

## Lâm sàng thống kê 23

### Kiểm định khác biệt giữa hai tỉ lệ

Nguyễn Văn Tuấn

*Hỏi: Xin hỏi thầy nếu chúng ta có 2 tỷ lệ mắc bệnh cho hai nhóm bệnh nhân, làm sao để biết sự khác biệt giữa hai tỷ lệ có ý nghĩa thống kê? Có thể sử dụng tỷ số chênh không? Trong bài trước, thầy có giải thích tỷ số chênh, nhưng không chỉ cách tính khoảng tin cậy 95%.*

Trong nhiều nghiên cứu lâm sàng, biến *outcome* -- tức là chỉ số hiệu quả lâm sàng -- thường chỉ có 2 giá trị (có hay không, mắc bệnh hay không mắc bệnh, sống hay tử vong, v.v...), và biến độc lập -- cũng có thể là yếu tố nguy cơ -- cũng chỉ có 2 giá trị (phơi nhiễm hay không phơi nhiễm). Trong các nghiên cứu lâm sàng đối chứng ngẫu nhiên (RCT), biến độc lập chính là nhóm bệnh nhân được điều trị và nhóm placebo. Trong các nghiên cứu như thế, kết quả nghiên cứu có thể tóm lược trong bảng số liệu với 2 cột và 2 dòng (tiếng Anh gọi là 2 x 2 contingency table).

Chẳng hạn như trong tai nạn tàu Titanic, số người tử vong có thể xem là *outcome*, và giới tính có thể xem là *biến độc lập*. Trong số 1313 hành khách, có 462 là nữ và 851 là nam. Trong số hành khách nữ, có 154 người chết; trong số nam, con số tử vong là 709 người. Chúng ta có thể tóm lược số liệu trên trong bảng 2x2 như sau:

<b>Bảng 1. Tử vong trong tai nạn tàu Titanic</b>		
	<b>Nữ</b>	<b>Nam</b>
Số bị chết	154	709
Số sống sót	308	142

Ở đây, chúng ta thấy tỉ lệ tử vong trong nhóm nữ có vẻ thấp hơn so với nhóm nam. (Có lẽ đàn ông lo cứu phụ nữ, và họ hi sinh trong nghĩa cử cao đẹp đó!) Câu hỏi đặt ra là sự khác biệt về tỉ lệ tử vong ở nam và nữ có ý nghĩa thống kê?

Một công trình nghiên cứu hiệu quả ngừa gãy xương của thuốc zoledronate, các nhà nghiên cứu chia bệnh nhân ra hai nhóm: nhóm 1 gồm 1065 bệnh nhân được điều trị bằng zoledronate, và nhóm 2 gồm 1062 bệnh nhân chỉ uống calcium và vitamin D (còn gọi là nhóm placebo). Sau 3 năm theo dõi, kết quả cho thấy 92 bệnh nhân nhóm 1 bị gãy

xương đốt sống, và 148 bệnh nhân trong nhóm placebo bị gãy xương đốt sống. Số liệu này có thể tóm lược trong một bảng như sau:

<b>Bảng 2. Kết quả điều trị bệnh nhân loãng xương với thuốc zoledronic acid</b>		
	<b>Zoledronic acid</b>	<b>Placebo</b>
Số bị gãy xương	92	148
Số không bị gãy xương	970	917

Chúng ta dễ dàng thấy trong nghiên cứu này, tỉ lệ gãy xương ở nhóm bệnh nhân được điều trị (8.6%) thấp hơn nhóm placebo (13.9%). Câu hỏi đặt ra là sự khác biệt này có ý nghĩa thống kê hay không?

Có ít nhất 3 cách trả lời câu hỏi loại này. Cách thứ nhất là ước tính tỉ lệ cho từng nhóm và so sánh bằng cách sử dụng kiểm định phân phối nhị phân (binomial test); cách thứ hai là tính *odds ratio* (OR); và cách thứ ba là tính *risk ratio* (RR). Tôi đã giải thích về OR và RR trong một bài trước; ở đây tôi sẽ tập trung vào cách tính khoảng tin cậy 95% cho OR và RR.

