EPID/CPH 573A - Basic Principles of Epidemiology

Week #1:

- **Introduction**
- **Historical perspective**
- **"Descriptive epi": person, place, time**
- **Measures of frequency ("risk")**

Slides for Week 1: developed by Dr. Eyal Shahar; minor modifications made by Dr. Leslie Dennis

Learning Objectives

1. What is epidemiology?
2. Know something about the history of our science
3. Begin to learn how to read graphs/tables that contain epidemiological data
4. Know how to compute three measures of frequency ("chance"; "risk"):
   - Probability (proportion)
   - Odds
   - Rate

Definitions of Epidemiology

Epi = over; about
Demos = people;
Logos = doctrine; rational arguments

- Lilienfeld, 1978: 'The study of the distribution of a disease or a physiologic condition in human populations and of the factors that influence this distribution'
- Last, 1988: 'The study of the distribution and determinants of health-related states or events in specified populations and the application of this study to control health problems'

Definitions of Epidemiology

- Hennekens, 1987: 'The study of the distribution and determinants of disease frequency in human populations'
- Hirsch, 1983: 'A **science** which will give, firstly, a picture of the occurrence, the distribution and types of diseases of mankind, in distinct epochs of time and at various points of the earth's surface; and secondly, will render an account of the relations of these diseases to the external conditions surrounding the individual and determining his manner of life'

Central Concepts of Epidemiology

- There are causes for whatever happens
  - Smoking causes lung cancer
  - Weight affects blood pressure
  - Aspirin prevents a heart attack
- Learn how to estimate their effects (methodology)
- Estimate their effects (science)

- Study designs
  - Cross sectional
  - Ecologic
  - Case-control
  - Cohort
  - Randomized experiments
- Analytical methods
  - Causal theories and diagrams
  - Theoretical basis of bias
  - Methods to deal with bias
  - Methods to estimate effects

The research method component of epidemiology

These are the major themes

Overlap with other disciplines

- Biostatistics
- Econometrics
- Health services research
- Educational psychology

How Epidemiology Contributes to Public Health

- Public health cores
  - Health services research
  - Biostatistics
  - Epidemiology
  - Behavioral sciences/health education
  - Environmental health sciences
  - Maternal and child health
- Public health practice
  - Departments of health
  - CDC
  - Public health service
  - Other public health organizations

Major Eras in Epidemiology

Era	Time Frame	Paradigm	Key Figures
"Sanitary Statistics"	Early 19th Century	Miasma Theory (bad air)	Pierre C. A. Louis William Farr
Infectious Disease Epidemiology	Late 19th to mid 20th Century	Germ Theory	Robert Koch John Snow Louis Pasteur Jonas Salk
Epidemiology	Latter half of 20th Century	Estimate causal effects	Too early for name dropping

Major Eras in Epidemiology

Era	Public Health Achievements	Methodological Advances
Sanitary Statistics	Drainage & sewage systems; Garbage collection; Clean water; Housing standards	Use of quantitative methods
Infectious Disease Epidemiology	Vaccination; Control & eradication of infectious diseases	Laboratory methods
Epidemiology	Many	Causal thinking Causal diagrams Study design Data analysis

EPID/CPH 573A

Week #1:

- **Introduction**
- **Historical perspective**
- **"Descriptive epi": person, place, time**
- **Measures of frequency ("risk")**

Classic Epidemiological Investigations

- Citrus Fruit & Scurvy (Lind)
- Smoking & Lung Cancer (Doll & Hill)
- Drinking Water & Cholera (Snow)
  (In-class exercise)

James Lind and Scurvy (vitamin C deficiency):

The first controlled clinical trial

18th Century Navy Life

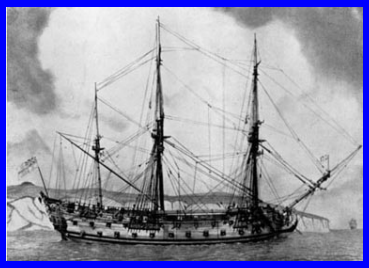
"the number of seaman in time of war who died of shipwreck, capture, famine, fire or sword, are but of inconsiderable in respect of such as are destroyed by the ship diseases, and by the usual maladies of intemperate climates."

