Measuring Associations Between Exposures and Outcomes

Liang-Yi Wang

3.1 INTRODUCTION

Absolute differences

- Absolute differences are often preferred when **public health or preventive activities** are contemplated, as their main goal is often an absolute reduction in the risk of an undesirable outcome.

Relative differences

- **Etiologic studies** that are searching disease causes and determinants usually rely on relative differences in the occurrence of discrete outcomes

Table 3–1 Types of Measures of Association Used in Analytic Epidemiologic Studies

Type	*Examples*		*Usual application*
Absolute difference	Attributable risk in exposed	AR, AR%	Primary prevention impact; search for causes
	Population attributable risk	PAR, PAR%	Primary prevention impact
	Effectiveness, Efficacy	Effectiveness Efficacy	Impact of intervention on recurrences, case fatality, etc
	Mean differences (continuous outcomes)	Mean difference	Search for determinants
Relative difference	Relative risk/rate	RR, OR	Search for causes
	Relative odds		Search for causes

3.2 Measuring Associations in a Cohort Study

- 3.2.1 Relative Risk (RR) and Odds Ratio (OR)
- 3.2.2 Attributable Risk (AR)
  - *Attributable Risk in* Exposed *Individuals* (AR_{exp})
  - *Percent Attributable Risk in Exposed Individuals* ($\%AR_{exp}$)
  - Levin's Population *Attributable Risk (Pop AR)*
  - Percent Population Attributable Risk *(%Pop AR)*

3.3 Cross-sectional Studies: Point Prevalence Rate Ratio

3.4 Measuring Associations in Case-control Studies

- 3.4.1 Odds Ratio (OR)
  - *OR in Matched Case-Control Studies* (Section 7.3.3.)
  - *OR as an Estimate of the Relative Risk in Case-Control Studies: The Rarity Assumption*
  - *When the Rarity Assumption Is Not Necessary: Selecting Population Controls*
  - *Influence of the Sampling Frame for Control Selection on the Parameter Estimated by the OR of Exposure: Cumulative Incidence Versus Density Sampling*
  - *Calculation of the OR When There Are More Than Two Exposure Categories*
- 3.4.2 Attributable Risk (AR) in Case-Control Studies

3.5 Assessing the Strength of Associations

3.2 Measuring Associations in a Cohort Study

3.2.1 Relative Risk (Risk Ratio) and Odds Ratio

q: incidence

Table 3–2 Cross-Tabulation of Exposure and Disease in a Cohort Study

Exposure	*Diseased*	*Nondiseased*	*Disease Incidence (Risk)*	*Probability Odds of Disease*
Present	a	b	$q_+ = \frac{a}{a+b}$	$\frac{q_+}{1-q_+} = \frac{\frac{a}{a+b}}{1-\left(\frac{a}{a+b}\right)} = \frac{a}{b}$
Absent	c	d	$q_- = \frac{c}{c+d}$	$\frac{q_-}{1-q_-} = \frac{\frac{c}{c+d}}{1-\left(\frac{c}{c+d}\right)} = \frac{c}{d}$

[Equation 3.1]

$$\text{Relative risk (RR)} = \frac{q_+}{q_-} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

[Equation 3.2]

$$\text{OR} = \frac{a \times d}{b \times c}$$

$$\text{Probability odds ratio (OR)} = \frac{\frac{q_+}{1-q_+}}{\frac{q_-}{1-q_-}} = \frac{\frac{\frac{a}{a+b}}{1-\left(\frac{a}{a+b}\right)}}{\frac{\frac{c}{c+d}}{1-\left(\frac{c}{c+d}\right)}} = \frac{\frac{\frac{a}{a+b}}{\frac{b}{a+b}}}{\frac{\frac{c}{c+d}}{\frac{d}{c+d}}} = \frac{\frac{a}{b}}{\frac{c}{d}}$$

Confidence Interval: Appendix A, section A.3,A.4

Rare disease

Table 3–3 Hypothetical Cohort Study of the 1-Year Incidence of Acute Myocardial Infarction in Individuals with Severe Systolic Hypertension (≥ 180 mm Hg) and Normal Systolic Blood Pressure (<120 mm Hg)

Blood Pressure Status	*Myocardial Infarction*				
	Number	*Present*	*Absent*	*Probability*	*Probability Odds$_{dis}$*
Severe hypertension	10,000	180	9820	180/10,000 = 0.0180	180/(10,000 − 180) = 180/9820 = 0.01833
Normal	10,000	30	9970	30/10,000 = 0.0030	30/(10,000 − 30) = 30/9970 = 0.00301

$$\text{RR} = \frac{\frac{180}{10{,}000}}{\frac{30}{10{,}000}} = \frac{0.0180}{0.0030} = 6.00$$

當rare disease時
OR 近似於RR

$$\text{Probability OR} = \frac{\frac{180}{9820}}{\frac{30}{9970}} = \frac{0.01833}{0.00301} = 6.09$$

Non-rare disease

Table 3–4 Incidence of Local Reactions in the Vaccinated and Placebo Groups, Influenza Vaccination Trial

		Local Reaction			
Group	*Number*	*Present*	*Absent*	*Probability*	*Probability Odds$_{dis}$*
Vaccine	2570	650	1920	650/2570 = 0.2529	650/(2570 − 650)= 650/1920 = 0.3385
Placebo	2410	170	2240	170/2410 = 0.0705	170/(2410 − 170) = 170/2240 = 0.0759

Note: Based on data for individuals 40 years old or older in Seltser et al.[3] To avoid rounding ambiguities in subsequent examples based on these data (Figure 3–4, Tables 3–7 and 3–9), the original sample sizes in Seltzer et al.'s study (257 vaccinees and 241 placebo recipients) were multiplied by 10.

Source: Data from R Seltser, PE Sartwell, and JA Bell, A Controlled Test of Asian Influenza Vaccine in Population of Families, *American Journal of Hygiene,* Vol 75, pp 112–135, © 1962.

