Measures of Disease

Basic definitions:

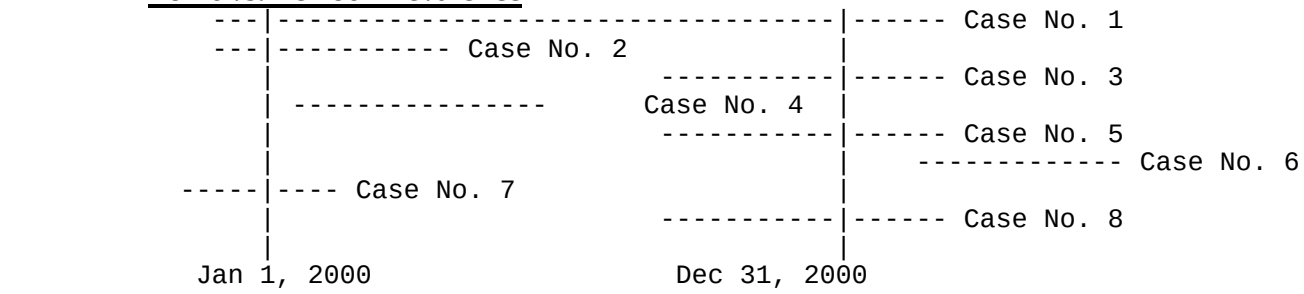
Incidence -measures the frequency of new disease occurrence in a defined population at risk observed for specific amount of time.

incidence cases = # of new cases in a given time period
incidence rate = (# of new cases in a given time period) / # at risk

Prevalence -measures the frequency of existing disease in a defined population at risk at a specified time

point prevalence - at a given point in time
period prevalence - during a given period

Point vs. Period Prevalence



Which cases would be counted for:
an annual incidence rate _____
a point prevalence at 1/1/00 _____
a point prevalence at 12/31/00 _____
period prevalence for 1/1-12/31/00_____

prevalence = incidence x (duration of disease)
Incidence is to prevalence as income is to wealth; production rate is to inventory

Incidence rate is direct estimate of the probability, or risk, of developing a disease during a specific time period. Whereas prevalence rate measures the number of cases that are present at, or during, a specified period of time.

Mortality

annual mortality rate = $\frac{\text{(\# deaths in one year)}}{\text{population}}$

case-fatality rate = $\frac{\text{(\# deaths due to specified disease during specified time)}}{\text{(\# of individuals with the disease)}}$

proportional mortality = $\frac{\text{(\# of deaths due to a specific disease in a year)}}{\text{(all deaths in that year)}}$

standardized mortality ratio = $\frac{\text{(observed \# of deaths per year)}}{\text{(expected \# of deaths per year)}}$

Incidence density (ID):

$$\frac{\textit{(No. of occurrences of new disease in a population observed over a specific amount of time)}}{\textit{(Total amount of person-time at risk experienced by the population during the time period)}}$$

(A). Interpretable as an instantaneous rate of new disease development over time, relative to the size of the population at risk

(B). Has units of time^{-1} and can range from 0 to 4

(C). Can be estimated from:

(1). Follow-up of a *fixed cohort*--i.e., a population at risk whose membership has no gains or losses during the period of surveillance for new cases. Denominator is calculated by summing the time at risk contributed by each person at risk, including that contributed by each case before development of disease

(2). Observation of a *dynamic population--i.e.*, one whose size and/or membership changes during the period of surveillance. Denominator is obtained by either:

(a). Summing up the amount of person-time contributed by each person while part of the population under surveillance, or

(b). Multiplying the average size of the population at risk by the duration of the surveillance period

(D). Examples:

(1). Incidence of colon cancer in Pima County during 2009 is approximately:

$$\frac{\textit{(No. of new colon cancer cases among Pima Co. residents during 2009)}}{\textit{(Average population of Pima Co. during the year, x 1 year at risk for each county resident)}}$$

(2). Incidence of colds among EPID573A students =

$$\frac{\textit{(No. of colds experienced by members of the class)}}{\textit{(Sum of each class member's duration of enrollment while cold-free)}}$$

(E). Commonly referred to as: incidence, **rate**, hazard rate, force of morbidity, instantaneous incidence

Cumulative incidence (CI):

(No. of persons who develop the disease during a specified time period)
(No. of persons initially at risk)

(A). Interpretable as the probability that a randomly chosen person from the population develops the disease during the time interval, assuming no loss to follow-up or changes in at-risk status

