

## CHƯƠNG 8

# NGHIÊN CỨU TRONG DỊCH TỄ QUAN SÁT

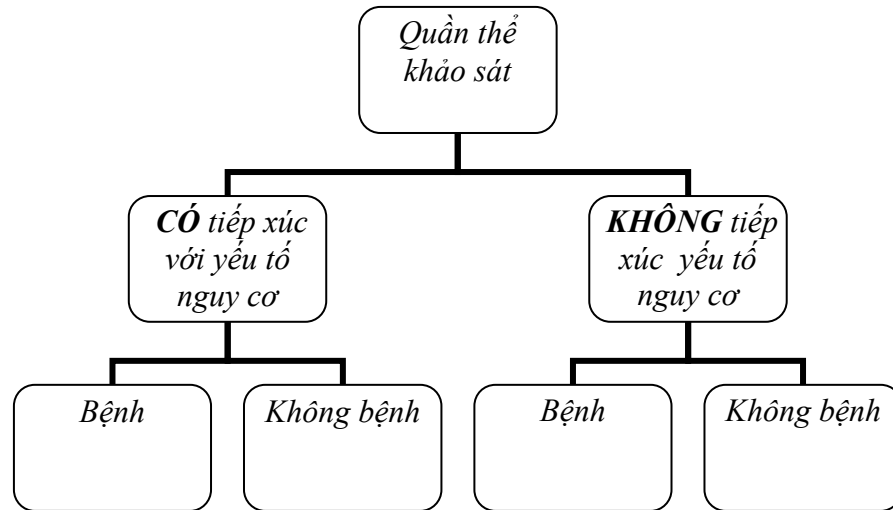
Bên cạnh các thử nghiệm lâm sàng trong việc phân tích yếu tố nguy cơ và nguyên nhân gây bệnh, các nghiên cứu dịch tễ quan sát là những nghiên cứu rất hữu dụng. Chúng khắc phục được những nhược điểm của thử nghiệm như đảm bảo được vấn đề về tính nhân đạo (ethic) và tận dụng được các số liệu thống kê, khảo sát. Chương này sẽ đề cập hai loại nghiên cứu quan sát, đó là nghiên cứu đoàn hệ và nghiên cứu bệnh-chứng.

### 1. Nghiên cứu đoàn hệ

Như đã đề cập, thuật ngữ đoàn hệ được dịch từ “cohort” trong tiếng Anh. Đây là thuật ngữ có nguồn gốc Latin, nghĩa là một nhóm chủ thể xác định có chung một đặc điểm. Ở đây người ta thường hình dung là các cá thể đưa vào nghiên cứu thuộc một quần thể trong đó chia ra thành hai nhóm, nhóm có cùng đặc tính là tiếp xúc với yếu tố nguy cơ và nhóm thứ hai là nhóm không tiếp xúc yếu tố nguy cơ.

Trong lĩnh vực thú y, người ta tiến hành nghiên cứu đoàn hệ bằng cách. Theo dõi một nhóm thú trong một quần thể. Khảo sát các cá thể xem có tiếp xúc với yếu tố nguy cơ hay không. Sau đó xác định được nhóm thú tiếp xúc và nhóm thú không tiếp xúc. Quan sát theo thời gian và ghi nhận lại sự xuất hiện bệnh ở hai nhóm thú trên. Tính toán giá trị RR cho phép người nghiên cứu kết luận được yếu tố nguy cơ quan sát có liên quan đến bệnh hay không.

Có hai loại nghiên cứu đoàn hệ: nghiên cứu đoàn hệ tiền cứu và hồi cứu. Trong nghiên cứu hồi cứu, sự phân nhóm thú tiếp xúc hay không tiếp xúc với yếu tố khảo sát dựa trên số liệu hoặc điều tra trong quá khứ. Sự xuất hiện bệnh xảy ra sau khi xác định sự tiếp xúc với yếu tố nguy cơ có thể thu thập từ quá khứ cho đến hiện tại, và có thể đến tương lai. Còn trong nghiên cứu tiền cứu, việc phân nhóm tiếp xúc với yếu tố nguy cơ được điều tra và xác định ngay trong hiện tại, việc quan sát xác định bệnh được thực hiện trong tương lai.



**Sơ đồ 12.1** Mô hình bố trí nghiên cứu đoàn hệ

Sau khi thu thập dữ liệu liên quan, các tham số thể hiện mối quan hệ giữa yếu tố nguy cơ và sự phát triển bệnh được tính toán. Với nghiên cứu đoàn hệ, chỉ số RR, OR và IR đều có thể chấp nhận được trong đó RR và IR được xem như mạnh hơn OR. Việc đánh giá và sử dụng các chỉ số này đã được thảo luận ở chương 9.

Để xác định dung lượng mẫu cho các nghiên cứu đoàn hệ, phải quan tâm đến một số tham số sau: ( $p_1$ ) ước lượng của tỷ lệ bệnh trong nhóm không tiếp xúc với yếu tố nguy cơ; (RR) giá trị RR mong muốn có ý nghĩa trong việc kết luận mối liên quan, thông thường người ta dùng 2 hoặc hơn 2; giá trị Z (phân phối chuẩn) của mức độ tin cậy và năng lực của trắc nghiệm thống kê (power of a test). Công thức tính như sau:

$$n = \frac{(p_1 \times q_1 + p_2 \times q_2)K}{(p_1 - p_2)^2} \quad \text{trong đó } p_2 = RR \times p_1$$