- James Lind, 1753

James Lind (1716-1794)

Some symptoms of Scurvy

- lethargy
- irritability
- joint pain
- spontaneous bruising and bleeding
- weight loss



H.M.S. Salisbury

Typical Diet: HMS Salisbury

Daily
1 lb., 4 oz. of cheese biscuits
1 gallon of beer

3 days / week
2 oz. dried fish and butter

2 days / week
2 lb. salted beef

4 days / week
8 oz. peas

Study Population

"On the 20th of May, 1747, I selected 12 patients in the scurvy, on board the Salisbury at sea. Their cases were as similar as I could have them. They all in general had putrid gums, the spots and lassitude, with weakness in their knees. They lay together in one place, being a proper apartment for the sick in the forehold...."

- James Lind, 1753, English naval surgeon

Study Design: usual diet + dietary supplement (6 options)

Quart of apple juice daily

25 drops of elixir vitriol (sulfuric acid and aromatics)

2 spoonfuls of vinegar 3 times a day

Concoction of herbs and spices

2 oranges and 1 lemon daily

half-pint of sea water daily (comparison group)

Results of the Experiment

"…the most sudden and visible good effects were perceived from the use of the oranges and lemons; one of those who had taken them being at the end of six days fit for duty….The other was the best recovered of any in his condition, and being now deemed pretty well was appointed nurse to the rest of the sick..."

- James Lind, 1753

Scurvy: A Timeline

1601 James Lancaster (sea captain) records lemon juice as protective against scurvy

1744 British Admiral George Anson completes trip around the world in the *Centurion*. Over half of the 2,000 sailors die from scurvy

1747 Lind's experiment on the HMS Salisbury

1754 Lind publishes "A Treatise of the Scurvy"

1775 Captain James Cook completes historic voyages. Sailors remained free from scurvy.

1795 British Navy supplies daily ration of lime or lemon juice

1932 Waugh and King (U of Pittsburgh) and Szent-Gyorgyi (Hungary) isolate and synthesize vitamin C

Classic Epidemiological Investigations

- Citrus Fruit & Scurvy (Lind)
- Smoking & Lung Cancer (Doll & Hill)
- Drinking Water & Cholera (Snow) (In-class exercise)

Smoking & Lung Cancer:

Risk Factor Epidemiology Comes of Age

Smoking and Health: A Timeline

1st century	Native Americans begin smoking tobacco
1492	Arawaks give Christopher Columbus dried tobacco leaves as a gift
1531	European cultivation of tobacco begins
1740	First American tobacco factories in Virginia

In 1947, "…generally, the medical profession in those days did not seriously think about smoking as a potential cause of major diseases." E.L. Wynder, *American Journal of Epidemiology* 1997

British Medical Journal
September 30, 1950

Smoking and Carcinoma of the Lung
Preliminary Report
by
Richard Doll, M.D., M.R.C.P.
A. Bradford Hill, Ph.D., D.Sc.

Lung Cancer: A Growing Epidemic

"…The rise [of lung cancer in England and Wales] seems to have been particularly rapid since the end of the first world war; the death rate of men ages 45 and over increased sixfold and of women of the same ages approximately threefold. This increase is still continuing. It has occurred, too, in Switzerland, Denmark, the U.S.A., Canada, and Australia, and has been reported from Turkey and Japan."

Possible Causes of the Epidemic (1950)

(1) Improved diagnosis

(2) Air pollution from various sources: the exhaust fumes of cars, the surface dust of tarred roads, industrial plants, and coal fires

(3) Tobacco smoking

Study Design: a case-control study (does it make sense?)

