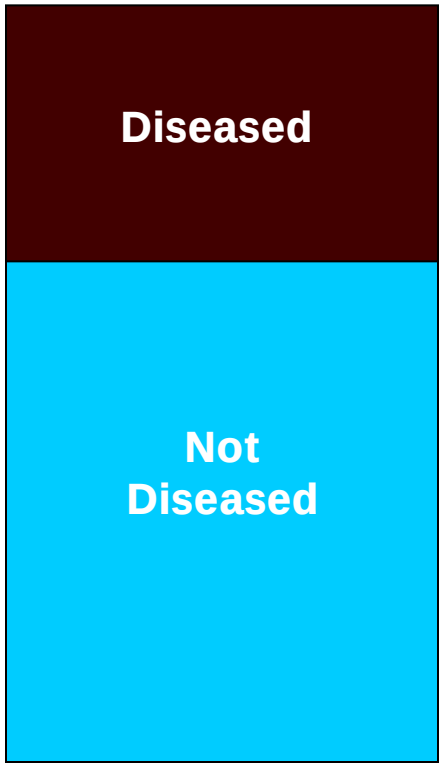
Introduction to Statistics

Learning Objectives

- Describe the indices used to measure health risk
- Define and calculate ratios, proportions, incidence, prevalence, attack rate
- State the reasons why rates are preferable to absolute numbers when describing a population
- Describe how to express a measures of disease using time, place, person

Measures of Frequency

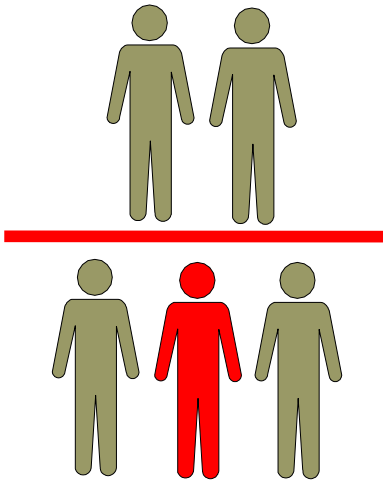


We often want to know:

1) How many people have a disease?

2) What proportion of the population has disease?

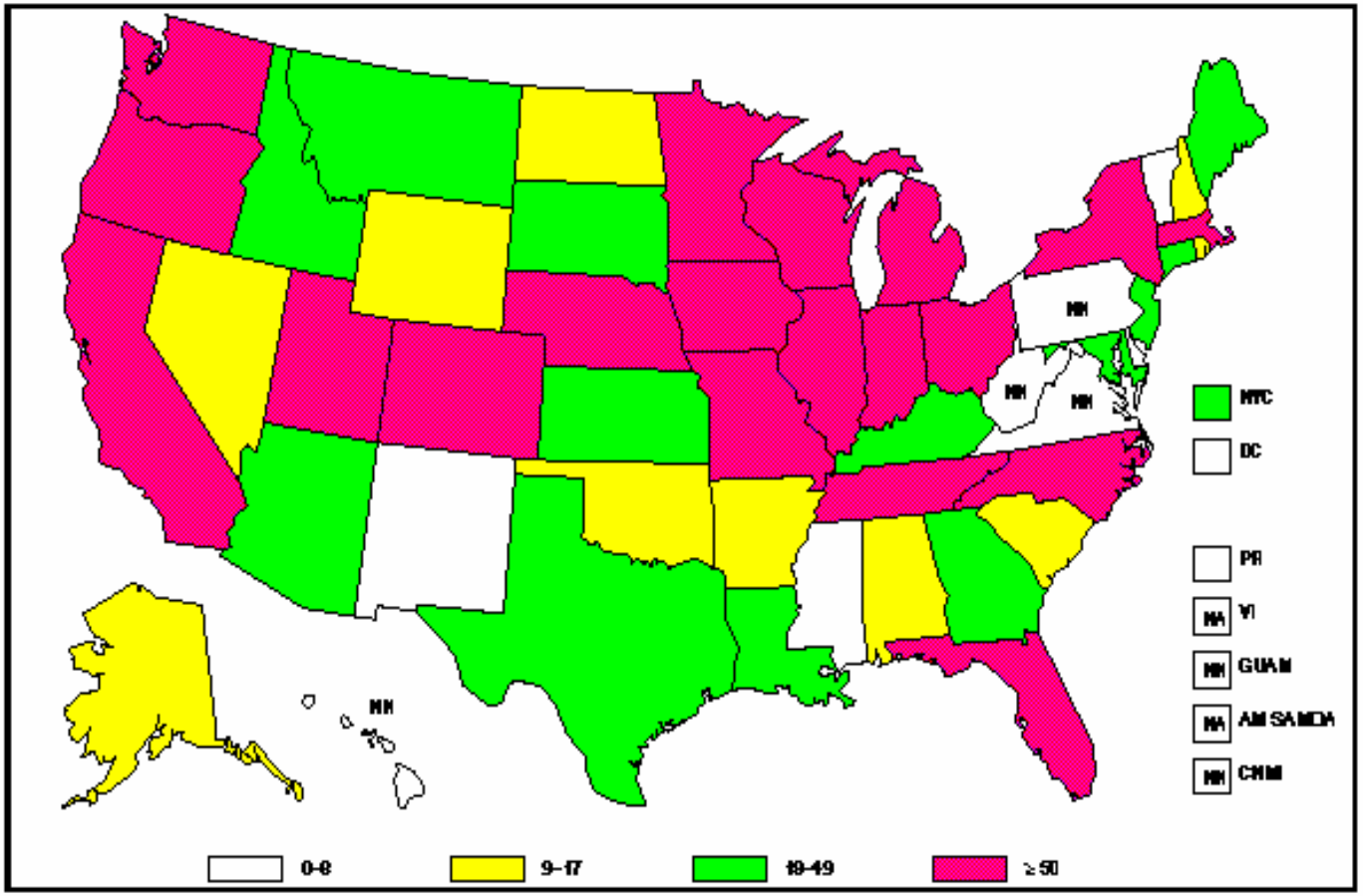
3) What proportion of the population could still get the disease?



