**21**

**Ước tính cỡ mẫu**

Mỗi nghiên cứu khoa học có liên quan đến quần thể và cỡ mẫu đều cần phải có ước tính số đối tượng cần thiết. Hai chữ “đối tượng” ở đây cần phải hiểu theo nghĩa rộng: với nghiên cứu lâm sàng, đối tượng có thể là bệnh nhân; với nghiên cứu mang tính điều tra xã hội, đối tượng có thể là cá nhân trong cộng đồng; với nghiên cứu trên động vật, đối tượng có thể là chuột. Mục tiêu của việc ước tính cỡ mẫu là ước tính số lượng đối tượng cần thiết (không quá thấp và cũng không quá cao) để kiểm định một hay nhiều giả thuyết khoa học. Số lượng đối tượng có liên quan đến đạo đức khoa học, qui mô nghiên cứu và chi phí, nên việc ước tính cỡ mẫu cần phải được xem xét và cân nhắc rất cẩn thận.

**21.1 Thông tin cần thiết cho ước tính cỡ mẫu**

Những yếu tố cần phải xem xét và cân nhắc trong ước tính cỡ mẫu là mô hình nghiên cứu, biến số phân tích, mức độ ảnh hưởng, và mức độ sai sót trong kiểm định giả thuyết. Nhà nghiên cứu cần phải có những thông tin này trước khi ước tính cỡ mẫu. Không có những thông tin trên thì không thể ước tính cỡ mẫu. Dưới đây là vài giải thích ngắn về 4 thông tin trên.

**21.1.1. Mô hình nghiên cứu**

Trong nghiên cứu y khoa, người ta phân biệt các mô hình nghiên cứu theo thời gian như sau:

*Nghiên cứu cắt ngang* (cross-sectional study) là mô hình nghiên cứu thường có mục tiêu ước tính một tỉ lệ hiện hành (prevalence) của bệnh ngay tại thời điểm thực hiện nghiên cứu. Ví dụ như nhà nghiên cứu muốn ước tính số người mắc bệnh xương khớp trong quần thể, nhà nghiên cứu có thể lấy mẫu gồm *n* cá nhân, và làm xét nghiệm để xác định có mắc bệnh hay không. Tỉ lệ mắc bệnh có thể ước tính từ xét nghiệm ngay tại thời điểm đó.

*Nghiên cứu bệnh chứng* (case-control study) là nghiên cứu ngược thời gian để đánh giá mối liên quan giữa một yếu tố nguy cơ và bệnh. Thông thường, nhà nghiên cứu chọn một nhóm bệnh nhân và một nhóm chứng, và thu thập thông tin về yếu tố nguy cơ *trong quá khứ* của mỗi cá nhân.

*Nghiên cứu đoàn hệ* (prospective study, có khi còn gọi là longitudinal study) là nghiên cứu thường có mục tiêu tìm hiểu yếu tố nguy cơ có ảnh hưởng đến nguy cơ mắc bệnh. Chẳng hạn như để đánh giá mối liên quan giữa tỉ trọng cơ thể (body mass index – BMI) và tử vong, nhà nghiên cứu có thể lấy mẫu *n* cá nhân, đo BMI lúc ban đầu, và sau đó theo dõi trong vòng [ví dụ như] 5 năm để biết bao nhiêu người sống và tử vong. Mối liên quan giữa BMI lúc ban đầu và nguy cơ tử vong có thể phân tích từ dữ liệu theo thời gian như thế. Nghiên cứu lâm sàng RCT (randomized controlled trials) có thể xem là một mô hình nghiên cứu đoàn hệ.

**21.1.2. Biến số kết quả**

Một nghiên cứu định lượng thường có một biến số chính, hay gọi là *outcome* hay *biến kết quả*. Biến kết quả có thể là biến liên tục (như huyết áp, nồng độ glucose, mật độ xương, v.v.) hoặc biến nhị phân (như sống / chết, bệnh / không bệnh). Phân biệt biến kết quả rất quan trọng vì sự phân biệt này xác định phương pháp ước tính cỡ mẫu.

**21.1.3. Mức độ ảnh hưởng**

Đây là một thông tin rất quan trọng cho ước tính cỡ mẫu. Không có thông tin này thì không thể nào ước tính cỡ mẫu. Trong thực tế, mức độ ảnh hưởng (còn gọi là *effect size* - SE) tuỳ thuộc vào mô hình nghiên cứu. Dưới đây là vài mô tả về cách xác định thông tin về mức độ ảnh hưởng đơn giản:

* Nếu là nghiên cứu cắt ngang, và mục tiêu là ước tính một tỉ lệ hiện hành (prevalence rate). Nhưng vì tỉ lệ có thể dao động giữa các mẫu nghiên cứu, nên một tham số khác cần phải xem xét đến: đó là sai số (gọi là *margin of error*). Gia dụ rằng chúng ta biết rằng tỉ lệ là *p* và sai số là *e,* thì ES = p / e.
* Nếu là nghiên cứu cắt ngang, và mục tiêu là ước tính một tham số liên tục. Chẳng hạn như để ước tính mật độ xương trong quần thể, nhà nghiên cứu có thể lấy mẫu *n* cá nhân, và đo mật độ xương. Trị số trung bình của mẫu là một ước số cho giá trị trung bình của quần thể. Do đó, trị số trung bình mẫu *m* vẫn chịu sự dao động ngẫu nhiên giữa các mẫu, và sai số để mô tả là *sai số e.* Trong trường hợp này, ES = *m / e*.
* Nếu là nghiên cứu có mục tiêu so sánh giữa hai nhóm, và trong trường hợp này nghiên cứu phải có hai tham số về trung bình (*T*1 và *T*2) và tham số về độ lệch chuẩn (*S*1 *và S*2), ES được định nghĩa là hiệu số giữa *T*1 và *T*2 chia cho *S*:



Nếu biến kết quả là biến liên tục, nghiên cứu sẽ có hai số trung bình *m*1 và *m*2, và trong trường hợp này,



Nếu biến kết quả là biến nhị phân, nghiên cứu sẽ có hai tỉ lệ *p*1 và *p*2, và trong trường hợp này, ES có thể ước tính bằng hiệu số ES = arcsin(*p*1) – arcsin(*p*2).

**21.1.4. Mức độ sai sót trong kiểm định giả thuyết**

Đứng trên quan điểm phương pháp luận, mỗi nghiên cứu thường có mục đích kiểm định một giả thuyết khoa học chính. Trong thống kê, người ta phân biệt 2 loại giả thuyết: giả thuyết vô hiệu (*null hypothesis,* thường kí hiệu là H0) và giả thuyết đảo (*alternative hypothesis*, kí hiệu HA). Giả thuyết vô hiệu, như tên gọi, phát biểu ngược lại với những gì nhà nghiên cứu muốn biết. Nếu giả thuyết đảo (hay giả thuyết chính) là bệnh nhân nhóm A có nguy cơ biến chứng cao hơn bệnh nhân nhóm B, thì giả thuyết vô hiệu sẽ phát biểu là “bệnh nhân nhóm A có nguy cơ biến chứng cao hơn bệnh nhân nhóm B”. Tương tự, nếu giả thuyết chính phát biểu rằng có mối tương quan giữa X và Y, thì giả thuyết vô hiệu phát biểu rằng không có mối tương quan giữa X và Y.

