).

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: The survey was conducted annually from 1973 through 1981 and once in 1985; it has been conducted annually since 1989.

Sample Size: The sample size for the years of data evaluated in this compendium ranges from 1,200 to 1,700 physicians and 23,000 to 35,000 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 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 3 was conducted from 1988–1994. NHANES is currently a continuing survey, with the latest data release covering 1999–2000.

Sample Size: The sample for NHANES 3 includes approximately 33,994 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.

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. And fourth, 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: The data have been collected annually since 1992.

Sample Size: The sample size for the years of data evaluated in this compendium is in the range of 22,000 to 30,000 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 Nursing Home Survey (NNHS)

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: NNHS is a continuing series of national sample surveys of nursing homes, their residents and their staff.

Time Frame: These surveys were conducted in 1973–1974, 1977, 1985, 1995, 1997, and 1999. The years of data used for this compendium are 1995 through 1999.

Sample Size: The sample includes approximately 1,500 facilities, where interviews (occasionally via self-administered questionnaires) were conducted with administrators and staff.

Use: The survey provides information from the perspectives of both the providers of service and the recipients. Data collected include information about the size and ownership of the facility, Medicare/Medicaid certification, occupancy rate, number of days of care provided, and expenses. Recipient data collected include demographic characteristics, health status, and services received.

Benefits: The dataset is unique in that information is solicited from both the provider and the recipient of care. It also targets a specific, useful population for study.

Limitations: The surveys do not contain information on the health services provided; they report only whether a patient received service within general categories. The records do not contain a facility number that would allow linkage of records to the facility.

Veterans Affairs Outpatient Clinic Dataset (VA-OPC)

Sponsor:

Austin Automation Center (AAC) Enterprise Business Office
Austin, TX (<http://www.aac.va.gov>)
(512) 326-6005

Design: The Department of Veterans Health Administration maintains a centralized data repository that contains computerized utilization data for all outpatient visits and acute care hospital stays, as well as other utilization datasets on nursing home stays, contract services paid for by the VA, etc. These datasets are integral to the National Patient Care Database (NPCD) in the VA.

Time Frame: The computerized outpatient clinic files (OPC) contain data from 1980 to the present. Ambulatory procedures were added in 1990, and outpatient diagnoses (ICD-9-CM) were added in FY1997. Patient treatment files (PTF) contain data from 1970 to the present.

Sample Size: The VA is the largest healthcare system in the United States, comprising more than 160 hospitals (>45,000 beds), more than 600 community-based outpatient clinics, and more than 100 nursing homes. The VA serves more than 3 million veterans annually. Each dataset within the NPCD contains records for the population of patients seen in all VA healthcare facilities, representing a comprehensive national record of the delivery of VA healthcare services to veterans.

Use: The OPC files include demographics, visits, and clinic stops. The PTF contains demographics and admission and discharge data, as well as diagnoses, DRGs, length of stay, transfers, and hospital-based procedures.

Benefits: The datasets represent the population of veteran users of VA healthcare for whom utilization data were recorded. They provide a rich resource for assessing prevalence of disease among healthcare users, as well as for evaluating patterns of care. Encrypted SSNs permit file linkage across

VA healthcare facilities and across settings within facilities, providing a relatively complete portrait of healthcare utilization in VA sites of care.

Limitations: The VA datasets do not provide comprehensive information about healthcare utilization obtained by veterans *outside* the VA healthcare system. Also,

the diagnosis codes are derived from outpatient visits from physician/patient encounters and thus *do not reflect all existing cases among veteran users*, but instead reflect the population for whom care was sought during the year under review.

VA Pharmacy Benefits Management version 3.0 database (PBM v3.0).

Sponsor:

Pharmacy Benefits Management Strategic Healthcare Group
VA Central Office
810 Vermont Ave. NW
Washington, DC 20420

Design: The PBM national database includes medication utilization information based on unique patients for every prescription filled in the VA. The database produces information on each occasion of drug dispensation, which helps to track total prescription usage on a gross or discrete level

Time Frame: 1998 to 2003

Sample Size: The VA serves more than 3 million veterans annually.

Use: The PBM v. 3.0 can be used to examine time trends and overall utilization for various medications used to treat urologic conditions.

Benefits: This comprehensive database provides a portrait of oral medication utilization for urologic conditions that often effect Medicare aged individuals, for whom such information is generally unavailable.

Limitations: Certain medications which are ordered by clinics or hospital wards (e.g., intravesical agents) may not be captured if no prescription is created prior to dispensation.

VA National Surgical Quality Improvement Program (NSQIP)

Sponsor:

Office of Patient Care Services
VA Central Office
810 Vermont Ave. NW
Washington, DC 20420

Design: The Department of Veterans Health Administration NSQIP collects pre-operative risk factors, intra-operative variables, and 30-day post-operative mortality and morbidity outcomes on most major surgical operations performed in the VA healthcare system. Data for each operative case are abstracted. Each year, risk-adjustment models are created and risk-adjusted 30-day mortality and morbidity surgical outcomes are reported.

