Epidemiologic Formulas and Terminology

Epidemiologic **terminology** is far from uniform. The chart clarifies some of the terms used in the course.

Measure	Synonyms (or nearly so)	Comment
Prevalence	Prevalence "rate" (misnomer)	Proportion of people with disease at a point in time.
Risk	Cumulative Incidence Incidence Proportion Probability of Disease	Number of disease onsets divided by the number of people exposed to risk.
Rate	Incidence Density Incidence Rate Central Rate Hazard Rate Force of morbidity / mortality	Number of disease onsets divided by sum of person-time.
Risk (or Rate) Ratio	Relative Risk Incidence Ratio Cumulative Incidence Ratio Incidence Density Ratio Hazard Ratio	*Ratio* of two risks or rates. Provides a relative measure of the effect of the exposure.
Risk (or Rate) Difference	Cumulative Incidence Difference Incidence Density Difference	*Difference* of two risks or rates. Provides an absolute measure of the effect of the exposure.
Attributable Fraction in the Population	Etiologic Fraction, Population	Expected % reduction in number cases following elimination of the exposure in population.
Attributable Fraction in Exposed Cases	Etiologic Fraction, Exposed Cases	Expected % reduction in number cases following elimination of the exposure in exposed case.
Odds ratio	Exposure odds ratio	Use primarily restricted to case-control studies. Also used in logistic models. Provides an estimate of the rate ratio.

Chapter 6: Incidence and Prevalence

Basics

$$\text{Prevalence} = \frac{\text{no. of existing cases on a specific date}}{\text{no. of people in the population on this date}}$$

$$\text{Risk} = \text{Cumulative Incidence} = \frac{\text{no. of disease onsets}}{\text{size of population initially exposed to risk}}$$

$$\text{Rate} = \text{Incidence density} = \frac{\text{no. of disease onsets}}{\text{sum of person-time}} \cong \frac{\text{no. of disease onsets}}{\overline{N} \cdot \Delta t}$$

where $\overline{N}$ represents the average ("central") population at risk and Δt represents the time of observations (e.g., a one-year study).

Examples of Specific "Rates"

$$\text{Birth rate per } m = \frac{\text{no. of births}}{\text{average population size}} \times m$$

where m is a population multiplier (e.g., per 1000 individuals).

$$\text{Crude death rate per } m = \frac{\text{no. of deaths}}{\text{avg. population size}} \times m$$

$$\text{Infant mortality rate per } m = \frac{\text{no. of deaths} < 1 \text{ yr of age}}{\text{no. of live births}} \times m$$

$$\text{Age-specific death rate per } m = \frac{\text{no. of deaths in age group}}{\text{no. of people in age group}} \times m$$

Chapter 7: Adjusted Rates

Notation: Capital letters (e.g., N_i) denote values from the reference population. Small letters (e.g., n_i) denote values from the study population.

Direct Adjustment

$$aR(direct) = \frac{\sum_{i=1}^{k} N_i r_i}{\sum_{i=1}^{k} N_i}$$

where $aR(direct)$ represents the directly adjusted rate, N_i represents the number of people in strata i in the reference population (there are k strata), and r_i represents the rate of disease in strata i of the study population.

Indirect Adjustment

Expected number of cases (μ)

$$\mu = \sum_{i=1}^{k} R_i n_i$$

where R_i represents the rate in the i^{th} stratum of the reference population and n_i represents the number of people in the i^{th} strata of the study population. The product $R_i n_i$ is the expected number of cases in the i^{th} stratum of the study population (μ_i).

Standardized Mortality Ratio

$$SMR = \frac{x}{\mu}$$

where x represents the observed number of cases in the study population and μ represents the expected number of cases as calculated above.

Indirectly Adjusted Rate

$$aR(indirect) = (cR)(SMR)$$

Chapter 8: Measures of Association

Two groups are considered: an exposed group (group 1) and an non-exposed group (group 0). The exposure may represent any modifiable or non-modifiable trait, intervention, characteristic, or environmental factor. Let:

R_1 represent the risk or rate of disease in exposed group
R_0 represent the risk or rate of disease in the non-exposed group
R represent the risk or rate of disease in the population as a whole (exposed + non-exposed group)

The **risk (or rate) difference** is the difference in the two risks (or rates):

$$RD = R_1 - R_0$$

For example, if the risk in the exposed group is 2 per 1000 and the risk in the non-exposed group is 1 per 1000, the risk difference = 1 per 1000.

The risk difference quantifies the effect of the exposure in absolute terms, i.e., the excess number of cases per *m* exposures. The risk difference may be positive (for a risk factor) or negative (for a protective factor).

The **risk (or rate) ratio** is the ratio of the two risks (or rates)

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

For example, if the risk in the exposed group is 2 per 1000 and the risk in the non-exposed group is 1 per 1000, then the risk ratio is 2.

