CDC
CENTERS FOR DISEASE CONTROL AND PREVENTION

Data and Disease Detectives:

Fundamentals of Biostatistics in Epidemiology
Governors' Honors Program, 1997
Valdosta, Georgia

When you can measure what you are speaking about, and express it in numbers, you know something about it. But when you cannot measure it, when you cannot express it in numbers,
your knowledge is of a meagre and unsatisfactory kind. --Lord Kelvin

Interpreting Risk

Think about events that may occur in your life:

- List one event that is certain.
- List one event that is impossible.
- List one event that is highly likely.
- List one event that is unlikely.

Which do you think caused more deaths in the United States last year, homicide or diabetes?

I offer you a game; which would you choose:

A: A gift of $240, guaranteed.

B: You have a 25% chance to win $1,000 and a 75% chance of winning nothing.

Faced with the following alternatives, which would you choose:

C: A sure loss of $740.

D: You have a 75% chance to lose $1,000 and a 25% chance to lose nothing.

Causality

In a 1993 study, scientists at the National Cancer Institute followed 48,000 men who had filled out dietary questionnaires in 1986. By 1990, 300 of these men had been diagnosed with prostate cancer and 126 had advanced cases. For the advanced cases, men who ate the most red meat
had a 164% higher risk than those with the lowest intake. Fats from dairy products, fish, and vegetable oils did not increase the risk. Is red meat a contributing cause of prostate cancer?

Legal sense: the immediate action (or lack thereof) producing an effect.

Epidemiologic sense:

- a factor that makes a difference;
- a factor that forms part of the circumstances that increase probability of occurrence of a disease;
- a condition with an essential role in producing occurrence of disease or injury.

Components:

- Strength of the association
- Temporality (does hypothesized cause precede effect)
- Biologic gradient (dose/response effect)
- Consistency of the evidence (do other studies agree)
- Plausibility
- Specificity of the association (occurs in every case of disease, occurs in no other disease)

Types of Epidemiologic Research

- Experimental
  - Laboratory trial
  - Clinical trial
  - Intervention
- Quasi-experimental
  - Clinical/laboratory
  - Program/policy
- Observational
  - Descriptive
  - Analytic

Role of Study Design in Establishing Causation: Cohort Study Approach

	Outcome	No Outcome
Exposed	a	b
Unexposed	c	d

Relative Risk = $\frac{\text{Incidence in exposed}}{\text{Incidence in unexposed}} = \frac{a/(a+b)}{c/(c+d)}$

	Outcome (Cases)	No Outcome (Controls)
Exposed	a	b
Unexposed	c	d

Odds Ratio = $\frac{\text{Odds of exposure in cases}}{\text{Odds of exposure in controls}} = \frac{a/c}{b/d} = \frac{ad}{bc}$

Measures of Association

- Relative Risk (RR): ratio of *rate* of disease among *exposed* to *rate* among *unexposed*
- Odds Ratio (RR): ratio of *odds* of exposure among *cases* to *odds* of exposure among *controls*.

Comparative Risk Estimates

0----------1----------> ∞

Risk, Probability, and Odds

A population contains 1,000 individuals; 400 of these carry the gene for a disease. For this population,

_____ % of all individuals carry the gene.

The proportion who carry the gene is _____.

The probability that someone carries the gene is _____.

The risk of carrying the gene is _____.

The odds of carrying the gene are _____.

Odds ratio

Compute the odds of disease for two groups (exposed and not exposed). Consider risk of developing breast cancer for women related to the age at which they had their first child (Pagano and Gauvreau, 1993):

First child at age 25 or older?	Breast Cancer	No Breast Cancer	Total
Yes	31 = a	1,597 = b	1,628
No	65 = c	4,475 = d	4,540
Total	96	6,072	6,168

Odds of br ca for women having first child at 25 or older = 31/1,597 = 0.0194

Odds of br ca for women having first child before age 25 = 65/4,475 = 0.0145

Odds ratio = 0.0194/0.0145 = 1.34. Shortcut formula: OR = bc/ad

Interpreting Risk: Missing Baseline Risk

Men who drank 500 ounces or more of beer a month (about 16 ounces a day) were three times
more likely to develop cancer of the rectum than nondrinkers. If you were a beer drinker reading about this study, would it encourage you to reduce your beer consumption?

Interpreting Risk: Risk over what time period?

"Italian scientists report that a diet rich in animal protein and fat increases a woman's risk of breast cancer threefold." -- (Prevention Magazine, 1991)

"One in nin e women in the US will get breast cancer." -- American Cancer Society

Does this mean US Women who eat a high-fat diet have a one-in-three chance of developing breast cancer?

Risk:	by age:	
1 in 50	50	
1 in 23	60	
1 in 13	70	
1 in 10	80	
1 IN 9	85	Lifetime risk

Testing a Hypothesis

- State the hypothesis and try to disprove it
  (Null hypothesis: no association)
- Collect data
- If the hypothesis is true, how likely is it that the data can be due to chance alone? (This is a p-value; requires special tables.)

P-Values

- Low $\rightarrow$ small chance that an association is found, if in truth there is no relationship. The evidence is strong enough that we are willing to rule out the possibility that chance alone produced the association.
- Medical analogy: false positives and false negatives
- Convention of 5%
- One of the factors determining statistical significance is sample size; even a trivial association may be significant if the sample is large enough.
- A significance test cannot tell us whether there is an association; it helps us decide whether to regard an observed association as nonfortuitous.

Data and Disease Detectives

A detective investigating a crime needs both tools and understanding. If he has no fingerprint powder, he will fail to find fingerprints on most surfaces. If he does not understand where the criminal is likely to have put his fingers, he will

not look in the right places.

--J. W. Tukey