(B). A proportion: has no units and can range from 0 to 1

(C). Directly estimable only from observation of a fixed cohort. Everyone in the fixed cohort must either get the disease during the surveillance period or remain at risk for disease throughout the period

(D). Examples: cumulative incidence of disease among participants in a vaccine trial; cumulative incidence among persons involved in a short-term disease outbreak

(E). Commonly referred to as: incidence, **risk**, attack rate

Example of Disease X

```
                    Person Years
 1  - - - - - - - - - - - - - -     1.0
 2  - - - - - -X                    0.5
 3  - - - - - - - - - - - - - -     1.0
 4  - - - - - - - - - - - - - -     1.0
 5  - - -X                         0.25
 6  - - - - - - - - - - - - - -     1.0
 7  - - - - - - - - - - - - - -     1.0
 8  - - - - - - - - - - - - - -     1.0
 9  - - - - - - - - -X              0.75
10  - - - - - - - - - - - - - -     1.0
```

So for: CI use the number at risk at the beginning of the interval = _____
ID use person time at risk = _____ =_____cases/ person years

Mortality--includes fatal disease only

(i). Can be considered a special type of incidence, in which only fatal disease occurrences are counted in the numerator

(ii). Mortality density and cumulative mortality can be calculated as described above for corresponding incidence measures

(iii). Should not be confused with case fatality, which is usually defined as:

(No. of fatal cases of a disease)
(Total no. of cases of the disease)

which is basically a measure of disease prognosis, not of disease occurrence.

Prevalence: Frequency of existing disease at a particular point in time

$$\frac{\textit{(No. of existing cases of disease in a specified population at a particular point in time)}}{\textit{(No. of persons in the numerator + no. of persons in the population who are at risk for the disease but who do not have it at the specified time)}}$$

a. A "snapshot" of disease frequency, relative to the size of the population at risk. at a particular moment
b. Example: prevalence of colds among CPH/EPID 573a students on last day of class =

$$\frac{\textit{(No. of students with colds on that day)}}{\textit{(CPH/EPID 573a enrollment on that day)}}$$

c. The "point in time" need not be a particular moment in calendar time; may be another time scale, such as:
   (i). A person's lifetime: e.g., prevalence of a condition at a given age, regardless of when one attains that age in calendar time
   (ii). Time since an event of interest: e.g., prevalence of depression among widows/widowers 6 months after death of a spouse
d. Above description of prevalence concern what is sometimes termed point prevalence. *Period prevalence includes in the numerator both new and pre-existing cases that are active anytime during a specified time period. Period prevalence is a less useful measure because it mixes incidence and (point) prevalence.

Other disease frequency measures that use a denominator, but which are not based on the true population at risk

1. Commonly used when information about population size at risk is not available
2. Denominator is usually a "proxy" measure which, under certain assumptions, may be approximately proportional to the size of the population at risk
3. Examples (fill in blank below):
   a. _____:

$$\frac{\textit{(No. of deaths from a certain disease in a population during a specified time period)}}{\textit{(Total no. of deaths in the population from all causes during the same period)}}$$

   b. _____:

$$\frac{\textit{(No. of fetal deaths in a population during a specified period)}}{\textit{(No. of live births in the population during that period)}}$$

4. Note that variations in frequency of the denominator event across comparison populations can cause apparent, but spurious, differences in the ratio itself across those populations

Subgroup-specific measures

A. Both the numerator and the denominator (if any) are confined to persons with some shared characteristic(s): e.g., same sex, or in same age range
B. Example: age-specific incidence (density) of MI among persons aged 50-59 =

$$\frac{\textit{(No. of myocardial infarctions experienced by persons aged 50-59 during a certain time period)}}{\textit{(No. of person-years at risk contributed by persons aged 50-59 during the period)}}$$

Comparison of Two Measures of Disease Incidence

	Cumulative Incidence	Incidence Density
Numerator:	No. of newly diseased persons	No. of new disease occurrences
Denominator:	Size of population at risk at start of time period	Total amount of person-time at risk contributed during time period
Units:	None	$time^{-1}$
Range:	0-1	$0\text{-}\infty$
Interpretation:	Probability that a randomly chosen person, at risk will develop disease during the time interval	Instantaneous rate of disease development in relation to size of population at risk

