H·CUP

HEALTHCARE COST AND UTILIZATION PROJECT

HCUP Methods Series

Contact Information:
Healthcare Cost and Utilization Project (HCUP)
Agency for Healthcare Research and Quality
540 Gaither Road
Rockville, MD 20850
http://www.hcup-us.ahrq.gov

For Technical Assistance with HCUP Products:

Email: hcup@ahrq.gov

or

Phone: 1-866-290-HCUP

Table of Contents

OVERVIEW

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

- The Nationwide Inpatient Sample (NIS) – the largest all-payer inpatient care database in the United States, containing data on nearly eight million hospital stays per year from approximately 1,000 hospitals.

- The Kids' Inpatient Database (KID) – the only all-payer inpatient care database for children in the United States.

- The Nationwide Emergency Department Sample (NEDS) – the largest all-payer emergency department database publicly available in the United States, containing information from over 27 million records for ED visits in about 1,000 hospitals

- The State Inpatient Databases (SID) – inpatient discharges from a census of hospitals in participating states.

- The State Ambulatory Surgery Databases (SASD) – ambulatory surgery encounters from hospital-affiliated and sometimes freestanding ambulatory surgery sites from selected states.

- The State Emergency Department Databases (SEDD) – hospital-affiliated emergency department visits that do not result in hospitalizations, collected from participating states.

The objective of this report is to identify relevant sources of population data that can be used with the HCUP databases to calculate rates of hospital care events per population. This compilation includes data sources that provide nationwide population counts based on people, and not hospitals, physicians, or local resources. Not included are data sources collected by associations (e.g., the March of Dimes) or data collected by individual states. State surveys can be a valuable resource for information on specific subpopulations.

When possible, collections with a mixture of information (i.e., demographic, health status, resource information) or other helpful resources are also mentioned.

Table 1 lists the array of population data sources referenced in this report.

Table 1. Population Denominator Data Consistent with Information in the HCUP Databases
This table lists possible sources of population denominator data. Please consult the description of the data source for more details. A list of abbreviations appears at the conclusion of this table.

	National	Region	State	County	ZIP Code
DEMOGRAPHIC					
Age and Gender	Census ACS Vendor	Census	Census ACS Vendor	Census ACS ARF	Vendor
Race (modified 1990 definitions)	Vendor		Vendor	ARF	Vendor
Race (2000 Census definitions)	Census ACS		Census ACS	Census ACS	
Median Household Income	Vendor CPS ACS		Vendor CPS ACS	ACS	Vendor
Personal Income, Wages, Salaries	BEA BLS SIPP		BEA BLS	BEA BLS	
Health Insurance Payer/Coverage	CMS CPS (coverage) SAHIE SIPP MEPS		CMS CPS (coverage) SAHIE	SAHIE	
AREA CHARACTERISTICS					
Urban-Rural Location	Census ERS	Census ERS		Census ERS	WWAMI RHRC
Health Professionals Shortage Areas				HRSA	
DISEASE PREVALENCE (FROM SURVEY AND DISEASE REGISTRIES)					
Disease Prevalence	NHIS BRFSS NHANES NSDUH NCS-R SLAITS NAS	SLAITS	BRFSS NSDUH SEER		
OTHER					
Births	NVSS - Natality		NVSS - Natality	NVSS - Natality	
Deaths	NVSS - Mortality		NVSS - Mortality	NVSS - Mortality	
COMBINATION – COMPILATION OF DATA FROM OTHER SOURCES					
Collection of demographic, health status, and resource information	Health System Measurement Project Health US KFF NWHIC US Stat	NWHIC US Stat	AARP Health US KFF NWHIC US Stat	ARF NWHIC	

Abbreviation	Data Source
AARP	AARP State Profiles
ACS	American Community Survey
ARF	Area Resource Files
BEA	Bureau of Economic Analysis
BLS	Bureau of Labor Statistics
BRFSS	Behavioral Risk Factors Surveillance System
Census	Census
CMS	Center for Medicare & Medicaid Services
CPS	Current Population Survey
ERS	Economic Research Service
Health US	Health, United States
HRSA	Health Resources and Services Administration
KFF	Kaiser Family Foundation
MEPS	Medical Expenditure Panel Survey
NAS	National Asthma Survey
NCS-R	National Comorbidity Survey – Replication
NHANES	National Health and Nutrition Examination Survey
NHIS	National Health Interview Survey
NSDUH	National Survey on Drug Use & Health
NVSS	National Vital Statistics System
NWHIC	National Women's Health Information Center
SAHIE	Small Area Health Insurance Estimates
SEER	Surveillance, Epidemiology, and End Result
SIPP	Survey of Income and Program Participation
SLAITS	State and Local Area Integrated Telephone Survey
US Stat	Statistical Abstract of the United States
Vendor	Third-party vendors for demographic and geographic data: Nielsen (formerly Claritas), ESRI Business Information Solutions, General Data Tech (GDT), Tele Atlas
WWAMI RHRC	Washington, Wyoming, Alaska, Montana, and Idaho Rural Health Research Center

POPULATION DATA SOURCES

The descriptions on the following pages provide details on the data sources listed in Table 1 in alphabetic order. For each data source, the following information is provided:

- Sponsor organization

- Data collection method

- Population targeted

- Types of information available

    o Demographic – age, gender, race/ethnicity, income level, payer type, etc.

    o Geographic aggregation – national, regional, state, county, and ZIP Code

- - - o Diseases and health statistics – health status, diabetes, asthma, tobacco use, etc.
- Update cycle for data
- Web references (when available) for the data, published statistics, and online query tools.

