Basic Epi Formulas

Measures of Disease Frequency

Incidence Proportion = $\dfrac{\text{onsets}}{\text{no. at risk at beginning of study}}$

- To convert a rate or proportion to "per *m* people," simply multiplying by *m.* For example, an incidence rate of 0.00877 per person-year = 0.008770 × 100,000 = 877 per 100,000 person-years.
- To report a risk or a rate as a unicohort (number of individuals expected to produce one case), take its reciprocal and report the rate as 1 per "unicohort."

Incidence Rate = $\dfrac{\text{onsets}}{\sum \text{person-time}}$

- In a cohort, Σperson-time can be estimated by summing individual person-time or by summing person-time within strata with or without an actuarial adjustment.
- In an open population, Σ person-time ≈ (average population size) × (duration of follow-up). Examples of open population rates follow:

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

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

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

Prevalence = $\dfrac{\text{cases}}{\text{No. in population}}$

- Prevalence ≈ (incidence rate) × (average duration of illness)

Measures of Association / Effect

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.

$RD = R_1 - R_0$ (Excess risk in absolute terms)

$RR = \dfrac{R_1}{R_0}$ (Excess risk in relative terms)

$SMR = \dfrac{\text{Observed}}{\text{Expected}}$ (Excess risk in relative terms)

Measures of Potential Impact

$$AF_e = \frac{RR - 1}{RR}$$

$$AF_p = \frac{R - R_0}{R}$$

$$PF_p = \frac{R_0 - R}{R_0} \qquad \text{Equivalently } PF_p = \frac{p_c(1 - RR)}{p_c(1 - RR) + RR}$$

2-by-2 Table (Two Independent Samples)

	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, 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. For **cohort** and **cross-sectional** data, $R_1 = \dfrac{A_1}{N_1}$ and $R_0 = \dfrac{A_0}{N_0}$. Then apply measures of effect and impact, as listed above.
- For **case-control data**, use $OR = \dfrac{A_1 B_0}{A_0 B_1}$ and ignore formulations based on rates and risks.

Matched Pair Case-Control Data

		Case pair-member	
Control pair-member		**Exposed**	**Nonexposed**
	Exposed	*t*	*u*
	Nonexposed	*v*	*w*

$$\widehat{OR} = \frac{v}{u}$$

For matched tuples, use WinPEPI's "PairsEtc" program E. 'Yes-no' variable: compare subjects with 2 or more controls.

Chapter 10: Screening for Disease

Inter-rater agreement (reproducibility)

		Rater B: +	Rater B: −	
Rater A	+	*a*	*b*	g_1
	−	*c*	*d*	g_2
		f_1	f_2	*N*

$$p_o = \frac{a+d}{N} \qquad p_e = \frac{f_1 g_1 + f_2 g_2}{N^2} \qquad \kappa = \frac{p_o - p_e}{1 - p_e}$$

See Table 4.4 in the 2nd edition (p. 82) for interpretation guidelines.

Validity statistics

	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)	*N*

SEN = (TP) / (those with disease) [note: TP = (SEN)(TP + FN)]
= (TP) / (TP + FN)

SPEC = (TN) / (those without disease) [note: TN = (SPEC)(FP + TN)]
= (TN) / (TN + FP)

PVP = (TP) / (those who test positive)
= (TP) / (TP + FP)

PVN = (TN) / (those who test negative)
= (TN) / (TN + FN)

True prevalence = (TP + FN) / N [True prevalence also known as *prior probability*]

Bayesian equivalents for determining predictive value based on prior probabilities and test parameters are available 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!!