- **Counts**
- **Ratio**
- **Proportion**
- **Percentages**
- **Rate**

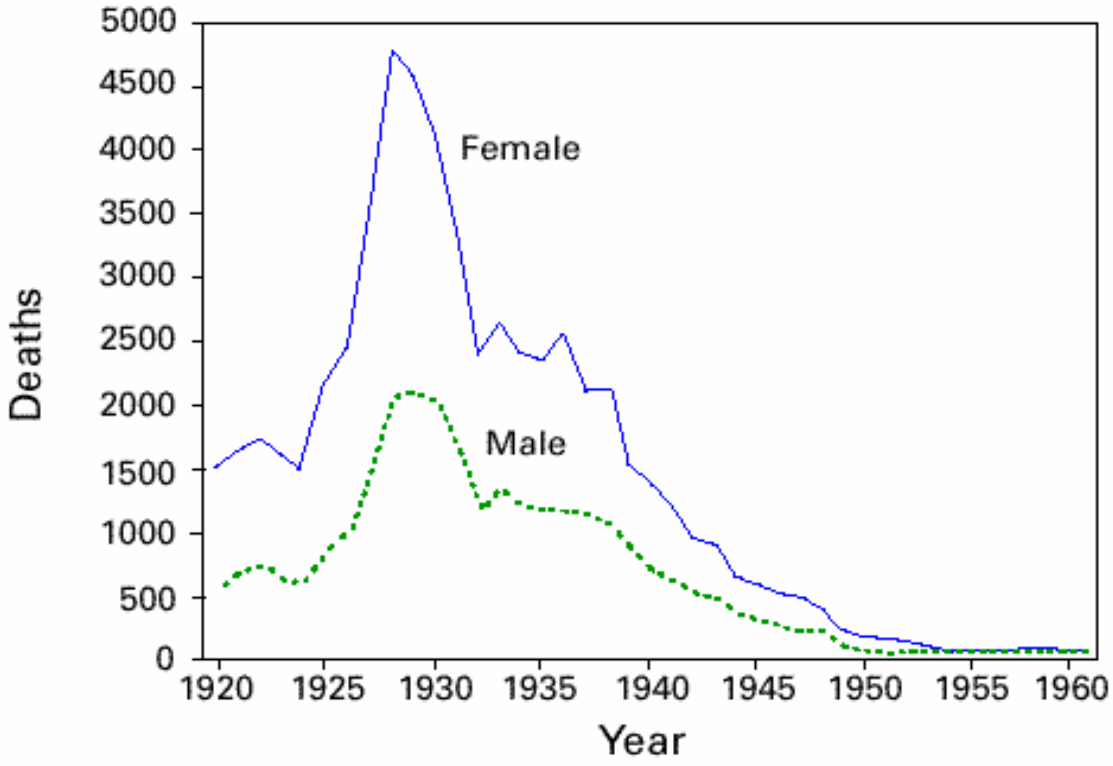
Counts

ESCHERICHIA COLI O157:H7 — reported cases, United States and territories, 1997



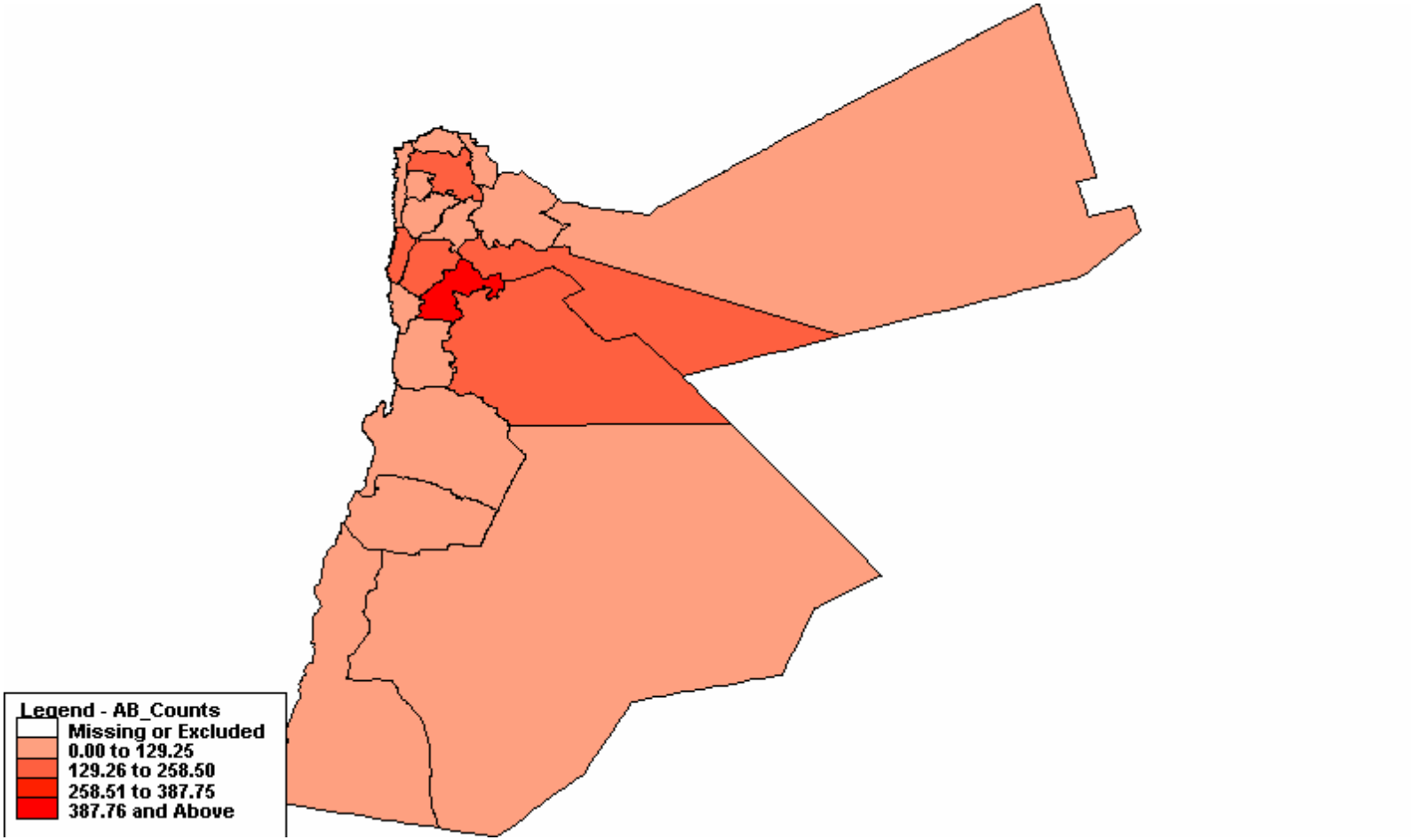
The number of states in which *E. coli* O157:H7 infection is a notifiable disease increased from 44 in 1996 to 45 in 1997. However, because <60% of clinical laboratories routinely test all stools — or even all bloody stools — for *E. coli* O157:H7, many infections are not recognized or reported.

