Overview of National Center for Health Statistics' Data Collections

Climate and Health Workshop
Jennifer D. Parker, Ph.D.
July 9, 2013
Objectives

• To understand the strengths and weaknesses of different types of national health data sources

• To be able to think about how these data can be used to assess possible effects of climate change on health
Outline

• National Center for Health Statistics

• Data Systems
  – Vital statistics
  – Population health surveys
  – Administrative Claims data

• Linked data files
National Center for Health Statistics

• 1960 - National Office of Vital Statistics and the National Health Survey merged to form National Center for Health Statistics

• 1987 - NCHS became part of the Centers for Disease Control and Prevention

• Official Federal Health Statistics Agency and part of the Federal Statistical System

• Data products
  – Public use micro-data: individual records to enable data users to calculate estimates
  – Integrated Data Tools: reports and online tools with pre-tabulated indicators
  – Restricted use micro-data: individual records with information that could potentially be used to identify an individual
    • Geographic identifiers and dates
    • Genetic information, detailed race and ethnicity
    • Can be accessed in the NCHS Research Data Center (RDC)
National Vital Statistics System (NVSS)  
Death and Birth Data

• State based system
  – More or less comparable data across the U.S.; practice patterns and coding can differ over time and place

• Data are available over decades

• About 2.4 million deaths and 4 million births each year

• Available information
  – Public-use Death: age, race, education, marital status, month and day of week of death, place and cause of death
  – Public use Birth: mother’s age, education, race and ethnicity, marital status, parity, prenatal care, some medical risk factors, gestational age, birth weight
  – Restricted-use: exact dates and county of residence (some states will provide more detailed geography)
National Vital Statistics System (NVSS) 
Death and Birth Data

• Vital statistics data can be used to assess events during specific time windows, at specific locations, aggregated locations, and/or for specific endpoints
  – Events during heat wave in a specific location compared to deaths outside of heat wave or in other locations
  – Events attributed to heat or other weather-related cause, regardless of time or location
  – Outcomes for births exposed by location or timing to excess heat (pollution etc) during some point during pregnancy

• National data released approximately 1.5 years after data year
Asthma deaths, all ages, 1980-2009

Rate per million

Black

White

Population Health Surveys
National Health Interview Survey

- Household interview survey – detailed demographic, health status, health insurance, access to care, economic etc.
  - Health outcomes generally more chronic than acute

- Approximately 35,000 – 40,000 households, 80,000 – 100,000 individuals per year
  - Data years can be combined to increase sample or study trends

- Complex sample design (stratified, clustered, oversampling)
- Collected since 1957

- Interview dates and geocoded residential locations available on restricted use files
  - Locations can be used to examine characteristics of locations, not actual places
Asthma period prevalence, 1980-96, asthma attack prevalence, 1997-2009, and current asthma prevalence 2001-2009, all ages, United States

Source: Akinbami, Moorman, and Liu, 2011, CDC/NCHS National Health Interview Survey
Population Health Surveys
National Health and Nutrition Examination Survey

- National examinations since 1960, since 1999 an ongoing survey
  - Physical measures, blood, urine, DNA
  - Household interview and 3-4 hr exam
- Complex survey design
- About 5,000 total survey respondents each year
- About 15 locations each year, 30 location in a two-year data release; data can be combined across releases
- Exam dates, interview dates and geocoded residential locations available on restricted use files
  - Locations can be used to examine characteristics of locations, not actual places
  - Data collected at each location over a few weeks
Measured and self-reported data

Percent adults with hypertension by data source

Percent adults with BMI $\geq 30$ by data source
National Health Care Surveys

- Provider surveys designed to provide a comprehensive look at medical care provided in the U.S

- Information generally obtained from administrative records (e.g. claims), not interview
  - Selected covariates: age, race, expected source of payment, procedures and diagnoses
  - Data can be combined across years to increase power
  - People without insurance or access to care may be less represented

- Restricted use files can contain dates, locations of hospitals, and zip codes of patients
  - Locations can be used to examine characteristics of locations, not actual places
Asthma ED visits among children 0-17 years

2007-2009 black/white ratio: 3.3
Asthma hospitalizations among children 0-17 years

Rate per 10,000 population

- White
- Black

2007-2009 black/white ratio: 3.1
Data Linkage

- Geographic-level: linked by locations (county, tract, block, latitude and longitude)
  - Air pollution, traffic, temperature, chemical exposures, pollen, frost-dates, etc
  - Area-level income, resources (e.g. health centers)

- Person-level: linked using respondent personal identifiers, such as social security number, date of birth, and name
  - Mortality - National Death Index
  - Medicare and Medicaid enrollment and claims
  - Social Security Administration Retirement and Disability
  - Housing and Urban Development (forthcoming)
Data Linkage

• External data should be as ‘national’ as possible
  • Sample weight adjustment may be needed when not all respondents link to external data

• Linkage increases risk of disclosure so linked data files are restricted

• Reference time may not be same for survey and linked data
  – Administrative data from multiple years can be linked to a single survey year
  – Multiple ways to characterize pollution, average (high, low) temperature and other exposures
    • Administrative (e.g. county, zip code), distance between measurement source and residence
  – Unit of time
    • Annual? Seasonal?
    • Relative to interview? Exam? Admission dates?
    • Residential mobility?
Adjusted\(^1\) odds ratios (95% confidence intervals) for 10 μg/m\(^3\) increase in ambient PM\(_{2.5}\) concentration and asthma outcomes, overall and stratified by race/ethnicity (NHIS 2002-2005)

<table>
<thead>
<tr>
<th>Race/ethnicity</th>
<th>Asthma attack in past year</th>
<th>Still has asthma</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AOR (95% CI)</td>
<td></td>
</tr>
<tr>
<td>All(^2)</td>
<td>0.90 (0.78 - 1.03)</td>
<td>0.97 (0.87 - 1.07)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.99 (0.73 - 1.34)</td>
<td>1.10 (0.85 - 1.43)</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>0.85 (0.72 - 1.01)</td>
<td>0.92 (0.81 - 1.04)</td>
</tr>
<tr>
<td>Non-Hispanic Black</td>
<td>1.76 (1.07 - 2.91)*</td>
<td>1.73 (1.17 - 2.56)**</td>
</tr>
</tbody>
</table>

* \(p < 0.05\), ** \(p < 0.01\)

\(^1\)Adjusted for sex, age group, smoking status, urbanicity, health insurance type, education, income, body mass index and exercise. Models for All include adjustment for race/ethnicity

\(^2\)Included adults who reported other race/ethnicity groups

Summary

• National health statistics come from a variety of sources, each source has strengths and limitations.
  – Estimates can vary by source, even for seemingly similar health conditions
• Good for population estimates, underlying prevalence; less useful for acute events or rare conditions
• Linkage between survey data and geographic/temporal variables are better for studies of location characteristics and changes in some chronic health conditions, rather than studies of specific locations or acute events
Contact:

Jennifer D. Parker
Chief, Special Projects Branch
Office of Analysis and Epidemiology
National Center for Health Statistics

jdparker@cdc.gov
(301) 458-4419