Ví dụ như nếu nhà nghiên cứu muốn xác định hiệu quả của thuốc chống loãng xương, và quyết định tiến hành một nghiên cứu theo mô hình RCT, với hai nhóm bệnh nhân: Một nhóm được điều trị bằng thuốc và một nhóm chứng, cả hai nhóm được theo dõi 3 năm. Biến kết quả là tỉ lệ gãy xương trong thời gian theo dõi. Gọi tỉ lệ gãy xương của hai nhóm là *p*1 và *p*2, giả thuyết vô hiệu (kí hiệu H0) là:

H0: *p*1 = *p*2

Và giả thuyết chính có thể là:

HA: *p*1 ≠ *p*2

Nên nhớ rằng nhà nghiên cứu không thể chứng minh H0; nhà nghiên cứu chỉ có thể bác bỏ H0 và do đó gián tiếp chấp nhận HA. Cơ sở để bác bỏ hay chấp nhận một giả thuyết là chỉ số thống kê (như t test, Ki bình phương, hồi qui tuyến tính, v.v.). Kết quả kiểm định thống kê có thể “*có ý nghĩa thống kê*” (*statistical significance*) hay không có ý nghĩa thống kê (statistical insignificance). Do đó, mỗi kiểm định giả thuyết, nhà nghiên cứu phải đối phó với 4 tình huống (xem bảng):

**Bảng 1: Kết quả kiểm định thống kê và giả thuyết vô hiệu**

|  |  |  |
| --- | --- | --- |
| **Kết quả kiểm định thống kê** | **Giả thuyết H0 đúng** | **Giả thuyết H0 sai** |
| Có ý nghĩa (tức bác bỏ H0) | Sai sót loại I (type I error, α) | Đúng |
| Không có ý nghĩa (chấp nhận H0) | Đúng | Sai sót loại II (type II error, β) |

* Kết quả kiểm định “có ý nghĩa thống kê”, và giả thuyết H0 là đúng, nhà nghiên cứu kết luận sai, và sai sót này tạm gọi là sai sót loại I (*type I error,* hay *alpha*);
* Kết quả kiểm định “có ý nghĩa thống kê”, và giả thuyết H0 là sai, nhà nghiên cứu kết luận đúng;
* Kết quả kiểm định “không có ý nghĩa thống kê”, và giả thuyết H0 là đúng, nhà nghiên cứu kết luận đúng;
* Kết quả kiểm định “không có ý nghĩa thống kê”, và giả thuyết H0 là sai, nhà nghiên cứu kết luận sai, và sai sót này tạm gọi là sai sót loại II (*type II error,* hay *beta*).

Nhà nghiên cứu phải xác định mức độ sai sót loại I và loại II trước khi ước tính cỡ mẫu. Thông thường, alpha được xác định ở mức 0.05 hay 0.01, còn đối với beta, các nhà nghiên cứu thường chấp nhận sai sót 0.20 hay 0.10. Cần nói thêm rằng, 1 – beta còn được đề cập đến là *power.* Power của một nghiên cứu là *độ nhạy* của nghiên cứu đó (cũng tương tự như độ nhạy hay *sensitivity* của một xét nghiệm chẩn đoán). Cần nhắc lại rằng, độ nhạy của một xét nghiệm có nghĩa là xác suất mà một cá nhân mắc bệnh có kết quả dương tính. Do đó, *power* của nghiên cứu là xác suất mà nghiên cứu cho ra kết quả có ý nghĩa thống kê nếu giả thuyết HA đúng (H0 sai).

Mỗi sai sót α và β đi kèm theo một hằng số của phân bố chuẩn (z), và hằng số này được sử dụng trong tính toán cỡ mẫu. Chẳng hạn như (trong trường hợp kiểm định 2 chiều) nếu α = 0.05 thì z = 1.96, nếu α = 0.01 thì z = 2.56. Sau đây là vài hằng số phổ biến:

**Bảng 2: Hằng số z liên quan đến sai sót loại I và loại II**

|  |  |
| --- | --- |
| **Hằng số z liên quan đến α (kiểm định 2 chiều)** | **Hằng số z liên quan đến β** |
| α = 0.01, z = 2.576 | β = 0.01, z = 2.326 |
| α = 0.05, z = 1.960 | β = 0.05, z = 1.645 |
| α = 0.10, z = 1.645 | β = 0.10, z = 1.282 |
| α = 0.20, z = 1.282 | β = 0.20, z = 0.842 |

**Chú thích**: có thể tính bằng hàm qnorm(x/2), trong đó x = 0.01, 0.05, 0.10, v.v.

**21.2 Ước tính cỡ mẫu bằng R**

Phần lớn các phương pháp ước tính cỡ mẫu có thể tóm gọn trong công thức sau:



Trong đó, *k* là hằng số cố định, *C*(α,β) là hằng số z liên quan đến sai sót loại I và sai sót loại II như sau:

|  |  |  |
| --- | --- | --- |
| **Sai số α** | **Sai số β = 0.10** | **Sai số β = 0.20** |
| 0.05 | C = 13.0 | C = 7.85 |
| 0.01 | C = 17.8 | C = 11.7 |

Trong công thức trên, vì ES là mẫu số, nên nếu *ES* thấp thì *n* sẽ tăng, và nếu ES cao thì *n* sẽ giảm. Tất cả các phương pháp ước tính cỡ mẫu qua công thức trên có thể triển khai trong R một cách dễ dàng. Tuy nhiên, trong R có một số package (như **samplesize, pwr, epiDisplay** có thể sử dụng cho ước tính cỡ mẫu, và người sử dụng chỉ cần gõ lệnh và cung cấp những thông tin vừa mô tả. Phần dưới đây, tôi sẽ minh hoạ một số trường hợp ước tính cỡ mẫu cho từng mô hình nghiên cứu và biến kết quả. Mỗi trường hợp sẽ được bắt đầu bằng vài điểm chính về lí thuyết và minh hoạ bằng một ví dụ cụ thể.

**21.2.1. Cỡ mẫu để ước tính một tỉ lệ**

Lí thuyết: Nhiều nghiên cứu có mục đích đơn giản là ước tính tỉ lệ. Chẳng hạn như các cuộc điều tra xã hội muốn biết bao nhiêu người thích hay không thích một đảng phái chính trị hay một món hàng, hoặc nghiên cứu y tế công cộng có mục đích ước tính tỉ lệ hiện hành (prevalence) của một bệnh.

