

The epidemiologic approach: Steps to public health action

Center for Infectious Disease Preparedness
UC Berkeley School of Public Health
URL: <http://www.idready.org>

Updated June 2006

Created using freely available, open source software:
<http://www.openoffice.org>

*Center for Infectious Disease Preparedness
UC Berkeley School of Public Health
www.idready.org* 

1

[INSTRUCTIONS ARE IN CAPITAL LETTERS. Talking points are in normal type.]

This lecture reviews the epidemiologic approach used to guide public health and clinical research, and interventions.

NOTE:

REFERENCES ARE GIVEN BY THE PUBMED IDENTIFICATION NUMBER (PMID). JUST COPY AND PASTE THE PMID INTO <http://www.pubmed.org> AND GET THE MEDLINE CITATION, ETC.

PLEASE SEND US GOOD EXAMPLES OR ARTICLES.

THANKS,

CIDP FACULTY

Learning objectives: Participants will be able to ...

- Define epidemiology;
- Describe at least five components of the epidemiologic approach to public health action;
- Describe at least three goals of public health surveillance;
- List the components of a case definition;
- Describe the five measures of occurrence;
- Describe the measures of association
- Describe at least 3 types of public health actions with examples.

2

*Center for Infectious Disease Preparedness
UC Berkeley School of Public Health
www.idready.org*



These are our learning objectives.

What is epidemiology?

Epidemiology is the study of the *distribution* and *determinants* of health-related states or events in specified populations, and the application of this study to the control of health problems.

3

Center for Infectious Disease Preparedness
UC Berkeley School of Public Health
www.idready.org



Here is one definition of epidemiology.

DON'T SPEND TOO MUCH TIME ON THIS SLIDE. IT'S SELF-EXPLANATORY.

By “study” we mean using quantitative, scientific research methods.

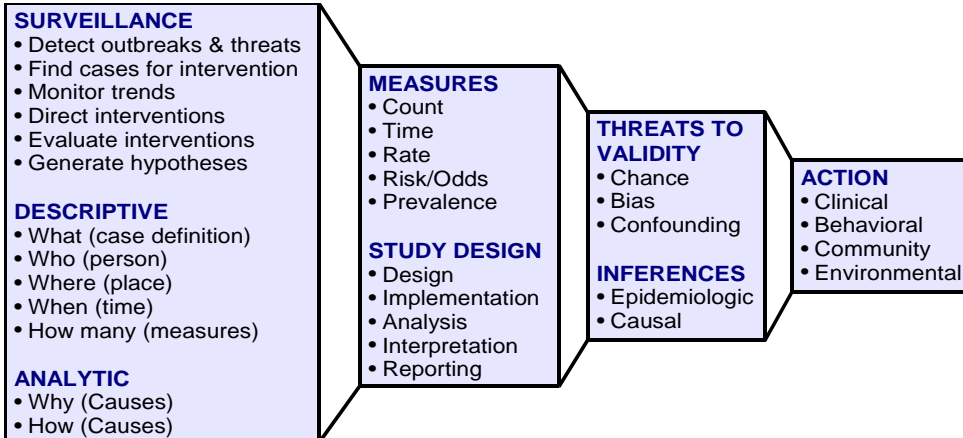
By “distribution” we mean how the amount or burden of risk factors and/or outcomes are distributed in a specified population.

By “determinants” we mean causal factors.

The outcome could be having a condition (“health-related states”) or experiencing a new event (e.g., myocardial infarction).

And, of course, the results should be applied to inform and guide public health actions (“evidence-based public health”).

The epidemiologic approach: Steps to public health action



4



This slide outlines this lecture.

This slide is important because it summarizes the approaches, tools, and methods used in epidemiology. In fact, we have used this slide by itself to give an overview of epidemiology to a lay audience.

What's is public health surveillance?

“Public health surveillance is the ongoing, systematic collection, analysis, interpretation, and dissemination of data about a health-related event for use in public health action to reduce morbidity and mortality and to improve health.”

5 PMID: 15129191

*Center for Infectious Disease Preparedness
UC Berkeley School of Public Health
www.idready.org*



Public health surveillance is the following:

- ongoing, systematic collection, analysis, interpretation, and dissemination of data;
- about a health-related event;
- for use in public health action
 - to reduce morbidity and mortality and
 - to improve health

Goals of surveillance system that guide public health actions

- Detect outbreaks
- Detect public health threats
- Detect infectious cases (case finding)
- Monitor trends in a target population
- Monitor exposed individuals for symptoms
- Monitor treated individuals for complications
- Direct public health interventions
- Evaluate public health interventions
- Generate hypotheses for further evaluation

6

Center for Infectious Disease Preparedness
UC Berkeley School of Public Health
www.idready.org



These are the goals of a public health surveillance system that guide public health actions.

The key words are

- Detect ...
- Monitor ...
- Direct ...
- Evaluate ...
- Generate hypotheses ...

The “detect” category results in a public health action.

The “monitor” category is just that—monitoring a situation.

The remaining categories support evidence-based public health.

Types of disease surveillance

- *Passive surveillance*
 - Title 17 reportable diseases in California
- *Enhanced passive surveillance*
- *Active surveillance*
 - California Emerging Infections Program (CEIP)

7



Passive and active surveillance is defined with reference to the public health authorities that receive (passive) or collect (active) the data. For non-epidemiologists, the most important distinction is knowing the difference between passive and active surveillance.

Passive surveillance is health departments receiving California Morbidity Reports (CMRs) from physicians and laboratories for legally reportable conditions. Passive surveillance occurs in the absence of any health department outreach, education, or request.

Passive surveillance can be “enhanced” through outreach and education to encourage reporting.

Active surveillance involves health department staff actively collecting the data, usually by field deployment to clinics, hospitals, and clinical laboratories.

Sources of surveillance data

- Mortality data
 - Death registry, medical examiner
- Morbidity data
 - Legally reportable diseases, including cancer
- Birth registry
- Hospital discharge diagnoses (utilization data)
- Special surveys (NHANES, NHIS, CHIS)

8

Center for Infectious Disease Preparedness
UC Berkeley School of Public Health
www.idready.org



There are only a few data sources that are population-based, ongoing, and systematic. These include birth and death registries, legally reportable diseases, and hospital discharge diagnoses.