Time Frame: 1998 to 2003

Sample Size: Approximately 160 hospitals (>45,000 beds). The VA serves more than 3 million veterans annually.

Use: The NSQIP can be used to examine complication rates and mortality for common urologic procedures in the VA population

Benefits: Uniformity of data collection minimizes selection bias. The database contains a wide variety of pre-operative clinical variables which may influence the outcomes of urologic surgery.

Limitations: For some urologic conditions, outcomes of importance to the patient (e.g., quality of life outcomes) are not collected.

NOTES

^a 2000 NAMCS Micro-data file documentation, Data Dissemination Branch, National Center for Health Statistics, 6525 Belcrest Road, Room 1064, Hyattsville, MD, 20782.

^b CPS Utilities, Unicon Research Corporation, March, 1992–2000.

^c The VA does not generate a claim or patient bill for eligible veteran users, with the exception of certain copayments or through medical care cost recovery of selected charges among coinsured veterans. As a result, the insurance categorization in the VA administrative databases may not be as accurate as those in private or other public sector healthcare organizations or systems for which financing is based entirely on reimbursement of charges. Prevalence estimates for private/HMO insurance may also be underestimates for these veterans, as the VA databases do not capture visits or diagnoses associated with visits to non-VA providers.

^d These files excluded anyone with health maintenance organization (HMO) experience during any years of our analysis.

^e 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.

^f Outpatient claims with facility type listed as skilled nursing facilities (SNF) or home health agencies (HHA) were excluded from analysis.

^g 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.

^h Operative procedures provided at these revenue centers were reviewed by clinical experts and were all considered to be appropriately categorized as ambulatory surgery.

ⁱ 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.

^j The HCFA Common Procedure Coding System.

^k 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.

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. Statistical Analysis System, Release 8.2. SAS Institute; Cary, NC.
4. Statistical Package for the Social Sciences for Windows, Release 9.0, 1998. SPSS Inc; Chicago, IL.
5. Rothenberg R, Bross DC, Vernon TM. Reporting of gonorrhea by private physicians: a behavioral study. *Am J Public Health* 1980;70:983-6.
6. Rothenberg RB. Analysis of routine data describing morbidity from gonorrhea. *Sex Transm Dis* 1979;6:5-9.
7. 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.
8. Chorba TL, Berkelman RL, Safford SK, Gibbs NP, Hull HF. Mandatory reporting of infectious diseases by clinicians. *Jama* 1989;262:3018-26.
9. 1998 guidelines for treatment of sexually transmitted diseases. Centers for Disease Control and Prevention. *MMWR Recomm Rep* 1998;47:1-111.
10. National Drug Data File Plus Tm reference software. First Data Bank Inc; San Bruno, CA.
11. *Drug Topics Redbook*. 107th ed. Montvale, NJ: Medical Economics Co., 2003.
12. 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.

GLOSSARY OF SELECTED TERMS

Race — The concept of race reflects self-identification by people according to the race or races with which they most closely identify. These categories are sociopolitical 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 to 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 indicate 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 indicate their race as “Black, African Am., or Negro,” 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, India, 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” and “Native Hawaiian and Other Pacific Islander” race categories described above. Respondents providing write-in entries such as multiracial, mixed, interracial, Wesort, 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 States are grouped into four regions corresponding to those used by the US Bureau of the Census:

- 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 Columbia, 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

Metropolitan Statistical Area — An MSA consists of a large population nucleus of 50,000 population or greater, together with adjacent communities having a high degrees of social and economic integration with that core. Metropolitan areas comprise at least

one county, except in New England, where cities and towns are the basic geographic unit.

Urban Area — Urban areas consist 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 population.

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 (formerly the Health Care Financing Administration).

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, casualty insurance companies, health insurance companies, and independent plans such as employer/union-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 visit were paid by the patient, spouse, family, or next-of-kin.

Other insurance — Includes any nonprofit 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

0.000-1.850 (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.

Nursing Home: In the National Nursing Home Survey, nursing homes are defined as facilities that routinely provide nursing care services and have three or more beds set up for residents. Facilities may be certified by Medicare or Medicaid or not certified but licensed by the state as a nursing home. The facilities may be freestanding or a distinct unit of a larger facility

Intermediate Care Facility: Institutions certified by the Medicaid program to provide health-related services on a regular basis to Medicaid-eligible individuals who do not require hospital or skilled nursing facility care, but do require institutional care above the level of room and board.

Skilled Nursing Facility: An institution (or a distinct part of an institution) that is primarily engaged in providing skilled nursing care and related services for residents who require medical or nursing care, or

rehabilitation services for the rehabilitation of injured, disabled, or sick persons, and is not primarily for the care and treatment of mental diseases.

Home Health: A collection of supportive care services focused on providing skilled nursing in the home, along with a range of the following services: personal care services; homemaker and companion services; physical therapy; medical social services; medical equipment and supplies; counseling; 24-hour home care; occupation and vocational therapy; dietary and nutritional services; speech therapy; audiology; and pharmacy care, such as intravenous

NIH Publication No. 07-5512