Harvard Pilgrim Health Care Institute

# CLINICAL EPIDEMIOLOGY AND POPULATION HEALTH

## Key Points – Frequency and Association

**Measures of frequency**

- Prevalence- how many people have an outcome (or risk factor) of interest
- Incidence- how many people develop an outcome of interest

**Measures of association**

- Absolute difference between groups
- Relative risk: proportion with an outcome in one group divided by proportion with the outcome in another. (The details of the calculation vary by study design.)

2 x 2 table is a standard representation of data with a dichotomous outcome:

| | Outcome | | |
|---|---|---|---|
| | **Yes** | **No** | |
| **Exposure Y** | a | b | a + b (all exposed) |
| **Exposure N** | c | d | c + d (all unexposed) |
| | a+c Total with outcome | b+d Total no outcome | a+b+c+d Entire population |

*Note: We typically set up a 2 x2 table with the "Disease" or "Outcome" in the columns and "Exposure" in the rows. Exposed=Y is the first row, and outcome=Y is the first column.*

**Cross-sectional studies: measures of frequency and association**

**Prevalence** (a snapshot in time) = # of persons with a particular attribute or outcome ("cases") / total # of persons in the population

**The measure of "relative risk" in a cross-sectional study is actually a prevalence ratio**

= prevalence among those with an exposure (attribute) / prevalence among those without an exposure (attribute)

**Cohort studies** Observational studies in which we follow a defined group of individuals:

- Prospective cohort study- individuals enrolled then followed forward in time (best)
- Retrospective cohort study- investigators go back in time to assemble the cohort or use data already collected.

Outcomes may be continuous variables (e.g. systolic blood pressure) but are often categorical (disease onset, death, etc.).

Follow-up time may be fixed (all individuals followed for 5 years) or variable (each individual contributes some amount of person-time to the denominator).

- **Cumulative incidence** = # of new events / total # persons at risk over a specified time period. (It is a unitless fraction, but the observation period is described (e.g. “the annual incidence of pancreatic cancer among men 40-50 years old”).

- **Incidence rate** (also called **Incidence Density**) = # of new events/ sum of total person-time at risk *(*e.g. 20 cases per 1000 person-years)

    - (Note: The denominator of an incidence rate is a combination of the number of people and the amount of time i.e. the sum of person-time contributed by each member of a population observed. For example, a person-year is one person followed for one year, or one person followed for 3 months together with another followed for 9 months, or 52 people each followed for a week, etc…)

Difference between cumulative incidence and incidence rate:
*Cumulative incidence* is usually used when all people are followed for an **equal time period** to see if they have an event, *incidence rate* allows for **different lengths of follow-up**.

**Measures of Association (Effect):** Differences between groups can be measured by:

- Relative Measures (e.g. smokers have 15 times the rate of lung CA as non-smokers) include risk ratios (a.k.a. relative risk) and odds ratios.

- Absolute Measures (e.g. difference of 3 cm in height; or, difference in mortality between 48% and 35%, for an absolute difference of 13%). Absolute measures provide the maximum information.

**Risk Ratio** (relative risk)
The risk ratio is a commonly used measure – it represents the risk in the exposed group compared with the risk (or odds) in the unexposed group

$$\text{Risk ratio (RR)} = \frac{\text{risk of disease in the exposed group}}{\text{risk of disease in the unexposed group}} = \frac{a/(a+b)}{c/(c+d)}$$

*In general, RR’s are reported for cohort studies.*

**Attributable risk (AR)** aka **Risk Difference (RD)** describes the increased risk of disease that results from the exposure (among the exposed). Even in the exposed group, not all disease results from the exposure of interest; some disease is the result of other causes.

AR = Risk of disease in the exposed – Risk of disease in the unexposed
= a/(a+b) - c/(c+d)

**Attributable fraction** (AF may also be expressed as AR%) is the proportion of disease in the exposed group that is due to the exposure.

$$AF = \frac{\text{Risk of disease in exposed – Risk of disease in the unexposed}}{\text{Risk of disease in exposed}}$$

$$= \frac{a/(a+b) - c/(c+d)}{a/(a+b)}$$

**Population attributable risk (PAR):** the risk of excess disease in the population due to the exposure.

PAR = Risk of disease in the total population – Risk of disease in the unexposed group
(PAR can also be calculated by: Attributable risk * Prevalence of the risk factor)

You can also calculate the proportion of risk in the entire population that is attributable to exposure. It is calculated as:
PAF (population attributable fraction) = PAR/risk of disease in total population
PAR% (population attributable risk percent) = PAFx100

**Odds, Odds Ratios, and Case Control Studies**

Any proportion (the number of times an event occurs divided by the number of opportunities) can also be expressed as odds (event occurrences divided by non-occurrences). A 50% chance of rain is also 1:1 odds.

If p= the proportion, odds = p/(1-p)

If o= the odds, p= o/(1+o)

Sometimes it's useful to use the odds ratio (OR) as a relative measure of effect (instead of a risk ratio):

$$\textbf{Odds ratio (OR)} = \frac{\text{odds of disease in the exposed group}}{\text{odds of disease in the unexposed group}} = \frac{a/b}{c/d} = \frac{ad}{bc}$$

Odds ratios are close approximations of risk ratios when the outcome is rare (a frequently used rule of thumb is <10%). When the outcome is more common and/or the association is stronger (RR further from 1.0) the OR will be a less good approximation of the RR. In these cases, the OR will be more extreme (higher than the RR if they are > 1.0 and lower than the RR if they are < 1.0).

Common outcome

| | Outcome + | Outcome - | Total |
|---|---|---|---|
| Exposure + | 95 | 5 | 100 |
| Exposure - | 90 | 10 | 100 |
| Total | 185 | 15 | 200 |

Risk in exposed = 95/(5+95) = 0.95 Odds in exposed = (95/5) = 19:1
Risk in unexposed = 90/(90+10) = 0.90 Odds in unexposed = (90/10) = 9:1
RR (exposed vs. unexp)=0.95/0.90 = 1.1 OR (exposed vs. unexp)= 19/9 =2.1

Rare outcome

| | Outcome + | Outcome - | Total |
|---|---|---|---|
| Exposure + | 8 | 92 | 100 |
| Exposure - | 7 | 93 | 100 |
| Total | 15 | 185 | 200 |

Risk in exposed = 8/(100) = 0.08 Odds in exposed = (8/92) = 0.09:1
Risk in unexposed = 7/(100) = 0.07 Odds in unexposed = (7/93) =0.08:1
RR (exposed vs unexp)= 0.08/.07 = 1.1 OR (exposed vs unexp) = .11/.053 = 1.1 (rounded)

*Note that in case-control studies, it is not possible to calculate risk ratios directly; OR's must be calculated as:*

$$\text{Odds ratio (OR)} = \frac{\text{odds of exposure among those with disease}}{\text{odds of exposure among those without disease}} = \frac{a/c}{b/d} = \frac{ad}{bc}$$