



NGUYỄN VĂN TUẤN

**Cẩm nang
Nghiên cứu khoa học**

Tủ ý tưởng đến công bố

NHÀ XUẤT BẢN TỔNG HỢP
THÀNH PHỐ HỒ CHÍ MINH





Tác giả Nguyễn Văn Tuấn là giáo sư Trường Y, Đại học New South Wales (UNSW Sydney), giáo sư Y khoa Tiên lượng (Predictive Medicine) thuộc Đại học Công nghệ Sydney (UTS), giáo sư adjunct Dịch tễ học và Thống kê học thuộc Trường Y, Đại học Notre Dame Australia. Tác giả là trưởng một labo nghiên cứu loãng xương thuộc Viện Nghiên cứu Y khoa Garvan (Úc), và giám đốc labo nghiên cứu cơ xương thuộc Trường Đại học Tôn Đức Thắng. Ông là người thứ 33 được UNSW Sydney trao bằng D.Sc (Tiến sĩ Khoa học) về những đóng góp xuất sắc cho chuyên ngành loãng xương trên thế giới. Ông đã công bố gần 300 công trình nghiên cứu trên các tạp san nổi tiếng trên thế giới, kể cả *Nature*, *Science*, *JAMA*, *BMJ*, *Lancet*, và *New England Journal of Medicine*. Tác giả là một trong những nhà nghiên cứu y khoa được trích dẫn nhiều nhất trên thế giới. Tác giả là thành viên hoặc biên tập trong hội đồng biên tập của các tạp san khoa học quốc tế như *Journal of Bone and Mineral Research*, *JBMR Plus*, *Journal of Clinical Endocrinology and Metabolism*, *Journal of the Endocrine Society*, *Osteoporosis International*, *PLoS*, *PeerJ*, và *BMC*. Ở Việt Nam, trong 15 năm qua ông đã giảng cho khoảng 20 workshop với tổng số hơn 2000 bác sĩ và nhà khoa học tham dự. Ông đã xuất bản 10 cuốn sách về khoa học, y tế và giáo dục ở trong nước. Ông được trao nhiều giải thưởng ở nước ngoài và trong nước về những thành tích khoa học và giáo dục.

**CẨM NANG NGHIÊN CỨU KHOA HỌC:
Từ ý tưởng đến công bố**

Nguyễn Văn Tuấn

Chịu trách nhiệm xuất bản

**Giám đốc - Tổng Biên tập
ĐINH THỊ THANH THỦY**

Biên tập: Cao Bá Định

Trình bày: Mạnh Hải

Bìa: Nguyễn Ngọc Khôi

NHÀ XUẤT BẢN TỔNG HỢP THÀNH PHỐ HỒ CHÍ MINH

62 Nguyễn Thị Minh Khai, Q.1, TP.HCM

ĐT: (028) 38225340 - 38296764 - 38247225 - Fax: 84.28.38222726

Email: tonghop@nxbhcm.com.vn

Sách online: www.nxbhcm.com.vn / Ebook: www.sachweb.vn

XNĐKXB số: 4898-2020/CXBIPH/46-332/THTPHCM

QĐXB số: 162/QĐ-THTPHCM-EBOOK2020 ngày 09/12/2020

ISBN: 978-604-312-279-4. Lưu chiều qui I/ 2021

ISBN: 978-604-312-279-4



9 786043 122794

NGUYỄN VĂN TUẤN

CẨM NANG NGHIÊN CỨU KHOA HỌC

Từ ý tưởng đến công bố



NHÀ XUẤT BẢN TỔNG HỢP THÀNH PHỐ HỒ CHÍ MINH

LỜI NÓI ĐẦU

Cuốn sách bạn đang cầm trên tay được soạn ra để giúp cho các bạn muốn dấn thân vào công việc nghiên cứu khoa học. Nhiều học sinh và sinh viên thường tự hỏi bắt đầu nghiên cứu khoa học từ đâu, ý tưởng nghiên cứu xuất phát từ nguồn nào, phương pháp nghiên cứu ra sao, và công bố kết quả nghiên cứu ở đâu. Đó là những câu hỏi cơ bản trong qui trình nghiên cứu khoa học. Qua 21 chương sách, tôi sẽ giải đáp những câu hỏi trên và kèm theo những ví dụ thực tế giúp bạn đọc hiểu rõ hơn về qui trình nghiên cứu khoa học.

Hai chữ “*Khoa học*” mà chúng ta dùng có nguồn gốc từ chữ “*science*” trong tiếng Anh (và Pháp); chữ *science* xuất phát từ chữ *scientia* trong tiếng Latin, có nghĩa là *kiến thức*. Nghiên cứu khoa học là một hoạt động quan trọng của con người nhằm mô tả các hiện tượng và qui luật tự nhiên bằng *phương pháp khoa học*. Mục tiêu của nghiên cứu khoa học là sản sinh ra dữ liệu mới, thông tin mới, và kiến thức mới. Cuốn sách này sẽ dẫn bạn đọc qua những mô hình nghiên cứu để thu thập dữ liệu cần thiết, những phương pháp phân tích để ‘hoán chuyển’ dữ liệu thô thành thông tin, và cách lí giải để biến thông tin

thành kiến thức. Những kiến thức mới giúp nâng cao phẩm chất cuộc sống và giúp cho xã hội tự tin hơn, giảm sự lệ thuộc vào những niềm tin thiếu căn cứ. Qua làm quen với nghiên cứu khoa học và phương pháp khoa học cũng giúp cho chúng ta tập thói quen phân tích và lí giải các hiện tượng tự nhiên một cách logic hơn, và xử lí thông tin tốt hơn.

Việc ứng dụng các nguyên lí khoa học vào mục tiêu cải thiện cuộc sống của con người có một lịch sử lâu dài. Theo một quan điểm khá phổ biến (nhưng vẫn còn tranh cãi) thì dấu vết của ứng dụng khoa học trong đời sống con người có thể tìm thấy trong nền văn minh Lưỡng Hà, tức khoảng 4.500 năm trước đây, khi mà người Sumerian sáng tạo ra chữ viết, toán học, lịch, và thiên văn học để cải tiến nông nghiệp, chính trị, và tôn giáo. Tuy nhiên, việc ứng dụng phương pháp khoa học vào đời sống con người chỉ thực sự phát triển mạnh kể từ thế kỷ 14 trở đi. Có thể gọi đó là một thời kì cách mạng khoa học. Nói “cách mạng” cũng không ngoa, bởi vì các tư tưởng khoa học đã giải phóng con người khỏi những tăm tối, ở thời kì mà các ý thức hệ tôn giáo mộng mị thống trị niềm tin của con người qua hàng ngàn năm. Cho đến nay, ai cũng phải công nhận rằng chính khoa học xây dựng nên thế giới hiện đại. Khoa học cho chúng ta xe ô-tô, xe tăng, máy bay, điện lực, điện thoại, máy điện toán, mạng, v.v. Khoa học giúp cho con người du hành lên mặt trăng. Khoa học góp phần làm tăng tuổi thọ của con người và con người sống lâu gấp hai lần so với 150 năm trước đây. Danh sách này có thể còn dài, nhưng thiết tưởng những lợi ích của khoa học như thế cũng đủ để thấy sự đóng góp của khoa học vào xã hội cực kì to lớn, và nó là một bộ phận không thể thiếu được trong xã hội hiện đại.

Trong thế kỉ 21, khoa học và công nghệ được xem là yếu tố hàng đầu để nâng tầm của một quốc gia trong nền kinh tế tri thức. Một nền kinh tế tri thức tốt, như tên gọi, được phát triển từ tri thức khoa học và công nghệ, hơn là lệ thuộc vào tài nguyên thiên nhiên và thể lực. Một ứng dụng máy tính hay một sản phẩm điện thoại thông minh có thể giúp cho quốc gia thoát nghèo và sánh vai cùng các nước tiên tiến. Nhưng trong thực tế cho đến nay, hầu hết các chỉ số về kinh tế tri thức của Việt Nam vẫn còn thấp so với các nước trong vùng. Tri thức khoa học, như đề cập trên, được sản sinh từ nghiên cứu khoa học. Do đó, Việt Nam cần nhiều người tham gia vào nghiên cứu khoa học để giúp nâng cao vị thế của nước nhà trong quá trình cạnh tranh trở thành một nền kinh tế tri thức. Cuốn sách nhỏ này có ý nguyện đóng góp một phần nhỏ vào công cuộc chung đó.

Cuốn sách này chỉ là một tập tài liệu ngắn nhằm tổng quan qui trình và nghiên cứu khoa học. Nội dung không đi vào chi tiết về phương pháp, vốn là chủ đề của các cuốn sách khác. Mục tiêu của tôi là cung cấp cho các bạn mới bắt đầu làm nghiên cứu khoa học thấy được hành trình trước mắt để chuẩn bị. Cuốn sách được soạn từ những bài giảng trong các lớp tập huấn về nghiên cứu khoa học do tôi thực hiện trong thời gian hơn 15 năm qua ở Việt Nam. Qua những lớp tập huấn này, tôi đã rút ra nhiều bài học từ học viên và qua hợp tác nghiên cứu. Những lớp tập huấn đó được sự giúp đỡ của rất nhiều người mà tôi muốn ghi nhận và cảm ơn. Tôi đặc biệt cảm ơn Tiến sĩ Trần Sơn Thạch (Viện Nghiên cứu Y khoa Garvan) và Tiến sĩ Hà Tấn Đức (Bệnh viện Đa khoa Trung ương Cần Thơ) đã giúp tôi và trợ giảng trong nhiều năm qua trong các lớp học về phương pháp nghiên cứu khoa học.

Khoa học không bao giờ dừng lại trong hành trình đi tìm sự thật, bởi vì khoa học không bao giờ tuyên bố nó đã hoàn thành sứ mệnh đó. Khoa học là một ngành nghề khai hóa, truyền bá cái mà chúng ta gọi là “văn minh”, và khoa học đặt sự thật trên hết và trước hết, trước tất cả những quyền lợi cá nhân. Đây là những lời phát biểu to tát mang tính nguyên tắc, nhưng trong thực tế, tất cả các nhà khoa học nghiêm túc đều chia sẻ. Tôi hi vọng rằng bạn đọc sẽ tìm thấy những ý tưởng và thông tin trong sách giúp các bạn dẫn thân vào con đường nghiên cứu khoa học.

Sydney 2/4/2018

NGUYỄN VĂN TUẤN

TẠI SAO NGHIÊN CỨU KHOA HỌC?

Trong nhiều buổi nói chuyện với sinh viên và các chương trình seminar ở Việt Nam, nhiều người thường đặt câu hỏi cho tôi là *"Tại sao cần nghiên cứu khoa học?"*. Có người trong ngành y nghĩ rằng không cần làm nghiên cứu khoa học, chỉ cần làm tốt điều trị lâm sàng là đủ. Tôi nghĩ câu trả lời có thể tiếp cận từ 4 cấp lý do: cá nhân, trường/viện, quốc gia, và quốc tế.

Đối với một số người, khoa học và nghiên cứu khoa học là một hoạt động xa rời thực tế, và từ đó thường đặt câu hỏi *"Tại sao cần nghiên cứu khoa học?"*. Khi đọc những bài báo khoa học, với những ý tưởng khó cảm nhận, những phương pháp có vẻ 'cầu kì', những kết quả phức tạp, đến những thuật ngữ khó hiểu, tất cả làm cho khoa học trở nên xa lạ với công chúng. Có khi nhìn chung quanh, người ta thấy những người không làm khoa học nhưng có thể sáng chế ra những phương tiện hữu dụng cho công việc đồng áng và cuộc sống hàng ngày, và từ đó người ta có vẻ nghi ngờ giá trị của khoa học và nghiên cứu khoa học. Hơn nữa, khoa học thường không có những sản phẩm cụ thể có thể sử dụng được ngay, nên câu hỏi 'tại sao cần nghiên cứu khoa học' thường được đặt ra.

Quay lại câu hỏi tại sao cần nghiên cứu khoa học, tôi nghĩ có thể tiếp cận qua cá nhân, trường/viện, quốc gia, và quốc tế. Ở mức độ cá nhân, có thể kể đến 5 lí do chính sau đây:

1. Nghiên cứu khoa học là một phương tiện để thu thập kiến thức và học hỏi. Nghiên cứu khoa học là một qui trình đi tìm sự thật và câu trả lời cho những câu hỏi có ích. Trong quá trình làm nghiên cứu khoa học, chúng ta phải thu thập những thông tin liên quan đến chủ đề nghiên cứu, nhận ra những vấn đề còn bỏ trống, và đặt câu hỏi mới. Qua khảo sát thông tin về những vấn đề liên quan, người làm nghiên cứu có cái nhìn rộng và lớn hơn những gì mình đang làm thực tế. Do đó, nghiên cứu khoa học có thể xem như là một phương tiện để hệ thống hóa kiến thức và cơ hội học hỏi chuyên sâu về một lĩnh vực hay một chuyên ngành.

2. Nghiên cứu khoa học là một phương tiện để đi tìm sự thật và bác bỏ kiến thức sai trái. Trong cuộc sống hàng ngày, chúng ta gặp rất nhiều những phát biểu đáng nghi ngờ mà người phát biểu khẳng định như là những sự thật. Chẳng hạn như có người cho rằng mỗi năm Việt Nam có 94.000 người chết vì bệnh ung thư, và con số này cao gấp 9 lần con số tử vong vì tai nạn giao thông (2), nếu chỉ mới nghe qua thì có vẻ phù hợp với cảm nhận của chúng ta, vì chúng ta ai cũng biết có ít nhất một người thân trong dòng họ bị bệnh ung thư và tử vong vì ung thư. Nhưng nếu suy nghĩ thêm, chúng ta sẽ tự hỏi những con số đó xuất phát từ đâu, làm thế nào để có thể biết một bệnh nhân ung thư chết vì ung thư mà không phải là vì một biến chứng khác. Những câu hỏi đó có thể không bao giờ có câu trả lời bởi vì những con số và suy đoán đó chỉ là những ý kiến cá nhân, chứ không phải từ nghiên cứu khoa học, và đó chính là một cơ hội để chúng ta thực hiện một nghiên cứu mới. Khi chúng ta làm nghiên cứu khoa học, kết quả có thể tiết lộ sự thật và phản bác (hay xác nhận) những phát biểu sai trái trước đây.

3. Nghiên cứu khoa học là một cách thể dục tinh thần và tâm thần. Nghiên cứu khoa học là một cơ hội để hiểu vấn đề và để khai thác những lựa chọn khả dĩ. Để hiểu và khai thác vấn đề, nhà nghiên cứu cần phải đọc và viết... rất nhiều. Nhiều người ngoài khoa học có lẽ không biết rằng đọc và viết là hai 'nhiệm vụ' và cũng là hai kỹ năng rất quan trọng của nhà khoa học. Đọc giúp cho nhà khoa học thu thập thông tin, và giúp mở mang trí não đến những khung trời kiến thức mới. Viết là kỹ năng quan trọng để nhà khoa học có thể chuyển tải thông tin đến công chúng.

Ngoài hai nhiệm vụ đọc và viết, nhà khoa học còn phải biết... nói và nghe. Ở một số chuyên ngành (như nghiên cứu khoa học xã hội và nghiên cứu định tính), nói và nghe là hai kỹ năng thiết yếu. Kỹ năng nói và nghe giúp nhà khoa học có thể thu thập thông tin qua phỏng vấn, tương tác với đối tượng nghiên cứu, và tương tác với đồng nghiệp trong các hội nghị. Nói và nghe cũng là hai phương thức để suy nghĩ có hệ thống, và qua đó nhà khoa học có thể sắp xếp thông tin và ý tưởng theo chủ đề. Tất cả những kỹ năng đọc, viết, nghe, và nói dĩ nhiên là những phương tiện luyện tập thể dục tinh thần rất hiệu quả.

4. Nghiên cứu khoa học là một cách để xây dựng sự nghiệp, thăng tiến và hội nhập quốc tế. Nghiên cứu khoa học là một cách giúp chúng ta đạt được mục tiêu trong đời. Những mục tiêu này có thể là có việc làm, có học bổng, được tài trợ cho nghiên cứu, hợp tác trong kinh doanh, và... du lịch. Đối với một bác sĩ hay nhà khoa học, nghiên cứu khoa học giúp họ trở thành người chủ động sáng tạo ra tri thức mới thay vì người tiếp thu kiến thức một cách thụ động. Đối với một chuyên gia nói chung, nghiên cứu khoa học giúp họ có thêm cái nhìn rộng lớn hơn về chuyên ngành của mình và quan trọng hơn là có thêm kỹ năng khảo sát, và qua đó nâng cao khả năng tìm được việc, và khi tìm được việc giúp họ có cơ may thăng tiến tốt hơn.

Trong hoạt động khoa học, nhà khoa học phải trình bày báo cáo nghiên cứu trong các hội nghị cấp quốc gia và quốc tế, và qua đó có cơ hội đi khám phá ở những nơi xa xôi và có cơ may tiếp cận với những nền văn hóa khác. Có thể nói không ngoa rằng cơ hội giao lưu với các đồng nghiệp nước ngoài cũng là một đặc lợi của giới khoa học.

Nghiên cứu khoa học thường dẫn đến công bố những bài báo khoa học trên các tập san quốc tế. Những bài báo khoa học được ví von như là những viên gạch để xây dựng sự nghiệp của một nhà khoa học. Trong thực tế, các đại học tuyển hay bổ nhiệm các chức vụ giáo sư thường dựa vào số ấn phẩm khoa học và phẩm chất của các ấn phẩm đó. Do đó, ở các nước tiên tiến có văn hóa “publish or perish” (có thể hiểu là “công bố hay là diệt vong”); có công bố khoa học là có cơ hội thăng tiến trong sự nghiệp, không có công bố khoa học thì xem như sự nghiệp khoa học chấm dứt. Để có công bố khoa học thì cần phải nghiên cứu khoa học. Đó cũng chính là lí do tại sao ở các đại học phương Tây (và cả nước ta nữa), nghiên cứu khoa học và công bố khoa học được xem là một tiêu chuẩn quan trọng trong việc để bạt và bổ nhiệm các chức vụ giáo sư.

5. Nghiên cứu khoa học là nhằm nâng cao phẩm chất giảng dạy. Đối với các giảng viên, làm mới bài giảng và giáo trình là một nhu cầu rất quan trọng, bởi vì nếu chỉ sử dụng một giáo trình suốt năm này sang năm kia thì sẽ dễ làm cho sinh viên nhàm chán. Thật ra, ngay cả người giảng cũng nhàm chán, vì họ cảm thấy mình như là một *thợ giảng*. Thợ khác với *thầy*, vì thầy là người ngoài giảng dạy còn sản sinh ra tri thức mới.

Do đó, nhu cầu đổi mới nội dung giảng, thêm ví dụ sinh động, thêm những ca nghiên cứu đặc thù, v.v. là rất cần thiết. Nghiên cứu khoa học giúp cho giảng viên có cơ hội cập nhật hóa kiến thức chuyên ngành. Nhưng ngoài việc cập nhật hóa

kiến thức và thông tin, giảng viên còn có thể sáng tạo ra tri thức mới, dữ liệu mới, tất cả những tri thức và dữ liệu đó có thể chia sẻ với sinh viên. Khi chia sẻ với sinh viên những dữ liệu của chính mình, giảng viên sẽ cảm thấy tự tin hơn (vì họ biết từ A đến Z sự xuất phát của các dữ liệu đó), chứ không chỉ đơn thuần chuyển tải dữ liệu của người khác. Nghiên cứu khoa học nâng vị thế của giảng viên là một người thầy hơn là một người thợ.

Nghiên cứu khoa học là vì uy tín của trường đại học. Bất cứ ai làm việc trong các trường đại học đều muốn đóng góp một phần của mình vào việc nâng cao “thương hiệu” và uy danh của trường, hay nói theo tiếng Anh là “visibility” (điểm nổi). Điểm nổi quan trọng của một trường đại học không chỉ là đào tạo nhân tài, mà còn sản sinh ra tri thức mới. Tri thức chỉ có thể đúc kết từ nghiên cứu khoa học. Qua nghiên cứu khoa học và công bố khoa học trên các tạp san quốc tế, nhà nghiên cứu góp phần nâng cao cái tên của trường đại học. Điểm nổi hay *visibility* của một đại học góp phần nâng cao hay giảm thấp thứ hạng của một đại học trên trường quốc tế.

Vì khoa học và cung cấp chứng cứ. Làm nghiên cứu khoa học thường có những phát hiện mới. Những phát hiện đó thường được công bố trên các tạp san quốc tế. Sự công bố kết quả hay phương pháp thể hiện sự đóng góp của nhà khoa học cho tri thức khoa học thế giới. Đó cũng là một cách thể hiện trách nhiệm của một thành viên trong cộng đồng khoa học quốc tế.

Bất cứ ai làm việc trong y khoa đều nghe biết đến trào lưu Y học Thực chứng (evidence based medicine). Y học Thực chứng được xen là một trường phái (hay cũng có thể xem là một ‘học thuyết’), mà theo đó các chuyên gia y tế phải thực hành dựa vào chứng cứ khoa học. Lí do của sự ra đời của trường phái Y học Thực chứng là trong quá khứ, rất nhiều liệu pháp điều trị (kể cả phẫu thuật) được sử dụng trên bệnh nhân nhưng chẳng có

chứng cứ khoa học nào cả. Một ngành nghề mang tiếng 'khoa học' mà lại thực hành không dựa vào khoa học! Do đó, trường phái Y học Thực chứng ra đời nhằm giảm thiểu những tai nạn đáng lẽ không xảy ra trong điều trị lâm sàng. Nhưng thực hành dựa vào chứng cứ khoa học thì cần phải có nghiên cứu khoa học. Điều này cũng có nghĩa là kết quả nghiên cứu khoa học được sử dụng như là những chứng cứ cho y học thực hành.

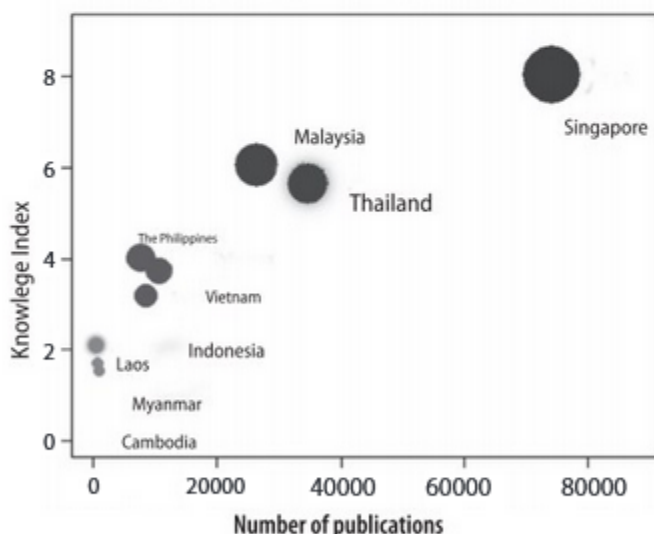
Nhưng không chỉ trong y khoa, mà các ngành nghề khác, kể cả giáo dục và chính sách công, cũng cần phải dựa vào chứng cứ. Trước khi triển khai một chương trình giáo dục mới, nhà chức trách cần phải có chứng cứ chứng minh rằng chương trình đó có hiệu quả. Trước khi ra một chính sách có ảnh hưởng đến hàng triệu người, nhà chức trách cũng cần phải thuyết phục công chúng bằng chứng cứ. Nhưng để có chứng cứ có chất lượng tốt thì cần phải thực hiện nghiên cứu khoa học.

Vì sự thịnh vượng của nước nhà. Có lẽ một trong những lý do để làm nghiên cứu khoa học là muốn đóng góp cho đất nước, và tôi muốn bàn thêm về cái ý quan trọng này. Trong thế kỉ 21, kinh tế tri thức được xem là một xu hướng chung, một mục tiêu mà các quốc gia muốn vươn tới. Kinh tế tri thức (knowledge economy) có thể hiểu là một nền kinh tế lấy tri thức khoa học và công nghệ làm động lực phát triển xã hội và tăng trưởng kinh tế. Một người công nhân trong một hãng sản xuất những con chip ở Đài Loan có thể có thu nhập cao gấp 10 lần một người nông dân ở nước ta, bởi vì hàm lượng tri thức trong sản phẩm của họ cao hơn ta. Do đó, các quốc gia trên thế giới cạnh tranh quyết liệt để xây dựng nền kinh tế tri thức, và qua đó có một vị trí "smart nation" (nước thông minh) trên thế giới.

Nói một cách ví von, khoa học là động cơ của sự phát triển và thịnh vượng của một quốc gia. Giới kinh tế học đã từng đánh giá rằng khoảng một phần ba đến phân nửa tỉ lệ tăng

trường kinh tế của Mỹ (tính từ Thế Chiến Thứ II) là nhờ vào nghiên cứu khoa học. Hầu như tất cả những sản phẩm công nghệ hiện đại mà chúng ta thụ hưởng ngày nay - từ xe ô-tô, được phẩm, đến điện thoại “thông minh” và thực phẩm - đều được sản xuất dựa trên thành tựu của nghiên cứu khoa học.

Chính vì thế mà khoa học và cách tân công nghệ đã giúp nhiều quốc gia và vùng lãnh thổ thoát khỏi tình trạng lạc hậu sang tình trạng tiên tiến. Quá trình phát triển của Hàn Quốc và Đài Loan là hai ví dụ điển hình về vai trò quan trọng của khoa học và công nghệ. Ngày nay, Hàn Quốc là một cường quốc khoa học và công nghệ ở Á châu, có thể nói là gần ngang hàng với Nhật Bản. Thành tựu ngoạn mục của Hàn Quốc do nhiều yếu tố, nhưng quan trọng nhất vẫn là đầu tư cho nghiên cứu khoa học và công nghệ.



Hình 1. Mối liên quan giữa chỉ số kinh tế tri thức (trục tung) và lượng ấn phẩm khoa học (trục hoành) trên các tạp san khoa học quốc tế của các quốc gia trong khối ASEAN.

Nguồn: Nguyen TV and Pham LT Scientometrics 2011.

Do đó, chúng ta có thể kì vọng một mối liên quan giữa lượng thông tin khoa học và mức độ tri thức hoá của nền kinh tế. Thật vậy, trong một nghiên cứu trước đây, chúng tôi phát hiện một mối tương quan khá chặt chẽ giữa số ấn phẩm khoa học và chỉ số kinh tế tri thức (1).

Để đạt được nền kinh tế tri thức, một quốc gia cần phải “sản xuất” ra tri thức mới, nâng cao cải tiến công nghệ, trình độ học vấn, và công nghệ viễn thông. Tuy nhiên, trong thực tế hiện nay, các chỉ số về kinh tế tri thức, cải tiến công nghệ và giáo dục của nước ta đều thấp hơn so với các nước trong khu vực như Thái Lan, Malaysia, Indonesia, Philippines, và Singapore. Sự thật này có nghĩa là muốn nâng cao nền kinh tế tri thức của Việt Nam, mỗi chúng ta - những người làm trong khoa học - phải có trách nhiệm nâng cao nghiên cứu khoa học và công bố khoa học. Do đó, nghiên cứu khoa học không chỉ là đem lại lợi ích cho cá nhân, cho trường đại học, mà còn là một đóng góp vào việc nâng cao nền kinh tế tri thức của nước nhà.

Tham khảo:

- (1) Nguyen TV, Pham LT. Scientific output and its relationship to knowledge economy: An analysis of ASEAN countries. *Scientometrics* 2011;89:101-117.
- (2) Báo Vietnamnet ngày 4/11/2017: <http://vietnamnet.vn/vn/suc-khoe/cac-loai-benh/94-000-nguoi-chet-do-ung-thu-moi-nam-gap-9-lan-tngt-408872.html>.

THẾ NÀO LÀ NGHIÊN CỨU KHOA HỌC?

Nghiên cứu khoa học (scientific research) là gì? Tuy câu hỏi rất đơn giản, nhưng trả lời thì không đơn giản chút nào. Không đơn giản là vì mỗi người, ngay cả trong giới khoa học, đều hiểu nghiên cứu khoa học khác nhau. Trong chương này, tôi sẽ giải thích **nghiên cứu khoa học** theo quan điểm của một người làm nghiên cứu thực nghiệm.

Nghiên cứu khoa học

Nghiên cứu có thể định nghĩa là một qui trình tìm hiểu một vấn đề để đi đến quyết định. Trong cuộc sống hàng ngày, chúng ta gặp rất nhiều vấn đề và sự kiện qua giao tiếp với các thành viên trong xã hội, qua đọc báo hoặc xem tivi, qua đọc các bài báo khoa học trước đây. Những vấn đề có thể làm cho chúng ta ngạc nhiên, có thể gây tác động tích cực hay tiêu cực đến đời sống chúng ta hay những người chung quanh. Chúng ta ngạc nhiên vì chưa từng thấy sự kiện đó trước đây. Chúng ta cảm thấy vấn đề có tác động, và do đó muốn tìm hiểu thêm nhằm có một giải pháp. Chẳng hạn như một bản tin viết rằng thức ăn có nhiều chất béo làm tăng nguy cơ bệnh tim mạch, nếu đó là sự thật thì đó là một mối liên hệ quan trọng, và chúng ta

cần tìm hiểu kỹ hơn. Tìm hiểu có thể là nguồn gốc của bản tin, cách thức mà nhà nghiên cứu đã thực hiện công trình nghiên cứu, dữ liệu đã được công bố ở đâu, ý kiến của các chuyên gia khác như thế nào, và thông tin nghiên cứu có liên quan gì đến chúng ta. Trả lời những câu hỏi trên đòi hỏi phải có *nghiên cứu*.

Do đó, có người từng phát biểu rằng chúng ta làm nghiên cứu khoa học hàng ngày. Theo quan điểm này, đôi trai gái tìm hiểu nhau để đi đến hôn nhân, bà nội trợ đi chợ khảo sát giá rau, hay người tài xế xe so sánh giá xăng dầu, v.v. là những “nghiên cứu khoa học”. Một số người khác hiểu nghiên cứu khoa học bao gồm cả những đánh giá về một chính sách công, hay khảo sát hành vi nịnh trong một ngôn ngữ, hay thái độ của quan chức đối với công chúng. Nhưng đó không thể xem là *nghiên cứu khoa học*, mà chỉ có thể xem là *nghiên cứu* mà thôi.

Đến đây thì chúng ta cần phải phân biệt giữa “*nghiên cứu*” và “*nghiên cứu khoa học*”. Nghiên cứu khoa học là chủ đề của chương này. Theo tôi, ***nghiên cứu khoa học* là một hoạt động của con người nhằm sáng tạo ra tri thức mới hay mở rộng tri thức hiện hành qua ứng dụng phương pháp khoa học.**

Mục tiêu và tri thức

Hai điều kiện của định nghĩa trên là mục tiêu và phương pháp. Mục tiêu là nhằm sáng tạo ra tri thức mới hoặc mở rộng ‘biên cương’ của tri thức hiện hành. Tri thức mới dĩ nhiên phải được chuyển hóa từ thông tin và dữ liệu. Khám phá ra một loài động vật mới, phát hiện một mối liên hệ mới giữa gen và bệnh tật, hay phát hiện một qui luật tự nhiên, tất cả những điều đó đều có thể xem là tri thức mới. Mở rộng kiến thức hiện hành bao gồm việc ứng dụng tri thức vào một lĩnh vực cụ thể. Chẳng hạn như chúng ta biết được cơ chế của các xương bào trong

xương là một tri thức tương đối mới, nhưng tìm cách ức chế các xương bào đó để phát triển thành thuốc điều trị cho bệnh nhân thì được xem là một sự mở rộng tri thức khoa học. Do đó, điều kiện thứ nhất của nghiên cứu khoa học là mục tiêu: tạo ra tri thức mới hay mở rộng tri thức hiện hành.

Tri thức ở đây phải đáp ứng 2 đặc tính chính: phổ quát và tái lập. Ở đây, tôi không có ý định bàn sâu về tri thức theo nghĩa triết học vì tôi không am hiểu triết học. Kiến thức phổ quát có thể hiểu là những tri thức mang tính qui luật tự nhiên. Chẳng hạn như hút thuốc lá gây ung thư phổi có thể xem là một tri thức phổ quát, vì tri thức này hiện diện ở bất cứ sắc dân nào và địa phương nào. Đặc tính thứ hai là tái lập (*reproducibility*). Một công trình nghiên cứu được xem là *khoa học* nếu kết quả từ nghiên cứu này có thể lặp lại trong nhiều nghiên cứu khác với điều kiện các nghiên cứu đó dùng cùng phương pháp với nghiên cứu gốc. Chẳng hạn như mối liên hệ giữa gen FTO và tiểu đường được quan sát ở rất nhiều nghiên cứu độc lập trên thế giới, và đây là một đặc tính quan trọng của khoa học. Nếu kết quả nghiên cứu không được tái lập thì đó chưa phải là tri thức khoa học. Do đó, tính tái lập có thể xem là nền tảng của khoa học, và là yếu tố để phân biệt khoa học thật với khoa học dỏm.

Phương pháp

Điều kiện thứ hai của nghiên cứu khoa học là *phương pháp khoa học*, còn gọi là *scientific method*. Phương pháp khoa học có thể định nghĩa là tập hợp một số kĩ thuật để điều nghiên các hiện tượng, thu thập dữ liệu mới, hay chỉnh sửa và tổng hợp các dữ liệu cũ. Hàm ý trong định nghĩa về phương pháp khoa học là đo lường hay định lượng hóa. Khác với những suy luận cá nhân hay phỏng đoán, phương pháp khoa học đòi hỏi

dữ liệu phải *chính xác* (accurate) và *tái lập* (precision) qua đo lường. Nếu tôi đo nồng độ HbA1c của một cá nhân nhiều lần và giá trị dao động trong khoảng 6,0 thì đó là phản ánh độ tái lập. Nếu tôi biết rằng nồng độ HbA1c chính xác của một cá nhân là 6,5 và phương pháp đo lường cho ra kết quả 6,5 thì đó là độ chính xác của phương pháp đó. Một phương pháp đo lường mang tính khoa học phải vừa chính xác và vừa mang tính tái lập.

Cần phải nói thêm là phương pháp khoa học không chỉ quan tâm đến đo lường, mà là một hệ thống phương pháp luận (methodology). Phương pháp luận từ cách đặt vấn đề, phát biểu giả thuyết, phương pháp chọn mẫu nghiên cứu, phương pháp đo lường, phương pháp phân tích dữ liệu đo lường, và logic trong diễn giải kết quả nghiên cứu.

Tóm lại, cần phải phân biệt giữa *nghiên cứu* và *nghiên cứu khoa học*. Nghiên cứu đơn thuần chỉ là thu thập dữ liệu hoặc thông tin mà không qua một phương pháp khoa học. Nghiên cứu khoa học đòi hỏi phải đáp ứng 2 điều kiện về mục tiêu liên quan đến tri thức khoa học và phương pháp khoa học. Do đó, những đánh giá về chính sách công, những khảo sát giá hàng hóa ở chợ, hay so sánh đơn thuần về thái độ ứng xử của quan chức với công chúng không thể xem là nghiên cứu khoa học, bởi không đáp ứng hai điều kiện về mục tiêu và phương pháp khoa học.

Mục đích của nghiên cứu khoa học có thể tóm lược và phân chia thành 3 nhóm: *mô tả* (description), *liên quan* (association), và *tiên lượng* (prediction). Nghiên cứu có mục đích mô tả bao gồm những nghiên cứu hệ thống hóa động vật, thực vật, hay sự kiện tự nhiên theo một tiêu chuẩn nào

đó. Nghiên cứu tìm hiểu mối liên quan bao gồm những phân tích tương quan giữa nhiều yếu tố nguy cơ và các nhóm bệnh lí. Nghiên cứu tiên lượng bao gồm phát triển các mô hình tiên lượng trong kinh tế và sinh học. Bất luận nghiên cứu vì mục đích nào, tính khoa học phải được đảm bảo qua mục tiêu tri thức và phương pháp khoa học.

Phân biệt khoa học cơ bản và khoa học ứng dụng

Một trong những vấn đề thường gây ra nhiều tranh cãi là định nghĩa thế nào là *khoa học cơ bản* (basic science) và thế nào là *khoa học ứng dụng* (applied science). Một số người nghĩ rằng khoa học cơ bản quan trọng hơn và “sang” hơn khoa học ứng dụng, nhưng cũng có người nghĩ ngược lại. Đối với những người làm việc trong các bộ môn khoa học mang tính thử nghiệm thì sự phân biệt quả thật là không cần thiết, có người còn lí giải rằng một phân biệt như thế là phản tác dụng và... nguy hiểm. Thậm chí có người còn nghĩ rằng một công trình nghiên cứu được các giới chức quản lí khoa học xếp vào nhóm khoa học ứng dụng thì sẽ không có yêu cầu công bố, nhưng nếu công trình được xếp vào nhóm khoa học cơ bản thì nhà khoa học được yêu cầu phải công bố kết quả nghiên cứu! Theo tôi, phân biệt giữa khoa học cơ bản và khoa học ứng dụng ngày nay có lẽ không cần thiết. Tất cả các công trình nghiên cứu khoa học đều cần phải có công bố dưới nhiều hình thức khác nhau.

Ở nước ngoài cũng từng có những cuộc tranh luận về phân biệt khoa học cơ bản và khoa học ứng dụng. Người ta đặt câu hỏi là có nên tài trợ cho nghiên cứu khoa học cơ bản quá nhiều như hiện nay hay không. Câu chuyện được đưa ra để làm “chất liệu” cho tranh luận là một nhóm nghiên cứu được

tài trợ khá nhiều tiền để nghiên cứu về... bộ phận sinh dục của vịt. Đó là một nghiên cứu thuộc loại khoa học cơ bản. Người chống những nghiên cứu cơ bản thì nói đó là một minh hoạ cho sự vô dụng của khoa học cơ bản. Và, trong thời “gạo châu củi quế” thì phải hạn chế tài trợ cho những nghiên cứu như thế, mà nên tập trung vào nghiên cứu có ích hơn. Người bênh thì nói đó là một mô hình nghiên cứu rất hay về ảnh hưởng của tiến hoá, biết được tại sao dương vật và âm vật của vịt có hình dạng đặc thù như ngày nay là một câu trả lời có thể quan trọng cho con người! Chỉ có vịt mới có thể hi sinh để chúng ta nghiên cứu, chứ làm sao mổ xẻ trên người được, và do đó, mô hình nghiên cứu trên vịt không hề vô dụng như những người “dốt” khoa học nói sai! Cuộc tranh luận cho thấy thỉnh thoảng vấn đề phân biệt khoa học cơ bản và khoa học ứng dụng vẫn được đặt ra.

Để phân biệt giữa hai thực thể thì chúng ta cần phải có tiêu chí. Theo tôi thấy, có thể dùng 3 tiêu chí sau đây để tạm phân biệt thế nào là khoa học cơ bản và thế nào là khoa học ứng dụng: động cơ nghiên cứu, sản phẩm nghiên cứu, và *giá trị nội tại*. Dĩ nhiên, có thể nghĩ đến vài tiêu chí khác nữa, nhưng tôi nghĩ 3 tiêu chí này là khá đầy đủ để phân định.

Về tiêu chí thứ nhất, động cơ của khoa học cơ bản là mở rộng tri thức con người. Tất cả hoạt động trong khoa học cơ bản chỉ nhằm vào mục tiêu tri thức chứ không có gì khác hơn. Có những nhà khoa học dành cả đời chỉ để theo đuổi tìm hiểu cấu trúc của một phân tử, mà họ có khi không biết thông tin đó sẽ ứng dụng cho cái gì. Còn khoa học ứng dụng thì có động cơ chính là ứng dụng tri thức, biến đổi hoặc cải tiến phương pháp, hay giải quyết một vấn đề thực tế. Khoa học cơ bản phát hiện ra gen, thì khoa học ứng dụng có thể phân tích xem gen

đó có liên quan đến bệnh tật hay không, và nếu có thì có thể can thiệp để giảm nguy cơ mắc bệnh. Khoa học ứng dụng cũng có thể là những nghiên cứu thực tế hơn như cải tiến một phương pháp hiện hành cho một điều kiện mới.

Tiêu chí thứ hai để phân biệt khoa học cơ bản và khoa học ứng dụng là *sản phẩm nghiên cứu*. Sản phẩm của khoa học cơ bản là tri thức mới mang tính lí thuyết và dữ liệu mới. Các nhà khoa học cơ bản không biết công trình và kết quả của họ sẽ ứng dụng cho cái gì. Họ có thể kiểm định một lí thuyết mới, phát triển một phương pháp mới, hay có nhiều khi phát hiện một cái mới hoàn toàn ngẫu nhiên (chẳng hạn như phát hiện X quang, penicillin và gen LRP5). Khoa học ứng dụng thì sản xuất ra công nghệ và tri thức thực tế. Còn khoa học ứng dụng là những công trình nghiêng về ứng dụng những tri thức hiện hành để có những kết quả cho một mục đích cụ thể. Theo cách hiểu này thì kĩ thuật (engineering) là một khoa học ứng dụng. Nghiên cứu sản xuất ra thuốc mới, phát triển phương pháp phân tích hàng triệu SNP trong hệ gen, v.v. cũng được xem là khoa học ứng dụng.