FIGURE 2. Number of reported pellagra deaths, by sex of decedent and year — United States, 1920–1960



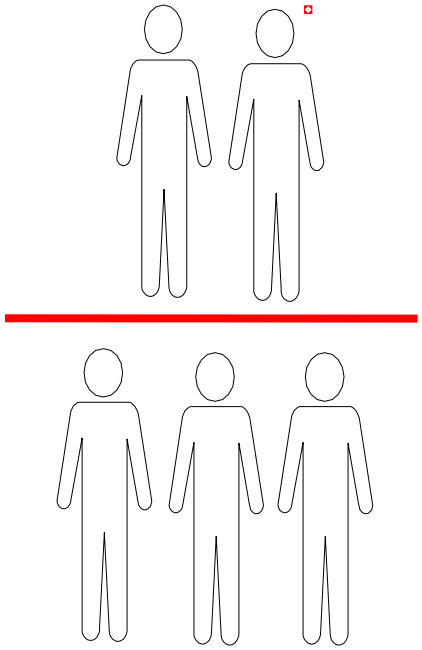
MMWR October 15, 1999 / 48(40);905-913

Counts of animal bites Jordan 2001



Ratio

- The quotient of two numbers
- Numerator **NOT necessarily INCLUDED** included in the denominator



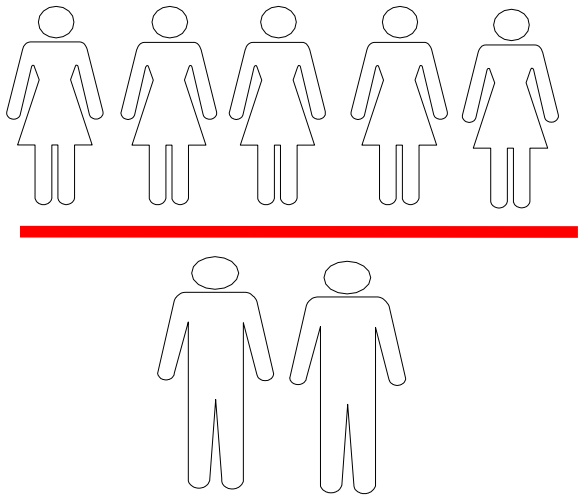
X 100

What, who is in the denominator ?
What, who is in the numerator?

Ratio Example

What is the number of females per one male?

$$\frac{\text{\# Females}}{\text{\# Males}} * 100$$



= 5 / 2 = 2.5 / 1

Sex Ratio

A typical epidemiological ratio

For example, in 1999 in country X what is the sex ratio of males-to-females in the age group 45-49 ?

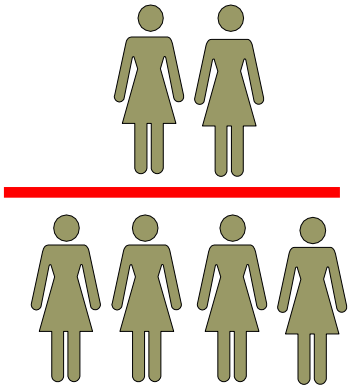
$$= \frac{76{,}875 \text{ males}}{72{,}470 \text{ females}} = 1.06 / 1$$

In the age group 65+?

$$= \frac{64{,}055 \text{ males}}{67{,}795 \text{ females}} = 0.94 / 1$$

Proportion

- The quotient of 2 numbers
- Numerator **IS INCLUDED** In the denominator
- Quantities have to be of the same nature
- Proportion always ranges between 0 and 1
- Percentage = proportion x 100



$$\frac{2}{4} = 0.5 = 50\%$$

Prevalence

Number of cases of disease at a specific time

Population at risk at that time

Proportion of a population affected by a disease at a given time.

Example:

$$\textit{Prevalence} = \frac{\text{100 people}}{\text{1000 people}} = .1$$

Prevalence will always fall between 0 and 1

Effect of prevalence on two populations January 1st, 2001

	Population 1	Population 2
Diabetes	1-1-2001	1-1-2001
Yes	250	250
No	750	750
Total	1 000	1 000
Prevalence	25%	25%

Effect of prevalence on two populations January 1st, 1999 and 2001

	Population 1		Population 2	
Diabetes	**1-1-1999**	1-1-2001	**1-1-1999**	1-1-2001
Yes	**50**	250	**150**	250
No	**950**	750	**850**	750
Total	**1000**	1 000	**1000**	1 000
Prevalence	**5%**	25%	**15%**	25%

Case-fatality

$$\frac{\text{Number of deaths due to disease X}}{\text{Number of diagnosed cases of disease X}}$$

This is a measure of the propensity of a disease to casue death of the affected individual.

