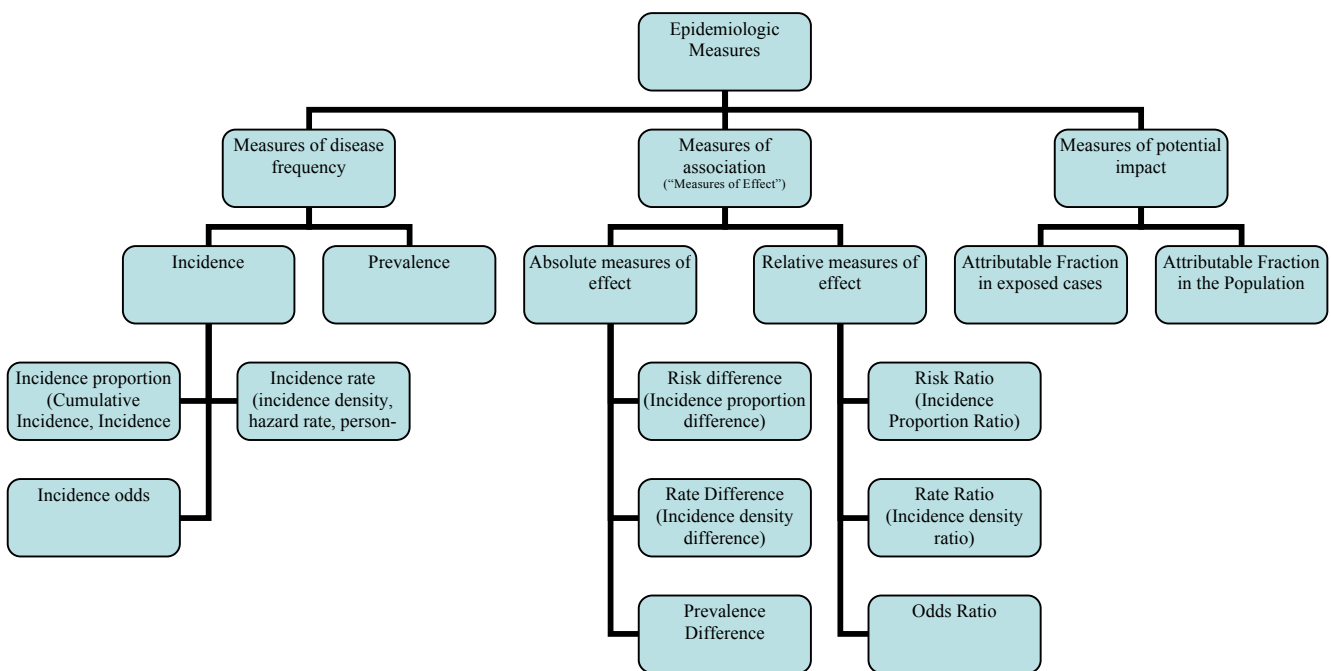


# FORMULAS FROM EPIDEMIOLOGY KEPT SIMPLE (3e)

## Chapter 3: Epidemiologic Measures

Basic epidemiologic measures used to quantify:

- frequency of occurrence
- the effect of an exposure
- the potential impact of an intervention.



### 3.1 Measures of Disease Frequency

$$\text{Incidence Proportion} = \frac{\text{No. of onsets}}{\text{No. at risk at beginning of follow-up}}$$

- Also called **risk**, **average risk**, and **cumulative incidence**.
- Can be measured in cohorts (closed populations) only.
- Requires follow-up of individuals.

$$\text{Incidence Rate} = \frac{\text{No. of onsets}}{\sum \text{person-time}}$$

- Also called **incidence density** and **average hazard**.
- When disease is rare (incidence proportion < 5%), incidence rate  $\approx$  incidence proportion.
- In cohorts (closed populations), it is best to sum individual person-time longitudinally. It can also be estimated as  $\sum \text{person-time} \approx (\text{average population size}) \times (\text{duration of follow-up})$ . Actuarial adjustments may be needed when the disease outcome is not rare.
- In an open populations,  $\sum \text{person-time} \approx (\text{average population size}) \times (\text{duration of follow-up})$ .  
Examples of incidence rates in open populations include:

$$\text{Crude birth rate (per } m) = \frac{\text{births}}{\text{mid-year population size}} \times m$$

$$\text{Crude mortality rate (per } m) = \frac{\text{deaths}}{\text{mid-year population size}} \times m$$

$$\text{Infant mortality rate (per } m) = \frac{\text{deaths } < 1 \text{ year of age}}{\text{live births}} \times m$$

$$\text{Prevalence Proportion} = \frac{\text{No. of cases}}{\text{No. of individuals in the study}}$$

- Also called **point prevalence** or just **prevalence**.
- The concept of **period prevalence** should be avoided when possible because it confuses the concepts of incidence and prevalence (Elandt-Johnson & Johnson, 1980).
- Prevalence dependence on the “inflow” and “outflow” of disease according to this formula  
Prevalence  $\approx$  (incidence rate)  $\times$  (average duration of illness).