OPTIONAL: What the limitations of hospital discharge data?

Therefore, because public health surveillance data is very limited, special studies are conducted periodically to fill in the gaps.

These include the National Health and Nutrition Examination Survey (NHANES), National Health Interview Survey (NHIS), Behavioral Risk Factor Surveillance System (BRFSS) and, in California, California Health Interview Survey (CHIS), to name just a few.

www.cdc.gov/nchs/nhanes.htm

www.cdc.gov/nchs/nhis.htm

<http://www.cdc.gov/BRFSS/>

<http://ww.chis.ucla.edu/>

Sources of surveillance data (continued)

- Outbreak reporting
- “Sentinel” systems
 - Influenza-like illness reporting/testing (Kaiser)
 - CDC Acute Hepatitis Sentinel County Study
 - West Nile Virus Surveillance Program
 - Encephalitis case surveillance
 - Larval and adult mosquito testing
 - Sentinel chicken testing
 - Dead bird surveillance

9

Center for Infectious Disease Preparedness
UC Berkeley School of Public Health
www.idready.org



Any outbreak is legally reportable.

Sentinel surveillance systems are not usually population-based, nor representative of a target population, but they are “good enough for government work.” For example, during “flu” season, Kaiser Permanente outpatient clinics provide to the California Dept of Health Service Public Health Labs viral throat culture specimens from patients presenting with an influenza-like illness. These cultures are used to type and subtype influenza strains circulating in the community. While the California Kaiser HMO population is clearly not representative of all Californians, it's “good enough.”

<http://www.dhs.ca.gov/dcdc/VRDL/html/FLU/Fluintro.htm>

The California West Nile virus surveillance system incorporates disease and virus surveillance in humans, animals (sentinel chickens and dead birds), and vectors (mosquitos).

<http://www.westnile.ca.gov/>

The case definition

- Inclusion criteria
 - Clinical criteria (symptoms, signs)
 - Epidemiologic criteria (person, place, and time)
 - Laboratory criteria
- Case classification
 - Suspect
 - Probable
 - Confirmed
- Exclusion criteria (for suspect and probable)
- Consider operating characteristics (sensitivity, specificity, predictive value)

10

Center for Infectious Disease Preparedness
UC Berkeley School of Public Health
www.idready.org



Setting up a surveillance requires creating a “case definition.”

A case definition is based on applying selection criteria: inclusion and exclusion criteria. Subjects that meet the inclusion criteria are considered a “case.” In general, confirmed cases are either laboratory confirmed or, lacking laboratory confirmation, have convincing inclusion criteria and are linked to a confirmed case. Probable cases have convincing inclusion criteria, but lack laboratory confirmation or a link to a confirmed case. Suspect cases have weak inclusion criteria, no laboratory confirmation, but may have epidemiologic risk factors (e.g., recently returned from a country with active cases). Suspect cases may also be a mild clinical case or early in symptom onset.

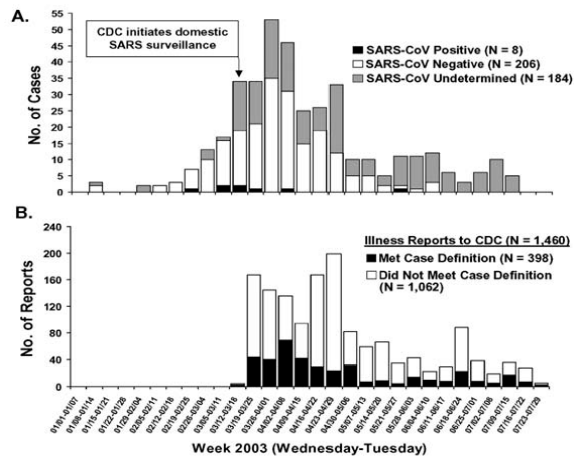
Sometimes exclusion criteria are used to “rule out” suspect or probable cases. For example, see the monkeypox case definition: <http://www.cdc.gov/ncidod/monkeypox/casedefinition.htm>

Finally, it's important to consider the sensitivity, specificity, and predictive value of case definition. A “loose” case definition is more sensitive (lower false negatives) and will pick up mild cases (but also more false positives). In contrast, a “tight” case definition is more specific (lower false positives), but will miss mild cases.

CDC Examples: <http://www.cdc.gov/mmwr/PDF/rr/rr4610.pdf>

U.S. severe acute respiratory syndrome (SARS) cases reported to CDC, January–July 2003.

Displayed in Figure B is the number of unexplained respiratory illness reports received by CDC by week of illness report (N = 1,460). However, only 398 met the U.S. SARS case definition and is displayed by week of illness onset in Figure A.



1 PMID: 15030681

Center for Infectious Disease Preparedness
UC Berkeley School of Public Health
www.idready.org

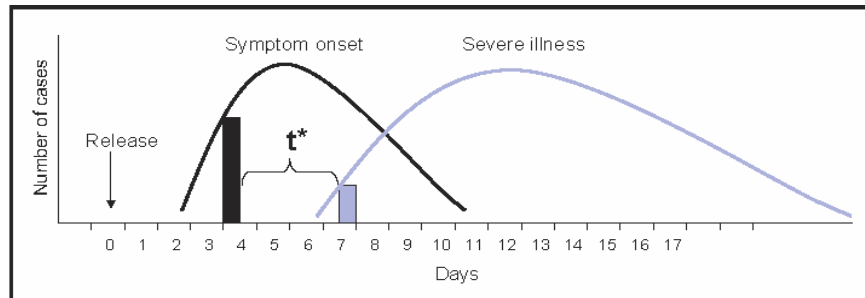
Here is a good example of public health surveillance in action during the SARS outbreaks. This is a retrospective analysis of possible SARS cases reported to the CDC.

The bottom panel (B) has all the possible cases that were reported (1,460). This epi curve is by date of report.

The top panel (A) has the possible cases that met the case definition (398 suspect or probable cases). This epi curve is by date of illness onset which is preferred. Suspect cases met the epi criteria (within 10 days of travel to SARS-affect country or contact to possible SARS case) plus mild/moderate respiratory symptoms, and probable cases met the epi criteria plus evidence of severe respiratory illness (pneumonia or hypoxemia).