Rate

- The quotient of 2 numbers
- **Time is always included in the denominator**
  - Something observed during some time
  - Measures the speed of occurrence of an event
  - Measure the probability to become sick by unit of time
  - Measures the risk of disease

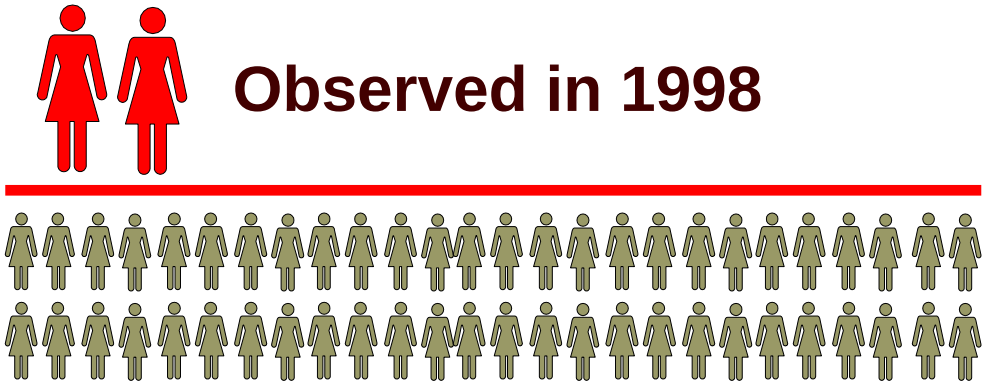
Rate Example

Numerator

- **number of NEW EVENTS observed for a given time**

Denominator

- **population in which the events occur (population at risk)**
- **includes time**



Observed in 1998

$$\frac{2}{100} = 0.02 \text{ / year}$$

Incidence (rate)

Cumulative Incidence

- Incidence Proportion
- Attack Rate

Incidence Density

- Person-time

Cumulative Incidence (CI)

Number of NEW cases of disease during a period

Population at risk during this period

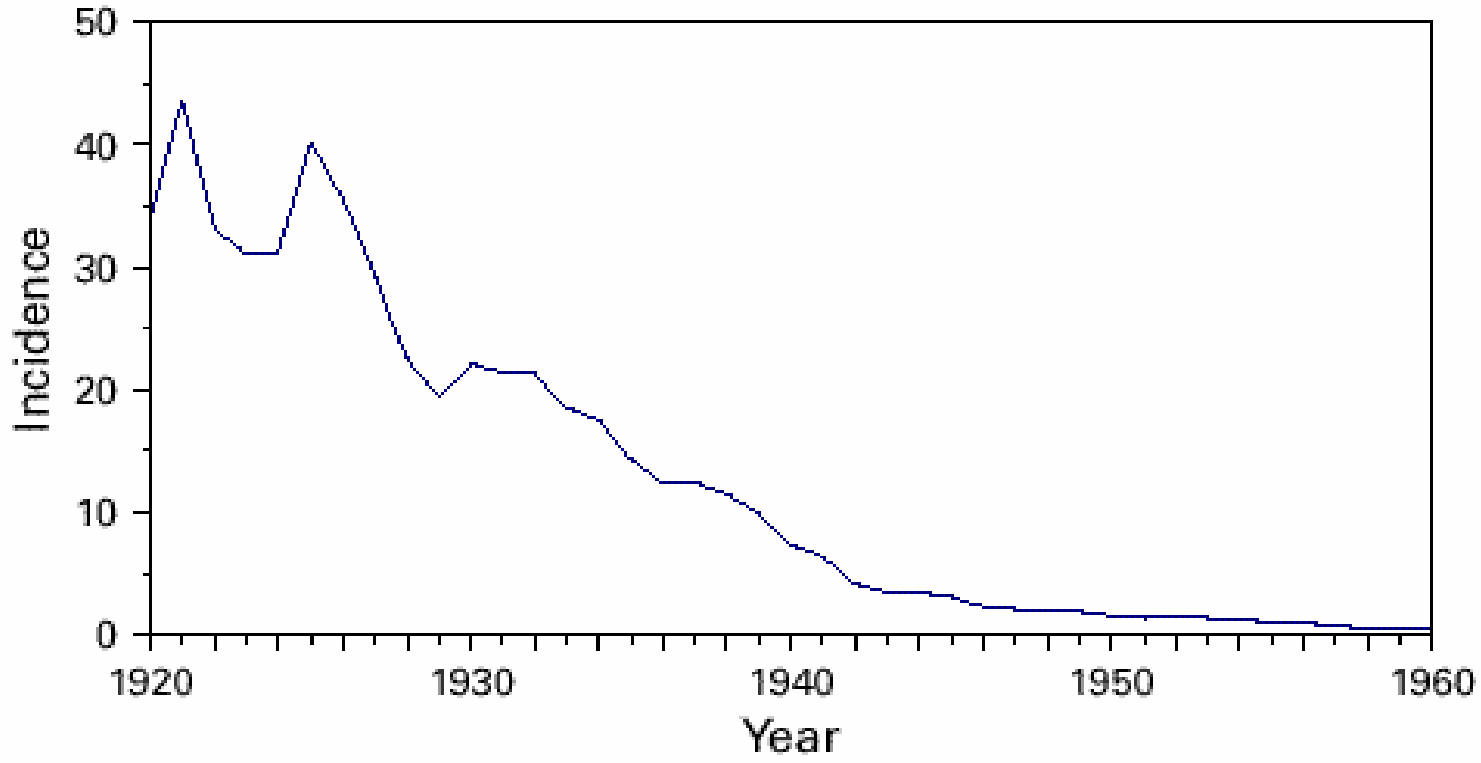
In Cumulative Incidence the denominator includes only disease-free (at risk), and excludes people who are pre-existing cases or unable to get the disease.

Example:

$$\textit{Cumulative Incidence} = \frac{\text{200 new cases}}{\text{1000 people}} = .20$$

Cumulative Incidence Example

FIGURE 1. Incidence* of typhoid fever, by year — United States, 1920–1960



*Per 100,000 population.

