

Bài giảng 11b:

So sánh 2 tỉ lệ bằng odds ratio

Nguyễn Văn Tuấn

Garvan Institute of Medical Research, Australia
Đại học Tôn Đức Thắng, Việt Nam

Khái niệm "risk" và "odds"

Định nghĩa *risk*

- Xác suất một biến cố bất lợi xảy ra trong một thời gian nhất định
- Ba yếu tố:
 - xác suất
 - biến cố bất lợi (adverse event)
 - thời gian

Định nghĩa *odds*

- The odds of an event is the number of times that the event has occurred over the the number of times that the event has not occurred
- Odds của một biến cố là **tỉ số** của *số lần biến cố đó xảy ra* trên *số lần biến cố đó không xảy ra*.

So sánh giữa *risk* và *odds*

Một nhóm bệnh nhân gồm **100** người được điều trị. Trong thời gian 3 năm sau đó, có 5 người tử vong

Nguy cơ (risk) tử vong:

$$p = 5 / 100 = 0.05$$

Odds tử vong:

$$\text{odds} = 5 / 95 = 0.053$$

Odds ratio

Jerome Cornfield

- Nhà thống kê học, Johns Hopkins Univ
- Chủ tịch American Statistical Association (1974)
- R A Lecturer (1973)
- Phát kiến khái niệm *odds ratio*



J Cornfield (1912-1979)

Nguy cơ tử vong của 2 nhóm

- Nghiên cứu lâm sàng đối chứng ngẫu nhiên (Randomized controlled trial)
- Hai nhóm bệnh nhân, theo dõi 3 năm
- Tiêu chí đánh giá: tử vong

	Placebo	Rx
Số bệnh nhân	1062	1065
Tử vong	141	101

Nguy cơ

	Rx	Placebo
Số bệnh nhân	1065	1062
Tử vong	101	141
Sống sót	964	921
Nguy cơ tử vong	0.095	0.133
Odds tử vong	0.105	0.153

- Câu hỏi nghiên cứu: Thuốc có hiệu quả giảm tử vong?
- Câu hỏi thống kê: nguy cơ tử vong khác nhau?

Trường hợp chung

	Rx	Control
Bệnh	<i>a</i>	<i>b</i>
Không bệnh	<i>c</i>	<i>d</i>
Cỡ mẫu	N_1	N_2

Nguy cơ

$$p_1 = \frac{a}{N_1}$$

$$p_2 = \frac{b}{N_2}$$

Odds

$$O_1 = \frac{a}{c}$$

$$O_2 = \frac{b}{d}$$

Hai cách để so sánh

- Hiệu số (absolute risk difference)

$$d = p_1 - p_2$$

- Tỷ số hai odds (**Odds Ratio**)

$$OR = \frac{O_1}{O_2}$$

Ý nghĩa của OR

- O_1 : odds nhóm *được điều trị*, O_2 odds nhóm *không điều trị*

$$OR = O_1 / O_2$$

- Ý nghĩa:
 - Nếu $OR = 1$, hai nhóm có hiệu quả như nhau
 - Nếu $OR > 1$, thuốc có hại
 - Nếu $OR < 1$, thuốc có lợi

Ước tính khoảng tin cậy 95% của OR

- Nhưng OR chỉ là số *trung bình*
- Nếu lặp lại nghiên cứu, OR có thể khác
- Lặp lại lần nữa, OR lại khác
- Nếu lặp lại 100 lần, và nếu tất cả hay 95% OR đều dưới 1 thì đó là chứng cứ khoa học
- Vấn đề: làm sao ước tính khoảng tin cậy 95%?

Qui trình ước tính khoảng tin cậy

- Cần phải ước tính phương sai của OR
- Sau đó, ước tính độ lệch chuẩn (căn số bậc 2 của phương sai)
- Áp dụng lí thuyết phân bố chuẩn: khoảng tin cậy 95% bằng

trung bình $\pm 1.96 \times$ độ lệch chuẩn

Ước tính phương sai của OR

- OR là một tỉ số
- Rất khó ước tính phương sai của một tỉ số!