Lung cancer cases - - - - -> Smoking history

Non-cancer controls* - - -> Smoking history

* Patient of same sex, 5-year age group at the same hospital

Results

	Cases	Controls
Smoker	688 (97%)	650 (92%)
Non-smoker	21 (3%)	59 (8%)
	709	709

Results

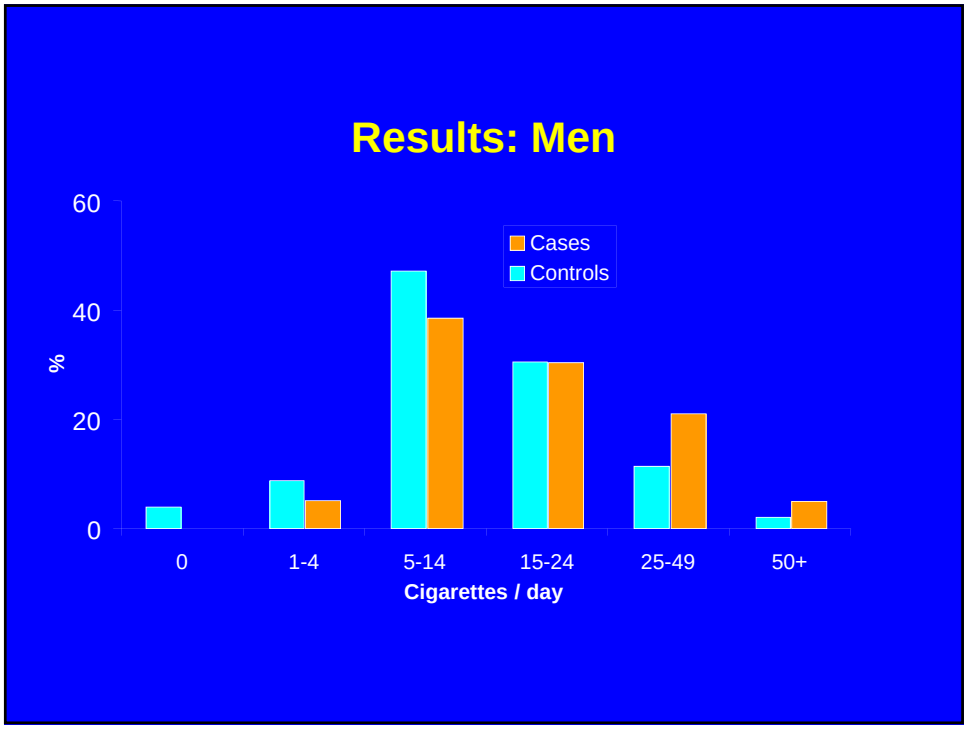
MEN

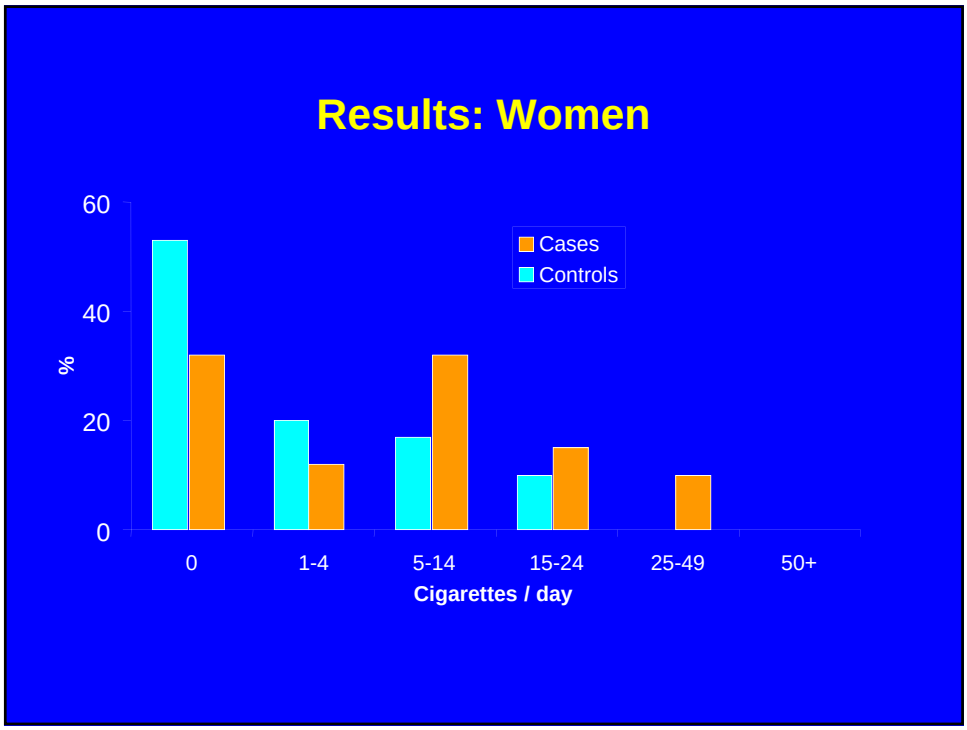
	Cases	Controls
Smoker	647 (99.7%)	622 (96%)
Non-smoker	2 (0.3%)	27 (4%)
	649	649

WOMEN

	Cases	Controls
Smoker	41 (68%)	28 (47%)
Non-smoker	19 (32%)	32 (53%)
	60	60

Why STRATIFY on gender?