NATIONAL POPULATION ESTIMATES

Appendix A provides commonly used estimates from the Bureau of the Census on the resident population for the U.S., for regions (Northeast, Midwest, South, and West), and for the States from 1990 to the most recent year for which data are available. In addition, estimates are provided by gender, five-year age groups, and race-ethnicity at the national level. Total resident population by community income quartile and urban-rural location are based on data from Nielsen, a vendor that compiles and adds value to the U.S. Bureau of Census data. Nielsen uses intra-census methods to estimate household and demographic statistics for geographic areas. Population estimates are available in Excel format on the HCUP User Support Web site under the Method Series (http://www.hcup-us.ahrq.gov/reports/methods.jsp).

AARP State Profiles

Sponsor:

AARP

Data Collection Method:

This is a secondary data source which is a compilation of data from Federal health agencies and private organizations

Population Targeted:

Varies

Types of Available Information:

Demographic Information:
Age distribution, race/ethnic composition, and poverty

Geographic Entity:
State

Diseases and Health:
Infant mortality rate, low birth weight infants, deaths per 100,000 population for various diseases, utilization of health services, health insurance, managed care, health expenditures, health resources

Update Cycle:

Annual, beginning in 1991

Data System Home Page:

http://www.aarp.org/research/

Published Statistics (State Profiles):

http://www.aarp.org/home-garden/livable-communities/info-09-2012/across-the-states-2012-profiles-of-long-term-services-supports-AARP-ppi-ltc.html

Online Query System:

None

American Community Survey (ACS)
Sponsor: United States Census Bureau
Data Collection Method: Survey forms mailed, computer assisted telephone interviewing, and computer assisted personal interviewing
Population Targeted: Resident population in the United States

Types of Available Information:
 	
    Demographic Information:
        Age, gender, race/ethnicity, annual household income, marital status, education, etc.
 	
    Geographic Entity:
        National, state, county
 	
    Diseases and Health:
        None

Update Cycle: Annual, beginning in 1999
Data System Home Page: http://www.census.gov/acs/www/
Published Statistics: http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml

Online Query System:
 	
    American Fact Finder: http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml
 	
    DataFerrett: http://dataferrett.census.gov/index.html

Area Resource File (ARF)
Sponsor: National Center for Health Workforce Analysis (NCHWA), Bureau of Health Professions (BHPr) within the Health Resources and Services Administration (HRSA)
Data Collection Method: Data integrated from more than 50 primary data sources, including the National Center for Health Statistics (mortality and natality records), the American Hospital Association (facilities statistics), and the American Medical Association (physician specialty data)
Population Targeted: Total U.S. population

Types of Available Information:
 	
    Demographic Information:
        Age, gender, race/ethnicity, employment and unemployment, housing statistics, distribution of families and individuals by income groups, and total, per capita, and median income
 	
    Geographic Entity:
        County. This information can be easily aggregated into larger geographic units
 	
    Diseases and Health:
        No specific disease information is available

Update Cycle: Annual, beginning in 1980
Data System Home Page: http://arf.hrsa.gov/
Published Statistics: None
Online Query System: None

Behavioral Risk Factors Surveillance System (BRFSS)
Sponsor: Centers for Disease Control and Prevention (CDC), National Center for Chronic Disease Prevention and Health Promotion
Data Collection Method: Cross-sectional survey using computer-assisted telephone interviewing
Population Targeted: Civilian non-institutionalized population residing in the United States

Types of Available Information:
 	
    Demographic Information:
     Age, gender, race/ethnicity, annual household income, marital status, education, etc.
 	
    Geographic Entity:
     National, state
 	
    Diseases and Health:
     Examples include: health status, immunization, diabetes, tobacco use, HIV/AIDS, arthritis, asthma, and cardiovascular disease

Update Cycle: Annual, beginning in 1984
Data System Home Page: http://www.cdc.gov/brfss/index.htm
Published Statistics: http://www.cdc.gov/brfss/pubs/index.htm

Online Query System:
 	
    Prevalence Data http://apps.nccd.cdc.gov/brfss/index.asp
    BRFSS Maps http://apps.nccd.cdc.gov/gisbrfss/default.aspx
    Trends Data http://apps.nccd.cdc.gov/brfss/
    SMART: Selected Metropolitan/Micropolitan Area Risk Trends
    http://apps.nccd.cdc.gov/brfss-smart/index.asp
 	
    DataFerrett: http://dataferrett.census.gov/index.html

Bureau of Economic Analysis (BEA)
Sponsor: The U.S. Department of Commerce
Data Collection Method: Integration of various economic measures from various sources, including the Census Bureau, the Department of Education, and the Bureau of Labor Statistics (BLS)
Population Targeted: Varies
Types of Available Information: Demographic Information: Personal income, wages, and salaries Geographic Entity: National, state, county Diseases and Health: None
Update Cycle: Quarterly
Data System Home Page: http://www.bea.gov/index.htm
Published Statistics: http://www.bea.gov/scb/index.htm
Online Query System: None

Bureau of Labor Statistics (BLS)
Sponsor: The U.S. Department of Labor
Data Collection Method: Quarterly tax reports submitted to State Employment Security Agencies
Population Targeted: Employed population in the U.S.

