Public Health Sciences 310
Epidemiologic Methods

Lecture 4
Measures of Association

January 16, 2024

Brian C.-H. Chiu

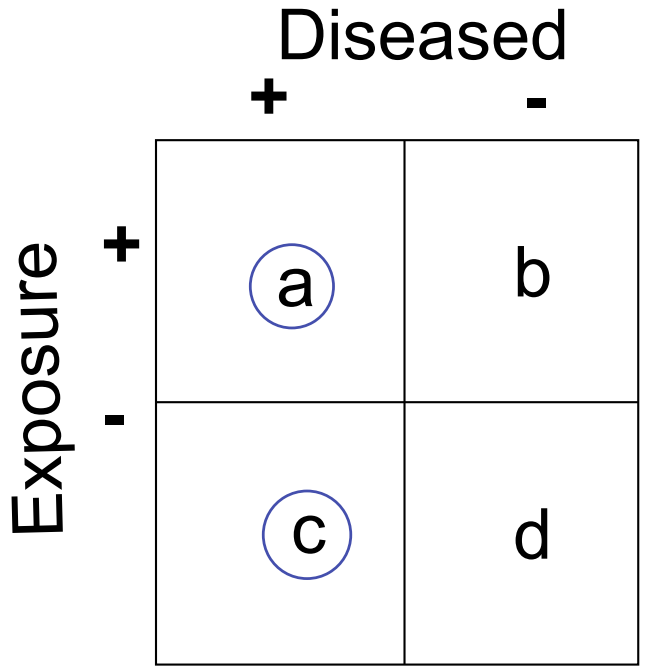
Measures of Association used in Epi Studies

- **Based on relative differences (or ratios):**
  - RR – Relative risk (or risk ratio) person; cumulative incidence
  - RR – Rate ratios person-time; incidence rate/density
  - OR – Odds ratios
- **Based on absolute differences:**
  - AR – Attributable risk or risk difference
  - AFP – Attributable fraction percent
  - PAR – Population attributable risk

Measures of Relative Effect

- **Relative risk (or risk ratio)** Cohort Study
  survival Analysis
  - The ratio of the cumulative incidence in exposed individuals to that in unexposed
- **Rate ratios** Cohort Study
  "The ratio of rate"
  - The ratio of the incidence rate/density in the exposed to that in unexposed
- **Odds ratios** Case Control Study
  - The ratio of the odds of exposure in the diseased to that in non-diseased
    - Odds: the ratio of the probability of the event of interest to that of the non event Determined by outcome, hence, c.c.
- Usual application: search for causes

2 x 2 Contingency Table



Compare between exposure +/−

$q^+_{exp.} = \frac{\text{numerator } a}{\text{denominator } a+b}$

$q^-_{nonexp.} = \frac{\text{numerator } C}{\text{denominator } C+d}$

- Similar contingency tables can represent cohort, case-control and cross-sectional studies
- Let $q+$ ($\frac{a}{a+b}$; numerator a, denominator $a+b$) be the disease incidence in exposed individuals, $q-$ ($\frac{C}{C+d}$; numerator C, denominator $C+d$) be the disease incidence in the unexposed.

Data from Cohort study

However, we can use q^+ and q^- to determine RR, which can use to estimate OR for C.C.

Why use C.C.? More feasable, cheaper for research.

- Can we determine q+ and q- from a case-control study?

NO. Since case control study begins w/ cases and controls (Retrospective) Can't determine incidence from C.C.

- Can we determine q+ and q- from a "snapshot" cross-sectional study?

No. C.S. measures exposure/outcome at the same time, so not possible to calculate incidence.

Relative Risk

- The ratio of the cumulative incidence in exposed individuals to that in unexposed

RR is for cohort study, and cohort study compare exposures.

- $\mathrm{RR} = \frac{q+}{q-} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$ (a: Diseased, Exposed; b: Exposed; c: Diseased, Unexposed; d: Unexposed; Exp; Unexp.)