Results: Men
60
40
20
0
%
Cases
Controls
0
1-4
5-14
15-24
25-49
50+
Cigarettes / day


Results: Women
60
40
20
0
%
Cases
Controls
0
1-4
5-14
15-24
25-49
50+
Cigarettes / day

Conclusions

"To summarize, it is not reasonable, in our view, to attribute the results to any special selection of cases or to bias in recording. In other words, it must be concluded that there is a real association between carcinoma of the lung and smoking."

"As to the nature of the carcinogen we have no evidence."

HPV & Cervical cancer

- 1952 study of 13,000 nuns found no cases of cervical cancer
- Thus lead to various studies around sexual activity and cervical cancer
- HPV association with cancer first proposed in 1972. Early 1980s was associated with cervical cancer

EPID/CPH 573A

Week #1:

- Introduction
- Historical perspective
- "Descriptive epi": person, place, time
- Measures of frequency ("risk")

The so-called "descriptive epidemiology"
Person, Place, Time

Person: distributions by personal characteristics (by gender, by age, etc.)

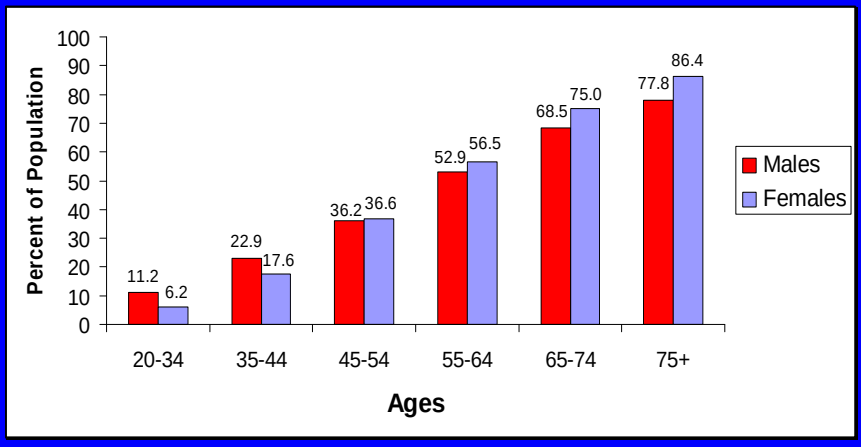
Place: distributions by location

Time: distribution over time

What is the purpose?

Prevalence of Cardiovascular Diseases in the US Ages 20 and Older, by Age and Sex

NHANES: 1999-2002

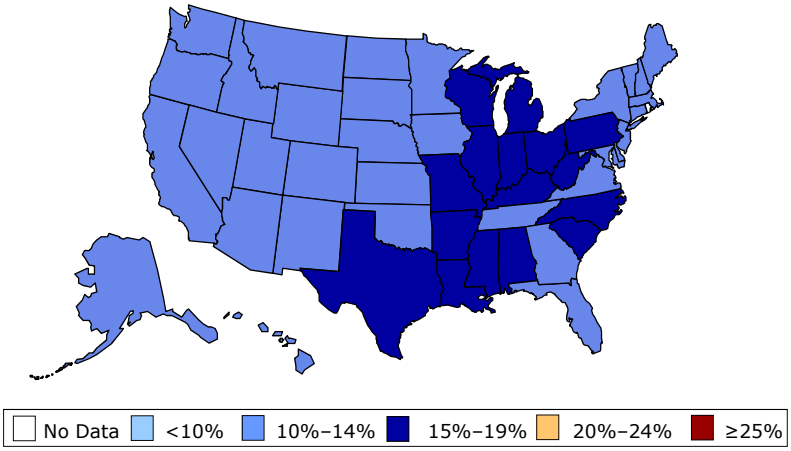


Source: CDC/NCHS and NHLBI. These data include coronary heart disease, heart failure, stroke and hypertension.