當non-rare disease時
OR 與RR越差越多

$$RR = \frac{\frac{650}{2570}}{\frac{170}{2410}} = \frac{0.2529}{0.0705} = 3.59 \qquad OR = \frac{\frac{650}{1920}}{\frac{170}{2240}} = \frac{0.3385}{0.0759} = 4.46$$

[Equation 3.3]

$$\text{OR} = \frac{\left(\frac{q_+}{1-q_+}\right)}{\left(\frac{q_-}{1-q_-}\right)} = \frac{q_+}{1-q_+} \times \frac{1-q_-}{q_-}$$

$$= \frac{q_+}{q_-} \times \left(\frac{1-q_-}{1-q_+}\right)$$

The term q_+/q_- in Equation 3.3 is the relative risk. Thus, the term

$$\left(\frac{1-q_-}{1-q_+}\right)$$

$$\text{OR} = \text{RR} \times \text{“built-in bias”} = 6.0 \times \frac{1-0.0030}{1-0.0180} = 6.0 \times 1.015 = 6.09$$

$$\text{OR} = 3.59 \times \frac{1-0.0705}{1-0.2529} = 3.59 \times 1.244 = 4.46$$

built-in bias responsible for the discrepancy between the relative risk and the odds ratio estimates

Odds ratio (OR) is valuable

- because it can be measured in case-control studies and because it is directly derived from **logistic regression models**
- Unlike the relative risk, the odds ratio of an event is the exact reciprocal of the odds ratio of the nonevent.

For example, in the study of local reactions to the influenza vaccine discussed previously,[3] the odds ratio of a local reaction

$$OR_{\text{local reaction (+)}} = \frac{\frac{650}{1920}}{\frac{170}{2240}} = 4.46$$

is the exact reciprocal of the odds ratio of not having a local reaction

$$OR_{\text{local reaction (−)}} = \frac{\frac{1920}{650}}{\frac{2240}{170}} = 0.224 = \frac{1}{4.46}$$

In addition, unlike the relative risk, the odds ratio of an event is the exact reciprocal of the odds ratio of the nonevent. For example, in the study of local reactions to the influenza vaccine discussed previously,[3] the odds ratio of a local reaction

$$\mathrm{OR}_{\text{local reaction (+)}} = \frac{\frac{650}{1920}}{\frac{170}{2240}} = 4.46$$

is the exact reciprocal of the odds ratio of not having a local reaction

$$\mathrm{OR}_{\text{local reaction (−)}} = \frac{\frac{1920}{650}}{\frac{2240}{170}} = 0.224 = \frac{1}{4.46}$$

This feature is not shared by the relative risk: using the same example

$$\mathrm{RR}_{\text{local reaction (+)}} = \frac{\frac{650}{2570}}{\frac{170}{2410}} = 3.59$$

$\mathrm{RR}_{\text{exposure}}$的倒數 ≠ $\mathrm{RR}_{\text{unexposure}}$

and

$$\mathrm{RR}_{\text{local reaction (−)}} = \frac{\frac{1920}{2570}}{\frac{2240}{2410}} = 0.8 \neq \frac{1}{3.59}$$

This seemingly paradoxical finding results from the sensitivity of the relative risk to the absolute frequency of the condition of interest, with relative risks associated with *very* common endpoints approaching 1.0. This is easily appreciated when studying the complement of rare outcomes. For example, if the case fatality rates of patients undergoing surgery using a standard surgical technique and a new technique were 0.02 and 0.01, respectively, the relative risk for the relatively rare outcome "death" would be 0.02/0.01 = 2.0. The relative risk for survival, however, would be 0.98/0.99, which is virtually equal to 1.0, suggesting that the new surgical technique did not affect survival. On the other hand, the odds ratio of death would be

$$RR_{death} = \frac{0.02}{0.01} = 2$$

$$RR_{survival} = \frac{0.98}{0.99} = 0.99 \approx \mathbf{1}$$

$$OR_{death} = \frac{\dfrac{0.02}{1.0 - 0.02}}{\dfrac{0.01}{1.0 - 0.01}} = 2.02$$

and that of survival would be

$$OR_{survival} = \frac{\dfrac{0.98}{1.0 - 0.98}}{\dfrac{0.99}{1.0 - 0.99}} = 0.495 = \frac{1.0}{2.02}$$

3.2.2 Attributable Risk (excess fraction)

A measure of **association** based on **the absolute difference** between two risk estimates.

[Equation 3.4]

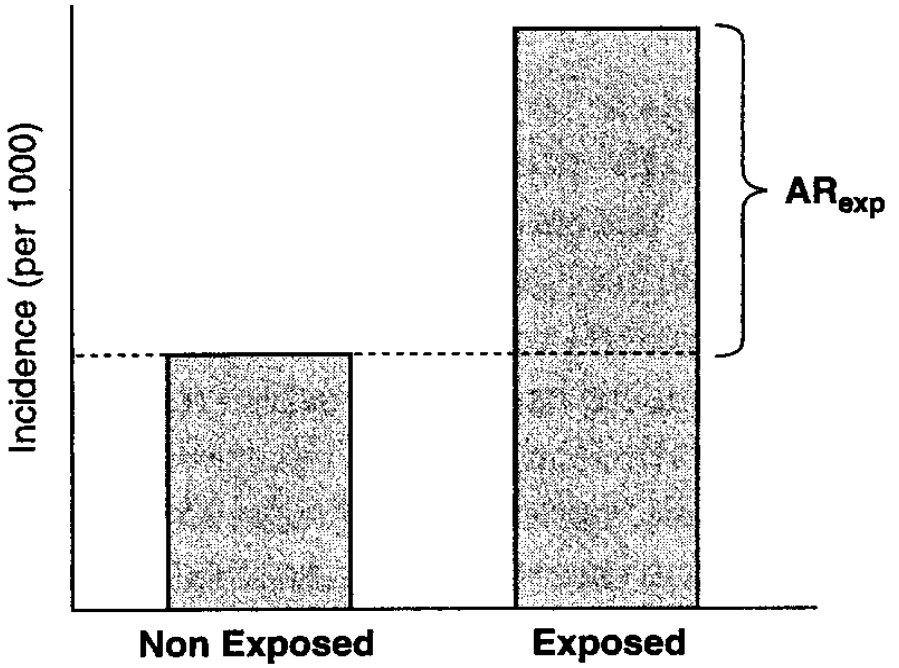
$$AR_{exp} = q_+ - q_-$$

[Equation 3.5]

$$\%AR_{exp} = \left(\frac{q_+ - q_-}{q_+}\right) \times 100$$

[Equation 3.6]

$$\%AR_{exp} = \left(\frac{RR - 1.0}{RR}\right) \times 100$$



Blood Pressure Status	Myocardial Infarction			
	Number	Present	Absent	Probability
Severe hypertension	10,000	180	9820	180/10,000 = 0.0180
Normal	10,000	30	9970	30/10,000 = 0.0030