MMWR October 15, 1999 / 48(40);905-913

Cumulative Incidence (CI)

Number of NEW cases of disease during a period

Population at risk during this period

Example:

***Cumulative Incidence* = $\frac{\text{200 new cases}}{\text{1000 people}}$ = .20**

But what if 100 people had pre-existing disease?

***Cumulative Incidence* = $\frac{\text{200 new cases}}{\text{1000-100 people}}$ = .222**

= 22.2 cases/100

Attack Rate (Incidence Proportion)

Number of new cases among the population during a limited time period

The population at risk at the beginning of the period

Attack Rate (Incidence Proportion)

For example, outbreak investigation at a church supper

- 46 ill and 75 attended

Attack rate (AR) = $\frac{46}{75}$ or AR of 61 percent

TIME

TIME

TIME

TIME

TIME

Rates

TIME

TIME

TIME

TIME

TIME

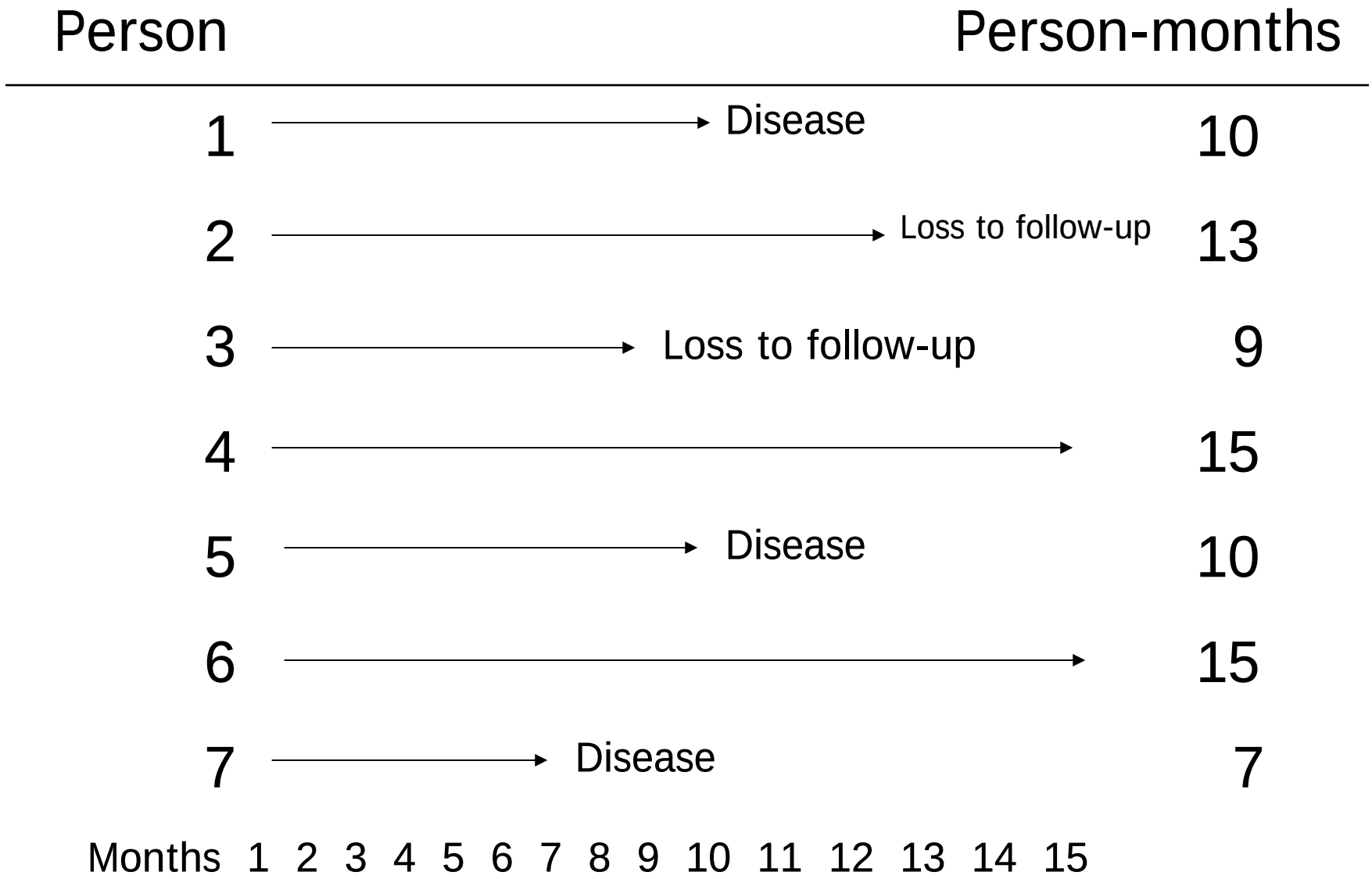
Incidence Rate (incidence density)

It is a measure of the instantaneous rate of the development of disease in a population

$$= \frac{\text{No. of new cases during time T}}{\text{Total person-time of observation}}$$

*accounts for varying time periods of follow-up

Person - Time


Person
Person-months
1
Disease
10
2
Loss to follow-up
13
3
Loss to follow-up
9
4
15
5
Disease
10
6
15
7
Disease
7
Months 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Incidence Rate using individual data

$$= \frac{\text{No. of events (x)}}{\text{person-time (t1 + t2 + . . .+ tN)}}$$

$$= \frac{\text{3 new cases of disease}}{\text{10 + 13 + 9 + 15 + 10 + 15 + 7 months of observation}}$$

$$= \frac{\text{3 new cases of disease x 100}}{\text{79 person-months}} = \frac{\text{3.8 cases}}{\text{person-months}}$$