Tiêu chí thứ ba là khái niệm “internal goods”, tôi tạm gọi là *giá trị nội tại*. Một cách ngắn gọn, giá trị nội tại là những giá trị bao hàm trong hành động thực tiễn. Đối với khoa học cơ bản, giá trị nội tại hiểu thế giới chung quanh. Đối với khoa học ứng dụng, giá trị nội tại là thay đổi thế giới (hoặc gìn giữ không cho thay đổi). Dự án *Human Genome Project* (dự án giải mã toàn bộ gen người) chẳng hạn là một công trình nhằm thu thập tri thức về con người (và một mô hình cho các sinh vật), và đó chính là giá trị nội tại.

Nhưng có những trường hợp mà lẫn ranh giữa khoa học cơ bản và khoa học ứng dụng không rạch ròi như tôi nói ở trên. Có thể nói rằng con đường từ khoa học cơ bản đến khoa học ứng dụng là một đường thẳng liên tục. Trong y khoa, nghiên cứu ở mức độ tế bào (như sinh học phân tử) được xem là nghiên cứu cơ bản, còn thử nghiệm lâm sàng được xem là nghiên cứu ứng dụng. Nhưng ngay cả những người làm trong lĩnh vực sinh học phân tử cũng phản đối rằng họ là giới nghiên cứu cơ bản; họ muốn nhìn mình là giới nghiên cứu ứng dụng, bởi vì kiến thức của họ có thể chuyển giao thành sản phẩm.

Nhưng còn nghiên cứu hệ gen thì sao? Nếu tôi phân tích trình tự hệ gen của một nhóm bệnh nhân thì công trình đó nên được xem là nghiên cứu cơ bản hay ứng dụng? Khó có thể phân biệt. Thật ra, không thể phân biệt, bởi vì công trình nghiên cứu có động cơ và sản phẩm vừa mang tính cơ bản lại vừa mang tính ứng dụng. Tôi có thể có thông tin của hàng triệu biến thể gen, nhưng chỉ có một số nhỏ là liên quan đến bệnh. Tuy nhiên, hàng triệu biến thể gen không liên quan đến bệnh không có nghĩa là thông tin vô dụng, vì chúng tôi *chưa* tìm thấy ứng dụng của chúng. Do đó, cái giá trị nội tại là mang tính cơ bản, nhưng sản phẩm thì lại mang tính ứng dụng vì tôi dùng các thông tin gen đó để chẩn đoán và điều trị bệnh. Trong trường hợp này, việc phân biệt nghiên cứu ứng dụng và cơ bản là không cần thiết. Trong thực tế rất nhiều công trình nghiên cứu khoa học ngày nay không cần đến sự phân biệt giữa cơ bản và ứng dụng.

Vai trò của “consumer” khoa học trong việc tài trợ cho nghiên cứu khoa học, nhất là ngành ung thư học, cũng quan trọng. Tôi phải để chữ *consumer* trong ngoặc kép, vì chữ đó không được hiểu theo nghĩa thông thường là “người tiêu thụ”,

mà là “người liên đới”. Trong trường hợp này, *consumer* có nghĩa là các tổ chức đại diện cho bệnh nhân ung thư. Tôi thấy đây là một mô hình rất có ích để Việt Nam tham khảo. Ở Việt Nam, có nhiều người phàn nàn rằng có nhiều dự án nghiên cứu... vô bổ. “Vô bổ” hiểu theo nghĩa không có ứng dụng, hay không nghiên cứu thì người ta cũng biết. Một đất nước còn nghèo mà tài trợ cho những dự án như thế (nếu đúng như thế) thì quả thật là phí tiền thuế của dân. Vấn đề ai là người đánh giá “vô bổ” hay “có ích”. Trong vấn đề này, tôi nghĩ sự hiện diện của “consumer” là một giải pháp có thể có ích.

Ý TƯỞNG NGHIÊN CỨU ĐẾN TỪ ĐÂU?

Nhiều bạn bày tỏ ý muốn làm nghiên cứu khoa học, nhưng không biết bắt đầu từ đâu. Trong chương này, tôi sẽ bàn về nguồn gốc của ý tưởng, làm sao đánh giá ý tưởng của mình là đáng theo đuổi, và nếu theo đuổi thì bước kế tiếp là làm gì. Đây là những câu hỏi rất căn bản, mà tôi đã có dịp đề cập đến rất nhiều lần trong các chương trình tập huấn về phương pháp nghiên cứu khoa học. Ở đây, tôi sẽ bàn cụ thể và chi tiết hơn với những ca nghiên cứu thú vị.

Nghiên cứu khoa học xuất phát từ ý tưởng. Ý tưởng xuất phát từ nhiều nguồn, nhưng chủ yếu là qua quan sát thực tế, đọc những bài nghiên cứu trước (hay trong y văn, đối với ngành y), tham dự hội nghị; chú ý đến kĩ thuật mới, ý tưởng ngành khác; và... một chút tưởng tượng. Quan sát thực tế là một nguồn ý tưởng rất quan trọng. Chẳng hạn như tôi chú ý thấy hình như số bệnh nhân tử vong ở bệnh viện hơi nhiều trong hai ngày cuối tuần, tôi nghĩ ngay đến lời giải thích rằng có thể hai ngày đó thiếu bác sĩ và y tá, và thế là câu hỏi nghiên cứu hình thành. Tham dự hội nghị và nghe các chuyên gia giảng cũng là dịp để có ý tưởng mới. Có khi có thể các bạn chẳng có ý tưởng gì mới, mà chỉ muốn lặp lại nghiên cứu của người khác ở bệnh viện mình. Loại nghiên cứu “*me too*” này chẳng có vấn đề gì, vì đó là cách các bạn học làm nghiên cứu.

Nhưng ý tưởng phải được “giải trình” trong một đề cương nghiên cứu. Trong đề cương, các bạn phải phác họa mình làm gì trong nghiên cứu. Nên nhớ rằng mục tiêu của nghiên cứu khoa học là tạo ra dữ liệu mới, từ dữ liệu mới qua phân tích chúng ta tạo ra thông tin mới, và qua lí giải thông tin, chúng ta tạo ra kiến thức mới. Quy trình là chuyển hóa từ *dữ liệu* (data) thành *thông tin* (information), và từ thông tin thành *tri thức* (knowledge). Do đó, “trái tim” của một nghiên cứu là dữ liệu. Điều này có nghĩa là trong đề cương nghiên cứu, bạn phải phác họa:

Ý tưởng của nghiên cứu và mục tiêu;

- Chọn mô hình thí nghiệm;
- Quan trọng: đo lường cái gì (đây là data);
- Phân tích dữ liệu ra sao;
- Kết quả dự kiến.

Dĩ nhiên, đề cương nghiên cứu phải được thông qua bởi hội đồng khoa học và ủy ban y đức. Chỉ khi nào được thông qua thì công trình nghiên cứu mới được bắt đầu. Nhưng giai đoạn thu thập dữ liệu (qua đo lường, thí nghiệm) là hết sức quan trọng, mà một mình nghiên cứu sinh thì không cách gì làm được. Cần phải có nhiều người tham gia và hỗ trợ. Do đó, với khoa học hiện đại, nghiên cứu sinh phải làm việc với một nhóm. Nếu không làm việc theo nhóm thì đó là một sai lầm mang tính hệ thống của chương trình đào tạo.

Tham gia nhóm

Ở Việt Nam có không ít người suy nghĩ đơn giản rằng khi họ có ý tưởng là cứ thế mà làm nghiên cứu. Đối với những bộ môn như toán học thì suy nghĩ đó có thể đúng (chỉ “có thể”

thời), vì chỉ cần một mình làm việc cũng có thể chứng minh một định đề nào đó. Nhưng đối với khoa học, nhất là khoa học thực nghiệm, thì làm khoa học kiểu “đơn thân độc mã” là việc... không thể.

Làm khoa học ngày nay là cả một qui trình. Qui trình từ lúc viết đề cương nghiên cứu (proposal) để trình bày ý tưởng và kế hoạch thí nghiệm, đến việc xin phép ủy ban y đức (ủy ban đạo đức), đến việc ngoại giao với đồng nghiệp khác để tìm nguồn hỗ trợ. Tất cả những bước này đều có sự tương tác giữa nghiên cứu sinh và người hướng dẫn hay các đồng nghiệp khác, chứ một mình thì không thể nào làm hết được.

Ý tưởng thì có thể nghiên cứu sinh có, nhưng ý tưởng đó có đáng để theo đuổi hay không thì là vấn đề khác. Để đánh giá ý tưởng nghiên cứu, cần phải có người trong ngành và rất am hiểu vấn đề, chứ không thể một sớm một chiều mà đánh giá được. Cái khó là một số thầy cô ở trong nước vì không am hiểu y văn trong lĩnh vực nghiên cứu, nên cũng khó có thể cho ra lời khuyên đúng đắn. Nhưng vẫn có những thầy cô am hiểu vấn đề và họ có thể trở thành người hướng dẫn tốt.

Một nghiên cứu sinh không thể nào có tư cách để đứng tên một dự án nghiên cứu dính dáng đến con người, nhất là nghiên cứu can thiệp randomized clinical trial (RCT). Nếu bệnh viện hay đại học nào mà cho phép nghiên cứu sinh đứng tên chủ trì đề tài thì có lẽ cần phải xem lại tư cách của bệnh viện hay đại học đó. Không có ủy ban y đức đúng nghĩa nào phê chuẩn cho một người mà lí lịch khoa học “quá mỏng” làm nghiên cứu khoa học, nhất là nghiên cứu trên người và động vật.

Những khó khăn và yếu tố “văn hóa” trên cho thấy nếu bạn muốn làm nghiên cứu khoa học thì phải tham gia một

nhóm nghiên cứu hay một labo. Khi tham gia nhóm, dưới sự hướng dẫn và dìu dắt của người đứng đầu thì nghiên cứu sinh mới học được nhiều. Một nhóm nghiên cứu hay labo tốt có hẳn một văn hóa khoa học riêng. Văn hóa đó bao gồm cả việc tham gia seminar, tương tác, họp lab hàng tuần. Qua những tương tác như thế thì nghiên cứu sinh mới được trưởng thành trong khoa học. Cũng cần nói thêm rằng có một số seminar không thuộc diện bắt buộc, nhưng có seminar thì bắt buộc. Nếu là nghiên cứu sinh mà không dự đủ giờ seminar thì có thể không được cho tốt nghiệp.

Như tôi từng nói trước đây, đào tạo ra tiến sĩ không có nghĩa là chỉ đào tạo ra một nhà khoa học chuyên nghiệp, mà còn vài yếu tố quan trọng khác. Cẩm cái bằng tiến sĩ cũng như có “licence” để hành nghề nghiên cứu khoa học, cũng y như cẩm cái bằng MD là có licence để điều trị bệnh nhân. Nhưng trước khi điều trị bệnh nhân, bác sĩ phải học nói, học cách suy nghĩ, học cách cư xử, thì người cẩm cái bằng PhD cũng vậy, tức là phải học nói, học viết, học suy nghĩ lớn (big thinking), học cách hành xử trong “bộ lạc khoa học”. Những cái học ngoài nghiên cứu đó chỉ có thể học trong lab.

Nhưng tình hình ở Việt Nam thì không được lí tưởng như thế. Nhiều đại học không có các labo nghiên cứu đúng nghĩa, còn bệnh viện thì quá bận rộn với điều trị lâm sàng nên càng không có labo nghiên cứu. Do đó, nghiên cứu sinh ở Việt Nam ít có cơ hội để học hỏi bài bản như các đồng môn ở nước ngoài.

Chúng tôi cố gắng tạo ra một nhóm nghiên cứu ở Đại học Tôn Đức Thắng, với những đóng góp từ Trường và sự tình nguyện của thành viên nghiên cứu. Tôi nghĩ nhiều khi nghiên cứu ở Việt Nam là loại “labor of love”, làm vì tình thương khoa

học, chứ chẳng vì danh lợi gì cả. Nhưng loại làm vì danh lợi thì họ rất lớn tiếng và góp phần làm suy thoái nền khoa học.

Bắt đầu nghiên cứu: Ý tưởng đến từ đâu?

Tôi đã nói rằng để làm nghiên cứu đẳng hoàng thì cần phải tham gia một lab. Nghiên cứu khoa học ngày nay không thể làm theo kiểu “đơn thân độc mã” như xưa được. Nhưng một trong những câu hỏi đối với người muốn làm nghiên cứu khoa học là ý tưởng nghiên cứu đến từ đâu. Đối với người làm việc trong lab thì ý tưởng thường đến từ sếp. Nhưng dù là làm việc trong lab hay ngoài lab, thì cá nhân nghiên cứu sinh cũng phải có ý tưởng. Quá trình hình thành ý tưởng không phải một sớm một chiều, mà nó trải qua một giai đoạn dài suy nghĩ và bàn luận với đồng nghiệp.

Để có ý tưởng tôi nghĩ việc đầu tiên là phải tham gia một hiệp hội khoa học quốc tế. Trong ngành y thì mỗi chuyên ngành đều có một hiệp hội chính. Nói là “quốc tế” nhưng tôi thì chỉ chọn hai nhóm hiệp hội, một là Mĩ, hai là Âu châu. Có nhiều khi chỉ cần tham gia hiệp hội Mĩ là đủ. Nội tiết thì có *Endocrine Society*, loãng xương thì có *American Society for Bone and Mineral Research*, xương khớp thì có *American College of Rheumatology*, di truyền thì có *American Society of Human Genetics*, Dịch tễ học thì có *American Epidemiological Society*, ngoại khoa thì nhiều nhưng tiêu biểu là *American College of Surgeons*, *American Surgical Association*, v.v. Nói theo ngôn ngữ dân dã thì đây là những “bộ lạc”. Mỗi bộ lạc có một văn hoá riêng và có khi ngôn ngữ riêng. Mình phải biết cái văn hoá và ngôn ngữ của bộ lạc.

Để bước chân vào bộ lạc, trước hết mình phải là thành

viên của bộ lạc. Để trở thành thành viên của bộ lạc, mình phải đóng niên liễm (mỗi năm khoảng 100 - 300 USD). Khi đã ở trong bộ lạc, mình có cơ hội biết các “trưởng lão”, các đồng môn, và có cơ hội đi dự hội nghị và qua đó mà hình thành một network. Thành viên của bộ lạc còn được nhận tập san của chuyên ngành, và đó là một nguồn thông tin rất quan trọng cho ý tưởng nghiên cứu (sẽ đề cập dưới đây).

Là nhà khoa học hay bác sĩ mà không nằm trong bộ lạc nào thì cái lí lịch khoa học rất tệ. Thử tưởng tượng bạn là người sắp quyết định nhận người vào lab, và phải xem xét rất nhiều lí lịch khoa học. Nhiều lí lịch rất tương đương nhau về bằng cấp, kinh nghiệm, nhưng cái yếu tố quan trọng mà bạn có khi phải hỏi “Người này có nằm trong bộ lạc của mình?”

Quay lại câu hỏi hỏi ý tưởng nghiên cứu đến từ đâu. Đây là một câu hỏi quan trọng, và trong các workshop về phương pháp nghiên cứu, tôi thường dành hẳn một bài nói chuyện 1 giờ đồng hồ về câu hỏi này. Một cách ngắn gọn, câu hỏi nghiên cứu đến từ 4 nguồn chính:

- Quan sát thực tế;
- Đọc y văn, tham dự hội nghị;
- Chú ý đến kĩ thuật mới, ý tưởng ngành khác; và
- Một chút tưởng tượng.

Quan sát thực tế

Có thể nói rằng từ thực tế hàng ngày, thực tế lâm sàng bất bình thường, chúng ta có thể hình thành những câu hỏi nghiên cứu rất hay. Bất bình thường ở đây có thể là những giá trị ngoại vi của một xét nghiệm, hay những hiện tượng bất thường. Một

trong những câu chuyện “favorite” tôi hay lấy làm minh họa cho ý này là câu chuyện khám phá gen LRP5 trong chuyên ngành loãng xương. Câu chuyện bắt đầu bằng một tai nạn giao thông ở Los Angeles, mà trong đó một người bị thương rất nặng, gãy nhiều xương, nhưng người kia thì chẳng hề hấn gì, không có xương nào bị gãy. Khi vào bệnh viện khám, bác sĩ rất ngạc nhiên về người “lành lặn” kia và nghi ngờ rằng có thể bị nội thương, nên ông cho đi chụp X-ray và nhiều xét nghiệm khác, kể cả MRI. Ông phát hiện từ phim X quang là người thanh niên này có xương rất dày! Để chắc ăn, ông cho người thanh niên đi đo mật độ xương bằng máy DXA. Kết quả DXA cho thấy người này có mật độ xương cao gấp 4 độ lệch chuẩn so với người cùng độ tuổi! Một chi tiết thú vị khác là anh ta không thể bơi lội, vì cứ xuống hồ bơi là... chìm. Xương nặng quá.

Ông bác sĩ “ngầu” được mùi khám phá, nên lần dò hỏi thêm về gia thế và các thành viên trong gia đình. Khi ông được phép của Hội đồng y đức của trường đại học, ông đo mật độ xương của tất cả thành viên trong đại gia đình. Có người đang làm việc ở tận Thổ Nhĩ Kỳ, ông bác sĩ cũng bay sang “bắt” cho được. Dĩ nhiên, ông lấy mẫu máu và tách DNA. Sau đó, ông hợp tác với một chuyên gia về di truyền học và làm một phân tích gọi là genomewide linkage analysis, tức là scan toàn bộ hệ gen (nói cho oai, nhưng thời đó chỉ khoảng 1000 markers mà thôi). Phân tích dữ liệu cho thấy một tín hiệu nằm ở nhiễm sắc thể 11. Phân tích thêm ở vị trí đó, ông phát hiện gen LRP5. Đột biến thêm chức năng (gain of function mutation) ở gen LRP5 làm cho xương dày lên. Thế là ông bác sĩ đã khám phá ra một gen quan trọng giải thích tại sao mật độ xương của người thanh niên này rất đặc.

Trong cùng một lúc, một nhóm nghiên cứu độc lập khác,

hoàn toàn không liên quan gì với nhóm của ông bác sĩ trên, nghiên cứu về hội chứng osteoporosis-pseudoglioma. Hội chứng osteoporosis-pseudoglioma tương đối hiếm trong cộng đồng, mà đặc điểm chính là xương mỏng, mật độ xương rất thấp, rối loạn thị giác và thường dẫn đến mù lòa. Nhóm nghiên cứu này cũng phát hiện gen LRP5, nhưng là đối nghịch. Tức là trẻ em mắc hội chứng osteoporosis-pseudoglioma có đột biến mất chức năng ở gen LRP5.

Vậy là kết quả nghiên cứu của hai nhóm bổ sung cho nhau một cách tuyệt vời. Từ câu chuyện trên, thế giới loãng xương có thêm một gen trong “gia đình”. Từ khám phá gen này, người ta khám phá ra con đường đi của gen có tên là Wnt. Hóa ra, Wnt “chịu trách nhiệm” cho sự hình thành của xương và cánh ở động vật. Do đó, các công ti dược bào chế thuốc để điều chế hệ thống Wnt và dùng cho điều trị loãng xương. Đó là một thành công rất ngoạn mục, một khám phá rất quan trọng xuất phát từ chỉ một ca tai nạn giao thông. Trong nhiều trường hợp khác, những khám phá tương tự cũng đã được đề cập trong y văn. Trong khoa học có câu “chúng ta khám phá điều bình thường từ sự việc bất bình thường”, và khám phá LRP5 là một ca tiêu biểu.

Đọc y văn, tham dự hội nghị

Một nguồn ý tưởng về nghiên cứu quan trọng khác là qua tham khảo (đọc) y văn. Ngày nay, y văn là một kho tàng kiến thức khổng lồ. Mỗi năm có hơn 1 triệu bài báo khoa học được công bố, và con số này vẫn tăng hàng năm. Do đó, một người bình thường không thể nào đọc hay theo dõi được tất cả thông tin y khoa. Chúng ta cần phải chọn lọc thông tin khả tín để đọc. Để có thông tin khả tín và liên quan, chúng ta cần phải biết mình

theo đuổi chuyên ngành gì (và từ khóa tiếng Anh là gì), chuyên ngành rộng và chuyên ngành hẹp. Một người có thể làm trong lĩnh vực nội tiết, nhưng chỉ quan tâm đến tiểu đường, lại có người quan tâm đến loãng xương. Mỗi chuyên ngành có những tập san riêng, tuy độc lập với nhau, nhưng cũng có những bài báo liên quan đến cả hai chuyên ngành.

Xác định chuyên ngành để chọn tập san mà đọc. Mỗi chuyên ngành chỉ có vài tập san, cao lắm là 10. Phần còn lại là những tập san làng nhàng, đa phần là không đáng quan tâm. Chọn những tập san có impact factor cao để theo dõi (Tuy impact factor bị chỉ trích nhiều, nhưng nó vẫn là thước đo tốt nhất để sàng lọc thông tin cho người mới vào nghề). Đó là những tập san trong danh mục ISI. Danh mục Scopus cũng được, nhưng trong đó có khá nhiều tập san làng nhàng, nên không cần tốn thì giờ. Có thể tham khảo trang scimagojr.com để biết tập san nào có uy tín tốt và tập san nào không có uy tín tốt.

Sau impact factor thì tác giả cũng là tiêu chuẩn sàng lọc rất tốt. Trong mỗi chuyên ngành, chỉ có vài mươi nhóm nghiên cứu uy tín rải rác trên thế giới. Khi thấy một bài báo, việc đầu tiên là xem tên tác giả đầu và tác giả cuối. Tác giả nổi tiếng thường chỉ công bố bài báo ở những tập san nổi tiếng, và họ thường được mời viết “review”. Đó là tín hiệu của tác giả có uy tín. Địa chỉ làm việc của tác giả cũng rất quan trọng. Thường các nhóm nghiên cứu nổi tiếng nằm trong các trường đại học loại “Top 200” trên thế giới, còn mấy trường làng nhàng, trường địa phương, hay trường mới lập, v.v. thì rất ít có lab nghiên cứu đẳng cấp quốc tế.

Hội nghị cũng là địa điểm rất tốt để hình thành ý tưởng

nghiên cứu. Trong các hội nghị cấp quốc tế (thường là ở Mỹ và Âu châu) của các hiệp hội chuyên khoa, người ta hay mời các chuyên gia nổi tiếng đến chia sẻ nghiên cứu. Mỗi bài giảng của họ là một bài review sống động, và cuối bài giảng thường có những vấn đề mà họ chưa giải quyết được. Người nghe chú ý đến những vấn đề đó để có thể có những ý tưởng tiên phong. Chẳng hạn như có lần tôi đi dự hội nghị và nghe người ta bàn về TBS (trabecular bone score) rất nhiều, nhưng một trong những vấn đề người ta đặt ra là TBS có chịu sự ảnh hưởng của di truyền. Thế là tôi về và “xúi” đồng nghiệp Việt Nam làm một nghiên cứu về vấn đề này, và kết quả là năm sau chúng tôi có một công trình đầu tiên trên thế giới về yếu tố di truyền trong TBS!

Chú ý đến kĩ thuật mới, ý tưởng ngành khác

Ý tưởng mới thường có tính cách liên ngành. Nếu mình chỉ vùi đầu vào chuyên ngành của mình thì có hai hệ quả xảy ra. Hệ quả thứ nhất là chúng ta sẽ bị bảo thủ, không chịu mở rộng tầm nhìn và tầm lòng đến bên ngoài. Hệ quả thứ hai là chúng ta sẽ cạn kiệt ý tưởng rất nhanh. Do đó, lúc nào cũng phải mở tầm nhìn sang các chuyên ngành khác, chú ý xem họ phát triển đến mức nào, họ có cái gì mới mà chúng ta có thể áp dụng cho chuyên ngành của mình.

Một trong những ca tiêu biểu là di truyền học. Nếu chúng ta nghiên cứu về bệnh Parkinson chẳng hạn và nếu chúng ta chỉ quanh đi quẩn lại làm những gì “kinh điển” thì chuyên ngành sẽ chẳng đi đến đâu. Nhưng nếu chúng ta biết rằng (và trong thực tế thì như thế) bệnh Parkinson là do một phần yếu tố di truyền gây nên, thì bắt buộc chúng ta phải để mắt đến công nghệ di truyền học họ phát triển đến đâu. Ngày xưa, có

lẽ chúng ta chỉ phân tích 1 vị trí gen một lần, nhưng với công nghệ mới, chúng ta có thể phân tích khoảng nửa triệu vị trí một lần! Mười năm trước, để giải mã toàn bộ hệ gen, chúng ta tốn 15 ngàn USD, nhưng ngày nay thì chỉ 1.500 USD. Với chi phí rẻ như thế, chúng ta có thể phân tích hàng trăm bệnh nhân, và sẽ có những khám phá thú vị.

Ý tưởng từ chuyên ngành khác dĩ nhiên không giới hạn trong công nghệ di truyền học, mà còn trong bất cứ chuyên ngành khác. Có thể kể đến các chuyên ngành hay công nghệ như vật lý y khoa, sinh hóa, tâm lý học, xã hội học, kỹ thuật, “công nghệ thống kê” (bao gồm các phương pháp phân tích tinh vi như *machine learning*), tất cả đều giúp chúng ta suy nghĩ về chuyên ngành mình. Điều này đòi hỏi chúng ta phải đọc nhiều. Ngoài các tập san y khoa, cần phải đọc những tạp chí có phẩm chất tốt như *American Scientist*, *Scientific American*, *Nature.com*, hay các trạm thông tin như *Science News*, *asianscientist.com*, *medscape.com*, v.v. Đây là những trạm thông tin đa ngành, nên những phát hiện và phát triển quan trọng đều được đề cập đến. Qua những bài báo để cập chúng ta có thể tìm bài báo gốc trên các tập san khoa học để tìm hiểu thêm chi tiết.

Một chút tưởng tượng

Thỉnh thoảng, ý tưởng đến với chúng ta rất tình cờ, qua tưởng tượng. Một câu của Albert Einstein mà tôi hay trích dẫn là “*If you have logic, you can go from A to B; if you have imagination, you can go anywhere*” (Nếu bạn có logic, bạn có thể đi từ A đến B; nếu bạn có trí tưởng tượng, bạn có thể đi bất cứ nơi nào). Tôi nghĩ câu này đúng với nhiều khám phá tình cờ. Đọc những cuốn sách về tiểu sử của các nhà khoa học lừng

danh, họ có xu hướng phát kiến ý tưởng từ những sự tưởng tượng có thể nói là... *fantasy*. Một nguồn “tưởng tượng có hệ thống” theo tôi là các tiểu thuyết và phim ảnh giả tưởng, vì những tác phẩm loại này giúp chúng ta bay bổng hơn. Và, bay bổng có khi là yếu tố rất cần thiết cho khoa học.

Một trong những ý tưởng nghiên cứu mang tính tưởng tượng là công trình nghiên cứu nổi tiếng (nhưng cũng khá tai tiếng) của Giáo sư Daryl Bem. Ông là một giáo sư tâm lý học lừng danh của Đại học Cornell (Mỹ). Ông tưởng tượng rằng con người chúng ta có một hệ tiềm thức giúp chúng ta đoán trước một tình huống tưởng như là hoàn toàn ngẫu nhiên, và nó giúp chúng ta có thể “cảm nhận được tương lai” (ông gọi là *feeling the future*)! Giả thuyết mà Giáo sư Bem muốn kiểm định là sự khoái lạc của người nhìn vào những hình ảnh ái ân trong tương lai sẽ giúp họ đi ngược về quá khứ và ảnh hưởng đến hiện tại. Cách ông thiết kế thí nghiệm như sau (tôi mô tả ngắn gọn lại cách hướng dẫn của ông để các bạn hiểu công trình độc đáo này):

“Đây là một thí nghiệm nhằm kiểm định giả thuyết ngoại cảm (ESP). Thí nghiệm cần khoảng 20 phút để hoàn tất và tất cả đều chạy trên máy tính. Trước hết, bạn sẽ trả lời một số câu hỏi ngắn. Sau đó, cứ mỗi lần thử, vài hình ảnh xuất hiện trên màn ảnh máy tính, mỗi hình có 2 bức màn song song nhau. Một bức màn có hình đằng sau, một bức màn không có hình. Việc của bạn là click vào bức màn mà bạn cảm thấy có hình đằng sau. Sau khi click bức màn sẽ mở ra để bạn biết mình đã chọn đúng (tức là bức màn có hình). Bạn sẽ làm 36 lần. Nhiều hình mang tính khiêu dâm (như những cặp trai gái trong tình huống ái ân). Nếu bạn không đồng ý với những hình ái ân như thế, bạn có quyền không tham gia thí nghiệm.”

Trên đây là 4 nguồn ý tưởng mà tôi hay bàn đến trong các workshop về nghiên cứu khoa học. Trong thực tế, mỗi labo có một hướng nghiên cứu riêng, và các trường labo đã có sẵn một chương trình nghiên cứu để đạt mục tiêu. Các chương trình nghiên cứu đó có khi bị “bê lái” vì những nguồn ý tưởng trên đây. Các nguồn ý tưởng trên đây đòi hỏi nhà khoa học phải có đầu óc quan sát tinh tế, phải chú ý đến chi tiết hoặc những điều gì có vẻ bất thường, phải đọc nhiều và đặt vấn đề trong bối cảnh chung, và phải có chút tính nghệ sĩ. Những cá tính đó, tôi nghĩ, cũng là nhân cách của một nhà khoa học thành công.

Nên nhớ rằng nghiên cứu khoa học là đi tìm qui luật tự nhiên. Tuyệt đối tránh những suy nghĩ kiểu xác định tỉ lệ này, ước tính tỉ suất kia, so sánh phương pháp nọ, v.v. vì những loại này là làm thống kê mô tả, chứ không phải nghiên cứu khoa học đúng nghĩa. Nghiên cứu khoa học là phải có giả thuyết (hypothesis), và giả thuyết chỉ được hình thành sau khi đã xem qua những bài báo trong chuyên ngành.

Thoát ra khỏi tư duy “tầm thường”

Nhiều nghiên cứu sinh ngành y ở các nước phát triển thường nghĩ đến câu hỏi nghiên cứu theo tư duy của những người làm về y tế công cộng. Những ý tưởng của y tế công cộng thường là tỉ lệ hiện hành, tỉ lệ phát sinh bệnh, và yếu tố nguy cơ. Có thời gian nhiều người theo đuổi những đề tài về kiến thức, hành vi, và thái độ. Điều này cũng dễ hiểu, bởi vì ở những nước phát triển, đó là những vấn đề quan trọng cần phải giải quyết. Nhưng để cho tư duy đó thống trị nghiên cứu y học thì sẽ có tác động không tốt cho y học nước nhà, bởi vì nghiên cứu y học không chỉ và không thể giới hạn trong tư duy và các phương pháp y tế công cộng. Những câu hỏi liên quan đến tỉ lệ

và yếu tố nguy cơ là quan trọng (có khi rất quan trọng) cho y tế, nhưng những câu hỏi đó không giúp cho y học tiến xa hơn, bởi vì nó chỉ là bề mặt của vấn đề.

Nghiên cứu y học cần phải đi sâu vào vấn đề. Những vấn đề cần đi sâu là cơ chế của bệnh, những mối liên hệ, can thiệp và hiệu quả của can thiệp. Cơ chế của bệnh có thể là yếu tố sinh học (ví dụ như *hormone*, các *markers* sinh hoá), là yếu tố gen, là yếu tố môi trường, là các yếu tố xã hội. Xin nói thêm rằng, khi nói “môi trường”, tôi không chỉ đơn giản đề cập đến lối sống (hút thuốc lá, bia rượu) và dinh dưỡng, mà là môi trường sống và nội môi trường (như microbiota hay hệ vi sinh vật trong đường ruột). Đó là những vấn đề đòi hỏi nhà khoa học phải tìm hiểu và nếu được triển khai thành nghiên cứu.

Dĩ nhiên, chúng ta đang làm khoa học theo mô hình “*reductionism*”, nên chúng ta chỉ có thể tập trung nghiên cứu vào một yếu tố hay một vấn đề cụ thể, chứ phương pháp hiện hành chưa cho phép chúng ta làm nghiên cứu tổng thể để có thể xem xét tất cả các yếu tố và cơ chế của bệnh. Do đó, còn có một lĩnh vực nghiên cứu khác liên quan đến phương pháp. Ngày nay, những chuyên ngành như thống kê, khoa học máy tính, vật lý cũng có những định hướng nghiên cứu phương pháp giúp cho nhà khoa học thực nghiệm giải quyết vấn đề tốt hơn và hiệu quả hơn. Trong thực tế, những ngưỡng trị số *P* (sẽ bàn trong một chương sau), những nghiên cứu về *false discovery rate*, phương pháp Bayes, phương pháp thể hiện dữ liệu sinh học, phương pháp phân tích đa biến, v.v. tất cả đều xuất phát từ tình huống thực tế y sinh học, và các nhà khoa học trong các ngành vừa kể đã có đóng góp quan trọng cho y sinh học. Hiện nay, vẫn còn rất nhiều vấn đề cần phải giải quyết, và y sinh học vẫn là một “lãnh địa” màu mỡ cho các nhà thống kê học, vật lý học, khoa học máy tính.

Nhưng có không ít người nghĩ rằng khoa học ở nước phát triển chưa phát triển đúng mức để có thể ứng dụng những thành tựu nghiên cứu của họ. Họ cũng nghĩ rằng đáng lí ra các nhà khoa học thực nghiệm phải tìm đến họ để họ có thể giúp. Nhưng suy nghĩ này rất ngược với suy nghĩ ở các nước như Mỹ và Úc. Ở Úc, các nhà nghiên cứu cơ bản, kể cả toán học, thống kê học, sinh hóa thường tìm đến giới nghiên cứu khoa học để tìm sự hợp tác và học hỏi. Họ phải học để hiểu về chuyên ngành khoa học và sau đó mới có thể hợp tác. Không ai dám nghĩ rằng nghiên cứu của mình quá cao siêu đến nỗi người ngoài ngành không sử dụng được.

Quay lại vấn đề mở rộng câu hỏi nghiên cứu. Tôi nghĩ có thể bắt đầu bằng những câu hỏi đơn giản, và sau đó mở rộng tầm nghiên cứu chuyên sâu hơn. Có thể lấy một nghiên cứu bệnh Parkinson làm một ví dụ minh họa. Thoạt đầu, có thể câu hỏi nghiên cứu chỉ giới hạn trong các vấn đề tiêu biểu như tỉ lệ và yếu tố nguy cơ như tuổi, lối sống, và nước. Nhưng chúng ta có thể mở rộng những yếu tố nguy cơ sang gen. Trong thực tế, người ta đã làm về gen khá lâu, những công nghệ mới cho phép chúng ta nghiên cứu về mối tương tác giữa gen và môi trường qua phân tích *epigenetics*. Thật ra, trong lĩnh vực Parkinson, y văn nói về *epigenetics* nhiều hơn là làm. Đọc những nghiên cứu trước đây có thể thấy nhiều bài tổng quan (*review*) hơn là nghiên cứu gốc (*original research*). Sự thật này là cơ hội tốt để Việt Nam làm nghiên cứu về *epigenetics* và qua đó có thể có đóng góp rất nguyên thủy cho y văn thế giới.

Nhưng câu hỏi kế tiếp là sau khi đã có *epigenetics* thì sẽ làm gì nữa? Đến đây thì vấn đề nghiên cứu ứng dụng (*translational research*) thành một câu hỏi khác. Câu hỏi là "*chúng ta có thể dùng các markers của mối tương tác giữa gen và môi trường để*

tạo ra một chữ kí gen để có thể tiên đoán bệnh Parkinson sớm hơn?”. Thế là nghiên cứu có thêm một câu hỏi thứ ba. Thật ra, còn phải suy nghĩ tiếp là sau khi xong epigenetics thì sẽ làm gì tiếp, và câu trả lời có lẽ là làm nghiên cứu cơ bản trên các dòng tế bào hay chuột. Kết cục lại là công trình nghiên cứu xuất phát từ một ý tưởng đơn giản, rất tiêu biểu y tế công cộng, trở thành một nghiên cứu tinh vi và chuyên sâu hơn.

Trong thực tế thì không chỉ dừng ở đó, mà phải nghĩ đến can thiệp. Câu hỏi là nếu chúng ta nhận ra bệnh nhân có nguy cơ cao sớm hơn bằng “chữ kí epigenetic” và nếu chúng ta can thiệp môi trường thì có làm giảm nguy cơ cho họ. Để trả lời câu hỏi này, chúng ta phải nghĩ đến can thiệp, chọn bệnh nhân, thiết kế nghiên cứu, v.v. và như thế là nghiên cứu được triển khai tiếp. Ngay cả sau khi triển khai, chúng ta phải nghĩ đến cách đánh giá “effectiveness” của chương trình can thiệp. Hay nếu nghiên cứu là phát triển mô hình tiên lượng, thì bước kế tiếp phải nghĩ đến là làm sao triển khai mô hình đó để nhiều người sử dụng và giúp cho bệnh nhân.

Không nên nghĩ rằng “Ở Việt Nam cũng đã làm rồi”. Đó là cách phát biểu “trùm mền” và tự tin thái quá. Mỗi một lĩnh vực nghiên cứu, người ta đi từ cơ bản đến lâm sàng, chứ không phải chỉ làm ngang. Ví dụ như nghiên cứu về microbiota, người ta bắt đầu từ giả thuyết, triển khai sang nghiên cứu cơ bản, sau đó là phát triển công nghệ, nghiên cứu trên chuột, và sau cùng là nghiên cứu lâm sàng. Nghiên cứu tế bào gốc cũng qua qui trình đó. Nhiều nhóm nhà khoa học khắp nơi cùng hợp tác làm, mỗi người một lĩnh vực, chứ không phải chỉ một vài người. Nếu chỉ làm ngang một khía cạnh nào đó có thể quá đơn giản, và suy nghĩ như thế không giúp gì cho khoa học về lâu dài cả.

Như vậy, hành trình suy nghĩ về câu hỏi nghiên cứu có thể xuất phát từ loại câu hỏi rất thô, rất đơn giản, nhưng để có câu hỏi chuyên sâu hơn, nhà nghiên cứu phải chịu khó đọc y văn, phải theo dõi sự phát triển của các chuyên ngành khác, và phải bàn luận với những người có kinh nghiệm trong chuyên ngành.

HÀNH TRÌNH KHÁM PHÁ GEN LRP5

Rất nhiều khám phá trong khoa học bắt đầu với một trường hợp bất bình thường. Từ một hiện tượng bất bình thường, các nhà khoa học có thể khám phá yếu tố... bình thường. Câu chuyện và quá trình khám phá gen LRP5 trong loãng xương bắt đầu từ một cá nhân với xương rất đặc, và khám phá này đang làm thay đổi định hướng nghiên cứu loãng xương. Khám phá còn cung cấp cho chúng ta nhiều bài học về quy trình nghiên cứu khoa học.

Từ một tai nạn giao thông

Năm 1994, một lực sĩ tên tạm gọi là John bị tai nạn xe hơi khá nghiêm trọng, nhưng anh ta không hề bị thương, xương không bị tổn hại. Qua xét nghiệm X quang, bác sĩ không thấy anh bị gãy đứt sống nào. Tuy nhiên, X quang cho thấy xương của anh khá đặc. Bác sĩ quang tuyến giới thiệu John đến bác sĩ Karl Insogna chuyên khoa về xương và cũng là giám đốc trung tâm nghiên cứu xương thuộc Đại học Yale. Sau khi đo mật độ xương (bone mineral density - BMD), bác sĩ Insogna ngạc nhiên thấy mật độ xương trong người John cao gấp 8 lần so với những người cùng độ tuổi! Nhưng sự thật này chẳng làm ngạc nhiên John, vì anh biết mình... nặng người, nặng đến nỗi khi anh ta không bơi lội được, vì mỗi khi xuống hồ tắm là người anh như cục đá chìm xuống sông. Bác sĩ Insogna không phát hiện bệnh hay bất cứ triệu chứng bất thường nào khác, nên quyết định không tiếp tục theo dõi.

Sáu năm sau, trong một lần tán gẫu cùng đồng nghiệp, bác sĩ Insogna được biết vị đồng nghiệp này từng biết đến một gia đình với mật độ xương rất cao. Hai người bèn lần dò theo sơ đồ gia hệ, và phát hiện chính John là thành viên của một đại gia đình có quê ở miền đông duyên hải nước Mỹ. Trong dòng họ này, ai cũng đều có mật độ xương rất cao, và cảm thường vuông một cách khác thường, nhưng ngoài đặc điểm đó, bộ xương của họ không có vấn đề gì. Một trong những thành viên của đại gia đình là một bác sĩ ở bang Alabama từng bị gãy xương đùi, nhưng các bác sĩ chấn thương chỉnh hình phải vất vả lắm mới lắp được xương giả vào vì họ gặp khó khăn trong việc vận vít vào xương.

