

# Measurements in Epidemiology

## Số đo trong dịch tễ học

Ho Chi Minh City, 16/10/2021

1

### Keywords

- Số đếm (count) Tần số (Frequency)
- Tỷ số (Ratio) Tỷ suất (Rate)
- Tỷ lệ % (Proportion percentage)
- Tỷ lệ mắc mới (Incidence) Tỷ lệ lưu hành (Prevalence)
- Tỷ số chênh (Odd ratio) Tỷ số nguy cơ (Risk ratio)

2

## Dẫn nhập

- Dịch tễ học nghiên cứu **phân bố** bệnh tật trong quần thể và các **yếu tố ảnh hưởng** đến bệnh tật
- Định nghĩa dẫn đến 2 câu hỏi:
  - Làm thế nào để đo lường bệnh tật và sự phân bố của bệnh tật?
  - Làm thế nào để đo lường sự kết hợp về sự xảy ra bệnh ở các quần thể khác nhau

3

## Nội dung trình bày

- I. Đo lường tần số
- II. Đo lường sự kết hợp



**Alden Henderson**

Active 11 h ago



26 Jul 2021, 21:56



Thank you for using count compare and communicate. Very simple words but complicated to do. We must communicate our studies to others

4

## Topic 1

### Measures of frequency

### Đo lường tần số

5

### Đo lường tần số

---

#### 1. Các phép đo tần số bệnh tật bằng toán học

##### 1.1. Đếm / tần số

##### 1.2. Phân số

- Tỷ số                      Ratio
- Tỷ suất                  Rate
- Tỷ lệ (%)                Proportion (percentage)

#### 2. Đo lường tần số bệnh tật trong dịch tễ học

- Tỷ lệ mắc mới
- Tỷ lệ hiện mắc

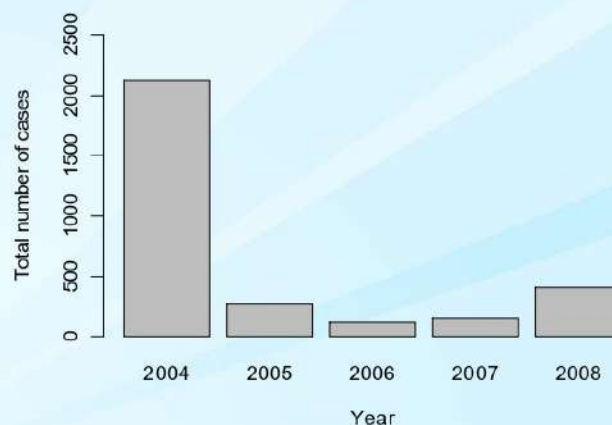
6

### 1.1. Số đếm

- Số đo cơ bản và đơn giản nhất – con số tuyệt đối của người mắc bệnh hay đặc điểm mà dịch tễ quan tâm
- Đếm số bệnh nhân mắc rất hữu ích cho công tác giám sát bệnh dịch để phát hiện sớm vụ dịch
- Hữu ích cho công tác hành chính và lập kế hoạch y tế
  - ✓ Số gói ORS cần để điều trị tiêu chảy trong 1 tuần/tháng
  - ✓ Số giường bệnh trong một tỉnh

7

### 1.1. Số đếm – Ví dụ



Pham HV, Doan HT, Phan TT, Minh NN.  
Ecological factors associated with dengue fever in a Central Highlands province, Vietnam. BMC Infect Dis. 2011;11:172

8

### 1.1. Số đếm – Ví dụ

Tỉnh/TP	Tổng số ca	Hôm nay	Tử vong
TP. Hồ Chí Minh	194.108	+3.934	7.810
Bình Dương	86.050	+4.868	716
Đồng Nai	20.471	+743	157

- Tỉnh/thành nào nghiêm trọng hơn ?
- Phụ thuộc vào:
  - Dân số nguy cơ
  - Phương pháp báo cáo
  - Định nghĩa ca bệnh