Tính phương sai của $\log(\text{OR})$

Cách tính "gián tiếp" hay "vòng"

1. Hoán chuyển OR sang $\log(\text{OR})$
2. Tính phương sai (và độ lệch chuẩn) của $\log(\text{OR})$
3. Tính khoảng tin cậy 95% của $\log(\text{OR})$
4. Hoán chuyển khoảng tin cậy $\log(\text{OR})$ sang OR

Ví dụ cách tính khoảng tin cậy 95%

	Rx	Control
Bệnh	a	b
Không bệnh	c	d
Cỡ mẫu	N_1	N_2

Odds

$$O_1 = \frac{a}{c}$$

$$O_2 = \frac{b}{d}$$

- $OR = O_1 / O_2$
- $L = \log(OR) = \log(O_1) - \log(O_2)$

Ví dụ cách tính khoảng tin cậy 95%

	Rx	Control
Bệnh	a	b
Không bệnh	c	d
Cỡ mẫu	N_1	N_2

Odds

$$O_1 = \frac{a}{c}$$

$$O_2 = \frac{b}{d}$$

- $OR = O_1 / O_2$
- $L = \log(OR) = \log(O_1) - \log(O_2)$
- Phương sai của L:

$$V = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

Ví dụ cách tính khoảng tin cậy 95%

- $OR = O_1 / O_2$
- $L = \log(OR) = \log(O_1) - \log(O_2)$

- Phương sai của L:

$$V = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

- Độ lệch chuẩn của L: $S = V^{1/2}$
- Khoảng tin cậy 95% của L: $L \pm 1.96xS$
- Khoảng tin cậy 95% của OR:
 $\exp(L-1.96xS)$ đến $\exp(L+1.96xS)$

Phân tích OR với R

Tính toán "thủ công"

	Rx	Placebo
Tử vong	101	141
Sống sót	964	921

- $OR = odds1 / odds2 = (101*921) / (141*964)$
- $L = \log(OR)$
- $V = 1/101 + 1/964 + 1/141 + 1/921$
- $S = \text{căn số bậc 2 của } V$
- KTC95: $L - 1.96S$ đến $L + 1.96S$
- KTC95% RR: $\exp(L - 1.96S)$ đến $\exp(L + 1.96S)$

```
a=101; b=141; c=964; d=921
OR = (a*d) / (b*c)
L = log(OR)
S = sqrt(1/a+1/b+1/c+1/d)
L95 = exp(L-1.96*S)
U95 = exp(L+1.96*S)
cbind(OR, L95, U95)
```

Kết quả tính toán thủ công

```
> cbind(OR, L95, U95)
```

OR

L95

U95

```
[1,] 0.6843604 0.5219082 0.8973784
```

Dùng package **epitools**

```
library(epitools)
```

```
data = matrix(c(101, 964, 141, 921), byrow=T,  
nrow=2)
```

```
oddsratio(data, rev="c")
```

Dùng package **epitools**

```
> oddsratio(data, rev="c")
```

Outcome

Predictor	Disease2	Disease1	Total
Exposed1	964	101	1065
Exposed2	921	141	1062
Total	1885	242	2127

\$measure

odds ratio with 95% C.I.

Predictor	estimate	lower	upper
Exposed1	1.000000	NA	NA
Exposed2	1.460349	1.114874	1.919114

\$p.value

two-sided

Predictor	midp.exact	fisher.exact	chi.square
Exposed1	NA	NA	NA
Exposed2	0.005904708	0.006268449	0.005874976

Tại sao OR = 1.46?

(Trước đó RR = 0.68)

Tóm lược

- OR = odds ratio (tỉ số odds)
- Một chỉ số đo lường độ ảnh hưởng (hay liên quan)
- Thường được sử dụng trong nghiên cứu bệnh chứng