Obesity Trends* Among U.S. Adults

BRFSS, 1994

(*BMI ≥30, or ~ 30 lbs overweight for 5' 4" person)

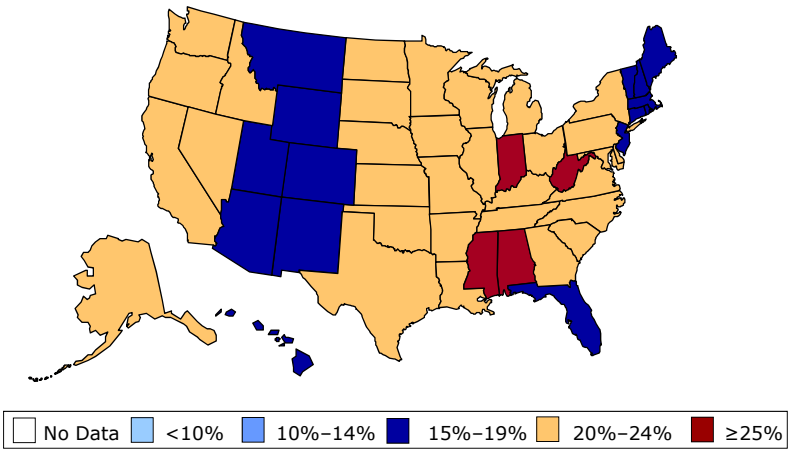


Source: Behavioral Risk Factor Surveillance System, CDC

Obesity Trends* Among U.S. Adults

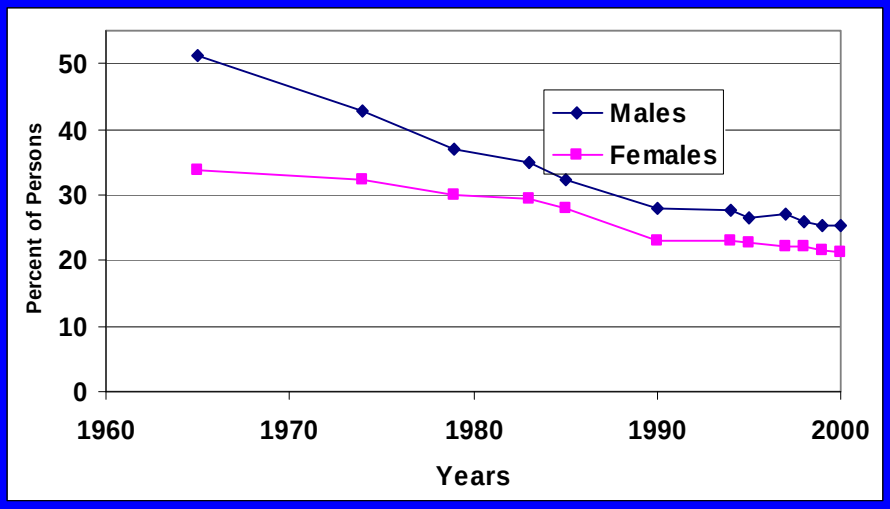
BRFSS, 2003

(*BMI ≥30, or ~ 30 lbs overweight for 5' 4" person)