Gọi π là tỉ lệ trong quần thể (population) mà chúng ta không biết, nhưng muốn ước tính dựa trên một mẫu nghiên cứu (sample). Gọi *p* là tỉ lệ có thể tính từ mẫu nghiên cứu gồm *n* đối tượng. Với giá trị *p* và *n* chúng ta có thể ước tính khoảng tin cậy (1-α)% của π dao động trong khoảng , trong đó là sai số mẫu – margin of error. Trong thực tế, *e* chính là sai số chuẩn [standard error] của *p*. Do đó, phát biểu ngược lại, số cỡ mẫu *n* cần thiết cho nghiên cứu phụ thuộc vào α, *e* và giá trị *p* như sau:



[1]



Ví dụ 1: Nhà nghiên cứu muốn ước tính cỡ mẫu cho một nghiên cứu có mục tiêu chính là xác định tỉ lệ bệnh tiểu đường. Theo y văn, tỉ lệ bệnh tiểu đường trong cộng đồng người trung niên là 10%. Nói cách khác, *p* = 0.10. Nhà nghiên cứu muốn ước tính số đối tượng cần thiết để ước tính tỉ lệ đó, và chấp nhận xác suất 95% là tỉ lệ có thể dao động trong khoảng 8% đến 12%. Nói cách khác e = (0.12 – 0.08) / 4 = 0.01, tức sai số là 1%.

Chúng ta có thể dùng công thức [1] để ước tính; trong đó, *z*a/2 = 1.96 (xem bảng 2), *p =* 0.10, và e = 0.01. Chúng ta có thể tính thủ công với R:

|  |
| --- |
| alpha=0.05; z = qnorm(0.05/2)  p=0.10; e = 0.01  n = (z/e)^2\*p\*(1-p) |

Hoặc dùng hàm **n.for.survey** trong **epicalc** như sau. Chú ý nếu chưa có epicalc trong R, cần phải cài đặt trước với lệnh **install.packages("epicalc")**. Trong epicalc, sai số e được gọi là delta.

|  |
| --- |
| library(epiDisplay)  n.for.survey(p=0.10, delta=0.01, alpha=0.05) |

Kết quả cho thấy n = 3457. Nói cách khác, nhà nghiên cứu cần tuyển ngẫu nhiên 3457 đối tượng cho công trình nghiên cứu.

Cách tính trên dành cho trường hợp dân số vô hạn (infinite). Trong trường hợp nhà nghiên cứu biết rằng một cộng đồng có dân số *N,* ví dụ như 100,000 người, thì cách tính có thay đổi:

**n.for.survey(p=0.10, delta=0.01, alpha=0.05, popsize=100000)**

Kết quả cho thấy n = 3342. Dĩ nhiên, thay đổi delta (tức e) sẽ thay đổi số cỡ mẫu khá lớn. Chẳng hạn như nếu nhà nghiên cứu chấp nhận sai số 2% (delta=0.02) thì số cỡ mẫu bây giờ là n = 857 người.

**21.2.2. Cỡ mẫu để ước tính số trung bình quần thể**

Lí thuyết: Chúng ta muốn ước tính giá trị trung bình μ của một biến trong quần thể (không biết). Tuy nhiên, chúng ta biết rằng độ lệch chuẩn của biến là σ. Với một mẫu gồm *n* đối tượng, chúng ta có thể ước tính giá trị trung bình mẫu *m,* và theo lí thuyết thống kê, giá trị trung bình quần thể sẽ nằm trong khoảng *m* ± *zα*/2(σ/√*n*) với khoảng tin cậy (1 – α)%. Gọi *zα*/2(σ/√*n*) là sai số *e*, giải phương trình trên để tính *n,* chúng ta có công thức ước tính cỡ mẫu đơn giản:

[2]



Ví dụ 2: Chúng ta muốn ước tính tuổi bị ung thư vú, và muốn làm một nghiên cứu cắt ngang. Thông tin mà chúng ta có là độ lệch chuẩn của độ tuổi ung thư là 12 tuổi, và chúng ta muốn ước tính độ tuổi ung thư chính xác trong khoảng ± 2.5 tuổi (tức độ rộng của khoảng tin cậy 95% là 5 tuổi). Nói theo ngôn ngữ thống kê của công thức [2], chúng ta có: *e* = 2.5 và σ = 12.

Dùng R:

|  |
| --- |
| alpha=0.05; z = qnorm(0.05/2)  sigma=12; e = 2.5  n = (z/e\*sigma)^2 ; n |

Kết quả là chúng ta cần n = 88 đối tượng cho công trình nghiên cứu.

**21.2.3. Cỡ mẫu để so sánh 2 tỉ lệ**

Lí thuyết: Trong trường hợp hai tỉ lệ π1 và π2, tham số mà chúng ta quan tâm là hiệu số δ = π1 – π2. Để kiểm định giả thuyết vô hiệu π1 = π2, chúng ta sẽ ước tính 2 tỉ lệ p1 và p2 từ mẫu nghiên cứu *n*1 và *n*2 đối tượng. Kiểm định *z* cho 2 tỉ lệ là:

Z = (*p*1 – *p*2) / *SD*

Trong đó, SD là độ lệch chuẩn của hiệu số:



và là tỉ lệ trung bình, tức = (*p*1 + *p*2) / 2.



Giả định rằng cỡ mẫu hai nhóm tương đương nhau và gọi chung là *n*, giải 2 phương trình trên, chúng ta sẽ có công thức ước tính cỡ mẫu (1):

[3]



Ví dụ 3: Một nghiên cứu lâm sàng có mục tiêu đánh giá hiệu quả của một thuốc điều trị bệnh ung thư vú. Mô hình nghiên cứu là RCT, nhà nghiên cứu chia bệnh nhân thành 2 nhóm: nhóm dùng thuốc hiện hành (nhóm 1) và nhóm dùng thuốc mới (nhóm 2). Biến kết quả là tỉ lệ sống sót sau 5 năm. Sau khi duyệt qua y văn, nhà nghiên cứu biết rằng tỉ lệ sống sót của nhóm 1 là *p*1 = 75%, và giả thuyết rằng thuốc mới sẽ tăng cải tiến tỉ lệ sống sót 20%, tức *p*2 = 90% (lấy 75\*1.2). Nhà nghiên cứu muốn ước tính cỡ mẫu sao cho nghiên cứu có xác suất 90% để “phát hiện” mức độ ảnh hưởng đó với xác suất sai sót loại I là 5%.

Sử dụng công thức [3] cho ví dụ trên, chúng ta có *p*1 = 0.75; *p*2 = 0.90; *z*α/2 = 1.96; zβ = 1.282. Chúng ta có thể dùng hàm **n.for.2p** trong **epicalc** để ước tính cỡ mẫu:

|  |
| --- |
| library(epiDisplay)  n.for.2p(p1=0.75, p2=0.90, power=0.9, alpha=0.05, ratio=1) |

Kết quả cho thấy n = 292 bệnh nhân (146 cho nhóm 1 và 146 cho nhóm 2). Chú ý trong hàm trên, power là 1 – β = 1 – 0.10 = 0.90. Tham số ratio = 1 có nghĩa là số cỡ mẫu của nhóm 1 bằng số cỡ mẫu của nhóm 2. Nếu chúng ta muốn số cỡ mẫu nhóm 2 hơn nhóm 1 gầp 2 lần thì ratio = 2.