<https://ncov.moh.gov.vn/vi/web/guest/trang-chu> - 26/08/2021

9

### 1.1. Số đếm – hạn chế

- Số đếm của một bệnh (ví dụ số mắc COVID-19, số mắc SXH) phụ thuộc vào **kích thước quần thể** có nguy cơ mắc bệnh trong **một địa phương**
  - Số người của quần thể có nguy cơ càng cao thì kỳ vọng số mắc càng cao
- **Thời gian quan sát** cũng ảnh hưởng đến tần số mắc bệnh. Thời gian quan sát càng dài thì số BN ghi nhận càng nhiều

**Số đếm không có thông tin của những nội dung trên!**

10



## 1.2. Phân số

- Có 3 số toán học thường sử dụng:
  - Tỷ số (ratio)
  - Tỷ suất (rate)
  - Tỷ lệ phần trăm (percentage)
- Thể hiện mối liên quan giữa số BN hay 1 kết quả (tỷ số) với kích thước của quần thể mà số BN xảy ra (mẫu số)
- Tỷ số và mẫu số cần được định nghĩa cụ thể
- Dịch tễ học luôn đề cập đến “số BN trong 1 quần thể + thời gian”

11

## 1.2. Phân số – Tỷ số (Ratio)

**Dictionary:** “The value obtained by dividing one quantity by another.”

- Là kết quả của 1 số chia cho 1 số (thương số)
- Tỷ số và mẫu số có thể độc lập hay có liên quan đến nhau
- “Tỷ số” là thuật ngữ bao gồm cả tỷ suất (rates) và tỷ lệ (proportions)
- Thường thể hiện là:  $\frac{x}{y}$
- Ví dụ: BMI, tỷ số nam/nữ

12

NHANES - National Health and Nutrition Examination Survey

CDC Centers for Disease Control and Prevention  
CDC 24/7: Saving Lives. Protecting People™

Search Search NCHS

National Center for Health Statistics

CDC - NCHS

National Health and Nutrition Examination Survey

About NHANES  
What's New  
Webinar  
Questionnaires, Datasets, and Related Documentation  
Survey Participants  
Biospecimen Program  
New Content and Proposal

**Survey Participants**  
If you were selected, learn more about participating

**Survey Data and Documentation**  
Access data, documentation & response rates

**Publications and Products**  
View health and nutrition reports & CDC Growth Charts

**Data Analysis Tutorials**  
Review step-by-step guidance on using NHANES data

13

**EXAMPLE: Calculating a Ratio — Different Categories of Same Variable**

Between 1971 and 1975, as part of the National Health and Nutrition Examination Survey (NHANES), 7,381 persons ages 40–77 years were enrolled in a follow-up study.<sup>1</sup> At the time of enrollment, each study participant was classified as having or not having diabetes. During 1982–1984, enrollees were documented either to have died or were still alive. The results are summarized as follows.

	Original Enrollment (1971–1975)	Dead at Follow-Up (1982–1984)
Diabetic men	189	100
Nondiabetic men	3,151	811
Diabetic women	218	72
Nondiabetic women	3,823	511

Of the men enrolled in the NHANES follow-up study, 3,151 were nondiabetic and 189 were diabetic. Calculate the ratio of

**Tỷ số không tiểu đường / tiểu đường  
Nam = 3.151/189 và Nữ = 3.823/218**

14

## 1.2. Phân số - Tỷ lệ (Proportion)

- Tỷ lệ: Tử số (a) bao gồm trong mẫu số (a+b)
- Thường thể hiện  $\frac{a}{a+b} \times 10^n$  ( $10^n$ , thường là tỷ lệ %)
- Ví dụ" Số trẻ tử vong sơ sinh trong tổng số trẻ sinh ra

$$\frac{\text{Số trẻ tử vong sơ sinh}}{\text{Tổng số trẻ sinh ra sống + trẻ tử vong sơ sinh}} \times 100$$

- Dictionary: ""Một loại tỷ lệ mà tử số bao gồm trong mẫu số"

Kleinbaum et al. ActivEpi

15

### EXAMPLE: Calculating a Proportion

**Example A:** Calculate the proportion of men in the NHANES follow-up study who were diabetics.

$$\begin{aligned} \text{Tử số} &= 180 \text{ nam giới tiểu đường} \\ \text{Mẫu số} &= \text{tổng số nam} = 189 + 3.151 = 3.340 \\ \text{Tỷ lệ} &= (189 / 3.340) \times 100 = 5,66\% \end{aligned}$$