Đến khám phá gen LRP5

Bác sĩ Insogna và đồng nghiệp quyết định tiến hành nghiên cứu trên các thành viên của đại gia đình này, xác định xem gen nào có ảnh hưởng đến mật độ xương cao. Sau một thời gian làm việc cực nhọc, nhóm nghiên cứu của bác sĩ Insogna phát hiện gen có dính dáng đến mật độ xương cao nằm trong nhiễm sắc thể 11. Nhưng nhiễm sắc thể này quá rộng và dài, và tìm gen trong một khoảng DNA dài như thế chẳng khác gì mò kim đáy biển. Nhóm nghiên cứu đành tạm thời bỏ cuộc vì không có phương tiện và kĩ thuật chuyên môn tiếp tục truy tìm gen.

Trong khi đó, một nhóm nghiên cứu của Matthew Warman thuộc Đại học Case Western Reserve cũng nghiên cứu gen trong nhiễm sắc thể 11. Nhóm của Warman nghiên cứu trên 28 gia đình, mà trong đó có người mắc hội chứng *osteoporosis-pseudoglioma syndrome* (OPPG) mang tính di truyền. Đặc điểm của hội chứng này là bệnh nhân có mật độ xương rất thấp, và họ dần dần trở nên mù. Bác sĩ Warman quyết định truy tìm cho được gen gây nên bệnh OPPG, bằng cách hợp tác với nhóm nghiên cứu thuộc Đại học Harvard của Giáo sư Bjorn Olsen. Sau một thời gian khá lâu phân tích, họ phát hiện một gen hoàn toàn nằm ngoài dự kiến: đó là gen LDL receptor-related protein 5 (viết tắt là LRP5) trong nhiễm sắc thể 11.

Gen LRP5 sản xuất một protein mà chức năng sinh học vẫn chưa được biết. Đột biến của gen LRP5 ở bệnh nhân OPPG mà hóa thụ thể của các tế bào tạo xương (osteoblasts). Những bệnh nhân OPPG với hai biến thể (allele) có mật độ xương rất thấp. Cha hay mẹ của bệnh nhân OPPG có một biến thể của gen cũng hay dễ mắc bệnh loãng xương. Điều này cho thấy protein mà gen này mã hóa có thể là một đối tượng quan trọng cho việc điều trị loãng xương.

Nghे tin này, bác sĩ Insogna nghĩ ngay rằng rất có thể cũng chính gen này đóng vai trò làm cho mật độ xương tăng cao. Sau khi phân tích gen LRP5, nhóm của bác sĩ Insogna phát hiện đúng là một đột biến khác của gen làm cho người ta có mật độ xương cao.

Trong cùng gần một lúc, một nhóm nghiên cứu của Mark Johnson thuộc Trung tâm nghiên cứu loãng xương tại Đại học Creighton University (bang Omaha), cũng qua nghiên cứu một gia đình gồm 21 người với mật độ xương cao khác thường, đã phát hiện chính đột biến của gen LRP5 (mà nhóm của bác sĩ Insogna vừa phát hiện) có liên quan đến mật độ xương cao. Trong gia đình này, tuổi từ 3 đến 93, không một ai từng bị gãy xương.

Chỉ vài tháng sau, LRP5 trở thành một tin tức khoa học thời sự. Nhóm của Warman công bố kết quả trên tập san danh tiếng Cell vào năm 2001 (4). Hai tháng sau, nhóm của Johnson công bố nghiên cứu của họ trên tập san *American Journal of Human Genetics* (5), và nhóm của Insogna cũng công bố kết quả trên tập san y học *New England Journal of Medicine* vào tháng 5/2002 (6).

Gen LRP5 nằm trong hệ thống tín hiệu có tên là Wnt (hay Wnt signaling pathway). Hệ thống tín hiệu Wnt là tập hợp nhiều phân tử được xem là đóng vai trò quan trọng trong việc kiểm soát tăng trưởng và phát triển của cơ thể. Bây giờ, các nhà khoa học còn biết thêm rằng hệ thống này cũng đóng vai trò quan trọng điều chế sự phát triển của xương. Vì lý do này, hệ thống Wnt là một đối tượng lý tưởng cho nghiên cứu tìm một thuật can thiệp chống loãng xương và các rối loạn xương khác. Đối với giới nghiên cứu xương, phát hiện này có ý nghĩa rất quan trọng, có thể làm thay đổi định hướng điều trị loãng xương. Từ trước đến nay, hầu hết các thuốc điều trị loãng xương có mục tiêu ức chế hoạt động của các tế bào hủy xương, nhưng khám phá mới này mở ra một hướng phát triển mới: đó là kích hoạt các tế bào tạo xương.

Kể từ khi khám phá gen LRP5, thế giới nghiên cứu loãng xương bắt đầu tập trung vào nghiên cứu chức năng và cơ chế của gen LRP5 ở chuột và con người. Hàng ngàn nghiên cứu và hàng ngàn bài báo khoa học khai thác đề tài này trong suốt 5 năm qua. Nhưng ngày nay, nghiên cứu gen không còn làm trong phòng thí nghiệm nữa, mà là trên máy tính. Một khi khoanh được một vùng có chứa gen cần tìm, các nhà khoa học chỉ việc truy tìm trong cơ sở dữ liệu trên mạng internet để có danh sách tất cả các gen trong vùng. Cơ sở dữ liệu gen này do Dự án phân tích bộ gen con người (Human Genome Project hay HGP) xây dựng và bất cứ ai cũng có thể sử dụng. Thành quả của dự án HGP có thể rút ngắn thời gian truy tìm gen từ 10 đến 20 năm cho các nhà khoa học chuyên tìm gen cho các bệnh phức tạp.

Và những bài học

Nhìn qua khám phá gen LRP5, chúng ta có thể thấy về mặt ý tưởng, một công trình nghiên cứu khoa học có khi bắt đầu bằng một quan sát "bất bình thường". Đánh rằng, biết được bao nhiêu người có mật độ xương rất cao hay rất thấp cũng là một thú vị, và biết được gen nào liên quan đến mật độ xương lại càng thú vị hơn, nhưng ý nghĩa của nó là gì? Bài học qua công trình nghiên cứu phải nói thú vị này là: một khám phá có thể bắt đầu từ những so sánh về khác biệt. Phải có người có mật độ xương cao bất thường (hay thấp bất thường) thì mới có đối tượng so sánh. Nói cách khác, nghiên cứu di truyền, cũng như các nghiên cứu lâm sàng dịch tễ học khác, cần một nhóm đối chứng (controls); không có đối chứng thì nghiên cứu sẽ không có ý nghĩa khoa học.

Một bài học khác là một khám phá mang tính khẳng định cần phải được lặp lại trong nhiều đối tượng khác nhau, chứ một phát hiện từ một mẫu vẫn chưa đủ tính thuyết phục. Trong trường hợp trên, ba nhóm nhà khoa học độc lập, và cả ba đều phát hiện gen LRP5 ngay tại một nhiễm sắc thể. Chính vì tính nhất quán này mà công trình nghiên cứu của họ có thể xem là một bài học quan trọng cho các nỗ lực truy tìm gen.

Về mặt phương pháp, công trình nghiên cứu của các nhà khoa học trên là một bài học về qui trình tìm gen cho một bệnh. Truy tìm gen liên quan đến bệnh tật là một nỗ lực lớn, là đề tài nghiên cứu khoa học "nóng" tại các nước đã phát triển. Tìm được gen cũng có nghĩa là có được cái chìa khóa để mở ra cánh cửa cho điều trị và thậm chí giúp cho chúng ta phát hiện bệnh sớm hơn. Nhưng tìm gen cũng không khác gì câu nói ví von ông bà ta thường nói: mò kim đáy biển. Trong cái biển 25000 hay 20000 gen và hơn 2 triệu marker, làm sao tìm cho được marker hay gen liên quan? Qui trình của các nhà khoa học này là họ làm từng bước. Bước đầu là tìm ra nhiễm sắc thể có triển vọng cao; bước kế tiếp là "khoanh vùng" những marker liên quan; và bước thứ ba là từng đoạn một siết chặt những vùng quan hệ và cuối cùng là "tóm" lấy cái marker liên quan nhất. Nỗ lực khoanh vùng và siết chặt này đòi hỏi chuyên môn cao của hai bộ môn khoa học chủ chốt: công nghệ sinh học phân tử và thống kê học. Để tìm ra những marker cần thiết chúng ta cần đến chuyên gia về công nghệ sinh học và di truyền học. Để xác định marker nào có liên quan đến bệnh, thống kê học đóng vai trò số một. Có thể nói không ngoa rằng, không có công nghệ sinh học và thống kê học, chúng ta sẽ không thể nào tìm gen.

Tuy nhiên, cũng như bất cứ vấn đề khoa học nào, một khám phá đều đặt ra nhiều câu hỏi, nhiều khả năng, hơn là giải quyết dứt điểm vấn đề. Cho đến nay, chúng ta vẫn chưa biết cơ chế của hệ thống tín hiệu Wnt hoạt động ra sao, và gen LRP5 đóng vai trò gì trong hệ thống này. Các nhà nghiên cứu loãng xương hy vọng sẽ có câu trả lời trong vòng 5 - 10 năm tới. Vẫn còn nhiều việc phải làm!

Tài liệu tham khảo:

1. Anonymous. Consensus Development Conference: diagnosis, prophylaxis, and treatment of osteoporosis. *Am J Med.* 1993; 94:646-650.
2. Nguyen ND, et al. Identification of high-risk individuals for hip fracture: a 14-year prospective study. *J Bone Miner Res* 2005; 20(11):1921-8.
3. Nguyen TV, et al. Bone mass, lean mass and fat mass: same gens or same environments. *Am J Epidemiol* 1998; 147:3-16.
4. Gong Y, et al. LDL receptor-related protein 5 (LRP5) affects bone accrual and eye development. *Cell* 2001; 107:513-23.
5. Little RD, et al. A mutation in the LDL receptor-related protein 5 gene results in the autosomal dominant high-bone-mass trait. *Am J Hum Genet* 2001;70:11-9.
6. Boyden LM, et al. High bone density due to a mutation in LDL-receptor-related protein 5. *N Engl J Med* 2002;346:1513-21.

MỘT VAI NGHIÊN CỨU THỦ VỊ

Dưới đây là vài thí nghiệm khoa học xã hội (KHXH) và tâm lí học mà tôi thấy rất thú vị. Tôi chọn ra vài thí nghiệm hay để chia sẻ cùng các bạn, và để nói rằng KHXH rất hay và rất phong phú.

Thí nghiệm về phân chia giai cấp

Thí nghiệm này được thực hiện trong một lớp học ở Iowa vào năm 1968. Nhà tâm lí học Jane Elliott lấy cảm hứng từ tư tưởng của Ts Martin Luther King, và bà muốn chứng minh rằng sự phân biệt giai cấp là do môi trường xã hội. Elliott chia học sinh trong lớp học của bà thành hai nhóm: nhóm mắt xanh và nhóm mắt nâu. Vào ngày đầu tiên, bà báo cho lớp học biết rằng nhóm học trò mắt xanh là giỏi giang hơn, và từ đó họ được hưởng nhiều đặc quyền đặc lợi, làm cho nhóm học trò mắt nâu trở thành thiểu số. Bà không cho hai nhóm tương tác với nhau, và cố tình tìm những học sinh trong nhóm thiểu số để trưng bày những cái xấu xa của họ. Ngay sau đó, hành vi và thái độ của học sinh thay đổi. Nhóm mắt xanh học hành tốt hơn và bắt đầu hành xử lưu manh với nhóm mắt nâu. Nhóm mắt nâu thì trở nên mất tự tin và học hành càng ngày càng tệ. Sau đó, bà đảo ngược thí nghiệm bằng cách cho nhóm mắt xanh trở thành thiểu số. Khi cuộc thí nghiệm kết thúc, học sinh cảm thấy khuây khỏa, và trở nên thân thiện với nhau hơn. Họ kết luận rằng không nên đánh giá con người dựa vào bề ngoài.

Thí nghiệm “Bobo Doll”

Thí nghiệm này do tiến sĩ Alburt Bandura thực hiện ở Đại học Stanford trong thời gian 1961-1963. Vào thời đó, có một cuộc tranh luận về sự ảnh hưởng của di truyền và môi trường đến sự phát triển của trẻ em. Tiến sĩ Albert Bandura thực hiện thí nghiệm dùng búp bê Bobo để chứng minh rằng hành vi của con người là do bắt chước (mô phỏng) chứ không phải do di truyền.

Trong thí nghiệm này, ông chia tình nguyện viên thành 3 nhóm: nhóm 1 được cho xem video mà trong đó người lớn có hành vi hung hăng với búp bê Bobo; nhóm 2 được cho xem video mà trong đó người lớn chơi với búp bê Bobo một cách thân thiện; và nhóm 3 là nhóm chứng. Sau khi xem video, trẻ em được cho vào một phòng với búp bê Bobo mà họ thấy trên video clip. Kết quả cho thấy trẻ em nhóm 1 trở nên hung hãn với búp bê, trong khi nhóm 2 và nhóm 3 thì hiền lành hơn. Nhà khoa học ước tính chỉ số hung hãn ở trẻ em nam trong nhóm 1 là 38,2 so với 12,7 ở nữ. Thí nghiệm này cho thấy khi trẻ em được “phơi nhiễm” hành vi hung hăng thì họ sẽ bắt chước làm theo, và nam có xu hướng hung hăng hơn nữ.

Thí nghiệm “Asch Conformity”

Năm 1951, Tiến sĩ Solomon Asch thuộc Trường Cao đẳng Swarthmore thực hiện một thí nghiệm nhằm chứng minh rằng con người có xu hướng tuân thủ theo chuẩn mực dù chẳng có áp lực nào để họ làm thế. Đây là một thí nghiệm được xem là đột phá trong tâm lý học. Một nhóm tình nguyện viên được cho xem những hình ảnh gồm nhiều đường kẻ thẳng nhưng chiều dài khác nhau. Họ được yêu cầu trả lời một câu hỏi đơn giản là đường kẻ nào dài nhất? Điều thú vị trong thí nghiệm này là trong nhóm tình nguyện viên đó, chỉ có 1 người duy nhất là tình nguyện viên, tất cả những người khác là diễn viên và họ đã được hướng dẫn trước về câu trả lời. Câu trả lời hướng dẫn có thể đúng nhưng đa số là sai. Kết quả thú vị thay là người tình nguyện viên thật sự lúc nào cũng đồng ý với đa số (mặc dù đa số là sai)! Kết quả thí nghiệm nổi tiếng này cho thấy khi chúng ta nghiên cứu về những mối tương tác xã hội, con người nói chung có xu hướng thích nghi hay làm theo số đông. Nói cách khác, người ta có xu hướng chú ý làm theo người khác hơn là hành xử đúng.

Hiệu ứng Hawthorne

Nghiên cứu này được thực hiện vào năm 1955 tại công ti Hawthorne Works thuộc thành phố Chicago. Người thực hiện là Henry Landsberger. Mục tiêu của thí nghiệm là chứng minh rằng con người trong môi trường thử nghiệm sẽ thay đổi hành vi chỉ vì họ đang được nghiên cứu (là đối tượng của nghiên cứu). Đạo đó, công ti Hawthorne Works muốn biết cường độ sáng của đèn điện có ảnh hưởng đến năng suất của công nhân. Elton Mayo làm thí nghiệm, nhưng ông không phát hiện mối liên hệ giữa đèn điện và năng suất lao động. Nhưng sau đó, Landsberger phân tích dữ liệu thì phát hiện rằng công nhân có xu hướng tự điều chỉnh năng suất chỉ vì họ biết họ đang là đối tượng của thí nghiệm, và điều này giải thích tại sao Mayo quan sát rằng không có mối liên hệ. Kết luận từ phân tích này là công nhân cảm thấy họ quan trọng (vì được chọn làm thí nghiệm) nên họ tự tăng năng suất lao động. Hiện tượng này được gọi là Hiệu ứng Hawthorne (hay Hawthorne Effect) và trở thành một dạng bias trong nghiên cứu khoa học.

Thí nghiệm nhạc công violin

Thí nghiệm này được Nhật báo Washington Post thực hiện tại một trạm xe điện ở Washington vào năm 2007. Mục tiêu của thí nghiệm là để tìm hiểu người ta hành xử như thế nào đối với những gì xảy ra quanh họ. Trong thí nghiệm, người qua đường tấp nập đến trạm xe điện nhưng không nhận ra có một người nhạc sĩ lừng danh Joshua Bell mới dành giải Grammy-winning [chỉ trước đó 2 ngày] đang chơi đàn violin. Bình thường trong nhà hát lớn nơi

Joshua Bell chơi nhạc, người ta phải trả \$100 mỗi vé để được thưởng thức. Nhưng tại trạm xe điện này ông đóng vai một nhạc sĩ đường phố chơi những bài nhạc tình vì với cái đàn violin trị giá 3,5 triệu USD! Trong 45 phút đầu tiên ông chơi đàn, chỉ có 6 người dừng lại và xem biểu diễn một chút. Khoảng 20 người cho tiền, nhưng không dừng lại xem trình diễn. Sau buổi trình diễn ông được người qua đường cho \$32. Đây là một thí nghiệm xã hội thú vị nhằm xem xét nhận thức, thị hiếu và ưu tiên của công chúng. Kí giả Gene Weingarten viết về thí nghiệm này cho tờ Washington Post và sau này được trao giải Pulitzer. Vài câu hỏi đặt ra là chúng ta cảm nhận về cái đẹp như thế nào? Chúng ta có dừng lại và ghi nhận cái đẹp? Chúng ta có nhận ra tài năng trong tình huống bất ngờ? Hóa ra, phần đông chúng ta không hẳn có cảm thụ được môi trường chung quanh.

Thí nghiệm kẹo Marshmallow

Thí nghiệm này do Tiến sĩ Walter Mischel thực hiện tại Đại học Stanford vào năm 1972. Mục tiêu của thí nghiệm là nhằm tìm hiểu mối liên quan giữa sự trì hoãn tưởng thưởng lúc còn nhỏ và khả năng thành công khi trưởng thành. Trẻ em tuổi 4 - 6 được dẫn đến một phòng, trong đó có một cái bàn trước mặt họ, trên bàn có một viên kẹo marshmallow. Trước khi rời phòng, người thí nghiệm cho các em biết rằng nếu sau 15 phút mà viên kẹo vẫn còn trên bàn, họ sẽ được cho 1 viên kẹo khác. Nghiên cứu viên sau đó quan sát các em "kháng" phần thưởng được bao lâu, và phân tích thời gian này với sự thành công khi trưởng thành. Trong số 600 em tham gia thí nghiệm, một nhóm nhỏ dùng ngay viên kẹo, và 1/3 số trẻ hoãn để được tưởng thưởng thêm 1 viên kẹo. Sau nhiều năm theo dõi, Mischel phát hiện rằng những em trong nhóm "trì hoãn" trở nên giỏi giang trong công việc và có điểm SAT cao hơn nhóm "ăn liền". Thí nghiệm có vẻ đơn giản, nhưng phát hiện này cho thấy những khác biệt giữa các cá nhân có thể đã được định hình ngay từ còn nhỏ.

Nguồn:

www.onlinepsychologydegree.info/influential-psychological-experiments/

CÁCH TIẾP CẬN CÂU HỎI NGHIÊN CỨU

Giả dụ như bạn đã có một ý tưởng để làm nghiên cứu, bước kế tiếp là phải suy nghĩ làm sao để thực hiện ý tưởng. Nhưng để cho việc suy nghĩ cách thực hiện tốt, bạn cần phải rất cụ thể về chi tiết, và phải bắt đầu với câu hỏi nghiên cứu và mục tiêu nghiên cứu. Cách tiếp cận vấn đề sẽ được xây dựng trên cơ sở của mục tiêu nghiên cứu. Chương này sẽ bàn về những mô hình nghiên cứu phổ biến để tiếp cận một câu hỏi nghiên cứu.

Nếu là nghiên cứu trên người, cách tiếp cận một câu hỏi nghiên cứu phải dựa trên nhiều yếu tố, nhưng tựu trung lại là 4 yếu tố chính như sau:

- Mô hình nghiên cứu là gì?
- Đối tượng nghiên cứu là ai?
- Đo lường những gì, tức là dữ liệu cần thu thập là gì?
- Phân tích dữ liệu ra sao?

Mô hình nghiên cứu

Trong nghiên cứu lâm sàng, hay nghiên cứu y học nói chung, việc chọn mô hình nghiên cứu là rất quan trọng, vì nó quyết định giá trị khoa học của nghiên cứu. Cùng là một câu hỏi nghiên

cứu, các bạn có nhiều mô hình để tiếp cận. Chẳng hạn như nếu các bạn muốn biết ăn chay có lợi hay hại đến sức khỏe, cụ thể là bệnh loãng xương, các bạn có thể chọn nghiên cứu so sánh hai nhóm người ăn chay và ăn mặn. Nhưng trong thực tế thì có nhiều mô hình so sánh. Nhìn chung, nghiên cứu y học / lâm sàng trên người có thể chia thành 5 mô hình, giá trị cao nhất đến thấp nhất:

- Phân tích tổng hợp các nghiên cứu RCT (randomized controlled trials);
- Nghiên cứu RCT: chia bệnh nhân một cách ngẫu nhiên thành hai nhóm, nhóm can thiệp và nhóm chứng;
- Nghiên cứu đoàn hệ (prospective cohort study): đây là nghiên cứu quan sát (không can thiệp), theo dõi một nhóm người theo thời gian;
- Nghiên cứu bệnh chứng (case - control study): đây cũng là nghiên cứu quan sát, nhà nghiên cứu chọn một nhóm bệnh và một nhóm chứng, rồi so sánh các yếu tố lâm sàng giữa hai nhóm;

Nghiên cứu thiết diện (hay cắt ngang, hay cross-sectional study): cũng là nghiên cứu không can thiệp, nói đúng ra nó là một dạng survey hay điều tra xã hội.

Trong thực tế thì còn nhiều mô hình nghiên cứu khác. Chẳng hạn như *case report*, *case series* (một số ca lâm sàng thú vị), nghiên cứu tương quan môi sinh (ecological correlation), và mô hình giao chéo (cross-over). Tuy nhiên, những mô hình trên đây giúp cho các bạn suy nghĩ về nghiên cứu khoa học và cách tiếp cận câu hỏi nghiên cứu.

Tùy vào mục tiêu, ngân sách, và thời gian, nhà nghiên cứu phải có cách chọn mô hình thích hợp. Nếu mục tiêu là đánh

giá hiệu quả của một thuật can thiệp (có thể là thuốc, thuật giải phẫu, chương trình y tế) thì có thể xem mô hình RCT hay biến thái của mô hình RCT. Nếu mục tiêu là đánh giá mối liên quan giữa một yếu tố nguy cơ đến một bệnh lí, thì nghiên cứu đoàn hệ có thể là lựa chọn tối ưu. Nếu mục tiêu là tìm hiểu mối liên quan giữa yếu tố nguy cơ và bệnh, nhất là bệnh hiếm, thì mô hình bệnh chứng có lẽ là một lựa chọn tốt. Còn nếu chỉ muốn tìm hiểu số người mắc bệnh hay yếu tố nguy cơ thì nghiên cứu thiết diện là một mô hình cần xem xét.

Mỗi mô hình nghiên cứu có những điểm mạnh và điểm yếu. Chẳng hạn như mô hình RCT thì rất đắt tiền, vì phải theo dõi bệnh nhân một thời gian, và phải can thiệp, nên cần hẳn một ban phụ trách thực hiện nghiên cứu. Đối với những nghiên cứu qui mô trên vài ngàn bệnh nhân, người ta phải chọn nhà nghiên cứu cẩn thận (với lí lịch khoa học “xem được” và thành tích khoa học xứng đáng để được làm nghiên cứu RCT). Nghiên cứu đoàn hệ tuy không can thiệp, nhưng cũng khó thực hiện, vì phải theo dõi bệnh nhân theo thời gian, có khi 5 hay 10 năm, mới có dữ liệu để phân tích. Nghiên cứu bệnh chứng tuy rất dễ làm nhưng giá trị khoa học thì không cao như các mô hình khác. Còn nghiên cứu thiết diện thì chỉ là một dạng điều tra xã hội, tuy dễ thực hiện, nhưng nếu thu thập dữ liệu không tốt thì cũng chẳng có giá trị gì. Do đó, cần phải cân nhắc cẩn thận việc chọn mô hình nghiên cứu để giải quyết câu hỏi nghiên cứu.

Tất cả các mô hình nghiên cứu có giá trị phải có cái gọi là nhóm chứng (control). Một nghiên cứu không có nhóm chứng thì rất khó diễn giải. Để biết hiệu quả của một phẫu thuật mới, mà chỉ có một nhóm bệnh nhân được mổ bằng phương pháp mới là chưa đủ, mà cần phải so sánh với một nhóm bệnh nhân

được mổ bằng phương pháp hiện hành. Dĩ nhiên, có khi rất khó tìm nhóm chứng, nhưng nhà nghiên cứu cần phải “sáng tạo” nghĩ ra cách làm sao cho có nhóm chứng.

Quay lại ví dụ về ăn chay, sau khi xem xét qua y văn, chúng tôi chú ý thấy các nghiên cứu trước đây là nghiên cứu quan sát hay nghiên cứu bệnh chứng. Do đó, để có cái mới trong nghiên cứu, chúng tôi quyết định chọn mô hình RCT. Theo mô hình này, chúng tôi sẽ chọn một số người (không hẳn là bệnh nhân, mà là người bình thường), sau đó tôi sẽ dùng phương pháp thống kê để chia thành hai nhóm ngẫu nhiên. Nhóm 1 sẽ được ăn chay suốt 3 năm liền, và nhóm hai là ăn mặn cũng 3 năm. Trong thời gian 3 năm chúng tôi sẽ theo dõi xem có bao nhiêu người bị gãy xương trong mỗi nhóm, và so sánh nguy cơ gãy xương giữa hai nhóm để đánh giá hiệu quả của ăn chay.

Trong một chương sau, tôi sẽ bàn cụ thể về cách chọn mô hình nghiên cứu, và những vấn đề liên quan đến thiết kế nghiên cứu.

Đối tượng nghiên cứu

Sau khi đã quyết định một mô hình nghiên cứu, bước kế tiếp là chọn đối tượng nghiên cứu. Nếu là nghiên cứu trên người, cần phải suy nghĩ trước hết là chọn người ở giới tính nào, nam hay nữ, hay cả nam và nữ, và người ở độ tuổi nào. Ngoài ra, cần phải suy nghĩ đến các tiêu chuẩn loại trừ, tức là những đối tượng không nên tuyển vào nghiên cứu. Ví dụ như trong nghiên cứu ăn chay, có thể chúng tôi chỉ chọn những phụ nữ trên 60 tuổi (vì nam thì khó ăn chay và họ ăn nhậu nhiều quá), những người không có tiền sử gãy xương, những người không mắc những bệnh có liên quan đến sự chuyển hóa của xương. Nhưng tôi sẽ loại bỏ những người đang dùng

thuốc có tác động đến sức khỏe của xương, hay những người bị đột quỵ, v.v. Đây là những tiêu chuẩn cần phải suy nghĩ cẩn thận, và đòi hỏi nhà nghiên cứu phải am hiểu vấn đề, am hiểu về bệnh lý loãng xương, và suy nghĩ về tương lai.

Chọn đối tượng nghiên cứu còn có ý nghĩa hợp lý ngoại tại (external validity). Một công trình nghiên cứu có giá trị khoa học phải đáp ứng hai tiêu chuẩn: hợp lý nội tại (internal validity) và hợp lý ngoại tại. Hợp lý nội tại (sẽ bàn dưới đây) liên quan đến đo lường, còn hợp lý ngoại tại là liên quan đến cách chúng ta chọn đối tượng nghiên cứu. Kết quả nghiên cứu, theo lý thuyết, chỉ được khái quát hóa cho những người đáp ứng tiêu chuẩn chọn vào và tiêu chuẩn loại trừ.

Một trong những nghiên cứu tâm đắc mà tôi hay nói với các bạn học viên là nghiên cứu để đánh giá sự ảnh hưởng của môi trường không trọng lực đến xương. Người trong chuyên ngành ai cũng biết xương là một mô động (chứ không phải tĩnh như người ta nghĩ). Khi chúng ta luyện tập thể dục thì xương cũng sẽ đáp ứng theo. Đó chính là lý do tại sao người ta khuyến cáo nên tập thể dục để giảm nguy cơ gãy xương. Nhưng làm sao để “chứng minh” được ý tưởng này? Các nhà nghiên cứu Mỹ nghĩ ra một cách làm rất thú vị. Họ chọn đối tượng là những phi hành gia. Phi hành gia khi bay lên vũ trụ, họ phải sống trong môi trường không có trọng lực, và nếu giả thuyết kia đúng thì khi họ đáp xuống mặt đất, mật độ xương phải giảm. Do đó, các nhà nghiên cứu đo mật độ xương của những phi hành gia trước khi bay và sau khi đáp xuống mặt đất. Kết quả cho thấy chỉ trong vòng 1 tháng hay ngắn hơn, mật độ xương của họ suy giảm đến 3-5%, có người giảm đến 10%. Ngoài ra, họ còn bị suy giảm về lượng cơ và trở nên yếu đuối. (Đó chính là lý do tại sao khi đáp xuống

mặt đất họ phải được khiêng, chứ không cho bước đi, vì rất dễ bị gãy xương)

Một nghiên cứu khác cũng thú vị không kém, là họ chọn các nữ sinh chơi quần vợt, họ đo mật độ xương ở tay phải và tay trái. Giả thuyết là nếu người thuận tay phải thì mật độ xương tay phải sẽ cao hơn tay trái, và quả thật như thế. Đó là một cách thiết kế nghiên cứu vừa đơn giản nhưng lại vừa thuyết phục (vì so sánh cùng một cá nhân). Nói chung, có rất nhiều cách tiếp cận câu hỏi nghiên cứu qua cách chọn đối tượng. Nhưng tất cả đều cần phải suy nghĩ và bàn bạc với người có kinh nghiệm.

Đo lường cái gì?

Dữ liệu là trái tim của một nghiên cứu. Điều này có nghĩa là đo lường đóng vai trò hết sức quan trọng trong một nghiên cứu. Khi nói đo lường, chúng ta nói đến hai yếu tố sau đây: số liệu nào cần thu thập, và phương pháp đo lường là gì.

Số liệu nào cần thu thập? Câu trả lời là cần phải chia các biến số (dữ liệu) thành 3 nhóm chính. Nhóm 1 là các số liệu liên quan đến *outcome* hay biến kết cục. Nhóm 2 là các số liệu liên quan đến yếu tố can thiệp hay *factors* - yếu tố. Nhóm 3 là các số liệu liên quan đến các yếu tố nhiễu hay *confounding factors* hay *covariates*. Một cách khác để biết mình cần đo lường những gì là đọc các bài báo khoa học trước đây, xem các tác giả họ đã đo lường những biến số nào và dùng phương pháp nào. Chúng ta cần phải biết những nghiên cứu trong quá khứ để kết quả của chúng ta có thể dùng làm so sánh với họ. Dĩ nhiên, ngoài những gì người ta đo lường, chúng ta cũng phải có cái mới của mình, chứ không phải chỉ thuần túy là “me too”.

Biến outcome là biến quyết định thành bại hay hiệu quả mối liên quan của một nghiên cứu. Ví dụ như nghiên cứu về loãng xương, biến outcome có thể là mật độ xương hoặc gãy xương. Vấn đề là chờ cho người ta bị gãy xương (xin lỗi, hơi vô duyên) thì rất lâu và cũng không có bao nhiêu người bị, nên có thể chúng ta phải dùng mật độ xương làm outcome. Nhưng để đảm bảo, chúng ta phải đo cả các marker chu chuyển xương. Lí do là nếu chúng ta quan sát rằng ăn chay có ảnh hưởng đến mật độ xương, có thể có người phản bác rằng đó là kết quả ngẫu nhiên, và để thuyết phục kết quả không phải do ngẫu nhiên, chúng ta phải có số liệu về bone marker để trình bày và lí giải. Điều này đòi hỏi nhà nghiên cứu phải tìm tư vấn từ một người có kinh nghiệm về chuyên ngành loãng xương để biết mình cần đo lường cái gì.

Biến *risk factors* hay yếu tố nguy cơ mà nhà nghiên cứu cần đo tùy thuộc vào mục tiêu nghiên cứu. Chẳng hạn như nếu tôi quan tâm đến mối liên quan giữa thuốc statin và loãng xương, thì các thông tin về statin là yếu tố nguy cơ. Những thông tin cần thu thập có thể là loại thuốc statin cụ thể, thời gian sử dụng cụ thể, liều lượng, chỉ định, v.v. Nếu tôi quan tâm đến béo phì và tử vong, thì yếu tố nguy cơ ở đây là béo phì. Tôi phải suy nghĩ dùng số liệu gì để phản ánh béo phì. Người “nhà nghèo” thì có thể dùng cân nặng và chiều cao để tính body mass index (BMI), nhưng người có “insight” và “serious” thì dùng tỉ trọng mỡ (percent body fat) để phản ánh cái gọi là “béo phì”. Dĩ nhiên, tỉ trọng mỡ đo bằng máy DXA là tiêu chuẩn vàng, còn BMI thì giá trị khoa học không cao, chủ yếu cho những người làm về y tế công cộng. Tóm lại, những biến *risk factors* cần phải suy nghĩ thấu đáo và chú ý đến giá trị khoa học của chúng.

Biến covariates hay tạm hiểu là biến confounding factors là những biến có liên quan đến outcome và risk factors. Đây là những yếu tố có thể làm cho nhà nghiên cứu... nhức đầu. Chẳng hạn như nếu tôi quan sát mối liên quan giữa bệnh thoái hóa khớp và loãng xương (ví dụ như người bị thoái hóa khớp có nguy cơ loãng xương cao hơn người không bị thoái hóa khớp), thì câu hỏi đặt ra là mối liên quan này có thể không thật, mà nó bị “gây rối” bởi một yếu tố khác. Yếu tố khác đó có thể đơn giản là độ tuổi. Người cao tuổi (trên 60 chẳng hạn) có nguy cơ bị thoái hóa khớp tăng, nhưng họ cũng có nguy cơ bị loãng xương cao; thành ra, mối liên quan giữa thoái hóa khớp và loãng xương mà chúng ta quan sát là không thật, mà là do độ tuổi. Do đó, chúng ta cần phải thu thập thông tin về độ tuổi, và câu hỏi đặt ra là: ở mỗi độ tuổi, người bị thoái hóa khớp có nguy cơ bị loãng xương cao hơn người không bị thoái hóa khớp hay không? Nếu câu trả lời là có, thì chúng ta mới có bằng chứng để nói rằng mối liên quan giữa hai bệnh là thật. Điều này có nghĩa là trước khi làm nghiên cứu, nhà nghiên cứu phải suy nghĩ tất cả các yếu tố nhiễu và phải đo lường tất cả.

Ngoài ra, còn phải suy nghĩ lâu dài, chứ không phải kiểu làm ngắn hạn. Có nhiều bạn chỉ muốn làm nghiên cứu để có một luận văn làng nhàng, làm cho có, làm cho qua, chứ không suy nghĩ xa. Nhưng các bạn cần phải luyện tập suy nghĩ xa. Phải suy nghĩ sau 10 năm hay 20 năm, các bạn vẫn có thể quay lại số liệu nghiên cứu để giải quyết một vấn đề nào đó. Điều này có nghĩa là trong nghiên cứu lâm sàng, các bạn cần phải lưu trữ sinh phẩm (như mẫu máu, mẫu nước tiểu, mô, tế bào, v.v.) hết sức kỹ lưỡng. Chẳng hạn như nếu 20 năm sau có người phát hiện gen có liên quan đến độ dày của động mạch, thì các bạn lúc đó chỉ cần lấy mẫu máu ra, tách DNA và phân tích gen

đó và kiểm định giả thuyết trong nghiên cứu của mình. Như thế là các bạn vẫn có khả năng đóng góp vào y văn dù nghiên cứu đã kết thúc 20 năm trước. Nên nhớ là các bạn phải suy nghĩ xa, chứ đừng suy nghĩ thiên cận.

Phẩm chất của số liệu và đo lường phản ánh giá trị hợp lí nội tại. Như tôi từng nói, cái yếu tố quan trọng nhất của nghiên cứu khoa học là hợp lí nội tại, chứ không phải hợp lí ngoại tại. Nếu phương pháp đo lường “dỏm”, phẩm chất số liệu thấp, và tính hợp lí nội tại rất thấp. Nếu hợp lí nội tại thấp, thì cho dù nghiên cứu có tính đại diện cũng chẳng có giá trị khoa học gì, mà chỉ là một đồng số liệu vô hồn.

Phân tích ra sao?

Sau khi nghĩ đến dữ liệu cần thu thập hay yếu tố cần đo lường, các bạn phải suy nghĩ làm gì với dữ liệu. Dĩ nhiên, các bạn nghĩ đến phân tích dữ liệu. Rất nhiều người không nghĩ đến khâu phân tích dữ liệu, nên khi thu thập xong dữ liệu, họ cảm thấy lúng túng, không biết phải làm cái gì trước, cái gì sau. Do đó, bất cứ đề cương nghiên cứu nào cũng phải có phần phân tích dữ liệu. Trong phần này, các bạn phải định nghĩa biến outcome là gì, biến risk factors là gì, biến confounders là gì, và xác định mô hình thống kê để phân tích.

Những vấn đề trong phân tích dữ liệu thì rất nhiều, nhà khoa học khó có thể lường trước tất cả tình huống. Tuy nhiên, những vấn đề thường nảy sinh trong thực tế khoa học là hoán chuyển dữ liệu (vì dữ liệu không tuân theo luật phân bố chuẩn), xử lí dữ liệu trống (missing values), phương pháp chọn biến số liên quan, và kiểm định kết quả phân tích (validation). Đối với các tập san lớn và uy tín, như *New England Journal of Medicine* chẳng hạn, họ yêu cầu tác giả phải nộp kế hoạch phân tích

dữ liệu và kế hoạch đó phải được phê chuẩn bởi một giáo sư, trước khi họ chịu xem qua bài báo hay không. Do đó, các bạn phải suy nghĩ và viết ra kế hoạch phân tích dữ liệu để thuyết phục người khác là mình biết mình làm cái gì.

Tôi phải nói thêm là phân tích dữ liệu càng ngày càng quan trọng, và nó là yếu tố quyết định sự thành bại của một bài báo khoa học. Ngày nay, các bạn đừng nghĩ rằng dùng SPSS hay Stata hay những nhu liệu như thế là đủ. Những nhu liệu đó chỉ thích hợp cho một đẳng cấp nghiên cứu mà thôi; các bạn muốn có đẳng cấp cao hơn thì phải dùng đến nhu liệu R. Các bạn có thể tham khảo các bài báo trên *PLoS Genetics*, *PLoS Medicine*, *PLoS Biology*, *PLoS ONE*, các tập san trong nhóm *Nature* (như *Nature Genetics*, *Nature Communications*, *Nature Scientific Reports*, v.v.) sẽ thấy những hình ảnh và biểu đồ phần lớn là từ R. Nghiên cứu khoa học tinh vi ngày nay đã đi xa cái thời của SPSS và Stata khá xa.

Trên đây là 4 yếu tố các bạn cần phải suy nghĩ khi bắt đầu nghiên cứu. Sau khi đã có ý tưởng, các bạn phải suy nghĩ đến một mô hình nghiên cứu thích hợp, đến đối tượng nghiên cứu, đến đo lường, và phân tích dữ liệu. Nhà nghiên cứu có kinh nghiệm có thể tiên đoán ngay từ lúc lên kế hoạch là kết quả này sẽ có thể đăng trên tập san nào. Những gì tôi mô tả trên đây chủ yếu là cho nghiên cứu y học. Trong thực tế, các nghiên cứu khoa học xã hội (KHXX) phong phú hơn nhiều và cách thiết kế thí nghiệm cũng đầy sáng tạo so với nghiên cứu y học.

THẾ NÀO LÀ “CÁI MỚI” TRONG KHOA HỌC?

Khi bàn về ý tưởng và sáng tác văn nghệ, Nhà thơ Nguyễn Bính từng nói rằng *“Làm văn nghệ khó lắm vì có nhiều ý, nhiều chữ quá tuyệt mà người khác đã viết trước mình. Mình chỉ nhái lại thôi”*. Chỉ cần thay đổi hai từ “khoa học” cho từ “văn nghệ” trong câu nói trên, chúng ta sẽ có một câu nói thích hợp cho tình hình nghiên cứu khoa học. Tuy nhiên, trong nghiên cứu khoa học, cái mới cần phải được hiểu một cách rộng hơn, bao gồm những cái mới không chỉ trong ý tưởng mà còn ở khâu phương pháp và ngay cả cách diễn giải kết quả nghiên cứu.

Vài thảo luận gần đây về luận án tiến sĩ đề cập đến “cái mới” như là một tiêu chuẩn để được cấp học vị cao nhất trong hệ thống giáo dục đại học. Trước những lem nhem về đề tài nghiên cứu cấp tiến sĩ, Bộ trưởng Bộ Giáo dục và Đào tạo đã từng phát biểu rằng nếu công trình nghiên cứu cấp tiến sĩ mà không có cái gì mới thì không nên làm. Lại có ý kiến cho rằng “cái mới” vẫn chưa đủ, mà “cái mới đó phải đúng, phải hay, phải dùng được” thì mới đủ điều kiện được cấp học vị tiến sĩ. Tuy nhiên, không ai đề cập đến thế nào là “cái mới” trong nghiên cứu khoa học, và do đó, vấn đề luận án tiến sĩ vẫn tiếp tục thu hút thảo luận mà không có kết quả sau cùng. Trong bài này tôi muốn trình bày một vài nhận xét qua kinh nghiệm cá

nhân về cái mới trong khoa học và tiêu chuẩn cho một học vị tiến sĩ.