- n = Số thú cần cho nghiên cứu trong mỗi nhóm
- $K = [Z(a) + Z(b)]^2$
- Z(a) = Giá trị Student's t ( $df=\infty$ ) ở độ tin cậy mong muốn ( $1-\alpha$ )
- Z(b) = Giá trị Student's t ( $df=\infty$ ) ở năng lực của trắc nghiệm thống kê mong muốn ( $1-\beta$ )
- $p_1$  = Tỷ lệ bệnh trong nhóm không tiếp xúc yếu tố nguy cơ
- $q_1 = 1 - p_1$
- $p_2$  = Tỷ lệ bệnh của nhóm tiếp xúc yếu tố nguy cơ (được dự đoán theo chỉ số RR)
- $q_2 = 1 - p_2$
- RR = Giá trị RR mong muốn

## 2. Ví dụ về nghiên cứu đoàn hệ

Tại một trại bò sữa, các con bò ở giai đoạn khô sữa được đưa đánh giá thể trạng mập hay bình thường. Người ta cho là thể trạng có liên quan đến sốt sữa sau khi sinh trên bò. Theo những nghiên cứu trước đây, tỷ lệ bệnh trong nhóm bình thường khoảng 15%, ước tính số mẫu như sau (với chỉ số RR mong muốn là 2 và độ tin cậy 95%, năng lực trắc nghiệm là 90%):

$$p_2 = 0,15 \times 2 = 0,3$$

$$K = (Z(0,95) + Z(0,90))^2 = (1,96 + 1,282)^2 = 10,51$$

$$n = \frac{(p_1 \times q_1 + p_2 \times q_2)K}{(p_1 - p_2)^2} = (0,15 \times 0,85 + 0,3 \times 0,7)10,51 / 0,15^2 = 157,65 \approx 158$$

Như vậy cần ít nhất 158 con bò trong mỗi nhóm. Việc tính toán này có thể được thực hiện bằng phần mềm WinEpiscope, vào menu “Sample” chọn “Cohort” sau đó điền thông số thích hợp.

**Hình 12.1** Ước lượng mẫu bằng WinEpiscope trong nghiên cứu đoàn hệ

Tổng số 400 bò được đưa vào khảo sát trong đó 200 con được đánh giá là mập và 200 con được xem là bình thường. Như vậy tình trạng mập được xem như là yếu tố nguy cơ và yếu tố này được xác định từ đầu. Các con bò này được tiếp tục quan sát cho đến khi đẻ (trong tương lai) và xem sự biểu hiện bệnh sốt sữa ở từng nhóm. Đây thật sự là nghiên cứu đoàn hệ tiên cứu. Kết quả như sau:

Kết quả từ WinEpiscope (menu Analysis/cohort cum. incidence) (hình 12.2) cho thấy RR biến động trong khoảng 1,1 đến 2,5, điều này có nghĩa là tình trạng mập ở giai đoạn khô sữa có liên quan đến bệnh sốt sữa.

**Bảng 12.1** Kết quả nghiên liên quan giữa thể trạng cơ thể và sốt sưa

|         |            | Yếu tố khảo sát |             |      |
|---------|------------|-----------------|-------------|------|
|         |            | Mập             | Bình thường | Tổng |
| Kết quả | Sốt sưa    | 50              | 30          | 80   |
|         | Không bệnh | 150             | 170         | 320  |
|         | Tổng       | 200             | 200         |      |

The screenshot shows the 'Analysis of Cohort (Cumulative Incidence) #1' window. It contains two main sections: 'Input of DATA' and 'RESULTS'.

**Input of DATA:**

- Observed frequencies:** A table with columns 'Exposed' (Yes, No, Total) and rows 'Diseased', 'Healthy', 'Total'. Values: Diseased (50, 30, 80), Healthy (150, 170, 320), Total (200, 200, 400).
- Level of Confidence:** 95 %
- Buttons:** Calculate, Close.

**RESULTS:**

- Expected frequencies:** A table with columns 'Exposed' (Yes, No) and rows 'Diseased', 'Healthy'. Values: Diseased (40.00, 40.00), Healthy (160.00, 160.00).
- Attributable Risk:** 0.100
- Attributable propor. among exposed:** 0.400
- Attributable proportion:** 0.250
- Confidence Limits:** A table with columns 'Logarithmic approx.' (Lower Lim., Upper Lim.) and 'χ² Approximation' (Lower Lim., Upper Lim.). Values: Odds Ratio (OR): (1.142, 3.124, 1.147, 3.112); Relative Risk (RR): (1.108, 2.506, 1.116, 2.489); Bias: 0.222.
- Status:** Limits are valid.

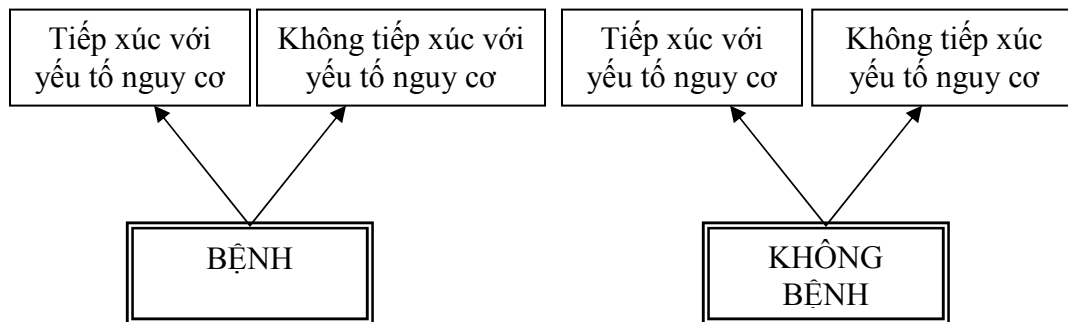
**Hình 12.2** Xử lý bằng WinEpiScope trong nghiên liên quan giữa thể trạng cơ thể và sốt sưa