**Example B:** Calculate the proportion of deaths among men.

$$\begin{aligned} \text{Numerator} &= \text{deaths in men} \\ &= 100 \text{ deaths in diabetic men} + 811 \text{ deaths in nondiabetic men} \\ &= 911 \text{ deaths in men} \end{aligned}$$

Notice that the numerator (911 deaths in men) is a subset of the denominator.

$$\begin{aligned} \text{Denominator} &= \text{all deaths} \\ &= 911 \text{ deaths in men} + 72 \text{ deaths in diabetic women} + 511 \text{ deaths in nondiabetic women} \\ &= 1,494 \text{ deaths} \end{aligned}$$

$$\text{Proportion} = 911 / 1,494 = 60.98\% = 61\%$$

16



## 1.2. Phân số - Tỷ suất (Rate)

- Số đo mức độ nhanh chậm mà vấn đề quan tâm xảy ra

- Thường thể hiện:

$$\frac{x}{y} \times 10^n$$

- VD: Số mắc mới Parkinson's trên 1,000 năm - người theo dõi

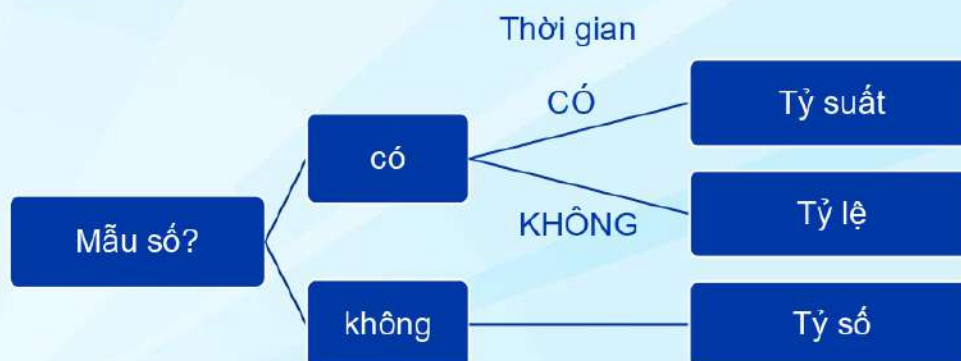
$$\frac{\text{Số mắc mới Parkinson's}}{\text{Tổng thời gian quan sát đối tượng không mắc bệnh}} \times 1000$$

- Thời gian, địa điểm và quần thể phải được nêu ra trong mỗi loại tỷ suất
- Trong tỷ suất thì tử số không bao gồm trong mẫu số

Kleinbaum et al. ActivEpi

17

## Thuật toán để phân biệt tỷ số, tỷ lệ và tỷ suất



18

## 2.1. Đo lường tần số bệnh tật – Tỷ lệ mắc mới

- Đo lường số mắc mới của 1 bệnh trong 1 khoảng thời gian
  - Hữu ích trong suy luận nhân/quả
  - Khó ước tính
  - Theo dõi theo thời gian (nghiên cứu cohort)
- Dictionary: “The number of new health-related events in a defined population within a specified period of time. May be measured as a frequency count, a rate or a proportion”

Kleinbaum et al. ActivEpi

19

## 2.1. Đo lường tần số bệnh tật – Tỷ lệ mắc mới

$\frac{\text{Số người mắc bệnh trong một khoảng thời gian nhất định}}{\text{Số người trong quần thể ở đầu kỳ}}$
---

- Một tỷ lệ
- Giá trị: 0 - 1

20

## 2.1. Đo lường tần số bệnh tật – Tỷ lệ mắc mới

- Ví dụ
  - Ở 1 địa phương, cứ 1000 người có 4 BN mắc bệnh A trong 1 năm:  $I = 4/1000 = 0.004$  or 0.4%
  - Ở 1 địa phương có 10.000 người, ghi nhận 11 người mắc bệnh B trong 1 năm:  $I = 11/10.000 = 0.0011$  or 0,11%
  - In a skin-cancer free population of 10,000 there are 11 new cases of skin cancer:  $I = 11/10,000 = 0.0011$  or 0.11%
- In a county with 2000 children within pre-school age there have occurred 20 new cases of leukemia within 10 years.
- Incidence (Cumulative incidence)