Vài năm trước đây, tôi được mời làm giám khảo luận án tiến sĩ của một nghiên cứu sinh ngành di truyền học ở Mi. Trong một công trình nghiên cứu khá qui mô và tốn kém, nghiên cứu sinh phân tích 24 gen và tìm hiểu mối liên hệ giữa các gen này với một bệnh ở trẻ em. Kết quả cho thấy tất cả 24 gen đều không có liên hệ gì với bệnh như giả thuyết đặt ra lúc ban đầu! Nói theo ngôn ngữ thời thượng [không đúng lắm] là công trình nghiên cứu hoàn toàn “negative” (âm tính). Tuy nhiên, một số nghiên cứu trước cho thấy một số gen này có liên quan đến bệnh. Khi trình bày kết quả nghiên cứu trước hội đồng giám khảo, nghiên cứu sinh tỏ ra rất lo lắng, vì đã bỏ ra 4 năm trời vất vả nghiên cứu mà kết quả thì không như “mong muốn”, và quan trọng hơn là có thể học vị tiến sĩ sẽ chẳng đến tay mình. Tuy nhiên, khi đánh giá luận án, các giám khảo không mấy quan tâm đến phần kết quả, nhưng lại chú tâm vào ý tưởng, phương pháp tiếp cận, khả năng phân tích dữ liệu, và cách diễn giải dữ liệu của nghiên cứu sinh. Dựa vào các tiêu chí này, hội đồng giám khảo đồng ý đề nghị cấp học vị tiến sĩ cho nghiên cứu sinh.

Mục tiêu chính của chương trình học tiến sĩ là nhằm đào tạo ra những nhà khoa học độc lập. Nhà khoa học độc lập là người có kiến thức chuyên sâu, có khả năng xử lý thông tin hữu hiệu, có khả năng phát hiện vấn đề, có kỹ năng nghiên cứu khoa học, và có thể sáng tạo ra những tri thức khoa học mới. Do đó, để xứng đáng với học vị tiến sĩ, nghiên cứu sinh phải đạt hai điều kiện chung: phải chứng tỏ rằng mình đã quán triệt và làm chủ được lĩnh vực nghiên cứu; và thứ hai là phải phát triển hay củng cố thêm được một cái gì mới cho tri thức khoa học.

Cái cốt lõi của học vị tiến sĩ (và cũng là khía cạnh dùng để phân biệt học vị tiến sĩ với các học vị đại học khác) có thể tóm lược bằng hai chữ: nghiên cứu. Cần phải nhấn mạnh rằng học tiến sĩ không chỉ đơn thuần là soạn một luận án, bởi vì luận án chỉ là một trong những điều kiện để tốt nghiệp. Nghiên cứu khoa học, nói cho cùng, có mục đích sản xuất ra tri thức mới, và tri thức không hẳn phải là “dương tính” hay “âm tính”. Do đó, kết quả nghiên cứu của nghiên cứu sinh tiến sĩ không nhất thiết phải “dương tính”, hay phải có khả năng ứng dụng vào thực tế.

Hiểu về “cái mới” trong khoa học

Chỉ cần tham khảo qua một vòng các qui định về luận án tiến sĩ ở các trường bên Mỹ, Úc, Canada, Âu châu, v.v... để thấy rằng cái tiêu chuẩn chính của một luận án tiến sĩ có thể tóm lược như sau: luận án tiến sĩ phải thể hiện một đóng góp nguyên thủy và có ý nghĩa (significant original contribution) vào tri thức khoa học của chuyên ngành, thích hợp cho việc công bố trên các tập san khoa học. Tiêu chuẩn này hàm ý rằng toàn bộ hay một phần nội dung của luận án tiến sĩ phải được công bố trên các diễn đàn khoa học nghiêm túc (tức có hệ thống bình duyệt) trước hay sau khi đệ trình luận án.

Cụm từ “đóng góp nguyên thủy” (original contribution) ở đây phải được hiểu là sáng tạo ra tri thức mới. Bởi vì một công trình nghiên cứu khoa học khởi đầu từ một ý tưởng (idea), đến phương pháp (methods), kết quả (results), và diễn dịch (interpretation), cho nên nói đến tri thức mới ở đây cũng có nghĩa là nói đến những cái mới có liên quan đến ít nhất là một trong 4 khía cạnh vừa kể. Cái mới trong nghiên cứu khoa học do đó là một sự liên kết giữa ý tưởng và phương pháp, hay một

cách hệ thống hóa dữ liệu để xây dựng những giả thuyết, suy luận, và kết luận. Vì dữ liệu được thu thập qua cách tiếp cận và phương pháp, do đó, sau cùng chúng ta có 2 khía cạnh chính là dữ liệu và ý tưởng. Theo đó, chúng ta có 4 trường hợp:

- Dữ liệu mới + ý tưởng mới hay cách diễn giải mới
- Dữ liệu mới + ý tưởng cũ
- Dữ liệu cũ + ý tưởng mới hay cách diễn giải mới
- Dữ liệu cũ + ý tưởng cũ

Trong 4 trường hợp trên, chỉ có trường hợp cuối cùng là không mới; 3 trường hợp còn lại đều có thể xem là có đóng góp mới cho khoa học. Do đó, khi nói đến “cái mới” trong luận án tiến sĩ, người ta nói đến cái mới hoặc về ý tưởng, hoặc phương pháp, hoặc kết quả, hoặc cách diễn dịch. Trong thực tế, có thể ý tưởng hay vấn đề rất cũ, nhưng nếu nghiên cứu sinh có một cách tiếp cận bằng một phương pháp mới thì vẫn được xem là đóng góp “cái mới” vào tri thức khoa học. Thậm chí, có khi ý tưởng và phương pháp không mới, nhưng cách diễn giải mới thì luận án vẫn được xem là một đóng góp mới vào tri thức khoa học.

Chẳng hạn như ý tưởng nghiên cứu đánh giá ảnh hưởng của các yếu tố di truyền đến xương là một ý tưởng không mới (và trong quá khứ đã có nhiều nghiên cứu như thế bằng cách xem xét mối tương quan về khối lượng chất khoáng trong xương ở những người thân trong cùng gia đình), nhưng khi có một công trình nghiên cứu tiếp cận vấn đề bằng cách so sánh mối tương quan về chất lượng xương giữa những cặp sinh đôi lớn lên trong cùng một gia đình và những cặp sinh đôi sống xa nhau trong thời kì trưởng thành, thì nghiên cứu vẫn được xem là mới. Cái mới ở đây là cách tiếp cận vấn đề qua một mô

hình độc đáo để giải đáp một câu hỏi cổ điển. Để đi đến mô hình này, đương nhiên tác giả phải lí giải lí do sinh học, và lí giải rằng thông số trong mô hình này chính là một phản ánh về ảnh hưởng của di truyền. Có thể nói không ngoa rằng phần lớn các nghiên cứu y sinh học ngày nay, đứng về mặt ý tưởng không có gì mới, nhưng cách tiếp cận vấn đề và phương pháp (một phần do thành quả của những tiến bộ trong công nghệ sinh học) thì rất mới.

Cách đây vài năm, một nhà khoa học ở Úc nghiên cứu về... cách buộc dây giày. Công trình nghiên cứu này thoát đầu mới nghe qua tưởng như khôi hài, nhưng thực chất đó là một công trình ứng dụng toán học. Kết quả được công bố trên tập san Nature (5/12/2002), một tập san khoa học số 1 trên thế giới, thường được mệnh danh là diễn đàn của các nhà Nobel tương lai. Để cập đến công trình này, tôi muốn nói rằng bất cứ vấn đề gì trong cuộc sống, cho dù là vấn đề thực tế nhất tưởng như chẳng có gì phải tìm hiểu thêm, nếu đặt dưới lăng kính khoa học một cách nghiêm chỉnh và có phương pháp đều có thể đem lại nhiều phát hiện thú vị.

Cái mới phải đẹp, hữu dụng, và đúng?

Về ý kiến cho rằng luận án tiến sĩ phải có cái mới phải đẹp, hữu dụng, và đúng, thì tôi e rằng có sự nhầm lẫn giữa học vị và giải thưởng ở đây. Tiến sĩ là một học vị, chứ không phải là một giải thưởng khoa học. Giải Nobel hay giải Fields là phần thưởng cho những cá nhân có đóng góp quan trọng vào một lĩnh vực chuyên môn. Công trình nghiên cứu cấp tiến sĩ không nhất thiết phải sản sinh ra một cái mới đẹp, đúng và có ích; nhưng giải thưởng Nobel thì đòi hỏi các tiêu chí này. Chẳng hạn theo di chúc của ông Alfred Nobel, giải thưởng

mang tên ông chỉ được trao tặng cho những người có những khám phá đem lại lợi ích lớn nhất cho con người.

Nhưng ngay cả tiêu chí của giải Nobel cũng rất khó định nghĩa. Thật vậy, những người thừa kế và thi hành bản di chúc Nobel đã từng tranh luận, bàn cãi chi tiết về hai tiêu chí này (khám phá mới và có ích lớn nhất cho nhân loại) sau khi Nobel qua đời, nhưng họ không đi đến một tiêu chuẩn nào cụ thể. Cuối cùng, họ để quyền quyết định cho Ủy ban Nobel diễn dịch hai cụm từ trên sao cho phù hợp nhất! Một vấn đề phức tạp khác là cách dùng chữ khác nhau của ông Nobel trong di chúc. Đối với ngành vật lí ông dùng chữ “discovery and invention” (khám phá và sáng chế), trong khi trong ngành hóa học ông dùng chữ “discovery or improvement” (khám phá hay cải tiến).

“Discovery” hay khám phá thì dễ định nghĩa trong các ngành khoa học cơ bản, nhưng trong các ngành như khoa học lâm sàng thì định nghĩa thế nào là một “khám phá” không dễ chút nào. Khám phá thường được định nghĩa là những phát hiện mang tính bất ngờ, đem lại kiến thức mới cho con người. Mặt khác, cụm từ “lợi ích lớn nhất cho con người” thì rất dễ định nghĩa trong khoa học lâm sàng, nhưng đối với khoa học cơ bản thì lại rất khó định nghĩa. Một khám phá về cơ cấu vận hành của một tế bào trong khoa học cơ bản có thể không có ý nghĩa lâm sàng nào trong vòng 20 năm, nhưng lại có thể đem đến lợi ích cho bệnh nhân về lâu về dài, trong khi đó một thuật điều trị mới có thể đem lại lợi ích trước mắt cho con người. Để dung hòa cái khó khăn này, Ủy ban Nobel thường trao giải thưởng cho những khám phá nào mang tính cơ bản nhất.

Tính hữu dụng của một khám phá rất khó đánh giá vì nó có thể thay đổi theo thời gian. Năm 1948, Paul Hermann Muller được trao giải Nobel do công trình nghiên cứu và phát hiện

hóa chất dichloro-diphenyl-trichloromethylmethane (DDT), một hóa chất diệt cỏ và sâu bọ. Trong và sau thế chiến thứ hai, DDT không những được dùng để chống lại các bệnh như bệnh sốt Rickettsia, mà còn là một loại vũ khí lợi hại chống lại bệnh sốt rét. Tổ chức Y tế Thế giới (World Health Organization) ước tính rằng trong thời gian dùng DDT, hơn 25 triệu người đã được cứu sống. Ngày nay, có người đề nghị tẩy chay DDT vì họ cho rằng nó có thể làm ô nhiễm và độc hại đến môi trường. Nhưng vào lúc mà giải thưởng được trao tặng, lợi ích của DDT cho nhân loại là một điều quá hiển nhiên.

Trong khoa học rất khó có những khám phá “đúng”, bởi vì cái đúng là một khái niệm tương đối và tùy thuộc vào mô thức khoa học (sẽ bàn chi tiết dưới đây). Năm 1949 giải Nobel được trao cho một bác sĩ giải phẫu (Moniz) vì phẫu thuật thùy não (hay còn gọi là leucotomy) trong việc chữa trị chứng loạn thần kinh (psychotic). Trước khi phẫu thuật được ứng dụng, bệnh nhân phải chịu nhiều cực hình như “nhốt” trong một lồng sắt (straitjackets), tắm bằng nước lạnh, rồi mới đến giải phẫu và giật điện. Đến khi thuốc được phát triển để điều trị bệnh, thì phẫu thuật này đã trở thành một phương pháp của quá khứ. Thật ra, ngày nay phẫu thuật thùy não được xem là một phẫu thuật vi phạm y đức.

“Khoa học bình thường”

Nghiên cứu khoa học ngày nay thường được thực hiện trong khuôn khổ của những mô thức (paradigm). Một mô thức bao gồm nhiều giả định, định luật, và phương pháp mà các thành viên trong cộng đồng khoa học phải tuân theo hay xem đó là chuẩn. Đại đa số các nhà khoa học trong một mô thức, dù là vật lý học hay sinh học, thực hành cái mà Thomas

Kuhn gọi là “Khoa học bình thường”. Những nhà khoa học bình thường làm việc trong khuôn khổ của một mô thức, cố gắng phản nghiệm giả thuyết (falsificationism), đóng góp vào việc phát triển cái mô thức có sẵn đó để giải thích các hiện tượng liên quan đến lĩnh vực chuyên môn của họ qua nghiên cứu. Theo thời gian, đến một thời điểm nào đó, nhà khoa học sẽ đương đầu với khó khăn. Nếu không có phương pháp nào khắc phục được khó khăn đó, khoa học lâm vào trạng thái khủng hoảng. Trạng thái khủng hoảng sẽ được giải quyết khi một mô thức hoàn toàn mới xuất hiện và thu hút được càng ngày càng nhiều nhà khoa học khác cho đến khi cái mô thức cũ bị hoàn toàn hủy bỏ. Sự gián đoạn này được gọi là một cuộc cách mạng khoa học. Cái mô thức mới cũng sẽ trải qua một qui trình như thế: khó khăn, khủng hoảng, cách mạng, và một mô thức mới ra đời. Chẳng hạn như trong lĩnh vực y sinh học, mô thức cũ về vi khuẩn là nguyên nhân gây bệnh được thay thế bằng mô thức gen và DNA và cho ra đời bộ môn sinh học phân tử. Phần lớn những nghiên cứu y sinh học ngày nay chỉ xoay quanh mô thức mới này.

Trong môi trường hoạt động khoa học bình thường của một mô thức, chúng ta không thể kì vọng có những khám phá hoàn toàn mới (ngoại trừ trường hợp một mô thức mới xuất hiện và nhà khoa học may mắn là người dẫn đầu mô thức đó). Đó cũng chính là lí do mà ngày nay chúng ta hiếm thấy những khám phá cơ bản như thời “vàng son” của khoa học trước đây gắn liền với những cái tên lừng danh như Federick Gauss, Albert Einstein, Marie Curie, Max Planck, Niels Bohr, Charles Darwin, Louis Pasteur, James Watson, v.v... Chúng ta không thể kì vọng bất cứ một nghiên cứu sinh tiến sĩ nào cũng sẽ trở thành những ngôi sao như vừa kể, mà chỉ mong các tân tiến sĩ

có những đóng góp mới vào tri thức khoa học chuyên ngành. Những đóng góp mới đó có thể - nhưng chúng ta không kì vọng hay đòi hỏi - đưa họ vào hàng ngũ các “đại thụ” trong khoa học trong tương lai.

Nói tóm lại, một công trình nghiên cứu cấp tiến sĩ phải thể hiện một đóng góp mới cho chuyên ngành, hay sáng tạo ra tri thức mới cho khoa học. Nhưng “cái mới” ở đây cần phải được hiểu trong khuôn khổ của một mô thức khoa học và được đa số cộng đồng khoa học công nhận. Nội dung một luận án tiến sĩ không nhất thiết phải có những cái mới toàn diện từ ý tưởng, phương pháp, kết quả đến diễn dịch, hay là những khám phá mang tính cách mạng một lĩnh vực, nhưng phải thể hiện một sự đóng góp có ý nghĩa vào tri thức của chuyên ngành. Lịch sử của phát triển khoa học cho thấy các lí thuyết thay đổi và tiến hóa không ngừng, bất kể các lí thuyết này mang tính khẳng định cỡ nào. Nếu chúng ta nhìn hệ thống tri thức khoa học như là một tòa nhà, thì những tri thức và đóng góp từ các nghiên cứu sinh tiến sĩ là những vật liệu dùng để xây dựng và sửa chữa tòa nhà khoa học hoàn hảo hơn.

Một số tiêu chuẩn để đánh giá “cái mới” trong nghiên cứu khoa học

- Phát triển một lí thuyết mới (rất hiếm trong nghiên cứu cấp tiến sĩ).
- Diễn dịch lại hay xét lại một lí thuyết hiện hành trong bối cảnh mới.
- Phát triển một công cụ nghiên cứu mới hay một kĩ thuật mới.
- Phát triển một mô hình mới để tiếp cận một vấn đề cũ.
- Thực hiện những nghiên cứu chuyên sâu về một vấn đề chưa từng biết đến trước đây hay chưa từng được điều tra trước đây.
- Thực hiện những phân tích chi tiết về một vấn đề chưa từng biết đến trước đây hay chưa từng được điều tra trước đây.
- Phát triển một sản phẩm mới, hay cải tiến sản phẩm hiện có.
- Cung cấp dữ liệu mới hay kết luận mới chưa từng có trước đây.

CHỌN MÔ HÌNH NGHIÊN CỨU VÀ NHỮNG VẤN ĐỀ LIÊN QUAN ĐẾN THIẾT KẾ

Sau khi đã xác định được giả thuyết khoa học và câu hỏi nghiên cứu, bước kế tiếp là chọn mô hình nghiên cứu. Một câu hỏi nghiên cứu có thể giải đáp (hay tiếp cận) bằng một hay nhiều mô hình nghiên cứu. Mỗi mô hình nghiên cứu có những điểm mạnh và điểm yếu, và việc chọn mô hình thích hợp tùy thuộc vào tính khả thi, ngân sách, và tài nguyên sẵn có. Chương này sẽ bàn qua những mô hình nghiên cứu phổ biến trong nghiên cứu khoa học.

Chọn mô hình nghiên cứu khoa học

Nghiên cứu khoa học được công bố trên các tập san khoa học rất đa dạng, và xuất hiện với nhiều danh từ, thuật ngữ không mấy quen thuộc. Một số nghiên cứu thường thấy là *meta-analysis* (phân tích tổng hợp), *randomized controlled clinical trial* (RCT, nghiên cứu lâm sàng đối chứng ngẫu nhiên), *prospective (longitudinal) study* (nghiên cứu theo thời gian), *cross-sectional study* (nghiên cứu tiêu biểu một thời điểm), và *case-control study* (nghiên cứu bệnh chứng). Giá trị khoa học của những nghiên cứu này không giống nhau. Bảng dưới đây mô tả vài nét chính của mỗi mô hình

nghiên cứu và giá trị khoa học xếp theo tứ tự thấp nhất đến cao nhất:

Nấc thang giá trị khoa học cho các loại nghiên cứu

Giá trị (thấp nhất đến cao nhất)	Tên loại nghiên cứu	Vài đặc điểm chính
1	Ý kiến của các chuyên gia, bài điểm báo, xã luận, bình luận.	Thường là các bài báo được các tạp san mời viết.
2	Nghiên cứu cơ bản trên chuột và động vật cấp thấp.	Các nghiên cứu sơ khởi trong phòng thí nghiệm về một phân tử hay tác nhân cụ thể.
3	Báo cáo các ca đặc biệt (case reports).	Phân tích chuyên sâu những trường hợp cá biệt trong xã hội. Kinh nghiệm điều trị về một hay vài trường hợp lâm sàng đặc biệt và hiếm thấy.
4	Nghiên cứu bệnh chứng (case-control study).	Là cách gọi chung cho nghiên cứu gồm hai nhóm: nhóm có đặc điểm (có bệnh) và nhóm không có đặc điểm (không có bệnh, còn gọi là 'nhóm chứng'). Trong y học, mục đích của nghiên cứu bệnh chứng là tìm hiểu mối liên hệ giữa một yếu tố nguy cơ và một bệnh lí.
5	Nghiên cứu tiêu biểu tại một thời điểm (cross-sectional study).	Trong khoa học xã hội, đây là một điều tra xã hội (survey) nhằm ước tính mức độ phổ biến của một hiện tượng. Trong y khoa, mục đích thường là ước tính tỉ lệ hiện hành của bệnh (prevalence) và các yếu tố liên quan đến bệnh trong một quần thể.

6	Nghiên cứu theo thời gian (prospective / longitudinal study).	Là những nghiên cứu theo dõi đối tượng theo thời gian nhưng không can thiệp. Mục đích thường là ước tính tỉ lệ phát sinh của bệnh (incidence) và các yếu tố liên quan đến bệnh trong một quần thể.
7	Thử nghiệm lâm sàng đối chứng ngẫu nhiên (randomized controlled clinical trial - RCT).	Sử dụng trong việc thẩm định mức độ hiệu nghiệm của một thuật điều trị lâm sàng trong một nhóm đối tượng cụ thể.
8	Phân tích tổng hợp (meta-analysis).	Tổng hợp các kết quả nghiên cứu RCT, cross-section, longitudinal và case-control.

Ý kiến cá nhân và điểm báo, xã luận

Một số bài báo trên các tập san y học là những ý kiến cá nhân của các giáo sư và chuyên gia trong ngành. Những ý kiến này thường xuất hiện dưới hình thức một bài xã luận (editorial), bình luận (commentary), hay điểm báo (review). Đây là những bài báo có giá trị thấp nhất trong khoa học, vì chúng chẳng cung cấp một bằng chứng khách quan, mà chỉ là những ý kiến cá nhân. Mà, ý kiến cá nhân thì thường, không ít thì nhiều, mang tính chủ quan. Đó là chưa kể một số giáo sư chỉ là những cái loa phát thanh cho các công ti dược, hay lợi dụng tập san để quảng cáo cho công ti của chính họ, chứ mục đích phục vụ khoa học chỉ là thứ yếu.

Nghiên cứu cơ bản (basic research)

Trong y học, các nghiên cứu cơ bản đóng vai trò rất quan trọng trong khám phá y khoa. Không có các nghiên cứu này, chúng ta chắc chắn không có một nền y học hiện đại như hiện nay. Những nghiên cứu này thường được tiến hành trên chuột hay thỏ, hay trong ống nghiệm. Đó là những nghiên cứu khởi

đầu để thử nghiệm một giả thiết khoa học cụ thể nào đó. Bởi vì chỉ là nghiên cứu sơ khởi và không tiến hành trên bệnh nhân, cho nên đối với y học thực chứng giá trị khoa học của các nghiên cứu này rất thấp, chỉ hơn ý kiến cá nhân một bậc mà thôi.

Báo cáo trường hợp lâm sàng (case reports)

Bằng chứng có giá trị cao hơn nghiên cứu cơ bản là các phân tích về các trường hợp lâm sàng. Trong nhiều tập san y học, giới nghiên cứu thường báo cáo trường hợp bệnh nhân rất đặc biệt, rất hiếm, những trường hợp bệnh khó chẩn đoán, hay những trường hợp bệnh lí chưa được mô tả trong sách giáo khoa, những ca giải phẫu ngoạn mục, v.v... Có khi bác sĩ thành công trong chẩn đoán và điều trị (nếu không thì chắc chắn sẽ không có bài báo!) và những thành công đó không thể xem là bằng chứng khoa học được, vì không có giá trị khái quát hóa đến các bệnh nhân khác.

Nghiên cứu bệnh chứng (case-control study)

Các nghiên cứu có giá trị khoa học “trung bình” là các case-control study. Trong các nghiên cứu này, mục đích chính là tìm hiểu mối liên hệ giữa một yếu tố nguy cơ (risk factor) và một bệnh rất cụ thể. Để tiến hành nghiên cứu này, nhà nghiên cứu phải “đi ngược thời gian” với những qui trình được minh họa qua một nghiên cứu tiêu biểu mối liên hệ giữa việc sử dụng hormone (hormone therapy hay HT) và gãy xương như sau:

Bước thứ nhất: chọn một nhóm đối tượng đã bị gãy xương (còn gọi là *cases*) mà nhà nghiên cứu muốn tìm hiểu;

Bước thứ hai: chọn nhóm đối chứng (gọi là *controls*) không bị gãy xương, nhưng mỗi người trong nhóm này phải có cùng độ tuổi, cùng giới tính, và các yếu tố lâm sàng khác với nhóm đối tượng;

Bước thứ ba: xem xét trong mỗi nhóm đối tượng và nhóm đối chứng có bao nhiêu người sử dụng HT và bao nhiêu người không sử dụng HT.

Bước thứ tư: ước tính odds ratio (hay thường viết tắt OR). Khi OR thấp hơn 1 có nghĩa là sử dụng HT giảm gây xương; khi $OR = 1$ có nghĩa là HT không có ảnh hưởng gì đến gây xương; và nếu OR cao hơn 1 thì đó là một “tín hiệu” cho thấy sử dụng HT có thể làm tăng nguy cơ gây xương.

Nghiên cứu tiêu biểu một thời điểm (cross-sectional studies)

Các nghiên cứu có giá trị khoa học trên trung bình là nghiên cứu tiêu biểu một thời điểm. Có người dịch “cross-sectional study” là “nghiên cứu cắt ngang”, nhưng tôi thấy cách dịch này không nói lên thực chất của nó. Trong thực tế, nghiên cứu này chọn một quần thể một cách ngẫu nhiên nhưng tiêu biểu cho một cộng đồng, tại một thời điểm nào đó. Mục đích chính của các nghiên cứu này là tìm hiểu tỉ lệ hiện hành (prevalence) của một bệnh nào đó, hay tìm hiểu mối tương quan giữa một yếu tố nguy cơ và một bệnh. Chẳng hạn như nếu chúng ta muốn tìm hiểu có bao nhiêu người trong độ tuổi 50 trở lên bị bệnh xơ vữa động mạch ở dân số sống trong nông thôn và thành phố, chúng ta có thể tiến hành như sau:

Bước thứ nhất: chọn một thành phố và một vùng nông thôn (với các tiêu chuẩn phù hợp cho định nghĩa thế nào là thành phố và thế nào là nông thôn), tìm hiểu xem có bao nhiêu quận và huyện;

Bước thứ hai: lập danh sách đối tượng nghiên cứu, gồm những đàn ông và phụ nữ trên 50 tuổi;

Bước thứ ba: sử dụng máy tính soạn một chương trình chọn (chẳng hạn như) 5% tổng số đối tượng đó một cách ngẫu nhiên;

Bước thứ tư: thu thập dữ liệu các đối tượng được chọn tại một thời điểm nào đó;

Bước thứ năm: xác định xem có bao nhiêu người bị xơ vữa động mạch, và phân tích mối liên hệ giữa các yếu tố lâm sàng và nguy cơ mắc bệnh.

Nghiên cứu theo thời gian (longitudinal studies hay prospective study)

Nghiên cứu có giá trị cao hơn nghiên cứu cross-section là các nghiên cứu theo thời gian. Mục đích của các nghiên cứu này là tìm hiểu mối liên hệ giữa một hay nhiều yếu tố nguy cơ và nguy cơ phát sinh bệnh tật (incidence). Khác với nghiên cứu cross-section chỉ ghi nhận sự kiện tại một thời điểm, các nghiên cứu longitudinal phải theo dõi đối tượng trong một thời gian có thể là nhiều năm tháng. Do đó, nghiên cứu longitudinal thường tốn tiền hơn và công phu hơn các nghiên cứu cross-section. Chẳng hạn như chúng ta muốn tìm hiểu mối liên hệ giữa tỉ trọng cơ thể (body mass index - BMI) và nguy cơ tử vong, hay ước tính khả năng tiên đoán tử vong của BMI, chúng ta có thể tiến hành nghiên cứu theo trình tự sau đây:

Bước thứ nhất: chọn một quần thể tiêu biểu gồm N người một cách ngẫu nhiên tại một địa phương (hay nhiều địa phương);

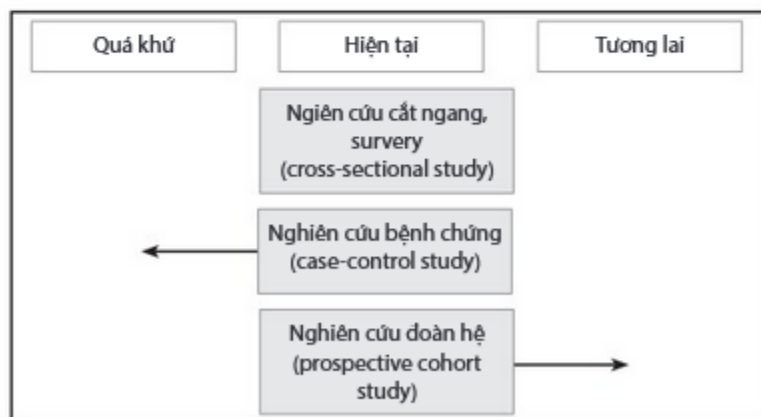
Bước thứ hai: thu thập dữ liệu (như BMI chẳng hạn) ban đầu tại thời điểm t_0 ;

Bước thứ ba: theo dõi quần thể trong một thời gian T và trong thời gian này, thu thập thêm dữ liệu và ghi nhận bao nhiêu người chết;

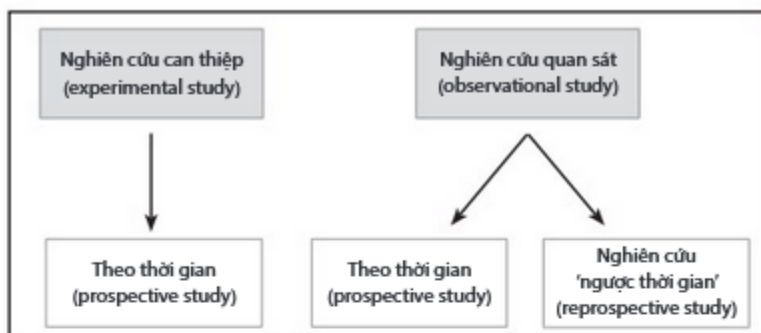
Bước thứ tư: phân tích mối liên hệ giữa BMI tại thời điểm t_0 và tỉ lệ tử vong trong thời gian T . Kết quả của một nghiên cứu như thế có thể tóm tắt bằng một bảng số liệu như sau:

Đến đây, chúng ta cần phải phân biệt hai chỉ số quan trọng: tỉ lệ phát sinh - *incidence* và tỉ lệ hiện hành - *prevalence*. Tỉ lệ phát sinh chỉ có thể ước tính cho các nghiên cứu theo thời gian (tức là *prospective study*), những nghiên cứu phải theo dõi đối tượng trong một thời gian và thu thập tần số phát sinh bệnh trong thời gian đó. Còn tỉ lệ hiện hành chỉ có thể ước tính cho các nghiên cứu *cross-section* mà thôi.

Vì nghiên cứu *cross-section* chỉ thu thập thông tin tại một thời điểm nhất định, tỉ lệ hiện hành thực chất là một chỉ số đi ngược thời gian. Trong nghiên cứu *cross-section* khi phỏng vấn đối tượng, chúng ta có thể hỏi - chẳng hạn như - “anh/chị từng bị bệnh xơ vữa động mạch bao giờ chưa”; do đó, tỉ lệ hiện hành không cung cấp cho chúng ta nguy cơ tương lai. Nhưng tỉ lệ hiện hành cho chúng ta biết có bao nhiêu người bị bệnh trong một quần thể, và do đó cho chúng ta biết *qui mô* của vấn đề.



Phân biệt loại hình nghiên cứu theo thời gian



Phân biệt loại hình nghiên cứu theo đặc tính

Ngược lại với qui mô và tính quá khứ của tỉ lệ hiện hành, tỉ lệ phát sinh cung cấp cho chúng ta thông tin về *nguy cơ* trong *tương lai* được giới hạn trong một thời gian nhất định. Tỉ lệ phát sinh thường thấp hơn tỉ lệ hiện hành. Biết được tỉ lệ phát sinh là một điều rất quan trọng và có ích cho việc lên kế hoạch để tiến hành những nghiên cứu lâm sàng đối chứng ngẫu nhiên.

Nghiên cứu lâm sàng đối chứng ngẫu nhiên (randomized controlled clinical trial - RCT)

Các nghiên cứu vừa nêu trên (cross-section, longitudinal, và case control studies) là những *nghiên cứu quan sát* (observational study), vì nhà nghiên cứu không can thiệp vào đối tượng. Ngược lại với các nghiên cứu quan sát, thử nghiệm RCT là những nghiên cứu nhằm thử nghiệm mức độ ảnh hưởng, mức độ hiệu nghiệm của một sự can thiệp (intervention). “Can thiệp” ở đây phải hiểu là những thuật điều trị, kể cả thuốc và phẫu thuật, nhưng tôi sẽ chú trọng vào các RCT thử nghiệm thuốc. Có khá nhiều cách thiết kế nghiên cứu RCT, nhưng tôi chỉ nói đến một dạng nghiên cứu

thông thường nhất: đó là nghiên cứu gồm hai nhóm bệnh nhân (còn gọi là *parallel design*). Để minh họa cho khái niệm RCT, tôi sẽ lấy ví dụ của công trình nghiên cứu *Women's Health Initiatives* làm ví dụ. Trong công trình nghiên cứu này, các nhà nghiên cứu muốn thử nghiệm xem mức độ hiệu nghiệm của calcium và vitamin D trong việc phòng chống gãy xương ở phụ nữ sau thời mãn kinh như thế nào. Họ tiến hành những bước sau đây:

Bước thứ nhất: chọn một quần thể gồm 36.282 phụ nữ sau thời mãn kinh (tuổi từ 50 đến 79), thu thập tất cả các dữ liệu lâm sàng liên quan;

Bước thứ hai: dùng máy tính để phân chia quần thể đó thành hai nhóm một cách ngẫu nhiên: nhóm 1 gồm có 18.176 phụ nữ được điều trị bằng calcium và vitamin D hàng ngày (nhóm *can thiệp*); nhóm 2 gồm 18.106 phụ nữ cùng độ tuổi nhưng không được bổ sung calcium hay vitamin D (còn gọi là nhóm đối chứng hay *placebo*);

Bước thứ ba: theo dõi hai nhóm can thiệp và đối chứng trong thời gian 10 năm (tính trung bình là 7 năm, vì một số qua đời, một số mất liên lạc, và một số không muốn tiếp tục tham gia công trình nghiên cứu). Trong thời gian này, các nhà nghiên cứu tiếp tục thu thập dữ liệu lâm sàng và đo lường sinh hóa để theo dõi tình trạng của bệnh nhân, cũng như ghi nhận số phụ nữ bị gãy xương;

Bước thứ tư: sau khi hết thời hạn theo dõi, các nhà nghiên cứu lên kế hoạch phân tích dữ liệu xem xét hiệu quả của vitamin D và calcium đến việc giảm nguy cơ gãy xương. Kết quả chính của công trình nghiên cứu có thể tóm lược bằng một bảng thống kê như sau:

	Nhóm can thiệp (calcium và vitamin D)	Nhóm chứng (placebo)
Số bệnh nhân	18.176	18.106
Thời gian theo dõi	7 năm	7 năm
Số phụ nữ bị gãy xương đùi	175	199
Tỉ lệ gãy xương đùi (%)	0,96	1,10

Nguồn: Jackson, RD, La Croix AZ, Gass M, et al. Calcium Plus Vitamin D Supplementation and the Risk of Fractures. *New Engl J Med* 2006; 354: 669-683

HOPE: một nghiên cứu RCT tiêu biểu

Chẳng hạn như trong công trình nghiên cứu HOPE (Heart Outcomes Prevention Evaluation) (3) tuyển chọn bệnh nhân trên 55 tuổi có nguy cơ bệnh tim mạch cao (tiền sử bệnh tim mạch, đột quỵ, tiểu đường, và ít nhất là một yếu tố nguy cơ như cao huyết áp, tăng cholesterol, hút thuốc lá, v.v...). Tính chung, công trình nghiên cứu tuyển chọn được 9297 bệnh nhân. Bệnh nhân được chia thành hai nhóm một cách ngẫu nhiên: nhóm dùng ramipril (10 mg/ngày) và nhóm chứng (placebo). Ngoài ra, bệnh nhân còn được ngẫu nhiên chia thành 2 nhóm: một nhóm sử dụng vitamin D và một nhóm chứng (giả vitamin E). Nói cách khác, đây là một công trình nghiên cứu được thiết kế theo mô hình “factorial” rất tinh vi và có hệ thống.

Bệnh nhân được theo dõi trong vòng 5 năm. Trong thời gian theo dõi, các nhà nghiên cứu ghi nhận các biến cố lâm sàng như nhồi máu cơ tim (myocardial infarction), đột quỵ, hay tử vong do bệnh tim mạch. Họ gọi chung các biến cố lâm sàng này là “composite outcome” (chỉ số hỗn hợp lâm sàng). Kết quả có thể tóm lược trong bảng số liệu sau đây:

Kết quả chính của công trình nghiên cứu HOPE

Biến cố lâm sàng	Ramipril (n = 4645) ^a	Placebo (n = 4652) ^a	Tỉ số nguy cơ (relative risk) và khoảng tin cậy 95%
Nhồi máu cơ tim	459 (9,9)	570 (12,3)	0,80 (0,70 – 0,90)
Đột quỵ	156 (3,4)	226 (4,9)	0,68 (0,56 – 0,84)
Tử vong do bệnh tim mạch	282 (6,1)	377 (8,1)	0,74 (0,64 – 0,87)
Chỉ số hỗn hợp	651 (14,0)	826 (17,8)	0,78 (0,70 – 0,86)

Chú thích:

- (a) Số ngoài ngoặc kép là số bệnh nhân có biến cố lâm sàng; số trong ngoặc kép là phần trăm tính trên tổng số bệnh nhân cho từng nhóm. Ví dụ: Trong số 4645 bệnh nhân sử dụng ramipril, có 459 người hay 9,9% bị nhồi máu cơ tim, trong số 4652 bệnh nhân nhóm chứng (placebo) có 570 người hay 12,3% bị nhồi máu cơ tim. Do đó, tỉ số nguy cơ là $9,9 / 12,3 = 0,80$, và tỉ số này dao động trong khoảng 0,70 đến 0,90 với xác suất 95%.

Số liệu trên cho thấy, ramipril giảm nguy cơ nhồi máu cơ tim 20%, giảm nguy cơ đột quỵ 32%, và giảm tử vong do bệnh tim mạch 25%. Tính chung các biến cố lâm sàng, ramipril giảm 22%. Do những kết quả khả quan trên đây, ủy ban thẩm định dữ liệu của công trình nghiên cứu đề nghị ngưng công trình nghiên cứu sớm hơn dự tính lúc ban đầu. Nói cách khác, vì các nhà nghiên cứu thấy thuốc ramipril có hiệu quả tốt, nên không có lí do để tiếp tục cho phân nửa bệnh nhân phải uống placebo.

Về tác động toàn cầu, theo tính toán của các nhà nghiên cứu, nếu 1/4 các bệnh nhân ở các nước đang phát triển (như Việt Nam chẳng hạn) và 1/2 bệnh nhân ở các nước kĩ nghệ tiên tiến được điều trị bằng ramipril, khoảng 2 triệu biến cố lâm sàng trên sẽ được ngăn ngừa.

Tuy nhiên, khi so sánh giữa hai nhóm vitamin E và placebo, không có khác biệt về tử vong, nhồi máu cơ tim hay đột quỵ giữa hai nhóm. Nói cách khác, vitamin E không có hiệu quả giảm các biến cố lâm sàng liên quan đến bệnh tim mạch.

Nói chung, đối với y học thực chứng, trong tất cả các nghiên cứu lâm sàng, kết quả từ các công trình thử nghiệm lâm sàng đối chứng ngẫu nhiên (RCT) được đánh giá là những nghiên cứu có giá trị khoa học cao nhất, nếu những nghiên cứu này được tiến hành nghiêm chỉnh và tuân theo các tiêu chuẩn về thiết kế, phương pháp phân tích, cũng như tuyển chọn đối tượng. Có thể nói không ngoa rằng công trình nghiên cứu HOPE đánh dấu một bước ngoặt trong nghiên cứu lâm sàng và RCT vì phương pháp thiết kế mới và logic, và kết quả của công trình này làm thay đổi thực hành lâm sàng.