### **Kiểm định nhị phân (cho hai tỉ lệ)**

Chúng ta quay lại với ví dụ 1 về tai nạn tàu Titanic. Gọi tỉ lệ tử vong trong nhóm nam nam là  $p_1$  và nữ là  $p_2$ . Chúng ta có thể ước tính xác suất tử vong cho hai nhóm như sau (bằng cách lấy số tử vong chia cho tổng số đối tượng):

	<b>Nữ</b>	<b>Nam</b>
Số bị chết	154	709
Số sống sót	308	142
<b>Tổng số đối tượng</b>	<b>462</b>	<b>851</b>
<b>Xác suất tử vong</b>	<b>0.333</b>	<b>0.833</b>

Như vậy  $p_1 = 0.33$  và  $p_2 = 0.833$ . Cách đơn giản nhất để ước tính khác biệt giữa hai nhóm là hiệu số. Gọi hiệu số là  $d$ , chúng ta có:

$$d = p_2 - p_1$$

$$= 0.833 - 0.333$$

$$= 0.50$$

Vấn đề đặt ra là tính phương sai cho  $d$ . Vì  $p_1$  và  $p_2$  độc lập, cho nên theo lí thuyết xác suất, phương sai của hiệu số bằng tổng phương sai của hai tỉ lệ. Phương sai của một tỉ lệ có thể tính dựa vào số cỡ mẫu. Gọi  $n$  là số cỡ mẫu và  $p$  là tỉ lệ thì phương sai của  $p$  (kí hiệu  $V$ ) có thể ước tính bằng công thức:

$$V = \frac{p(1-p)}{n}$$

Độ lệch chuẩn của  $p$  (kí hiệu bằng  $S$ ) chính là căn số bậc hai của  $V$ :  $S = \sqrt{p(1-p)/n}$ . Trong ví dụ trên, chúng ta có thể tính phương sai cho  $p_1$  và  $p_2$  như sau:

$$V_1 = \frac{p_1(1-p_1)}{n_1} = \frac{0.333(1-0.333)}{462} = 0.000481$$

$$V_2 = \frac{p_2(1-p_2)}{n_2} = \frac{0.833(1-0.833)}{851} = 0.000163$$

Theo lí thuyết xác suất, phương sai của hiệu số  $d$  chính là tổng phương sai của hai tỉ lệ. gọi phương sai của  $d$  là  $V_d$ , chúng ta có:

$$V_d = V_1 + V_2$$

$$= 0.000481 + 0.000163$$

$$= 0.000644$$

Và độ lệch chuẩn của hiệu số  $d$  là căn số bậc hai của  $V_d$ :

$$S_d = \sqrt{V_d}$$

$$= \sqrt{0.000644}$$

$$= 0.0255$$

Kiểm định nhị phân (bionomial test) có thể tính bằng cách lấy hiệu số  $d$  chia cho độ lệch chuẩn  $S_d$ :

$$\begin{aligned}
 z &= \frac{d}{S_d} \\
 &= 0.50 / 0.0255 \\
 &= 19.7
 \end{aligned}$$

Vì  $z$  cao hơn 1.96 (theo phân phối chuẩn), nên chúng ta có bằng chứng để kết luận rằng tỉ lệ tử vong ở nam cao hơn nữ, và sự khác biệt này có ý nghĩa thống kê. Có thể ước tính khoảng tin cậy 95% của hiệu số thật như sau:

$$\begin{aligned}
 95\% \text{ CI} &= d \mp 1.96 \times S_d \\
 &= 0.50 - 1.96 \times 0.0255 \text{ đến } 0.50 + 1.96 \times 0.0255 \\
 &= 0.450 \text{ đến } 0.550
 \end{aligned}$$

Nói cách khác, nếu nghiên cứu này lặp lại 100 lần, thì tỉ lệ tử vong ở nam cao hơn nữ dao động trong khoảng 45% đến 55%. (Cố nhiên, chẳng ai lặp lại tai nạn 100 lần, nhưng ở đây chúng ta nói chuyện ... lí thuyết).

## Risk ratio

Cách kiểm định như trình bày trên là lấy hiệu số của hai tỉ lệ và tính khoảng tin cậy 95% cho hiệu số thật. Một cách kiểm định khác là lấy tỉ số của hai tỉ lệ. Bởi vì tỉ lệ là xác suất (hay nguy cơ, *risk*), cho nên cách này còn có tên là *risk ratio* (hoặc *relative risk*) với kí hiệu thông thường là RR. Sử dụng kí hiệu  $p_1$  và  $p_2$  trên, định nghĩa vừa đề cập có thể tóm lược như sau:

Hiệu số:  $d = p_1 - p_2$

Tỉ số nguy cơ:  $RR = \frac{p_1}{p_2}$

Do đó, nếu  $RR = 1$ , thì hai nguy cơ bằng nhau ( $p_1 = p_2$ ); nếu  $RR > 1$  thì nguy cơ trong nhóm 1 cao hơn nhóm 2 ( $p_1 > p_2$ ); và nếu  $RR < 1$ , nguy cơ nhóm 1 thấp hơn nhóm 2 ( $p_1 < p_2$ ). Trong ví dụ 2, chúng ta có thể ước tính RR như sau:

### Chi tiết tính toán RR

	Zoledronic acid	Placebo
Số bị gãy xương	92	148
Số không bị gãy xương	970	917
Xác suất gãy xương	0.087	0.139
RR	0.087 / 0.139 = 0.626	

- Tỷ lệ (nguy cơ) gãy xương trong nhóm zoledronic acid là:  $92 / (92 + 970) = 0.087$ , hay 8.7%;
- Tỷ lệ (nguy cơ) gãy xương trong nhóm placebo là:  $148 / (148 + 917) = 0.139$ , hay 13.9%;
- Do đó,  $RR = 0.087 / 0.139 = 0.626$ . Nói cách khác, nguy cơ gãy xương trong nhóm điều trị bằng 62.6% so với nhóm placebo. Một cách hiểu khác là: thuốc zoledronate giảm nguy cơ gãy xương 37.4% (lấy 100 trừ cho 62.6).

Để biết  $RR$  có ý nghĩa thống kê hay không, chúng ta cần phải tính khoảng tin cậy 95%. Để tính khoảng tin cậy của  $RR$ , chúng ta cần phải tính độ lệch chuẩn. Nhưng để tính độ lệch chuẩn, chúng ta cần phải tính phương sai. Do đó, vấn đề đặt ra là tìm một phương pháp để ước tính phương sai của  $RR$ .

Trong xác suất học, tính phương sai cho một hiệu số (hay tổng số) rất đơn giản, nhưng tính phương sai của một tỉ số thì cực kỳ phức tạp và nhiều khi. Do đó, để tính phương sai cho  $RR$ , chúng ta cần hoán chuyển  $RR$  từ tỉ số sang hiệu số. Chúng ta biết

rằng cho hai số  $x_1$  và  $x_2$  thì  $\log\left(\frac{x_1}{x_2}\right) = \log(x_1) - \log(x_2)$ . Do đó, “vấn đề” trở nên đơn

giản hơn nếu chúng ta tính phương sai cho  $\log(RR)$ . Nói cách khác, thay vì tính phương sai cho  $RR$ , chúng ta tính phương sai cho  $\log RR$ , và sau đó hoán chuyển ngược lại cho  $RR$ . Có thể minh họa cách tính này qua 3 bước cụ thể như sau:

**Bước 1:** hoán chuyển  $RR$  sang đơn vị  $\log RR$ ; gọi kết quả là  $L$

$$\begin{aligned} L &= \log(RR) \\ &= \log(0.626) \\ &= -0.468 \end{aligned}$$

**Bước 2:** tính phương sai của  $L$ . Công thức tính phương sai cho  $L$  có tên là công thức Greenwood, và có thể tính dựa vào bảng số liệu trên như sau (kí hiệu phương sai là  $V$ ):

$$V = \frac{92/970}{92+970} + \frac{148/917}{148+917}$$

$$= 0.00024$$

Chú ý công thức trên cho thấy về thứ nhất của phương trình chính là phương sai của  $\log(p_1)$  nhóm 1, và về thứ hai là phương sai của  $\log(p_2)$ . Do đó, phương sai của  $\log(p_1) - \log(p_2)$  chính là tổng phương sai của hai nhóm.

Độ lệch chuẩn của  $\log(RR)$  do đó chỉ đơn giản là căn số bậc hai của phương sai  $V$ .

$$S = \sqrt{0.00024}$$

$$= 0.016$$

**Bước 3:** tính khoảng tin cậy 95% cho  $\log RR$ . Vì  $\log RR$  tuân theo luật phân phối chuẩn, cho nên chúng ta có thể ước tính khoảng tin cậy 95% cho  $\log RR$  như sau:  $L \mp 1.96 \times S$ , hay cụ thể hơn:

$$95\% \text{ CI } \log RR = -0.543 - 1.96 \times 0.016 \text{ đến } -0.543 + 1.96 \times 0.016$$

$$= -0.503 \text{ đến } -0.442$$

**Bước 4:** hoán chuyển sang đơn vị tỉ số. Chúng ta biết rằng [nhớ đại số học lớp 9?] nếu  $\log(x) = y$  thì  $x = e^y$ . Trong trường hợp trên chúng ta biết  $\log(x) = -0.503$ ; do đó, thì  $x = e^{-0.503} = 0.605$ . Ở đây  $x$  chính là phần thấp của khoảng tin cậy 95% cho  $RR$ . Tính tương tự chúng ta có phần cao của khoảng tin cậy 95% cho  $RR$  là 0.643.