AR = 0.018-0.003 = 0.015

怎麼解釋?

the cessation of the exposure (severe systolic hypertension) would lower the risk in the exposed group from 0.018 to 0.003.

In the example shown in Table 3–3, the percent attributable risk in the exposed is

$$\%AR_{exp} = \frac{0.018 - 0.003}{0.018} \times 100 = 83.3\%$$

$$\%AR_{exp} = \left(\frac{q_+ - q_-}{q_+}\right) \times 100 = \left(1 - \frac{1}{RR}\right) \times 100 = \left(\frac{RR - 1.0}{RR}\right) \times 100$$

$$\%AR_{exp} = \left(\frac{6.0 - 1.0}{6.0}\right) \times 100 = 83.3\%$$

AR % : the percentage of the total risk of myocardial infarction among hypertensives that is attributable to hypertension

The AR% in the exposed is analogous to percentage **efficacy** when assessing an **intervention** such as a vaccine.

[Equation 3.7]

$$\text{Efficacy} = \left(\frac{q_{\text{cont}} - q_{\text{interv}}}{q_{\text{cont}}}\right) \times 100$$

For example, in a randomized trial to evaluate the efficacy of a vaccine, the risks in persons receiving the vaccine and the placebo are 5% and 15%, respectively. Using Equation 3.7, efficacy is found to be 66.7%:

$$\text{Efficacy} = \left(\frac{15\% - 5\%}{15\%}\right) \times 100 = 66.7\%$$

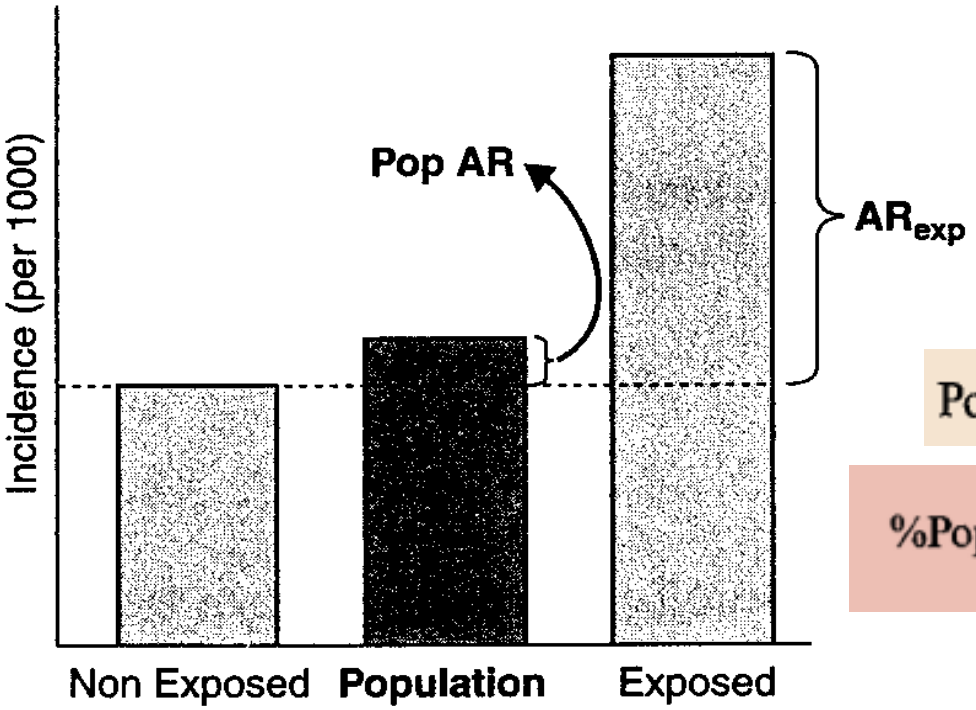
Alternatively, Equation 3.6 can be used to estimate efficacy. In the previous example, the relative risk (placebo/vaccine) is 15% ÷ 5% = 3.0. Thus,

$$\text{Efficacy} = \left(\frac{3.0 - 1.0}{3.0}\right) \times 100 = 66.7\%$$

$$\text{Efficacy} = \left[1.0 - \left(\frac{5\%}{15\%}\right)\right] \times 100 = 66.7\% \qquad = 1\text{-}\frac{1}{RR}$$

Levin's Population Attributable Risk (PAR)

The proportion of the disease risk in the total population associated with the exposure



$$\text{Pop AR} = q_{pop} - q_{-}$$

$$\%\text{Pop AR} = \frac{(q_{pop} - q_{-})}{q_{pop}} \times 100$$

For example, let the exposure prevalence in the target population (p_e) be 0.40 (and, thus, prevalence of *nonexposure*, $[1 - p_e]$, be 0.60), and the risks in exposed and unexposed be $q_+ = 0.20$ and $q_- = 0.15$, respectively.