Measures of Excess Risk

Computation of Excess Risk

1. Risk ratio (relative risk)

	Disease				Lung Cancer		
Exposure	+	-	Total		+	-	Total (1 yr)
+	a	b	a + b	Smokers	432	229,335	229,787
-	c	d	c + d	Nonsmokers	27	142,078	142,105

$$RR = \frac{a/(a+b)}{c/(c+d)} = \frac{I_e}{I_o} \qquad \text{RR} = \frac{432 / 229{,}787}{27 / 142{,}106} = 9.9$$

2. I_e = incidence in exposed = a/(a+b) = absolute risk

   I_o = incidence in nonexposed = c/(c+d)

   I = total incidence

   a. Attributable risk (AR) = incidence of disease due to exposure among exposed individuals.
   = $I_e - I_o$ = 432/229,787 - 27/142,105 in one year
   = 188/100,000 - 19/100,000 in one year
   = 169/100,000 in one year

   b. Attributable risk percent (AR%) = percentage of disease in exposed individuals due to exposure.
   $$= \frac{Ie - Io}{Ie} * 100 = \frac{188 - 19}{188} * 100 = 89.9\%$$

c. Population attributable risk (PAR) = risk of disease incidence in population as a whole due to exposure.
= $I - I_o$
Given an annual risk of lung cancer of 113 per 100,000 for all men in this age group:

PAR = 113/100,000 - 19/100,000 in one year
= 94/100,000 in one year

d. Population attributable risk percent (PAR%, etiologic fraction) = percent of disease incidence in population as a whole due to exposure.
$$= \frac{I - Io}{I} * 100 = \frac{113 - 19}{113} * 100 = 83\%$$

3. If instead of risks (cumulative incidence), the data are in the form of rates (instantaneous incidence), then the following terms apply (although they are not often used, even by those who know enough to do so):
   Rate ratio
   Rate difference
   Attributable rate
   Attributable rate percent
   Population attributable rate
   Population attributable rate percent

Measures of Excess Risk

Measure of risk	Magnitude depends on:	Helps to answer the questions:
Risk ratio (relative risk, rate ratio)	Strength of association	Does exposure (E) cause disease (D)?
Risk difference (attributable risk to the exposed)	a. Risk ratio b. Frequency of disease	(If it is inferred that E causes D) Among persons exposed to E, how much of D is E responsible for? Should anything be done to modify or eliminate E?
Attributable risk (%)	a. Risk ratio	(If it is inferred that E causes D) How likely is it that the occurrence of disease in an exposed individual was due to the exposure?
Attributable risk to the population	a. Risk ratio b. Frequency of exposure c. Frequency of disease	(If it is inferred that E causes D) What rate of D in the population is caused by E? Should resources be allocated to controlling E or, instead, to exposures causing greater health problems in the population?
Attributable risk to the population (%)	a. Risk ratio b. Frequency of exposure	(If it is inferred that E causes D) What portion of D in the population is caused by E? Should resources allocated to combating D be directed toward etiologic research or control of known etiologies (e.g., E)?
(Statistical significance)	a. Risk ratio b. Size of sample c. Frequency of disease (or exposure, depending on study design)	Could association between E and D have occurred by chance?

APPENDIX:
ESTIMATION OF RELATIVE RISK IN CASE-CONTROL STUDIES

Consider a prospective cohort study in which a large number of exposed and unexposed people are followed for a defined time period. with no losses to follow-up. The results can be expressed as follows:

		Disease	
		Yes	No
	Yes	a	b
Exposed	No	c	d

The (cumulative) incidence of disease in each group is:

$$CI_{exposed} = \frac{a}{a + b}$$
$$CI_{unexposed} = \frac{c}{c + d} .$$

From these data, the relative risk (*RR*) can be calculated directly as

$$\frac{CI_{exposed}}{CI_{unexposed}} = \frac{a/(a + b)}{c/(c + d)} .$$

If the disease is rare in both the exposed and unexposed groups, then $b>>a$, and $d>>c$, so that $a+b=b$, and $c+d=d$.

$$\frac{ad}{bc} = \frac{ad}{cb} = \frac{\frac{a}{c}}{\frac{b}{d}}$$

Thus:

$$RR = \frac{a/(a + b)}{c/(c + d)} \approx \frac{a/b}{c/d} = \frac{ad}{bc} .$$

Now note that -

the numerator of this last fraction, *a/c*, is the odds of exposure among people who develop the disease--i.e., among the cases. The denominator, *b/d*, is the odds of exposure among people who do not develop the disease-i.e., among the non-cases. The true value of each of these two factors can be estimated by studying only a representative sample of each respective group, since neither quantity depends on the relative frequency of cases and non-cases in the study population.