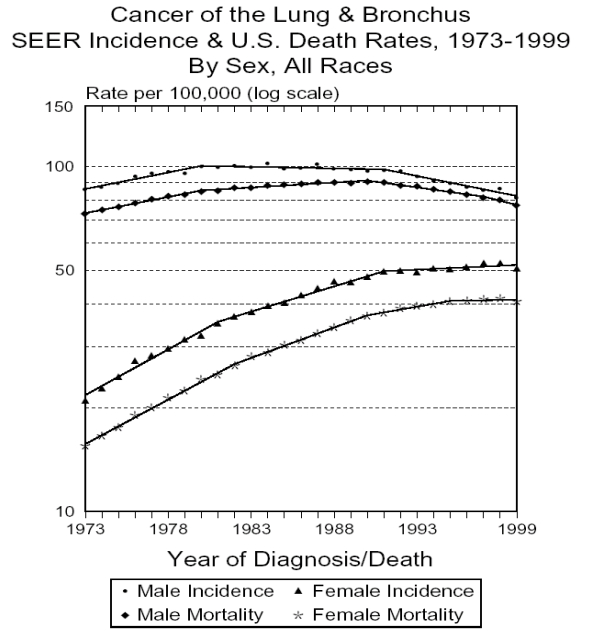


Source: Behavioral Risk Factor Surveillance System, CDC

Current Cigarette Smoking by Persons 18 and Over 1965-2000



Source: Health, United States, 2002. US Centers for Disease Control and Prevention. http://www.cdc.gov/nchs/hus.htm



Source: SEER Cancer Statistics Review. http://seer.cancer.gov/

EPID/CPH 573A

Week #1:

- Introduction
- Historical perspective
- "Descriptive epi": person, place, time
- Measures of frequency ("risk")

Measures of frequency (measures of "risk") (measures of "chance")

Suppose:

- Epi I class has 45 students of whom 16 will get an A
- What is the frequency ("risk", "chance") of getting an A?
- Count (16) is not good enough
- To calculate frequency/risk/chance, we need a denominator!

Measures of frequency (measures of "risk") (measures of "chance")

Count of interest

DENOMINATOR

Probability **Proportion**	16 (A)	**Odds**
16/45 = 0.36	29 (not A)	16/29 = 0.55
or 36%		or 1 to 1.8
or 36 per 100		or 5 to 9

Odds (formally defined)

Odds the probability of having an attribute
divided by the probability of not having that attribute

The probability of getting an A
Pr(A) = 16/45 = 0.36

The probability of not getting an A
Pr(not A) = 29/45 = 0.64

Odds of getting an A = 0.36 / 0.64 = 0.55 (=16/29)

Odds versus probability

Probability	Odds
Range: 0 to 1	Range: 0 to +infinity
Pr(not A) = 1 – Pr (A)	Odd(not A) = 1/odds(A)
29/45 = 1 – 16/45	29/16 = 1/(16/29)