#### Additional Notes

- **Terminology:** The term “rate” is often used loosely, to refer to any of the above measures of disease frequency (even though the only true rate is the incidence density rate)
- **Odds:** Both prevalence and incidence proportions may be addressed in terms of odds. Let  $p$  represent the incidence proportion or prevalence proportion of disease and  $o$  represent the odds of disease. Thus, odds  $o = p / (1 - p)$ .
- **Reporting:** To report a risk or rate “per  $m$ ,” simply multiply it by  $m$ . For example, an incidence proportion of  $0.0010 = 0.0010 \times 10,000 = 10$  per 10,000.
- **Uni-cohort:** To report a risk or rate as a unihort, take its reciprocal and report it as 1 in “unihort.” For example, an incidence proportion of  $0.0025 = 1$  in  $\frac{1}{0.0025}$  or “1 in 400.”

### 3.2 Measures of Association (Measures of Effect)

**Notation and terminology:** Concepts apply to incidence proportions, incidence rates, and prevalence proportions, all of which will be loosely called “rates.” Let  $R_1$  represent the rate or risk of disease in the exposed group and let  $R_0$  represent the rate or risk of disease in the non-exposed group.

#### Absolute Measure of Effect (Rate Difference)

$$RD = R_1 - R_0$$

#### Relative Measure of Effect (Rate Ratio)

$$RR = \frac{R_1}{R_0}$$

The relative effect of an exposure can also be captured by the SMR (see section on Rate Adjustment)

#### 2-by-2 Cross-Tabulation

	D+	D-	Total
E+ (Group 1)	$A_1$	$B_1$	$N_1$
E- (Group 0)	$A_0$	$B_0$	$N_0$
	$M_1$	$M_0$	$N$

- For **person-time data** (incidence rates/densities) ignore cells  $B_1$  and  $B_0$  and let  $N_1$  and  $N_0$  represent the person-time in group 1 and group 0, respectively.
- **Rates, Rate Ratio, and Rate Difference:**  $R_1 = \frac{A_1}{N_1}$ ,  $R_0 = \frac{A_0}{N_0}$ ,  $RR = \frac{A_1 / N_1}{A_0 / N_0}$ , and  $RD = (A_1 / N_1) - (A_0 / N_0)$  (cohort and cross-sectional data)
- **Odds ratio:**  $OR = \frac{A_1 B_0}{A_0 B_1}$  (independent samples only; for matched-pairs and tuples data, see text)
- **Rounding:** Basic measures should be reported with 2 or 3 significant digit accuracy. Carry 4 or 5 significant digits to derive a final answer that is accurate to 2 or 3 significant digits, respectively.

### 3.3 Measures of Potential Impact

- The **attributable fraction in exposed cases**  $AF_e = \frac{R_1 - R_0}{R_1}$ , or equivalently,  $AF_e = \frac{RR - 1}{RR}$ .
- The **attributable fraction in the population**  $AF_p = \frac{R - R_0}{R}$   
Equivalent,  $AF_p = AF_e \times p_c$  where  $p_c$  represents proportion of population cases that are exposed

### 3.4 Rate Adjustment (“Standardization”)

For uniformity of language, the term rate will be used to refer to any incidence or prevalence measure.

#### Direct Standardization

The directly adjusted rate ( $aR_{direct}$ ) is a weighted average of strata-specific rates with weights derived from a reference population:

$$aR_{direct} = \sum w_i r_i$$

where  $w_i = \frac{N_i}{\sum N_i}$

$N_i$  represents the size of strata  $i$  of the reference population

$r_i$  represents rate in strata  $i$  of the study population.

Note that capital letters denote values that come from the reference population and lower case letters denote values that come from the study population.

#### Indirect Standardization

Indirect standardization is based on the **Standardized Mortality Ratio (SMR)**

$$SMR = \frac{\text{Observed}}{\text{Expected}}$$

where “Observed” is the observed number of cases and

“Expected” is the expected number of cases in the population based on this formula:

Expected =  $\sum R_i n_i$  where  $R_i$  represents the rate in strata  $i$  of the reference population and  $n_i$  represents the number of people strata  $i$  of the study population.