21

## 2.1. Đo lường tần số bệnh tật – Tỷ lệ hiện mắc

- Đo lường số trường hợp **bệnh hiện có** tại một **thời điểm** hoặc trong một **khoảng thời gian** cụ thể
- Rất hữu ích để định lượng gánh nặng bệnh tật (ví dụ: sức khỏe cộng đồng)
- Tương đối dễ ước tính
- Thiết kế nghiên cứu cắt ngang
- **Dictionary:** "A measure of disease occurrence: the total number of individuals who have an attribute or disease at a particular time (or period) divided by the population at risk of having the disease at that time or midway through the period. It is a proportion, not a rate."

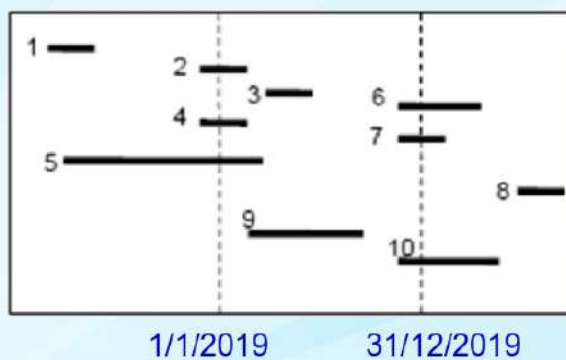
22

## 2.1. Đo lường tần số bệnh tật – Tỷ lệ hiện mắc

- Ví dụ:
  - Ở 1 địa phương có 1000 người, hiện có 4 BN mắc bệnh A trong 1 năm:  $\Rightarrow p = 4/1000 = 0.004$  or 0.4%
  - Trong cộng đồng với 10.000 người và hiện có 4 BN mắc ung thư da:  $\Rightarrow p = 4/10,000 = 0.0004$  or 0.04%
- Trong dịch tễ học, sự xuất hiện của bệnh thường **liên quan** đến quy mô dân số
- Trong một quận với 3.000 dân, ghi nhận xảy ra 3 trường hợp mắc bệnh bạch cầu. **Tỷ lệ hiện mắc?**

23

## 2.1. Đo lường tần số bệnh tật – Tỷ lệ hiện mắc

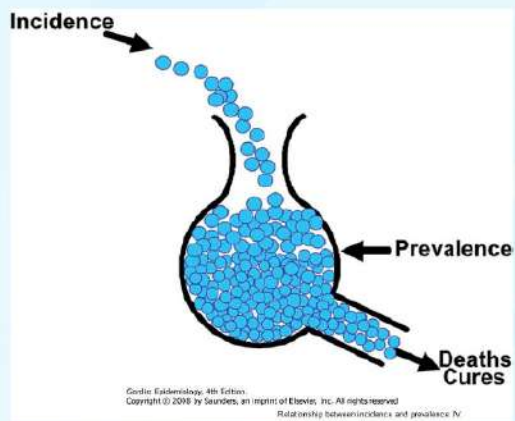


- Tỷ lệ hiện hành điểm (point prevalence):
- Tỷ lệ hiện hành khoảng thời gian (period prevalence)

24



### Tỷ lệ hiện mắc và tỷ lệ mắc mới



- Tỷ lệ hiện mắc (Prevalence)
- Tỷ lệ mắc mới (Incidence )
- Tử vong và khỏi bệnh

25

### Tỷ lệ hiện mắc và tỷ lệ mắc mới

- Chung
- Phân theo giới
- Phân theo trường
- Phân theo giới và trường

26



### Tỷ lệ hiện mắc và tỷ lệ mắc mới

- Một BV có 10 phòng đơn điều trị nội trú
- Mỗi ngày đều có 2 BV đến nhập viện
- Giả sử tất cả đều điều trị trong 5 ngày
- Tỷ lệ mắc mới ?
- Tỷ lệ hiện hành?