$$\text{Pop AR} = q_{\text{pop}} - q_-$$

[Equation 3.8]

$$q_{\text{pop}} = [q_+ \times p_e] + [q_- \times (1 - p_e)]$$

$$q_{\text{pop}} = (0.20 \times 0.40) + (0.15 \times 0.60) = 0.17$$

$$\text{Pop AR} = q_{\text{pop}} - q_-$$

[Equation 3.9]

$$\%\text{Pop AR} = \frac{(q_{\text{pop}} - q_-)}{q_{\text{pop}}} \times 100$$

[Equation 3.10]

$$\%\text{Pop AR} = \frac{p_e \times (\text{RR} - 1)}{p_e \times (\text{RR} - 1) + 1} \times 100$$

$$\%\text{Pop AR} = \frac{0.40 \times (1.33 - 1.0)}{0.40 \times (1.33 - 1.0) + 1.0} \times 100$$

$$= \frac{0.40 \times 0.33}{0.40 \times 0.33 + 1.0} = 12\%$$

Confidence Interval: Appendix A, section A.5

Exposure is rare

Exposure is common

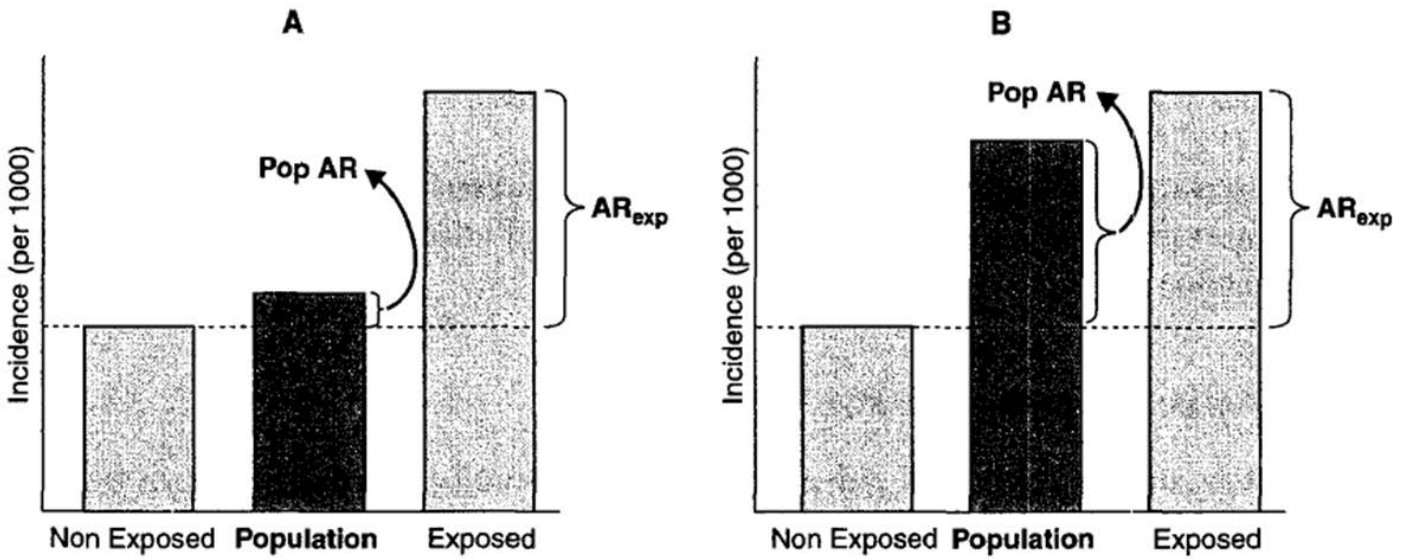


Figure 3–2 Population attributable risk and its dependence on the population prevalence of the exposure. As the population is composed of exposed and unexposed individuals, the incidence in the population is similar to the incidence in the unexposed when the exposure is rare **(A)** and is closer to that in the exposed when the exposure is common **(B)**. Thus, for a fixed relative risk (eg, RR ≈ 2 in the figure) the population attributable risk is heavily dependent on the prevalence of exposure.

$$\text{Pop AR} = q_{\text{pop}} - q_-$$

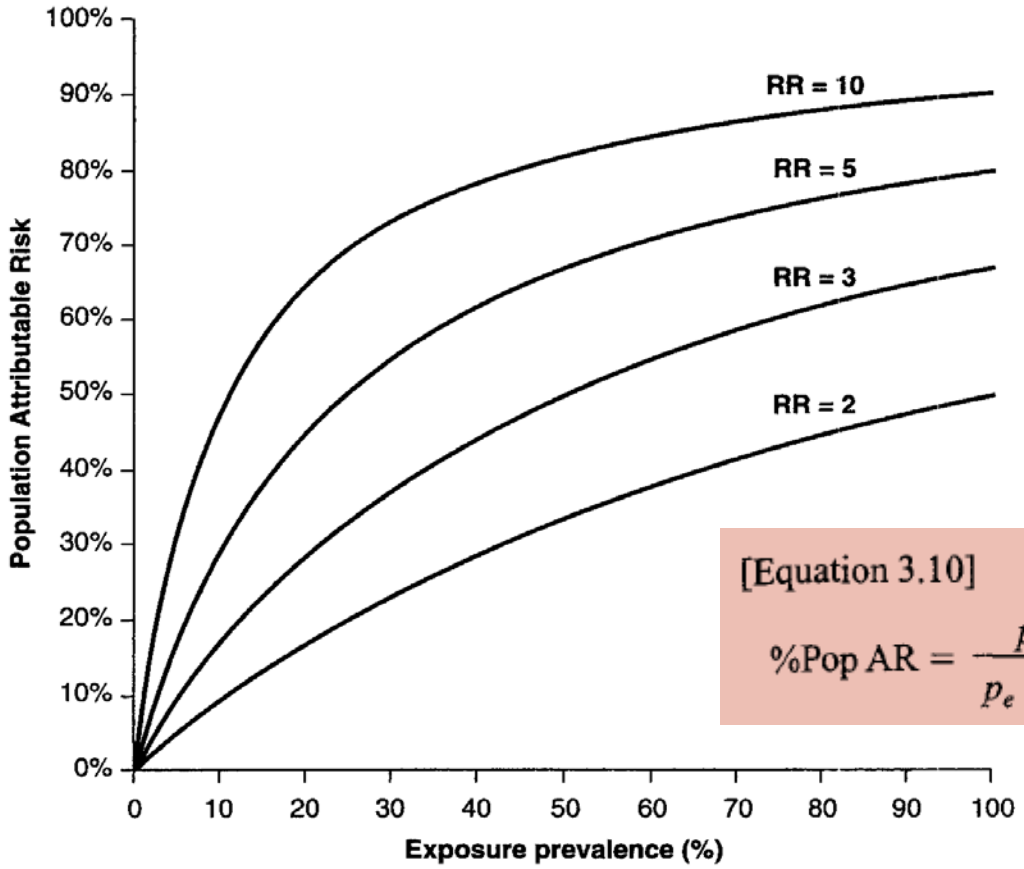
[Equation 3.8]

