

Sources of Data on Disease Frequency

Seattle Epidemiology and
Biostatistics Summer Session
June, 2004

Numerator Data

Sources of numerator data

- Vital records 4
- Disease reports 9
- Registries 12
- Medical records 14
- Health surveys 19
- Other sources

Vital records

- Death certificates
- Birth certificates
- Fetal death certificates (rarely used)

(View death certificate)

Death certificates

<u>Strengths</u>	<u>Weaknesses</u>
<ul style="list-style-type: none"> • Virtually complete • Mortality can be useful proxy for incidence if case fatality is high 	<ul style="list-style-type: none"> • Reflect frequency of fatal cases only • Accuracy of cause of death (and other data items) may be questionable

(View birth certificate)

Birth certificates

Strengths

- Virtually complete
- Can provide data on both cases and population at risk
- Rich data on pregnancy complications, maternal and child outcomes

Weaknesses

- Reporting of prenatal and perinatal conditions incomplete
- Content varies somewhat by state

Examples of reportable diseases in Washington state

Immediately	Within 1 working day	Within 7 working days
Anthrax Botulism Cholera Diphtheria Measles Plague Polio	Hepatitis A & B Meningococcal disease Salmonellosis Syphilis	Giardiasis Gonorrhea Genital herpes Lyme disease Reye syndrome

Disease reports

Strengths

- Fairly complete for certain rare diseases of obvious public health importance
- Degree of incompleteness may be stable over time

Weaknesses

- Reporting often woefully incomplete
- Miss cases who don't seek care or have poor access
- List of reportable diseases varies by area

Laboratory-reportable diseases in Washington state

Within 2 working days	Within 7 working days
Anthrax Botulism Cholera Diphtheria Hepatitis A Malaria Measles Meningococcus Salmonellosis	Gonorrhea

Examples of well-developed registries

- **Cancer** – National Cancer Institute's *Surveillance, Epidemiology, and End Results (SEER)* program
 - Implemented in 14 areas in U.S.
 - Covers about 28% of U.S. population
- **Birth defects** – CDC's *Metropolitan Atlanta Congenital Defects Program*
 - Covers 5 counties comprising greater Atlanta, GA
 - About 50,000 births annually

Disease registries

Strengths	Weaknesses
<ul style="list-style-type: none"> • Diagnostic data often of high quality • Some (but not all) are population-based 	<ul style="list-style-type: none"> • Expensive to establish and maintain • Exist for only a few diseases

3

Medical records

- Hospital discharge data
- Clinical records
- Laboratory records
- Pharmacy records

Hospital discharge data

Strengths	Weaknesses
<ul style="list-style-type: none"> • In some areas, all hospitals participate (e.g., Washington state's CHARS system) • Diagnostic data usually good 	<ul style="list-style-type: none"> • If only some hospitals, may be no clear denominator • Miss cases not hospitalized • Can be hard to identify readmissions as such

Clinical records

Strengths	Weaknesses
<ul style="list-style-type: none"> • Especially useful if a defined population receives all its care from one source • Diagnostic data usually good 	<ul style="list-style-type: none"> • If only some providers, may be no clear denominator • Miss cases who do not seek care • Costly to gather data • HIPAA

Laboratory records

Strengths	Weaknesses
<ul style="list-style-type: none"> • Useful if a certain lab test is almost always done to establish a diagnosis 	<ul style="list-style-type: none"> • Often no clear denominator • Miss cases who do not seek care • Many private labs exist; not all are cooperative • HIPAA

Pharmacy records

Strengths	Weaknesses
<ul style="list-style-type: none"> • Useful if disease of interest is tightly linked with a certain drug or drug class 	<ul style="list-style-type: none"> • Often no clear denominator • Miss cases who do not get treated with medications • HIPAA

Selected U.S. national health surveys

- National Health Interview Survey
- National Health and Examination Survey
- National Ambulatory Medical Care Survey
- Behavioral Risk Factor Surveillance System

www.cdc.gov/nchs

The screenshot shows the homepage of the National Center for Health Statistics (NCHS). The header includes the CDC logo and navigation links like 'Home', 'Search', and 'Health Topics A-Z'. The main content area features several sections: 'What's New' with a 'Senior Level Position' announcement, 'Featured Event' for the '2004 NCHS Data Users Conference', 'Information Showcase' with links to 'Health, United States' and 'NHIS Early Release of Selected Estimates', and 'Top 10 Links'. A left sidebar contains a menu with categories such as 'About NCHS', 'FASTSTATS A to Z', 'Help', 'Events', 'Surveys and Data Collection Systems', 'Initiatives', and 'Research and Development'.