Of the 398 patients that met the case definition, only 8 were confirmed (black bars in panel A) to have a positive diagnostic test for the SARS coronavirus. This meant that there were up to 390 false positives, suggesting that the CDC used a “sensitive” case definition. We will study this example in more detail at a later lecture and evaluate the implication of using different case definitions.

Syndromic surveillance — rationale for early detection



* t = time between detection by syndromic (prediagnostic) surveillance and detection by traditional (diagnosis-based) surveillance.

Centers for Disease Control and Prevention. Syndromic Surveillance: Reports from a National Conference, 2003. MMWR 2004;53(Suppl).

1P MID: 15714620

Center for Infectious Disease Preparedness
UC Berkeley School of Public Health
www.idready.org

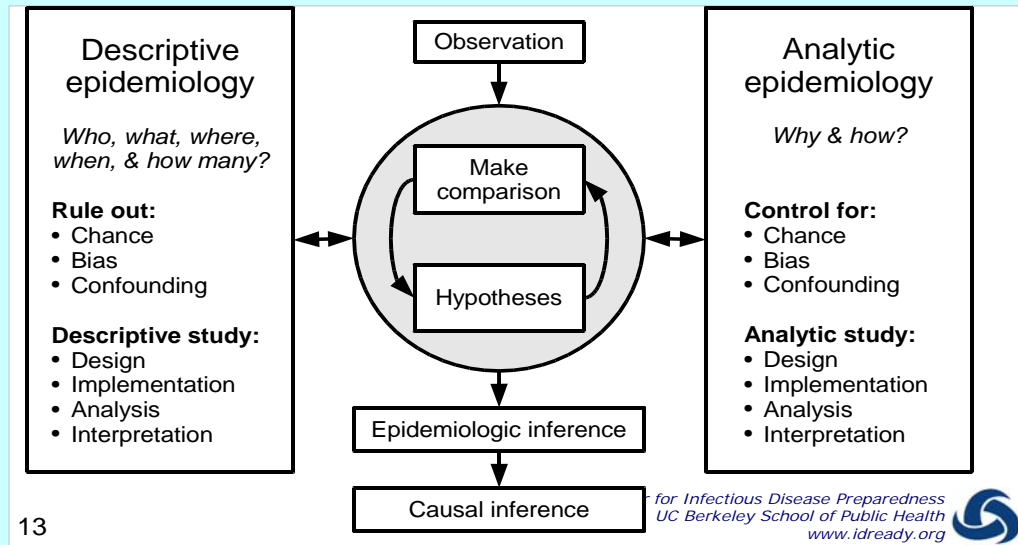
Traditional public health surveillance relies on diagnoses, usually laboratory confirmed. However, with an large intentional release of a bioterror agent, such as *Bacillus anthracis* spores, traditional surveillance systems will only detect advanced disease, meaning that most of those exposed are destined to die from inhalational anthrax, even with antibiotic therapy.

To reduce morbidity and mortality, we would need to detect exposed persons with early, pre-diagnostic symptoms to initiate life-saving therapy. Unfortunately, early symptoms are nonspecific and might only be detectable by evaluating non-traditional systems (e.g., pharmacy sales). These types of systems have been called “syndromic surveillance.”

“The fundamental objective of syndromic surveillance is to identify illness clusters early, before diagnoses are confirmed and reported to public health agencies, and to mobilize a rapid response, thereby reducing morbidity and mortality.”

“...syndromic surveillance might help determine the size, spread, and tempo of an outbreak after it is detected, or provide reassurance that a large-scale outbreak is not occurring, particularly in times of enhanced surveillance (e.g., during a high-profile event).”

Inferences in epidemiology



13

In general, epidemiologic studies are either descriptive or analytic. Descriptive studies measure the burden and distribution of health-related states; i.e., describing the what (case definition), who (person), where (place), when (time), and how many (measure of occurrence). Study findings inform and guide public health services.

When we are interested in determining the underlying causes (the why and how) of health-related event states, we design and conduct analytic studies (analytic epidemiology). Analytic studies involve generating and testing causal hypotheses. These hypotheses are often generated from descriptive studies.

We can think of descriptive and analytic epidemiology as two sides of the same coin. In both types, we are always interested in minimizing threats to making valid conclusions (inferences).

We'll come back to threats to making valid inferences, for now, let's look at the process of descriptive epidemiology more carefully.

NEXT SLIDE

Descriptive epidemiology

- Study of the *distribution* of health-related states or events
 - For a specific outcome, make comparisons and note differences across one or more dimensions (e.g. time series)
 - Seek known explanations to account for observed differences (rule out chance, bias, confounding as explanations)
 - Draw conclusions from descriptive study (epidemiologic inference #1)

14

Center for Infectious Disease Preparedness
UC Berkeley School of Public Health
www.idready.org



It is human nature that when we observe occurrence of health-related states, we make comparisons across some dimension; for example, by neighborhood, by sex, by environmental exposures, etc. We might notice that they are not randomly distributed, and we imagine reasons to explain our observations (hypothesis generation). A layperson may prematurely conclude that their “causal” explanation is correct; in contrast, an epidemiologist first seeks alternative explanations (chance, bias, confounding) for the observations, and, if none are found, may then entertain new causal hypotheses that may need to be tested in an analytic study.

Therefore, epidemiologists and non-epidemiologists share similar thinking processes when observing/describing data. However, an epidemiologist will first rule out alternative explanations (chance, bias, and confounding) before considering new causal hypotheses. A layperson is more likely to infer causality without considering threats to validity and without hypothesis-testing studies.

Sometimes, these different approaches is a source of tension and misunderstanding between public health epidemiologists and community advocates who disagree on what conclusions can be drawn from descriptive studies.

Descriptive epidemiology in outbreak investigation

- Describe cases
 - What (case definition) - - - - -> Outcome (Case status)
 - How many? (measures of occurrence)
 - When? (time)
 - Where? (place)
 - Who? (person)
 - Rule out chance, bias, confounding - - - - -> Outcome (Case status)
 - Generate causal hypotheses - - - - -> Exposures (Causes)
 - Known
 - Suspected
-
- ```
graph TD; A[Outcome Case status] --- B[Exposures Causes Known Suspected]; A -.-> C[What case definition]; B -.-> D[Generate causal hypotheses];
```

15

In outbreak investigations, the process of descriptive epidemiology occurs when we describe identified cases. Immediately, we start to consider alternative explanations (chance, bias, and confounding). If the outbreak is real, we generate causal hypotheses base on our observations and on current fund of knowledge.