Types of Available Information:
 	
    Demographic Information:
        Wages and salaries
 	
    Geographic Entity:
        National, state, county
 	
    Diseases and Health:
        None

Update Cycle: Quarterly
Data System Home Page: http://www.bls.gov/cew/home.htm
Published Statistics: http://www.bls.gov/opub/
Online Query System: None

Census
Sponsor: United States Census Bureau
Data Collection Method: Mailed survey forms
Population Targeted: Resident population in the United States

Types of Available Information:
 	
    Demographic Information:
      Age, gender, race/ethnicity, annual household income, marital status, education, etc.
 	
    Geographic Entity:
      National, state, county
 	
    Diseases and Health:
      None

Update Cycle: Decennial census with annual estimates. Decennial census long form will be replaced in 2010 census by the American Community Survey (ACS)
Data System Home Page: http://www.census.gov/
Published Statistics: http://www.census.gov/prod/www/
Online Query System: American Fact Finder http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml

Centers for Medicare & Medicaid Services (CMS)
Sponsor: Department of Health & Human Services
Data Collection Method: Enrollment data and survey
Population Targeted: Medicare and Medicaid beneficiaries

Types of Available Information:
 	
    Demographic Information:
        Age, gender, birth dates, race, residence (Medicare Utilization & Enrollment, Medicaid Utilization & Enrollment)
        Socioeconomic and demographic characteristics (Medicare Current Beneficiary Survey)
 	
    Geographic Entity:
        National, state
 	
    Diseases and Health:
        Health status and functioning, health care use and expenditures, health insurance coverage (Medicare Current Beneficiary Survey)
        Health spending by service type and state (National Health Expenditure Data)

Update Cycle: Fiscal year

Data System Home Page:
 	
    CMS home page: www.cms.gov
 	
    Research Data Assistance Center (ResDAC), CMS-sponsored Website for data files and documentation: http://www.resdac.umn.edu/Available_CMS_Data.asp

Published Statistics:
 	
    CMS Research, Statistics, Data & Systems: http://www.cms.hhs.gov/home/rsds.asp
 	
    Knowledge-based articles available through ResDAC: http://www.resdac.umn.edu/Tools/tech_pubs.asp

Online Query System: None

Current Population Survey (CPS)
Sponsor: Bureau of Labor Statistics and Bureau of the Census
Data Collection Method: Phone survey, with periodic surveys conducted by an interviewer who visits the sample unit
Population Targeted: Civilian non-institutionalized population in the United States
Types of Available Information: Demographic Information: Age, gender, race/ethnicity, marital status, education, etc. Geographic Entity: National, state Diseases and Health: Annual Demographic Survey (the March CPS supplement) includes income and health insurance coverage. See Appendix B for a discussion on using payer population estimates from the CPS with HCUP data.
Update Cycle: Monthly
Data System Home Page: http://www.census.gov/cps/
Published Statistics: http://www.census.gov/cps/data/
Online Query System: DataFerrett: http://dataferrett.census.gov/index.html CPS Table Creator: http://www.census.gov/cps/data/cpstablecreator.html

Economic Research Service (ERS)
Sponsor: United States Department of Agriculture
Data Collection Method: Urban-rural classifications based on Census data
Population Targeted: Resident population in the United States

Types of Available Information:
 	
Urban-rural classification methods:
   • Rural-Urban Continuum Codes (RUCC) – classifies U.S. counties by urbanization and nearness to a metropolitan area.
   • Urban Influence Codes (UIC) – classifies U.S. counties by size of the largest city and nearness to metropolitan and micropolitan areas.
   • Rural-Urban Commuting Area Codes (RUCA) – classifies U.S. census tracts using measures of urbanization, population density, and daily commuting.
 	
Geographic Entity:
National, state, county

Update Cycle: A 10-year cycle; updates appear a few years after each decennial census
Data System Home Page: ERS Home Page: http://www.ers.usda.gov Urban-rural classifications are accessed through http://www.ers.usda.gov/briefing/Rurality/
Published Statistics: www.ers.usda.gov/Publications/
Online Query System: None

Health, United States
Sponsor: National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), and the Department of Health and Human Services
Data Collection Method: This is a secondary data source which is a compilation of data from Federal health agencies and private organizations
Population Targeted: Varies
Types of Available Information: Demographic Information: Age distribution, race/ethnic composition, and poverty Geographic Entity: National, state Diseases and Health: Health insurance, preventive care, risk factors, limitation of activity, and mortality
Update Cycle: Annual, beginning in 1975
Data System Home Page: http://www.cdc.gov/nchs/hus.htm
Published Statistics: http://www.cdc.gov/nchs/hus.htm
Online Query System: None

Health Resources and Services Administration (HRSA)
Sponsor: The U.S. Department of Health and Human Services
Data Collection Method: Data extracted from various other data sources for demographic, spatial data, and health systems information
Population Targeted: Total U.S. population
Types of Available Information: Demographic Information: Age, race, gender, marital status, urban-rural location, income, poverty status, and the combinations of these demographic characteristics Geographic Entity: County Diseases and Health: Health Professional Shortage Areas (HPSAs) – shortages of primary medical care, dental, or mental health providers
Update Cycle: On-going
Data System Home Page: http://datawarehouse.hrsa.gov/ Health Professional Shortage Areas: http://bhpr.hrsa.gov/shortage/
Published Statistics: http://datawarehouse.hrsa.gov/hpsadetail.aspx#Reports
Online Query System: None

Health System Measurement Project

Sponsor:

Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services

Data Collection Method:

The Health System Measurement Project builds primarily on data collected from existing HHS and other Federal surveys. Data have been provided by the following agencies: Agency for Healthcare Research and Quality (AHRQ), Centers for Disease Control and Prevention (CDC), Centers for Medicare and Medicaid (CMS), Food and Drug Administration (FDA), Health Resources & Services Administration (HRSA), Office of the Assistant Secretary for Planning and Evaluation (ASPE), Office of the National Coordinator (ONC), and Substance Abuse and Mental Health Services Administration (SAMHSA).