$$95\% \text{ CI } RR = e^{-0.503} \text{ đến } e^{-0.442}$$

$$= 0.605 \text{ đến } 0.643$$

Đến đây thì chúng ta đã có đáp số:  $RR = 0.626$  với khoảng tin cậy 95% dao động từ 0.605 đến 0.643. Kết quả này có thể diễn giải như sau: điều trị bệnh nhân loãng xương bằng zoledronic acid có hiệu quả giảm nguy cơ gãy xương đốt sống 37.4%, và nếu nghiên cứu này lặp lại 100 lần, thì hiệu quả này có thể dao động trong vòng 35.7% đến 39.5%.

## Odds ratio

Thay vì tính RR, nhiều nhà nghiên cứu chọn cách tính odds ratio (OR). Để minh họa cho cách tính này, tôi sẽ sử dụng số liệu của nghiên cứu zoledronic acid trong ví dụ 2. Gọi  $p$  là xác suất một sự kiện, thì *odds* được định nghĩa là:

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

Nếu chúng ta có hai nhóm, thì sẽ có hai odds:  $odds_1$  và  $odds_2$ . “Tỉ số chênh” (mà một số người dịch – không hợp lí mấy) chính là tỉ số của hai odds:

$$OR = \frac{odds_1}{odds_2}$$

Quay lại ví dụ 2, chúng ta tính toán như sau. Vì  $p_1 = 0.087$  và  $p_2 = 0.139$ , cho nên odds gãy xương trong nhóm zoledronic acid là  $odds_1 = 0.087 / (1 - 0.087) = 0.095$ . Tương tự, odds gãy xương trong nhóm placebo là  $odds_2 = 0.139 / 0.861 = 0.161$ . Do đó,  $OR = 0.095 / 0.161 = 0.587$ .

Chi tiết tính toán OR và khoảng tin cậy 95%		
	Zoledronic acid	Placebo
Số bị gãy xương	92	148
Số không bị gãy xương	970	917
<b>Xác suất gãy xương</b>	<b>0.087</b>	<b>0.139</b>
<b>Odds gãy xương</b>	<b>0.095</b>	<b>0.161</b>
<b>OR</b>	<b>0.587</b>	

Chú ý rằng chúng ta có thể tính OR một cách đơn giản hơn bằng cách nhân chéo các phần tử số trong bảng 2x2 trên:

$$OR = \frac{92 \times 917}{148 \times 970} = 0.587$$

Để biết *OR* này có ý nghĩa thống kê hay không, chúng ta phải tính phương sai và độ lệch chuẩn cho *OR*. Nhưng vì *OR* là một tỉ số, và việc tính phương sai cho tỉ số cực kì phức tạp, nên chúng ta phải tính gián tiếp: đó là hoán chuyển sang đơn vị log và tính

phương sai cho log *OR*. Có thể mô tả bước tính này bằng các bước cụ thể như sau (tương tự như cách tính khoảng tin cậy 95% cho *RR*):

**Bước 1:** hoán chuyển *OR* sang đơn vị log:

$$\begin{aligned}\log OR &= \log(OR) \\ &= \log(0.587) \\ &= -0.532\end{aligned}$$

**Bước 2:** Ước tính phương sai cho logOR. Công thức phương sai cho logOR rất đơn giản vì chỉ là hàm số của các phần tử trong bảng 2x2 trên. Gọi phương sai là *V*, công thức của *V*, bốn số trong bảng 2x2 là *a*, *b*, *c*, và *d*, chúng ta có:

$$\begin{aligned}V &= \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \\ &= \frac{1}{92} + \frac{1}{148} + \frac{1}{970} + \frac{1}{917} \\ &= 0.0197\end{aligned}$$

Theo đó, độ lệch chuẩn của logOR (kié hiệu *S*) là căn số bậc hai của phương sai:

$$\begin{aligned}S &= \sqrt{0.0197} \\ &= 0.141\end{aligned}$$

**Bước 3:** Ước tính khoảng tin cậy 95% của logOR.

$$\begin{aligned}95\% \text{ CI logOR} &= \log OR \pm 1.96 \times S \\ &= -0.532 \pm 1.96 \times 0.141 \\ &= -0.807 \text{ đến } -0.256\end{aligned}$$