Exposure	Diseased +	Diseased -
+	a	b
-	c	d

Framingham Cohort: Association between systolic blood pressure ≥ 165 mm Hg & 18-year incidence of CHD.

	Event	No event	Total
Exposed	95	201	296
Unexposed	173	894	1,067

Risk Ratio

$$RR = \frac{\text{Event exp} / \text{Total exp.}}{\text{No event unexp.} / \text{Total unexp.}} = \frac{95/296}{173/1,067} = 1.98$$

Person

- Assume no follow-up, all participant complete study.

	Event	Person Time
Exposed	95	3,511
Unexposed	173	14,865

Rate Ratio

True Incidence

$$\star RR = \frac{95/3,511}{173/14,865} = 2.32$$

Person-time

- Take into account of follow-up, censoring event.

Contribution of different time. In real world, not everyone contributed entirely to study, especially with long follow-up time ex) 18yrs.

What do these two results suggest about the association of SBP and CHD risk in the Framingham Cohort? Why are the two results different?

Interpretation of Results

- **Example - Risk Ratio**

  The 18-year risk of CHD is 1.98 times higher among persons with SBP 165 mm Hg or higher compared to those with a lower SBP.

- **Example – Rate Ratio**

  The 18-year incidence of CHD is 2.32 times higher among persons with SBP 165 mm Hg or higher compared to those with a lower SBP.

Odds Ratio (Case-Control Study)

- One cannot determine incidence in a c-c study.
  - Why not?

→ Since case control study begins w/ cases and controls (Retrospective) Can't determine incidence from C.C.

- Instead, one compares the ODDS of exposure in the cases and controls.
- Consider a c-c study of bladder cancer and smoking:
- The odds (the ratio of the probability of the event to that of the nonevent) of smoking among the cases is 40/60=.67 (exp odds in cases)
- The odds of smoking among the controls is 36/164=.22 (exp odds in ctls)

	BlCa+ Cases	BlCa- Control
E+ Smoker	40 a	36 b
E- Nonsmoker	60 c	164 d

or $\frac{40 \cdot 164}{60 \cdot 36}$

The ratio of the odds is .67/.22=3.05 Unitless
This "odds ratio" has the same scale as a RR.

"1" means the odds are the same. $\frac{\text{same}}{\text{same}} = 1$

"The odds of smoking among bladder cancer cases is three times greater than controls."

How is OR related to the RR?

Relative Risk and Odds Ratio

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

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

Cases Controls → Cases / Controls

Controls Cases

TABLE 2 Cross-tabulation of exposure and disease in a cohort study.

Exposure	Diseased	Nondiseased	Disease incidence (risk)	Probability odds of disease
Present	a	b	$q_+ = \frac{a}{a+b}$ Diseased / Exposed	$\frac{q_+}{1-q_+} = \frac{\frac{a}{a+b}}{1-\left(\frac{a}{a+b}\right)} = \frac{a}{b}$ Diseased / Nondiseased } Exp
Absent	c	d	$q_- = \frac{c}{c+d}$ Diseased / Nonexposed	$\frac{q_-}{1-q_-} = \frac{\frac{c}{c+d}}{1-\left(\frac{c}{c+d}\right)} = \frac{c}{d}$ Diseased / Nondiseased } Nonexp.

Hypothetical Cohort Study

TABLE 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 (Exp. +)	Absent (Nonexp. −)	Probability	Probability odds$_{dis}$
Severe hypertension (Cases +)	10,000	180 (a)	9820 (c)	180/10,000 = 0.018	180/(10,000 − 180) = 180/9820 = 0.01833
Normal (Control −)	10,000	30 (b)	9970 (d)	30/10,000 = 0.003	30/(10,000 − 30) = 30/9970 = 0.00301

$$RR = \frac{\frac{180}{10000}}{\frac{30}{10000}} = \frac{0.0180}{0.0030} = 6.00$$

Risk Ratio

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

Cohort begins with Exposed/Nonexposed. ⟷ ✓ good estimate of one another ⟷ Case study begins with Cases/Control.