Population Targeted:

Total U.S. population

Types of Available Information:

Diseases and Health:
The Project focuses on ten critical dimensions of our health care system covering the availability, quality, and cost of care, the overall health of Americans, and the dynamism of the system. The Project examines the evolution of these aspects of our system over time. It also assesses the status of these dimensions of the system with respect to subgroups of the population, with a particular emphasis on vulnerable populations. The ten topical areas include: access to care, cost and affordability, coverage, health information technology, innovation, population health, prevention, quality, vulnerable populations, and workforce.

Update Cycle:

Annually

Data System Home Page:

https://healthmeasures.aspe.hhs.gov/

Published Statistics:

https://healthmeasures.aspe.hhs.gov/

Online Query System:

None

Kaiser Family Foundation (KFF) State Health Facts
Sponsor: Henry J. Kaiser Family Foundation
Data Collection Method: This is a secondary data source which is a compilation of data from Federal health agencies and private organizations
Population Targeted: Varies

Types of Available Information:
 	
    Demographic Information:
        Age distribution, race/ethnic composition, poverty, income, etc.
 	
    Geographic Entity:
        National, state
 	
    Diseases and Health:
        Health status, insurance coverage, health costs and budgets, utilization, minority health, women's health, HIV/AIDS, etc.

Update Cycle: Varies
Data System Home Page: http://www.statehealthfacts.org/
Published Statistics: http://www.statehealthfacts.org/comparecat.jsp?cat=1&rgn=6&rgn=1
Online Query System: None

Medical Expenditure Panel Survey (MEPS)
Sponsor: Agency for Healthcare Research & Quality, U.S. Department of Health and Human Services
Data Collection Method: Survey with computer assisted personal interviewing
Population Targeted: U.S. civilian non-institutionalized population
Types of Available Information: Demographic Information: Age, race, sex, marital status, education, income, poverty status, employment status, etc. Geographic Entity: National, census region, and Metropolitan Statistical Area (MSA) Diseases and Health: Health status, mental health status, extensive information pertaining to health care utilization and expenditure. The Household Component provides data from individual households and their members. The Insurance Component is a separate survey of employers that provides data on employer-based health insurance.
Update Cycle: Annual
Data System Home Page: http://www.meps.ahrq.gov/
Published Statistics: http://www.meps.ahrq.gov/mepsweb/data_stats/publications.jsp
Online Query System: None

National Asthma Survey (NAS)
Sponsor: National Center for Environmental Health (NCEH) and the Centers for Disease Control and Prevention (CDC)
Data Collection Method: Random-Digit-Dial (RDD) telephone survey as part of the 2003 State and Local Interview Integrated Telephone Survey (SLAITS)
Population Targeted: Civilian non-institutionalized population residing in the United States
Types of Available Information: Demographic Information: Age, sex, race/ethnicity Geographic Entity: National and four states (AL, CA, IL, TX) Diseases and Health: Asthma
Update Cycle: One time, 2003
Data System Home Page: http://www.cdc.gov/nchs/slaits/nas.htm
Published Statistics: None
Online Query System: None

National Comorbidity Survey Replication (NCS-R)
Sponsor: National Institute of Mental Health
Data Collection Method: Survey using face-to-face computer-assisted personal interviews (CAPI)
Population Targeted: Adult and youth population in the United States

Types of Available Information:
 	
    Demographic Information:
        Race/ethnicity, marital status, education, income, etc.
 	
    Geographic Entity: National
 	
    Diseases and Health: Mental disorders

Update Cycle: The baseline NCS-1 was fielded in 1990-92. The NCS-1 respondents were reinterviewed in 2001-02 for the NCS-2. The NCS Replication Survey (NCS-R) was carried out in a new national sample of 10,000 respondents. A survey of 10,000 adolescents (NCS-A) was added.

Data System Home Page:
 	
    NIMH page http://www.nimh.nih.gov/index.shtml
 	
    Substance Abuse and Mental Health Data Archive (SAMHDA)
        http://www.icpsr.umich.edu/icpsrweb/SAMHDA/
 	
    National Comorbidity Survey Program
        http://www.hcp.med.harvard.edu/ncs/

Published Statistics: NCS 1993 forward http://www.hcp.med.harvard.edu/ncs/publications.php
Online Query System: http://www.icpsr.umich.edu/icpsrweb/SAMHDA/sdatools/resources

National Health and Nutrition Examination Survey (NHANES)
Sponsor: National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC)
Data Collection Method: Cross-sectional population-based survey
Population Targeted: Civilian non-institutionalized population residing in the United States
Types of Available Information: Demographic Information: Age, gender, education, race/ethnicity, language, marital status, language, etc. Geographic Entity: National Diseases and Health: Health examinations (e.g., blood pressure, lower extremity disease, obesity, etc.) and laboratory tests (e.g., hepatitis, Human Immunodeficiency Virus, measles, etc.)
Update Cycle: Annual, since 1999, but released in two-year increments (e.g., NHANES 2003-2004)
Data System Home Page: http://www.cdc.gov/nchs/nhanes.htm
Published Statistics: http://www.cdc.gov/nchs/nhanes/nhanes_products.htm
Online Query System: DataFerrett: http://dataferrett.census.gov/index.html