The Expected in the population can be understood in terms of the **expected number of cases within strata  $i$** , which is:  $\text{Expected}_i = R_i n_i$ . Thus:  $\text{Expected} = \sum \text{Expected}_i$ .

The SMR is a population-based relative risk estimate in which “1” represents a population in which the observed rate equals the expected rate.

*Optional:* Use the SMR to derive the indirectly adjusted rate via this formula:

$$aR_{indirect} = (\text{crude rate}) \times SMR$$

# Chapter 10: Screening for Disease

## Reproducibility (Agreement)

		Rater B		
		+	-	
Rater A	+	<i>a</i>	<i>b</i>	<i>g</i> <sub>1</sub>
	-	<i>c</i>	<i>d</i>	<i>g</i> <sub>2</sub>
		<i>f</i> <sub>1</sub>	<i>f</i> <sub>2</sub>	<i>N</i>

$$p_{\text{obs}} = \frac{a + d}{N}$$

$$p_{\text{exp}} = \frac{f_1 g_1 + f_2 g_2}{N^2}$$

$$\kappa = \frac{p_{\text{obs}} - p_{\text{exp}}}{1 - p_{\text{exp}}}$$

## Validity (Sensitivity, Specificity, PVPT, PVNT)

	Disease +	Disease -	Total
Test +	TP	FP	TP + FP (those who test positive)
Test -	FN	TN	FN + TN (those who test negative)
Total	TP + FN (those with disease)	FP + TN (those w/out disease)	<i>N</i>

$$\text{SEN} = \frac{\text{TP}}{\text{(those with disease)}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad [\text{note: TP} = (\text{SEN})(\text{TP} + \text{FN})]$$

$$\text{SPEC} = \frac{\text{TN}}{\text{(those without disease)}} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad [\text{note: TN} = (\text{SPEC})(\text{FP} + \text{TN})]$$

$$\text{PVP} = \frac{\text{TP}}{\text{(those who test positive)}} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{PVN} = \frac{\text{TN}}{\text{(those who test negative)}} = \frac{\text{TN}}{\text{TN} + \text{FN}}$$

$$\text{True prevalence} = \frac{\text{TP} + \text{FN}}{N} \quad [\text{also known as "prior probability"}]$$

Bayesian equivalents for PVP and PVN are presented in the text.



# TEN COMMANDMENTS FOR DEALING WITH CONFOUNDING



Source: EPIB-601 McGill University, Montreal, Canada madhukar.pai@mcgill.ca,  
<http://www.teachepi.org/documents/courses/Ten%20Commandments%20for%20Dealing%20with%20Confounding.pdf>

- I. Always worry about confounding in your research, especially at the design/protocol stage. Try to use design elements (e.g. randomization) that will help reduce potential confounding.
- II. Prior to the study, review the literature and consider the underlying causal mechanisms (e.g. draw causal diagrams such as directed acyclic graphs [DAGs]). Then make sure you collect data on all potential confounders; otherwise you will not be able to adjust for them in your analyses.
- III. Know your field or collaborate with an expert who does! Subject-matter knowledge is important to recognize (e.g. draw causal diagrams) and adjust for confounding.
- IV. Use a priori and data-based methods to check if the potential confounders are indeed confounders that should be adjusted for.
- V. Use stratified analyses and multivariable methods to handle confounding at the analysis stage. Choose the multivariate model that best suits the type of data (e.g. dichotomous vs. continuous) you collected and the design you employed (e.g. case-control vs. cohort).
- VI. Do not adjust for covariates that may be intermediate causes (on the causal pathway between the exposure and disease). Do not adjust for covariates that may not be genuine confounders. And beware of time-varying covariates that will need special approaches.
- VII. Use matching with great caution. Use analytic methods that are appropriate for the design used; for example, if matching was done, use methods that take matching into account (e.g. conditional logistic regression, matched pairs analyses).
- VIII. Always consider effect measure modification, but perform and interpret subgroup analyses with caution. The subgroup analysis should be one of a small number of hypotheses tested, and the hypothesis should precede rather than follow the analysis (i.e. subgroups must be pre-specified).
- IX. Always remember that adjustment for confounding can be inadequate due to residual confounding because of unmeasured confounders, misclassification of confounders, and inadequate adjustment procedures (e.g. model misspecification, categorization of continuous covariates).
- X. If conventional methods prove to be inadequate, consider using newer approaches such as propensity scores, matched sampling, instrumental variables and marginal

structural models. However, make sure you work with statisticians who understand these new methods (not many do).

*When all else fails, pray! If prayer fails, consider changing professions!!*