$$q_{\text{pop}} = [q_+ \times p_e] + [q_- \times (1 - p_e)]$$

$$q_{\text{pop}} = (0.20 \times 0.40) + (0.15 \times 0.60) = 0.17$$

$$\text{Pop AR} = q_{\text{pop}} - q_-$$

Levin 's formula underscores the importance of the two critical elements contributing to the magnitude of the population attributable risk: **the relative risk** and the **prevalence of exposure**.



[Equation 3.10]

$$\%\text{Pop AR} = \frac{p_e \times (\text{RR} - 1)}{p_e \times (\text{RR} - 1) + 1} \times 100$$

Figure 3–3 Population attributable risk: dependence on prevalence of exposure and relative risk.

Walter's extension for multilevel exposures

When exposure has more than just two categories……

$$\%\text{Pop AR} = \frac{p_i \times (\text{RR}_i - 1)}{1 + \sum_{i=0}^{k} p_i \times (\text{RR}_i - 1)} \times 100$$

The subscript i denotes each exposure level; p_i is the proportion of the study population in the exposure level i, and "RR_i" is the relative risk for the exposure level i compared with the unexposed (reference) level.

Assumption of PAR

- Both Levin's formula and Walter's extension for multilevel exposures assume that there is no confounding.
- If confounding is present, it’s not appropriate to calculate the adj-RR and plug it into the formulas in order to obtain an “adj-PAR”.

3.3 Cross-sectional Studies: Point Prevalence Rate Ratio

3.3 CROSS-SECTIONAL STUDIES:
POINT PREVALENCE RATE RATIO

$$\text{Point Prevalence} = \text{Incidence} \times \text{Duration} \times (1 - \text{Point Prevalence})$$

*The derivation of this formula is fairly straightforward. Under the assumption that the disease is in steady state, the incidence and the number of **existing cases (x) at any given point are approximately constant**.

Incidence_rate*(N-X)=X*Fatality_rate,

X/N=Incidence*D*((N-X)/N)

Prevalence=Incidence*D*(1-prevalence)

Fatality_rate=1/Duration

*The derivation of this formula is fairly straightforward. Under the assumption that the disease is in steady state, the incidence and the number of existing cases at any given point (e.g., X) are approximately constant. For an incurable disease, this implies that the number of *new cases* during any given time period is approximately equal to the *number of deaths* among the cases. If N is the population size, I is the incidence, and F is the case fatality rate, the number of new cases can be estimated by multiplying the incidence times the number of potentially "susceptible" $(N - X)$; in turn, the number of deaths can be estimated by multiplying the case fatality rate (F) times the number of prevalent cases. Thus, the above assumption can be formulated as follows: $I \times (N - X) \approx F \times X$. If there is no immigration, the case fatality rate is the inverse of the duration (D).[3] Thus, after a little arithmetical manipulation and dividing numerator and denominator of the right-hand side term by N:

$$I \times D \approx \frac{X}{(N - X)} = \frac{\text{Prevalence}}{(1 - \text{Prevalence})}$$

p.72

An analogous reasoning can be applied to nonfatal diseases, for which F is the proportion cured.

3.3 CROSS-SECTIONAL STUDIES:
POINT PREVALENCE RATE RATIO

$$\text{Point Prevalence} = \text{Incidence} \times \text{Duration} \times (1 - \text{Point Prevalence})$$

Using the notations "Prev" for point prevalence, "q" for incidence, and "Dur" for duration and denoting presence or absence of a given exposure by "+" or "–," the point prevalence rate ratio (PRR) can be formulated as follows:

疾病的prevalence (E組)

$$\text{PRR} = \frac{\text{Prev}_+}{\text{Prev}_-} = \frac{q_+ \times \text{Dur}_+ \times [1.0 - \text{Prev}_+]}{q_- \times \text{Dur}_- \times [1.0 - \text{Prev}_-]}$$

疾病的prevalence (nonE組)

[Equation 3.11]

$$\text{PRR} = \text{RR} \times \left(\frac{\text{Dur}_+}{\text{Dur}_-}\right) \times \left(\frac{1 - \text{Prev}_+}{1 - \text{Prev}_-}\right)$$

Two types of BIAS, (ch 4)

3.4 Measuring Associations in Case-control Studies

3.4 Measuring Associations in **Case-control Studies**

- 3.4.1 Odds Ratio
  - *Odds Ratio as an Estimate of the Relative Risk in Case-Control Studies: The Rarity Assumption*
  - *When the Rarity Assumption Is Not Necessary: Selecting Population Controls*
  - *Influence of the Sampling Frame for Control Selection on the Parameter Estimated by the Odds Ratio of Exposure: Cumulative Incidence Versus Density Sampling*
  - *Calculation of the Odds Ratio When There Are More Than Two Exposure Categories*
- 3.4.2 Attributable Risk in Case-Control Studies

3.4.1 Odds Ratio

rare disease

Cornfield pointed out that the odds ratio of disease and the odds ratio of exposure are mathematically equivalent.