RR vs. OR: not as rare events including all cases and all non-cases

	Event	No event	Total
Exposed	180	9820	10,000
Unexposed	30	9970	10,000

$$RR_{dis} = \frac{\frac{180}{10{,}000}}{\frac{30}{10{,}000}} = 6.0$$

Common Disease

	Event	No event	Total
Exposed	1800	8200	10,000
Unexposed	300	9700	10,000

$$OR_{dis} = \frac{\frac{1800}{8200}}{\frac{300}{9700}} = 7.10$$

disease

x Huge difference

In c.c., OR is not a good estimate for RR for common disease

fraction huge difference

WHY?

- RR = (a/a+b) ÷ (c/c+d)

  Exp. Diseased / Total Exp.

  Nonexp. Diseased / Total Nonexp.

  Exp

  Nonexp.

- Rare disease assumption: OR≈RR

  Diseased ↓↓↓

  Exp. Diseased a <<<< b Exp. Undiseased

  Much smaller

  Nonexp. Diseased c <<<< d Nonexp. Undiseased

  therefore:

  (a/a+b) ÷ (c/c+d) ≈ (a/b) ÷ (c/d) = (ad) ÷ (bc)

- The formula for an OR

  Cases

  (a/c) ÷ (b/d) = (ad) ÷ (bc)

OR for c.c. can still determine a valid association for exposure and disease (cohort).

OR and RR

Useful for Common Disease: Prone to large error/difference.

- When the OR is used to estimate the RR, there is a "built-in" bias:

$$\mathrm{OR} = \frac{\left(\frac{q_+}{1-q_+}\right)}{\left(\frac{q_-}{1-q_-}\right)} = \frac{q_+}{1-q_+} \times \frac{1-q_-}{q_-} = \frac{q_+}{q_-} \times \left(\frac{1-q_-}{1-q_+}\right) \quad \text{(Eq. 3.3)}$$

RR

Built-in bias

easier to coduct c.c. study
↑
Use OR to estimate RR

OR and RR

Vice Versa

$$OR = RR \times \frac{1 - q_-}{1 - q_+}$$

d (Nonexp) / b (Exp.) — Unexp

C.C. — Cohort
OR tends to overestimate RR
b/c presence of built-in bias

- The OR is always further away from 1.0 than the RR.
  - The higher the disease or event incidence, the higher the discrepancy.

Common Disease

- Examples:
  - For risk factors ($q_+ > q_-$, RR>1.0):

    ($1 – q_-$) > ($1 – q_+$) and the resulting bias is > 1.0, therefore OR > RR.

    "Risk" Exp. Nonexp.

    Bias ↑ ↓ Big # > 1
  - For protective factors ($q_+ < q_-$, RR<1.0):

    ($1 – q_-$) < ($1 – q_+$) and the resulting bias is < 1.0, therefore OR < RR.

    Bias ↓ ↑ Small # < 1

OR and RR

- Regardless of whether OR can properly estimate RR, it is a genuine measure of association
- When probability of the event (q) is low in both exposed and unexposed: Rare event

$$\frac{q}{1-q} \approx q$$

  - And thus, the "built-in-bias" term is also ≈ 1.0, and OR ≈ RR.
- When probability (Common) of the event (q) is high, the bias can be quite large. OR > RR
- The built-in bias exist (or is a concern) only when OR is used as an estimate of RR

If directly calculate RR, no need for bias

OR vs. RR: Advantages

- OR can be estimated from a case-control study.
- OR can be estimated from logistic regression.
- OR allows the estimation of RR

Case Control

Cohort

Exposure vs. Disease Odds Ratio C.C.

Same OR, different interpretation

	Case	Control
Exposed	a	b
Not Exposed	c	d

Exposure Odds Ratio

$$OR_{exp} = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{a\,d}{b\,c}$$

(Case: a exp, c nonexp; Control: b exp, d nonexp)

Case study begins with Cases/Control.