	d1	d2	d3	d4	d5
<b>Incidence</b>	2	2	2	2	2
<b>Prevalence</b>	2	4	6	8	10

27

## Topic 2

Measures of association  
Đo lường sự kết hợp

28

	Bệnh	Không bệnh	Tổng
Phơi nhiễm (+)	a	b	a+b
Phơi nhiễm (-)	c	d	c+d
Tổng	a+c	b+d	a+b+c+d

- |                                   |                     |
|-----------------------------------|---------------------|
| a. Phơi nhiễm và bệnh             | a+b: Phơi nhiễm     |
| b. Phơi nhiễm và KHÔNG bệnh       | c+d: NOT phơi nhiễm |
| c. KHÔNG phơi nhiễm và bệnh       | a+c: Bệnh           |
| d. KHÔNG phơi nhiễm và KHÔNG bệnh | b+d: NOT disease    |

- Nguy cơ (Risk): Khả năng xảy ra một sự kiện không mong muốn liên quan đến một mối nguy cụ thể và dẫn đến hậu quả
- Hậu quả (Outcome): Mức độ trầm trọng của một sự cố

### Nguy cơ (Risk)

- Khả năng xảy ra một sự kiện (mong muốn/không mong muốn) liên quan đến một mối nguy hiểm cụ thể và dẫn đến hậu quả



31

31

### Kiểm định Khi bình phương (Chi-square)

- Kiểm định khi bình phương để xem có sự kết hợp giữa 2 biến phân loại (phơi nhiễm và hậu quả)
- Giả thuyết ( $H_0$  và  $H_1$ )
- $H_0$ : Không có sự kết hợp giữa các biến số phơi nhiễm (hàng) & bệnh
- $H_1$ : Có sự kết hợp giữa các biến số phơi nhiễm (hàng) & bệnh

32

### Kiểm định Khi bình phương (Chi-square test)

$$X^2 = \sum \frac{(O-E)^2}{E}$$

- O: Số quan sát
- E: Số kỳ vọng
- Khi giá trị khi bình phương lớn hơn giá trị tới hạn hoặc  $p < 0,05 \Rightarrow$  có sự kết hợp
- Thống kê khi-bình phương chỉ cho biết có sự kết hợp hay không. Nó không cho biết mức độ mạnh của sự kết hợp

33

### Tỷ số nguy cơ (RR) – Khái niệm

- Nguy cơ xuất hiện bệnh ở nhóm phơi nhiễm (a + b) so với nhóm không phơi nhiễm (c + d)
- Tỷ số nguy cơ (**Risk Ratio**)

$$RR = \frac{\text{Tỷ số mắc mới (nguy cơ) ở nhóm phơi nhiễm}}{\text{Tỷ số mắc mới (nguy cơ) ở nhóm không phơi nhiễm}} = \frac{a/(a+b)}{c/(c+d)}$$

a. Phơi nhiễm và bệnh

a+b: Phơi nhiễm

b. Phơi nhiễm và KHÔNG bệnh

c+d: NOT phơi nhiễm

c. KHÔNG phơi nhiễm và bệnh

a+c: Bệnh

d. KHÔNG phơi nhiễm và KHÔNG bệnh

b+d: NOT disease

34

### Tỷ số nguy cơ (RR) – Khái niệm

- Theo dõi 2 nhóm đối tượng ( $n_1$  và  $n_2$ ) sau 1 thời gian
- Nhóm 1 phơi nhiễm với 1 yếu tố nguy cơ nào đó, có  $k$  người mắc bệnh ( $k_1$ )
- Tỷ lệ mắc bệnh ở nhóm 1:  $p_1 = \frac{k_1}{n_1}$
- Nhóm 2 không phơi nhiễm với 1 yếu tố đó, có  $k$  người mắc bệnh ( $k_2$ )
- Tỷ lệ mắc bệnh ở nhóm 2:  $p_2 = \frac{k_2}{n_2}$

$$RR = \frac{p_1}{p_2}$$

35

### Diễn giải RR

- $RR > 1$  ( $p_1 > p_2$ ): Yếu tố nguy cơ **tăng khả năng** mắc bệnh
- $RR = 1$  ( $p_1 = p_2$ ): Yếu tố nguy cơ **không liên quan** đến khả năng mắc bệnh
- $RR < 1$  ( $p_1 < p_2$ ): Yếu tố nguy cơ **giảm khả năng** mắc bệnh