Comparing Incidence and Prevalence

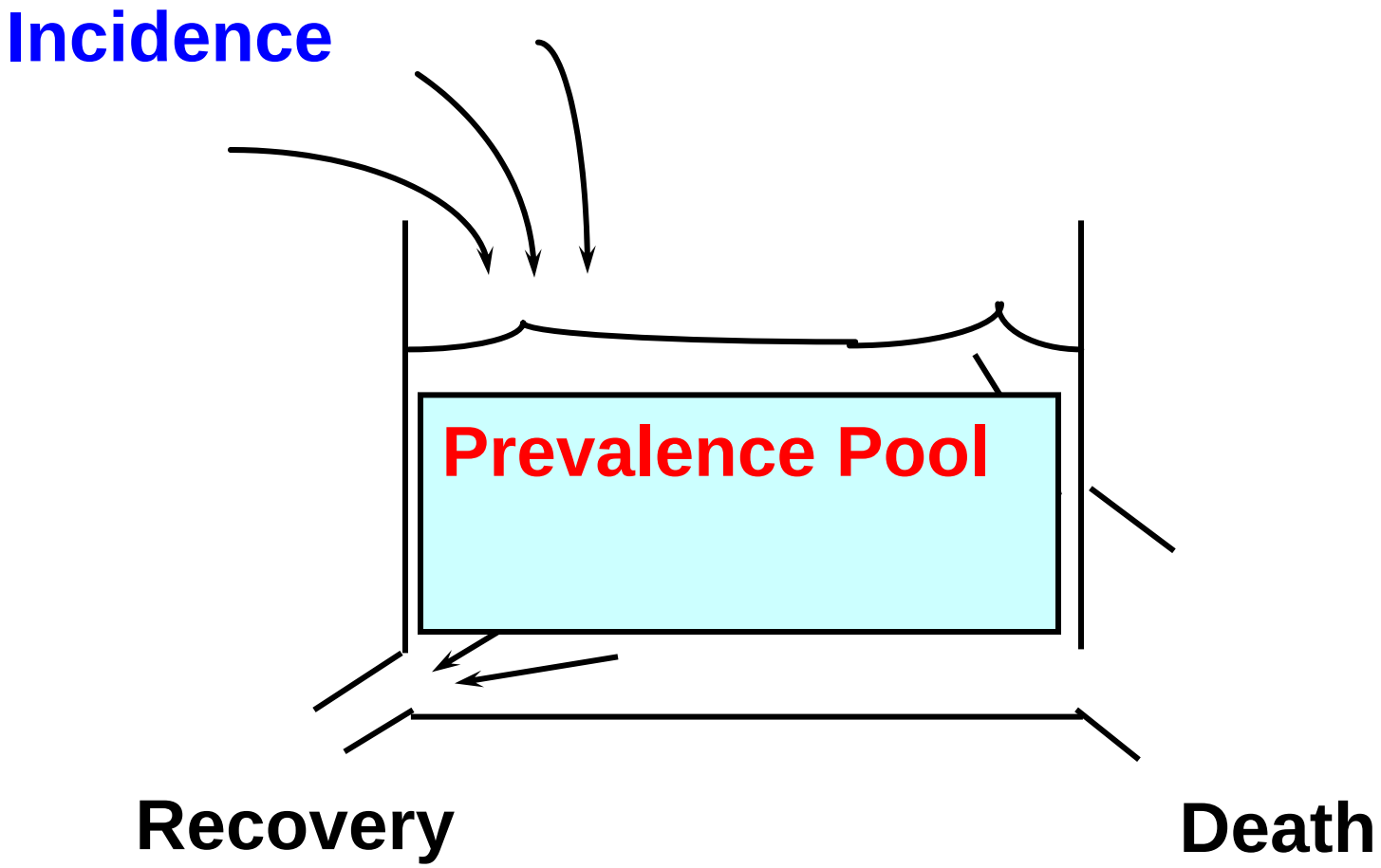
Incidence

- **NEW** cases or events over period of time
- Useful studying factors causing disease, disease "risk"

Prevalence

- **ALL** cases at point/period of time
- Useful for measuring size of problem and planning