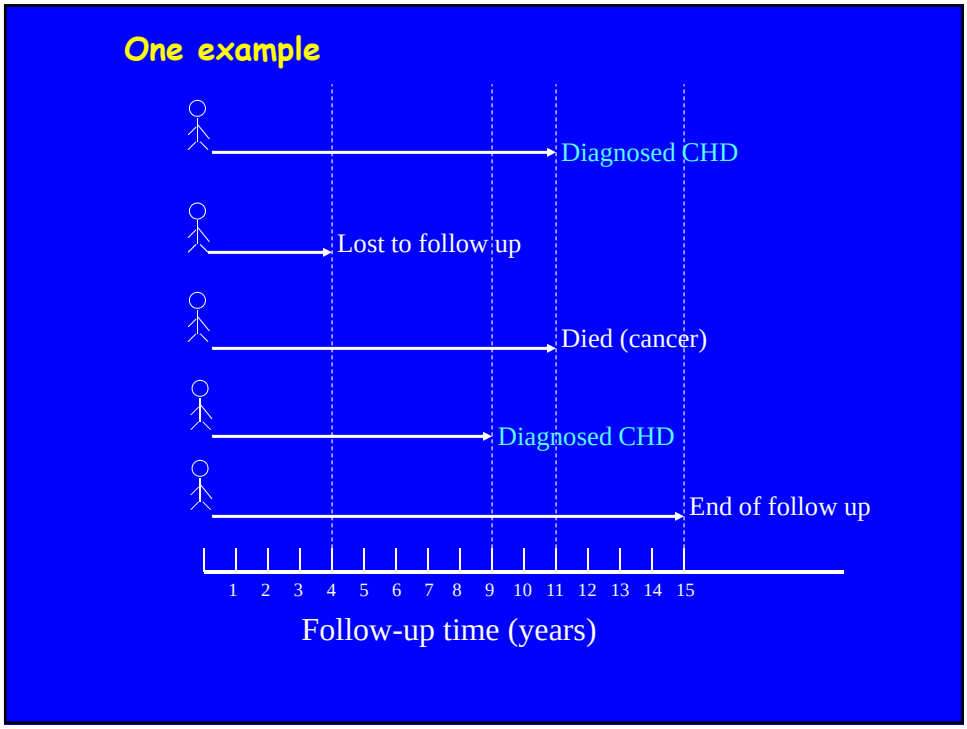
Some problems of probability and odds

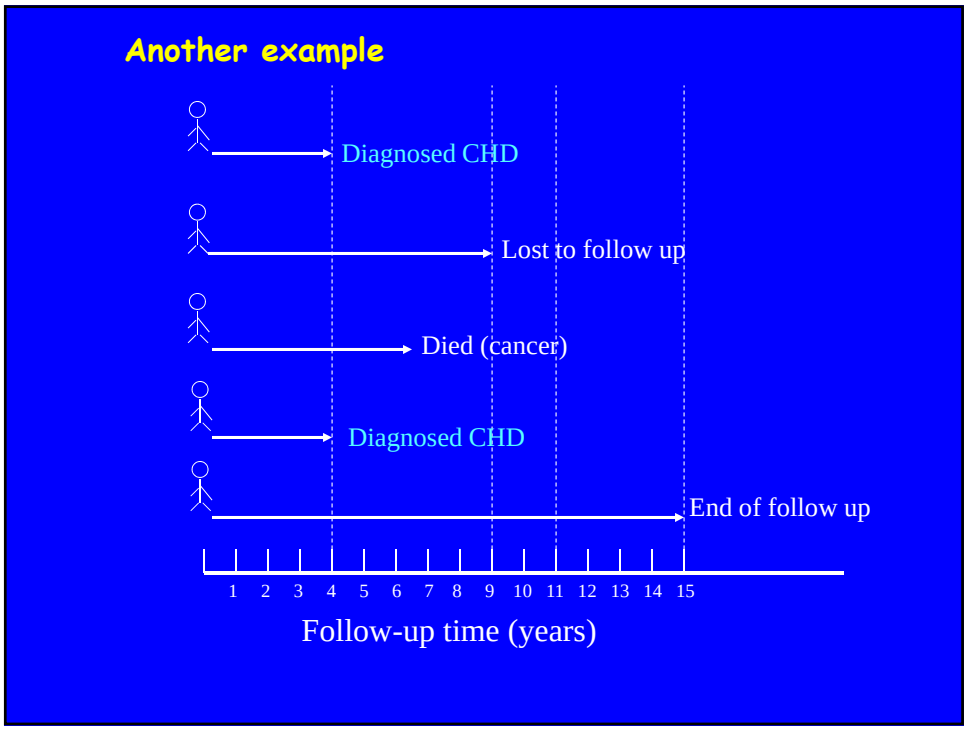
- To estimate the chances of developing coronary heart disease (CHD), a group of 5 people were observed for 15 years
- None had CHD at the beginning
- By the end of the study, CHD was diagnosed in 2 of these 5 people
- What was the frequency of CHD?

(or: what were the chances of developing CHD?)

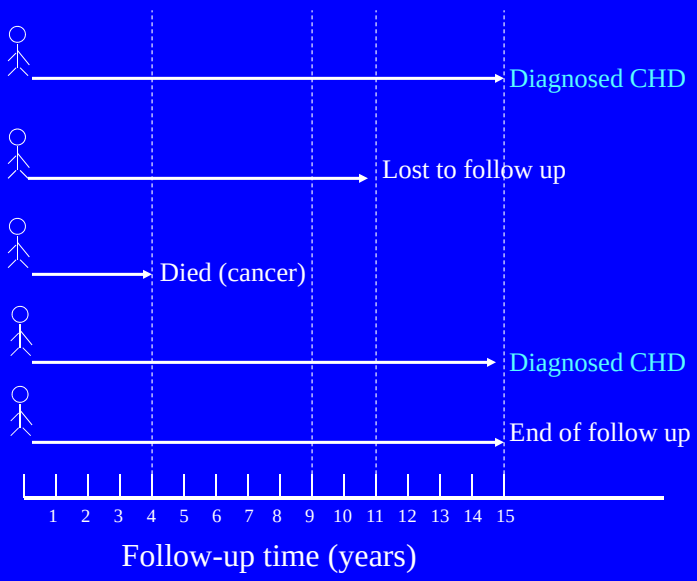
Probability and Odds

- Pr (CHD) = 2/5 = 0.4 (40%)
- Odds (CHD) = 2/3 = 0.67
- Let's examine more closely what has happened to these 5 people