之前提過的cohort study 例子

Table 3–3 Hypothetical Cohort Study of the 1-Year Incidence of Acute Myocardial Infarction in Individuals with Severe Systolic Hypertension (≥ 180 mm Hg) and Normal Systolic Blood Pressure (<120 mm Hg)

Blood Pressure Status	*Myocardial Infarction*				
	Number	*Present*	*Absent*	*Probability*	*Probability Odds$_{dis}$*
Severe hypertension	10,000	180	9820	180/10,000 = 0.0180	180/(10,000 − 180) = 180/9820 = 0.01833
Normal	10,000	30	9970	30/10,000 = 0.0030	30/(10,000 − 30) = 30/9970 = 0.00301

$$\text{RR} = \frac{\frac{180}{10{,}000}}{\frac{30}{10{,}000}} = \frac{0.0180}{0.0030} = 6.00$$

$$\text{Probability OR} = \frac{\frac{180}{9820}}{\frac{30}{9970}} = \frac{0.01833}{0.00301} = 6.09$$

Table 3–5 Hypothetical Case-Control Study of Myocardial Infarction in Relation to Systolic Hypertension, Based on a 1-Year Complete Follow-up of the Study Population from Table 3–3

*Systolic Blood Pressure Status**	*Myocardial Infarction*			
	Present		*Absent*	
Severe hypertension	180	(*a*)	9820	(*b*)
Normal	30	(*c*)	9970	(*d*)
Total	210	(*a* + *c*)	19790	(*b* + *d*)

*Severe systolic hypertension ≥180 mm Hg, and normal systolic blood pressure < 120 mm Hg.

$$\text{Odds}_{\text{exp cases}} = \frac{\frac{a}{a+c}}{1-\left(\frac{a}{a+c}\right)} = \frac{a}{c}$$

$$\text{Odds}_{\text{exp controls}} = \frac{\frac{b}{b+d}}{1-\left(\frac{b}{b+d}\right)} = \frac{b}{d}$$

[Equation 3.12]

$$\text{OR}_{\text{exp}} = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{a \times d}{b \times c} = \frac{\frac{a}{b}}{\frac{c}{d}} = \text{OR}_{\text{dis}}$$

$$\text{OR}_{\text{exp}} = \frac{\frac{180}{30}}{\frac{9820}{9970}} = \frac{180 \times 9970}{9820 \times 30} = 6.09 = \text{OR}_{\text{dis}}$$

Table 3–6 Case-Control Study of the Relationship of Myocardial Infarction to Presence of Severe Systolic Hypertension Including All Cases and a 10% Sample of Noncases from Table 3–5

*Systolic Blood Pressure Status**	*Myocardial Infarction*			
	Present		*Absent*	
Severe hypertension	180	(*a*)	982	(*b*)
Normal	30	(*c*)	997	(*d*)
Total	210	(*a* + *c*)	1979	(*b* + *d*)

*Severe systolic hypertension ≥180 mm Hg, and normal systolic blood pressure < 120 mm Hg.

$$OR_{exp} = \frac{\frac{180}{30}}{\frac{982}{997}} = \frac{180 \times 997}{982 \times 30} = 6.09 = OR_{dis}$$

This example underscores the notion that the sampling fractions do not have to be the same in cases and controls. To obtain unbiased estimates of the absolute odds of exposure for cases and controls, however, sampling fractions must be independent of exposure: that is, they should apply equally to cells (*a*) and (*c*) for cases and cells (*b*) and (*d*) for controls.

Non-rare disease

之前提過的cohort study 例子

Table 3–4 Incidence of Local Reactions in the Vaccinated and Placebo Groups, Influenza Vaccination Trial

		Local Reaction			
Group	*Number*	*Present*	*Absent*	*Probability*	*Probability Odds$_{dis}$*
Vaccine	2570	650	1920	650/2570 = 0.2529	650/(2570 − 650)= 650/1920 = 0.3385
Placebo	2410	170	2240	170/2410 = 0.0705	170/(2410 − 170) = 170/2240 = 0.0759
		820	**4160**		

$$RR = \frac{\frac{650}{2570}}{\frac{170}{2410}} = \frac{0.2529}{0.0705} = 3.59$$

$$OR = \frac{\frac{650}{1920}}{\frac{170}{2240}} = \frac{0.3385}{0.0759} = 4.46$$