In the ideal scenario, we have sufficient information to implement control measures. Sometimes we don't have enough information to implicate a cause, or control measures are failing. In this case, we might conduct an analytic study to test causal hypotheses.

We'll cover this approach in detail in another lecture.

## Analytic epidemiology

- Study of the *determinants* of health-related states or events
  - Generate hypotheses from descriptive studies
  - Design and conduct studies to test hypotheses (control for chance, bias, confounding)
  - Draw conclusions from analytic study (epidemiologic inference #2)

16

Center for Infectious Disease Preparedness  
UC Berkeley School of Public Health  
[www.idready.org](http://www.idready.org)



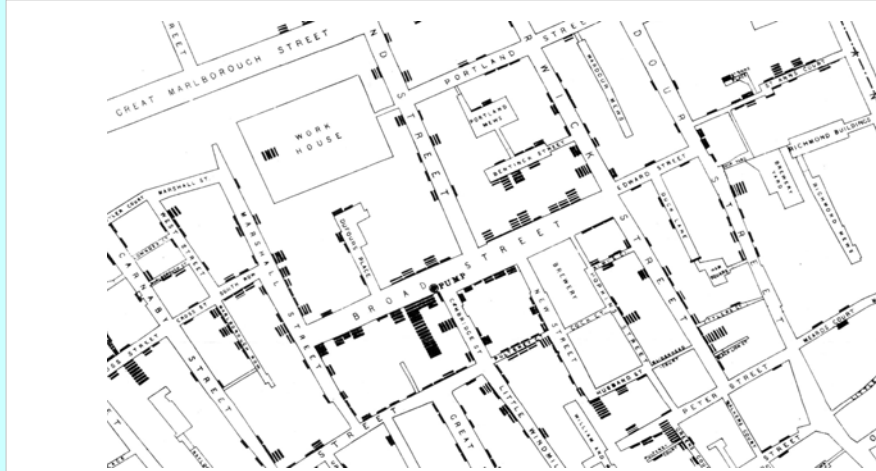
Analytic epidemiology is the complementary process of descriptive epidemiology. Whereas in descriptive epidemiology we measure and describe differences across strata, in analytic epidemiology we measure and test differences across strata (exposed vs non-exposed, for example) in order to test a causal hypothesis.

In descriptive epidemiology, epidemiologists rule out chance, bias, and confounding as explanations of observed differences. In analytic epidemiology, epidemiologist control for chance, bias, and confounding in study design and data analysis (more on these later).

The conclusions we draw from descriptive and analytic studies is called “epidemiologic inference.” Chance, bias, and confounding are always threats to making valid inferences in both types of studies.

Descriptive epidemiology is backbone of studies (including public health surveillance) conducted at health departments. Analytic studies are more commonly conducted at academic institutions. However, both require require scientific rigor.

## John Snow's cholera map, 1854



17

Center for Infectious Disease Preparedness  
UC Berkeley School of Public Health  
[www.idready.org](http://www.idready.org)



Here's a classic example of “analytic epidemiology.”

John Snow hypothesized that cholera was waterborne (presumably from the process of descriptive epidemiology). Two companies provided water to households. The Lambeth Company recently switched to a cleaner river source. He recognized this as a natural experiment to test his causal hypothesis. He predicted in the next cholera outbreak more cholera deaths would occur in households that received water from the Southwark and Vauxhall Company that used river water contaminated with human sewage. In contrast, households that received water from the Lambeth Company would have lower cholera death rates.

The cholera outbreak occurred in 1854. In the 40,046 households supplied by the S & V Co, there were 1,263 deaths (315 deaths per 10,000 households); and in the 26,107 households supplied by the Lambeth Co, there were 98 deaths (37 deaths per 10,000 households).

He made the epidemiologic inference that the cholera death rates differed. He also made the causal inference that cholera was waterborne and recommended the removal of the Broad St. pump.

[http://www.ph.ucla.edu/epi/snow/socbulletin34\(2\)3\\_7\\_2001.pdf](http://www.ph.ucla.edu/epi/snow/socbulletin34(2)3_7_2001.pdf)

## Describing disease occurrence

- Measures of occurrence
  - Count, time, rate, risk, prevalence
- Describing disease occurrence by
  - Person (Age, ethnicity, sex/gender, risk factors)
  - Place (Country, state, county, city, census tract)
  - Time (Calendar time, seasonality)



In order to conduct descriptive epidemiology, we need to measure the occurrence of disease cases.

We now turn our focus to understanding epidemiologic measures of occurrence.

## Epidemiologic measures: Overview

- Types of measures
- How we combine numbers
  - Measures of occurrence
  - Measures of association
  - Measures of attribution (not today)

19

Center for Infectious Disease Preparedness  
UC Berkeley School of Public Health  
[www.idready.org](http://www.idready.org)



THIS IS AN OVERVIEW SLIDE

We will discuss types of epidemiologic measures.

We will discuss how we combine numbers into new measures.

## Types of measures

- Quantitative
  - Continuous numbers
    - Time, height, weight, etc.
  - Discrete numbers
    - Integers (... , -4, -3, -2, -1, 0, 1, 2, 3, 4, ...)
    - Counting (natural) numbers (0, 1, 2, 3, 4, ...)
- Qualitative
  - Nominal categorical (non-ordered)
    - Male vs. female; ethnicity (white, Latino, Asian, etc.)
    - Outcomes are usually binary (event vs. no event)
  - Ordinal categorical (ordered)
    - Low, medium, high



THIS SLIDE DESCRIBES THE TYPES OF MEASURES

## Epidemiologic Measures: How we combine numbers

|            |                                               |                                        |
|------------|-----------------------------------------------|----------------------------------------|
| Rate       | $r = \frac{\Delta x}{\Delta t}$               | Change in x per change in t            |
| Proportion | $p = \frac{a}{a+b}$                           | Numerator is part of the denominator   |
|            | $P = \left( \frac{a}{a+b} \right) \times 100$ | Percent                                |
| Ratio      | $R = \frac{x}{y}$                             | x and y are different and are compared |

21

Center for Infectious Disease Preparedness  
UC Berkeley School of Public Health  
[www.idready.org](http://www.idready.org)



In epidemiology we often combine numbers to construct a rate, proportion, or ratio.