**21.2.4. Cỡ mẫu để so sánh 2 số trung bình**

Lí thuyết: Đối với nghiên cứu so sánh hai nhóm độc lập với biến kết quả là biến liên tục, tham số quan tâm là hiệu số của hai số trung bình quần thể π1 và π2, tức δ = π1 –π2. Giả thuyết đặt ra là:

H0: π1 = π2 hay δ = 0

HA: π1 ≠ π2 hay δ ≠ 0

Dĩ nhiên, chúng ta không biết π1 và π2, mà chỉ có thể ước tính hai số trung bình *m*1 và *m*2 qua mẫu *n*1 và *n*2 đối tượng. Phương pháp căn bản để kiểm định giả thuyết là t-test. Gọi *m*1 và *m*2là số trung bình mẫu của hai nhóm, *s*1 và *s*2là độ lệch chuẩn, và *n*1 và *n*2là số đối lượng cho hai nhóm so sánh, kiểm định t được định nghĩa là hiệu số của hai số trung bình chia cho độ lệch chuẩn của hiệu số:



Trong đó,



Giả dụ rằng cỡ mẫu của hai nhóm bằng nhau *n*1 = *n*2  và gọi chung là *n = n*1 = *n*2 . Vấn đề là tìm *n* sao cho giá trị của *t* có nghĩa thống kê, tức sao cho trị số *P* của *t* thấp hơn α. Giải phương trình trên, chúng ta có (1):

[4]



Nhớ rằng mức độ ảnh hưởng ES là:



Do đó, công thức [4] cũng có thể viết lại đơn giản hơn:

[5]



Nếu α = 0.05 và β = 0.20 (tức power bằng 0.8), thì công thức trên còn đơn giản hơn:

[6]



Ví dụ 4: Để đánh giá hiệu quả của thuốc chống loãng xương, nhà nghiên cứu thiết kế một nghiên cứu lâm sàng gồm 2 nhóm bệnh nhân: Nhóm 1 được điều trị bằng thuốc mới, nhóm 2 là nhóm chứng (placebo). Biến kết quả là mật độ xương hay BMD (tính bằng g/cm2). Qua nghiên cứu sơ khởi, nhà nghiên cứu biết rằng BMD trung bình ở phụ nữ sau mãn kinh là *m*1 = 0.80 g/cm2 và độ lệch chuẩn là *s* = 0.12 g/cm2. Qua y văn, nhà nghiên cứu biết rằng thuốc có thể tăng BMD 5% sau 1 năm điều trị. Do đó, có thể nói mục tiêu là BMD trung bình của nhóm điều trị là *m*2 =0.8 x 1.05 = 0.84. Giả dụ rằng độ lệch chuẩn của nhóm 2 cũng là 0.12 g/cm2. Nói cách khác, độ ảnh hưởng ES là: ES = (0.84 – 0.80) / 0.12 = 0.33. Nhà nghiên cứu muốn có power là 80% và sai sót loại I là 5%. Công thức ước tính cỡ mẫu cho trường hợp này là:



Chúng ta có thể dùng hàm **n.for.2means** trong **epicalc** để ước tính cỡ mẫu:

|  |
| --- |
| library(epiDisplay)  n = n.for.2means(mu1=9.908, mu2=9.656, sd1=0.14, sd2=0.14, power=0.9, alpha=0.05, ratio=1) |

Trong hàm trên **mu1** và **mu2** là chỉ số trung bình của hai nhóm, **sd1** và **sd1** là độ lệch chuẩn của hai nhóm, và giả dụ rằng cỡ mẫu nhóm 1 bằng nhóm 2 (ratio = 1).

Kết quả cho thấy nhà nghiên cứu cần tuyển n = 284 bệnh nhân (142 cho mỗi nhóm) cho công trình nghiên cứu.

**21.2.5. Cỡ mẫu so sánh hai xác suất sống sót (survival curves)**

Lí thuyết: Đối với nghiên cứu so sánh hai nhóm độc lập với biến kết quả là biến thời gian sống sót (survival time), cách ước tính cỡ mẫu còn tuỳ thuộc vào số biến cố xảy ra. Do đó, có hai ước tính: số biến cố xảy ra cần quan sát và số đối tượng (bệnh nhân). Trong phân tích sống còn, chỉ số quan trọng nhất là hazard ratio (HR, tức tỉ số rủi ro). HR là tỉ số của hai *xác suất biến cố xảy ra* tại một thời điểm (kí hiệu là λ1 và λ2 cho 2 nhóm so sánh):



Nếu giả định rằng (và trong thực tế cũng phù hợp cho nhiều trường hợp) xác suất sống còn tuân theo luật phân bố mũ (exponential distribution), mối liên quan giữa λ và tỉ lệ sống sót π tại thời điểm *T* và trung vị thời gian sống còn *M* có thể tóm tắt:



Nhưng trong thực tế, rất khó có được λ1 và λ2, nên HR thường được ước tính dựa vào *xác suất sống sót* (kí hiệu π1 và π2):



Dĩ nhiên, qua công thức trên, chúng ta có thể phát biểu rằng: .



Trong nhiều trường hợp, chúng ta có thể nghĩ đến số trung vị thời gian sống sót (median survival time) *M*1 và *M*2 cho hai nhóm, và HR có thể định nghĩa là:



Với những lí giải trên, và nếu chúng ta giả định rằng số bệnh nhân nhóm 1 và nhóm 2 theo tỉ lệ 1/r, thì số biến cố cần quan sát cho nhóm 1 (với sai sót α và β) là (2):

[7]



và số biến cố cần quan sát cho nhóm 2: .



Tương tự, chúng ta có thể tính số bệnh nhân cho nhóm 1 là:

[8]



và số bệnh nhân cho nhóm 2: .