The relative risk quantifies the effect of the exposure in relative terms, i.e., the relative strength of the effect. Notice that when $R_1 = R_0$, $RR = 1$, indicating no association between the exposure and disease. Relative risks greater than 1 indicate a positive association, and relative risks less than 1 indicate a negative association. The risk ratio is a risk multiplier (e.g., a RR of 2 indicates that the exposed group is at twice the risk of the non-exposed group).

If we define the *relative risk difference* (RRD) as the absolute effect (i.e., risk difference) compared to baseline risk, then $RRD = \frac{R_1 - R_0}{R_0} = \frac{R_1}{R_0} - \frac{R_0}{R_0} = \frac{R_1}{R_0} - 1 = RR - 1$. That is, the relative risk difference = $RR - 1$. By subtracting 1 from the risk ratio, we are left with its segment that is above baseline. This allows us to say that a relative risk of 2 suggests that the exposure *increases* risk by 2 – 1 = 1(100%) = 100%.(It would *not* be correct to say risk is 200% greater, since this would imply a risk ratio of 3!)

The **attributable fraction in the population** is:

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

This quantifies the expected proportional reduction in risk if the exposure were eliminated from the population. For example, a population attributable fraction of 50% suggests that eliminating the exposure from the population would eliminate half the cases.

The **attributable fraction in exposed cases** is:

$$AF_e = \frac{R_1 - R_0}{R_1}$$

This quantifies the expected proportional reduction in cases had the exposed cases not been exposed (it is a liability measure). For example, an attributable fraction in exposed cases of 75% would suggest that three-quarters of the exposed cases would have been avoided had they not been exposed.

Chapter 9: Case-Control Studies

Notation for 2-by-2 Cross-Tabulations:

	Exposed	Not Exposed	
Cases	a	b	m_1
Controls	c	d	m_0
	n_1	n_0	n

The case-control method precludes absolute direct estimation of risk, but allows risk to be estimated in relative terms through a statistic known as the **odds ratio**:

$$OR = \frac{p_1/(1-p_1)}{p_0/(1-p_0)} = \frac{(a/m_1)/(b/m_1)}{(c/m_0)/(d/m_0)} = \frac{a/b}{c/d} = \frac{ad}{bc}$$

This statistic is equivalent to a rate ratio from a cohort study when density sampling. Therefore, the odds ratio is a measure of relative incidence (not unlike the risk ratio). Thus, an odds ratio of 1 indicates no association between the exposure and disease, an odds ratio of 2 indicates a doubling of the rate, and so on. For example, as case-control study with the following data:

	Exposed+	Exposed-
Case	647	2
Cntl	622	27

has OR =(647)(27)/(622)(2) = 14.0. This indicates that the exposed group has a rate of disease that is 14 times that of the unexposed group (equivalently, a 1300% *increase* in risk).

Attributable fractions in exposed cases can be determined from case-control studies as:

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

For example, when the OR = 14.0, AF_e = (14.0 – 1)/(14.0) = .929.

The **attributable fraction in the population** is

$$AF_p = \frac{p_0(OR-1)}{p_0(OR-1)+1}$$

where p_0 represent the exposure proportion in controls, which is equal to $p_0 = c / m_0$. For the above data, p_0 = 622 / 649 = .9584 and AF_p = [(.9584)(14.0 – 1)] / [(.958)(14.0 – 1) + 1] = .926.

Chapter 4: Reproducibility and Validity

Reproducibility Statistics

Rater A	Rater B +	Rater B −	
+	a	b	p_1
−	c	d	q_1
	p_2	q_2	N

Overall agreement = $(a + d) / N$

Agreement in Subjects w/ At Least One Positive Diagnosis = $a / (a + b + c)$

Kappa

$$\hat{e} = \frac{2(ad - bc)}{p_1 q_2 + p_2 q_1}$$

Iinterpreting ê: ê ≈ 0 indicates random agreement; ê < .4 represents poor agreement; .4 ≤ ê < .7 represents moderate agreement; ê > .7 represents excellent agreement; ê = 1 indicates perfect agreement.

Validity Statistics

	Disease +	Disease −	
Test +	True Positives (*TP*)	False Positive (*FP*)	n_1
Test −	False Negative (*FN*)	True Negative (*TN*)	n_2
	m_1	m_2	N

$SENsitivity = TP / m_1$ Note: $TP = (SEN)(m_1)$

$SPECificity = TN / m_2$ Note: $TN = (SPEC)(m_2)$

PVP = $(TP) / n_1$ Bayesian: $PVP = \frac{(P)(SEN)}{(P)(SEN) + (1-SPEC)(1-P)}$

PVN = $(TN) / n_2$ Baysian: $PVN = \frac{(1-P)(SPEC)}{(1-P)(SPEC) + (1-SEN)(P)}$

$P = m_1 / N$ where P represents the true prevalence of disease. This allows us to calculate the number of [true] cases, $m_1 = (P)(N)$

$P^* = n_1 / N$ where P^* represents the apparent prevalence of disease