Prevalence and Incidence

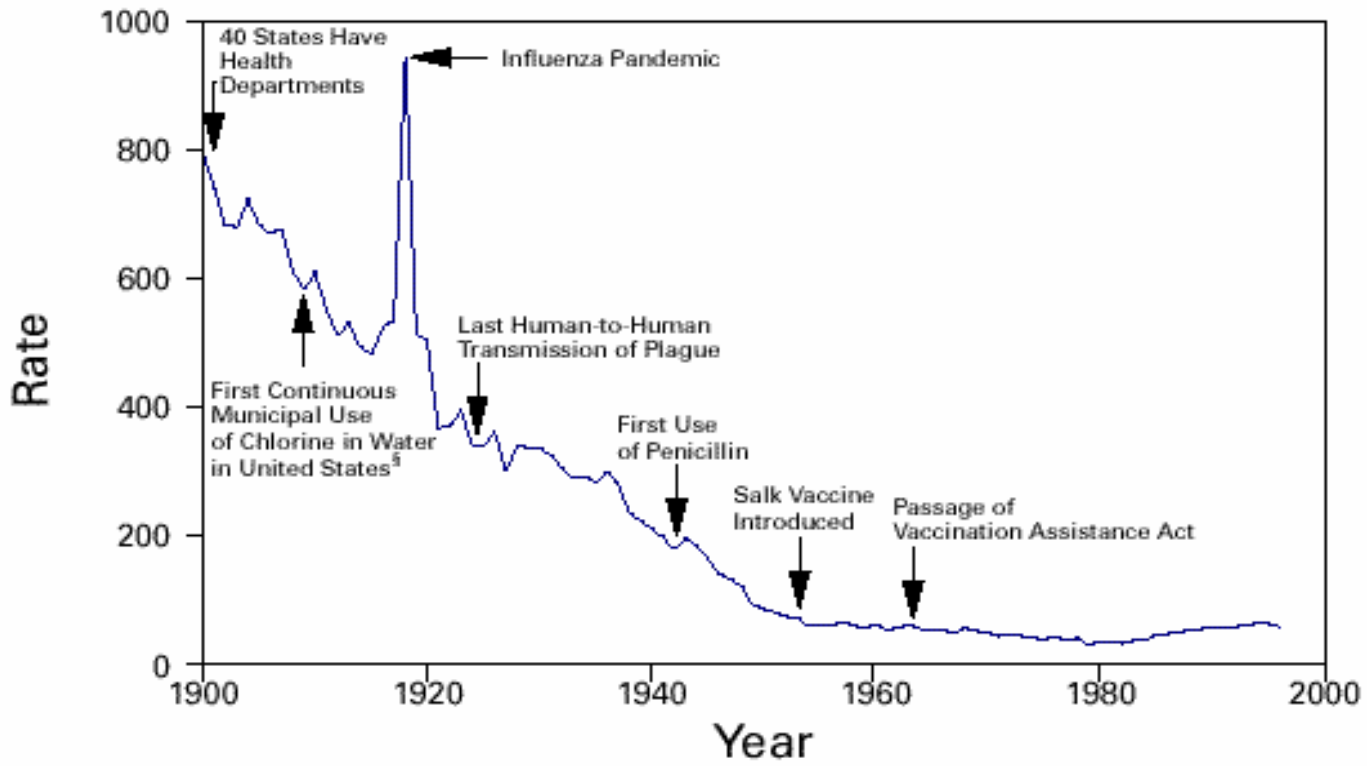


Crude mortality rate

$$\frac{\text{Total number of deaths from all causes in a population}}{\text{Number of the total population}}$$

Mortality rates

FIGURE 1. Crude death rate* for infectious diseases — United States, 1900–1996†



*Per 100,000 population per year.

MMWR July 30, 1999 / 48(29);621-629

Infant mortality rate

Number of deaths among children under 1 year of age reported during a time period

Number of live births reported during the same period

Example: U.S. Infant Mortality Rate

$$= \frac{\text{Number of Deaths <1yr in 1989}}{\text{Number of Live Births in 1989}}$$

$$= \frac{39{,}601}{4{,}040{,}958} = .0098$$

= **9.8 per 1000 Live Births**

Practice

In a Central Asian country with a population of six million people, there were 60,000 deaths during the year ending December 31, 1997. These included 30,000 deaths occurring in 100,000 people who were sick with cholera.

Case fatality rate from cholera in 1997 ?

Cause specific mortality rate from cholera in 1997 ?

Comparison of Calculations

Ratio	**Division of two unrelated numbers**
Proportion	**Division of two related numbers; numerator is a subset of denominator**
Rate	**Division of two related numbers; numerator is a subset of denominator; time is always in denominator**

Using measures of disease frequency

When a statistic, for example prevalence or incidence, is being described or interpreted, we need to reference

Who -- what population or subgroup

When -- what time point or time period

Where -- the geographical location

Expressing a measure of disease frequency

In 1998, the estimated cumulative incidence of men who ever smoked a cigar was 25.5% in the United States.

The estimated cumulative incidence was 25.5% among men

Counts or rates?

Comparing populations by PERSON

Person	Women aged 15-29 years and women aged 30-44 years	Different populations
Place	In Epiland	Same place
Time	In the year 2000	Same time