Risk in exposed =  $RR \times$  Risk in non-exposed

36



### RR - Example

- Tìm hiểu mối liên quan giữa ung thư và chất độc A
- 11.000 người chia thành 2 nhóm và được theo dõi trong 10 năm
- Nhóm 1: 1.000 người phơi nhiễm với chất A, có 20 người bị ung thư (2%)
- Nhóm 2: 10.000 người không phơi nhiễm với chất A, có 100 người bị ung thư (1%)

37

### RR - Example

	Ung thư	Không ung thư	Tổng số
Phơi nhiễm	20 (2%)	980	1,000
Không phơi nhiễm	100 (1%)	9900	10,000

$$RR = \frac{p_1}{p_2} = \frac{0,02}{0,01} = 2$$

- Nguy cơ xảy ra ung thư ở nhóm phơi nhiễm với chất độc A cao gấp 2 lần so với không phơi nhiễm

38



## Knowledge Check

- Một nghiên cứu cohort về mối liên quan giữa hút thuốc và ung thư phổi ghi nhận RR=17. Điều nào sau đây là cách giải thích tốt nhất về RR này?

  1. Nhóm hút thuốc bị ung thư phổi nhiều hơn 17% so với nhóm không hút thuốc
  2. Nhóm hút thuốc có nguy cơ mắc ung thư phổi cao gấp 17 lần so với nhóm không hút thuốc
  3. 17% trường hợp ung thư phổi ở những người hút thuốc là do hút thuốc
  4. Có thêm 17 trường hợp ung thư phổi ở nhóm hút thuốc

39

## Khái niệm odd

- Odd là tỉ số của 2 xác suất
- Nếu p là xác suất mắc bệnh thì 1-p là xác suất không mắc bệnh

$$odd = \frac{p}{1 - p}$$

- odd > 1: khả năng mắc bệnh cao hơn
- odd = 1: khả năng mắc bệnh bằng không mắc bệnh
- odd < 1: khả năng không mắc bệnh cao hơn

40

### Tính toán OR

- Odd mắc bệnh ở nhóm phơi nhiễm (Odds of disease among exposed)  $odd_1 = \frac{p_1}{1 - p_1}$
- Odd mắc bệnh ở nhóm không phơi nhiễm (Odds of disease among non-exposed)  $odd_2 = \frac{p_2}{1 - p_2}$

$$OR = \frac{odd_1}{odd_2} = \frac{p_1/(1 - p_1)}{p_2/(1 - p_2)}$$

41

### Odds ratio (OR)

- RR có thể ước tính tốt từ OR khi (*can be best estimated by OR if the following conditions are fulfilled*)
- Nhóm chứng đại diện cho quần thể (*Controls are representative of general population*)
- Nhóm bệnh đại diện cho tất cả bệnh nhân (*Selected cases are representative of all cases*)
- Bệnh hiếm gặp (*The disease is rare*)

42

### Odds ratio (OR)

- RR chỉ có thể tính trong các nghiên cứu theo thời gian (cohort)
- OR có thể tính cho tất cả các loại nhưng chủ yếu là nghiên cứu bệnh chứng (case - control)

43

### OR – ví dụ

- Tìm hiểu mối liên quan giữa ung thư và chất độc A
- Chọn 100 BN ung thư và 100 không ung thư,
  - Nhóm 1: 100 BN có người 10 người phơi nhiễm với chất A
  - Nhóm 2: 100 người không ung thư có 5 người phơi nhiễm với chất A