### 3. Nghiên cứu bệnh-chứng

Đối với nghiên cứu bệnh-chứng, người thực hiện bắt đầu từ những ca bệnh ở các bệnh viện hay bệnh xá thú y, thu thập thông tin về các yếu tố nghi ngờ. Tìm thú đối chứng (không bệnh) thích hợp, thu thập dữ liệu của thú đối chứng. Kết quả tổng hợp được phân tích để xác định mối liên quan. Do bắt đầu từ những cá thể bệnh nên việc xác định tỷ lệ bệnh trong các nhóm có hay không có tiếp xúc yếu tố nguy cơ là không có ý

nghĩa, chính vì lý do đó mà giá trị RR không được sử dụng. Để đánh giá mức liên quan người ta dùng chỉ số OR.

Nghiên cứu bệnh-chứng thích hợp cho việc nghiên cứu những bệnh hiếm. Những ca bệnh có thể gặp ở các bệnh xá là trường hợp đặc biệt để đưa vào nghiên cứu. Bên cạnh đó, nghiên cứu bệnh-chứng còn rất thích hợp cho các bệnh thông thường ở giai đoạn đầu của nghiên cứu, khi mà những nghiên cứu cơ bản chưa được khảo sát, những yếu tố nguy cơ không được xác định. Dùng nghiên cứu này để giới hạn các yếu tố nguy cơ cần khảo sát trước khi thực hiện nghiên cứu đoàn hệ. Bố trí nghiên cứu bệnh-chứng như sau.



**Sơ đồ 12.2** Bố trí nghiên cứu bệnh chứng

Việc chọn lựa thú đưa vào nhóm bệnh và nhóm không bệnh là rất quan trọng và chúng quyết định tính chính xác của nghiên cứu. Các ca bệnh nên được chọn từ nhiều nơi, còn các nhóm đối chứng thì phải tương đồng về các yếu tố không phải là yếu tố nguy cơ với nhóm thú bệnh. Thông thường thì một thú bệnh sẽ có một hay nhiều thú làm đối chứng. Nếu các thú đối chứng chỉ là những thú không bệnh ngẫu nhiên, việc chọn lựa không theo nguyên tắc gắt gao nào thì được gọi là “không tương xứng” hay “không bắt cặp” (un-match). Ở đây, dùng từ bắt cặp cho dễ hiểu nhưng cần lưu ý là không phải bắt cặp theo kiểu 1 bệnh và 1 đối chứng mà có thể nhiều hơn. Trong khi đó, một số nghiên cứu khác người ta chọn thú đối chứng phải có một đặc điểm nào đó tương đồng với thú bệnh để loại trừ những sai lệch do yếu tố nhiễu. Ví dụ, thú đối chứng phải cùng mẹ với con bệnh, hay cùng giới ... trong trường hợp đó được gọi là nghiên cứu bệnh-chứng tương xứng (match). Kiểu tương xứng sẽ được thảo luận ở phần đề cập đến việc khắc phục yếu tố nhiễu.

Trước khi thực hiện nghiên cứu bệnh-chứng, người ta cũng dựa vào một số dữ kiện để ước tính dung lượng mẫu cần thiết cho nghiên cứu theo công thức sau:

$$n = \frac{\left[ Z_{\alpha} \sqrt{\left(1 + \frac{1}{c}\right) p q} + Z_{\beta} \sqrt{p_1 q_1 + \frac{p_o q_o}{c}} \right]^2}{(p_1 - p_o)^2}$$

Trong đó

n = số mẫu cần cho mỗi nhóm

$Z_{\alpha}, Z_{\beta}$  = Giá trị t cho độ tin cậy và năng lực của trắc nghiệm thống kê  
( $df=\infty$ )

$c$  = số lượng thú đối chứng cho một đơn vị thú bệnh

$p_o$  = tỷ lệ thú tiếp xúc với yếu tố nguy cơ trong nhóm thú đối chứng;

$q_o = 1 - p_o$

$p_1$  = tỷ lệ thú tiếp xúc với yếu tố nguy cơ trong nhóm thú bệnh (tính toán dựa trên OR mong muốn)

$q_1 = 1 - p_1$

$\bar{p}$  = tỷ lệ trung bình thú có tiếp xúc với yếu tố nguy cơ trong quần thể

$\bar{q} = 1 - \bar{p}$

$$\bar{p} = \frac{(p_1 + cp_o)}{1 + c} \quad p_1 = \frac{OR \times p_o}{1 + p_o \times (OR - 1)}$$

OR = chỉ số OR mong muốn

#### 4. Ví dụ về nghiên cứu bệnh-chứng

Giả thiết cho rằng cường độ chiếu sáng trong chăn nuôi gà thịt có liên quan đến sự xuất hiện bệnh tích trên ruột do cầu trùng (coccidiosis). Theo số liệu điều tra, tỷ lệ trại sử dụng chế độ chiếu sáng liên tục chiếm khoảng 30%.