Suppose, for example, that only 1% of the non-cases were studied. We would expect the sample to contain about (.01) (*b*) exposed non-cases and (.01)(*d*) unexposed non-cases. The odds of exposure in this sample of non-cases would thus be (.01)(*b*)/(.01)(*d*) = *b/d*.

In a case-control study, exactly this strategy is followed: a representative sample of non-cases is studied to determine the frequency of exposure among them. This is then compared with exposure frequency data on all (or a random sample of) cases in the same population.

Example: Say that in a community of 200,000 persons, 50% have a certain characteristic X, and 50% do not. In a one-year period, 200 cases of disease Y occur in persons with X and 100 cases in persons without X:

		Disease Y		
		Yes	No	Total
	Yes	200	99,800	100,000
X	No	100	99,900	100,000
	Total	300	199,700	200,000

The relative risk is (200/100,000)/(100/100,000) = 2.0. (Note that the odds ratio for these data is (200)(99,900)/(99,800)(100) = 2.002, very close to 2.0.) Now suppose that 50% of the cases and 0.1% of the non-cases are chosen at the end of the year for a case-control study. The expected results would be:

		Disease Y		
		Yes	No	Total
	Yes	100	100	200
X	No	50	100	150
	Total	150	200	350

The odds ratio is (100)(100)/(100)(50) = 2.0.

Thus the odds ratio calculated on case-control data can be relied upon to provide a good estimate of relative risk if (a) the disease is rare among both exposed and unexposed individuals, (b) exposure frequency among the cases studied is representative of exposure frequency among all cases in the population of interest, and (c) exposure frequency among the controls studied is representative of exposure frequency among all non-cases in the population of interest.

Under certain circumstances, the odds ratio from case-control data provides a good estimate of the incidence density ratio (= rate ratio) even if assumption (a) in the previous paragraph is not satisfied. For details, see the references cited below.

1. Rothman KJ. Modern epidemiology. Boston: Little, Brown, 1986, pp. 62-64.

2. Greenland S, Thomas DC. On the need for the rare disease assumption in case-control studies. Am J Epidemiol 1982; 116:547-553.

REFERENCES from recommended books:

Causal Inference:

Lilienfeld &Stolley 1994 - Chapter 12, and pages 279-83
Kelsey 1986 - pg 31-3
Rothman 1986 - Chapter 2

Excess Risk:

Lilienfeld &Stolley 1994 - pg 200-2.
Kelsey 1986 - pg 36-41
Rothman 1986 - Chapter 4
Mausner & Bahn - Chapter

Bias

Various Definitions

Deviation of results or inferences from the truth or processes leading to such deviation.

Any trend in the collection, analysis, interpretation, publication, or review of data that can lead to conclusions that are systematically different from the truth.

Deviation of inferences from the truth as a result of flaws in study design, data collection, or the analysis or interpretation of results.

A tendency of procedures (in study design, data collection, analysis, interpretation, review or publication) to yield results or conclusions that depart from the truth

Forms of bias (not all inclusive)

- Selection bias
- Information bias
  - interviewer bias,
  - observer bias,
  - reporting bias ,
  - recall bias,
  - response bias (recall & other informant bias)
- Confounding
- Other
  - screening,
  - publication bias

Criteria of Causation

Hills Criteria of Causation outlines the minimal conditions needed to establish a causal relationship between two items. These criteria were originally presented by Austin Bradford Hill (1897-1991), a British medical statistician, as a way of determining the causal link between a specific factor (e.g., cigarette smoking) and a disease (such as emphysema or lung cancer). *Hill's Criteria* form the basis of modern epidemiological research, which attempts to establish scientifically valid causal connections between potential disease agents and the many diseases that afflict humankind.

1. **Temporal Relationship**: Exposure always precedes the outcome
2. **Strength**: the size of the association
3. **Dose-Response Relationship**: An increasing amount of exposure increases the risk.
4. **Consistency**: The association is consistent when results are replicated in studies in different settings using different methods.
5. **Plausibility**: The association agrees with currently accepted understanding of pathological processes.
6. **Coherence**: The association should be compatible with existing theory and knowledge.
7. **Consideration of Alternate Explanations**
8. **Specificity**: This is established when a single putative cause produces a specific effect
9. **Experimental evidence**: rarely available for human conditions

Austin Bradford Hill, "The Environment and Disease: Association or Causation?" Proceedings of the Royal Society of Medicine, 1965; 58: 295-300.