44

### Tính OR

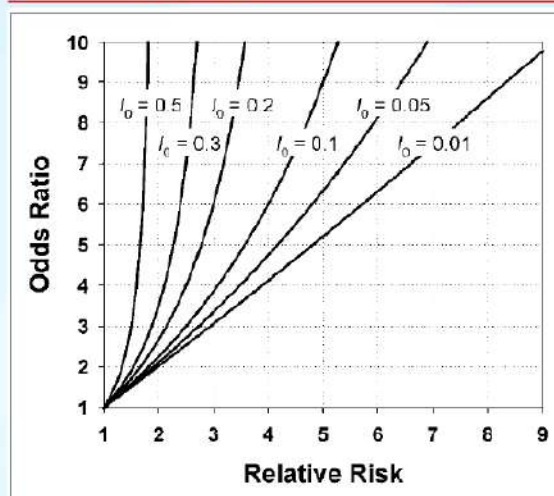
	Ung thư	Không ung thư
Phơi nhiễm	10	5
Không phơi nhiễm	90	95
Tổng số	100	100

$$OR = \frac{10/90}{5/95} = \frac{0,111}{0,052} = 2,11$$

Khả năng xảy ra ung thư ở nhóm phơi nhiễm với chất A cao gấp 2,11 lần so với không phơi nhiễm

45

### Mối liên quan giữa OR và RR theo tỷ lệ mắc mới



Carsten Oliver Schmidt, Thomas Kohlmann. When to use the odds ratio or the relative risk?  
*Int J Public Health* 53 (2008) 165–167

46



Tỉ lệ (nguy cơ) mắc bệnh		Odds mắc bệnh		So sánh 2/1	
Nhóm 1 ( $p_1$ )	Nhóm 2 ( $p_2$ )	Nhóm 1 (odds <sub>1</sub> )	Nhóm 2 (odds <sub>2</sub> )	RR	OR
0,001	0,003	0,001	0,003	3	3,01
0,01	0,03	0,01	0,03	3	3,06
0,02	0,06	0,02	0,06	3	3,13
0,05	0,15	0,05	0,18	3	3,35
0,10	0,30	0,11	0,43	3	3,86
0,15	0,45	0,18	0,82	3	4,64
0,20	0,60	0,25	1,50	3	6,00
0,25	0,75	0,33	3,00	3	9,00
0,30	0,90	0,43	9,00	3	21,0
0,33	0,99	0,49	99,0	3	2101

47

	Bệnh	Không bệnh	Tổng số
Phơi nhiễm	a	b	a+b
Không phơi nhiễm	c	d	c+d
Tổng số	a+c	b+d	a+b+c+d

$$RR = \frac{p_1}{p_2} = \frac{a/(a+b)}{c/(c+d)} \quad OR = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} = \frac{a/b}{c/d}$$

48

Formal display

		Disease	
		yes	no
Exposed	yes	a	b
	no	c	d

Prevalence<sup>1a)</sup> / Incidence<sup>1b)</sup> among exposed,  $I_1 = \frac{a}{a+b}$  and unexposed  $I_0 = \frac{c}{c+d}$

Relative Risk (RR) und aprocimation to Odds Ratio (OR) under the rare disease assumption

$$RR = \frac{I_1}{I_0} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} \xrightarrow{\text{if } a, c \text{ small}} \approx \frac{\frac{a}{b}}{\frac{c}{d}} = OR$$

Conversion of Odds Ratio into Risk Ratio (Holland, 1989; Zhang und Yu 1998)

$$RR = \frac{OR}{1 - I_0 + I_0 * OR} \quad \text{alternatively } RR = OR * \frac{1 - I_1}{1 - I_0}$$

<sup>a)</sup> using cross-sectional data; <sup>b)</sup> using longitudinal data

Carsten Oliver Schmidt, Thomas Kohlmann. When to use the odds ratio or the relative risk? *Int J Public Health* 53 (2008) 165–167

49

### RR and OR

- RR và OR là 2 chỉ số dịch tễ học phản ánh mối liên hệ giữa 1 yếu tố nguy cơ và bệnh tật
- RR là giá trị cần tìm
- OR là ước tính của RR
- Khi tỷ lệ mắc bệnh cao (10%) thì OR thường cao hơn RR (ước tính quá mức)

50

50