Measures of association: comparing disease frequencies

Gustaf Edgren, PhD
Karolinska Institutet

Outline

- *Repetition of measures of disease occurrence*
- Relative association measures
- Absolute association measures
- Extended examples
- Summary

Repetition – measures of disease occurrence

- Incidence – how many people developed the disease?
  - Cumulative incidence – what proportion of the population got the disease?
  - Incidence rate – how many get the disease per unit time?
- Prevalence – how many have the disease?
  - Point prevalence – … at a certain time point?
  - Period prevalence – … over a certain time period?

Objective

- Compare disease frequencies in different populations (or different groups in the same population)
- Summarize this comparison in one measure signifying the strength of an association between an exposure and an outcome
  - Absolute or relative measures
- Interpreting this measure

Definitions

- Absolute measures
  - Difference of two measures (i.e. subtract one measure of association from another)
- Relative measures
  - Ratio of two measures (i.e. divide one measure by another)
- Choice of association measure depends on type of study and objective

Relative measures (relative risks)

- Relative risks measure the strength of an association between a given exposure and an outcome
  - Large relative risks implies a dramatic effect of the exposure in question on the risk of the outcome and vice versa
  - Large relative risks say nothing about the importance of the exposure (e.g. Asbestos and mesothelioma in the 20th century)

Relative measures (2)

- For count data:
  - Cumulative incidence ratio (CIR) = risk ratios = relative risks (discouraged)
  - Odds ratios (OR)
- For person-time data:
  - Incidence rate ratios (IRR) = rate ratios = relative rate
  - Hazard ratios = approximate IRRs calculated primarily using Cox regression

Relative measures (3)

- For person-time data, we can also calculate standardized incidence ratios (SIRs).
- SIRs are the relative risk of a certain disease comparing our study population to the background population.
- SIRs are commonly used in occupational epi when special populations are investigated
- *Note: requires access to detailed background rates*

Relative measures (4)

- A relative risk of X is interpreted as:
  - Those who are exposed have a risk of the outcome that is X times as high as the unexposed
- A relative risk of 1.00 is interpreted as no association between the exposure and the outcome
- A relative risk has no unit of measure
- The range of relative risks is from 0 to ∞

Epi tool nr 1: The 2-by-2 table

Exposure \ Outcome	+	-	
+	a	b	a + b
-	c	d	c + d

Relative risks – general form

$$RR = \frac{Risk_{\text{exposed}}}{Risk_{\text{unexposed}}}$$

- Where risk signifies cumulative incidence, incidence rate, odds, etc.

Relative risks – specifics

Cumulative incidence ratio

$$CIR = \frac{CI_{\text{exposed}}}{CI_{\text{unexposed}}} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

Exposure \ Outcome	+	-
+	a	b
-	c	d

Relative risks – specifics (2)

Incidence rate ratio

$$IRR = \frac{IR_{\text{exposed}}}{IR_{\text{unexposed}}} = \frac{\frac{a}{pyrs_{\text{exposed}}}}{\frac{c}{pyrs_{\text{unexposed}}}}$$

Examples – asbestos

- The incidence of mesothelioma is approximately 1 per 1,000,000 person years
- Among heavily exposed asbestos workers it is higher, approximately 50-100 per 1,000,000 person years
- What is the relative risk?

$$RR = \frac{Risk_{\text{exposed}}}{Risk_{\text{unexposed}}}$$

$$IRR = \frac{50-100/1{,}000{,}000}{1/1{,}000{,}000}$$

$$IRR = 50-100$$

Examples – infectious disease

- What is the relative risk of rabies in relation to infection with the rabies virus?
- Say, as a numeric example, that of 200 patients bitten by wild dogs, 50% were infected and 50% were not…

Rabies virus	Rabies +	Rabies -	
+	90	10	100
-	?	?	100

Absolute measures

- Describe differences in risk between populations
- Typically give information about the public health impact of an exposure
- Examples
  - Cumulative incidence difference (aka. attributable risk or risk difference)
  - Incidence rate difference (aka. attributable rate or rate difference)

Odds ratios – brief intro

- As an alternative to risk ratios (which are based on proportions), we can also calculate ratios of odds – odds ratios (OR).

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

Exposure \ Outcome	+	-	
+	a	b	a+b
-	c	d	c+d

Odds ratios (2)

- Odds ratios approximate relative risks
- They are typically calculated from case-control studies (come Thursday!), but can also be derived using cohort data
- Rare disease assumption…
  - The more common the disease, the further from relative risks odds ratios deviate
  - Often a break point of 10% is used

ABSOLUTE RISKS

Absolute measures – general form

$$RD = Risk_{\text{exposed}} - Risk_{\text{unexposed}}$$

- Where RD is the risk (or rate) difference
- And $Risk_{exposed}$ and $Risk_{unexposed}$ is the risk (or rate) among the exposed and unexposed, respectively

Absolute measures – specifics

- Cumulative incidence difference:

$$CID = CI_{\text{exposed}} - CI_{\text{unexposed}} = \frac{a}{a+b} - \frac{c}{c+d}$$

- Incidence rate difference:

$$IRD = IR_{\text{exp}} - IR_{\text{unexp}} = \frac{a}{pyrs_{\text{exposed}}} - \frac{c}{pyrs_{\text{unexposed}}}$$

Absolute measures – specifics (2)

- The risk difference (CID), which is a difference of proportions can take any value between +1 and -1
- Rate differences (IRD) can take any value between -∞ and +∞
- Negative values mean a protective effect of the exposure
- What is a null finding?