- Ước lượng mẫu ở độ tin cậy 95% và năng lực trắc nghiệm 90% như sau:

$$Z_{\alpha} = 1,96 \quad Z_{\beta} = 1,282$$

$$c = 1$$

$$p_o = 0,3$$

$$q_o = 1 - 0,3 = 0,7$$

$$OR = 2$$

$$p_1 = \frac{OR \times p_o}{1 + p_o \times (OR - 1)} = 2 \times 0,3 / 1,3 = 0,46$$

$$\bar{p} = \frac{(p_1 + cp_o)}{1 + c} = (0,46 + 0,3) / 2 = 0,38$$

$$n = \frac{\left[ Z_{\alpha} \sqrt{\left(1 + \frac{1}{c}\right) \bar{p} \bar{q}} + Z_{\beta} \sqrt{p_1 q_1 + \frac{p_o q_o}{c}} \right]^2}{(p_1 - p_o)^2} =$$

$$= \frac{\left[ 1,96 \sqrt{\left(1 + \frac{1}{1}\right) 0,38 \times 0,62} + 1,282 \sqrt{0,46 \times 0,54 + \frac{0,3 \times 0,7}{1}} \right]^2}{(0,46 - 0,3)^2} \approx 191$$

Cách ước lượng mẫu trong WinEpiScope như sau: vào menu “samples” chọn “unmatched case-control” (Hình 12.3)

Nghiên cứu thực tế tiến hành như sau. Số liệu từ 208 trại gà được thu thập, trong đó 99 đàn có bệnh tích. Điều tra cho thấy gần 50% đàn có bệnh này sử dụng chế độ đèn chiếu sáng gián đoạn, trong khi đó chỉ có 28,4% số đàn gà âm tính dùng chế độ này.

Sử dụng WinEpiScope ,vào menu “Analysis” chọn “ Case-control”

**Bảng 12.1** Kết quả nghiên cứu liên quan giữa thời gian chiếu sáng và coccidiosis

|                   |                  | Coccidiosis |              |      |
|-------------------|------------------|-------------|--------------|------|
|                   |                  | <i>Có</i>   | <i>Không</i> | Tổng |
| Chế độ chiếu sáng | <i>Gián đoạn</i> | 49          | 31           | 80   |
|                   | <i>Liên tục</i>  | 50          | 78           | 128  |
|                   | Tổng             | 99          | 109          | 208  |

**Sample Size: Unmatched Case-Control #1**

**Sample Size** | Minimal Detectable OR

**Input of DATA:**

N° of controls per case:

Estimated exposure rate Controls (%):

Minimal OR to be detected signif.:

Level of Confidence (%):

Power (%):

**RESULTS:**

Estimated exposure rate Cases (%):

**Sample size per group:**

|             | Cases | Controls |
|-------------|-------|----------|
| One-Tailed: | 153   | 153      |
| Two-Tailed: | 188   | 188      |

**Hình 12.3** Ước lượng mẫu trong ví dụ của nghiên cứu bệnh chứng bằng WinEpiScope

**Analysis of Case-Control #2**

**Input of DATA:**

**Observed frequencies:**

|         | Exposed |     | Total |
|---------|---------|-----|-------|
|         | Yes     | No  |       |
| Case    | 49      | 50  | 99    |
| Control | 31      | 78  | 109   |
| Total   | 80      | 128 | 208   |

**Level of Confidence**

☐ 90 %  
☒ 95 %  
☐ 97.5 %  
☐ 99 %  
☐ 99.5 %

---

**RESULTS:**

**Expected frequencies:**

|         | Exposed |       |
|---------|---------|-------|
|         | Yes     | No    |
| Case    | 38.08   | 60.92 |
| Control | 41.92   | 67.08 |

**Confidence Limits of Odds Ratio**

|                        | Lower Lim. | Upper Lim. |
|------------------------|------------|------------|
| Logarithmic approx.    | 1.390      | 4.374      |
| $\chi^2$ Approximation | 1.396      | 4.355      |

Odds Ratio (OR): 2.466

Attributable propor. among exposed: 0.594

Attributable proportion: 0.294

**Limits are valid.**

**Hình 12.4** Xử lý số liệu ví dụ của nghiên cứu bệnh chứng bằng WinEpiScope

Kết quả  $OR = 2,466$  ( $1,39 - 4,37$ ) cho thấy bệnh coccidiosis có nguy cơ cao xảy ra ở những đàn sử dụng chế độ chiếu sáng gián đoạn, gấp 2,5 lần so với những đàn sử dụng chế độ liên tục

## 5. Sự sai lệch trong nghiên cứu quan sát

Trong các nghiên cứu dịch tễ học đã được trình bày ở chương trước, một vấn đề cần được quan tâm là độ chính xác (validity) của các nghiên cứu. Dĩ nhiên trong các nghiên cứu quan sát không thể nào tránh khỏi những sai lệch do các yếu tố tác động không kiểm soát được. Nội dung chương này sẽ đề cập đến các loại sai lệch đó (bias) và các cách để khắc phục. Các loại sai lệch bao gồm:

- + Sai lệch do chọn lựa (selection bias)
- + Sai lệch thông tin (information bias)
- + Sai lệch do yếu tố nhiễu (confounding bias)

### 5.1. Sai lệch do chọn lựa

Sai lệch loại này xảy ra khi thực hiện những nghiên cứu về mối liên quan giữa yếu tố nguy cơ và bệnh, các cá thể bệnh và cá thể đối chứng hoặc cá thể tiếp xúc với yếu tố nguy cơ và cá thể không tiếp xúc được chọn, kết quả cho thấy có mối liên quan ý nghĩa, tuy nhiên sự thật là chúng không liên quan nhau. Trong trường hợp đó, chúng ta đã mắc một sai lầm trong việc chọn các cá thể đưa vào khảo sát và sai lệch này được gọi là sai lệch do chọn lựa.