Disease Odds Ratio

$$OR_{dis} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a\,d}{b\,c}$$

(Exp.: a case, b control; Unexp.: c case, d control)

Exposure vs disease OR

	Event	No event	Total
Exposed	1800	8200	10,000
Unexposed	300	9700	10,000

$$OR_{exp} = \frac{\frac{1800}{300}}{\frac{8200}{9700}} = 7.10$$

$$OR_{dis} = \frac{\frac{1800}{8200}}{\frac{300}{9700}} = 7.10$$

Interpretation of Results

- **Example – Odds Ratio for exposure**

  These data suggest that the odds of exposure among cases is 7.1 times higher than controls

- **Example – Odds Ratio for disease**

  These data suggest that the odds of disease is 7.1 times higher among exposed compared to unexposed

Which OR is more appropriate for case-control studies? ORexp.

C.C. starts with disease, then determine exposure.

Summary: Relative Measures of Association

- Cohort Study
  - May report risk ratio or rate ratio (if data available)
  - Risk ratio and rate ratio both are often called "relative risk" (RR) as the interpretation is the same
  - Difference is whether it is based on cumulative incidence or incidence rate/density

- Case-Control Study
  - Odds ratio is the only option because incidence and prevalence cannot be determined (investigator has selected on the basis of event/disease and thus has controlled the ratio of cases to non-cases
  - The rare-disease assumption applies only to situations in which the OR is used to estimate the relative risk. When OR is used as a measure of the association itself, the rare-disease assumption is not needed.
  - Regardless of whether OR can properly estimate RR, it is a *bona fide* measure of association.

Measures of Absolute Difference

• Absolute difference based on incidence → cohort study ✓.

However, many research studies use C.C. to est. abs. diff, and they ignore built-in bias.

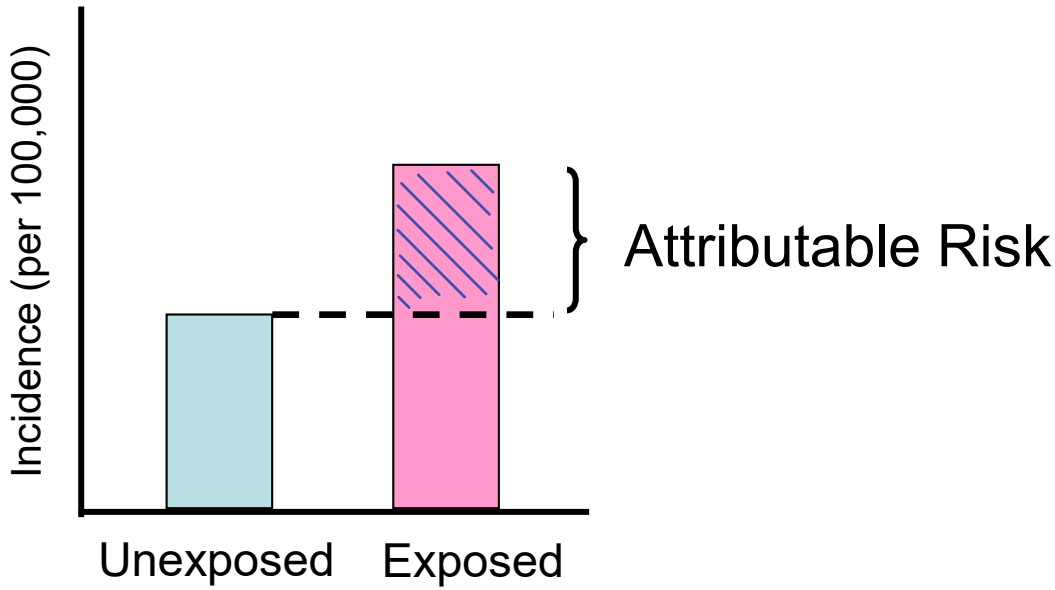
- **Attributable risk or risk difference**
  - The difference between the risk in the exposed and the unexposed (absolute excess incidence which would be eliminated were the exposure removed)
  - KEY ASSUMPTION: a true, causal association
  - $Incidence_{exposed} - Incidence_{unexposed}$
- **Attributable fraction percent**
  - Unitless percentage of the total risk in the **exposed** attributable to the exposure