**Bước 4:** Hoán chuyển khoảng tin cậy 95% của logOR sang khoảng tin cậy 95% của *OR* (từ đơn vị log sang đơn vị tỉ số) bằng hàm số mũ:

$$\begin{aligned}95\% \text{ CI OR} &= e^{-0.807} \text{ đến } e^{-0.256} \\ &= 0.45 \text{ đến } 0.77\end{aligned}$$

Như vậy *OR* = 0.59 với khoảng tin cậy 95% dao động từ 0.45 đến 0.77. Ở đây, vì đơn vị là odds, nên chúng ta không thể sử dụng thuật ngữ “nguy cơ” (risk) được. Do đó,



cách diễn giải *OR* cũng khá khó khăn. Một cách nghiêm chỉnh, ý nghĩa thật sự của kết quả trên có thể diễn giải như sau: *odds* gãy xương trong nhóm bệnh nhân được điều trị với zoledronic acid giảm 41% (lấy 1 trừ cho 0.59 và nhân cho 100), với khoảng tin cậy 95% dao động từ 23% đến 55%.

Xin nói thêm rằng tiếng Việt chúng ta (cũng như nhiều tiếng gốc Latin khác như Pháp, Ý, Tây Ban Nha, v.v...) không có từ tương đương với *odds* của tiếng Anh, nên phần lớn sách giáo khoa trên thế giới đều sử dụng *odds* và *odds ratio* mà không dịch sang ngôn ngữ địa phương.

## So sánh

Bây giờ chúng ta có thể tính toán kiểm định nhị phân, *OR* và *RR* cho cả hai ví dụ mà tôi trình bày trong phần đầu. Kết quả có thể tóm lược trong bảng sau:

Chỉ số	Kí hiệu	Ví dụ 1 (Tai nạn tàu Titanic): nam và nữ	Ví dụ 2 (thuốc zoledronic acid): zoledronate vs placebo
Tỉ lệ biến cố nhóm 1	$p_1$	0.83	0.087
Tỉ lệ biến cố nhóm 2	$p_2$	0.33	0.139
Hiệu số và khoảng tin cậy 95%	$d = p_1 - p_2$	0.50 (0.45, 0.56)	-0.052 (-0.079, -0.026)
Odds ratio và khoảng tin cậy 95%	OR	0.10 (0.08, 0.13)	0.59 (0.45, 0.77)
Relative risk và khoảng tin cậy 95%	RR	0.40 (0.34, 0.47)	0.63 (0.61, 0.64)

Tất cả 3 phân tích đều nhất quán với kết luận. Trong ví dụ 1, tỉ lệ tử vong ở nam hành khách cao hơn nữ hành khách. Trong ví dụ 2, thuốc zoledronic acid có hiệu quả giảm nguy cơ gãy xương đốt sống.

Tuy nhiên, về chi tiết thì có sự khác biệt khá lớn giữa 3 phân tích. Trong ví dụ 1, nếu diễn tả bằng chỉ số tuyệt đối (absolute risk), chúng ta thấy tỉ lệ tử vong ở nữ thấp hơn nam đến 50% (83% trừ cho 33%). Nhưng nếu diễn tả bằng chỉ số tương đối, chúng ta phát biểu rằng tỉ lệ tử vong ở nữ thấp hơn nam 60% (lấy 1 trừ cho 0.4 và nhân cho 100). Nếu diễn tả bằng *OR*, thì *odds* tử vong ở nữ thấp hơn nam đến 90%! Trong ví dụ này, có sự khác biệt rất lớn giữa *OR* và *RR*. Tuy nhiên, trong ví dụ 2, mức độ khác biệt giữa *OR* và *RR* không đáng kể.

Vấn đề đặt ra là chỉ số nào (hiệu số, OR hay RR) thích hợp? Đây là vấn đề đã được trả lời trong một bài trước. Ở đây tôi chỉ tóm lược câu trả lời như sau: (a) hiệu số và OR có thể tính toán cho bất cứ mô hình nghiên cứu nào (như nghiên cứu bệnh chứng, nghiên cứu cắt ngang, nghiên cứu xuôi thời gian, v.v...); và (b) RR chỉ có thể tính toán cho nghiên cứu nghiên cứu cắt ngang và nghiên cứu xuôi thời gian. Tuy nhiên, OR chỉ nên sử dụng khi tần số bệnh (hay biến cố) xảy ra thấp, như dưới 10%, và trong những trường hợp này, OR tương đương với RR. Nhưng nếu tần số biến cố cao hơn 10% (như ví dụ 1) thì OR thường có xu hướng cao hơn RR. Trong cả hai trường hợp, cần phải rất cẩn thận khi diễn giải ý nghĩa của OR.