Giả sử trong một nghiên cứu, tổng số cá thể theo từng nhóm và số cá thể trong quần thể được đưa vào nghiên cứu được trình bày theo bảng 12.2:

**Bảng 12.2** Phân bố các cá thể đưa vào khảo sát so với quần thể

| Cả quần thể khảo sát |                |                |                | Các cá thể được đưa vào nghiên cứu |                |                |                |
|----------------------|----------------|----------------|----------------|------------------------------------|----------------|----------------|----------------|
|                      | E+             | E-             |                |                                    | E+             | E-             |                |
| Bệnh                 | A <sub>1</sub> | A <sub>o</sub> | M <sub>1</sub> | Bệnh                               | a <sub>1</sub> | a <sub>o</sub> | m <sub>1</sub> |
| Không bệnh           | B <sub>1</sub> | B <sub>o</sub> | M <sub>o</sub> | Không bệnh                         | b <sub>1</sub> | b <sub>o</sub> | m <sub>o</sub> |
|                      | N <sub>1</sub> | N <sub>o</sub> | N              |                                    | n <sub>1</sub> | n <sub>o</sub> | n              |

E+: Tiếp xúc yếu tố nguy cơ

E- : Không tiếp xúc yếu tố nguy cơ

Nhóm cá thể đưa vào khảo sát là một dạng mẫu lấy từ quần thể khảo sát. Giả sử việc chọn mẫu được thực hiện ngẫu nhiên thì tỷ lệ phân mẫu (sf) ở các nhóm thú như sau:

$$sf_{11} = a_1/A_1$$

$$sf_{12} = a_0/A_0$$

$$sf_{21} = b_1/B_1$$

$$sf_{22} = b_0/B_0$$

$$OR_{sf} = sf_{11} \times sf_{22} / sf_{12} \times sf_{21}$$

Nếu tất cả các tỷ lệ sf này đều bằng nhau, xem như sai lệch do chọn lựa không xảy ra. Trong thực tế, rất khó lấy mẫu theo cùng tỷ lệ như vậy. Mục đích của khảo sát là xác định mối liên quan thông qua chỉ số OR, do đó nếu tỷ số OR của tỷ lệ phân mẫu (OR<sub>sf</sub>) bằng 1 thì cũng có thể được cho rằng không có sai lệch chọn lựa.

Nếu OR của tỷ lệ phân mẫu (OR<sub>sf</sub>) lớn hơn 1 thì sai lệch có xảy ra và có xu hướng làm cho kết quả khảo sát có OR lớn hơn thực tế (lệch khỏi 1). Ngược lại nếu OR của tỷ lệ phân mẫu (OR<sub>sf</sub>) nhỏ hơn 1 thì kết quả OR khảo sát có xu hướng bị kéo về 1 (nhỏ hơn thực tế).

Ví dụ: nghiên cứu về mối quan hệ giữa yếu tố nguy cơ và bệnh trong quần thể thật và kết quả lấy mẫu như sau

**Bảng 12.3** Ví dụ về phân bố các cá thể đưa vào khảo sát so với quần thể

| Cả quần thể khảo sát |                         |      | Các cá thể được đưa vào nghiên cứu |                     |    |
|----------------------|-------------------------|------|------------------------------------|---------------------|----|
|                      | E+                      | E-   |                                    | E+                  | E- |
| Bệnh                 | 247                     | 853  | Bệnh                               | 13                  | 57 |
| Không bệnh           | 1089                    | 7911 | Không bệnh                         | 10                  | 80 |
| OR                   | 247×7911/853×1089=2,104 |      | OR                                 | 13×80/57×10 = 1,825 |    |

$$sf_{11} = a_1/A_1 = 13/247 = 0,053$$

$$sf_{12} = a_0/A_0 = 57/853 = 0,067$$

$$sf_{21} = b_1/B_1 = 10/1089 = 0,009$$

$$sf_{22} = b_o/B_o = 80/7911 = 0,01$$

$$OR_{sf} = sf_{11} \times sf_{22} / sf_{12} \times sf_{21} = (0,053 \times 0,01) / (0,067 \times 0,009) = 0,88$$

Trên đây chỉ trình bày về nguyên tắc tại sao sai lệch do chọn lựa lại ảnh hưởng đến kết quả khảo sát. Trên thực tế, không phải lúc nào chúng ta cũng xác định được các tỷ lệ phân mẫu. Do đó cách lấy mẫu ngẫu nhiên được xem như là thích hợp nhất để không chế sai lệch này. Việc chọn quần thể lấy mẫu cũng khá quan trọng sao cho đại diện được cho một quần thể lớn.

## 5.2. Sai lệch thông tin

Trong các nghiên cứu quan sát, thông tin thu thập được đôi khi không phải là thông tin trực tiếp đo lường mà chỉ liên quan đến sự đánh giá hoặc chỉ là thông tin hồi cứu. Trong trường hợp này, thông tin cung cấp cho khảo sát có thể bị sai lệch. Trong sai lệch thông tin, người ta có thể kể đến sai lệch thông tin tóm tắt, sai lệch thông tin do phỏng vấn, sai lệch thông tin do hồi cứu, và sai lệch thông tin do báo cáo.

Bên cạnh đó, người ta cũng đề cập đến một dạng sai lệch thông tin ở các chỉ tiêu được trực tiếp xác định bởi nhà nghiên cứu, đó là “sai lệch do phân loại sai” (misclassification bias). Sai lệch này xảy ra khi ca bệnh không được nhận xét đánh giá, chẩn đoán đúng, và ngược lại. Sai lệch này đặc biệt xảy ra khi các chẩn đoán xét nghiệm có độ nhạy và độ chuyên biệt thấp hơn 100%. Ngoài sự phân loại sai ca bệnh, thì việc tiếp xúc với yếu tố nguy cơ cũng có thể bị phân loại sai. Cả hai loại phân loại sai này đều ảnh hưởng trực tiếp đến kết quả về mối quan hệ cần xác định.