Absolute measures – specifics (3)

- Risk (rate) differences are often interpreted as attributable risks (rates):
  - Signifying, what proportion of the risk (incidence) can be attributed to a certain exposure
  - Assumes causal association between the exposure and the outcome

Examples – risk difference

RR =(40/40000)/(4/40000)
=10

- Interpretation?

RD=40/40000 – 4/40000
=10/10000 – 1/10000
=9/10000 = 0.09%

- Interpretation?

Smoking	Lung cancer +	Lung cancer -	
+	40	39,960	40,000
-	4	39,996	40,000
			80,000

Examples – risk difference (2)

RR =(50/1,000,000)/ (1/1,000,000) =50

RD=50/1,000,000 – 1/1,000,000 =49/1,000,000

- Interpretation?

Asbestos	Mesothelioma +	Mesothelioma -	
+	50	999,950	1,000,000
-	1	999,999	1,000,000
			2,000,000

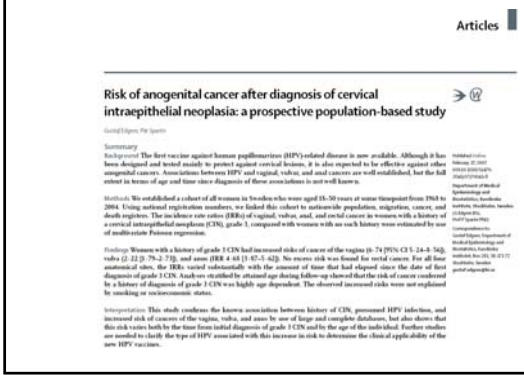
Numbers needed to treat?

- Using the cumulative incidence difference, we can calculate the "numbers needed to treat" (NNT)

$$NNT = \frac{1}{CID}$$

- NNT gives us the number of individuals which would have to experience the exposure of interest (or treatment) for 1 outcome to occur

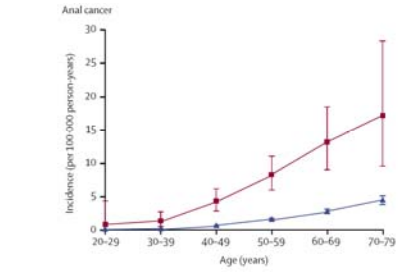
Absolute vs. relative risks



Articles

Risk of anogenital cancer after diagnosis of cervical intraepithelial neoplasia: a prospective population-based study

Absolute vs. relative risks



CIN3 history	0·79	1·35	4·32	8·30	13·21	17·20
No CIN3 history	0·03	0·17	0·66	1·58	2·79	4·48
Risk difference	0·76	1·18	3·66	6·72	10·42	12·72

Absolute risks – extension

- Risk differences are often expressed as percentages: Attributable percent among the exposed:

$$AP_{\text{exposed}} = \frac{RD}{R_{\text{exposed}}} \times 100 = \frac{R_{\text{exposed}} - R_{\text{unexposed}}}{R_{\text{exposed}}} \times 100$$

- Signifies the proportion of cases among the exposed that is due to the exposure
- Key point: assumes causal relationship
- Often referred to as etiologic fraction

Attributable percent – example

AP_{exp} $=100\times RD/R_{exp}$
$=100\times(49/1{,}000{,}000)/(50/1{,}000{,}000)$
$=98\%$

- Interpretation?
  98% of mesothelioma cases among those exposed to asbestos arise from asbestos exposure

		Mesothelioma +	Mesothelioma -	
Asbestos	+	50	999,950	1,000,000
	-	1	999,999	1,000,000
				2,000,000

Absolute risks – extension (2)

- In a further extension, we can calculate the attributable percent in the total population – the proportion of all cases in the population due to the exposure:

$$AP_t = \frac{PRD}{R_{total}} \times 100 = \frac{R_{total} - R_{unexposed}}{R_{total}} \times 100$$

- Where R_{total} is the risk in the total population and PRD (population risk difference) is the difference between R_{total} and $R_{unexposed}$

Attributable percent – example

$AP_t = 100 \times PRD / R_{total}$
$= 100 \times (44/80{,}000 - 4/40{,}000)/44/80{,}000$
$= 81.8\%$

- Interpretation?
  82% of all lung cancer cases in the population arise du to exposure to smoking

		Lung cancer +	Lung cancer -	
Smoking	+	40	39,960	40,000
	-	4	39,996	40,000
				80,000

Summary – general

- Getting the numbers right is the most important step in understanding the data!
  1. What exposure and outcome are you interested in?
  2. Summarize the data by exposure and outcome
  3. Calculate basic measures of disease occurrence and fitting association measures
  4. Interpret the results!

Summary – specifics

- Relative risks are often used to investigate etiology:
  - Does the exposure cause the disease?
- Absolute risks are often used from a public health perspective
  - How important is this exposure for the total disease burden? How much of the disease is caused by the exposure?

Summary – specifics (2)

$$RR = \frac{Risk_{\text{exposed}}}{Risk_{\text{unexposed}}}$$

$$RD = Risk_{\text{exposed}} - Risk_{\text{unexposed}}$$