**Bảng 3: Số biến cố cần quan sát để so sánh hai hàm số sống còn với α = 0.05 và β = 0.20. Kết quả chỉ tính cho nhóm 1 (*e*1). Nhóm 2 là e2 = r\**e*1**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| π2 | π1 | | | | | | | | | |
| 0.05 | 0.10 | 0.15 | 0.20 | 0.25 | 0.30 | 0.35 | 0.40 | 0.45 | 0.50 |
| 0.10 | 230 |  |  |  |  |  |  |  |  |  |
| 0.15 | 78 | 422 |  |  |  |  |  |  |  |  |
| 0.20 | 44 | 126 | 584 |  |  |  |  |  |  |  |
| 0.25 | 30 | 64 | 163 | 708 |  |  |  |  |  |  |
| 0.30 | 22 | 40 | 79 | 189 | 793 |  |  |  |  |  |
| 0.35 | 17 | 29 | 48 | 89 | 206 | 839 |  |  |  |  |
| 0.40 | 14 | 22 | 33 | 53 | 95 | 214 | 851 |  |  |  |
| 0.45 | 12 | 17 | 24 | 35 | 55 | 96 | 213 | 832 |  |  |
| 0.50 | 11 | 14 | 19 | 25 | 36 | 55 | 94 | 205 | 787 |  |
| 0.55 | 9 | 12 | 15 | 19 | 25 | 35 | 53 | 89 | 191 | 721 |
| 0.60 | 8 | 10 | 12 | 15 | 19 | 25 | 33 | 49 | 82 | 172 |
| 0.65 | 8 | 9 | 10 | 12 | 15 | 18 | 23 | 31 | 44 | 73 |
| 0.70 | 7 | 8 | 9 | 10 | 12 | 14 | 17 | 21 | 27 | 39 |
| 0.75 | 6 | 7 | 8 | 9 | 10 | 11 | 13 | 15 | 18 | 23 |
| 0.80 | 6 | 6 | 7 | 7 | 8 | 9 | 10 | 11 | 13 | 15 |
| 0.85 | 5 | 6 | 6 | 6 | 7 | 7 | 8 | 9 | 9 | 11 |
| 0.90 | 5 | 5 | 5 | 6 | 6 | 6 | 6 | 7 | 7 | 8 |
| 0.95 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 6 | 6 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| π2 | π1 | | | | | | | | | |
| 0.50 | 0.55 | 0.60 | 0.65 | 0.70 | 0.75 | 0.80 | 0.85 | 0.90 | 0.95 |
| 0.55 | 721 |  |  |  |  |  |  |  |  |  |
| 0.60 | 172 | 638 |  |  |  |  |  |  |  |  |
| 0.65 | 73 | 149 | 544 |  |  |  |  |  |  |  |
| 0.70 | 39 | 62 | 125 | 444 |  |  |  |  |  |  |
| 0.75 | 23 | 32 | 51 | 99 | 343 |  |  |  |  |  |
| 0.80 | 15 | 19 | 26 | 39 | 74 | 246 |  |  |  |  |
| 0.85 | 11 | 12 | 15 | 20 | 29 | 51 | 159 |  |  |  |
| 0.90 | 8 | 9 | 10 | 11 | 14 | 19 | 31 | 87 |  |  |
| 0.95 | 6 | 6 | 6 | 7 | 8 | 9 | 11 | 15 | 33 |  |

Ví dụ 5: Một nghiên cứu lâm sàng theo mô hình RCT nhằm mục đích đánh giá hiệu quả của levamisole (một loại thuốc điều trị ung thư ruột) so với nhóm chứng (placebo). Qua y văn, biết rằng tỉ lệ sống sót trong vòng 1 năm của nhóm chứng là 50%, và hi vọng rằng nhóm được điều trị là 60%. Số biến cố cần quan sát để đạt α = 0.05 và β = 0.20 là 172 cho nhóm 1 và 172 cho nhóm 2.

Có thể áp dụng công thức [7] để tính. Trong đó, chúng ta có π1 = 0.5 và π2 = 0.60, HR = log(0.6) / log(0.5) = 0.7369, và nếu r = 1 (cỡ mẫu hai nhóm bằng nhau):



Cũng có thể tính bằng hàm epi.sscomps trong package epiR như sau:

|  |
| --- |
| library(epiR)  epi.sscomps(treat=0.60, control=0.50, n=NA, power=0.80, r=1, conf.level=0.95, sided.test=2) |

**21.2.6. Cỡ mẫu cho nghiên cứu bệnh chứng**

Trong nghiên cứu bệnh chứng, như đề cập trong phần đầu, mục tiêu là so sánh tỉ lệ phơi nhiễm yếu tố nguy cơ giữa nhóm bệnh và nhóm chứng. Gọi *E* là có yếu tố nguy cơ, *NE* là không có yếu tố nguy cơ, *D* là bệnh, *ND* là không mắc bệnh (nhóm chứng). Trong nghiên cứu bệnh chứng, chúng ta có thể ước tính π1 là xác suất *E* trong nhóm bệnh, và π2 là xác suất *E* trong nhóm chứng:

π1= P(E | D)

π2 = P(E | ND)

Giả thuyết đặt ra là:

H0: π1 = π2

HA: π1 / π2 = λ

Xin nhắc lại rằng chúng ta không biết π1 và π2 trong quần thể. Tuy nhiên, với mẫu nghiên cứu, chúng ta có thể ước tính hai chỉ số tương ứng là *p*1 và *p*2. Gọi *P* = xác suất phơi nhiễm yếu tố nguy cơ = P(*E*), chúng ta có thể nói rằng tỉ lệ quan sát *p*1 *P*. Với giả thuyết HA, chúng ta có tỉ lệ phơi nhiễm yếu tố nguy cơ trong nhóm chứng là:



Số cỡ mẫu *n* có thể ước tính từ các lí giải trên như sau (1):

[9]



Trong đó: r là tỉ số cỡ mẫu nhóm bệnh trên nhóm chứng (thường r = 1), và

[10]



Ví dụ 6: Một nghiên cứu bệnh chứng về mối liên quan giữa hút thuốc lá và bệnh tim mạch được hoạch định. Theo kế hoạch, một nhóm bệnh nhân và một nhóm chứng (cỡ mẫu như nhau, tức r = 1) sẽ được so sánh về tỉ lệ hút thuốc lá. Các nhà nghiên cứu đặt giả thuyết là tỉ số nguy cơ là 2 Khoảng 30% người trong cộng đồng hút thuốc lá. Câu hỏi đặt ra là cần bao nhiêu người cho nghiên cứu để có power 90% để phát hiện tỉ số đó có ý nghĩa thống kê ở mức độ 5%.

Trong trường hợp này, chúng ta có: *P =* 0.30, λ = 2, zα/2 = 1.96, và zβ = 1.2816. Thay các giá trị này vào phương trình [10]:



và số cỡ mẫu là:



Trong hàm trên, nhóm control và treat được cho giá trị sao cho λ = 2, sigma là tỉ lệ hút thuốc lá (tức yếu tố phơi nhiễm trong nhóm chứng).

**21.2.7. Cỡ mẫu cho nghiên cứu đoàn hệ**

Lí thuyết: Nghiên cứu đoàn hệ (prospective cohort study) thường có mục tiêu tiên lượng và đánh giá yếu tố nguy cơ liên quan đến bệnh. Ngược lại với nghiên cứu bệnh chứng (đi ngược thời gian), nghiên cứu đoàn hệ đi xuôi theo thời gian. Theo đó, nhà nghiên cứu chọn một nhóm đối tượng, đo lường các yếu tố tiên lượng hay yếu tố nguy cơ lúc đầu, và sau đó theo dõi một thời gian để thu thập số ca mắc bệnh. Chỉ số thống kê để đánh giá mối liên quan giữa một yếu tố nguy cơ (như hút thuốc lá chẳng hạn) và nguy cơ mắc bệnh là *tỉ số nguy cơ* (còn gọi là *risk ration* hay *relative risk –* RR). Gọi xác suất mắc bệnh (còn gọi là *nguy cơ – risk*) trong thời gian theo dõi ở nhóm có yếu tố nguy cơ là *p*1, và nhóm không có yếu tố nguy cơ là *p*2, RR được định nghĩa đơn giản là:

RR = *p*1 / *p*2

Vấn đề thống kê là ước tính số đối tượng và số bệnh cần thiết sao cho RR có ý nghĩa thống kê (tức *P* < α) với power 1-β.