Phân tích tổng hợp (Meta-analysis)

Thông thường với bất cứ một thuốc nào hay một thuật điều trị nào, một nghiên cứu RCT đơn lẻ không thể nào giải quyết dứt khoát vấn đề hiệu nghiệm, mà thường cần phải có nhiều nghiên cứu độc lập nhau. Nhiều nghiên cứu trên nhiều quần thể khác nhau, dù có cùng mục đích và tiến hành với một phương pháp chuẩn, có thể cho ra nhiều kết quả khác nhau, thậm chí trái ngược nhau. Trước sự khác biệt về kết quả như thế, vấn đề đặt ra là làm sao hệ thống hóa tất cả các kết quả nghiên cứu này để cho ra một “đáp số”, một “phán quyết” sau cùng. Gần đây, có một nghiên cứu mới ra đời có tên là “**meta-analysis**” mà tôi tạm dịch ra “**phân tích tổng hợp**”, vì thực chất của các nghiên cứu này là phân tích tất cả các kết quả nghiên cứu trong quá khứ. Trong y học thực chứng, kết

quả phân tích tổng hợp được đánh giá là có giá trị khoa học cao nhất.

Cần phải nói thêm rằng, phân tích tổng hợp không chỉ áp dụng cho việc hệ thống hóa các nghiên cứu RCT, mà còn có thể áp dụng để tổng hợp kết quả từ các nghiên cứu cases-control, cross-section, và longitudinal. Trong mấy năm gần đây, trước tình trạng các nghiên cứu đơn lẻ cho ra quá nhiều kết quả khác nhau khó diễn dịch, phân tích tổng hợp đã trở thành một phương tiện nghiên cứu rất thông dụng. Không chỉ trong lĩnh vực nghiên cứu y khoa, mà các lĩnh vực khác như sinh học, nông nghiệp, tâm lý học, xã hội học, kinh tế học, v.v... đều bắt đầu sử dụng phân tích tổng hợp rất nhiều.

Trên đây chỉ là khái quát về những nghiên cứu y học mà bạn đọc hay tìm thấy trong các tập san y khoa trên thế giới. Hiểu biết được sự khác nhau giữa các nghiên cứu và nấc thang giá trị khoa học của chúng là một bước đầu trong việc thẩm định thông tin khoa học. Ngoài việc đó, người đọc còn phải đặt một số câu hỏi thông thường như: nghiên cứu này được tiến hành ở đâu, do ai làm, đối tượng trong nghiên cứu thuộc độ tuổi và giới tính nào, có bệnh gì, thời gian theo dõi bao lâu, các dữ liệu được phân tích như thế nào, cách diễn dịch có phù hợp với dữ liệu hay không, và quan trọng là nghiên cứu đã được công bố chưa, ở đâu. Tất cả các câu hỏi này sẽ giúp cho người tiếp nhận thông tin quyết định xem các kết quả nghiên cứu có giá trị gì hay không, và nếu có thì chúng có áp dụng cho cá nhân được hay không.

Ưu điểm và yếu điểm của vài mô hình nghiên cứu

Nghiên cứu cắt ngang (cross-sectional study)

Ưu điểm

- Có thể nghiên cứu toàn bộ một quần thể;
- Cung cấp ước tính tỉ lệ hiện hành và các yếu tố nguy cơ;
- Có khả năng khái quát hóa rộng hơn;
- Tương đối dễ thực hiện.

Yếu điểm

- Rất dễ bias (như “survival bias”, hiện tượng bệnh nhân sống lâu và có thể làm cho việc ước tính tỉ lệ hiện hành không đúng với thực tế);
- Dễ bị sai lầm trong phân nhóm (ví bệnh nhân... quên);
- Không thể phát biểu về nguyên nhân và hệ quả (vì chỉ 1 thời điểm);
- Không thực tế đối với các bệnh hiếm.

Nghiên cứu bệnh chứng (case-control study)

Ưu điểm

- Rất hữu hiệu cho các bệnh hiếm;
- Không tốn kém;
- Có thể thực hiện nhanh.

Yếu điểm

- Dễ bị “selection bias” và chọn sai đối tượng nghiên cứu;
- Dễ bị sai lầm trong việc phân bệnh;
- Không thể phát biểu về nguyên nhân và hệ quả.

Nghiên cứu xuôi thời gian (longitudinal study)

Ưu điểm

- Có thể phát biểu về nguyên nhân và hệ quả (vì thông tin về yếu tố nguy cơ được thu thập trước khi bệnh phát sinh);
- Có thể nghiên cứu các bệnh hiếm hay với tần số thấp.

Yếu điểm

- Tốn kém thời gian và tiền;
- Đối tượng “bỏ cuộc” hay “lost to follow-up”.

Nghiên cứu lâm sàng đối chứng ngẫu nhiên (randomized controlled clinical trial)

Ưu điểm

- Giảm các yếu tố “nhiều” có thể ảnh hưởng đến bệnh như trong các nghiên cứu cắt ngang;
- Có chọn nhóm ngẫu nhiên;
- Có thể phát biểu về nguyên nhân và hệ quả.

Yếu điểm

- Rất tốn kém thời gian và tiền;
- Đối tượng “bỏ cuộc” hay “lost to follow-up”.

Những 'biases' trong nghiên cứu khoa học

Bias là một khái niệm hơi khó dịch một cách gọn gàng sang tiếng Việt. Chúng ta có thể hiểu bias là sự sai lệch giữa giá trị quan sát và giá trị thật. Nếu giá trị thật là T, và giá trị quan sát là O thì hiệu số giữa T và O được gọi là *bias*. Khái niệm *bias* áp dụng cho một chỉ số thống kê (như trung bình, độ lệch chuẩn), nhưng cũng áp dụng cho các qui luật tự nhiên. Khi chúng ta nói một chỉ số thống kê là *biased*, chúng ta ám chỉ rằng ước số của chỉ số này **lệch so với giá trị thật**. Tương tự, nếu một nghiên cứu cho ra kết quả lệch so với thực tế (có nghĩa là từ quần thể) thì nghiên cứu đó được xem là *biased*.

Survival bias

Trong nghiên cứu y khoa, survival bias có nghĩa là nghiên cứu chỉ tuyển dụng được những người khỏe mạnh và còn sống, còn những người kém khỏe mạnh thì đã chết. Do đó, ước số của mẫu nghiên cứu có xu hướng lạc quan hơn thực tế. Trong kiểm định giáo dục tôi cũng nghĩ có hiện tượng survival bias, vì chúng ta chỉ đánh giá được những em còn theo đuổi việc học, chứ không đánh giá được những em đã bỏ học. Tỷ lệ bỏ học ở Việt Nam ngày càng cao như báo chí hay đề cập đến. Những em bỏ học có thể do học lực kém và do tình hình kinh tế gia đình không tốt. Ngoài tình trạng bỏ học, còn có tình trạng ghi danh cấp trung học. Theo một số liệu tôi đọc được trước đây của Ngân hàng Thế giới, tỷ lệ học sinh ghi danh lên trung học cấp I dao động trong khoảng 90-92%, còn đến cấp II thì giảm còn 65% mà thôi. Tỷ lệ ghi danh trung học cấp II ở Mỹ và Úc là khoảng 90-98%. Do đó, kết quả kiểm định PISA cũng bị bias hướng lạc quan hơn thực tế.

Bias trong cách lấy mẫu

Đây là một bias phổ biến và rất khó tránh khỏi. Theo báo cáo của PISA thì họ chọn mẫu phân tầng, và theo đó, họ chọn trường một cách ngẫu nhiên, và mỗi trường được chọn họ cũng ngẫu nhiên chọn học sinh. Đó là cách chọn mẫu hợp lý và khoa học. Nhưng cái khó khăn ở đây là phải cân đối vùng miền (Nam, Trung, Bắc), nơi cư trú (nông thôn, thành thị), loại trường công và trường bán công, v.v. Dĩ nhiên, ai cũng biết lấy mẫu đại diện thì kết quả sẽ đáng tin cậy hơn là lấy mẫu thiếu đại diện.

Có tất cả 188 trường tham gia vào chương trình kiểm định PISA 2015. Nhưng số học sinh không cân đối giữa vùng miền. Chẳng hạn như tỉ lệ học sinh vùng thành thị ở các trường miền Bắc là 48%, và tỉ lệ này cao hơn miền Trung (43%), nhưng thấp hơn miền Nam (56%). Tương tự, tỉ lệ trường công cũng không đồng đều giữa các miền. Tỉ lệ học sinh trường bán công ở miền Bắc là 12%, cao gần gấp 2 lần so với miền Trung (6,7%) và miền Nam (7%). Bởi vì học sinh thành thị và học sinh trường công có khả năng học tập tốt hơn học sinh vùng nông thôn và trường bán công, nên cách lấy mẫu thiếu cân đối như trên có thể dẫn đến bias.

Hawthorne bias

Trong y khoa, thuật ngữ Hawthorne bias (còn gọi là Hawthorne effect) thường được dùng để đề cập đến hiện tượng bệnh nhân hay đối tượng nghiên cứu tự điều chỉnh trong môi trường bị quan sát / theo dõi. Trong kiểm định giáo dục hay thi cử, khi học sinh ý thức được tầm quan trọng của cuộc thi nên họ có thể thay đổi thói quen học tập để đạt được điểm tốt. Tình trạng này dĩ nhiên cũng dẫn đến bias một cách lạc quan.

Horse-racing bias

Hiện tượng “đua ngựa” này cũng hay xảy ra trong nghiên cứu y khoa. Đây là hiệu ứng được dùng để giải thích các số liệu của nghiên cứu quan sát. Trong điều kiện không có sai số đo lường, các yếu tố ảnh hưởng đến sự tăng trưởng nhanh trước khi tham gia PISA cũng là những người có xu hướng có năng lực tốt trong kì thi. Chúng ta biết rằng khi trường hay học sinh được chọn tham gia thì họ được chọn để huấn luyện và cho làm bài tập thử trước. Do đó, có thể đoán tại sao phương sai trong điểm PISA của Việt Nam nằm trong nhóm thấp nhất so với các nước khác. Đó cũng là một dạng bias làm lệch kết quả của kì thi.

Regression-toward-the-mean bias (Hồi qui về trung bình)

Đây là một hiện tượng được quan sát từ thế kỉ 18 bởi Francis Galton. Lúc đó, ông Galton làm nghiên cứu về ảnh hưởng của yếu tố di truyền đến trí thông minh, và ông lấy chiều cao làm chỉ số đo lường thông minh. Ông phát hiện rằng chiều cao của người con có xu hướng tương quan cao với chiều cao trung bình của cha và mẹ. Sau này, người ta cũng quan sát hiệu ứng này trong nhiều lĩnh vực khác như tâm lí học và giáo dục. Trong tâm lí học, Giáo sư Daniel Kahneman (giải Nobel kinh tế 2002) giải thích trong cuốn “Thinking Fast, Thinking Slow” rất hay rằng hiện tượng “Hồi qui về trung bình” có thể giải thích tại sao trừng phạt có hiệu quả tăng khả năng, nhưng khen thưởng có thể gây phản tác dụng. Trong giáo dục, những học sinh khởi đầu với điểm thấp thường có xu hướng tăng về hướng trung bình quần thể sau khi thi vài kì; ngược lại, học sinh có điểm khởi đầu cao lại có xu hướng giảm về điểm trung

binh quân thể. Nhưng đây là tác động chủ yếu do sai số trong đo lường, và nó có ảnh hưởng trực tiếp đến cách tính điểm trung bình cho một quần thể.

Đó chỉ là những bias tôi có thể nghĩ đến, chứ trong thực tế còn có nhiều bias khác mà chúng ta không kiểm soát được. Chẳng hạn như bias về văn hoá, phong tục, cách học, v.v. cũng có thể ảnh hưởng đến điểm PISA. Một số bias có thể điều chỉnh, nhưng một số bias thì không thể điều chỉnh, và chúng ta phải chấp nhận như là những điểm yếu. Đó là chưa kể đến những vấn đề nghiêm trọng về phương pháp thống kê mà đã có nhiều người chỉ ra trong quá khứ. Những bias này nói lên rằng chúng ta rất khó xếp hạng các nước dựa vào điểm PISA, và chúng ta cần phải cẩn thận trong diễn giải kết quả PISA.

Có lẽ Nhà nước không nên tốn tiền triệu USD để tham gia vào một chương trình kiểm định có nhiều vấn đề về phương pháp. Đã qua 2 kì kiểm định, và chúng ta đã có chút dữ liệu cho biết học trò Việt Nam đang ở đâu trên bản đồ học thế giới. Nếu tham gia lần nữa thì có lẽ kết quả cũng chẳng khác gì, nhất là hai lần kiểm định tốn kém đều cho ra kết quả gần như giống nhau. Chúng ta cần những nghiên cứu qui mô theo thời gian (longitudinal research) để đánh giá học sinh theo thời gian, chứ không phải những loại kiểm định mang tính cắt ngang như PISA. Nghiên cứu theo thời gian có giá trị hơn nghiên cứu cắt ngang, và điều này thì ai cũng biết.

Tôi nghĩ các quan chức giáo dục có lí do để theo đuổi PISA, vì kết quả kiểm định làm cho họ cảm thấy an lòng. Nó cũng giống như các đại học phương Tây chạy theo các bảng xếp hạng, vì họ làm thế là để củng cố cái ghế của hiệu trưởng và ban quản lí đại học. Nhưng nếu chúng ta muốn xây dựng một

nền giáo dục với ba đặc điểm dân tộc, nhân bản, khai phóng, thì về lâu dài chúng ta không có lí do phải tham gia vào PISA. Tôi không cần phải chi ra vài triệu USD để có vài con số hay một thứ hạng làm cho tôi cười hài lòng, trong khi hàng trăm ngàn con em phải đi du học ở nước có hạng thấp hơn mình.

Tài liệu tham khảo:

1. Sackett DL. Evidence based medicine: what it is and what it isn't. BMJ 1996;312:71-72.
2. Mulrow CD. Systematic Reviews: Rationale for systematic reviews. BMJ 1994;309:597-599.
3. The Heart Outcomes Prevention Evaluation Study Investigators. Effects of an Angiotensin-Converting-Enzyme Inhibitor, Ramipril, on Cardiovascular Events in High-Risk Patients. New Engl J Med 2000; 342:145-153

CỖ MẪU VÀ HUYỄN THOẠI CON SỐ 30

Bất cứ một công trình nghiên cứu nào cũng cần đến ước tính cỡ mẫu. Cỡ mẫu là số đối tượng nghiên cứu cần thiết để nhà nghiên cứu đạt được một kết quả đáng tin cậy và không vi phạm đạo đức khoa học. Chương này không bàn về kĩ thuật và phương pháp ước tính cỡ mẫu, mà bàn về ý nghĩa và lí thuyết đằng sau việc ước tính cỡ mẫu cho một nghiên cứu khoa học. Tôi cũng bàn về con số 30 thường được xem là một “huyền thoại” đã được lưu truyền quá lâu (nhưng không đúng hay bị hiểu lầm) ở Việt Nam và nhiều nơi trên thế giới.

Ước tính cỡ mẫu

Một công trình nghiên cứu thường dựa vào một mẫu (sample). Một trong những câu hỏi quan trọng nhất trước khi tiến hành nghiên cứu là cần bao nhiêu mẫu hay bao nhiêu đối tượng cho nghiên cứu. “Đối tượng” ở đây là đơn vị căn bản của một nghiên cứu, là số bệnh nhân hay số tình nguyện viên. Ước tính số lượng đối tượng cần thiết cho một công trình nghiên cứu đóng vai trò cực kì quan trọng, vì nó có thể là yếu tố quyết định sự thành công hay thất bại của nghiên cứu. Nếu số lượng đối tượng không đủ thì kết luận rút ra từ công trình nghiên cứu không có độ chính xác cao, thậm chí không thể kết luận gì được. Ngược lại, nếu số lượng đối tượng quá nhiều hơn số

cần thiết thì tài nguyên, tiền bạc và thời gian sẽ bị hao phí. Do đó, vấn đề then chốt trước khi nghiên cứu là phải ước tính cho được một số đối tượng vừa đủ cho mục tiêu của nghiên cứu.

- Số lượng đối tượng “vừa đủ” tùy thuộc vào loại hình nghiên cứu và 3 yếu tố chính:
- Mô hình nghiên cứu và tiêu chí lâm sàng (outcome measure).
- Hệ số ảnh hưởng (effect size);
- Sai lầm mà nhà nghiên cứu chấp nhận, cụ thể là sai lầm loại I và II (power);

Mô hình nghiên cứu

Thông tin thứ nhất trong qui trình ước tính cỡ mẫu là mô hình nghiên cứu. Thông thường, các nghiên cứu có thể chia thành X mô hình dựa vào thời gian tính:

- Các nghiên cứu thu thập dữ liệu tại một thời điểm hiện tại được gọi là cross-sectional study (nghiên cứu tiêu biểu một thời điểm hay nghiên cứu cắt ngang);
- Các nghiên cứu có định hướng theo dõi đối tượng trong một thời gian, tức thu thập dữ liệu từng đối tượng nhiều lần (hiện tại và tương lai) được gọi là prospective (hay longitudinal study) mà tôi gọi là nghiên cứu theo thời gian;
- Các nghiên cứu được tiến hành hiện tại nhưng có định hướng tìm hiểu quá khứ được gọi là retrospective study. Mô hình này chủ yếu là nghiên cứu bệnh chứng (case control study).

Nghiên cứu tại một thời điểm hay cross-sectional study là một thiết kế mà các nhà nghiên cứu chọn một quần thể một

cách ngẫu nhiên nhưng tiêu biểu cho một cộng đồng, tại một thời điểm nào đó. Nói cách khác, nhà nghiên cứu thu thập dữ liệu chỉ một lần duy nhất của các đối tượng ngay tại thời điểm đó (hiện tại). Mục đích chính của các nghiên cứu này là tìm hiểu tỉ lệ hiện hành (prevalence) của một hiện tượng, hay tìm hiểu mối tương quan giữa một yếu tố nguy cơ và một bệnh. “Hiện tượng” ở đây có thể hiểu là bệnh tật, ý kiến về một chính sách, bầu cử, v.v. Do đó, mô hình nghiên cứu này thường được sử dụng trong điều tra xã hội (survey).

Ví dụ: nhà nghiên cứu muốn biết tỉ lệ người trong thành phố mắc bệnh tiểu đường. Nhà nghiên cứu có thể lấy mẫu ngẫu nhiên từ cộng đồng khoảng 1000 người, và xác định ai mắc bệnh qua xét nghiệm đường huyết trong máu. Số người mắc bệnh chia cho tổng số mẫu 1000 chính là tỉ lệ hiện hành. Đó là một nghiên cứu cắt ngang tiêu biểu.

Nghiên cứu xuôi thời gian (longitudinal studies hay prospective study). Khác với nghiên cứu cắt ngang vốn chỉ ghi nhận sự kiện tại một thời điểm, các nghiên cứu theo thời gian phải theo dõi đối tượng trong một thời gian có thể là nhiều năm tháng. Mục đích của các nghiên cứu xuôi thời gian thường là ước tính tỉ lệ phát sinh (incidence) của một hiện tượng trong một thời gian (điều này khác với mục đích của nghiên cứu tại một thời điểm là ước tính tỉ lệ hiện hành của hiện tượng). Trong y học, các nghiên cứu theo thời gian còn cho phép nhà nghiên cứu tìm hiểu mối liên hệ giữa một hay nhiều yếu tố nguy cơ và nguy cơ phát sinh bệnh tật.

Ví dụ: nhà nghiên cứu muốn xác định tỉ lệ phát sinh người trong thành phố mắc bệnh tiểu đường. Nhà nghiên cứu có thể lấy mẫu ngẫu nhiên từ cộng đồng khoảng 1000 người, và xác định ai mắc bệnh qua xét nghiệm đường huyết trong máu; và

phát hiện 50 người bị tiểu đường. Họ (50 người) sẽ được loại ra khỏi nghiên cứu. Số người còn lại (950) sẽ được theo dõi 12 tháng; sau 12 tháng họ được đo đường huyết một lần nữa. Nhà nghiên cứu phát hiện có 10 người mới mắc bệnh tiểu đường trong 12 tháng qua. Con số tỉ lệ 10/950 chính là *incidence* hay tỉ lệ phát sinh.

Nghiên cứu bệnh chứng hay case-control study.

Ngược lại với nghiên cứu theo thời gian, nghiên cứu bệnh chứng đi ngược thời gian. Mô hình nghiên cứu bệnh chứng thường được áp dụng trong nghiên cứu y khoa. Mục đích của nghiên cứu bệnh chứng là đánh giá mối liên quan giữa một yếu tố nguy cơ và bệnh lí. Với mô hình nghiên cứu bệnh chứng, nhà nghiên cứu biết ai mắc bệnh và ai không mắc bệnh, nhưng họ không biết ai có yếu tố nguy cơ và do đó phải truy tìm trong quá khứ về sự hiện diện của yếu tố nguy cơ.

Ví dụ: nhà nghiên cứu muốn tìm hiểu mối liên quan giữa bệnh ung thư máu và hút thuốc lá. Bởi vì ung thư máu là bệnh rất hiếm, nên mô hình nghiên cứu theo thời gian sẽ khó có đủ số bệnh nhân. Thay vào đó, nhà nghiên cứu chọn mô hình nghiên cứu bệnh chứng. Theo mô hình này, nhà nghiên cứu chọn [ví dụ] 20 người có bệnh ung thư máu và 20 người không mắc bệnh (gọi là 'nhóm chứng'). Sau đó, nhà nghiên cứu sẽ hỏi mỗi cá nhân về thói quen hút thuốc lá trong quá khứ [trước khi mắc bệnh]. Nhà nghiên cứu sẽ so sánh tỉ lệ hút thuốc lá giữa nhóm bệnh và nhóm chứng, và nếu sự khác biệt có ý nghĩa thống kê, đó là tín hiệu về mối liên quan giữa hút thuốc lá và ung thư máu.

Biến outcome

Sau khi đã xác định mô hình nghiên cứu, nhà nghiên cứu cần phải quyết định chọn một tiêu chí đánh giá (tôi tạm gọi là

evaluative outcome hay *primary outcome measure*) để căn cứ vào đó mà ước tính cỡ mẫu. Quyết định chọn tiêu chí đánh giá là một quyết định vừa mang tính lâm sàng, vừa mang tính khoa học. Bởi vì mục tiêu tối hậu của nghiên cứu y khoa là đem lại lợi ích cho bệnh nhân hay cộng đồng, cho nên tiêu chí được chọn phải có ý nghĩa thực tế đối với bệnh nhân. Chẳng hạn như trong việc thẩm định hiệu quả của các phương pháp truy tìm ung thư, thì tỉ lệ phát hiện ung thư và điều trị không phải là tiêu chí có ý nghĩa thực tế, nhưng tỉ lệ tử vong và thời gian sống sót sau khi truy tìm ung thư mới là tiêu chí có ý nghĩa lâm sàng và thực tế. Mặc khác, tiêu chí phải đáp ứng các tiêu chuẩn khoa học về độ tin cậy và độ chính xác. Nếu một nghiên cứu có mục tiêu tìm hiểu hiệu quả của một loại thuốc phòng chống bệnh xơ vữa động mạch, thì độ cholesterol trong máu không thể được xem là tiêu chí có ý nghĩa lâm sàng, dù nó đáp ứng yêu cầu khoa học tính. Do đó, việc chọn một tiêu chí lâm sàng cho nghiên cứu cần phải cân nhắc rất cẩn thận.

Quyết định chọn tiêu chí đánh giá là một quyết định quan trọng, bởi vì nó có ảnh hưởng đến cỡ mẫu rất lớn. Chẳng hạn như trong các nghiên cứu loãng xương, các nhà nghiên cứu có thể so sánh mật độ xương hay tỉ lệ gãy xương giữa hai nhóm can thiệp để biết hiệu quả của thuốc. Nếu chọn mật độ xương làm tiêu chí lâm sàng thì số lượng cỡ mẫu có thể sẽ là con số vài trăm bệnh nhân, nhưng nếu chọn tỉ lệ gãy xương con số cỡ mẫu có thể lên đến vài chục ngàn đối tượng.

Sai sót loại I và II

Thống kê học là một phương pháp khoa học có mục đích phát hiện, hay đi tìm những cái có thể gộp chung lại bằng cụm từ “chưa được biết”. Cái chưa được biết ở đây là những hiện tượng chúng ta không quan sát được, hay quan sát được nhưng

không đầy đủ. Cái chưa biết có thể là một ẩn số (như chiều cao trung bình ở người Việt Nam, hay trọng lượng một phần tử), hiệu quả của một thuật điều trị, tỉ lệ lưu hành, tỉ lệ phát sinh của bệnh, v.v. Chúng ta có thể đo chiều cao, hay tiến hành xét nghiệm để biết hiệu quả của thuốc, nhưng các nghiên cứu như thế chỉ được tiến hành trên một nhóm đối tượng, chứ không phải toàn bộ quần thể của dân số. Vấn đề là sử dụng kết quả của một nhóm đối tượng để suy luận cho một quần thể lớn hơn. Mục đích của ước tính cỡ mẫu là tìm số lượng đối tượng sao cho suy luận đó đạt độ chính xác cao nhất và đầy đủ nhất.

Ở mức độ đơn giản nhất, những cái chưa biết này có thể xuất hiện dưới hai hình thức: hoặc là có, hoặc là không. Chẳng hạn như một thuật điều trị có hay không có hiệu quả chống gãy xương. Bởi vì không ai biết hiện tượng một cách đầy đủ, chúng ta phải đặt ra giả thiết. Giả thiết đơn giản nhất là *giả thiết đảo* (hiện tượng không tồn tại, kí hiệu Ho) và *giả thiết chính* (hiện tượng tồn tại, kí hiệu Ha).

Chúng ta sử dụng các phương pháp kiểm định thống kê (statistical test) như kiểm định t , F , z , χ^2 , v.v... để đánh giá khả năng của giả thiết. Kết quả của một kiểm định thống kê có thể đơn giản chia thành hai giá trị: hoặc là *có ý nghĩa thống kê* (statistical significance), hoặc là *không có ý nghĩa thống kê* (non-significance). Có ý nghĩa thống kê ở đây thường dựa vào trị số P: thông thường, nếu $P < 0,05$, chúng ta phát biểu kết quả có ý nghĩa thống kê; nếu $P > 0,05$ chúng ta nói kết quả không có ý nghĩa thống kê. Cũng có thể xem có ý nghĩa thống kê hay không có ý nghĩa thống kê như là có tín hiệu hay không có tín hiệu. Hãy tạm đặt kí hiệu T+ là kết quả có ý nghĩa thống kê, và T- là kết quả kiểm định không có ý nghĩa thống kê.

Hãy xem xét một ví dụ cụ thể: để biết thuốc risedronate có hiệu quả hay không trong việc điều trị loãng xương, chúng ta tiến hành một nghiên cứu gồm 2 nhóm bệnh nhân (một nhóm được điều trị bằng risedronate và một nhóm chỉ sử dụng giả dược placebo). Chúng ta theo dõi và thu thập số liệu gãy xương, ước tính tỉ lệ gãy xương cho từng nhóm, và so sánh hai tỉ lệ bằng một kiểm định thống kê. Kết quả kiểm định thống kê hoặc là có ý nghĩa thống kê ($P < 0,05$) hay không có ý nghĩa thống kê ($P > 0,05$). Xin nhắc lại rằng chúng ta không biết risedronate thật sự có hiệu nghiệm chống gãy xương hay không; chúng ta chỉ có thể đặt giả thiết. Do đó, khi xem xét một giả thiết và kết quả kiểm định thống kê, chúng ta có bốn tình huống:

- Giả thuyết H_0 đúng (thuốc risedronate có hiệu nghiệm) và kết quả kiểm định thống kê $P < 0,05$.
- Giả thuyết H_0 đúng, nhưng kết quả kiểm định thống kê không có ý nghĩa thống kê;
- Giả thuyết H_0 sai (thuốc risedronate không có hiệu nghiệm) nhưng kết quả kiểm định thống kê có ý nghĩa thống kê;
- Giả thuyết H_0 sai và kết quả kiểm định thống kê không có ý nghĩa thống kê.

Ở đây, trường hợp (a) và (d) không có vấn đề, vì kết quả kiểm định thống kê nhất quán với thực tế của hiện tượng. Nhưng trong trường hợp (b) và (c), chúng ta phạm sai lầm, vì kết quả kiểm định thống kê không phù hợp với giả thiết. Trong ngôn ngữ thống kê học, chúng ta có vài thuật ngữ:

- Xác suất của tình huống (b) xảy ra được gọi là *sai lầm loại II* (type II error), và thường kí hiệu bằng β .

- Xác suất của tình huống (a) được gọi là *Power*. Nói cách khác, *power* chính là xác suất mà kết quả kiểm định thống kê cho ra kết quả $P < 0,05$ với điều kiện giả thiết H_a là thật. Nói cách khác: $power = 1 - \beta$;
- Xác suất của tình huống (c) được gọi là *sai lầm loại I* (type I error, hay significance level), và thường kí hiệu bằng α . Nói cách khác, α chính là xác suất mà kết quả kiểm định thống kê cho ra kết quả $P < 0,05$ với điều kiện giả thiết H_a sai;
- Xác suất tình huống (d) không phải là vấn đề cần quan tâm, nên không có thuật ngữ, dù có thể gọi đó là kết quả âm tính thật (hay true negative).

Có thể tóm lược 4 tình huống đó trong bảng sau đây:

**Các tình huống trong việc thử nghiệm
một giả thiết khoa học**

Kết quả kiểm định thống kê	Giả thuyết H_a	
	Đúng (thuốc có hiệu nghiệm)	Sai (thuốc không có hiệu nghiệm)
Có ý nghĩa thống kê ($P < 0,05$)	Dương tính thật (power) $1 - \beta = P(S H_a)$	Sai lầm loại I (type I error) $\alpha = P(S H_o)$
Không có ý nghĩa thống kê ($P > 0,05$)	Sai lầm loại II (type II error) $\beta = P(NS H_a)$	Âm tính thật (true negative) $1 - \alpha = P(NS H_o)$

Chú thích: kí hiệu S trong bảng này có nghĩa là "significant" (tức $P < 0,05$); NS = "non-significant" (tức $P > 0,05$). Do đó, có thể mô tả 4 tình huống trên bằng ngôn ngữ xác suất có điều kiện như sau: $Power = 1 - \beta = P(S | H_a)$; $\beta = P(NS | H_a)$; và $\alpha = P(S | H_o)$. Xin nhắc lại rằng kí hiệu toán học " $P(A | B)$ " có nghĩa là một xác suất có điều kiện, cụ thể hơn kí hiệu $P(S | H_a)$ có nghĩa là "xác suất S xảy ra nếu (hay với điều kiện) H_a là đúng."

Cũng như các phương pháp xét nghiệm y khoa không bao giờ hoàn hảo, các phương pháp kiểm định thống kê cũng có sai sót. Và do đó, kết quả nghiên cứu lúc nào cũng có độ bất định (như sự bất định trong một chẩn đoán y khoa vậy). Vấn đề là chúng ta phải thiết kế nghiên cứu sao cho *sai sót* loại I và II thấp nhất. Thông thường, sai sót loại I là 5% đến 1%, còn sai sót loại II dao động từ 20% đến 10%.

Sau khi đã xác định được mô hình nghiên cứu, tiêu chí outcome, và sai sót I và II, việc ước tính cỡ mẫu chỉ là áp dụng các công thức thống kê. Chương này không bàn về các công thức đó, vì các bạn có thể tham khảo trong một sách chuyên sâu khác, chẳng hạn như cuốn “Phân tích dữ liệu với R: Hỏi và Đáp” (Nhà xuất bản Tổng hợp Thành phố Hồ Chí Minh, 2017).

Về con số 30

Bàn đến vấn đề ước tính cỡ mẫu không thể bỏ qua huyền thoại con số 30. Nhiều người hay cho rằng nghiên cứu phải có cỡ mẫu trên 20 hay trên 30 mới có ý nghĩa. Nhưng tôi e rằng đây là một sự hiểu lầm. Theo tôi biết, cũng chẳng có con số huyền thoại 30 hay 20 nào cả trong thống kê học. Số 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 vài 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.

Biến đánh giá (còn gọi là *evaluative outcome*). Mỗi nghiên cứu đều có một biến 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 biến đánh giá có thể là mật độ xương hay tình trạng gãy xương. Có hai loại biến đánh giá: biến liên tục (continuous variable) và biến phân loại (categorical variable). Trong ví dụ vừa kể, mật độ xương là biến liên tục, còn gãy xương là biến phân loại (hoặc là bị gãy, hoặc là không gãy). Phương pháp ước tính cỡ mẫu cho biến liên tục rất khác với phương pháp cho các biến 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 tham 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 $0,05 \text{ g/cm}^2$, với độ lệch chuẩn là $0,12 \text{ g/cm}^2$. Ngoài ra, nhà nghiên cứu sẵn sàng chấp nhận sai sót loại I là 5% và sai sót loại II là 20%. Với các tham 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 $0,15 \text{ g/cm}^2$ (thay vì $0,05 \text{ g/cm}^2$), 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 cỡ mẫu cho nghiên cứu khoa học.

Có thể một hiểu lầm khác là liên quan đến Định lý giới hạn trung tâm (Central Limit Theorem - CLT). Một cách ngắn gọn, CLT phát biểu rằng bất cứ chỉ số thống kê nào (như trung bình, phương sai, trung vị) đều tuân theo luật phân bố chuẩn hay gần với phân bố chuẩn nếu cỡ mẫu đủ. “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.

Tôi sẽ lấy một ví dụ để minh họa cho định lý CLT trên. Chúng ta biết rằng thu nhập của người dân có phân bố lệch, tức không tuân theo luật phân bố chuẩn, vì nhiều người có thu nhập thấp và rất ít người có thu nhập cao. Giả dụ như chúng ta lấy mẫu từ Việt Nam theo qui trình (tưởng tượng) như sau:

- (a) lấy mẫu lần thứ nhất, chọn ngẫu nhiên n người; tính thu nhập trung bình;
- (b) lấy mẫu lần 2, y như lần 1, nhưng dĩ nhiên lần này chúng ta có n người khác; chúng ta tính thu nhập trung bình;
- (c) lặp lại k lần như thế (k có thể là vài ngàn đến vài triệu lần), chúng ta sẽ có k số trung bình.

Định lý CLT nói rằng k số trung bình đó - mỗi số trung bình tính từ n đối tượng - sẽ tuân theo luật phân bố chuẩn khi $n > 30$. Nếu n dưới 30 thì có thể k số trung bình không phù hợp với luật phân bố chuẩn. Ví dụ trên giải thích rằng $n > 30$ chẳng có liên quan gì đến cỡ mẫu để nghiên cứu khoa học cả, mà nó liên quan để phân bố lý thuyết của một chỉ số thống kê (trong trường hợp này là số trung bình).

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.

NGỘ NHẬN VỀ TÍNH ĐẠI DIỆN TRONG NGHIÊN CỨU KHOA HỌC

Hầu hết những ai làm nghiên cứu lâm sàng đều từng trải qua các hội đồng xét duyệt đề tài hay nghiệm thu, và tuyệt đại đa số đều nhận được lời phê bình là nghiên cứu không mang tính đại diện. Nhận xét này hàm ý nói kết quả nghiên cứu không có giá trị khoa học, và có khi nó được dùng như là một cách bác bỏ nỗ lực của nhà nghiên cứu. Nhưng trong thực tế, nhiều nghiên cứu không cần tính đại diện. Trong chương này tôi sẽ giải thích tại sao một nghiên cứu y học nghiêm chỉnh không nhất thiết phải có mẫu mang tính đại diện.

Những cách hiểu về khái niệm “đại diện”

Một trong những nguyên lý cổ điển về nghiên cứu y học được giảng dạy cho sinh viên là mẫu nghiên cứu phải mang tính đại diện. Nói theo tiếng Anh là *representativeness*. Theo nguyên lý này, mẫu mang tính đại diện thì kết quả nghiên cứu mới có giá trị khoa học, dù nhiều người không định nghĩa rõ “giá trị khoa học” có nghĩa gì trong thực tế. Nhưng từ suy nghĩ đó, đối với những người đã qua đào tạo hay chịu ảnh hưởng bởi nguyên lý cổ điển đó đòi hỏi nghiên cứu phải mang tính “đại diện” cho quần thể mà mẫu được lấy từ đó. Nhưng khái niệm “đại diện” được hiểu rất khác nhau giữa các nhà khoa

học, và có nhiều hiểu lầm. Bốn cách hiểu phổ biến về tính đại diện là như sau:

1. Đại diện mang ý nghĩa thống kê

Một nghiên cứu y học thường lấy mẫu (sample) từ một quần thể (population). Từ một quần thể, chúng ta có thể lấy rất nhiều mẫu khác nhau, và dĩ nhiên mỗi lần lấy mẫu thì mỗi lần chúng ta có kết quả hơi khác. Nếu quần thể gồm 100 người có chiều cao 160 cm và độ lệch chuẩn 6,5 cm, thì nếu chúng ta lấy mẫu ngẫu nhiên 10 người từ quần thể đó thì chiều cao trung bình của 10 người đó có thể dao động trong khoảng 157 đến 162 cm. Ở đây, tính đại diện có nghĩa là phân bố chiều cao của mẫu nghiên cứu giống như phân bố chiều cao của quần thể.

2. Đại diện về một đặc điểm đối tượng

Đối với nhiều người, cách hiểu về tính đại diện có vẻ đơn giản hơn, và đó là kì vọng rằng mẫu nghiên cứu có những đặc điểm giống như đặc điểm của quần thể. Theo cách hiểu này, nếu quần thể có 51% dân số là phụ nữ, 30% nam giới hút thuốc lá, 15% người có thu nhập cao, 1% người ở tuổi 20, v.v. thì mẫu nghiên cứu cũng phải có những đặc điểm tương tự.

3. Đại diện về kết quả

Một cách hiểu khác là đại diện tính có nghĩa là kết quả, ví dụ như tỉ lệ mắc bệnh, trong mẫu nghiên cứu gần bằng với tỉ lệ của quần thể. Nói cách khác, nếu trong quần thể có 10% người mắc bệnh tiểu đường, thì mẫu nghiên cứu cũng có khoảng 10% người mắc bệnh tiểu đường. Nếu mẫu nghiên cứu có 20% người mắc bệnh thì được xem là không mang tính đại diện.

4. Đại diện về phân bố địa lí

Một cách hiểu khác mà nhiều người khi đề cập đến tính đại diện họ hàm ý nói đến tính địa lí và đặc điểm của đối tượng nghiên cứu. Rất nhiều người nghĩ rằng một nghiên cứu y học, đặc biệt là dịch tễ học, phải mang tính đại diện địa lí, có nghĩa là chọn ngẫu nhiên trong cộng đồng. Họ nghĩ rằng để đảm bảo tính đại diện của mẫu nghiên cứu, phải chọn mỗi tỉnh thành, chọn ngẫu nhiên vài quận/huyện, mỗi quận/huyện chọn ngẫu nhiên vài phường/xã, mỗi phường/xã chọn ngẫu nhiên vài ấp, v.v. Nói cách khác, nếu Quận I có dân số chiếm 10% dân số của Thành phố, thì mẫu nghiên cứu cũng nên có khoảng 10% cá nhân đến từ Quận I.

Nhưng cả 4 cách hiểu về đại diện trên là 4 ngộ nhận về nghiên cứu khoa học. Cách hiểu thứ nhất là một sự hiểu lầm giữa đại diện mang tính thống kê và đại diện mang tính khoa học, và tôi sẽ bàn thêm dưới đây. Một cách ngắn gọn trong suy luận khoa học, chúng ta không nhất thiết phải cần đến đại diện thống kê.

Cách hiểu thứ hai và thứ ba là rất phản khoa học, bởi vì để kiểm định một giả thuyết khoa học nhà khoa học không cần chọn đối tượng mang tính đại diện, mà có thể chọn một nhóm đối tượng với những đặc điểm đặc biệt. Chẳng hạn như để kiểm định giả thuyết về ảnh hưởng của thiếu vận động đến loãng xương, chúng ta có thể chọn bệnh nhân đột quỵ và những phi hành gia.

Cách hiểu thứ tư là một hiểu lầm về điều tra dân số, lấy ý kiến quần chúng và nghiên cứu khoa học. Khi lấy ý kiến trong công chúng, người ta có thể [chỉ “có thể” thôi] cần đến mẫu đại diện, nhưng nghiên cứu khoa học không phải là điều tra

xã hội. Ngay cả điều tra xã hội, kết quả là một tỉ lệ (ví dụ như) ủng hộ ứng viên nào đó cũng có thể là... vô nghĩa. Cần phải nói thêm rằng nếu phải lấy mẫu đại diện địa lí thì có thể nói rằng 100% các nghiên cứu y học đều vô giá trị! Nếu phải đáp ứng đại diện địa lí thì tất cả những nghiên cứu nổi tiếng như Framingham, Rotterdam (Hà Lan), CaMOS (Canada), Dubbo Osteoporosis Epidemiology Study (của tôi), tất cả các công trình nghiên cứu lâm sàng ngẫu nhiên (RCT), v.v. đều không có giá trị khoa học?! Do đó, những ngộ nhận này tuy đáng tiếc, nhưng lại rất nguy hiểm, vì làm chùn bước rất nhiều người muốn làm nghiên cứu khoa học có ý nghĩa. Những hiểu lầm này có thể xuất phát từ sự ngộ nhận về suy luận thống kê và suy luận khoa học mà tôi giải thích dưới đây.