## 5.3. Yếu tố nhiễu

Một vấn đề thường gặp trong nghiên cứu dịch tễ là sự liên quan giữa yếu tố khảo sát và bệnh đôi khi bị một yếu tố khác ảnh hưởng làm cho kết quả bị sai lệch. Chính vì vậy mà thuật ngữ “yếu tố nhiễu” được đề cập để đưa ra khái niệm về các yếu tố này. Đây cũng là một vấn đề chính trong nghiên cứu phân tích dịch tễ.

Về mặt định nghĩa, một yếu tố X được gọi là yếu tố nhiễu trong mối quan hệ của yếu tố nguy cơ A và bệnh B khi thỏa mãn 2 điều kiện sau: (1) yếu tố X có liên quan đến nguy cơ bệnh, có nghĩa là X được biết làm tăng nguy cơ bệnh nhưng không phải là nguyên nhân, cách phát triển, và diễn tiến của bệnh B; (2) yếu tố X liên quan đến yếu tố A nhưng không là kết quả của yếu tố A.

Một ví dụ điển hình khi nghiên cứu trên người về tác dụng của cà phê và bệnh ung thư tuyến tụy trên người, yếu tố gây nhiễu chính là hút. Người ta quan tâm đến mối liên hệ giữa cà phê (yếu tố A) và bệnh ung thư tuyến tụy (bệnh B) và nhận thấy là (1) yếu tố hút thuốc (X) được biết là một nguy cơ dẫn đến ung thư tuyến tụy và (2) những người sử dụng nhiều cà phê thì thường có kèm theo hút thuốc. Như vậy, yếu tố X đã có 2 đặc điểm để được xem là một yếu tố gây nhiễu.

Trong thí nghiệm, khi khảo sát về thể trạng của bò sữa ở giai đoạn khô sữa (A) với bệnh sốt sữa (X), chúng ta nhận thấy yếu tố tuổi hay chu kỳ cho sữa ảnh hưởng rất lớn đến bệnh vì những con có chu kỳ mới thường xảy ra tình trạng này. Bên cạnh đó số chu kỳ cho sữa cũng ảnh hưởng đến tình trạng cơ thể (điểm thể trạng) do đó tuổi của bò sữa có thể là một yếu tố nhiễu khi xét mối quan hệ này.

Giả sử kết quả thu thập từ khảo sát trên như sau

**Bảng 12.4** Bảng 2x2 tính chung cho kết quả khảo sát liên quan giữa thể trạng và sốt sữa

|           |             | Sốt sữa | Không sốt sữa |
|-----------|-------------|---------|---------------|
| Thể trạng | Mập         | 50      | 150           |
|           | Bình thường | 30      | 170           |

$OR_o = 1,89$ . Điều này nghĩa là có sự liên quan giữa chỉ số tình trạng cơ thể và bệnh sốt sữa. Tuy nhiên khi xét về chu kỳ cho sữa thì kết quả như sau:

**Bảng 12.5** Kết quả khảo sát liên quan giữa thể trạng và sốt sữa theo các chu kỳ cho sữa

| Chu kỳ 1-2  |                                   |           | Chu kỳ 3 trở lên |                                    |           |
|-------------|-----------------------------------|-----------|------------------|------------------------------------|-----------|
|             | Sốt sữa                           | Không sốt |                  | Sốt sữa                            | Không sốt |
| Mập         | 5                                 | 45        | Mập              | 45                                 | 105       |
| Bình thường | 15                                | 135       | Bình thường      | 15                                 | 35        |
| OR =        | $5 \times 135 / 15 \times 45 = 1$ |           | OR =             | $45 \times 35 / 15 \times 105 = 1$ |           |

Như vậy, OR theo từng nhóm lại cho thấy kết quả bằng 1, nghĩa là không có sự liên quan giữa thể trạng với bệnh sốt sữa. Điều này một lần nữa cho thấy tầm quan trọng của việc xác định yếu tố nhiễu. Trong thực tế, các yếu tố ảnh hưởng có thể dự đoán được. Nhà nghiên cứu sẽ tìm cách kiểm soát yếu tố nhiễu bằng những cách sau: giới hạn khảo sát - có nghĩa là khi biết được yếu tố nhiễu thì người ta sẽ lấy mẫu và chọn thú khảo sát theo một tiêu chuẩn xác định. Ví dụ trong trường hợp trên, người ta chỉ chọn thú có 1-2 chu kỳ, lúc đó sẽ không chế được yếu tố nhiễu. Đây là phương pháp hợp lý làm đơn giản hoá phân tích dịch tễ. Trong các nghiên cứu bệnh-chứng, việc lấy mẫu, chọn thú bệnh và đối chứng theo kiểu bắt cặp cũng là phương pháp kiểm soát yếu tố nhiễu.

Tuy nhiên trong nghiên cứu người ta cũng muốn đưa yếu tố nhiễu vào phân tích thêm các mối quan hệ. Trong trường hợp đó, người ta có thể dùng phương pháp phân tầng để khảo sát yếu tố nhiễu. Ngoài ra còn có phương pháp phân tích đa biến ví dụ như logistic cũng là một cách để hiệu chỉnh tác động của yếu tố gây nhiễu.