Giả thuyết đặt ra là:

H0: RR = 1 (hay π1 = π2)

HA: RR ≠ 1 (hay π1 ≠ π2)

Trong đó, π1 và π2 là nguy cơ mắc bệnh ở nhóm có yếu tố nguy cơ và nhóm không có yếu tố nguy cơ trong quần thể. Tuy nhiên, với mẫu nghiên cứu, chúng ta có thể ước tính hai chỉ số tương ứng là *p*1 và *p*2. Phương pháp ước tính cỡ mẫu có thể dựa vào phương pháp ước tính cỡ mẫu cho hai tỉ lệ theo hướng RR. Có thể chứng minh qua một số thao tác toán, số cỡ mẫu cần thiết là (1):



[11]

Trong đó, π là tỉ lệ mắc bệnh trung bình của hai nhóm:



Trong trường hợp *r* = 1



Ví dụ 7: Nhà nghiên cứu muốn đánh giá mối liên quan giữa hút thuốc lá và bệnh tim mạch qua một nghiên cứu đoàn hệ mà đối tượng sẽ được theo dõi 5 năm. Trước đây, có nghiên cứu cho thấy tỉ lệ tử vong vì bệnh tim mạch mỗi năm ở người không hút thuốc lá là 413 / 100,000. Nhà nghiên cứu giả định rằng hút thuốc lá sẽ tăng nguy cơ tử vong 40% (nói cách khác, RR = 1.4). Nhà nghiên cứu muốn ước tính số cỡ mẫu (số người hút thuốc lá và không hút thuốc lá) cần thiết với điều kiện α = 0.05 (một chiều) và power = 90% (tức beta = 0.1).

Ở đây, chúng ta có thể ước tính vài thông số cần thiết như sau: tỉ lệ tử vong 5 năm nhóm không hút thuốc là π2 = 5 x 413/100000 = 0.02065, π1 = 0.02065 x 1.4 = 0.02891, z0.05 = 1.6449, z0.10 = 1.2816. Nếu r = 1, chúng ta có:



và:



Nói cách khác, nhà nghiên cứu cần tuyển 12130 cá nhân (6065 người hút thuốc và 6065 người không hút thuốc) cho công trình nghiên cứu.

Có thể dùng hàm **epi.studysize** trong **epiR** và có cùng kết quả:

|  |
| --- |
| library(epiR)  smoke=1.4\*(5\*413)/100000  nonsmoke=(5\*413)/100000  epi.sscohortc(irexp1 = irexp1, irexp0 = irexp0, n = NA, power = 0.90, r = 1, design = 1, sided.test = 1, conf.level = 0.95) |

**21.2.8. Cỡ mẫu cho nghiên cứu chẩn đoán**

Lí thuyết: Trong nghiên cứu chẩn đoán, nhà nghiên cứu hay quan tâm đến hai chỉ số quan trọng: độ nhạy (sensitivity) và độ đặc hiệu (specificity). Độ nhạy, có khi hiểu là *dương tính thật* hay *true positive*, là xác suất bệnh nhân có kết quả dương tính (+ve). Độ đặc hiệu, có khi hiểu là *âm tính thật* hay *true negative,* là xác suất một người không có bệnh có kết quả xét nghiệm âm tính (-ve). Gọi D là bệnh, ND là không bệnh, định nghĩa trên có thể viết theo ngôn ngữ xác suất như sau:

Sensitivity = P(+ve | D)

Specificity = P(-ve | ND)

Do đó, trong nghiên cứu chẩn đoán, chúng ta muốn biết cần bao nhiêu bệnh nhân (*nd*) để “chứng minh” một độ nhạy khác với độ nhạy. Thông thường, chúng ta muốn so sánh độ nhạy của nghiên cứu (*Se*) và với độ nhạy hiện tại (*Se*0). Phương pháp ước tính *nd* có thể dựa vào phương pháp ước tính cỡ mẫu cho 2 tỉ lệ (2):

[12]



Nếu nghiên cứu muốn so sánh hai độ nhạy (*Se*1 và *Se*2), công thức trên có thể chỉnh sửa thành (cho hai nhóm độc lập):

[13]



Công thức trên ước tính cỡ mẫu cho 2 nhóm độc lập. Trong nhiều trường hợp, nhà nghiên cứu có thể so sánh độ nhạy trong cùng một nhóm, và số bệnh cần thiết là:

[14]



trong đó,



Ví dụ 8: Nhà nghiên cứu muốn đánh giá độ nhạy của một xét nghiệm mới trong việc chẩn đoán rối loạn cường dương. Biết rằng trong cộng đồng có khoảng 55% người rối loạn cường dương. Nghiên cứu trước đây cho thấy độ nhạy của phương phát xét nghiệm hiện hành là 85%. Nhà nghiên cứu hi vọng rằng xét nghiệm mới sẽ có độ nhạy 95%, và muốn ước tính số cỡ mẫu cần thiết cho nghiên cứu với điều kiện α = 0.05 (một chiều) và β = 0.20.

Với những thông tin trên, chúng ta có z0.05 = 1.645, z0.20 = 0.842.

Dùng công thức [12], số cỡ mẫu (bệnh nhân) cần thiết là:



Nhà nghiên cứu cần tuyển 75 bệnh nhân và số không mắc bệnh (tính theo tỉ lệ hiện hành 55%) là 75 / 0.55 = 136. Tính chung, công trình nghiên cứu cần 211 cá nhân.

Ví dụ 9: Một nghiên cứu có mục tiêu là so sánh hai độ nhạy của hai phương pháp xét nghiệm để chẩn đoán bệnh viêm răng với mức độ ý nghĩa thống kê 5% (kiểm định hai chiều) và power 80%. Nhà nghiên cứu giả định rằng xét nghiệm 1 có độ nhạy là 66%, và xét nghiệm 2 có độ nhạy 27%. Số bệnh nhân cần thiết có thể ước tính như sau:



Nếu tỉ lệ viêm răng trong cộng đồng là 25%, số nhóm chứng có thể ước tính là 49 / 0.25 = 200. Tính chung, nhà nghiên cứu cần tuyển mộ 49 bệnh nhân và 200 người thuộc nhóm chứng cho công trình nghiên cứu.

Tuy nhiên, nếu hai xét nghiệm có thể làm trên 1 nhóm đối tượng thì:



và



Công thức [14] sẽ cho ra kết quả cỡ mẫu cần thiết:



Do đó, số cỡ mẫu cần xét nghiệm là 26 / 0.25 = 92 (giả định rằng tỉ lệ viêm răng là 25%).

