

Cỡ mẫu: huyền thoại con số 30



Mỗi lần nói chuyện về phương pháp ước tính cỡ mẫu cho nghiên cứu lâm sàng là mỗi lần có người

muốn tôi bình luận câu phát biểu [đại khái] rằng có nhiều thầy cô trong nước nói rằng cỡ mẫu tối thiểu cho nghiên cứu khoa học là 30 đối tượng. Thật ngạc nhiên! Theo tôi hiểu, không có ai có thể cho ra một con số về cỡ mẫu tùy tiện như thế cả. Có lẽ đó chỉ là một hiểu lầm, và khi tích lũy lâu ngày nó trở thành một huyền thoại.

Trong nghiên cứu lâm sàng, hay nghiên cứu khoa học nói chung, cỡ mẫu có khi đóng vai trò quan trọng. Quan trọng là vì ngân sách của công trình nghiên cứu tùy thuộc vào số lượng cỡ mẫu; nghiên cứu với số đối tượng cao thì sẽ đắt tiền hơn nghiên cứu với số đối tượng thấp. Có khi các nhà tài trợ quyết định cung cấp kinh phí cho nghiên cứu dựa vào số lượng đối tượng cần thiết cho nghiên cứu. Tính khả thi của một công trình nghiên cứu có thể đánh giá qua số lượng đối tượng cần tuyển và khả năng mà nhà nghiên cứu có thể quản lý đối tượng nghiên cứu.

Hơn thế nữa, trong nghiên cứu y khoa, số cỡ mẫu còn có ý nghĩa y đức. Nếu nghiên cứu có quá ít bệnh nhân thì sẽ rất khó đi đến một quyết định đáng tin cậy. Nếu nghiên cứu có quá nhiều bệnh nhân thì có lẽ không cần thiết. Trong cả hai trường hợp (nhiều và ít) đều có thể xem là vi phạm y đức. Nếu nghiên cứu

nhiều hơn cần thiết có thể gây phiền toái cho bệnh nhân và có thể họ ở trong tình huống nguy hiểm (nếu là nghiên cứu can thiệp mang tính xâm phạm).

Ước tính cỡ mẫu

Nhưng cỡ mẫu là một ước tính. Phương pháp ước tính tùy thuộc vào nhiều giả định và thông số. Nói một cách ngắn gọn, số cỡ mẫu tùy thuộc vào 5 yếu tố chính như sau:

- *Mô hình nghiên cứu.* Trong nghiên cứu y khoa, có nhiều mô hình nghiên cứu (như mô hình bệnh chứng, nghiên cứu theo thời gian, nghiên cứu cắt ngang, v.v.) Mỗi mô hình có một cách tính cỡ mẫu khác nhau, và mức độ phức tạp của tính toán càng tăng theo sự phức tạp của mô hình nghiên cứu.

- *Biên đánh giá* (còn gọi là evaluative outcome). Mỗi nghiên cứu đều có một biên số hay chỉ tiêu chính để đánh giá / kiểm định giả thuyết. Nếu nghiên cứu đánh giá hiệu quả của thuốc điều trị loãng xương, thì biên đánh giá có thể là mật độ xương hay tình trạng gãy xương. Có hai loại biên đánh giá: biên liên tục (continuous variable) và biên phân loại (categorical variable). Trong ví dụ vừa kể, mật độ xương là biên liên tục, còn gãy xương là biên phân loại (hoặc là bị gãy, hoặc là không gãy). Phương pháp ước tính cỡ mẫu cho biên liên tục rất khác với phương pháp cho các biên phân loại.

- *Mức độ ảnh hưởng.* Nghiên cứu lâm sàng thường có mục đích đánh giá ảnh hưởng của một can thiệp hay tìm mối liên quan. Mức độ ảnh hưởng cao thì số lượng đối tượng cần thiết sẽ thấp. Nếu mức độ

ảnh hưởng thấp hay mối tương quan yếu thì số lượng đối tượng sẽ tăng.

- *Độ biến thiên của độ ảnh hưởng.* Biết mức độ ảnh hưởng vẫn chưa đủ, mà còn phải biết mức độ dao động của ảnh hưởng, gọi tắt là biến thiên (variability). Độ biến thiên thường đo bằng một chỉ số thống kê có tên là độ lệch chuẩn (standard deviation). Độ lệch chuẩn càng cao, số cỡ mẫu càng lớn; ngược lại, độ lệch chuẩn thấp thì số cỡ mẫu cần thiết cũng thấp.

- *Mức độ sai sót mà nhà nghiên cứu chấp nhận.* Đó là sai sót loại I (type I error) và sai sót loại II (type II error). Sai sót loại I là xác suất nhà nghiên cứu kết luận có ảnh hưởng trong khi trong thực tế thì không có, tương tự như xác suất dương tính giả trong chẩn đoán bệnh (bệnh nhân không có bệnh, nhưng kết quả xét nghiệm dương tính). Sai sót loại II là xác suất nhà nghiên cứu kết luận không có ảnh hưởng nhưng trong thực tế thì có. Thông thường, sai sót loại I có thể chấp nhận được là <5%, và sai sót loại II là <20%.