National Health Interview Survey (NHIS)
Sponsor: National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC)
Data Collection Method: Cross-sectional survey consisting of computer-assisted individual interviews
Population Targeted: Civilian non-institutionalized population residing in the United States, all ages
Types of Available Information: Demographic Information: Gender, age, race/ethnicity, income, marital status Geographic Entity: National Diseases and Health: Adult conditions (e.g., hypertension, coronary heart disease, diabetes, cancer, asthma, alcohol use, smoking, AIDS, etc.) and pediatric conditions (e.g., sickle cell anemia, autism, diabetes, stuttering, etc.). Also insurance coverage and health care use (number of physician visits, dental visits, etc.).
Update Cycle: Annual, beginning in 1969
Data System Home Page: http://www.cdc.gov/nchs/nhis.htm
Published Statistics: Vital and Health Statistics Series: http://www.cdc.gov/nchs/products/series.htm#sr10 Advance Data Reports: http://www.cdc.gov/nchs/nhis/nhis_ad.htm NCHS Health e-Stats: http://www.cdc.gov/nchs/products/hestats.htm
Online Query System: DataFerrett: http://dataferrett.census.gov/index.html

National Survey on Drug Use and Health (NSDUH) (formerly called the National Household Survey on Drug Abuse (NHSDA))
Sponsor: Substance Abuse and Mental Health Services Administration (SAMHSA), Office of Applied Studies, U.S. Department of Health and Human Services
Data Collection Method: In-person computer assisted interviewing or computer assisted self-interviewing
Population Targeted: Civilian, non-institutionalized population, aged 12 or older

Types of Available Information:
 	
    Demographic Information:
        Gender, age, race/ethnicity, education, family income
 	
    Geographic Entity:
        National, regional
 	
    Diseases and Health:
        Mental illness and the use of alcohol, tobacco, marijuana, cocaine, prescription-type drugs used nonmedically (pain relievers, tranquilizers, stimulants, and sedatives), etc.

Update Cycle: Annual, beginning in 1990
Data System Home Page: http://www.oas.samhsa.gov/nsduh.htm
Published Statistics: http://www.oas.samhsa.gov/nsduh.htm
Online Query System: http://www.oas.samhsa.gov/samhda.htm

National Vital Statistics System (NVSS)
Sponsor: National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC)
Data Collection Method: Standardized form
Population Targeted: Total U.S. population
Types of Available Information: Demographic Information: Age, gender, race, education, marital status   Geographic Entity: National, state, city (100,000 persons or more), county   Diseases and Health: Births, deaths (including cause), fetal deaths, linked birth/infant death, matched multiple births
Update Cycle: Annual, beginning in 1968
Data System Home Page: http://www.cdc.gov/nchs/nvss.htm
Published Statistics: http://www.cdc.gov/nchs/data_access/Vitalstatsonline.htm
Online Query System: None

National Women's Health Information Center (NWHIC)
Sponsor: Office of Women's Health, U.S. Department of Health and Human Services
Data Collection Method: This is a secondary data source which is a compilation of data from Federal health agencies and private organizations
Population Targeted: Varies

Types of Available Information:
 	
    Demographic Information:
        Age, gender, race/ethnicity
 	
    Geographic Entity:
        National, regional, state, county
 	
    Diseases and Health:
        Variety of infectious and chronic diseases, mental health, reproductive health, maternal health, illness prevention, mortality and indicators of access to care

Update Cycle: Varies
Data System Home Page: http://womenshealth.gov/
Published Statistics: http://womenshealth.gov/publications/
Online Query System: None

Small Area Health Insurance Estimates (SAHIE)
Sponsor: Housing and Household Economic Statistics Division, U.S. Census Bureau
Data Collection Method: Health insurance coverage is estimated using the 3-year average of values from the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS)
Population Targeted: Insured individuals, both adults and children, in the U.S.

Types of Available Information:
 	
    Demographic Information:
        Age groups, insurance status (whether insured or uninsured)
 	
    Geographic Entity:
        National, state, county
 	
    Diseases and Health:
        None

Update Cycle: Only available for 2000
Data System Home Page: http://www.census.gov/hhes/www/sahie/index.html
Published Statistics: http://www.census.gov/did/www/sahie/publications/index.html
Online Query System: http://www.census.gov//did/www/sahie/data/interactive/

State and Local Interview Integrated Telephone Survey (SLAITS)
Sponsor: National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC)
Data Collection Method: Random-Digit-Dial (RDD) telephone survey
Population Targeted: Civilian non-institutionalized population residing in the United States
Types of Available Information: Demographic Information: Age, gender, race/ethnicity, household income Geographic Entity: National, regional Diseases and Health: Varies by year
Update Cycle: Annual, beginning in 1997. The U.S. Census Bureau terminated the collection of data for the Statistical Compendia program effective October 1, 2011.
Data System Home Page: http://www.cdc.gov/nchs/slaits.htm
Published Statistics: http://www.cdc.gov/nchs/slaits/slaits_products.htm
Online Query System: None

Statistical Abstract of the United States (US Stat)
Sponsor: United States Census Bureau
Data Collection Method: This is a secondary data source which is a compilation of data from Census Bureau, other Federal agencies, and private organizations
Population Targeted: Varies
Types of Available Information: Demographic Information: Age, gender, race/ethnicity, marital status, education, income, etc.   Geographic Entity: Nation, region, state   Diseases and Health: Health care resources and utilization, in addition to health conditions, diseases, expenditures, insurance
Update Cycle: Annual, beginning in 1878
Data System Home Page: http://www.census.gov/compendia/statab/
Published Statistics: http://www.census.gov/compendia/statab/
Online Query System: None