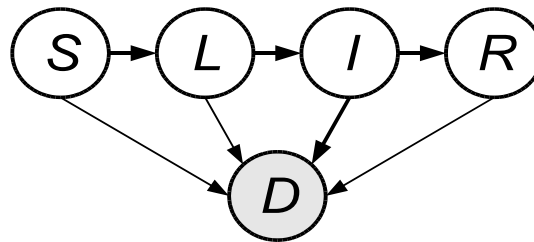
A rate is the change in one measure (usually counts) per change in another measure (usually person-time).

In a proportion, the numerator (usually a count) is part of the denominator. We use proportions to estimate risk (“2 out of 10 exposed subject got ill”) or prevalence (“8 out of 10 injection drug users are anti-hepatitis C antibody positive”).

We use ratios to compare two measures. For example, a ratio of rates is called a rate ratio, and a ratio of risks is called a risk ratio.

## Epidemiologic Measures: Measures of occurrence

- Count
- Time
- Rate
- Risk (or Odds)
- Prevalence



*S* = Susceptible  
*L* = Latent infection  
*I* = Infective  
*R* = Recovered  
*D* = Dead

Center for Infectious Disease Preparedness  
UC Berkeley School of Public Health  
[www.idready.org](http://www.idready.org)



22

There are **ONLY** five types of epidemiologic measures of occurrence. That's it! No more, no less.

The multi-state model depicts a susceptible person becoming infected but not infectious (latent state), becoming infectious (infective state), then recovering (recovery state). At any time the subject can die.

We can count the number of events (new infections) during some time interval, or we can count the number of subjects in a health state (e.g., susceptibles) at a point in time.

We can measure the waiting time until a subject becomes newly infected. This is also called time-to-event.

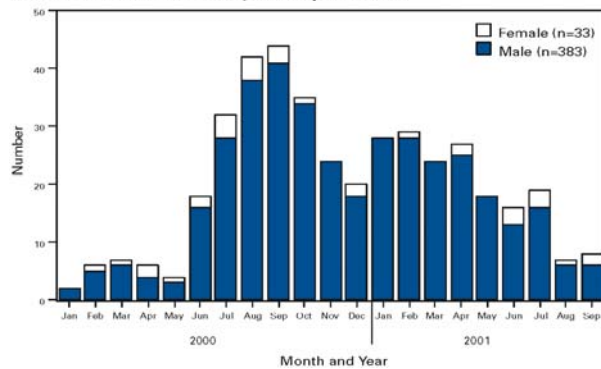
We can calculate the average (per capita) rate of new infections as the count of new infections divided by person-time at risk.

We can calculate the risk for new infection (for a specific time interval) as the count of new infections divided by the number of subjects infection-free at the start of the time interval. Odds is just the numerical transformation of risk: (more on this later).

Prevalence is the proportion that are in a specific state (e.g., susceptible) at a point in time. For epidemics, prevalence measures can change significantly over time.

# Community outbreak of shigellosis, San Francisco, 2000

FIGURE 1. Number of adult *Shigella sonnei* infections, by month, year, and sex — San Francisco, California, January 2000–September 2001



During June–December 2000, 230 cases of culture-confirmed\* *S. sonnei* infection were reported to the San Francisco Department of Public Health; an average of 21 cases (range: 13–29 cases) occurred during the same period from 1996 to 1999. [MMWR 2004;50(42):922]

Center for Infectious Disease Preparedness  
UC Berkeley School of Public Health  
[www.idready.org](http://www.idready.org)



23 \*Residents of San Francisco County aged  $\geq 15$  years.

This “epidemic curve” depicts a community outbreak of shigellosis in San Francisco. Epidemic curves most commonly use counts of newly identified cases (also call “incident cases” or “incidence”).

We avoid the term “incidence” because there are other terms that use the word “incidence” but have very different meanings:

- incidence (count of new cases)
- incidence rate (a rate)
- incidence density (a rate)
- cumulative incidence (a risk)

Unfortunately, epidemiology terminology can sloppy. So our best recommendation is to use precise terms and to read methods sections carefully to understand precisely the actual measures used in papers.

## Epidemiologic Measures: Measures of occurrence - Rate

$$\text{rate} = \frac{\text{Number of events}}{\text{Person-time at risk}}$$

$$\text{Period rate}_{1989-1991} = \frac{D_{1989} + D_{1990} + D_{1991}}{K_{1989} + K_{1990} + K_{1991}}$$

$D_x$  = Events in year  $x$

$K_x$  = Midyear population estimate in year  $x$

24

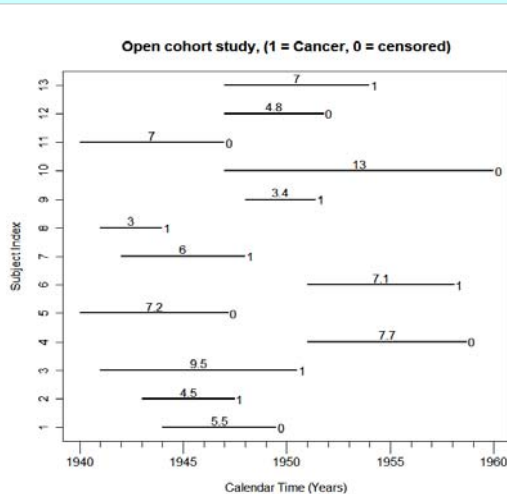
Center for Infectious Disease Preparedness  
UC Berkeley School of Public Health  
[www.idready.org](http://www.idready.org)



The average or per-capita rate is calculated by dividing the number of events divided by the person-time at risk. Rates can be calculated for an open (dynamic) cohort or a closed (fixed) cohort. Average rates are most useful when the actual rate is approximately constant during the observation period, or the observation period is sufficiently short so that the rate is approximately constant during that time interval.

For geographic regions, where it is not practical to measure individual person times, person-time at risk is approximated using population estimates. For example, if a population size is approximately constant during a 2-year period, then the person-years at risk is approximated by the mid-interval population size times 2 years. If population estimates are available every year, use all the data. Rates calculated using this method are called “period rates” because we are using data that is defined by calendar periods instead of individual observation times.