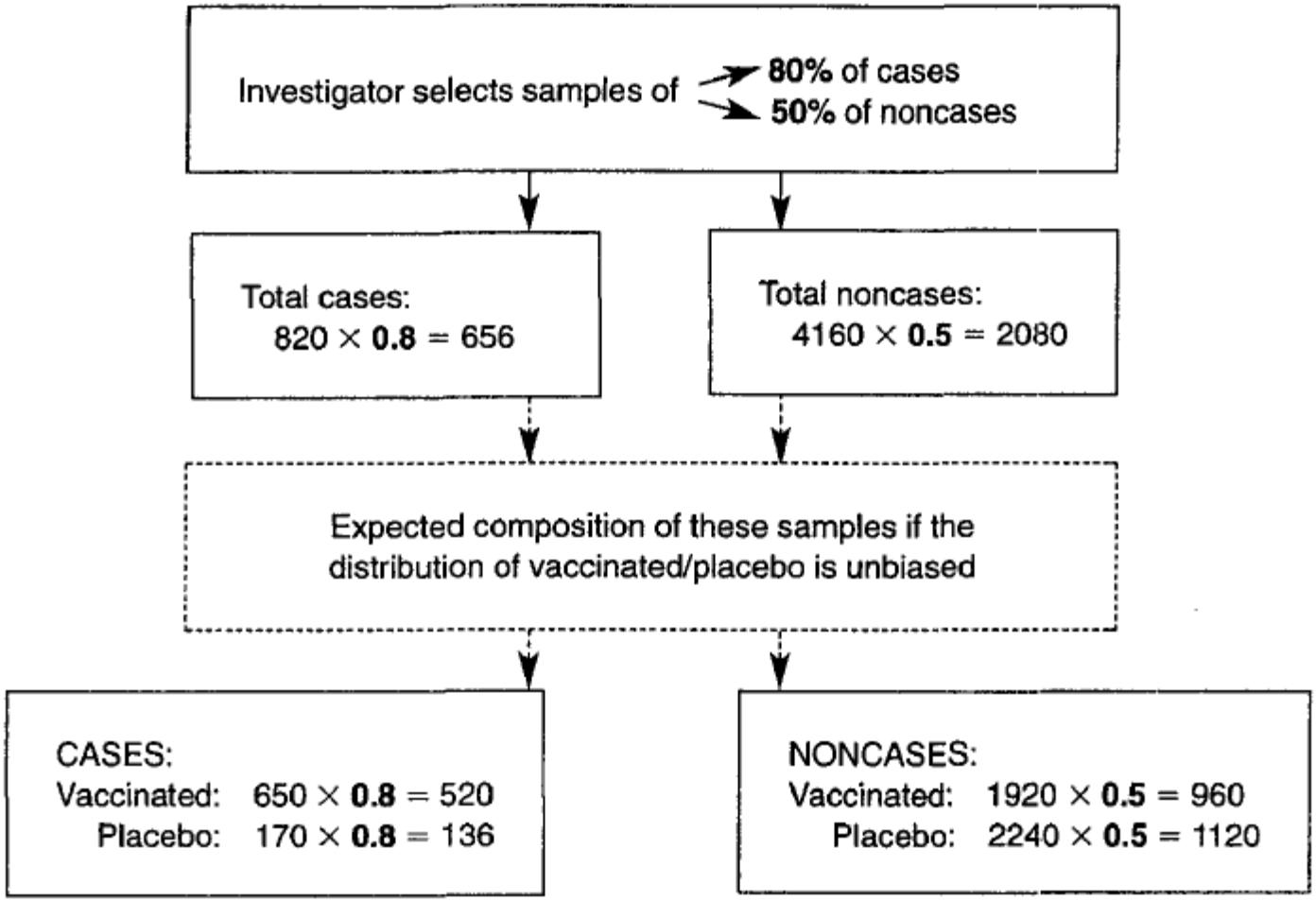


Figure 3–4 Selection of 80% of total cases and 50% of noncases in a case-control study from the study population shown in Table 3–4. Expected composition assumes no random variability.

Table 3–7 Case-Control Study of the Relationship Between Occurrence of Local Reaction and Previous Influenza Immunization

Vaccination	*Cases of Local Reaction*	*Controls Without Local Reaction*
Yes	520	960
No	136	1120
Total	$820 \times 0.8 = 656$	$4160 \times 0.5 = 2080$

Note: Based on a perfectly representative sample of 80% of the cases and 50% of the controls from the study population shown in Table 3–4 (see Figure 3–4).

$$OR_{exp} = \frac{\frac{520}{136}}{\frac{960}{1120}} = 4.46 = OR_{dis}$$

the odds of developing local reactions is 4.46 times greater than the odds for those who received the placebo

The fact that the odds ratio of exposure is identical to the odds ratio of disease permits a "prospective" interpretation of the odds ratio in case-control studies

Odds Ratio as an Estimate of the Relative Risk in Case-Control Studies:

The Rarity (稀有的) Assumption

In a case-control study, the use of the odds ratio to estimate the relative risk is based on the assumption that the disease under study has a low incidence, thus resulting in a small built-in bias (Equation 3.3)

[Equation 3.3]

$$OR = \frac{\left(\frac{q_+}{1-q_+}\right)}{\left(\frac{q_-}{1-q_-}\right)} = \frac{q_+}{1-q_+} \times \frac{1-q_-}{q_-}$$

$$= \frac{q_+}{q_-} \times \left(\frac{1-q_-}{1-q_+}\right)$$

The term q_+/q_- in Equation 3.3 is the relative risk. Thus, the term

$$\left(\frac{1-q_-}{1-q_+}\right)$$

q_- = the incidence in the unexposed

[Equation 3.13]

$$RR = \frac{OR}{1 - [q_- - (OR \times q_-)]}$$

Table 3–8 Relative risk equivalent to a given odds ratio as a function of the incidence of the condition that defined case status in a case-control study

Incidence in the unexposed population	*Odds ratio = 0.5*	*Odds ratio = 1.5*	*Odds ratio = 2.0*	*Odds ratio = 3.0*
		Relative Risk Equivalent		
0.001	0.50	1.50	2.00	2.99
0.01	0.50	1.49	1.98	2.94
0.05	0.51	1.46	1.90	2.73
0.1	0.53	1.43	1.82	2.50
0.2	0.56	1.36	1.67	2.14
0.3	0.59	1.30	1.54	1.88
0.4	0.63	1.25	1.43	1.67*

*Example of calculation: for an OR = 3 and $q_- = 0.4$ and using Equation 3.13, the relative risk is:

$$RR = \frac{3}{1 - (0.4 - 3 \times 0.4)} = 1.67$$

[Equation 3.13]

$$RR = \frac{OR}{1 - [q_- - (OR \times q_-)]}$$

When the Rarity Assumption Is Not Necessary Selecting Population Controls

The rare-disease assumption is irrelevant in situations in which the control group is a sample of the total population,[13] which is the usual strategy in case-control studies within a defined cohort (Chapter 1, Section 1.4.2). In this situation, the odds ratio is a *direct estimate of the relative risk*, irrespective of the frequency of the outcome of interest.

[Equation 3.14]

$$OR_{exp} = \frac{Odds_{exp\ cases}}{Odds_{exp\ total\ population}} = \frac{\left(\frac{a}{c}\right)}{\left(\frac{a+b}{c+d}\right)} = \frac{\left(\frac{a}{a+b}\right)}{\left(\frac{c}{c+d}\right)} = RR$$

selecting non-cases as controls may not be always the best option.