Surveillance, Epidemiology, and End Results (SEER)
Sponsor: National Cancer Institute, U.S. National Institutes of Health
Data Collection Method: Population-based cancer registries
Population Targeted: Cancer patients
Types of Available Information: Demographic Information: Race, sex, age Geographic Entity: States Diseases and Health: Cancer incidence and survival
Update Cycle: Annual
Data System Home Page: http://seer.cancer.gov/
Published Statistics: Finding Cancer Statistics: http://seer.cancer.gov/statistics/ Statistical Resources: http://seer.cancer.gov/resources/
Online Query System: http://statecancerprofiles.cancer.gov/

Survey of Income and Program Participation (SIPP)
Sponsor: U.S. Census Bureau
Data Collection Method: Panel survey using personal and phone interviews
Population Targeted: U.S. civilian noninstitutionalized population, age 15 years and older

Types of Available Information:
 	
    Demographic Information:
        Sex, age, marital status, education, extensive information of employment, earning, income sources
 	
    Geographic Entity:
        National
 	
    Diseases and Health:
        Topical modules sometimes include information on health, disability, and physical well-being

Update Cycle: 3 to 4 years, beginning in 1993
Data System Home Page: http://www.census.gov/sipp/
Published Statistics: Publications: http://www.census.gov/sipp/pubs.html
Online Query System: DataFerrett: http://dataferrett.census.gov/index.html

Vendor   Third-party private organizations that provide demographic and/or geographic data
Sponsor: Varies
Data Collection Method: Varies
Population Targeted: Varies
Types of Available Information: Demographic Information: Nielsen, formerly Claritas   Geographic Information: General Data Tech (GDT), Tele Atlas   Diseases and Health: None
Update Cycle: Varies
Data System Home Page: Nielsen: http://www.claritas.com/sitereports/default.jsp   General Data Tech: http://www.gdt.com/   Tele Atlas: http://www.teleatlas.com/Pub/Home

Washington, Wyoming, Alaska, Montana, and Idaho Rural Health Research Center (WWAMI RHRC)

Sponsor:

    Federal Office of Rural Health Policy, Health Resources and Service Administration

Data Collection Method:

    Urban-rural classification based on Census data

Population Targeted:

    Populations in urban and rural areas

Types of Available Information:

    Urban-rural classification methods:
        Rural-Urban Commuting Area Codes (RUCA) – classifies U.S. census tracts using measures of urbanization, population density, and daily commuting

    Geographic Entity:
        ZIP Code

Update Cycle:

    10-year cycle; updates appear a few years after each decennial census

Data System Home Page:

    WWAMI RHRC: http://depts.washington.edu/uwrhrc/

    RUCA Project: http://depts.washington.edu/uwruca/

Published Statistics:

    Study findings: http://depts.washington.edu/uwrhrc/rhrc-studies.php

    RUCA data demographics: http://depts.washington.edu/uwruca/ruca-demographics.php

Online Query System:

    None

APPENDIX A. NATIONAL POPULATION DATA

Appendix A includes a collection of national population tables to facilitate the use of this type of information. *The information is provided in a separate Excel file to facilitate use.*

The following population data from the U.S. Census Bureau are available:

- Table A.1 Annual estimates of the resident population for the U.S., census regions, and states, July 1 1990, to July 1, 2012

- Table A.2 Annual estimates of the resident U.S. population by gender and selected age groups, July 1 2000, to July 1 2012

- Table A.3 Annual estimates of the resident U.S. population by gender and age in years

    - Table A.3.1 Annual Estimates for July 1, 2010 to July 1, 2012

    - Table A.3.2 Annual Estimates for July 1, 2000 to July 1, 2009

    - Table A.3.3 Annual Estimates for July 1, 1990 to July 1, 1999

- Table A.4 Annual estimates of the resident U.S. population by gender, race, and Hispanic origin

    - Table A.4.1 Annual Estimates for July 1, 2010 to July 1, 2012

    - Table A.4.2 Annual Estimates for July 1, 2000 to July 1, 2009

    - Table A.4.3 Annual Estimates for July 1, 1990 to July 1, 1999

National population counts from Nielsen that are consistent with the HCUP data elements are also available:

- Table A.5 Annual estimates of the resident U.S. population by national income quartiles, 2000 to 2012.

    - In the HCUP databases, the data element ZIPINC_QRTL contains the national quartile based on the median household income of the ZIP Code of the patient's residence.

- Table A.6 Annual estimates of the resident U.S. population by a four-category urban-rural designation based on Urban Influence Codes, 2000 to 2012.

    - In the HCUP databases, the data element PL_UR_CAT4 is a four-category urban-rural designation for the patient's county of residence. The categorization is a simplified adaptation of the 2003 version of the Urban Influence Codes (UIC). The 12 categories of the UIC are combined into four broader categories that differentiate between large and small metropolitan, micropolitan, and a non-urban residual.

- Table A.7 Annual estimates of the resident U.S. population by a six-category urban-rural classification scheme developed by the National Center for Health Statistics (NCHS), 2000 to 2012.

    - In the HCUP databases, the data element PL_NCHS2006 is a six-category urban-rural classification scheme for U.S. counties developed by the National Center for Health Statistics (NCHS) especially for use in health care research. The classification emphasizes urban distinctions and is unique in differentiating between central and fringe counties of large metropolitan areas. Smaller metropolitan counties are subdivided by population. Non-metropolitan counties are divided simply into micropolitan and non-core categories.