Suy luận thống kê và suy luận khoa học

Để hiểu tại sao khái niệm đại diện không cần thiết trong nghiên cứu khoa học, chúng ta cần phải biết mục tiêu thật sự của một nghiên cứu khoa học. Ở đây, tôi không bàn đến những cuộc điều tra xã hội để biết ý kiến của công chúng về một vấn đề nào đó (ví dụ như bầu cử) hay để biết tỉ lệ mắc bệnh trong quần thể, vì đó không phải là nghiên cứu khoa học. Thường, các nghiên cứu khoa học có mục tiêu chính là kiểm định một giả thuyết khoa học. Giả thuyết khoa học thường mang tính tiên lượng, giống như qui luật, trong thế giới tự nhiên. Do đó, mục tiêu chính của khoa học là đi tìm một qui luật của thế giới tự nhiên (law of nature).

Qui luật này có thể thể hiện qua hiện tượng chúng ta gọi là **liên quan** (1). Ý tưởng “liên quan” ở đây có thể là mối liên hệ nhân quả (causal relationship) hay mối tương quan thống kê (correlation). Dĩ nhiên, tương quan thống kê chỉ là điều kiện

cần, chứ chưa đủ, để nói lên mối liên hệ nhân quả. Chẳng hạn như để kiểm định giả thuyết thiếu vận động thể lực dẫn đến tăng nguy cơ loãng xương, nhà nghiên cứu có thể đặt mục tiêu nghiên cứu là tìm mối liên quan giữa vận động thể lực và loãng xương. Vận động thể lực có thể được đo bằng chỉ số tiêu hao năng lượng (biến liên tục), nhưng cũng có thể so sánh giữa nhóm bất động và nhóm bình thường. Đó là nghiên cứu tìm mối liên quan.

Khía cạnh quan trọng nhất trong nghiên cứu khoa học là kết quả (qui luật quan sát trong nghiên cứu) phù hợp với thực tế của **qui luật tự nhiên**. Chẳng hạn như trong nghiên cứu về mối liên quan giữa vận động thể lực và loãng xương, chúng ta có thể thiết kế nghiên cứu bằng cách so sánh mật độ xương của các phi hành gia và người bình thường, và nếu kết quả cho thấy phi hành gia có mật độ xương thấp hơn người bình thường, thì đó là một qui luật, một mối liên quan. Mối liên quan đó đúng hay sai không phải tùy thuộc vào tính đại diện của đối tượng nghiên cứu, mà phụ thuộc vào kiến thức sinh học về xương và những nghiên cứu trước đây trong y văn.

Từ những cân nhắc trên, chúng ta có thể phân biệt khái niệm suy luận thống kê (**statistical inference**) và suy luận khoa học (**scientific inference**). Suy luận thống kê dựa trên kết quả phân tích trong một mẫu nghiên cứu để phát biểu về quần thể. Những kết quả này bao gồm các chỉ số kiểm định thống kê (như t-test, Ki bình phương, mô hình hồi qui tuyến tính) kèm theo trị số P và khoảng tin cậy 95%. Suy luận khoa học liên quan đến khái quát hoá mối liên quan quan sát từ nghiên cứu đến thế giới tự nhiên. Suy luận khoa học dựa trên kết quả suy luận thống kê, nhưng đó chỉ là điều kiện cần, chứ chưa đủ, vì ngoài thống kê ra, chúng ta cần có kiến thức chuyên ngành trước khi phát biểu một suy luận khoa học.

Hợp lí nội tại và hợp lí ngoại tại

Khi nói một nghiên cứu khoa học có giá trị, người ta nghĩ đến hai tính hợp lí: hợp lí nội tại (internal validity) và hợp lí ngoại tại (external validity). Hiểu một cách đơn giản, hợp lí ngoại tại là đảm bảo “chất lượng” (quality assurance) của nghiên cứu. Trong thực tế, **hợp lí nội tại** có nghĩa là chọn phương pháp đo lường chính xác và có độ tin cậy cao, là sử dụng phương pháp phân tích tốt và thích hợp, là kiểm soát và giảm thiểu tối đa các yếu tố nhiễu, để kết quả nghiên cứu có tính hợp lí. Chẳng hạn như nghiên cứu về loãng xương mà dùng phương pháp siêu âm thì có nghĩa là tính hợp lí nội tại chưa cao.

Hợp lí ngoại tại (còn gọi là khái quát hoá) là mức độ mà kết quả của nghiên cứu có thể áp dụng, hoặc liên quan, hoặc khái quát hoá cho một quần thể lớn hơn hoặc cho những người không tham gia vào công trình nghiên cứu. Khái quát hoá theo cách hiểu này không phải chỉ đơn thuần là dựa vào thống kê (mẫu đại diện) mà *quan trọng hơn* là còn phải dựa vào kiến thức chuyên ngành và các nghiên cứu trong quá khứ.

Trong suy luận khoa học, tính hợp lí nội tại đóng vai trò quan trọng nhất, còn hợp lí ngoại tại chỉ là phụ. Nếu một nghiên cứu mà phương pháp đo lường không chuẩn mực thì cũng chẳng có giá trị gì về hợp lí ngoại tại. Chẳng hạn như nghiên cứu về béo phì mà chỉ đơn thuần dựa vào các chỉ số như chiều cao và cân nặng (hay body mass index - BMI) thì giá trị khoa học và tính hợp lí nội tại không cao (vì phương pháp chuẩn là đo lường tỉ trọng mỡ trong cơ thể). Do đó, xin nhắc lại: nếu một nghiên cứu thiếu tính hợp lí nội tại thì kết quả cũng không có ý nghĩa hợp lí ngoại tại. Đảm bảo tính hợp lí nội tại là rất quan trọng trong bất cứ công trình nghiên cứu y học nào.

Tại sao nghiên cứu y học không cần mang tính đại diện?

Nghiên cứu khoa học không cần mẫu nghiên cứu phải mang tính đại diện, bởi vì mục tiêu chính là suy luận khoa học. Để đánh giá sự ảnh hưởng của yếu tố di truyền đến một bệnh lí, chúng ta có thể nghiên cứu những thành viên trong gia đình hay những cặp sinh đôi, và cách chọn hoàn toàn không cần phải mang tính đại diện địa lí hay bất cứ chỉ số lâm sàng nào. Để đánh giá hiệu quả của một loại thuốc điều trị bệnh loãng xương, chúng ta có thể thực hiện một nghiên cứu lâm sàng đối chứng ngẫu nhiên (RCT), và bệnh nhân được tuyển chọn theo hàng chục tiêu chuẩn thu nhận và tiêu chuẩn loại trừ. Nghiên cứu RCT không cần đến tính đại diện của mẫu nghiên cứu. Ngay cả công trình vĩ đại UKBiobank với mục tiêu tư vấn cho chính sách y tế ở Anh, với gần nửa triệu người tham gia, hoàn toàn không mang tính đại diện. Những nghiên cứu như thế đòi hỏi tính hợp lí nội tại rất cao.

Cần nói thêm rằng đối với các nghiên cứu khoa học thì các phát hiện phải mang tính tái lập (repeatability). Tính tái lập cực kì quan trọng, vì nó phân biệt giữa khoa học thật và khoa học dỏm. Tái lập không chỉ quần thể này mà còn ở quần thể khác, và thường không tùy thuộc vào thời gian. Khi chúng ta phát hiện người hút thuốc lá có nguy cơ tử vong cao, thì mối liên quan này gần như được lặp lại ở bất cứ quần thể lớn nào và ở thời điểm nào cũng thế. Đó chính là tính đại diện. Và, tính đại diện này không phải do chọn đối tượng nghiên cứu mang tính đại diện theo nghĩa xác suất, càng không có gì liên quan đến đại diện địa lí.

Tính đại diện khoa học hay tái lập, một cách lí tưởng, không tùy thuộc vào thời gian và không gian. Trong y học, chúng ta đã thấy mối liên quan giữa cholesterol và bệnh tim

mạch xuất phát từ một nghiên cứu không mang tính đại diện địa lí bên Mỹ, nhưng kết quả đó áp dụng cho toàn thế giới. Một nghiên cứu về ảnh hưởng của di truyền đến xương ở Việt Nam cũng được lặp lại ở tất cả các nhóm khác trên thế giới. Nếu giả dụ như mẫu nghiên cứu “đại diện” theo cách hiểu đó thì mối liên quan mà chúng ta quan sát được có lẽ không áp dụng cho bất cứ nhóm nào, nam hay nữ, người giàu hay người nghèo. Cái mối liên quan giữa x và y , ví dụ như thể hiện qua tỉ số nguy cơ, đã được trọng số hoá và trung bình hoá từ nhiều nhóm nhỏ, nên nó không thể khái quát hoá cho nhóm nào cả. Sự thật này nói lên một lần nữa, trong nghiên cứu kiểm định giả thuyết khoa học, chúng ta không cần chọn mẫu đại diện.

Nhưng có trường hợp tính đại diện có thể cần thiết. Trưng cầu dân ý, điều tra xã hội để tìm hiểu qui mô của một bệnh (như tiểu đường) hay thói quen (như hút thuốc lá, uống rượu bia) rất khó có thể xem là nghiên cứu, vì không có giả thuyết đằng sau. Ở đây, chúng ta không nói những điều tra như thế là không quan trọng; thật ra, kết quả của các điều tra đó rất quan trọng cho việc hoạch định chính sách. Những dạng điều tra xã hội này có thể đòi hỏi phải có mẫu đại diện. Xin nhắc lại: “có thể” mà thôi, chứ không phải đúng.

Tuy nhiên, nếu mẫu điều tra mang tính đại diện theo nghĩa thứ hai (tức các đặc điểm của mẫu nghiên cứu giống như đặc điểm của quần thể) thì kết quả của một nghiên cứu như thế sẽ không có tính hợp lí ngoại tại. Chẳng hạn như nếu một mẫu điều tra đại diện về cấu trúc độ tuổi trong dân số, và phát hiện tỉ lệ mắc bệnh là 10%, thì con số này không đại diện cho cái gì cả. Tại sao? Tại vì con số 10% là số trung bình của nhiều độ tuổi, và tỉ lệ này khác biệt giữa các độ tuổi. Ở đây, chúng ta thấy cái gọi là tính “đại diện” là rất vô nghĩa!

Một ảo tưởng (và cũng là sai lầm) khác là cho rằng nếu mẫu nghiên cứu đại diện cho quần thể về vài đặc điểm nào đó thì kết quả hoàn toàn đại diện cho quần thể. Những kết quả điều tra như thế (tỉ lệ mắc bệnh) chỉ giá trị trong vòng một thời gian nhất định, và phụ thuộc vào không gian. Nói cách khác, nghiên cứu tìm hiểu tỉ lệ của một bệnh trong cộng đồng cũng chỉ có giá trị ngắn hạn ngay cả được thực hiện trong một quần thể cụ thể, và không thể khái quát hoá thậm chí cho quần thể khác. Do đó, tính đại diện ở đây cũng rất vô nghĩa.

Khái niệm đại diện theo tỉ lệ càng vô lí. Theo khái niệm và cách hiểu này, nếu quần thể có 5% người tuổi 70-79 thì mẫu cũng nên có 10% người trong độ tuổi đó. Đây là một hiểu lầm tai hại. Một nghiên cứu nghiêm chỉnh với cỡ mẫu gần bằng nhau cho từng nhóm tuổi có độ nhạy cao hơn nghiên cứu đảm bảo tỉ lệ của mẫu nghiên cứu và quần thể bằng nhau. Thử tưởng tượng, nếu chúng ta có 3 nhóm tuổi 50-59, 60-69, và 70-79, nghiên cứu với cỡ mẫu bằng nhau (ví dụ như 100, 100, 100 cho 3 nhóm tuổi) có độ nhạy cao hơn là theo tỉ lệ (mà kết quả có thể là 100, 60, 40). Do đó, tính đại diện cho mỗi nhóm là rất phản tác dụng.

Nhưng tính đại diện bị hiểu lầm khá nhiều bởi những người thiếu kinh nghiệm về nghiên cứu khoa học và các nhà dịch tễ học hiểu chưa đến vấn đề. Ở nước ngoài, thỉnh thoảng cũng có những phê bình sai về tính đại diện, vì người phát biểu hiểu sai. Tuy nhiên, họ không dám bác kết quả nghiên cứu, có lẽ một phần là do kinh nghiệm thực tế nên họ cũng dè dặt trong phê phán. Nhưng ở Việt Nam thì xuất phát từ ngộ nhận và hiểu sai về khái niệm đại diện nên những phê phán cũng sai và gây tác hại cho không ít nhà nghiên cứu chân chính.

Khái niệm “đại diện” theo cách hiểu 1-4 như trên đã được bàn luận trong y văn từ hơn 30 năm nay. Mới đây, một số nhà dịch tễ học lừng danh trên thế giới tranh luận trên tập san *International Journal of Epidemiology* (2-7), và ngoại trừ 1 người phản đối yếu ớt, tất cả các nhà dịch tễ học còn lại đều đồng ý rằng tính đại diện không cần cho nghiên cứu y học nhằm kiểm định một giả thuyết (8) liên quan đến một qui luật tự nhiên. Bất cứ ai làm nghiên cứu khoa học với giả thuyết khoa học đều chẳng ngạc nhiên về sự đồng ý đó, bởi vì tính đại diện theo cách hiểu 1-4 là không cần thiết và có thể gây tác hại cho sự tiến bộ của y học.

Chú thích:

- (1) Cần nói thêm rằng chữ “liên quan” ở đây là tương đương với thuật ngữ “*association*” trong tiếng Anh. Thuật ngữ *association* trong nghiên cứu khoa học dùng để chỉ những hiện tượng thống kê học như *correlation* (tương quan), *relationship* (liên hệ thường là nhân quả), *linkage* (trong di truyền học), và khác biệt (*difference*). Một mối liên quan có thể là nhân quả hay không nhân quả. Khi tỉ lệ mắc bệnh ở nam và nữ khác nhau, chúng ta cũng có thể nói có “liên quan” giữa giới tính và nguy cơ mắc bệnh. Các chỉ số thống kê như tỉ số odds, tỉ số nguy cơ, tất cả đều phản ánh mối liên quan. Do đó, liên quan dùng ở đây là một cách nói cấp cao hay “high level language” (chứ không phải chỉ đơn thuần theo nghĩa thống kê).
- 2 Rothman KJ, Gallacher JEJ, Hatch E. Why representativeness should be avoided. *Int J Epidemiol* 2013;42:1012-4.
- 3 Elwood M. Commentary: On representativeness. *Int J Epidemiol* 2013;42:1014-5.
- 4 Nohr EA, Olsen J. Commentary: Epidemiologists have debated representativeness for more than 40 years – has the time come to move one? *Int J Epidemiol* 2013;42:1016-7.

- 5 Richiardi L, Pizzi C, Pearce N. Commentary: Representativeness is usually not necessary and should be avoided. *Int J Epidemiol* 2013;42:1018-22.
- 6 Ebrahim S, Smith GD. Commentary: Should we always deliberately be non-representative? *Int J Epidemiol* 2013;42:1022-6.
- 7 Rothman KJ, Gallacher JEJ, Hatch E. Rebuttal: When it comes to scientific inference, sometimes a cigar is just a cigar. *Int J Epidemiol* 2013;42:1026-8.
- 8 Trích dẫn vài bình luận của các chuyên gia:
 - “Representativeness is time and place specific, and will therefore be a historical concept.”
 - “We apply knowledge of genetics from fruit flies to humans because the biological relationships are generalizable although the individuals studied are not.”
 - “Representativeness is not desirable even if the goal is to study effect-measure modification.”

CÁCH THIẾT KẾ BỘ CÂU HỎI

Bộ câu hỏi (hay questionnaire) là một công cụ rất quan trọng trong các nghiên cứu xã hội và dịch tễ học. Bộ câu hỏi được thiết kế tốt sẽ giúp nhà nghiên cứu thu thập dữ liệu có ích và giúp cho việc phân tích có ý nghĩa hơn. Rất dễ để viết ra một câu hỏi, nhưng rất khó để viết ra một câu hỏi tốt và có ý nghĩa. Trong chương này, tôi bàn về những nguyên tắc trong cách thiết kế một bộ câu hỏi có ý nghĩa.

Bộ câu hỏi là một công cụ để thu thập dữ liệu. Những câu trả lời trong bộ câu hỏi được xem là dữ liệu thật mà đối tượng nghiên cứu cung cấp. Do đó, một bộ câu hỏi nếu được thiết kế tốt sẽ cho ra những dữ liệu chính xác và có liên quan đến nghiên cứu. Ngoài ra, bộ câu hỏi còn cho phép nhà nghiên cứu tìm hiểu những tâm sự, tâm tư của đối tượng nghiên cứu mà các phương tiện đo lường sinh hóa không thể nào đo được. Điều này cũng có nghĩa là bộ câu hỏi là một công cụ bổ sung, làm cho công trình nghiên cứu thêm phong phú và đa dạng dữ liệu.

Sử dụng bộ câu hỏi để thu thập dữ liệu có nhiều lợi thế nhưng cũng có vài điểm yếu. Lợi thế của bộ câu hỏi là nó giúp nhà nghiên cứu có thể thu thập dữ liệu từ nhiều đối tượng, chứ không hạn chế như các phương tiện máy móc. Bộ câu hỏi cung cấp cho nhà nghiên cứu những câu trả lời

rõ ràng và đơn giản (nếu bộ câu hỏi được chuẩn hoá tốt). Những dữ liệu từ bộ câu hỏi có thể giúp cho nhà nghiên cứu biết sự việc gì đã xảy ra và xảy ra như thế nào. Nhưng bộ câu hỏi cũng có vài giới hạn, và giới hạn lớn nhất là nó chỉ cung cấp những thông tin mang tính bề mặt của sự việc. Ngoài ra, bộ câu hỏi có khi không phản ánh cái bối cảnh của nghiên cứu, không nói lên được nguyên nhân và hệ quả. Cũng cần ghi nhận thêm rằng tất cả các câu trả lời của đối tượng nghiên cứu là “tự báo cáo” (self-report), và do đó độ chính xác có thể không cao.

Việc thu thập dữ liệu bằng bộ câu hỏi có thể thực hiện qua một trong 4 hình thức: qua bưu điện, website, điện thoại, hay phỏng vấn trực tiếp. Ngày xưa (và ngay cả ngày nay), bộ câu hỏi thường được gửi qua đường bưu điện đến đối tượng nghiên cứu và thỉnh cầu họ điền vào bộ câu hỏi, và gửi trả lại cho nhà nghiên cứu. Cách làm này thuận tiện, chi phí thấp, và có thể làm trên rất nhiều đối tượng, nhưng có nhiều hạn chế như các đối tượng không gửi lại bộ câu hỏi hay họ trả lời không trung thực. Phương thức thứ hai là dùng website để thu thập dữ liệu, và đây cũng là một cách có thể thu thập dữ liệu từ rất nhiều đối tượng. Tuy nhiên, thu thập dữ liệu qua website có nhiều hạn chế về tính hợp lý cũng như *bias* trong nghiên cứu. Nhiều nhóm nghiên cứu dùng điện thoại để phỏng vấn đối tượng nghiên cứu, nhưng tỉ lệ đối tượng nghiên cứu đồng ý trả lời thường không cao, và do đó có thể dẫn đến những kết quả không chính xác. Cách thứ tư là dùng bộ câu hỏi để phỏng vấn đối tượng nghiên cứu. Cách phỏng vấn trực tiếp này có lợi thế là thu thập dữ liệu tương đối chính xác, có độ tương tác cao, nhưng phương pháp này khá tốn kém và khó có thể thực hiện trên bình diện rộng lớn. Nói tóm lại, không

có phương pháp nào là hoàn chỉnh trong việc thu thập dữ liệu cả, vì phương pháp nào cũng có ưu điểm và hạn chế. Việc lựa chọn một phương pháp phù hợp phải là quyết định cân bằng giữa khoa học, cỡ mẫu, và chi phí.

Để soạn bộ câu hỏi cho tốt, nhà nghiên cứu cần phải tìm hiểu và làm nghiên cứu sơ khởi. Nhà nghiên cứu cần phải đọc và tham khảo các công trình nghiên cứu trước đây trên các tạp san khoa học. Tham khảo về các chủ đề họ quan tâm, những dữ liệu cần phải thu thập, và cách họ phân tích các dữ liệu đó. Ngoài ra, nhà nghiên cứu cũng cần phải làm một nghiên cứu sơ khởi bằng cách phỏng vấn một số đối tượng, và qua đó thu thập những dữ liệu mà họ quan tâm. Những dữ liệu này có thể đúc kết thành những chủ đề cho nghiên cứu và cho việc thiết kế bộ câu hỏi.

1. Thể loại câu hỏi

Một bộ câu hỏi thường thu thập dữ liệu liên quan đến 4 yếu tố sau đây: *kiến thức; thái độ, niềm tin; hành vi; và đặc điểm (attributes)*. Kiến thức là *những gì đối tượng hiểu biết về vấn đề nằm trong tầm của công trình nghiên cứu*. Ví dụ như câu hỏi “Theo bạn thời gian cần thiết để tái khám bệnh tiểu đường là bao lâu?” là loại câu hỏi về kiến thức, và câu trả lời có thể là:

- ☐ 6 tháng
- ☐ 1 năm
- ☐ 2 năm
- ☐ Không chắc, không biết

Thái độ và niềm tin là *những giá trị mà đối tượng nghiên cứu nghĩ đến*. Thường, đây là những câu hỏi mang tính cho ý kiến. Ví dụ như câu hỏi “Bạn nghĩ gì về các vấn đề lớn về giáo

dục mà nước ta đang đối phó hiện nay?” là một câu hỏi mang tính thái độ và niềm tin.

Hành vi là những gì họ làm trong cuộc sống hàng ngày hay những quyết định mà họ đi đến trong việc đối phó với một sự việc. Ví dụ:

Bạn đã lên kế hoạch học tập cho năm nay?

☐ Đã lên kế hoạch

☐ Chưa lên kế hoạch

Sau cùng là những dữ liệu liên quan đến *đặc điểm của đối tượng nghiên cứu* (như giới tính, độ tuổi, tiền sử lâm sàng, v.v.) Chẳng hạn như câu hỏi “*Bạn tốt nghiệp đại học từ năm nào?*” là một câu hỏi về đặc điểm.

Câu hỏi “Mở” và “Đóng”

Trong mỗi loại câu hỏi, có hai thể loại câu hỏi: câu hỏi Mở (Open-Ended) và câu hỏi “Đóng” (Close-Ended). Những câu hỏi thuộc thể loại Mở thường là những câu hỏi cái gì, tại sao, như thế nào. Đó là những câu hỏi có sẵn câu trả lời, và nó cho phép đối tượng nghiên cứu có thể dùng từ ngữ của chính họ. Đây là những câu hỏi giúp cho nhà nghiên cứu khai thác dữ liệu và tìm thêm ý tưởng. Tuy nhiên, nó có hạn chế là câu trả lời thiếu tính đồng dạng (un-uniformity) và do đó khó cho phân tích. Ví dụ như câu hỏi:

Bạn nghĩ gì về sự phục vụ của cơ quan công chứng?

thì đối tượng nghiên cứu có thể trả lời bất cứ gì họ muốn / thích. Nhưng chính vì sự ‘tự do’ đó mà nhà nghiên cứu sẽ rất khó định lượng trong phân tích.

Đối với câu hỏi Đóng, các câu hỏi được thiết kế để thu thập những câu trả lời có sẵn (Có/Không; Đúng/Sai; Đồng Ý/Không Đồng Ý). Những dữ liệu loại này rất dễ phân tích và do đó giới nghiên cứu định lượng rất thích. Một lợi thế khác là các câu hỏi Đóng thường dễ diễn giải, dù nhà nghiên cứu có thể không thu thập tất cả các câu hỏi khả dĩ. Ví dụ:

Bạn hãy đánh giá chất lượng của tờ giấy xuất viện của Bệnh viện ABC:

- ☐ Nghèo nàn
- ☐ Trung bình
- ☐ Tốt

Có một thể loại câu hỏi khác mang tính sàng lọc (filter questions). Đây là những câu hỏi mang tính thăm dò, rất có ích cho việc đảm bảo người trả lời cung cấp những trả lời liên quan đến câu hỏi. Thường, các câu hỏi này được thiết kế theo kiểu 2 tầng. Ví dụ:

Bạn có dùng phần mềm phân tích dữ liệu?

- ☐ Không – xin mời trả lời câu hỏi kế tiếp
- ☐ Có – phần mềm nào?

2. Các bước cần thiết để thiết kế một bộ câu hỏi

Kinh nghiệm cá nhân của tôi cho thấy để thiết kế một bộ câu hỏi tốt, cần phải trải qua 5 bước sau đây:

- Viết ra mục tiêu chính và mục tiêu phụ của nghiên cứu;
- Viết ra các khái niệm, thông tin cần thu thập;
- Kiểm tra nghiên cứu trong quá khứ xem có bộ câu hỏi nào chưa;

- Soạn bản nháp bộ câu hỏi;
- Chỉnh sửa và hoàn thiện.

Xác định mục tiêu. Bước đầu tiên là xác định và viết ra mục tiêu chính và mục tiêu phụ của nghiên cứu. Cần phải xác định vấn đề chính là gì, kèm theo những mục tiêu cụ thể. Mỗi mục tiêu cần có một câu văn, và lên kế hoạch phân tích cho mỗi mục tiêu. Ngoài ra, còn phải xác định đối tượng nghiên cứu cụ thể. Kế hoạch phân tích giúp cho nhà nghiên cứu bám sát mục tiêu và biết trước phương pháp để thu thập dữ liệu cần thiết.

Xác định biến số. Bước thứ hai trong việc soạn bộ câu hỏi là xác định các biến số cần thu thập. Trong bước này, nhà nghiên cứu cần phải liệt kê các dữ liệu cần thu thập cho nghiên cứu. Hãy tự hỏi: nghiên cứu cần biết về thái độ, hành vi, nhu cầu, đặc tính nhân trắc, hay tất cả các dữ liệu trên. Một cách để suy nghĩ về biến số cần thiết là phân biệt những biến nào được xem là biến tiên lượng (predictors), biến nhiễu (confounders) và biến phụ thuộc (dependent variable). Sau khi đã có các biến số cần thiết, triển khai thành những câu hỏi có thể có câu trả lời.

Tham khảo nghiên cứu trước. Bước ba là kiểm tra các nghiên cứu trước đã được công bố. Trong hầu hết các lĩnh vực nghiên cứu, rất có thể đã có các nhóm nghiên cứu trên thế giới thực hiện và công bố. Do đó, một việc làm rất cần thiết là phải điểm qua những nghiên cứu trước đây xem tác giả dùng bộ câu hỏi nào để thu thập dữ liệu. Tham chiếu các bộ câu hỏi trước giúp cho việc so sánh kết quả với các nghiên cứu trước dễ dàng hơn. Một số bộ câu hỏi đã được đăng kí bản quyền, và trong tình huống này, nhà nghiên cứu cần phải liên lạc tác giả để có bộ câu hỏi.

Soạn bản nháp. Sau khi đã xác định được mục tiêu, biến số cần thu thập, bước thứ tư là soạn bản nháp bộ câu hỏi. Trong bước này, nhà nghiên cứu cần phải xác định cách triển khai bộ câu hỏi (qua phỏng vấn trực tiếp, điện thoại, website, hay bưu điện). Sau đó, nhà nghiên cứu viết xuống những câu hỏi cụ thể sao cho đảm bảo các câu hỏi phải “trôi chảy” một cách tự nhiên và logic theo chủ đề.

Chỉnh sửa. Bước sau cùng là chỉnh sửa cho đến khi bộ câu hỏi có thể triển khai trong thực tế. Không có bộ câu hỏi nào là hoàn chỉnh, và quá trình hoàn thiện bộ câu hỏi có khi cần đến nhiều nỗ lực và thời gian. Trước khi triển khai đại trà, bộ câu hỏi cần phải được thử nghiệm trên vài chục đối tượng để đánh giá thời gian cần thiết cho việc trả lời câu hỏi, tính khả thi, và những rắc rối có thể xảy ra trong quá trình triển khai.

3. Nguyên tắc soạn bộ câu hỏi

Nguyên tắc 1: Câu hỏi 1 chiều và đơn giản

Câu hỏi phải đơn giản để người tham gia nghiên cứu chọn câu trả lời. Tuyệt đối tránh câu hỏi đa chiều, vì nó thiếu tính minh bạch, và người trả lời cảm thấy khó chọn câu trả lời đúng. Ví dụ như câu hỏi:

Mỗi ngày bạn uống bao nhiêu tách cà phê và trà?

vì có hai loại dữ liệu, trà và cà phê, nên nếu người chỉ uống cà phê thì khó trả lời. Giải pháp là nên chia câu hỏi thành 2 phần như sau:

Trong một ngày tiêu biểu, bạn uống bao nhiêu tách cà phê?

và

Trong một ngày tiêu biểu, bạn uống bao nhiêu tách trà?

Nguyên tắc 2: Câu hỏi không có giả định ngầm (hidden assumptions).

Câu hỏi giả định ngầm là câu hỏi giới hạn người trả lời vào một số đối tượng mà người hỏi nghĩ rằng đó là tất cả. Chẳng hạn như xuất phát từ suy nghĩ rằng đa số công chúng sở hữu máy tính của hãng Apple và IBM, nên câu hỏi đặt ra là:

Bạn đang sở hữu máy tính hiệu gì?

(A) IBM PC

(B) Apple

Câu hỏi này không tốt, vì người ta có thể sở hữu máy tính hiệu khác. Ngoài ra, người ta cũng có thể sở hữu cả hai máy tính IBM và Apple. Có thể sửa lại thành hai câu riêng:

Bạn đang sở hữu máy tính hiệu IBM?

☐ Có

☐ Không

Bạn đang sở hữu máy tính hiệu Apple?

☐ Có

☐ Không

Một cách khác là cho ra nhiều câu trả lời cho một câu hỏi:

Bạn đang sở hữu máy tính hiệu gì (bạn có thể tick nhiều hiệu)?

☐ Không có máy tính

☐ IBM PC

☐ Apple

☐ Dell

☐ Khác

Nguyên tắc 3: Câu hỏi và câu trả lời phải ăn khớp.

Đây là nguyên tắc rất cơ bản trong việc soạn bộ câu hỏi. Nếu câu hỏi và câu trả lời ăn khớp thì người hỏi rất dễ trả lời; nếu không thì sẽ làm cho người tham gia nghiên cứu cảm thấy lúng túng. Ví dụ như câu hỏi

Bạn có đau trong tuần vừa qua?

- ☐ Không ☐ Thỉnh thoảng
☐ Thường xuyên ☐ Rất thường xuyên

Chữ “có đau” hàm ý có/không, và do đó làm cho người trả lời khó chọn câu trả lời thích hợp. Tuy nhiên, chỉ cần thay đổi một chữ là câu hỏi rõ ràng hơn:

Bạn thường hay bị đau trong tuần vừa qua?

- ☐ Không ☐ Thỉnh thoảng
☐ Thường xuyên ☐ Rất thường xuyên

Nguyên tắc 4: Câu trả lời phải độc lập với nhau.

Trong nhiều trường hợp, khi câu hỏi nhập nhằng với câu trả lời sẽ làm cho người trả lời khó lựa chọn. Một ví dụ tiêu biểu của câu hỏi nhập nhằng là:

Bạn lớn lên ở đâu?

- ☐ Vùng quê
☐ Nông trại
☐ Thành phố

Hai lựa chọn “vùng quê” và “nông trại” có thể làm cho người trả lời cảm thấy lúng túng và dẫn đến chọn sai câu trả lời. Một cách tốt hơn là soạn câu hỏi sao cho lựa chọn độc lập nhau:

Bạn lớn lên ở đâu?

- ☐ Nhà ở miền quê
☐ Nông trại ở miền quê
☐ Thành phố

Nguyên tắc 5: Viết câu hỏi để có câu trả lời chính xác, tránh bất định

Nguyên tắc này cũng rất quan trọng, vì câu hỏi không rõ ràng sẽ làm cho người trả lời chọn câu trả lời không đúng. Chẳng hạn như câu hỏi:

Trong những yếu tố sau đây, yếu tố nào bạn nghĩ sẽ tăng nguy cơ bị đột quỵ?

☐ Hút thuốc lá ☐ Quá cân ☐ Căng thẳng

sẽ làm cho người trả lời lúng túng. Nếu họ không am hiểu về y tế, thì rất có thể câu trả lời không phản ánh đúng kiến thức của họ. Một giải pháp tốt hơn là soạn câu hỏi theo hướng chia lẻ:

Trong những yếu tố sau đây, yếu tố nào bạn nghĩ sẽ tăng nguy cơ bị đột quỵ?

Hút thuốc lá: ☐ Tăng ☐ Không ☐ Không biết

Quá cân: ☐ Tăng ☐ Không ☐ Không biết

Căng thẳng: ☐ Tăng ☐ Không ☐ Không biết

Nguyên tắc 6: Tránh “chi nhánh hoá”

Câu hỏi “chi nhánh hóa” sẽ làm cho người tham gia nghiên cứu mất thì giờ, và họ có thể bỏ sót câu trả lời. Chẳng hạn như:

(1) *Bạn đang có bảo hiểm nhân thọ? (Có / Không)*

Nếu “Không”, xin trả lời câu hỏi 3

[...]

(3) *Bạn trả tiền niên liễm hàng năm bao nhiêu?*

Một cách tốt hơn là chỉ viết một câu hỏi, ví dụ như:

Năm ngoái, bạn chi tiêu bao nhiêu tiền cho bảo hiểm nhân thọ? (Nếu không có bảo hiểm thì viết 0)

4. Những câu hỏi cần nên tránh

Tránh những câu hỏi dẫn (leading question)

Câu hỏi dẫn là những câu hỏi mang tính “điều dắt” người trả lời theo giả định của nhà nghiên cứu. Chẳng hạn như câu hỏi “*Bạn muốn được khám bởi bác sĩ cùng giới tính với bạn?*” thoát đầu thì không có vấn đề gì, nhưng nếu người đồng tính thì sẽ khó lựa chọn. Do đó, thay vì viết như trên, câu hỏi có thể viết lại dễ hiểu hơn:

Bạn muốn được khám bởi:

☐ Nam bác sĩ ☐ Nữ bác sĩ

☐ Nam hay nữ bác sĩ đều được

Tránh câu hỏi mù mờ (vague question)

Câu hỏi làm cho người trả lời khó lựa chọn hay khó trả lời là những câu hỏi không rõ ràng về ý nghĩa. Ví dụ như hai câu hỏi sau đây:

Tổng hợp lại, bạn có hài lòng trong thời gian nằm viện?

Trong năm qua, bạn đã đến bác sĩ để khám bệnh lần nào?

sẽ làm cho người tham gia nghiên cứu lúng túng. Chữ “hài lòng” cũng rất khó trả lời, vì hài lòng về khía cạnh nào. Câu “thời gian nằm viện” không rõ ràng vì thiếu sự giới hạn thời gian cụ thể. Có thể viết lại câu hỏi như sau:

Trong thời gian nằm bệnh vừa qua, nói chung, bạn đánh giá chất lượng chăm sóc của bệnh viện như thế nào?

hay thay vì viết “*Trong năm qua, bạn đã đến bác sĩ để khám bệnh lần nào?*” có thể soạn lại rõ ràng hơn:

“Trong thời gian 12 tháng qua, bạn đã đến khám phòng mạch bác sĩ bao nhiêu lần?”

Tránh câu hỏi thiên lệch (biased or value-laden question)

Đó là những câu hỏi lồng vào giá trị của người hỏi. Chẳng hạn như câu hỏi sau đây có thể xem là thiên lệch:

Bạn có nghĩ rằng y học thực chứng là việc vô bổ, phí thì giờ vì nó mang tính “nhét chữ vào miệng” người đọc. Do đó, cần phải viết lại cho trung dung hơn như:

Bạn nghĩ gì về y học thực chứng?

Tránh câu hỏi đe dọa (threatening question)

Cũng là một dạng câu hỏi thiên lệch và hàm ý dẫn dắt người đọc theo ý của người hỏi. Ví dụ:

Bạn có thường xuyên đánh con?

Bạn có đủ kiến thức để điều trị bệnh loãng xương?

Những câu hỏi này tuyệt đối cần phải tránh. Nên viết lại một cách thích hợp và “khoa học” hơn:

Bạn có thường xuyên dùng các hình thức sau đây để kỉ luật con?

hay

Bạn tự đánh giá kiến thức về điều trị bệnh loãng xương như thế nào?

Tránh câu hỏi nước đôi (double-barrelled question)

Đó là những câu hỏi có hai ý hay khái niệm trong một câu hỏi. Ví dụ:

Bạn có bị đau cổ hay đau lưng trong thời gian qua?

Ý tưởng “đau cổ” và “đau lưng” lồng vào một câu hỏi là không nên, vì một cách hiển nhiên người trả lời sẽ rất khó mà trả lời đúng. Cần phải viết lại như:

Trong thời gian 12 tháng qua, bạn có những triệu chứng nào dưới đây (xin ghi dấu vào hộp)

☐ Đau cổ ☐ Đau lưng ☐ Nhức đầu

Tránh câu hỏi phức hợp (complex question)

Cũng giống như câu hỏi nước đôi, những câu hỏi phức hợp lồng nhiều đối tượng và ý tưởng vào một câu hỏi. Ví dụ:

Trên thang điểm từ 1 đến 10, bạn hãy cho điểm mỗi trong số 12 thể loại dưới đây về kiến thức, sự tự tin, và kinh nghiệm của bạn.

Thay vì viết như trên, câu hỏi có thể soạn lại như sau:

Làm ơn điền vào bảng dưới đây về kiến thức, sự tự tin và kinh nghiệm của bạn trong mỗi lĩnh vực:

Đọc _____

Viết _____

Hiểu _____

Tránh câu hỏi hàm ý tiêu cực (negative question)

Câu hỏi tiêu cực thường dùng những chữ mang tính tiêu cực hay những chữ như “không”, “hiếm”, “không bao giờ”, v.v..

Bác sĩ không cần thiết phải khám bệnh nhân sau giờ làm việc:

☐ Đồng ý ☐ Không đồng ý

Câu hỏi trên có thể viết lại như sau:

Bác sĩ cần phải khám bệnh nhân sau giờ làm việc nếu cần thiết:

☐ Đồng ý ☐ Không đồng ý

Tránh câu hỏi phân nhóm

Thỉnh thoảng có những câu hỏi định lượng (như độ tuổi, chiều cao, cân nặng) nhưng lại được cho câu trả lời theo nhóm. Ví dụ như:

Xin cho biết tuổi của bạn

☐ 18 - 25

☐ 26 - 30

☐ 30 trở lên

Trong trường hợp trên, nếu một người ở độ tuổi 25 và 26 trở thành hai nhóm khác nhau, và cũng như 18 và 30. Do đó, cách phân nhóm làm mất thông tin định lượng. Một cách để khắc phục tình trạng trên là viết lại câu hỏi:

Ngày tháng năm sinh của bạn: ____/ ____/ ____

5. Thiết kế câu trả lời

Dữ liệu thu thập từ bộ câu hỏi có thể phân chia thành 3 nhóm chính: định danh (*nominal*), bán định lượng (*ordinal*), và định lượng (*continuous*). Các biến định danh có 'giá trị' như nam / nữ, tình trạng hôn nhân (độc thân / có gia đình / góa phụ / li dị / li thân / khác), v.v. Những biến này không thể dùng cho phân tích định lượng, nhưng có thể dùng cho phân tích đếm.

Những biến số bán định lượng thường được dùng để mô tả giai đoạn của bệnh lý như (ung thư ở giai đoạn tai chỗ / vùng / di căn), hay thái độ (hài lòng / không hài lòng). Một ví dụ tiêu biểu là câu hỏi sau đây:

Mức độ mà bạn đồng ý hay không đồng ý với phát biểu "Thầy/cô giáo có quyền phạt học sinh bằng roi":

☐ Rất đồng ý

☐ Đồng ý

☐ Trung dung

☐ Không đồng ý

☐ Rất không đồng ý

Những dữ liệu loại này cũng có giới hạn về lựa chọn, thường thiếu tính nhất quán.

Sau cùng là dữ liệu mang tính liên tục và định lượng. Tiêu biểu cho loại dữ liệu này là thang điểm VAS (visual analogue scales), độ tuổi, chiều cao, cân nặng, v.v. Những dữ liệu này giàu thông tin nhất và giúp cho việc phân tích thống kê có ý nghĩa hơn các biến bán định lượng.

Tóm lại, soạn bộ câu hỏi là một khâu quan trọng trong việc thiết kế nghiên cứu khoa học. Một bộ câu hỏi tốt phải đáp ứng 6 đặc điểm sau đây:

- Phải có tính hợp lí nội tại, có nghĩa là dữ liệu thu thập được phản ánh biến số mà nhà nghiên cứu muốn thu thập.
- Người trả lời biết câu trả lời. Không có lí do gì hỏi những câu hỏi mà đối tượng nghiên cứu không thể trả lời. Ví dụ như hỏi về thời còn 1-2 tuổi thì rất khó có câu trả lời đúng.
- Câu hỏi phải rõ ràng và đơn giản. Mỗi câu hỏi chỉ có một nghĩa. Không để một câu văn có thể được hiểu hơn 1 ý nghĩa. Chẳng hạn như chữ “độc thân” có thể hiểu theo nghĩa ở chung nhau nhưng chưa chính thức thành hôn, nhưng cũng có nghĩa là chưa thành hôn và chưa có người phối ngẫu.
- Câu hỏi phải mang tính thân thiện. Những câu hỏi như “cân nặng của bạn là bao nhiêu (lấy cân nặng pound nhân cho 0,45 để tính ra kilogram)” là chỉ làm cho người tham gia nghiên cứu thêm gánh nặng.
- Câu hỏi không tuyệt đối không mang tính xúc phạm.
- Công bằng và trung dung, không thiên lệch theo một quan điểm nào.