**21.2.9. Cỡ mẫu cho đường biểu diễn ROC (receiver operating characteristic curve)**

Lí thuyết: Đường biểu diễn ROC là một chỉ số thống kê phản ảnh mức độ phân biệt (discrimination) bệnh và không bệnh của một phương pháp xét nghiệm. Trong chẩn đoán, khi phương pháp xét nghiệm tăng độ nhạy thì độ đặc hiệu sẽ giảm, và ngược lại, khi độ nhạy giảm thì độ đặc hiệu tăng. ROC là một “thỏa hiệp” giữa độ nhạy và độ đặc hiệu, để có độ nhạy và độ đặc hiệu tối ưu nhất.

Đường ROC được vẽ với trục tung là độ nhạy, và trục hoành là tỉ lệ dương tính giả (FPR – false positive rate). FPR thật sự là 1 trừ cho độ đặc hiệu. Diện tích dưới đường ROC được gọi là *area under the curve* (AUC). Mục tiêu của nghiên cứu chẩn đoán thường là (a) ước tính AUC, hoặc (b) so sánh các phương pháp xét nghiệm là so sánh AUC sao cho có ý nghĩa thống kê.

Đối với nghiên cứu có mục tiêu ước tính AUC trong một khoảng tin cậy *w,* công thức ước tính số bệnh nhân cần thiết là (3-4):



Trong đó:



Chú ý: FPR = 1 – độ đặc hiệu, và TPR = độ nhạy.

Ví dụ 10: Nhà nghiên cứu muốn ước tính số cỡ mẫu cần thiết để đánh giá độ chính xác của MRI trong việc chẩn đoán viêm khớp xương. Khoảng 40% cá nhân có viêm khớp xương. Qua y văn, nhà nghiên cứu biết rằng độ nhạy là 45% và độ đặc hiệu là 90%, và khoảng tin cậy của AUC *w =* 10% (tức 0.1).

Với các thông tin trên, chúng ta có thể tính: R = 0.6 / 0.4 = 1.5, FPR = 0.1, TPR = 0.45. Do đó, z0.1 = 1.2816, và z0.45 = 0.1257, A = 1.2816 – 0.1257 = 1.1559, σ = 0.3265 (các bạn có thể kiểm tra). Từ đó, số cỡ mẫu cần thiết là:



Số đối tượng nhóm chứng là 164 x 1.5 = 246. Tính chung, nhà nghiên cứu cần 164 + 246 = 410 đối tượng.

**R codes**

ci=0.1

sens = 0.80; qsens = qnorm(sens)

spec=0.85; qspec=qnorm(spec)

prev=0.60; r=(1-prev)/prev

a = qspec-qsens

s = exp(-a^2/4)/(2\*sqrt(3.1416))\*sqrt(1+1/r+(5\*a^2)/8+a^2/(8\*r))

n=4\*1.96\*(s^2/ci^2)

**21.3 Vài nhận xét**

Trong nghiên cứu khoa học, đặc biệt là những công trình nghiên cứu liên quan đến con người và động vật, việc xác định cỡ mẫu là một bước cực kì quan trọng. Quan trọng vì số lượng đối tượng nghiên cứu quyết định ngân sách và qui mô của nghiên cứu. Ngoài ra, đứng trên quan điểm y đức, số lượng đối tượng cũng liên quan đến đạo đức khoa học. Một nghiên cứu có quá ít đối tượng sẽ không có khả năng kiểm định giả thuyết và kết quả chẳng có ý nghĩa gì, và nghiên cứu đó được xem là vi phạm đạo đức. Ngược lại, một nghiên cứu có quá nhiều đối tượng hơn số cần thiết cũng là một vi phạm đạo đức, vì gây bất tiện không cần thiết cho những người tham gia. Đối với các nghiên cứu về thuốc, xác định cỡ mẫu còn có ý nghĩa pháp lí trong việc đăng kí thuốc. Hội nghị ICH (International Conference on Harmonization) (5) có qui định rằng các công ti dược muốn đăng kí dược phẩm phải đệ trình đầy đủ hồ sơ và dữ liệu, trong đó phải có phần phương pháp ước tính cỡ mẫu. Do đó, tuy là bước đầu trong nghiên cứu, nhưng việc xác định cỡ mẫu là một bước trung tâm trong thiết kế nghiên cứu.

Xác định cỡ mẫu là một khoa học, nhưng cũng có thể xem là một nghệ thuật. Tính khoa học của ước tính cỡ mẫu là cơ sở lí thuyết xác suất và giả định đằng sau mỗi công thức tính. Chính vì cơ sở khoa học này mà các phép tính phải mang tính tái thiết lập (repeatability). Nếu phương pháp tính toán được mô tả đầy đủ và minh bạch, người khác có thể đi đến cùng kết quả, và điều đó rất quan trọng trong tính minh bạch. Tuy nhiên, trong thực tế thì không hẳn như thế. Trong một thẩm định và phân tích 446 đề cương nghiên cứu đệ trình cho uỷ ban y đức ở Anh, các nhà nghiên cứu phát hiện chỉ có 134 (30%) đề cương mô tả cách tính đầy đủ và người khác có thể lặp lại kết quả tính toán, 70% còn lại thì kết quả không tái lập được (6). Đó là một kết quả bất ngờ và đáng quan tâm vì đây là những công trình nghiên cứu có liên quan đến con người.

Tính nghệ thuật trong việc xác định cỡ mẫu là cách chọn mức độ ảnh hưởng cho tính toán. Trong rất nhiều trường hợp, nhà nghiên cứu không biết được mức độ ảnh hưởng là bao nhiêu, không biết được độ nhạy hay độ chính xác của một xét nghiệm. Nhưng những thông số này lại rất cần cho việc tính toán cỡ mẫu! Do đó, nhà nghiên cứu phải vận dụng suy nghĩ để tìm một thông số cho tính toán. Một nghiên cứu phải cho ra kết quả có ích, do đó, thông số đó phải có ý nghĩa thực tế hay ý nghĩa lâm sàng. Cần phải phân biệt ý nghĩa thống kê và ý nghĩa lâm sàng. Khi một nghiên cứu có cỡ mẫu lớn, bất cứ một khác biệt nhỏ nào cũng có thể có ý nghĩa thống kê (P < 0.05), nhưng không có ý nghĩa lâm sàng. Ngược lại, một kết quả có thể có ý nghĩa lâm sàng nhưng không có ý nghĩa thống kê khi cỡ mẫu quá nhỏ. Do đó, nghệ thuật ở đây là phải chọn một thông số về ảnh hưởng vừa có ý nghĩa thống kê và có ý nghĩa lâm sàng. Như là một hướng dẫn chung, nếu mức độ ảnh hưởng (ES) khoảng 0.1-0.3 là “nhỏ”, 0.3-0.5 là “trung bình”, và >0.5 được xem là “lớn” (7). Việc chọn mức độ ảnh hưởng cho tính toán cỡ mẫu đòi hỏi nhà nghiên cứu phải suy nghĩ về vấn đề mình quan tâm và phù hợp với tình hình thực tế.