#### 5.4 Phương pháp hiệu chỉnh yếu tố nhiễu

##### (1) Phương pháp bắt cặp (matching)

Phương pháp này thường được sử dụng trong các nghiên cứu bệnh-chứng. Để không chế yếu tố nhiễu, người ta thực hiện bắt cặp như sau: cứ 1 ca bệnh được xác định mang một giá trị nào đó của yếu tố nhiễu thì người ta chọn R thú đối chứng (không bệnh)

cùng mang giá trị tương ứng của yếu tố nhiều. Ví dụ, yếu tố nhiều là giới tính và người nghiên cứu muốn bắt cặp 1:2 thì cứ 1 ca bệnh là thú đực thì phải có 2 con đực làm đối chứng. Trong trường hợp đó, cách tính OR như sau

**Bảng 12.6** Phân bố trong nghiên cứu bệnh-chứng có bắt cặp

|         |                               | Số lượng đối chứng bắt cặp |                 |                 |                 |                 |
|---------|-------------------------------|----------------------------|-----------------|-----------------|-----------------|-----------------|
|         |                               | 0                          | 1               | 2               | 3               | R               |
| Ca bệnh | Tiếp xúc yếu tố nguy cơ       | F <sub>10</sub>            | F <sub>11</sub> | F <sub>12</sub> | F <sub>13</sub> | F <sub>1R</sub> |
|         | KHÔNG tiếp xúc yếu tố nguy cơ | F <sub>00</sub>            | F <sub>01</sub> | F <sub>02</sub> | F <sub>03</sub> | F <sub>0R</sub> |

F<sub>1R</sub> : số lượng của bộ số liệu (record) có ca bệnh tiếp xúc yếu tố nguy cơ và đối chứng nhóm R tiếp xúc yếu tố nguy cơ

F<sub>0R</sub> : số lượng của bộ số liệu (record) có ca bệnh không tiếp xúc yếu tố nguy cơ và đối chứng nhóm R không tiếp xúc yếu tố nguy cơ

$$OR = \frac{\sum_{i=1}^R (R+1-i) \times F_{1,i-1}}{\sum_{i=1}^R i \times F_{0,i}}$$

Ví dụ như trong trường hợp khảo sát mối quan hệ giữa yếu tố A và bệnh B trong đó yếu tố giới tính là yếu tố nhiễu. Thực hiện một nghiên cứu bệnh-chứng trong đó có tiến hành bắt cặp theo giới tính với một ca bệnh và 2 đối chứng. Kết quả được trình bày trong bảng 12.7.

**Bảng 12.7** Kết quả số liệu thô của ví dụ về nghiên cứu bệnh-chứng có bắt cặp

| Male | Case | Control 1 | Control 2 |
|------|------|-----------|-----------|
| 1    | 1    | 0         | 1         |
| 1    | 1    | 0         | 0         |
| 1    | 1    | 1         | 0         |
| 1    | 0    | 0         | 0         |
| 1    | 0    | 0         | 1         |
| 1    | 1    | 0         | 1         |
| 1    | 0    | 1         | 0         |
| 0    | 0    | 1         | 0         |
| 0    | 0    | 0         | 0         |
| 0    | 0    | 0         | 0         |
| 0    | 1    | 1         | 0         |
| 0    | 1    | 0         | 0         |

Male: với (1) là con đực và (0) là con cái

Case: là những con bệnh trong đó (1) là tiếp xúc yếu tố nguy cơ A và (0) là không tiếp xúc yếu tố A

Control: đối chứng tiếp xúc hay không tiếp xúc yếu tố nguy cơ.

Nếu không phân tích dạng bắt cặp mà phân tích theo nghiên cứu bệnh-chứng bình thường (không quan tâm yếu tố giới tính và bắt cặp) thì kết quả như sau :

**Bảng 12.8** Kết quả của ví dụ về nghiên cứu bệnh-chứng có bắt cặp nhưng không quan tâm yếu tố giới tính và bắt cặp

|            | E (+) | E (-) |    |
|------------|-------|-------|----|
| Bệnh       | 6     | 6     | 12 |
| Không bệnh | 7     | 17    | 24 |

$$OR = 2,49 \quad (0,58 - 10,19)$$

Nếu phân tích theo kiểu bắt cặp kết quả như sau:

**Bảng 12.9** Kết quả của ví dụ về nghiên cứu bệnh-chứng có bắt cặp

|                                      | Các đối chứng bắt cặp |   |   |
|--------------------------------------|-----------------------|---|---|
|                                      | 0                     | 1 | 2 |
| Ca bệnh có mang yếu tố nguy cơ       | 6                     | 2 | 2 |
| Ca bệnh KHÔNG có mang yếu tố nguy cơ | 6                     | 4 | 5 |

$$OR = \frac{\sum_{i=1}^R (R+1-i) \times F_{1,i-1}}{\sum_{i=1}^R i \times F_{o,i}} = \frac{(2+1-1) \times 6 + (2+1-2) \times 2}{1 \times 4 + 2 \times 5} = 1$$

Có thể dùng WinEpiscope bằng cách chọn menu “Analysis”, “Analysis matched case-control”

**Analysis of Matched Case-Control #1**

**Input of DATA:**

|     |          | Controls |   |   |   |   |   |
|-----|----------|----------|---|---|---|---|---|
|     |          | 0        | 1 | 2 | 3 | 4 | 5 |
| R=1 | Exp.     | 0        | 0 |   |   |   |   |
|     | Non Exp. | 0        | 0 |   |   |   |   |
| R=2 | Exp.     | 6        | 2 | 2 |   |   |   |
|     | Non Exp. | 6        | 4 | 5 |   |   |   |
| R=3 | Exp.     | 0        | 0 | 0 | 0 |   |   |
|     | Non Exp. | 0        | 0 | 0 | 0 |   |   |
| R=4 | Exp.     | 0        | 0 | 0 | 0 | 0 |   |
|     | Non Exp. | 0        | 0 | 0 | 0 | 0 |   |
| R=5 | Exp.     | 0        | 0 | 0 | 0 | 0 | 0 |
|     | Non Exp. | 0        | 0 | 0 | 0 | 0 | 0 |