## Measures of occurrence – Cancer rates in an open cohort study, Person-time data



$$\begin{aligned}
 \text{rate} &= \frac{\# \text{cases}}{\sum \text{person-time}_i} \\
 &= \frac{7 \text{ cases}}{85.7 \text{ person-years}} \\
 &= 0.08168028 \\
 &= 8.2 \text{ cases per } 100 \text{ py}
 \end{aligned}$$



Here is an example of calculating an average rate from 13 subjects. This is an open (dynamic) cohort because subjects are moving in and out of the study. Subjects that leave the study before developing cancer are “censored” at the last time their status was known.

For the numerator we count the number of cancer cases.

For the denominator we sum the individual person-times at risk.

## Measures of occurrence - Period rates: Female breast cancer deaths, San Francisco

### Annual period rates

| CATEGORIES                | YEAR    |         |         |
|---------------------------|---------|---------|---------|
|                           | 1989    | 1990    | 1991    |
| Breast cancer deaths      | 125     | 130     | 131     |
| Female population         | 361,975 | 361,401 | 366,613 |
| Rate per 100,000 per year | 34.5    | 36.0    | 35.7    |

### Period rate

$$r_{1989-1991} = \frac{125 + 130 + 131}{361,975 + 361,401 + 366,613}$$

$$= 35.4 \text{ per } 100,000 \text{ per year}$$

Center for Infectious Disease Preparedness  
UC Berkeley School of Public Health  
[www.idready.org](http://www.idready.org)

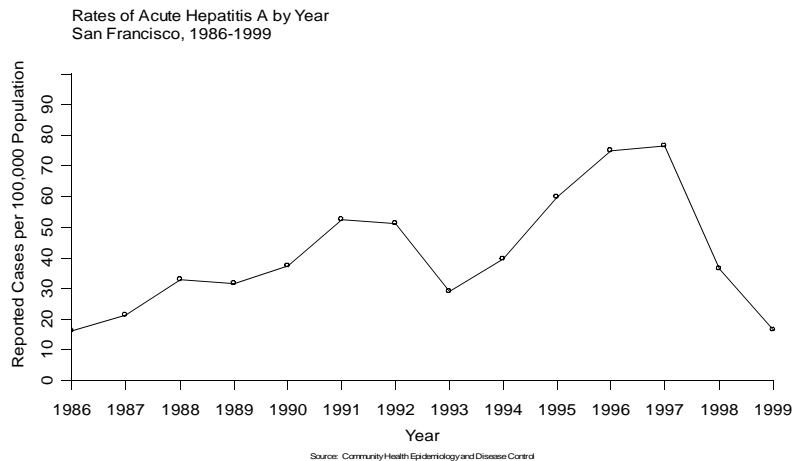


26

Here is an example using period data to calculate either (1) annual period rates for 3 years; or (2) a period rate combining 3 years of data.

The midyear population estimates are used to estimate the person-years at risk per year.

## Rates of acute hepatitis A, San Francisco, 1986-1999



Center for Infectious Disease Preparedness  
UC Berkeley School of Public Health  
[www.idready.org](http://www.idready.org)



27

Here is a graphical display of annual period rates (time series) for hepatitis A cases reported in San Francisco.

## Epidemiologic Measures: Measures of occurrence - Risk

- Risk is a probability
- Probability models are used to estimate risk
- Here is useful and common approach\*:
  - Consider N subjects as disease free at time 0
  - From time 0 to t, x subjects develop disease
  - Proportion  $x/N$  estimates the risk of disease

\*Binomial probability model

28

Center for Infectious Disease Preparedness  
UC Berkeley School of Public Health  
[www.idready.org](http://www.idready.org)



In epidemiology we commonly estimate risks.

Risks are probabilities.

To estimate probabilities we use probability models.

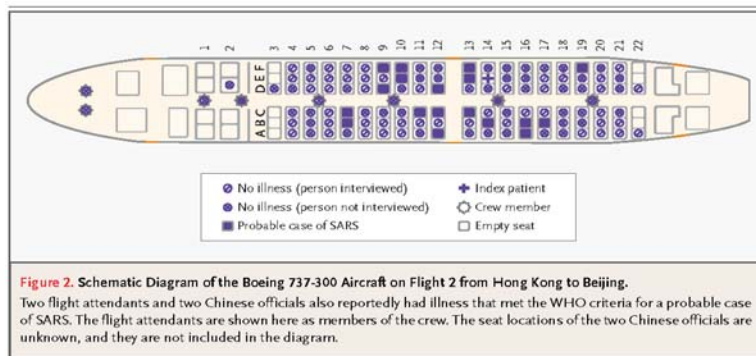
The simplest, familiar model we use is the binomial model.

ADVANCED: Other probabilities models for estimating risk:

- Binomial model (our focus)
- Hazard-based models (not today)
  - Constant hazard (constant rate) model
  - Non-constant hazard model
    - Exponential formula method (fixed time intervals)
    - Kaplan-Meier method (time-to-event data)

## Measures of occurrence – Risk estimation with binomial data

$$R(0,t) = \frac{\#events\ in(0,t)}{Population\ at\ risk\ at\ time\ 0} = \frac{18}{111} = 0.1622$$



29PMID: 14681502

Center for Infectious Disease Preparedness  
UC Berkeley School of Public Health  
www.idready.org

In March, 2003, a single passenger with SARS boarded an airplane that flew 3 hours from Hong Kong to Beijing. Out of the 111 passengers that were disease-free at the start of the flight, 18 developed SARS.

The risk of infection was 18/111 or 16.2%.

QUESTION: What conclusions can we draw from this data?

ANSWER: SARS can be transmitted during an airplane flight.

## Epidemiologic Measures: Measures of occurrence - Odds

Odds is a transformed risk (R) estimate

$$\text{Odds}(0,t) = \frac{R(0,t)}{1-R(0,t)}$$

Example

*If  $R = 1/4$*

$$\text{Then, Odds}(0,t) = \frac{1/4}{3/4} = \frac{1}{3}$$



An odds is just the numerical transformation of a risk estimate.  
Gamblers think of risk in terms of odds.  
Epidemiologists use odds for mathematical convenience.  
For now, we only need to understand the definition of odds.