One example
Diagnosed CHD
Lost to follow up
Died (cancer)
Diagnosed CHD
End of follow up
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Follow-up time (years)


Another example
Diagnosed CHD
Lost to follow up
Died (cancer)
Diagnosed CHD
End of follow up
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Follow-up time (years)

Another example



Some problems of probability and odds

- All three scenarios (and many more) are summarized by a single proportion (0.4) and a single odds (0.67)
- But they tell very different stories. Not every person was observed for 15 years.
- Each person "contributed" different time-at-risk of CHD
- How can we compute the frequency (chance) of CHD in this group, taking the time-at-risk into account?
- Answer: compute the rate of CHD

Measures of frequency

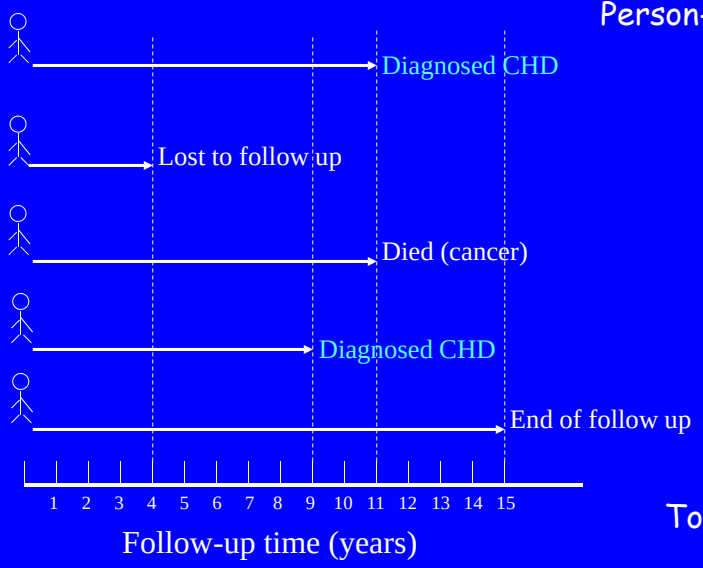
$$\frac{\text{Count of interest}}{\text{DENOMINATOR}}$$

$$\textbf{Rate} = \frac{\text{Count of interest}}{\text{Person-time at-risk}}$$

Rate, like speed, is measuring the velocity at which an event occurs.

Probability and odds are measuring cumulative events within arbitrary time intervals.

Let's look again at the first example



Rate of CHD

$$\frac{\text{Count of interest}}{\text{Person-time at-risk}} = \frac{2}{50 \text{ person-years}} = \frac{4}{100 \text{ person-years}}$$

Note:
We prefer to convert the denominator to some power of 10

We say: "4 cases of CHD per 100 person-years at-risk"

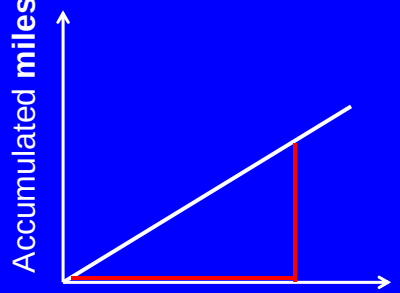
100 person-years could be generated in many ways:
1 person observed for 100 years
2 persons observed for 50 years each
100 persons observed for 1 year each

Velocity

- "rapidity of motion"
- "the speed at which something happens"

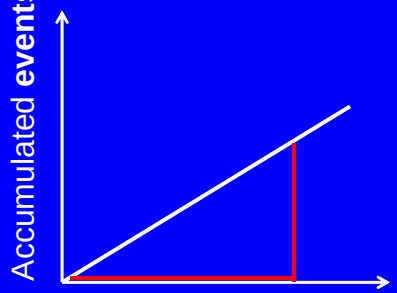
Rate

Velocity (speed)
of a car



Speed = Miles / Time

Velocity (speed)
of the occurrence of an
event



Rate = Events / person-time at risk