National Health Interview Survey (NHIS)

- Multi-stage sample of households to represent non-institutionalized civilians
- 106,000 people in 43,000 households
- Oversamples blacks, Hispanics
- Done continuously, with annual changes to questions
- Stable “core” questions, supplemented by “topic of the year” questions

NHIS (cont.)

Strengths	Weaknesses
<ul style="list-style-type: none"> • Captures many cases who do not seek care • Good representativeness for civilian, non-institutionalized U.S. population 	<ul style="list-style-type: none"> • Respondents often have limited medical sophistication • Subject to incomplete recall • Can't get statistically stable estimates for areas smaller than a region

National Health and Nutrition Examination Survey

- Uses mobile vans to do in-person physical exams and lab tests on sample of civilian, non-institutionalized U.S. residents
- About 5,000 people examined/year
- Oversamples blacks, Hispanics, adolescents, elderly
- Done continuously; exam protocol changed periodically
- Detailed nutritional assessment

NHANES (cont.)

Strengths	Weaknesses
<ul style="list-style-type: none"> • Excellent standardization of exam procedures • Can detect subclinical disease • Good representativeness for civilian, non-institutionalized U.S. population 	<ul style="list-style-type: none"> • Can't get statistically stable estimates for areas smaller than a region

Behavioral Risk Factor Surveillance System (BRFSS)

- World's largest continuous telephone survey
- Covers civilian, non-institutionalized adults age 18+ years in all 50 states
- About 1500–6000 respondents *per state*, about 185,000 total
- Focus on health behavior; also asks about a few chronic conditions
- Includes “core” questions for all states, plus optional “modules” used by some states

BRFSS (cont.)

Strengths	Weaknesses
<ul style="list-style-type: none"> • Large sample size supports state-specific estimates • Respondent can combine experience across multiple care providers 	<ul style="list-style-type: none"> • Telephone coverage poor for some subpopulations • Response rate <50% and falling • Depends on validity of self-reports

3

Denominator Data

Common sources of denominator data

- **U.S. Census**
- **Administrative records**
 - HMO enrollment records
 - Employment or labor union records
 - Alumni rosters
 - Etc.
- **Birth certificates** for perinatal epidemiology

www.census.gov

U.S. census data

Strengths

- Fairly complete
- Available at many geographic levels of aggregation

Weaknesses

- Only done every 10 years
- Underenumeration varies among population subgroups
- Content limited to sociodemographic characteristics

Introduction to Epidemiologic Methods — Summer, 2004
Discussion Questions: Sources of Data on Disease Frequency

For each of the following research topics, can you suggest a suitable source of existing data on disease occurrence?

1. In recent years, there has been considerable interest in biomarkers of inflammation, including blood levels of C-reactive protein (CRP), as possible predictors of future coronary heart disease. Although associations between CRP levels and other risk factors have been studied in adults, little is known about these associations in childhood. How might they be investigated relatively quickly and inexpensively?
2. In part because of alarming increases in the prevalence of obesity, Americans are increasingly being encouraged to get up off the couch and engage in more exercise. A potential adverse effect of increased physical activity, however, is sports and recreational injury. What is the incidence of medically treated sports and recreational injuries in the U.S., and which population subgroups appear to be at highest risk?
3. Some women develop breast cancer during their childbearing years but still want to have children. Little is known about whether the hormonal changes that occur during a subsequent pregnancy may affect the risk of breast cancer recurrence or progression. How does survivorship after a breast cancer diagnosis differ between women who subsequently bear children compared with women who do not?
4. The proportion of pregnant mothers whose child is delivered by Cesarean section has begun to rise in recent years in the U.S. One hypothesis is that a larger percentage of pregnant women are gaining too much weight during pregnancy, which is associated with increased frequency of fetal macrosomia—unusually large babies who may not fit through the birth canal and have to be delivered by Cesarean section. How might this hypothesis be tested?
5. As an epidemiologist with a state health department, you are interested in estimating the prevalence of diabetes among adults in your state, in determining how the prevalence compares with other states, and in examining how the frequency of diabetes varies among major demographic groups of the state's population. Where might you find this information?