## Measures of occurrence - Prevalence

Point prevalence

$$P = \frac{\text{Number of existing cases}}{\text{Number in total population}}, \text{ at a point in time}$$

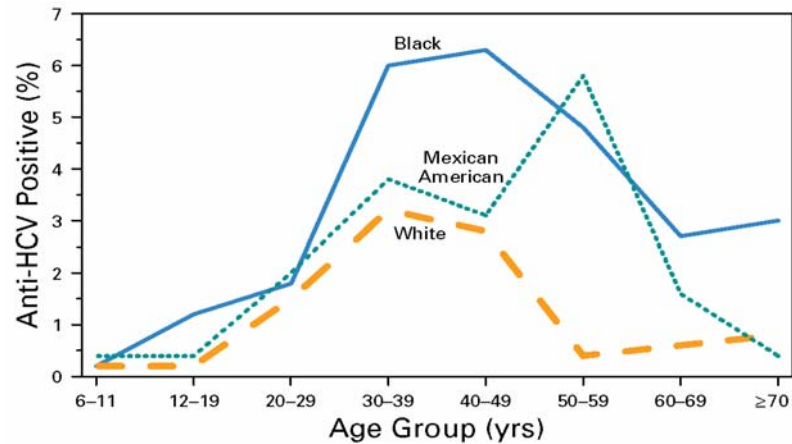
31

Center for Infectious Disease Preparedness  
UC Berkeley School of Public Health  
[www.idready.org](http://www.idready.org)



Prevalence is proportion of a target population with an existing condition at a point in time.

## Prevalence of hepatitis C virus (HCV) infection by age and race/ethnicity — United States, 1988–1994



Source: Third National Health and Nutrition Examination Survey, CDC.

Center for Infectious Disease Preparedness  
UC Berkeley School of Public Health  
[www.idready.org](http://www.idready.org)



32

Here is a graphical display of the prevalence of hepatitis C infection (indicated by anti-HCV antibodies) by age group and ethnicity.

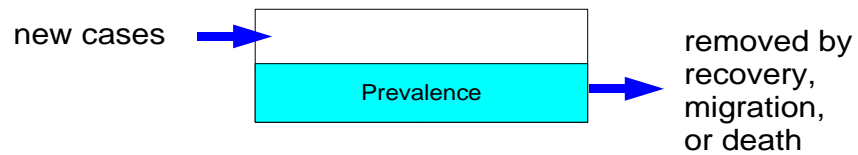
Note:

Age group is a ordinal categorical variable.

Ethnicity is a nominal categorical variable.

Prevalence of anti-HCV positivity is a proportion.

## Measures of occurrence: Prevalence, rate, and duration



Under steady state conditions and small P,

$$Prevalence \approx Rate \times Duration$$



It is useful to know that the prevalence of a condition is a function of rate of new cases and the duration of illness.

## Measures of association

| Measures of occurrence | Measures of association |
|------------------------|-------------------------|
| Rate                   | Rate ratio              |
| Risk                   | Risk ratio              |
| Odds                   | Odds ratio              |

34

Center for Infectious Disease Preparedness  
UC Berkeley School of Public Health  
[www.idready.org](http://www.idready.org)



Measures of association are used to compare a measure of occurrence in an exposed group vs. a non-exposed group.

For example, if  $R_1$  is the rate of lung cancer among smokers, and  $R_0$  is the rate of cancer among non-smokers, the  $RR = r_1/r_2$  is the rate ratio comparing smokers to non-smokers.

Interpretation:

$RR > 1$  means smoking is associated with lung cancer

$RR < 1$  means not smoking is associated with lung cancer (that is, smoking is protective)

$RR \approx 1$  means no association

## Outbreak of nodding off at CIDP lectures

The Centers for Disease Control (CDC) is responding to an outbreak of nodding off at training conferences provided by a west coast CDC Center for Public Health Preparedness. The UC Berkeley Center for Public Health Preparedness trainees are nodding off during lectures at an alarming rate. Local public health authorities have requested assistance from the CDC to determine the cause of this problem.

35

*Center for Infectious Disease Preparedness  
UC Berkeley School of Public Health  
[www.idready.org](http://www.idready.org)*



Here is an example of a fictitious outbreak.

## Outbreak of nodding off at CIDP lectures

Based on initial interviews, the CDC field investigators suspect that the lectures given by Dr. Tomás Aragón, Director of the Center, is putting his students to sleep. At the next 3-day conference, investigators set up a surveillance system to measure students nodding off and who was lecturing when the nodding off occurred.



## Outbreak of nodding off at CIDP lectures

Investigators set up hidden cameras in the lecture hall with unobstructed views of all the students and the lecturers. A case of nodding off was classified as probable if the student appeared to be sleeping with their eyes closed for 2 or more minutes; a case was classified as confirmed if the student snored audibly for 15 or more seconds. All student attendance times, and who was lecturing and for how long was measured.



## Outbreak of nodding off at CIDP lectures

Investigators hypothesize that students exposed to Dr. Aragón were at higher risk for nodding off.

The 3-day conference had 100 attendees. There were 6 hours of lecture per day. Dr. Aragón lectured 6 of 18 hours. Risk of nodding off was estimated as the proportion that nodded off at each lecture. Rates were also calculated.