Table 3–9 Cross-Tabulation of a Defined Population by Exposure and Disease Development

Exposure	*Cases*	*Noncases*	*Total Population (Cases + Noncases)*
Present	a	b	a + b
Absent	c	d	c + d

Non-rare disease

之前提過的cohort study 例子

Table 3–4 Incidence of Local Reactions in the Vaccinated and Placebo Groups, Influenza Vaccination Trial

		Local Reaction			
Group	*Number*	*Present*	*Absent*	*Probability*	*Probability Odds$_{dis}$*
Vaccine	2570	650	1920	650/2570 = 0.2529	650/(2570 − 650)= 650/1920 = 0.3385
Placebo	2410	170	2240	170/2410 = 0.0705	170/(2410 − 170) = 170/2240 = 0.0759

$$\text{RR} = \frac{\frac{650}{2570}}{\frac{170}{2410}} = \frac{0.2529}{0.0705} = 3.59 \qquad \text{OR} = \frac{\frac{650}{1920}}{\frac{170}{2240}} = \frac{0.3385}{0.0759} = 4.46$$

$$\text{OR}_{\text{exp}} = \frac{\text{Odds}_{\text{exp cases}}}{\text{Odds}_{\text{exp pop}}} = \frac{\left(\frac{650}{170}\right)}{\left(\frac{2570}{2410}\right)} = \frac{\left(\frac{650}{2570}\right)}{\left(\frac{170}{2410}\right)} = \frac{q_+}{q_-} = 3.59 = \text{RR}$$

unbiased samples of 40% of the cases and 20% of the total population

Table 3–10 Case-Cohort Study of the Relationship of Previous Vaccination to Local Reaction

Previous Vaccination	*Cases of Local Reaction*	*Cohort Sample*
Yes	260	514
No	68	482
Total	328	996

Note: Based on a random sample of the study population in Table 3–4, with sampling fractions of 40% for the cases and 20% for the cohort.

$$OR_{exp} = \frac{\frac{260}{68}}{\frac{514}{482}} = 3.59 = RR$$

case-cohort studies

One of the advantages of the case-cohort approach is that it allows direct estimation of the relative risk and thus does not have to rely on the rarity assumption. Another advantage is that because the control group is a sample of the total reference population, an unbiased estimate of the exposure prevalence (or distribution) needed for the estimation of Levin's population attributable risk (Equation 3.10) can be obtained.

[Equation 3.10]

$$\%\text{Pop AR} = \frac{p_e \times (\text{RR} - 1)}{p_e \times (\text{RR} - 1) + 1} \times 100$$

Influence of the Sampling Frame for Control Selection on the Parameter Estimated by the Odds Ratio of Exposure:

Cumulative Incidence Versus Density Sampling

Table 3–11 Summary of the Influence of Control Selection on the Parameter Estimated by the Odds Ratio of Exposure in Case-Control Studies Within a Defined Cohort

Design	*Population Frame for Control Selection*	*Exposure Odds Ratio Estimates*
Nested case-control	Population at approximate times when cases occur during follow-up	Rate (density) ratio
	(Population during follow-up minus cases)	(Density odds ratio)
Case-cohort	Total cohort at baseline	Cumulative incidence ratio (relative risk)
	(Total cohort at baseline minus cases that develop during follow-up)	(Probability odds ratio)

Calculation of the Odds Ratio

When There Are More Than Two Exposure Categories

Table 3–12 Distribution of Cases of Craniosynostosis and Normal Controls According to Maternal Age

Maternal Age (Years) (1)	*Cases (2)*	*Controls (3)*	*Odds of Specified Maternal Age vs Reference in Cases (4)*	*Odds of Specified Maternal Age vs Reference in Controls (5)*	*Odds Ratio (6) = (4)/(5)*
<20*	12	89	12/12	89/89	1.00*
20–24	47	242	47/12	242/89	1.44
25–29	56	255	56/12	255/89	1.63
>29	58	173	58/12	173/89	2.49

*Reference category.

3.4.2 Attributable Risk in Case-Control Studies

[Equation 3.15]

$$\%AR_{exp} = \left(\frac{OR - 1.0}{OR}\right) \times 100$$

[Equation 3.16]

$$\%\text{Pop AR} = \frac{p_e^* \times (OR - 1)}{p_e^* \times (OR - 1) + 1} \times 100$$

p_e^*: exposure prevalence among controls.

As shown by Levin and Bertell,[17] if the odds ratio is used as the relative risk estimate, Equation 3.16 reduces to a simpler equation:

$$\%\ Pop\ AR = \frac{p_{e\ case} - p_{e\ control}}{1.0 - p_{e\ control}} \times 100$$

where $p_{e\ case}$ represents the prevalence of exposure among cases—that is, $a/(a + c)$ in Table 3–9—and $p_{e\ control}$ represents the prevalence of exposure among controls—that is, $b/(b + d)$ in Table 3–9.

3.5 Assessing The Strength of Associations

- It's hard to compare association strengths
  - It's absurdity of saying that systolic blood pressure (a 50-mm/Hg increase) is more important for MRI than total cholesterol (a 1-mg/dL increase).
- estimate the exposure intensity necessary for that factor to produce an association of the same magnitude as that of well-established risk factors or vice-versa.

A relative risk of 2.2 for coronary heart disease mortality comparing men drinking 9+ or more cups of coffee per day versus < one cup per day corresponds to:

Smoking:	4.3 cigarettes/day
Systolic blood pressure:	6.9 mm/Hg
Total serum cholesterol:	0.47 mmol/L
Serum high-density lipoprotein:	−0.24 mmol/L

Source: Data from A Tverdal et al, Coffee Consumption and Death from Coronary Heart Disease in Middle-Aged Norwegian Men and Women, *British Medical Journal*, Vol 300, pp 566–569, © 1990.

Exhibit 3–1 A Possible Way to Describe the Strength of an Association Between a Risk Factor and an Outcome

頸動脈壁內膜中層厚度

Exhibit 3–2 Cross-Sectionally Determined Mean Intima-Media Thickness (IMT) of the Carotid Arteries (mm) by Passive Smoking Status in Never-Active Smokers, the Atherosclerosis Risk in Communities Study, 1987–1989

動脈粥樣硬化

	Passive Smoking Status in Never-Active Smokers		
	Absent (n = 1,774)	Present (n = 3,358)	Estimated Increase by Year of Age
Mean IMT (mm) →	0.700	0.711	0.011

Age-equivalent excess attributable to passive smoking:
(0.711 − 0.700)/0.011 = 1 year

Source: Data from G. Howard et al, Active and Passive Smoking Are Associated with Increased Carotid Wall Thickness. The Atherosclerosis Risk in Communities Study, Archives of Internal Medicine, Vol 154, pp 1277–1282, © 1994, American Medical Association.