Level of Confidence:

- ☐ 90 %
- ☒ 95 %
- ☐ 97.5 %
- ☐ 99 %
- ☐ 99.5 %

**RESULTS:**

Odds Ratio (OR):

Attributable propor. among exposed:

Attributable proportion:

**Confidence Limits of Odds Ratio**

|                        | Lower Lim. | Upper Lim. |
|------------------------|------------|------------|
| Logarithmic approx.    | 0.404      | 2.478      |
| $\chi^2$ Approximation | 1.000      | 1.000      |

**Hình 12.5** Kết quả xử lý nghiên cứu bệnh-chứng có bắt cặp bằng WinEpiScope

Đặc biệt khi bắt cặp đôi (pair matching), nghĩa là cứ một ca bệnh sẽ được lấy một đối chứng có cùng yếu tố nhiễu, công thức tính OR sẽ như sau:

**Bảng 12.10** Bảng tổng hợp cho nghiên cứu bệnh-chứng bắt cặp đôi

|         |       | Đối chứng |       |
|---------|-------|-----------|-------|
|         |       | E (+)     | E (-) |
| Ca bệnh | E (+) | a         | b     |
|         | E (-) | c         | d     |

$$OR = b/c \text{ và } 95\%CI(OR) = e^{[\log OR \pm (1,96 \times \sqrt{1/b + 1/c})]}$$

## (2) Phương pháp phân tích phân tầng (stratification)

Được thực hiện theo công thức Mantel-Haenszel. Giả sử khảo sát chia số liệu theo từng nhóm thú theo yếu tố nhiễu (phân tầng) thành nhiều tầng (ví dụ theo lứa tuổi). Ở mỗi tầng, chúng ta đều có thể tổng hợp lại thành dạng bảng 2x2.

**Bảng 12.9** Bảng tổng hợp phân bố phân tầng

|      | Tiếp xúc yếu tố nguy cơ | Không tiếp xúc | Tổng cộng |
|------|-------------------------|----------------|-----------|
| Bệnh | $A_i$                   | $B_i$          | $M_{1i}$  |

|            |          |          |          |
|------------|----------|----------|----------|
| Không bệnh | $C_i$    | $D_i$    | $M_{0i}$ |
|            | $N_{1i}$ | $N_{0i}$ | $T_i$    |

$$OR_{MH} = [\Sigma(A_i D_i / T_i)] / [\Sigma(B_i C_i / T_i)]$$

$$\text{và } RR_{MH} = (\Sigma A_i N_{0i} / T_i) / (\Sigma B_i N_{1i} / T_i)$$

Việc tính khoảng biến động của các chỉ số này khá phức tạp, có thể dùng WinEpiscope để tính một cách dễ dàng. Dựa theo kết quả của ví dụ phần 5.3 (Bảng 12.5), OR hiệu chỉnh có thể được tính như sau:

$$OR_{MH} = [\Sigma(A_i D_i / T_i)] / [\Sigma(B_i C_i / T_i)] = [(5 \times 135 / 200) + (45 \times 35 / 200)] / [(15 \times 45 / 200) + (15 \times 105 / 200)] = 1$$

Trong WinEpiscope, vào menu “Analysis”, chọn “Analysis of stratified case-control study”.

**Analysis of Stratified Case-Control #6**

**Input of DATA:**

| Stratum | Cases |          | Controls |          | Total |
|---------|-------|----------|----------|----------|-------|
|         | Exp.  | Non Exp. | Exp.     | Non Exp. |       |
| 1       | 5     | 15       | 45       | 135      | 200   |
| 2       | 45    | 15       | 105      | 35       | 200   |
| 3       | 0     | 0        | 0        | 0        | 0     |
| 4       | 0     | 0        | 0        | 0        | 0     |
| 5       | 0     | 0        | 0        | 0        | 0     |
| 6       | 0     | 0        | 0        | 0        | 0     |
| 7       | 0     | 0        | 0        | 0        | 0     |
| 8       | 0     | 0        | 0        | 0        | 0     |
| 9       | 0     | 0        | 0        | 0        | 0     |
| 10      | 0     | 0        | 0        | 0        | 0     |
| Total   | 50    | 30       | 150      | 170      | 400   |

Level of Confidence: 90 %, 95 %, 97.5 %, 99 %, 99.5 %

**RESULTS: Odds Ratio and Confidence Limits**

|                   | OR    | Logarithmic approx. |            | $\chi^2$ Approximation |            | $\chi$ |
|-------------------|-------|---------------------|------------|------------------------|------------|--------|
|                   |       | Lower Lim.          | Upper Lim. | Lower Lim.             | Upper Lim. |        |
| Crude             | 1.889 | 1.142               | 3.124      | 1.709                  | 2.088      | 12.459 |
| Pooled (Direct)   | 1.000 | 0.557               | 1.794      |                        |            |        |
| Pooled (Mantel-H) | 1.000 | 0.557               | 1.794      | Error                  | Error      | 0.000  |
| Stratum 1         | 1.000 | 0.344               | 2.906      | 1.000                  | 1.000      | 2.701  |

**Breslow-Day test for homogeneity of OR's across strata using MH-OR:**  
 Q(BD) = 20345.00 with 1 degree of freedom.  
 p-value: <0.0001 Level of Confidence (%): > 99.9

**Hình 12.6** Kết quả xử lý nghiên cứu bệnh-chứng có phân tầng bằng WinEpiscope

**(3) Phương pháp phân tích đa biến (multivariable analysis):** sẽ được đề cập ở chương sau.