## Outbreak of nodding off at CIDP lectures: Risk ratio & Odds ratio

|                           | Exposure |       |
|---------------------------|----------|-------|
|                           | Tomás    | Other |
| Individuals nodding off   | 19       | 12    |
| Cases of any nodding off  | 21       | 13    |
| Number started conference | 100      | 100   |

$$R_1 = \frac{19}{100} = 0.19$$

$$R_0 = \frac{12}{100} = 0.12$$

Why is this analytic approach not recommended?

$$RR = \frac{R_1}{R_0} = 1.58$$

$$OR = \frac{R_1/(1-R_1)}{R_0/(1-R_0)} = 1.72$$

$p$  value = 0.24

Center for Infectious Disease Preparedness  
UC Berkeley School of Public Health  
[www.idready.org](http://www.idready.org)



## Outbreak of nodding off at CIDP lectures: Risk ratio & Odds ratio

|                      | Exposure |       |
|----------------------|----------|-------|
|                      | Tomás    | Other |
| Cases of nodding off | 21       | 13    |
| No. started lecture  | 559      | 1151  |

$$R_1 = \frac{21}{559} = 0.0376$$

$$R_0 = \frac{13}{1151} = 0.0113$$

$$RR = \frac{R_1}{R_0} = 3.33$$

$$OR = \frac{R_1/(1-R_1)}{R_0/(1-R_0)} = 3.42$$

$$p \text{ value} = 0.0006$$



## Outbreak of nodding off at CIDP lectures: Rate ratio

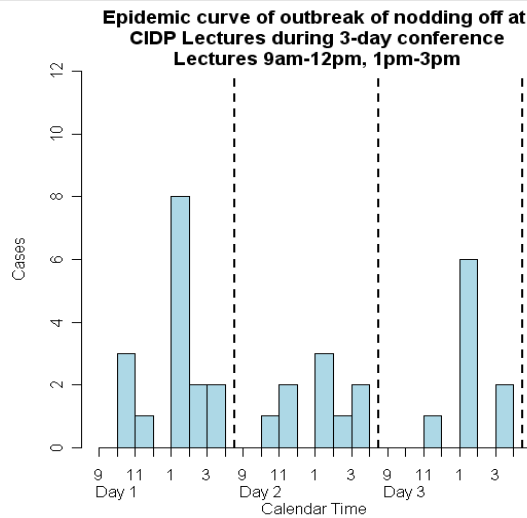
|                         | Exposure |        |
|-------------------------|----------|--------|
|                         | Tomás    | Other  |
| Cases of nodding off    | 21       | 13     |
| Person-hours at lecture | 522.7    | 1047.3 |

$$r_1 = \frac{21}{522.7} \quad r_0 = \frac{13}{1047.3} \quad rr = \frac{r_1}{r_0} = 3.27$$

*p* value = 0.0008



# Epidemic curve of nodding off



For 3 days, Dr. Aragón lectured the hour before *and* the hour after lunch.

What else could explain the nodding off?

What is this called?



## Epidemiologic jargon!!!

- Ambiguous terminology
  - Attack rate
  - Case fatality rate, Case fatality ratio
  - Survival rate
  - Prevalence rate
- The many lives of “incidence”
  - Incidence
  - Incidence rate
  - Incidence density
  - Cumulative incidence

43

Center for Infectious Disease Preparedness  
UC Berkeley School of Public Health  
[www.idready.org](http://www.idready.org)



Avoid sloppy terminology. If you call a measure a rate, be sure it is actually a rate.

Be aware that the word “incidence” is used in epidemiologic terms with very different meanings.

“Incidence” means new case counts. “Incidence rate” and “incidence density” mean rate. “Cumulative incidence” means risk. Instead just precise, unambiguous terminology:

- Count: e.g., number of new cases
- Count: e.g., number of prevalent cases
- Time: e.g., median survival time
- Rate
- Risk (or odds)
- Prevalence

# Epidemiologic study designs

- Study designs
  - Experimental (Randomized control trial)
  - Observational (Cohort, Case-control)
- Steps
  - Design (study protocol)
  - Implementation (operations manual)
  - Analysis
  - Interpretation



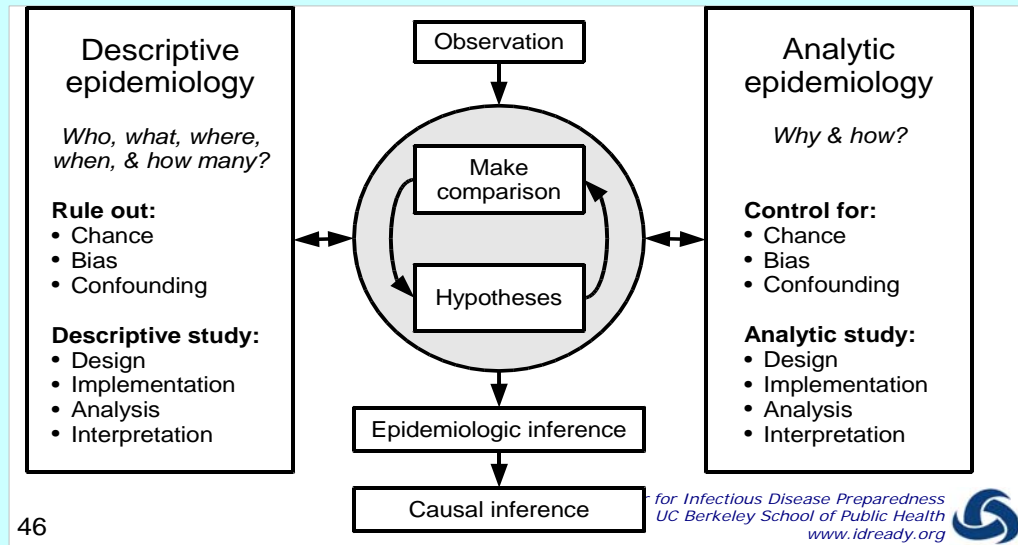
## Threats to validity

- Chance: Random error
  - Confidence interval (precision)
  - P value ( $\Pr\{X \geq x\}$ , under null hypothesis)
- Bias: Systematic error
  - Selection bias
  - Measurement bias (information bias)
- Confounding



notes pending

# Inferences in epidemiology



We have already discussed the process of descriptive and analytic epidemiology.

In both approaches we need to address threats to making valid epidemiologic inferences (chance, bias, and confounding) from our studies.

Causal inference is concluding that an exposure is “causal.” This involves incorporating current investigation findings with existing evidence from other studies and scientific disciplines.

In public health, when findings are inclusive, we generally err on the side of protecting the public health.

## Public health action

- Clinical
- Behavioral
- Community
- Environmental

Adapted from Haddix AC, et al. Prevention Effectiveness: A Guide to Decision Analysis and Economic Evaluation. Oxford University Press 2003, 2nd Edition

47

Center for Infectious Disease Preparedness  
UC Berkeley School of Public Health  
[www.idready.org](http://www.idready.org)



Public health interventions occur at four levels:

- In the clinic with patients.
- Changing individual behaviors
- Changing community norms.
- Changing the environment so that it is safer.