APPENDIX B. USING PAYER POPULATION ESTIMATES FROM THE CPS WITH HCUP DATA

Rosanna M. Coffey, Ph.D., Katharine Levit, and Marguerite Barrett

During the development of AHRQ Quality Indicator (QI) estimates from HCUP data for the first National Healthcare Quality Report (NHQR), the possibility of using the Current Population Survey (CPS) for population denominators related to QIs stratified by hospital bill payer was considered and abandoned. The difficulty was that the NHQR measures were categorized by the primary expected payer of the hospital bill (a mutually exclusive concept in HCUP), while the CPS captures the health insurance coverage of the population, where one respondent can have multiple types of coverage. Our assessment was that there was not a ready translation from CPS counts by health insurance coverage (including uninsured) to HCUP primary payer categories (including no payment, government subsidy programs, and liability insurance) and that a quick attempt at such a translation was not defensible. When we attempted to create mutually exclusive payer categories from CPS data, the payer categories underestimated some payer populations.

Below we raise some of the differences between HCUP and CPS that we are aware of. We present them here, in case HCUP data users are considering using the CPS as population denominators. Solutions to these issues would require an investment of resources and more exploration.

To understand the following discussion, it is important to be familiar with how HCUP retains information on the expected primary payer. The coding schemes utilized by state-specific data sources are retained as provided in the HCUP data element PAY1_X. During the processing of the data into HCUP uniform files, the state-specific coding in PAY1_X is mapped into a uniform coding scheme in the HCUP data element PAY1. For example, any state-specific values in PAY1_X that refer to either fee-for-service and managed care Medicare patients are mapped to the value one (1) for PAY1. The uniform coding scheme of PAY1 simplifies the analysis of the expected primary payer across states, but often obscures the additional detail available in PAY1_X.

Payer vs. Insurance Conceptual Differences

Problem: Because the concepts of HCUP "payer" and CPS "insurance" are inherently different, there will never be an exact match between the two. CPS represents health insurance coverage (including multiple coverages) plus an estimate of the uninsured. HCUP retains the hospital primary bill payers (including a category for when there is no payment). The devil is in the details.

Potential Solution: To address the issue of conceptual differences in payers, one could impute 'primary' insurance coverage for individuals in CPS who have multiple coverages. This would be done by employing a set of rules as to which insurance takes precedence over other insurance coverages when it comes to paying bills. Such an order of precedence would most likely be Medicare, other government insurance (CHAMPVA, CHAMPUS), private insurance, and Medicaid. After each individual with multiple coverages in CPS is assigned a primary insurer, then the CPS data can be tabulated to create "primary payer" denominators for Medicare, CHAMPVA/CHAMPUS, private insurance, and Medicaid.

Tabulations could also be made of the uninsured, which for CPS would be any person not covered by one of these previously mentioned insurance programs. Similar counts of the uninsured in HCUP include "no charge," "self-pay," and government program payments sometimes identified in the state-specific payer field (PAY1_X). In addition, CPS tabulations could be made of workers, which would be an appropriate denominator for analysis of certain states where the data element PAY1_X indicates workers' compensation as the primary payer. (Ninety-six percent of all workers are covered by workers' compensation, so this is a reasonable denominator for this payer type.)

Note that CPS would not be useful in creating population denominators similar to state-specific payer categories (PAY1_X) for programs that pay for services but are not insurance (e.g., maternal and child health programs, state-county run mental health and substance abuse programs, black lung program, corrections system, etc.) or for homeowners or automobile liability insurance payments. CPS respondents who only have coverage through these programs would be considered uninsured.

Other Payer vs. Uninsured

Problem: HCUP uniform coding of payer (PAY1) identifies the following government subsidy programs (other than Medicaid) in an "other" payer group: CHAMPUS, CHAMPVA, Indian Health Service, child health insurance programs, maternal and child health programs, state-county run mental health and substance abuse programs, black lung program, corrections system, and other general assistance state and county programs. Also included in the "other" payer group are non-government payers such as workers' compensation[1], accident insurance, etc. CPS, which aims to measure health insurance coverage of the population, counts people in government assistance programs (other than Medicaid) as part of the uninsured population if they have no other health insurance coverage. HCUP uniform coding (PAY1) includes a category for "self-pay" and "no charge" (uninsured from the hospital perspective).

Potential Solution: For analyses of specific states, it would be possible to regroup the state-specific payer categories (PAY1_X) involving government subsidy programs (other than Medicare, Medicaid, and CHAMPUS/CHAMPVA) as uninsured for comparability with the CPS. This may or may not be helpful, depending on the purpose of the analysis.

Total vs. Non-Institutionalized Populations

Problem: The CPS universe covers only the non-institutionalized population, whereas HCUP covers hospital stays for the entire population. The most obvious segment of the population missing from the CPS universe but included in HCUP is the aged population residing in nursing homes.

Potential Solution: This mismatch is difficult to solve because the data provided to HCUP does not contain accurate information on which admissions are from or to nursing homes. Because many patients from nursing homes are admitted to the hospital through the emergency department, identifying admissions from and discharges to nursing homes from the HCUP data elements for the patient admission route and discharge disposition underestimates the actual number.

---

[1] Workers' Compensation is state mandated, but is fully funded by employers. It becomes the sole payer of services once liability is established.

Time Dimension Differences

Problem: Another issue to consider is the time dimension of the two measures – a point in time for a HCUP discharge versus recall for a year in the CPS. Even if the two sources could be joined, the estimates will undoubtedly differ by a factor related to time dimension because some people change insurance categories during the year (e.g. uninsured to Medicaid).