$$\frac{Incidence_{exposed} - Incidence_{unexposed}}{Incidence_{exposed}}$$

- **Population attributable risk (Levin's population attributable risk)**
  - the % of all cases that would be were the population entirely unexposed, compared with its current exposure pattern (need data about the current prevalence of the exposure)
- Usual application: public health or preventive activities

Attributable Risk

- The difference between the risk in the exposed and the unexposed
- Information about the absolute effect of the exposure or excess risk of disease among the exposed
- Computed: $AR_{exp} = I_{exp+} - I_{exp-}$



Framingham Example: Attributable Risk

	Event	Person Time
Exposed	95	3,511
Unexposed	173	14,865

- Using incidence density data
- Computed: $AR_{exp} = I_{exp+} - I_{exp-}$

Exp. Incidence rate Unexp.

$$AR_{exp} = 95/3511 - 173/14865$$
$$= 0.027 - 0.0116$$
$$= 0.0154$$
$$= 154/10{,}000$$

Interpretation: *Among individuals with SBP ≥ 165 mm Hg, the excess incidence of CHD is 154 per 10,000 person-years.*

Percent Attributable Risk (Etiologic Fraction)

- Using incidence density data

- Computed: $\%AR_{exp} = (I_{exp+} - I_{exp-}) / I_{exp+}$

  $= (0.0154/0.02705) \times 100$

  $= 57\%$

  I_{exp+}

Interpretation:

Among individuals with SBP ≥ 165 mm Hg, approximately 57% of the CHD incidence can be attributed to high SBP

Population Attributable Risk

- The proportion of disease in a population attributable to a particular exposure.

- Dependent on the relative risk (RR_{exp+}) disease and the prevalence of the exposure (p_e).

$$\text{Pop AR} = I_{total} - I_{unexp}$$ = "I exp for the population"

Example: Population Attributable Risk

	Event	Person Time
Exposed	95	3,511
Unexposed	173	14,865
TOTAL	268	18,376

I_{total} $I_{unexp.}$

Pop AR = 268/18,376 - 173/14,865

= 0.0001698

= 16.98/100,000

Interpretation: *The excess annual incidence rate of CHD that is attributable to SBP > 165 in the population is approximately 17 per 100,000.*

Interpreting the relative risk and attributable risk for Cohort Study

- Relative risk
  - Measures strength of the association between exposure and disease
  - Provides information that can be used to judge whether a valid observed association is likely to be causal
- Attributable risk
  - Provide a measure of the public health impact of an exposure
  - Assume the exposure is causal

Interpreting the relative risk and attributable risk

- Magnitude of the relative risk does not predict the magnitude of the attributable risk

Table 4-18. Relative and attributable risks of mortality from lung cancer and coronary heart disease among cigarette smokers in a cohort study of British male physicians

	Annual mortality rate per 100,000	
	Lung cancer (Rare)	Coronary heart disease (Common)
Cigarette smokers	140	669
Nonsmokers	10	413
Relative risk	14.0 (larger)	1.6 (smaller)
Attributable risk	130/10⁵/year (smaller)	256/10⁵/year (larger) → ↑PH impact for mortality

R. Doll and R. Peto, Mortality in relation to smoking: Twenty years' observations on male British doctors. *Br. Med. J.* 2:1525, 1976.

- Generally, relative risk is used most commonly by epidemiologists whereas attributable risk is used most commonly by public health administrators and policy makers (PH scale:)

ex) Allocate Resources more towards CHD.