Xin nhắc lại như là một nhấn mạnh: có những “nên” và “không nên” trong việc soạn bộ câu hỏi. Nên dùng chữ đơn giản, ngắn gọn, và cụ thể. Không nên dùng những câu văn hay chữ mù mờ, không nên dùng kí hiệu và viết tắt. Tuyệt đối tránh câu hỏi xúc phạm người trả lời. Như đề cập ngay từ đầu, viết một câu hỏi thì rất dễ, nhưng viết ra một câu hỏi tốt thì rất khó, đòi hỏi suy nghĩ nhiều và đối chiếu với các nguyên tắc trình bày ở trên.

PHÂN TÍCH DỮ LIỆU

Sau khi thu thập dữ liệu, bước kế tiếp là phân tích dữ liệu. Đây là bước chuyển hóa từ dữ liệu sang thông tin. Phân tích dữ liệu là một khâu quan trọng trong nghiên cứu khoa học, bởi vì khoảng 50% bài báo khoa học bị từ chối là do phân tích sai. Phân tích dữ liệu bao gồm hai lĩnh vực liên quan với nhau: ước lượng (estimation) và kiểm định giả thuyết (test of hypothesis). Trong chương này, tôi muốn chia sẻ với bạn đọc các ý tưởng đằng sau hai lĩnh vực vừa kể.

Vai trò của phân tích thống kê

“Thống kê” là một thuật ngữ được dịch từ chữ “*statistics*” (tiếng Anh). *Statistics* có xuất xứ từ chữ “*status*” (tiếng Latin), có nghĩa là “nhà nước”. Trong cách dùng của tiếng Anh thời xưa, *statistics* (hay thống kê) là những bảng tóm tắt bằng con số nhằm mô tả những đặc điểm định lượng của nhà nước. Những “đặc điểm định lượng” này bao gồm số lượng đất đai của nhà vua, thuế má thu được, lương lậu cho lính, v.v. Nhưng ngày nay, thống kê được dùng để chỉ một ngành toán học ứng dụng, nghiên cứu và đánh giá mức độ biến thiên (hay thay đổi) và xác suất của các hiện tượng xã hội và sinh học. Những đánh giá này phải dựa vào logic và toán học. Do đó, muốn hiểu thống kê một cách chính xác, người tiếp nhận cần phải có một số kiến thức tối thiểu về logic học và toán học.

Khi ứng dụng trong các môi trường phức tạp, có nhiều nguyên nhân và hậu quả, như sinh học và xã hội, thống kê là một phương tiện rất hữu hiệu để mô tả và phân tích hiện tượng. Chúng ta có thể cảm nhận sự phong phú của thống kê qua một ví dụ sau đây: Trong một nghiên cứu về ảnh hưởng của việc li dị đến con cái, được giới báo chí loan tin khá rộng rãi gần đây. Theo nghiên cứu này, 80% con cái của những cặp cha mẹ li dị không có vấn đề gì khi trưởng thành và thích nghi tốt với xã hội. Kết quả này, thoát đầu mới nghe qua, thì quả là tốt; nhưng nếu chúng ta chịu khó suy nghĩ lại thì 20% con em (hay 1 trong 5 trẻ em) trong những gia đình li dị có vấn đề khi trưởng thành và không thích nghi tốt với xã hội. Qua ví dụ đơn giản này chúng ta có thể thấy những dữ kiện thống kê có thể được “chế biến”, đóng hộp cho những mục đích cụ thể, và nếu người tiếp nhận không chịu khó lần mò, tra cứu nguồn gốc thì những thông tin thống kê có thể mất đi ít nhiều ý nghĩa của chúng. Có một câu nói ví von và dí dỏm về thống kê như sau: Những dữ kiện thống kê như là những bộ bikini, những gì chúng phô trương thì hấp dẫn đấy, song những gì chúng che giấu mới là quan trọng.

Trong nghiên cứu khoa học, phân tích thống kê là nhằm mục đích chuyển hóa dữ liệu (data) thành thông tin (information). Dữ liệu là những gì nhà nghiên cứu thu thập được từ đo lường, phỏng vấn, hay thu thập qua các bộ câu hỏi. Vấn đề quan trọng hơn là biến những dữ liệu đó thành thông tin để nhà nghiên cứu có thể cảm nhận được ý nghĩa của chúng. Làm sao chúng ta biết phẫu thuật A có hiệu quả tốt hơn phẫu thuật B? Làm sao chúng ta biết thuốc aspirin có thể đem lại lợi ích cho bệnh nhân? Trong số 25 ngàn gen trong cơ thể con người, gen nào có khả năng gây ra ung thư,

tiểu đường, loãng xương? Làm sao chúng ta biết một giống lúa mới có sản lượng cao hơn giống lúa cũ? Làm sao chúng ta biết được quá trình học vấn ở cấp phổ thông có ảnh hưởng đến kết quả học tập ở bậc đại học? Tại sao trẻ học tiểu học ở nông thôn hay bỏ học? Làm sao chúng ta biết những đặc tính nào của cà phê được người tiêu thụ ưa chuộng, và có sự khác biệt về sở thích giữa nam và nữ hay không? Một số du khách vào Việt Nam có xu hướng “một đi không trở lại”, vậy yếu tố nào đã làm cho họ có xu hướng đó? Làm sao chúng ta biết người dân ủng hộ chính sách A mà không là chính sách B? Vân vân và vân vân. Đó là những vấn đề mà thống kê học có thể cung cấp câu trả lời khách quan và đáng tin cậy nhất.

Bốn thể loại phân tích thống kê

Các phương pháp phân tích thống kê có thể chia thành 3 dạng: mô tả, so sánh hoặc liên quan, và tiên lượng. Trong phần này tôi sẽ tổng quan về mục đích, giả định và nguyên lý của mỗi nhóm phương pháp, những chi tiết về kỹ thuật có thể tham khảo trong một cuốn sách về phương pháp phân tích dữ liệu khác.

Phân tích mô tả (descriptive analysis) như tên gọi thường có mục đích chính là ước tính một đại lượng. Chẳng hạn như chiều cao trung bình của quần thể, nồng độ estrogen của các bệnh nhân loãng xương, tỉ lệ mắc bệnh tiểu đường trong cộng đồng, tỉ lệ công chúng đồng ý với một chính sách, số người hài lòng với một kiến nghị, v.v. có thể ước tính bằng phương pháp thống kê. Phân tích mô tả không có mục đích kiểm định một giả thuyết khoa học.

Mô tả một định lượng như chiều cao hay tỉ lệ mắc bệnh trong cộng đồng thì rất đơn giản. Đối với biến liên tục như

chiều cao, **số trung bình** (*average* hay *mean*) có thể ước tính để mô tả ‘số tiêu biểu’ của một dãy số liệu. Theo Francis Galton, số trung bình có thể xem như là một “*wisdom of the crowd*” (tri thức của đám đông). Nhưng số trung bình có vấn đề, bởi vì nó không phản ánh được độ dao động của biến số. Thử tưởng tượng có hai dãy số của hai nghiên cứu có cùng đơn vị như sau:

- **Nghiên cứu 1:** số liệu thu thập từ 6 người: 6, 7, 8, 4, 5, và 6.
- **Nghiên cứu 2:** số liệu của 4 người: 10, 2, 3, và 9.

Cả hai biến số đều có số trung bình là 6. Nhưng nếu chú ý kỹ, chúng ta dễ dàng thấy phân bố của nghiên cứu 1 có vẻ ít biến thiên hơn so với nghiên cứu 2. Trong nghiên cứu 1, giá trị thấp nhất và cao nhất là 4 và 8; trong nghiên cứu 2, giá trị thấp nhất và cao nhất là 2 và 10. Do đó, số trung bình không phản ánh được độ biến thiên của dữ liệu. Chúng ta cần một chỉ số khác, và đó chính là độ lệch chuẩn (*standard deviation* hay SD). Độ lệch chuẩn là chỉ số phản ánh mức độ khác biệt giữa mỗi giá trị (của mỗi cá thể) và số trung bình (của tất cả cá thể) sau khi đã hiệu chỉnh cho số cỡ mẫu. Trong ví dụ trên, SD của nghiên cứu 1 là 1,41 và của nghiên cứu 2 là 4,1, và kết quả này phản ánh được ‘trực giác’ của chúng ta, tức là nghiên cứu 2 có độ biến thiên cao hơn nghiên cứu 1 cho dù cả hai có cùng số trung bình.

Với độ lệch chuẩn và số trung bình, và nếu biến số tuân theo luật phân bố chuẩn (*standardized normal distribution*) nhà nghiên cứu có thể ước tính **khoảng tin cậy 95%** (95% confidence interval) của các giá trị. Chẳng hạn như trong nghiên cứu 1, khoảng tin cậy 95% là $\text{mean} \pm 1.96 \times \text{SD} = 6 \pm$

$1,96 \times 1,41 = 3,2$ đến 8,8. Khoảng tin cậy 95% này có nghĩa là 95% các cá thể có giá trị dao động từ 3,2 đến 8,8. Với một giả định “mạnh” (không bàn ở đây), nhà nghiên cứu cũng có thể phát biểu rằng xác suất các cá thể trong quần thể nghiên cứu có giá trị từ 3,2 đến 8,8 là 0,95.

Nếu hai nghiên cứu có hai đơn vị khác nhau, thì rất khó so sánh độ dao động. Chẳng hạn như số trung bình và SD của chiều cao là 160 cm và 6 cm, của trọng lượng cơ thể là 65 kg và 3 kg, chúng ta không thể nói độ dao động của chiều cao cao hơn độ dao động của trọng lượng, bởi vì hai đơn vị khác nhau. Chú ý rằng độ lệch chuẩn có cùng đơn vị với số trung bình. Do đó, một cách khác để hóa giải vấn đề khác đơn vị đo lường là lấy SD chia cho số trung bình và nhân kết quả cho 100 ($SD / \text{mean} \times 100$) thì sẽ cho ra **hệ số biến thiên** (*coefficient of variation* hay CV). Trong ví dụ trên, chúng ta có thể ước tính CV cho chiều cao là 3,75% và trọng lượng là 4,61%. Hệ số biến thiên, do đó, là một chỉ số phản ánh độ dao động của một biến số sau khi đã hiệu chỉnh cho số trung bình.

Những mô tả trên áp dụng cho biến liên tục, chứ không thể áp dụng cho biến phân nhóm hay biến nhị phân. Biến nhị phân chỉ có hai giá trị (có/không, sống/chết, 0/1, v.v.). Đối với biến nhị phân, cách mô tả đơn giản nhất là tính tỉ lệ (hay xác suất). Chẳng hạn như một nghiên cứu có mục tiêu xác định tỉ lệ mắc bệnh tiểu đường qua xét nghiệm HbA1c, nhà nghiên cứu có thể lấy số người xét nghiệm dương tính chia cho tổng số đối tượng được xét nghiệm. Nhưng tỉ lệ đó chỉ là số trung bình, nó không phản ánh được độ dao động mẫu của nghiên cứu. Thử tưởng tượng có hai nghiên cứu với kết quả và cỡ mẫu như sau:

Nghiên cứu 1: cỡ mẫu $n = 100$ người, và số người tiểu đường là 10.

Nghiên cứu 2: cỡ mẫu $n = 1000$ người, và số người tiểu đường là 100.

Cả hai nghiên cứu đều cho ra tỉ lệ 10%. Nói cách khác, xác suất mắc bệnh tiểu đường trong cộng đồng là 0,10. Nhưng rõ ràng ‘độ tin cậy’ của hai nghiên cứu rất khác nhau. Chúng ta kì vọng rằng nghiên cứu với cỡ mẫu lớn (nghiên cứu 2) phải có độ tin cậy cao hơn nghiên cứu với cỡ mẫu thấp (nghiên cứu 1). Một cách để đánh giá độ tin cậy là tính độ lệch chuẩn. Độ lệch chuẩn của một tỉ lệ có thể ước tính từ tỉ lệ p và số cỡ mẫu n theo công thức $s = \sqrt{p(1-p)/n}$. Theo công thức này, độ lệch chuẩn của nghiên cứu 1 và 2 lần lượt là 0,03 và 0,009.

Phương pháp thống kê cho phép nhà khoa học có thể ước tính và đánh giá độ bất định của một ước số, và cho phép nhà khoa học trả lời câu hỏi “nếu tôi lặp lại nghiên cứu này 100 lần thì ước số này sẽ dao động như thế nào?” Chẳng hạn như trong nghiên cứu trên, nhà nghiên cứu có thể ước tính **khoảng tin cậy 95%** để trả lời câu hỏi. Dựa vào giả định phân bố chuẩn, khoảng tin cậy 95% của tỉ lệ mắc bệnh tiểu đường cho nghiên cứu 2 là: $0,10 \pm 1,96 \times s = 0,10 \pm 1,96 \times 0,009 = 0,08$ đến $0,12$. Chú ý rằng trong công thức đó, 1,96 là hằng số của phân bố chuẩn tương ứng với xác suất 95% của phân bố. Nói cách khác, nhà nghiên cứu không biết tỉ lệ mắc bệnh tiểu đường trong quần thể là bao nhiêu, nhưng nếu nghiên cứu này được lặp lại 100 lần thì 95 nghiên cứu sẽ cho ra kết quả với tỉ lệ dao động từ 8 đến 12%.

Tóm lại, các chỉ số thống kê thường dùng trong phân tích mô tả bao gồm số trung bình (mean), trung vị (median), độ

lệch chuẩn (standard deviation), khoảng tin cậy 95% (95% confidence interval), bách phân vị (percentile), v.v. Mỗi chỉ số thống kê có thể áp dụng cho số liệu thực tế, nhưng mỗi chỉ số đều có một số giả định đi kèm. Ví dụ như số trung bình và độ lệch chuẩn chỉ có ý nghĩa nếu biến số tuân theo luật phân bố chuẩn hay không quá lệch so với phân bố chuẩn. Giả định về phân bố chuẩn của một biến số liên tục có thể nói là nền tảng rất quan trọng trong khoa học thống kê.

Một hình thức phân tích mô tả khác là phân nhóm (cluster analysis). Chẳng hạn như nhà nghiên cứu dùng máy giải trình tự gen có thể giải mã những vi khuẩn liên quan đến bệnh viêm phổi. Hàng ngàn chủng loại của vi khuẩn, với những cấu trúc di truyền có thể giống nhau nhưng cũng có thể khác nhau. Phương pháp phân tích thống kê có thể giúp cho nhà khoa học nhận dạng các loài vi khuẩn có cùng “dòng họ”, và qua đó, có thể hệ thống hóa các chủng loại một cách khoa học.

Phân tích so sánh (comparison analysis)

Rất nhiều nghiên cứu khoa học đi ra ngoài việc mô tả, và thường có mục tiêu kiểm định một giả thuyết khoa học. Giả thuyết khoa học thường mang tính so sánh. Để biết thuốc A có hiệu quả hơn thuốc B, nhà khoa học đặt giả thuyết vô hiệu là hiệu quả hai thuốc tương đương nhau, và sau đó dùng phân tích thống kê để bác bỏ hay chấp nhận giả thuyết vô hiệu.

Việc ứng dụng các phương pháp phân tích so sánh tùy thuộc loại biến số và mô hình nghiên cứu. Chẳng hạn như để so sánh hai biến số liên tục tuân theo luật phân bố chuẩn, nhà nghiên cứu có thể sử dụng phương pháp kiểm định t (t-test). Nhưng nếu biến số là định tính hay nhị phân (chỉ có hai giá trị như có/không, sống/chết) thì phương pháp kiểm định Ki bình phương (Chi squared analysis) có thể áp dụng.

Phân tích liên quan (association analysis)

Rất nhiều nghiên cứu khoa học có mục tiêu so sánh. So sánh giữa 2 nhóm hay nhiều hơn 2 nhóm. So sánh giữa hai nhóm trong khoa học có thể đặt dưới lăng kính của kiểm định giả thuyết. Giả dụ rằng trong một nghiên cứu đánh giá hiệu quả của một loại thuốc trong điều trị loãng xương, nhà nghiên cứu cho một nhóm bệnh nhân dùng thuốc và một nhóm dùng giả dược (placebo). Sau khi phân tích mô tả, nhà nghiên cứu có kết quả mật độ xương trung bình của nhóm placebo và nhóm dùng thuốc lần lượt là 0,85 và 0,92 g/cm². Câu hỏi đặt ra là sự khác biệt trên có phải là do ngẫu nhiên hay do ảnh hưởng của thuốc. Nếu không có sự khác biệt giữa hai nhóm bệnh nhân (tức thuốc không có hiệu quả), thì hiệu số trung bình (tạm kí hiệu là *delta*) phải là 0. Do đó, giả thuyết vô hiệu là $\delta = 0$. Phân tích so sánh có mục tiêu kiểm định giả thuyết vô hiệu đó.

Để kiểm định giả thuyết vô hiệu về sự khác biệt giữa hai nhóm và biến số kết cục (outcome) là biến liên tục (mật độ xương), ngoài số trung bình, nhà nghiên cứu còn phải có thêm hai chỉ số khác: độ lệch chuẩn và cỡ mẫu. Với 3 thông tin (trung bình, độ lệch chuẩn, và cỡ mẫu), nhà nghiên cứu có thể ước tính kiểm định t:

$$t = \frac{\delta}{SD(\delta)}$$

Trong đó, *delta* là hiệu số giữa 2 số trung bình, và $SD(\delta)$ là độ lệch chuẩn của *delta*. Có thể xem kiểm định t là một đo lường về tín hiệu và nhiễu. Hiệu số giữa hai số trung bình có thể xem là tín hiệu (*signal*) của dữ liệu. Độ lệch chuẩn của hiệu số phản ánh độ nhiễu của dữ liệu (*noise*). Do đó, kiểm định t

là một tỉ số của tín hiệu trên nhiễu. Có thể nói rằng đại đa số các phương pháp kiểm định thống kê đều được xây dựng theo công thức tín hiệu trên nhiễu.

Nếu t có giá trị 0 thì dữ liệu chỉ là nhiễu. Nếu $t = |1|$ thì dữ liệu cho thấy tín hiệu và nhiễu bằng nhau, không thể kết luận gì được. Nếu t dưới $|1|$ thì dữ liệu có nhiễu nhiều hơn là tín hiệu, cũng không kết luận gì được. Nếu t cao hơn 1 thì dữ liệu có nhiều tín hiệu hơn nhiễu. Vấn đề đặt ra là $|t|$ phải là bao nhiêu thì mới kết luận rằng hai nhóm thật sự khác nhau. Để có thể kết luận sự khác biệt giữa hai nhóm, chỉ số t cần phải so sánh với một chỉ số chuẩn trong phân bố t . Phân bố t có hình thức giống như phân bố chuẩn, nhưng nó tùy thuộc vào cỡ mẫu. Thường thì khi t cao hơn $|2|$, nhà nghiên cứu có thể phát biểu rằng giả thuyết vô hiệu là không khả dĩ, hay dữ liệu quan sát không nhất quán với giả thuyết vô hiệu.

Một cách khác để đánh giá hiệu quả của thuốc theo thiết kế so sánh hai nhóm là ước tính khoảng tin cậy 95% của hiệu số δ . Xin nhắc lại rằng nếu thuốc không có hiệu quả, tức hiệu số trung bình giữa hai nhóm là 0. Nhưng nhà nghiên cứu quan sát hiệu số $0,92 - 0,85 = 0,07$. Ước số 0,07 được ước tính từ dữ liệu của một mẫu, và mẫu được chọn ngẫu nhiên từ một quần thể. Nếu nhà nghiên cứu lấy một mẫu khác cũng từ quần thể, và đo mật độ xương thì hiệu số có thể là 0,10. Nếu lấy mẫu một lần khác thì hiệu số trung bình có thể là -0,12. Nói chung, cứ mỗi lần lấy mẫu, nhà nghiên cứu sẽ có một kết quả khác. Thử tưởng tượng, nhà nghiên cứu lấy 100 mẫu (tức 100 nghiên cứu), và tính hiệu số của 100 nghiên cứu, câu hỏi đặt ra là hiệu số δ thấp nhất và cao nhất là bao nhiêu? Để minh họa cho ý tưởng và cách diễn giải, thử tưởng có 2 tình huống giả thuyết như sau: 95% giá trị δ dao động từ:

(a) -0,03 đến 0,14

(b) 0,04 đến 0,10

Trường hợp (a) cho thấy nếu lặp lại nghiên cứu 100 lần thì sẽ có 95 nghiên cứu cho ra kết quả nhóm dùng thuốc có mật độ xương trung bình thấp hơn ($-0,03 \text{ g/cm}^2$) nhưng cũng thể cao hơn ($0,14 \text{ g/cm}^2$) nhóm chứng. Nói cách khác, nhà nghiên cứu không có chứng cứ để kết luận rằng thuốc có hiệu quả. Kết quả trên bất định.

Trường hợp (b) cung cấp nhà nghiên cứu chứng cứ để kết luận rằng thuốc có hiệu quả. Lí do là 95% nghiên cứu cho ra kết quả nhất quán rằng nhóm dùng thuốc có mật độ xương trung bình cao hơn nhóm chứng. Nói cách khác, khoảng tin cậy 95% đều dương tính, nên chứng cứ rất nhất quán.

Nếu giá trị trung bình của mật độ xương khác biệt giữa nhóm dùng thuốc và nhóm chứng, nhà nghiên cứu có thể kết luận rằng thuốc có ảnh hưởng. Khái niệm khác biệt cũng có thể diễn giải như là sự hiện hữu của một **mối liên quan (association)**. Chẳng hạn như nhà nghiên cứu quan sát rằng tỉ lệ hút thuốc lá ở nhóm bệnh nhân ung thư phổi cao hơn tỉ lệ hút thuốc lá trong nhóm không mắc bệnh, nhà nghiên cứu có thể phát biểu rằng có mối liên quan giữa hút thuốc lá và ung thư phổi.

Định lượng liên quan giữa hai biến nhị phân

Thước đo của mối liên quan rất đa dạng. Hiệu số *delta* như trình bày trong ví dụ trên là một thước đo, nhưng ngoài *delta* ra còn có một số thước đo phổ biến khác như **tỉ số odds (odds ratio hay OR)**, **tỉ số nguy cơ (risk ratio hay relative risk hay RR)**, **hệ số tương quan (correlation coefficient)**. Tỉ số odds và tỉ số nguy cơ áp dụng cho các nghiên cứu mà biến độc lập

(X) và biến phụ thuộc (Y) là nhị phân. Chẳng hạn như công trình nghiên cứu về hiệu quả của thuốc zoledronic acid, các nhà nghiên cứu cho một nhóm bệnh nhân gồm 1062 người dùng thuốc, và một nhóm gồm 1065 dùng giả dược cùng với calcium và vitamin D (nhóm chứng). Sau 3 năm theo dõi, các nhà nghiên cứu ghi nhận có 92 người trong nhóm dùng thuốc và 139 người trong nhóm chứng bị gãy xương:

Nhóm bệnh nhân	Gãy xương	Không gãy xương	Nguy cơ gãy xương
Nhóm zoledronic acid	970	92	$92/1062 = 0,087$
Nhóm chứng	926	139	$139 / 1065 = 0,131$

Nguồn: Lyles KW, et al. Zoledronic acid and clinical fractures and mortality after hip fracture. *N Engl J Med* 2007;357.

Tính trung bình, mức độ khác biệt giữa hai nhóm hay delta là 0,0439 (hay 4,4%). Với những dữ liệu trên, câu hỏi đặt ra là sự khác biệt về nguy cơ gãy xương giữa hai nhóm (8,7% trong nhóm điều trị và 13,1% trong nhóm chứng) có thật sự khác nhau hay chỉ là do yếu tố ngẫu nhiên? Cách đơn giản nhất là tính khoảng tin cậy 95% của delta. Phương pháp này rất đơn giản và có thể mô tả như sau:

$$\text{Tính delta} = (139 / 1065) - (92 / 1062) = 0,044$$

Tính phương sai của nhóm điều trị:

$$p(1 - p) / N = 0,087 \times 0,913/1062 = 0,0000745.$$

$$\text{Tính phương sai của nhóm chứng} = 0,131 \times 0,869/1065 = 0,000106.$$

Tính phương sai của delta là tổng phương sai của hai nhóm: $0,0000745 + 0,000106 = 0,00018$.

Tính độ lệch chuẩn của delta = $\sqrt{0,00018}=0,013$

Tính khoảng tin cậy 95% của delta = $0,044 \pm 1,96 \times 0,013$
= 0,018 đến 0,069.

Nói cách khác, nếu nghiên cứu lặp lại 100 lần (không thể!) thì sẽ có 95% nghiên cứu cho ra kết quả hiệu số nguy cơ gãy xương giữa hai nhóm dao động từ 1,8% đến 6,9%, nhưng trung bình là 4,4%. Nói cách khác nữa, xác suất mà thuốc zoledronic acid giảm nguy cơ gãy xương từ 1,8 đến 6,9% là 0,95 (phát biểu này có một giả định ‘mạnh’ đằng sau). Bởi vì 95% của hiệu số delta không bao gồm 0, nên nhà nghiên cứu cũng có thể phát biểu rằng sự khác biệt có ý nghĩa thống kê (statistically significant).

Kết quả trên cũng có thể mô tả bằng các ‘thước đo’ như tỉ số odds (OR) và tỉ số nguy cơ (RR). Trong nghiên cứu trên, nguy cơ gãy xương ở nhóm điều trị là $p_1 = 8,7\%$ và nhóm chứng là $p_2 = 13,1\%$. Chữ “nguy cơ” trong y học chính là xác suất; p_1 và p_2 là hai xác suất. Tỉ số nguy cơ được định nghĩa như là tỉ số của p_1 trên p_2 :

$$RR = \frac{p_1}{p_2}$$

So sánh RR với $delta = p_1 - p_2$, chúng ta thấy rằng hai chỉ số này (RR và $delta$) tuy hai mà chỉ một. Khác nhau là cách tính. Nếu $delta = 0$ thì $RR = 1$, tức thuốc không có hiệu quả. Nếu $delta < 0$ thì $RR < 1$, tức thuốc giảm nguy cơ gãy xương (có lợi). Nếu $delta > 0$ thì $RR > 1$, tức thuốc tăng nguy cơ gãy xương (có hại). Chúng ta cũng có thể tính khoảng tin cậy 95% của RR , và nếu khoảng tin cậy 95% không bao gồm 1 thì kết quả đó có nghĩa là thuốc có hiệu quả (nếu $RR < 1$) hay thuốc có hại ($RR > 1$).

Một chỉ số khác để định lượng mối liên quan giữa biến X và Y là dùng tỉ số odds. Khái niệm odds rất đặc thù trong tiếng Anh, và các ngôn ngữ khác không có odds, nên tôi không dịch sang tiếng Việt. Odds là tỉ số của xác suất biến cố xảy ra trên xác suất biến cố không xảy ra. Gọi p_1 là xác suất gãy xương của nhóm điều trị, odds gãy xương được định nghĩa như sau:

$$odds_1 = \frac{p_1}{1 - p_1}$$

tương tự, odds cho nhóm chứng là

$$odds_2 = \frac{p_2}{1 - p_2}$$

Tỉ số odds hay OR được định nghĩa là tỉ số của hai odds:

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

Trong ví dụ trên, với $p_1 = 0,087$ và $p_2 = 0,131$, $RR = p_1 / p_2 = 0,664$. Nói cách khác, nguy cơ gãy xương của nhóm được điều trị bằng 66,4% nguy cơ nhóm chứng (không điều trị). Kết quả này có nghĩa là thuốc zoledronic acid giảm nguy cơ gãy xương 33,6%.

Nếu mô tả bằng OR, chúng ta có: $Odds_1 = 0,087 / 0,913 = 0,0952$ và $Odds_2 = 0,131 / 0,869 = 0,151$. Do đó, $OR = 0,952 / 0,151 = 0,632$. Kết quả này có thể diễn giải rằng odds gãy xương ở bệnh nhân được điều trị bằng 63,2% odds của nhóm chứng. Nói cách khác, thuốc zoledronic acid giảm odds gãy xương 36,8%. Chú ý rằng, chúng ta chỉ có thể nói “odds gãy xương” chứ không phải “nguy cơ gãy xương”.

Chú ý rằng kết quả trên cho phép chúng ta phát biểu 3 cách khác nhau. Cách thứ nhất là qua delta, chúng ta có thể

phát biểu rằng thuốc zoledronic acid giảm nguy cơ gãy xương tuyệt đối 4,4% (so với nhóm chứng). Cách thứ hai là thuốc zoledronic acid giảm nguy cơ gãy xương tương đối 33,6%. Cách thứ ba là thuốc zoledronic acid giảm odds gãy xương tương đối 36,8%.

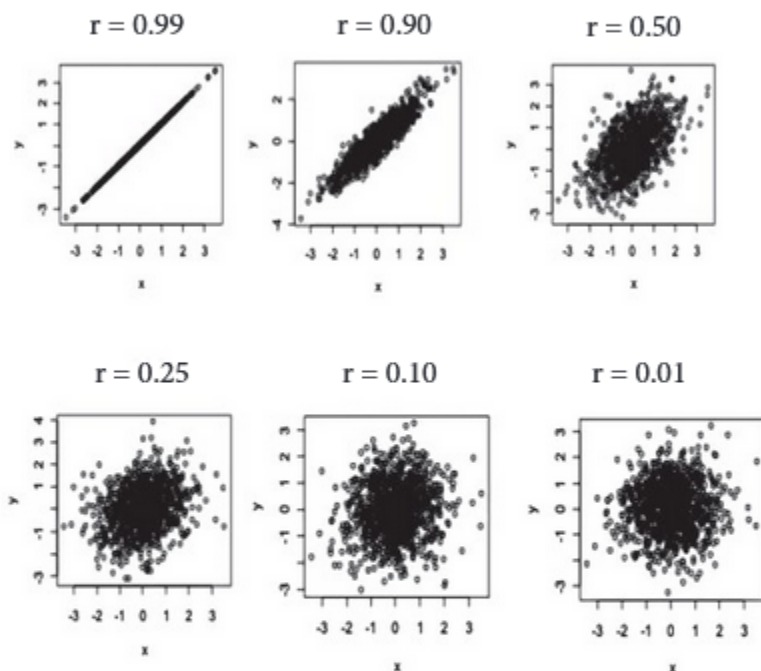
Định lượng liên quan giữa hai biến liên tục

Nếu biến độc lập X và biến phụ thuộc Y đều là biến liên tục thì hệ số tương quan (coefficient of correlation) có thể dùng để “đo lường” mối liên quan giữa X và Y . Hệ số tương quan (thường kí hiệu là r) được ước tính từ phương sai (variance) của X và Y , và hiệp biến (covariance) giữa X và Y . Hiệp biến giữa X và Y là tích số $X \cdot Y$ sau khi đã hiệu chỉnh cho số cỡ mẫu. Hệ số tương quan cũng có thể mô tả một cách đơn giản như sau:

$$r = \frac{\text{cov}(x,y)}{\sqrt{\text{var}(x) \times \text{var}(y)}} = \frac{\text{cov}(x,y)}{SD_x \times SD_y}$$

Trong đó, $\text{var}(x)$, $\text{var}(y)$, $\text{cov}(x,y)$ lần lượt là phương sai của biến X , phương sai của Y , và hiệp biến giữa X và Y . Khi lấy căn số bậc 2 của phương sai thì kết quả là độ lệch chuẩn (SD). Công thức trên có nghĩa rằng hệ số tương quan r chỉ là tỉ số của hiệp biến trên tích số của hai độ lệch chuẩn. Hệ số tương quan có giá trị tối thiểu là 0 và tối đa là $|1|$. Khi $r = 0$ có nghĩa là hai biến X và Y không có liên quan với nhau. Khi $r = 1$ thì hai biến X và Y có liên quan tuyệt đối. Trong thực tế, mối liên quan giữa X và Y thường dao động trong khoảng trên 0 và thấp hơn $|1|$. Nếu r có giá trị dương (ví dụ như $r = 0,7$), thì kết quả này có nghĩa rằng khi giá trị của X tăng thì giá trị của Y cũng tăng (tương quan thuận). Nếu r có giá trị âm (ví dụ như

$r = -0,6$), có nghĩa rằng khi giá trị của X tăng thì giá trị của Y giảm (tương quan nghịch đảo). Biểu đồ dưới đây minh họa cho 6 mối liên quan thuận giữa X và Y .



Phân tích tiên lượng (prediction analysis).

Một trong những mục tiêu chung của nghiên cứu khoa học là tìm ra những quy luật tự nhiên và những mối liên quan. Khi một mối liên quan giữa một hay nhiều biến độc lập (còn gọi là biến tiên lượng - *predictor variables*) và biến phụ thuộc Y đã được phát hiện, bước kế tiếp là xây dựng mô hình tiên lượng. Chẳng hạn như nếu nhà nghiên cứu khám phá hàng loạt gen có liên quan đến bệnh loãng xương, đó là một bước rất quan trọng; nhưng bước kế tiếp là làm thế nào để triển khai

những mối liên quan này thành một mô hình tiên lượng. Câu hỏi là nếu một cá nhân có những gen này thì nguy cơ (hay xác suất) mà cá nhân mắc bệnh loãng xương là bao nhiêu. Mô hình tiên lượng có thể trả lời câu hỏi đó.

Dùng dữ liệu của biến X (hay nhiều biến X) để tiên lượng giá trị của Y là một trong những «sứ mệnh» của nghiên cứu khoa học, nên rất nhiều nghiên cứu ngày nay dùng mô hình tiên lượng. Trong y học, sinh học, khoa học xã hội, vật lý, v.v. đã có rất nhiều mô hình tiên lượng được xây dựng và ứng dụng trong thực tế. Chẳng hạn như khi bệnh nhân nhập viện trong tình trạng cấp cứu, bác sĩ có thể sử dụng các thông tin của bệnh nhân (như tuổi, giới tính, các xét nghiệm thường qui, tiền sử lâm sàng) để tiên lượng nguy cơ... tử vong trong một thời gian ngắn như 30 ngày. Dĩ nhiên, bác sĩ có kinh nghiệm về hồi sức cấp cứu cũng có thể tiên lượng nguy cơ tử vong, nhưng có nhiều nghiên cứu cho thấy tiên lượng của bác sĩ thường thiếu tính nhất quán và do đó độ tin cậy không cao. Lí do đơn giản là bác sĩ tuy giàu kinh nghiệm lâm sàng, nhưng vẫn không thể nào xử lí nhiều thông tin và đánh giá trọng số của thông tin. Một giải pháp thay thế là dùng mô hình thống kê, và việc áp dụng mô hình thống kê trong tiên lượng rất thành công trong quá khứ và hiện tại.

Có nhiều phương pháp thống kê có thể dùng để xây dựng mô hình tiên lượng. Tùy vào mối liên quan giữa biến tiên lượng X và biến kết cục Y , mô hình tiên lượng có thể trong nhóm tuyến tính hay phi tuyến tính. Các mô hình tuyến tính bao gồm hồi qui tuyến tính. Các mô hình phi tuyến tính bao gồm hồi qui phi tuyến tính (nonlinear regression), hồi qui logistic, mô hình Cox, v.v. Một số mô hình “machine learning” (sẽ bàn dưới đây) cũng dựa trên cơ sở các mô hình thống kê vừa kể.

Nếu biến phụ thuộc là biến liên tục (như chiều cao, trọng lượng, nồng độ HbA1c, mật độ xương, thu nhập, v.v.) mô hình hồi qui tuyến tính đa biến và các 'biến thể' của mô hình có thể áp dụng. Nếu biến phụ thuộc là biến phân nhóm (như giai đoạn tiến triển của bệnh) hay biến nhị phân (có/không), mô hình hồi qui logistic, mô hình phân định (discriminant analysis), mô hình phân tích cụm (cluster analysis) có thể áp dụng. Các mô hình phân tích này có thể dùng dữ liệu của các biến độc lập để ước tính xác suất của biến nhị phân (ví dụ như xác suất tử vong).

Mô hình tiên lượng có thể phát triển và triển khai theo qui trình 3 bước như sau:

Bước 1: xác định các biến có liên quan. Bước này, nhà nghiên cứu dùng các phương pháp thống kê trong phân tích liên quan để nhận ra biến nào có liên quan độc lập đến biến phụ thuộc.

Bước 2: xây dựng mô hình tiên lượng (development). Nhà nghiên cứu chọn một phương pháp phân tích (hồi qui tuyến tính, hồi qui logistic, mô hình Cox) để ước tính các tham số cần thiết cho mô hình. Mô hình tiên lượng thường có dạng: $f(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k$ trong đó Y là biến phụ thuộc, $X_1, X_2, X_3, \dots, X_k$ là các biến tiên lượng; k là số biến tiên lượng.

Bước 3: kiểm tra mô hình tiên lượng trong một quần thể độc lập (validation). Nhà nghiên cứu ứng dụng mô hình phát triển trong bước 2 trên một quần thể khác, hoàn toàn độc lập với mẫu gốc.

Các thước đo để đánh giá một mô hình tiên lượng là độ nhạy, độ đặc hiệu, độ phân định (discrimination), độ chính

xác (calibration), và tái phân nhóm (reclassification). Trong đó, độ nhạy, độ đặc hiệu và chỉ số phân định là quan trọng nhất. Kết quả tiên lượng dùng mô hình có thể là dương tính hoặc âm tính. Trong thực tế, một cá nhân hoặc là có bệnh hay không có bệnh. Do đó, chúng ta có 4 tình huống chính:

(a) **độ nhạy**: bệnh nhân có bệnh, và kết quả tiên lượng dương tính.

(b) **âm tính giả**: bệnh nhân có bệnh, nhưng kết quả tiên lượng âm tính.

(c) **dương tính giả**: bệnh nhân không có bệnh, nhưng kết quả tiên lượng dương tính.

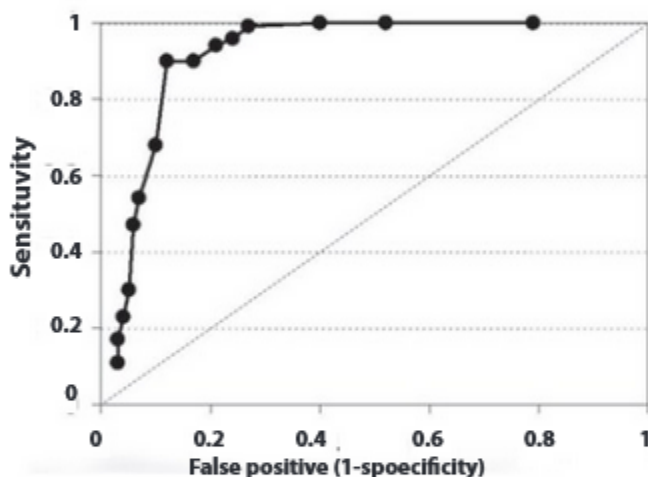
(d) **độ đặc hiệu**: bệnh nhân không có bệnh, và kết quả tiên lượng âm tính.

Bốn tình huống trên có thể tóm lược trong bảng dưới đây:

Kết quả tiên lượng	Tình trạng "thật" của bệnh	
	Có bệnh	Không có bệnh
Dương tính (+ve)	Dương tính thật Sensitivity, true positive	Dương tính giả False positive
Âm tính (-ve)	Âm tính giả False negative	Âm tính thật Specificity, true negative

Một mô hình tiên lượng tốt phải có độ nhạy và độ đặc hiệu cao (gần 1). Nhưng trong thực tế thì nếu chúng ta điều chỉnh ngưỡng xác suất của mô hình tiên lượng để tăng độ nhạy thì sẽ làm giảm độ đặc hiệu, hoặc tăng độ đặc hiệu thì sẽ làm giảm độ nhạy. Do đó, một cách để dung hòa hai độ nhạy

và độ đặc hiệu là dùng chỉ số AUC (area under the curve), tức diện tích dưới đường biểu diễn ROC (receiver operating characteristic curve). Biểu đồ ROC có trục hoành là tỉ lệ dương tính giả và trục tung là tỉ lệ dương tính thật (tức độ nhạy). Chẳng hạn như biểu đồ dưới đây cho thấy độ nhạy và dương tính giả của xét nghiệm PSA trong việc tiên lượng bệnh ung thư tiền liệt tuyến.