Để ước tính cỡ mẫu đúng, cần phải biết 5 thông tin vừa mô tả. Không có một trong những thông tin trên thì không thể nào ước tính cỡ mẫu được. Do đó, không có một con số cụ thể về cỡ mẫu. Hoàn toàn không có một con số cố định hay phỏng đoán như 30. Một nghiên cứu có thể cần 20 bệnh nhân, nhưng cũng có thể cần đến hàng ngàn bệnh nhân, tùy vào giả định và các thông số trên. Sau đây là hai ví dụ để minh chứng cho phát biểu trên.

Ví dụ 1: Nhà nghiên cứu muốn đánh giá hiệu quả của một loại thuốc mới chống loãng xương. Nghiên cứu có 2 nhóm: nhóm chứng và nhóm can thiệp. Giả

định rằng thuốc sẽ tăng mật độ xương khoảng 0,05 g/cm², với độ lệch chuẩn là 0,12 g/cm². Ngoài ra, nhà nghiên cứu sẵn sàng chấp nhận sai sót loại I là 5% và sai sót loại II là 20%. Với các thông số trên, có thể ước tính rằng nhà nghiên cứu cần 180 bệnh nhân (90 cho nhóm chứng và 90 cho nhóm điều trị).

Ví dụ 2: Tiếp theo ví dụ trên, giả dụ nhà nghiên cứu nghĩ rằng thuốc có hiệu quả rất cao. Mức độ khác biệt giữa hai nhóm là 0,15 g/cm² (thay vì 0,05 g/cm²), và các giả định / thông số khác không thay đổi. Với kì vọng mới, nhà nghiên cứu chỉ cần 20 bệnh nhân (10 người cho mỗi nhóm).

Hiểu lầm?

Không ai biết rõ con số cỡ mẫu 30 được lưu truyền ở Việt Nam từ lúc nào, và tại sao nhiều người tin vào nó. Tuy nhiên, con số này xuất phát từ một sự hiểu lầm, hay hiểu chưa đúng về phân bố thống kê và cỡ mẫu. Một số sách giáo khoa thường có một phát biểu chung chung về số cỡ mẫu “lớn” và “nhỏ”. Chẳng hạn như cuốn Probability and Statistical Inference của Hogg và Tanis có viết rằng cỡ mẫu dưới 25 hay dưới 30 được xem là “nhỏ”, và trên con số đó là “lớn”. Nhưng ngưỡng nhỏ/lớn này không phải là cỡ mẫu cho nghiên cứu khoa học, mà là ngưỡng để tính xấp xỉ giữa phân bố chuẩn (normal distribution) và phân bố t (t distribution). Chúng ta biết rằng phân bố t là xấp xỉ phân phối chuẩn. Khi cỡ mẫu lớn, phân bố t và phân bố chuẩn gần như giống nhau. Ở đây, “lớn” có nghĩa là trên 30. Con số này, do đó, chẳng liên quan gì đến số cỡ mẫu cho nghiên cứu khoa học.

(xem tiếp trang 14)

(tiếp theo trang 9)

Có thể một hiểu lầm khác là liên quan đến Định lý giới hạn trung tâm (central limit theorem - CLT). Đại khái, CLT phát biểu rằng bất cứ chỉ số thống kê nào đều tuân theo luật phân bố bình thường hay gần bình thường nếu số cỡ mẫu đủ. (Định lý này không dễ hiểu và thường bị diễn giải sai, nhưng đây là đề tài tôi sẽ trở lại sau). “Đủ” ở đây thường được hiểu là trên 30. Cần nhấn mạnh rằng đây là một xấp xỉ về phân bố của chỉ số thống kê (statistic) như trung bình, trung vị, tỉ lệ, độ lệch chuẩn, v.v. chứ chẳng liên quan gì đến cỡ mẫu nghiên cứu.

Nói tóm lại, số cỡ mẫu cho một nghiên cứu khoa học tùy thuộc vào một số thông số và giả định. Không bao giờ có một con số 30 cỡ mẫu (hay bất cứ con số

cố định nào) áp dụng cho tất cả nghiên cứu. Có thể nói rằng những lưu truyền ở Việt Nam về con số 30 như là số đối tượng cần thiết cho nghiên cứu khoa học chỉ là một huyền thoại, và huyền thoại này có thể xuất phát từ một hiểu lầm.

Ghi thêm:

Bài này không bàn về phương pháp ước tính cỡ mẫu, nhưng các bạn có thể tham khảo một số bài trước đây để biết thêm về phương pháp:

Những sai sót phổ biến trong ước tính cỡ mẫu

Phương pháp ước tính cỡ mẫu cho một nghiên cứu y học.

Nguyễn Văn Tuấn (Tác giả) - Đoàn Dũng (sưu tầm)

Nguồn: <http://statistics.vn/index.php?>