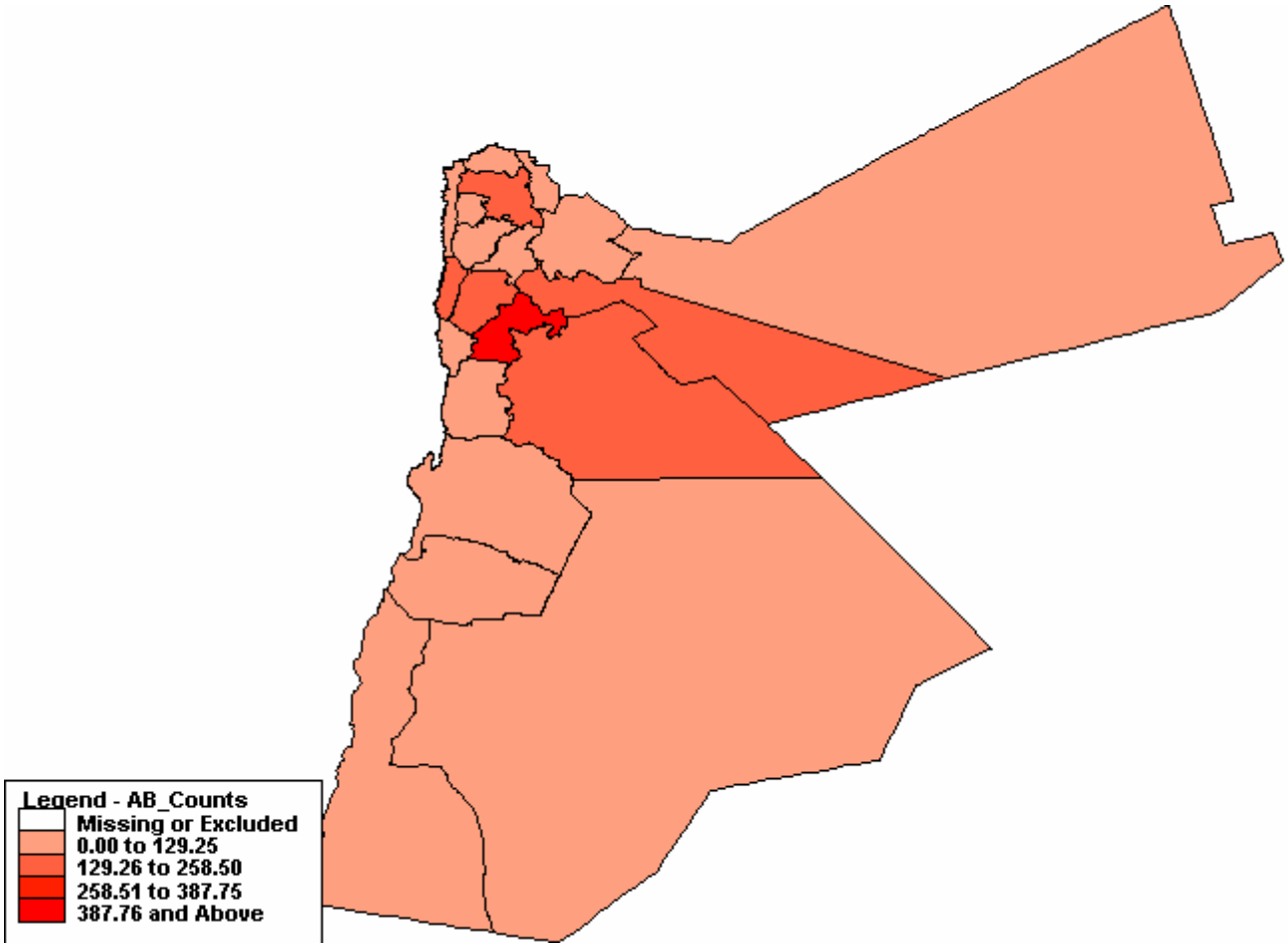
Comparing populations by TIME

Person	All residents	Same population
Place	of Epiland	Same place
Time	In the years 1990 and 2000	Different times

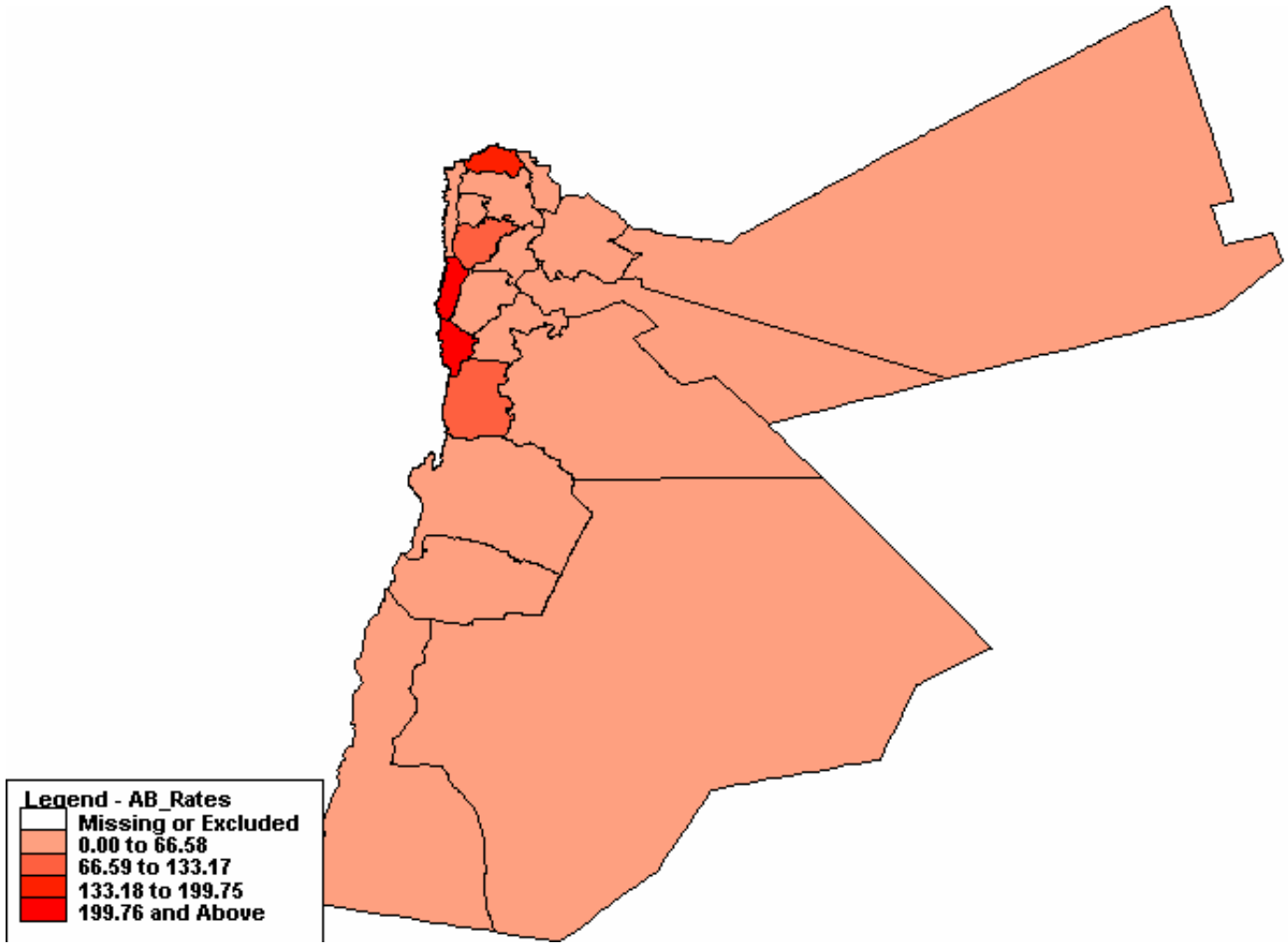
Comparing populations by PLACE

Person	Women aged 15--44 years	Same population
Place	in Epiland and Demiland	Different places
Time	in the year 2000	Same time

Counts of animal bites Jordan 2001



Incidence of animal bites Jordan 2001



Final Thoughts

- Epidemiological terms can be confusing
- For frequency measures know what is in the numerator, denominator, and the time period
- Reminder:
  - Numerator in prevalence is new and pre-existing cases
  - Numerator in incidence is new cases

Acknowledgements

- For Excellent Slides
  - Material modified from Epi course 503, Univ. Michigan accessed 2/16/2002 at:
    http://www.sph.umich.edu/epid503/LecutreJan2201.ppt
  - TEPHINET CD Presentations

- Reference books
  - J. Last, *A Dictionary of Epidemiology 4th ed*
  - C Hennekens, *Epidemiology in Medicine, 1987*
  - B. Gertsman, *Epidemiology Kept Simple, 1998*
  - T. Timmreck, *An Introduction to Epidemiology, 3rd ed, 2002*