Potential Solution: No immediate solution.

Uniform Payer Tangent

Problem: A tangential issue that is worth considering is the inconsistency in the state-specific reporting of payer categories which limits the creation of truly consistent uniform HCUP categories for PAY1. For example, in scrolling through state-specific HCUP documentation for the uniformly coded expected payer PAY1, Hill-Burton charity care is often classified as "no-charge," but sometimes included under "other payer" when the state-specific data source combines Hill-Burton and other government programs into one "other government" category. In addition, state-specific data sources sometimes do not distinguish other government programs from other non-government programs. This distinction would be beneficial for HCUP and its consistency with CPS.

Solution: No immediate solution.

Other Considerations

There are two other considerations for better aligning HCUP to CPS and for using other, more definitive sources for some denominators:

Using HCUP Secondary Payer. There may be other ways to use HCUP payer data, depending on the purpose of the analysis. For example, analysts might use both the primary and secondary payer categories (available for some states) in HCUP to get a better idea of the number and types of discharges that are covered by various insurance payers and programs. All of the combinations of payers in the CPS could be compared with the relevant combinations in HCUP. This would allow for more careful construction of payer denominators for studies of specific populations. Rules (such as who is likely to pay the bill, as noted above) also could still be used to collapse to more manageable groups for analysis. Again the other payers would have to be grouped into no insurance, and people covered in the hospital by workers' compensation would have to be paired with CPS "counts-of-workers" denominators or left out.

More Definitive Sources. Alternatively, beneficiary counts for major payers (Medicare, private insurance, Medicaid, no charge, and self-pay) should be able to be tied to population estimates nationally and at the state level from definitive programmatic and other survey sources. It may be best to use CMS Medicare and CMS or state Medicaid enrollment as the definitive source for those programs and to use CPS for private insurance and uninsured. For other payers, depending on the analytic need, it might be best to obtain counts (person-years) for many of these programs directly from the agencies involved. For example, the National Academy for Social Insurance develops counts of people covered by workers' compensation in recent years (http://www.nasi.org/publications2763/publications_show.htm?doc_id=385937), while the Indian Health Services counts people eligible for IHS services. Counts for other populations might be explored (e.g., the nursing home population).

Some of the issues presented above need to be assessed. The more important issues that would have a major impact on population estimates would need to be dealt with in some effective way for HCUP numbers to be related to population insurance estimates in some reasonable manner.

APPENDIX C. CALCULATING STANDARD ERRORS FOR POPULATION-BASED RATES FROM HCUP DATA

When calculating population-based rates using HCUP data as the numerator and Census population data as the denominator, standard errors for each component must be carefully calculated. For estimates based on the HCUP nationwide databases – Nationwide Inpatient Sample (NIS), Kids' Inpatient Sample, and Nationwide Emergency Department Sample (NEDS) – the standard errors should be calculated as described in the HCUP report entitled *Calculating Nationwide Inpatient Sample (NIS) Variances* (Houchens, et al., 2005). This report will simply be referred to as the NIS Variance Report throughout this appendix. The method for calculating standard errors takes into account the cluster and stratification aspects of the NIS, KID, and NEDS sample design when calculating these statistics using the SAS procedure PROC SURVEYMEANS. For estimates based on the HCUP state databases – State Inpatient Databases (SID), State Emergency Department Databases (SEDD), and State Ambulatory Surgery Databases (SASD) – the same procedure omitting the cluster and stratification features should be used. For population counts based on Census data, there is no sampling error.

Step-By-Step Example

Consider the following example of calculating a population-based rate and standard error using the NIS as the numerator and Census data as the denominator.   The rate for the adult admissions for diabetes in the US per 100,000 U.S. population, age 18 and above is defined as follows:

> Number of weighted NIS discharges age 18 and above with a diagnosis of diabetes * 100,000
> U.S. adult population count from the Census for the same data year

Step1: Define the population of interest in the NIS. Create a 0/1 variable named DIABETES in which the value 1 indicates adult discharges with a diagnosis of diabetes and the value 0 is used for all other discharges.

Step 2: Calculate the numerator and standard deviation of the numerator using the NIS.
The example below for calculating the numerator and standard deviation uses the SAS procedure PROC SURVEYMEANS. It is important to include statements for cluster, strata, and weight because the NIS is a stratified sample. The NIS Variance report provides more detail on the purpose of these SAS statements and also provides example code for other statistical software packages.

```
ods listing close;
proc surveymeans data=NISdataset SUM STD ;
  cluster hospid ;
  strata  nis_stratum;
  var    DIABETES;
  weight  discwt;
  ods output statistics=NIS_COUNTS;
  title2 "NIS weighted counts for 0/1 variable DIABETES";
ods listing;
```

In the PROC SURVEYMEANS output data set, the <u>sum</u> is the numerator of the rate (call this "NIS_TOP" for this example) and <u>std</u> is the standard deviation of numerator (call this

"NISTOP_SD" for this example). The actual configuration of the output data set depends on the version of SAS.

Step 3: Determine the appropriate population denominator from Census data. There is no error associated with the Census population count. For this example, call the population count "ADULT_US_POP"

Step 4: Calculate rate and standard error of the rate. Calculate the population-based rate per 100,000 population and the standard error of the rate as follows:

$$\text{Rate} = \frac{\text{NIS\_TOP}}{\text{Adult\_US\_POP}} * 100,000$$

$$\text{SE of Rate} = \frac{\text{NISTOP\_SD}}{\text{Adult\_US\_POP}} * 100,000$$