Tính toán cỡ mẫu phải được thực hiện trước khi triển khai nghiên cứu. Trong thực tế, có tình huống xảy ra khi kết quả nghiên cứu không có ý nghĩa thống kê, nhà nghiên cứu bèn tính toán power và số cỡ mẫu cần thiết để phát hiện mức độ ảnh hưởng quan sát. Cách làm như thế có khi được gọi là *post-hoc analysis of power.*  Nhưng ước tính cỡ mẫu và tính toán power sau khi công trình nghiên cứu đã hoàn tất hay sau khi xem qua kết quả hoàn toàn không có giá trị khoa học và phi logic (8-9). Cần nhớ rằng power là xác suất mà một kiểm định thống kê bác bỏ giả thuyết vô hiệu khi giả thuyết vô hiệu sai; do đó, power *chỉ* có ý nghĩa khi giả thuyết vô hiệu sai. Nhưng vấn đề là một kết quả không có ý nghĩa thống kê không nói cho chúng ta biết giả thuyết vô hiệu là đúng hay sai. Do đó, tính cỡ mẫu và power sau khi đã quan sát kết quả là hàm ý giả định rằng giả thuyết vô hiệu sai nhưng lại không phù hợp với dữ liệu thực tế. Cần phải tuyệt đối tránh tính toán cỡ mẫu và power sau khi quan sát một kết quả không có ý nghĩa thống kê. Khoảng tin cậy 95% cung cấp cho chúng ta nhiều thông tin hơn là một phân tích hậu định về power.

Một số nhà nghiên cứu (đặc biệt là ở Việt Nam) có thói quen tính toán cỡ mẫu sao cho ăn khớp với số đối tượng họ có thể tiếp cận được hay đã có sẵn. Chẳng hạn như nhà nghiên cứu đã có 30 bệnh nhân, và họ cố gắng điều chỉnh các thông số trong công thức ước tính cỡ mẫu sao cho kết quả là 30. Việc làm này dẫn đến nhiều kết quả vô lí, chẳng hạn như sai số mẫu gần bằng với tham số chính như tỉ lệ hiện hành! Đây là một thói quen, nếu không muốn nói là một sự “lươn lẹo” và phản khoa học không thể chấp nhận trong văn hoá khoa học. Như đề cập trên, việc ước tính cỡ mẫu phải được thực hiện trước khi triển khai nghiên cứu và phải được tiến hành một cách *có đầu tư suy nghĩ* và nghiêm chỉnh.

Xin nhấn mạnh một lần nữa là phương pháp ước tính cỡ mẫu phụ thuộc vào mô hình nghiên cứu, biến kết quả, sai sót trong kiểm định giả thuyết, và mức độ ảnh hưởng. Mỗi mô hình nghiên cứu với mỗi loại biến kết quả có một công thức ước tính cỡ mẫu đặc thù. Chẳng hạn như không thể dùng công thức [1] để ước tính cỡ mẫu cho nghiên cứu so sánh một biến liên tục giữa hai nhóm nghiên cứu. Tuy nhiên, trong thực tế ở Việt Nam, rất nhiều nhà nghiên cứu tính toán cỡ mẫu chỉ dựa vào công thức [1] cho hầu hết các mô hình nghiên cứu! Việc áp dụng công thức tính toán một cách tuỳ tiện như thế là một sai lầm rất căn bản và khó chấp nhận được. Sai lầm này có khi nguy hiểm, vì cho ra kết quả sai, làm ảnh hưởng đến chất lượng nghiên cứu.

Như có thể thấy qua các trường hợp vừa trình bày, phương pháp ước tính cỡ mẫu dựa trên một giả định rất đơn giản: cái gì nhỏ thì rất khó tìm (cần nhiều cỡ mẫu), và cái gì lớn thì dễ tìm (cần ít cỡ mẫu). Gắn kết với cỡ mẫu là khái niệm power (mà có nhiều người dịch là *lực nghiên cứu*, nhưng trong thực tế là *độ nhạy của nghiên cứu*). Độ nhạy của nghiên cứu là xác suất bác bỏ giả thuyết vô hiệu khi giả thuyết vô hiệu sai. Cỡ mẫu càng cao, độ nhạy càng cao; cỡ mẫu càng thấp, độ nhạy cũng giảm theo. Trong thực tế, có tình trạng thiếu nhất quán giữa công thức ước tính cỡ mẫu trong đề cương nghiên cứu và trong báo cáo khoa học. Một phân tích mới công bố trên BMJ (6) cho thấy chỉ có 11 trong số 62 nghiên cứu mà mô tả về cách tính cỡ mẫu nhất quán giữa đề cương và bài báo! Sự thiếu nhất quán như thế dẫn đến khó khăn trong diễn giải kết quả. Do đó, những thông số được sử dụng trong ước tính cỡ mẫu cần phải được mô tả cẩn thận và rõ ràng, vì những thông tin này rất có ích cho việc đánh giá tính khả tín của kết quả nghiên cứu. Những ví dụ trong bài này chủ yếu liên quan đến nghiên cứu y khoa (vốn đa dạng), nhưng cũng áp dụng cho các nghiên cứu xã hội học. Hi vọng rằng những chỉ dẫn trên đây sẽ giúp cho các nhà nghiên cứu hiểu hơn về ý tưởng và ý nghĩa của việc ước tính cỡ mẫu cho một công trình nghiên cứu khoa học.

**Tài liệu tham khảo**

(1) Woodward M. Epidemiology: Study Design and Data Analysis, Third Edition, Chapter 8. Chapman & Hall 2013.

(2) Machin D, Campbell MJ, Tan SB, Tan SH. Sample size tables for clinical studies, 3rd Edition, Chapter 13. Blackwell Publishing 2009.

(3) Obuchowski NA, McClish DW. Sample size determination for diagnostic accuracy studies involving binomial ROC curve indices. Stat Med 1997;16:1529-42.

(4) Li J, Fine J. One sample size for sensitivity and specificity in prospective diagnostic accuracy studies. Stat Med 2004;23:2537-50.

(5) Clark T, Berger U, Mansmann U. Sample size determinations in original research protocols for randomised clinical trials submitted to UK research ethics committees: review. BMJ 2013 Mar 21;346:f1135.

(6) Chan AW, Hróbjartsson A, Jørgensen K, Gøtzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. BMJ 2008;337:a2299

(7) Cohen, J. Statistical power analysis for the behavioral sciences (2nd ed.). New Jersey: Lawrence Erlbaum, 1988.

(8) Hoenig JM, Heisey DM. The abuse of power: the pervasive fallacy of power. The American Statistician 2001;55:1-6.

(9) Onwuegbuzie AJ, Leech NL. Post-hoc power: A concept whose time has come. Understanding Statistics 2004;3(4):201-230.