Biểu đồ ROC thể hiện dương tính thật (trục tung) và dương tính giả (trục hoành) của xét nghiệm PSA trong phát hiện ung thư tiền liệt tuyến ở đàn ông Mĩ da đen trong độ tuổi 70

Diện tích AUC tối thiểu là 0,5 và tối đa là 1. AUC có nghĩa là xác suất mà một cặp đôi, một với bệnh và một bình thường, được chọn ngẫu nhiên trong cộng đồng, và kết quả tiên lượng của bệnh nhân cao hơn người bình thường. Chúng ta có thể lấy một ví dụ để minh họa: nếu chúng ta chọn một cặp đối tượng (chọn một cách ngẫu nhiên từ một quần thể), một người có bệnh và một người không có bệnh; nếu cả hai người đều được tiên lượng bằng một mô hình có AUC = 0,85, thì điều này có

nghĩa là xác suất mà người có bệnh có kết quả tiên lượng với giá trị cao hơn người không mắc bệnh là 85%.

Một mô hình tiên lượng có ích phải có diện tích AUC trên 0,5. Không có ngưỡng nào của AUC để xác định là một mô hình tiên lượng tuyệt vời. Tuy nhiên theo qui ước thì một mô hình với AUC trên 0,8 được xem là tốt hay rất tốt; còn AUC dưới 0,6 được xem là không tốt và không thể áp dụng vào lâm sàng được:

Diễn giải ý nghĩa của diện tích dưới đường biểu diễn ROC (AUC)

AUC	Ý nghĩa
>0,90	Rất tốt (Excellent)
0,80 đến 0,90	Tốt (Good)
0,70 đến 0,80	Trung bình (Fair)
0,60 đến 0,70	Không tốt (Poor)
0,50 đến 0,60	Vô dụng (Fail)

Machine Learning

Trong vài năm gần đây, “machine learning” (ML) được đề cập đến nhiều lần, và gây ra nhiều lẫn lộn và hiểu lầm. Trong phần này, tôi sẽ cố gắng giải thích mối tương quan giữa machine learning và phân tích thống kê.

Có người dịch Machine Learning là “Học Máy”, nhưng tôi thì nghĩ dịch đúng hơn là “Mô hình hoá bằng máy”. Lí do là ở ý nghĩa của chữ “learning”, không hẳn có nghĩa đen là ‘học’, mà là mô hình hoá. Tại sao learning? Thực chất của ML là tìm mô hình thích hợp từ dữ liệu mẫu (sample data), và dùng cái

mô hình đó để tiên lượng cho quần thể (population) hay các mẫu dữ liệu khác.

Quá trình “learning” gồm 3 bước như mô tả trên: bước 1 là nhận dạng biến liên quan; bước 2 là xây dựng hệ thống phụ thuộc giữa các biến số (dependencies); và bước 3 là dùng hệ phụ thuộc để tiên lượng cho quần thể. Ở đây, cái gọi là “hệ thống phụ thuộc” chính là mô hình hoá dữ liệu, tức là “data modelization”. Nói cách khác nữa, “**learning**” trong ML chính là “**fitting**” trong thống kê học. Hiểu đúng như thế thì chúng ta dễ dàng thấy Machine Learning thực chất là “**mô hình hoá bằng máy tính.**”

Có hai phương pháp mô hình hoá: có giám sát (supervised learning) và không có giám sát (unsupervised learning). Mô hình hoá có giám sát bao gồm các mô hình như hồi qui tuyến tính, linear classification, logistic regression, Cox’s regression. Mô hình hoá không giám sát bao gồm phân tích cụm (cluster analysis) và phân tích đa biến như PCA và factor analysis. Bảng dưới đây sẽ phân biệt 4 nhóm mô hình / phương pháp phổ biến:

Biến phụ thuộc	Supervised Learning	Unsupervised Learning
Phân nhóm (discrete outcome)	Classification	Clustering
Liên tục (continuous outcome)	Regression	Dimensionality reduction

Để đánh giá độ chính xác, “bí quyết” của ML là huấn luyện (training hay development) và kiểm định (testing hay validation). Nói cách khác, ML chia một dataset ra thành

hai nhóm: nhóm huấn luyện để mô hình hoá dữ liệu, và sau khi mô hình xong thì áp dụng cho nhóm thử nghiệm. Qui trình này lặp đi lặp lại rất nhiều lần theo nguyên lí bootstrap trong thống kê học. Bảng dưới đây sẽ giúp bạn đọc phân biệt sự tương đồng giữa thuật ngữ thống kê học và ML (tôi dùng tiếng Anh):

Ngôn ngữ bình thường	Thống kê học	Machine learning
Modeling (mô hình)	Fitting	Learning
New data (dữ liệu mới)	New data	Generalization
Depenent variable (biến phụ thuộc)	Outcome, Response	Target
Independent variable (biến độc lập)	Predictor, risk factor	Feature
Baseline (điểm khởi đầu)	Intercept	Bias
Regression coefficient (hệ số hồi qui)	Beta coefficient	Weights
Error (sai số)	Residual / deviation	Loss function
Accuracy of prediction (độ chính xác)	Classification table	Confusion matrix
Selection of models (chọn mô hình)	Variable selection	Feature selection

Thật vậy, nhiều khái niệm của ML không có gì khác với thống kê học; nếu khác thì chỉ là cách gọi. Chẳng hạn như ML gọi là target, thì thống kê học gọi là outcome hay response; bias của ML chính là intercept trong thống kê; weight trong

ML là beta coefficient trong thống kê; loss function trong ML là residual trong thống kê; feature selection trong ML thì chính là variable selection trong thống kê; vân vân. Hiểu như thế thì chúng ta thấy rằng ML cũng có thể xem là một cách gọi khác của thống kê học hiện đại.

Machine learning và statistical learning (SL)

Trong thực tế, chúng ta thấy không có sự khác biệt lớn giữa hai thuật ngữ này. Nhưng có một chút khác biệt giữa ML và khoa học thống kê. Các mô hình thống kê chính là cơ chế của ML. Thống kê học đặt nặng vấn đề giả định (assumptions) trong việc xây dựng mô hình, nhưng ML thì đặt nặng độ chính xác của mô hình. Chẳng hạn như để xây dựng mô hình tiên lượng hồi qui tuyến tính, thống kê học rất quan tâm đến các giả định như phân bố chuẩn, đồng dạng phương sai, và tính độc lập giữa các giá trị. Nhưng ML quan tâm nhiều đến việc huấn luyện mô hình sao cho mô hình đó có thể tiên lượng chính xác khi áp dụng cho một quần thể khác. Nói như thế không có nghĩa là thống kê học không quan tâm đến độ chính xác của mô hình, mà chỉ muốn nói về ưu tiên của hai cách tiếp cận về mô hình tiên lượng.

Ngoài ra, thống kê học quan tâm đến mô hình để đánh giá tầm quan trọng của yếu tố tiên lượng (qua các kiểm định thống kê), còn ML thì không quan tâm đến ý nghĩa thống kê của các yếu tố trong mô hình. Đối với ML, một yếu tố có thể không có ý nghĩa thống kê nhưng có thể là một yếu tố tiên lượng tốt. Như vậy, thống kê học và ML khác biệt chủ yếu ở yếu tố “văn hoá”: một bên là quan tâm đến *association* (liên quan) và một bên là *prediction* (tiên lượng).

Phân biệt giữa association và prediction

Trong nhiều nghiên cứu khoa học, mục tiêu là giải thích mối liên quan giữa các yếu tố. “Yếu tố” ở đây có thể là gen, là bệnh lí, hay nói theo ngôn ngữ thống kê là biến độc lập và biến phụ thuộc, hay nói theo ngôn ngữ dịch tễ học là yếu tố nguy cơ và outcome.

Để giải thích mối liên quan giữa bệnh lí (Y) và yếu tố nguy cơ (X), chúng ta cần phải phát biểu một giả thuyết khoa học. Giả thuyết là một phát biểu mang tính tiên lượng về mối liên quan giữa Y và X. Từ giả thuyết khoa học, chúng ta triển khai thành giả thuyết thống kê. Từ giả thuyết thống kê, chúng ta triển khai thành mô hình thống kê. Đã nói là “mô hình” thì phải có hàm số, và ở đây là tìm hàm số của X để giải thích mối liên quan giữa Y và X. Hàm số có thể là tuyến tính (ví dụ như mô hình hồi qui tuyến tính) hay phi tuyến tính. Sau đó, chúng ta dùng dữ liệu thu thập từ thí nghiệm để ước tính tham số của mô hình thống kê. Qua ước tính tham số và kiểm định thống kê (test of significance), chúng ta có thể biết X có liên quan với Y hay không, và mức độ liên quan cao thấp ra sao. Như vậy, toàn bộ qui trình nghiên cứu, từ phát biểu giả thuyết đến kiểm định thống kê, là một quá trình đi tìm mối liên quan, hay nói theo thuật ngữ dịch tễ học lâm sàng là đi tìm *association* (hoặc có khi để cập đến là *relationship*).

Với ML, chúng ta có thể đi một bước xa hơn association. Bước đó là prediction – tiên lượng (hay dự báo). Với ML, chúng ta có thể tiên lượng biến cố xảy ra trong tương lai. Phân tích association cho phép chúng ta phát biểu chung về mối liên quan giữa X và Y. Phân tích prediction cho phép chúng ta tiên lượng giá trị của Y cho mỗi cá thể (*individual*) nếu chúng ta biết giá trị X của cá thể đó.

Những tiên lượng tương lai có thể đúng, có thể sai. Do đó, ML có hẳn một loạt “thủ thuật” để đánh giá khả năng tiên lượng của một mô hình thống kê. Đối với biến liên tục, cách đánh giá mô hình của ML bao gồm những thủ thuật có tên khó hiểu là “loss function”, nhưng trong thực tế đó là phân tích dao động dư (residual analysis) mà ai học thống kê cũng đều biết. Đối với biến outcome là biến phân nhóm (như có hay không có bệnh), ML tính các chỉ số gọi là “accuracy” trong cái gọi là “confusion matrix”, mà người làm dịch tễ học ai cũng biết đó chính là độ nhạy (sensitivity) và độ đặc hiệu (specificity). Cũng như dịch tễ học, ML cũng dùng AUC để đánh giá mô hình tiên lượng. Nhưng ML chưa/không quan tâm đến calibration và reclassification trong việc đánh giá mô hình tiên lượng.

Những ML có những đóng góp quan trọng trong việc kiểm định mô hình tiên lượng. Trước đây, thống kê học chỉ dừng ở những phương pháp như testing / validation hay bootstrap, nhưng ML cung cấp cho chúng ta hàng loạt phương pháp rất hấp dẫn và cực kì có ích cho việc phân tích dữ liệu lớn. Những phương pháp đó bao gồm random forest, neural network, k-nearest neighbours, partial least squares, boosted generalized additive model, support vector machine (thực chất là hồi qui tuyến tính mở rộng), classification, v.v. Trong thực tế, những phương pháp này cũng xuất phát hoặc xây dựng trên lý thuyết thống kê, nhưng có giao thoa với khoa học máy tính.

Những phương pháp trên rất có ích trong thời đại dữ liệu lớn. Ngày xưa, giới khoa học thường làm nghiên cứu ở qui mô tương đối mà theo đó số biến độc lập (independent variables) ít hơn số cỡ mẫu. Nhưng ngày nay, nhiều nghiên

cứu mà số biến độc lập cao gấp ngàn lần số cỡ mẫu. Chẳng hạn như trong nghiên cứu gồm 100 bệnh nhân và 100 người không bệnh, mỗi cá nhân được phân tích khoảng 500.000 biến thể gen. Đây là vấn đề được gọi tắt là $p \gg n$ hay “curse of dimensionality”. Trong các nghiên cứu này (ngày càng phổ biến), các phương pháp thống kê cổ điển không còn thích hợp nữa, và các phương pháp ML cung cấp cho chúng ta một số giải pháp rất hấp dẫn.

HAI TRƯỜNG PHÁI PHÂN TÍCH DỮ LIỆU

Trong phân tích dữ liệu bằng mô hình thống kê có hai cách tiếp cận, hay cũng có thể nói là hai trường phái chính: ước tính và kiểm định giả thuyết. Hai trường phái này tồn tại song song nhau. Đối với trường phái ước tính (thuật ngữ tiếng Anh là *estimation*), nhà phân tích chỉ quan tâm đến ước tính tham số của mô hình và khoảng tin cậy (KTC) của tham số. Đối với trường phái kiểm định giả thuyết, nhà phân tích quan tâm đến ước tính trị số P của tham số dưới điều kiện giả thuyết vô hiệu là đúng. Trong thực tế, cả hai trường phái tồn tại song song nhau, và bổ sung cho nhau.

1. Trường phái ước tính

Triết lý thống kê hiện đại dựa vào khái niệm *tần số* (frequency) và *lấy mẫu* (sampling). Để biết chiều cao của người Việt, một cách chính xác nhất là đo chiều cao của tất cả dân số Việt Nam, và tính giá trị trung bình. Trong trường hợp này, chúng ta không cần mô hình phân tích thống kê, và cũng không cần suy luận thống kê bởi vì chúng ta đã có trong tay toàn bộ quần thể.

Nhưng đo lường cả một dân số là điều khó khả thi, và cũng không cần thiết. Thay vì đo lường chiều cao tất cả người

Việt, chúng ta có thể chọn một nhóm người một cách ngẫu nhiên, và đo lường chiều cao, tính trung bình và độ lệch chuẩn, và sau đó là dùng các chỉ số thống kê quan sát được (như trung bình và độ lệch chuẩn) để suy luận cho dân số. Lúc đó, chúng ta cần phân tích và suy luận thống kê.

Để hiểu kết quả phân tích tốt hơn, chúng ta cần phải phân biệt hai phạm trù: mẫu và quần thể. *Mẫu* ở đây là sample, là số người chúng ta chọn ngẫu nhiên. *Quần thể* trong thuật ngữ thống kê là *population*. Tuy gọi là *population* nhưng nó không có nghĩa thông thường là dân số. Như vậy, *mẫu* được lấy từ *quần thể*. Dĩ nhiên, chúng ta có thể chọn rất nhiều mẫu từ một quần thể, và phân bố của các chỉ số thống kê (được tính toán từ mẫu) cho phép chúng ta có thể suy luận về phân bố của quần thể.

Phân biệt mẫu và quần thể giúp cho chúng ta hiểu rõ hơn khái niệm thống kê, nhưng một phân biệt khác là các *chỉ số thống kê mẫu* và *thông số* cũng cần thiết. Khi chúng ta lấy mẫu, chúng ta chỉ có thể tính toán những chỉ số thống kê, như trung bình, trung vị, phương sai, độ lệch chuẩn, tỉ lệ hay xác suất, v.v. Tất cả những chỉ số đó (được tính từ mẫu), thuật ngữ tiếng Anh gọi là *statistic*. Thuật ngữ *thông số* hay *parameter* là những chỉ số thống kê liên quan đến quần thể. Như vậy, chúng ta có hai nhóm chỉ số tồn tại song song nhau, một nhóm đề cập đến mẫu (*statistic*), một nhóm liên quan đến quần thể (chúng ta gọi là *parameter*). Theo qui ước chung, kí hiệu của các *statistic* được viết bằng mẫu tự tiếng Latin, và các *thông số* được viết bằng mẫu tự Hi Lạp:

Vài kí hiệu phổ biến cho mẫu và quần thể

Chỉ số	Mẫu (statistic)	Quần thể (parameter)
Trung bình	m	μ
Phương sai	s^2	σ^2
Độ lệch chuẩn	s	σ
Xác suất	p	π
Tỉ số nguy cơ	RR	θ
Tỉ số odds	OR	θ

Ở đây, có lẽ tôi cần giải thích thêm hai thuật ngữ: “estimate” và “estimation”. Tôi tạm dịch *estimate* là ước số, vì estimate là một con số ước tính từ mẫu. Ví dụ như trung bình chiều cao của 25 thanh niên Việt là 167 cm, thì con số 167 được gọi là *estimate*. Nói cách khác, $m = 167$. *Estimation* là qui trình hay cũng có thể gọi là thuật toán để có ước số. Một thuật ngữ khác cũng hay sử dụng là *test statistic*, có nghĩa là một chỉ số kiểm định thống kê tính từ mẫu. Ví dụ như $t = x / \sqrt{s/n}$ thì t được xem là một *test statistic*. Nói cách khác, cũng có thể xem *test statistic* như là một công thức của chỉ số thống kê.

Dĩ nhiên, chúng ta không biết giá trị của các thông số trong quần thể. Nhưng chúng ta biết (vì có thể tính toán) các chỉ số thống kê từ quần thể. Các phương pháp phân tích thống kê cung cấp cho chúng ta những phương tiện để dùng các chỉ số thống kê mà suy đoán [hay suy luận] giá trị của thông số trong quần thể.

Một phần những cái hay và cái đẹp của thống kê là ở đây: dùng giá trị của mẫu để suy luận cho quần thể. Chúng ta không biết chiều cao trung bình của dân số Việt Nam, nhưng chúng ta có thể lấy mẫu và qua ước tính, có thể suy đoán khá chính xác về chiều cao của dân số. Chúng ta không biết thuốc strontium ranelate có hiệu quả giảm gãy xương hay không, nhưng chúng ta có thể làm nghiên cứu trên mẫu và suy luận cho hiệu quả của thuốc. Chúng ta không biết gen LRP5 có liên quan đến mật độ xương hay không, nhưng chúng ta có thể nghiên cứu trên một số đối tượng và suy luận về mối liên quan.

Trong trường phái ước tính, nhà phân tích quan tâm đến ước tính chỉ số thống kê và sự bất định của nó. Chẳng hạn như để ước tính giá trị trung bình chiều cao, chúng ta có thể làm như sau:

- Lấy mẫu ngẫu nhiên gồm n người, và đo chiều cao;
- Dùng phương pháp bình phương tối thiểu (least square method) hay phương pháp tối đa khả dĩ (maximum likelihood) để ước tính giá trị trung bình m ;
- Ước tính phương sai độ lệch chuẩn trong mẫu;

Dùng các chỉ số trên để tính toán khoảng tin cậy, và dùng khoảng tin cậy để suy luận cho thông số μ trong quần thể (dân số).

Một trong những chỉ số quan trọng của phân tích mẫu là khoảng tin cậy (KTC), thuật ngữ tiếng Anh gọi là *confidence interval*. Khoảng tin cậy có thể tính cho xác suất 90, 95 hay 99%. Nhưng thông thường người ta hay lấy KTC 95% làm chuẩn để suy luận về thông số.

Chẳng hạn như qua nghiên cứu trên 100 người, và chúng ta có trung bình chiều cao là $m = 167$ cm, và độ lệch chuẩn

$s = 6$ cm. Giả định rằng chiều cao tuân theo luật phân bố chuẩn, chúng ta có thể ước tính khoảng tin cậy 95% như sau:

$$m \pm 1,96 \times s = 167 \pm 1,96 \times 6$$

dao động trong khoảng 155,2 đến 178,8 cm. Nói cách khác, chúng ta có thể suy luận rằng 95% chiều cao của 100 người đó dao động trong khoảng 155,2 đến 178,8 cm.

Nhưng còn suy luận cho quần thể (thông số)? Xin nhắc lại rằng chúng ta không biết giá trị trung bình chiều cao của dân số Việt; chúng ta chỉ có thể suy luận về chiều cao quần thể qua các chỉ số thống kê số tính toán từ mẫu. Một cách để suy luận về thông số là tính KTC của thông số đó. Khoảng tin cậy 95% của trung bình có thể ước tính từ độ lệch chuẩn và số cỡ mẫu qua công thức:

$$m \pm 1,96 \left(\frac{s}{\sqrt{n}} \right) = 167 \pm 1,96 \left(\frac{6}{\sqrt{100}} \right)$$

Trong đó, $s / \sqrt{n}$ được gọi là *sai số chuẩn* (standard error). Sai số chuẩn, trong thực tế, là độ lệch chuẩn của các số trung bình mẫu. Thử tưởng tượng, nếu chúng ta lấy k mẫu, và mỗi mẫu chúng ta tính số trung bình, như vậy chúng ta có tập hợp k số trung bình mẫu: $x_1, x_2, x_3, \dots, x_k$. Độ lệch chuẩn của $x_1, x_2, x_3, \dots, x_k$ được gọi là *sai số chuẩn*.

Trong ví dụ trên, chúng ta có KTC95% của số trung bình là 165,8 đến 168,2 cm. Con số KTC95% này có nghĩa là nếu chúng ta lặp lại nghiên cứu rất nhiều lần, và mỗi lần lấy mẫu 100 người, thì 95% các nghiên cứu đó sẽ cho ra số trung bình mẫu trong khoảng 165,8 đến 168,2 cm.

Một cách diễn giải khác, tuy không chính xác lắm nhưng lại được sử dụng phổ biến, là có 95% xác suất mà thông số μ của quần thể dao động trong khoảng 165,8 đến 168,2 cm.

Chúng ta không biết chính xác μ là gì, nhưng rất có thể là 167 cm (trung bình mẫu), và cũng có thể dao động trong khoảng 165,8 đến 168,2 cm với xác suất 95%. Thực tế cách diễn giải này chỉ đúng với khoảng tin cậy (95% credible interval) ước tính trong trường phái Bayes (xem thêm phần 3 dưới đây).

Các phương pháp phân tích dữ liệu theo trường phái ước tính, như mô tả trên, tập trung vào việc ước tính các chỉ số thống kê mẫu, và khoảng tin cậy, và dùng các chỉ số này để suy đoán cho giá trị của quần thể.

2. Trường phái kiểm định giả thuyết

Một trường phái khác trong phân tích thống kê là kiểm định giả thuyết (test of hypothesis). Đây là trường phái được Ronald A. Fisher, cha đẻ của khoa học thống kê hiện đại, đề xướng và quảng bá rất thành công. Phương pháp kiểm định giả thuyết chịu ảnh hưởng của triết lý phản nghiệm (falsificationism) của Karl Popper. Xin nói thêm rằng Popper là một nhà triết học khoa học rất nổi tiếng và có ảnh hưởng lớn đến khoa học. Theo Popper, chúng ta không thể nào chứng minh một giả thuyết là đúng hay sai; chúng ta chỉ có thể bác bỏ (falsify) giả thuyết với dữ liệu thí nghiệm, và khoa học tiến bộ nhờ bác bỏ những giả thuyết không phù hợp với dữ liệu thực tế.

Có lẽ chịu ảnh hưởng bởi ý tưởng của Popper, nên Fisher đề ra phương pháp “**Test of Significance**” (tôi tạm dịch là: **phương pháp kiểm định ý nghĩa thống kê**). Theo Fisher, thống kê là một phương pháp suy luận theo phép qui nạp (inductive inference), tức là phương pháp suy luận dựa vào quan sát từ các mẫu (sample) và khái quát cho một quần thể

(population). Phương pháp kiểm định ý nghĩa thống kê được tiến hành theo 3 bước như sau:

Bước 1, phát biểu một giả thuyết vô hiệu (null hypothesis).

Giả thuyết vô hiệu là giả thuyết ngược lại với giả thuyết mà nhà nghiên cứu muốn kiểm định. Chẳng hạn như nếu giả thuyết điều trị bằng thuốc Ramipril làm giảm nguy cơ tử vong, thì giả thuyết vô hiệu sẽ phát biểu là “tỉ lệ tử vong ở bệnh nhân được điều trị bằng Ramipril bằng với nhóm giả dược. Gọi giả thuyết vô hiệu là H_0 .”

Bước 2, thu thập dữ liệu liên quan đến giả thuyết. Hai chữ “thu thập” ở đây nghe qua có lẽ là quá đơn giản, nhưng trong thực tế đây là bước quan trọng nhất. Thu thập dữ liệu phải qua thí nghiệm hoặc nghiên cứu, và nghiên cứu phải được thiết kế một cách rất cẩn thận. Trong ví dụ trên, số liệu sẽ là số trường hợp tử vong. Sau đó, tính toán một chỉ số thống kê (như t-test, ki bình phương, tỉ số odds, v.v.), gọi chung chỉ số này là D .

Bước 3, ước tính xác suất quan sát dữ liệu D nếu giả thuyết H_0 đúng. Nói cách khác và viết theo ngôn ngữ xác suất, bước này chúng ta ước tính $P(D | H_0)$. Đây chính là **trị số P (P-value)**.

Fisher đề nghị dùng trị số P để bác bỏ hay chấp nhận giả thuyết vô hiệu. Nên nhớ rằng P là xác suất dữ liệu xảy ra nếu giả thuyết vô hiệu đúng; do đó, nếu P thấp thì đó là chứng cứ để bác bỏ giả thuyết vô hiệu. Thấp như thế nào? Fisher đề nghị rằng nếu trị số P thấp hơn 0,05 thì giả thuyết vô hiệu không phù hợp với dữ liệu quan sát được, và do đó phải bác bỏ. Đối với Fisher, không có chuyện “bác bỏ giả thuyết” hay “chứng minh giả thuyết” mà chỉ có số liệu có phù hợp, có nhất quán với giả thuyết hay không mà thôi. Cần nói thêm rằng Fisher

để nghị báo cáo trị số P một cách chính xác, tức là không có những cách viết như $P < 0,05$ hay $P > 0,01$ mà phải là $P = 0,043$ hay $P = 0,002$.

Mô hình kiểm định giả thuyết

Jerzy Neyman (1894 - 1981) là một nhà toán học xuất sắc gốc Ba Lan và Egon Pearson (1895 - 1980) là một nhà thống kê học (con của giáo sư Karl Pearson, cha đẻ của lý thuyết Chi-square và hệ số tương quan) cùng lúc với Fisher, phát triển một phương pháp rất khác với Fisher, mà hai ông gọi là **Test of Hypothesis (Kiểm định giả thuyết)**. Neyman và Pearson bác bỏ khái niệm suy luận theo qui nạp; hai ông nghĩ rằng thống kê học là một phương pháp hay cơ chế để hướng dẫn chúng ta đi đến một quyết định đúng về lâu về dài. Nói cách khác, Neyman và Pearson cho rằng phương pháp của Fisher vô nghĩa!

Một cách đơn giản, mô hình kiểm định giả thuyết của Neyman và Pearson có thể thực hiện qua các bước như sau:

Bước 1, phát biểu giả thuyết chính (H_1) và giả thuyết vô hiệu (H_0). Ví dụ để kiểm định giả thuyết thuốc có hiệu quả giảm tử vong, chúng ta có thể phát biểu 2 giả thuyết như sau:

H_0 : thuốc không có hiệu quả

H_1 : thuốc có hiệu quả

Nếu nghiên cứu có hai nhóm bệnh nhân (một nhóm được điều trị bằng thuốc A và một nhóm không được điều trị bằng thuốc A), và nếu tỉ lệ tử vong của hai nhóm là π_1 và π_2 , chúng ta có thể phát biểu 2 giả thuyết rằng:

$H_0: \pi_1 = \pi_2$

$H_1: \pi_1 \neq \pi_2$

Hay, nếu chúng ta gọi hiệu số của hai tỉ lệ là $\delta = \pi_1 - \pi_2$, thì giả thuyết có thể viết:

$$H_0: \delta = 0$$

$$H_1: \delta \neq 0$$

Bước 2, quyết định mức độ α và β có thể chấp nhận được và ước tính cỡ mẫu cần thiết. α là xác suất bác bỏ giả thuyết H_1 nhưng đó là giả thuyết đúng. β là xác suất bác bỏ H_0 trong khi H_0 đúng.

Bước 3, thu thập dữ liệu liên quan đến giả thuyết. Như đề cập trên, thu thập dữ liệu để cập ở đây chính là thực hiện thí nghiệm hay nghiên cứu. Một khía cạnh rất quan trọng trong bước 3 là xác định số đối tượng nghiên cứu cần thiết (sample size) dựa vào α và β , và mức độ ảnh hưởng (tức δ).

Bước 4, nếu dữ liệu nằm trong khoảng bác bỏ giả thuyết H_0 , thì chấp nhận giả thuyết H_1 ; nếu không thì chấp nhận giả thuyết H_0 . Chú ý rằng “chấp nhận” một giả thuyết không có nghĩa là chúng ta tin vào giả thuyết đó, mà chỉ có nghĩa là chúng ta hành động với điều kiện đó là giả thuyết đúng.

Nguyên lý của mô hình Neyman và Pearson là chúng ta dựa vào dữ liệu để chọn một giả thuyết sao cho về lâu về dài chúng ta không quá sai. Chính vì thế mà ngày nay chúng ta thường chọn $\alpha = 5\%$ và $\beta = 10\%$ đến 20% .

Fisher bác bỏ hoàn toàn mô hình của Neyman và Pearson. Ông cho rằng đó là một mô hình... vô duyên. Fisher nhạo báng rằng các nhà toán học (ám chỉ Neyman và Pearson) “chẳng hiểu gì về thực nghiệm và để ra một mô hình quá phi thực tế”.

Trong những năm sau đó (thập niên 1930s) cộng đồng thống kê học chứng kiến một cuộc tranh luận dai dẳng và đôi khi nóng bỏng giữa Fisher và Neyman-Pearson trên các tập san thống kê học ở Anh.

Trở trêu thay, mấy mươi năm sau, hai mô hình của Fisher và Neyman-Pearson được “hun đúc” thành một mô hình tổng hợp mà chúng ta ứng dụng ngày nay trong nghiên cứu y học. Mô hình này sử dụng kết quả kiểm định thống kê của Fisher để đi đến quyết định chấp nhận hay bác bỏ giả thuyết vô hiệu H_0 hay giả thuyết chính H_1 theo mô hình của Neyman và Pearson. Tiêu biểu cho mô hình này là nghiên cứu lâm sàng đối chứng ngẫu nhiên (randomized controlled clinical trial hay RCT). Theo đó, một nghiên cứu lâm sàng được tiến hành theo các bước như sau:

Bước 1, định nghĩa một giả thuyết vô hiệu và một giả thuyết chính. Thí dụ trong một nghiên cứu lâm sàng, gồm hai nhóm bệnh nhân: một nhóm được điều trị bằng thuốc A, và một nhóm được điều trị bằng placebo, nhà nghiên cứu có thể phát biểu giả thuyết vô hiệu rằng độ hiệu nghiệm thuốc A tương đương với placebo.

Bước 2, xác định xác suất α (còn gọi là sai số loại I) và β (còn gọi là sai số loại II), và ước tính cỡ mẫu dựa vào hai xác suất này.

Bước 3, tiến hành thí nghiệm để thu thập dữ liệu liên quan đến giả thuyết. Dữ liệu ở đây có nghĩa là các chỉ số thống kê mẫu như t-test, ki bình phương, tỉ số nguy cơ, v.v.. Gọi dữ liệu là D .

Bước 4, sử dụng phương pháp kiểm định ý nghĩa thống kê của Fisher ước tính xác suất $P(D | H_0)$. Gọi trị số này là P .

Bước 5, nếu $P < \alpha$, bác bỏ giả thuyết H_0 . Chú ý, bác bỏ H_0 không có nghĩa là chúng ta chấp nhận giả thuyết H_1 .

Có thể minh họa cho các bước trên bằng một ví dụ về nghiên cứu hiệu quả của thuốc Ramipril trong việc phòng chống tử vong và đột quỵ. Với giả thuyết rằng thuốc có hiệu nghiệm giảm nguy cơ tử vong và đột quỵ, các nhà nghiên cứu so sánh tỉ lệ tử vong và đột quỵ giữa hai nhóm bệnh nhân: nhóm 1 được điều trị bằng Ramipril và nhóm 2 là nhóm giả dược (placebo). Bắt đầu bằng cách xác định $\alpha = 0,05$ và $\beta = 0,80$. Các nhà nghiên cứu ước tính số lượng bệnh nhân cần thiết. Sau ba năm thu thập số liệu, kết quả có thể tóm lược trong bảng số liệu sau đây:

Hiệu quả của Ramipril giảm nguy cơ nhồi máu cơ tim, đột quỵ, tử vong và tiểu đường thể hiện qua tỉ số nguy cơ và trị số P

	Tỉ số nguy cơ (relative risk) và khoảng tin cậy 95%	Trị số P
Nhồi máu cơ tim, đột quỵ, tử vong	0,83 (0,75 – 0,91)	0,0002
Tỉ lệ mới mắc bệnh tiểu đường	0,69 (0,56 – 0,85)	0,0006

Nguồn: HOPE/HOPE-TOO Study Investigators. Long-term effects of Ramipril on cardiovascular events and diabetes. Results of the HOPE Study Extension. Circulation 2005; 112:1339-1346.

Bởi vì trị số P thấp hơn mức α (0,05) mà các nhà nghiên cứu đề ra từ lúc đầu (trước khi thu thập số liệu); cho nên, các nhà nghiên cứu kết luận rằng sự khác biệt về tỉ lệ tử vong và đột quỵ giữa hai nhóm có ý nghĩa thống kê. Tất nhiên, *trị số P*

trên không có nghĩa là nghiên cứu đã chứng minh rằng thuốc Ramipril có hiệu quả giảm nguy cơ tử vong và đột quỵ. Nó có nghĩa là nếu thật sự thuốc Ramipril không có hiệu quả giảm nguy cơ tử vong và đột quỵ thì xác suất mà các nhà nghiên cứu quan sát các số liệu trên là 0,0002.

3. Trường phái Bayes

Ở trên, tôi nói có hai trường phái chính trong phân tích dữ liệu, nhưng trong thực tế còn có một trường phái khác: Bayes. Thật ra, trường phái Bayes không phải là một cái gì mới, vì nó đã tồn tại từ thế kỉ 18 và thịnh hành trong thế kỉ 19. Đến thế kỉ 20 thì trường phái Bayes bị làm cho lu mờ bởi trường phái ước tính và kiểm định giả thuyết. Nhưng với những khiếm khuyết của trị số P cùng vấn đề ước tính, trường phái Bayes đang rất thịnh hành. Cộng đồng thống kê dự đoán rằng trong thế kỉ 21 trường phái Bayes sẽ trở thành phổ biến hơn như thời “huy hoàng” của nó vào thế kỉ 19.

Trường phái Bayes do linh mục Thomas Bayes đề xuất vào năm 1763. Linh mục Bayes thường được xem là một nhà toán học tài tử. Suốt cuộc đời ông chỉ có một bài báo duy nhất, và bài báo chỉ công bố sau khi ông đã qua đời. Sở dĩ có tình trạng đó là vì ông là người rất cẩn thận, sợ công bố sớm mà có sai sót thì sau này rất khó nói. Tuy nhiên, bài báo sau khi được công bố đã đặt nền móng cho một trường phái thống kê hoàn toàn mới: *Bayesian statistics*. Chỉ một công trình mà ông đã làm thay đổi cả thế giới khoa học, thay đổi cách suy nghĩ về sự bất định trong khoa học, và chỉ ra một phương pháp suy luận hoàn toàn logic. Ngày nay, phương pháp Bayes được ứng dụng trong hầu hết tất cả lĩnh vực khoa học, kể

cả trong công nghệ thông tin (ứng dụng Bayes trong việc ngăn chặn những thư rác điện tử), tiên lượng kinh tế, phân tích các mối liên hệ xã hội, và lí giải qui trình suy nghĩ của con người. Ngày nay, suy luận theo trường phái Bayes được nhắc đến trên báo chí đại chúng chứ không chỉ trong báo khoa học. Những tờ báo lớn như *New York Times*, *Economist*, *Guardian*, v.v. đều thường xuyên nhắc đến phương pháp suy luận Bayes.

Suy luận Bayes dựa vào định lí Bayes (Bayesian Theorem). Có thể phát biểu định lí Bayes theo ngôn ngữ hàng ngày như sau: những gì chúng ta *biết* là tổng hợp những gì chúng ta đã biết cộng với *chứng cứ* thực tế. Có thể nói rằng Định lí Bayes thể hiện cách suy nghĩ rất phổ biến của tất cả chúng ta: đó là chúng ta tiếp thu kiến thức một cách tích lũy.

Để minh họa cho ý tưởng trên, chúng ta thử xem qua ví dụ sau đây. Giả dụ rằng chúng ta có số liệu về một hormone (tính bằng mmol/L) của 10 người như sau:

156, 158, 160, 162, 164, 166, 168, 170, 172, 175

Với số liệu trên, chúng ta có thể tính trị số trung bình là 165,1 mmol/L.

Nhưng nếu chúng ta biết rằng số liệu đó được thu thập trên 670 người, và trong đó có 20 người với giá trị 156 mmol/L, 30 người có giá trị 158 mmol/L, 50 người với giá trị 160 mmol/L, v.v.. (xem bảng dưới đây). Thì số trung bình bây giờ chắc chắn không phải là 165,1 mmol/L, vì nó còn tùy thuộc vào thông tin về số đối tượng cho từng giá trị hormone ghi nhận.

Giá trị hormone (mmol/L)	Số người	Phần trăm (Xác suất)
156	20	0,030
158	30	0,045
160	50	0,075
162	100	0,149
164	150	0,224
166	130	0,194
168	100	0,149
170	50	0,075
172	30	0,045
175	10	0,015
Tổng	670	1,000

Nếu chúng ta diễn tả số người bằng xác suất (cột 3 trong bảng trên), thì xác suất đó có thể xem như là một trọng số. Bây giờ, giá trị trung bình có thể tính như là tổng tích số của giá trị và trọng số:

$$(156 \times 0,030) + (158 \times 0,045) + \dots + (175 \times 0,015) = 164,85$$

Giá trị hormone (x_i)	Số người	Xác suất (p_i)	$x_i \times p_i$
156	20	0,030	4,66
158	30	0,045	7,07
160	50	0,075	11,94
162	100	0,149	24,18

164	150	0,224	36,72
166	130	0,194	32,21
168	100	0,149	25,07
170	50	0,075	12,69
172	30	0,045	7,70
175	10	0,015	2,61
Tổng	670	1,000	164,85

Như vậy, giá trị trung bình bây giờ chính xác là 164,85 mmol/L, chứ không phải 165,1 mmol/L.

Ví dụ trên đây cho thấy kết quả phân tích có thể thay đổi khi chúng ta có thông tin. Trong trường hợp 1, chúng ta không có thông tin nào khác là 10 giá trị, và chỉ có thể dựa vào đó để tính giá trị trung bình. Nhưng trong trường hợp 2, chúng ta có thông tin tiền định về phân bố của hormone trong cộng đồng, nên số trung bình chính xác hơn.

Suy luận Bayes có thể phát biểu một cách đơn giản như sau:

$$[\text{Thông tin đang có}] \times [\text{Thông tin trước đây}] = [\text{Thông tin mới}]$$

Trong ví dụ trên, thông tin đang có là 10 giá trị của hormone; thông tin trước đây chính là xác suất phân bố của từng giá trị; và thông tin mới là tích số của hai thông tin đang có và thông tin trước đây.

Như nói trên, suy luận Bayes dựa vào định lý Bayes. Định lý Bayes dĩ nhiên cũng có thể mô tả một cách đơn giản qua xác suất. Gọi H là giả thuyết, và D là dữ liệu thực tế, Định lý Bayes phát biểu rằng xác suất H với điều kiện D xảy ra – kí hiệu $P(H | D)$ – là:

$$P(H | D) = \frac{P(D | H) \times P(H)}{P(D)}$$

Trong đó

- $P(H)$ là xác suất giả thuyết trước khi làm thí nghiệm; và
- $P(D | H)$ là xác suất dữ liệu xảy ra với điều kiện giả thuyết H là đúng;
- $P(D)$ là phân bố của dữ liệu, có thể xem là một hằng số.

Thay vì dùng H như là một giả thuyết, chúng ta có thể thay H bằng θ như là một thông số chúng ta muốn biết. Trường phái cổ điển hay frequentist (tần số) xem θ là cố định, nhưng trường phái Bayes xem θ có một phân bố riêng, và do đó, không cố định. Chúng ta cũng có thể viết theo ngôn ngữ của hàm số:

$$f(\theta | D) = \frac{f(D | \theta) f(\theta)}{f(D)}$$

hay

$$f(\theta | D) \propto f(D | \theta) \times f(\theta)$$

Một cách khác để phát biểu Định lý Bayes là:

$$\begin{aligned} & \text{Likelihood} \times \text{Prior Probability of Hypothesis} \\ &= \text{Posterior Probability of Hypothesis} \end{aligned}$$

Trong đó,

- Likelihood = $P(D | H)$, hay $f(D | \theta)$, tức phân bố của dữ liệu;
- Prior Probability of Hypothesis = $P(H)$ hay $f(\theta)$, có thể xem là **xác suất tiên định**;
- Posterior Probability of Hypothesis = $P(H | D)$, hay $f(\theta | D)$, có thể xem là **xác suất hậu định**.

Nhìn qua định lý trên, chúng ta thấy suy luận Bayes có 3 thông tin. Thông tin thứ nhất là thông tin mà chúng ta muốn biết, thuật ngữ tiếng Anh gọi là *posterior information* – thông tin hậu định. Thông tin thứ hai là thông tin chúng ta đã biết, tiếng Anh là *prior information* – thông tin tiền định. Và, thông tin thứ ba là thông tin thực tế, thuật ngữ tiếng Anh là *likelihood*. Ở đây, “thông tin” có nghĩa là xác suất hay phân bố.

Nguyên lý của phương pháp Bayes xem các thông số là bất định. (Điều này phải phân biệt với các phương pháp tần số xem các thông số là cố định, bất biến). Chính vì quan điểm này, đối với Bayes, mỗi thông số có một phân bố gắn liền với nó. Vì có phân bố, chúng ta có thể ước tính xác suất cho bất cứ giá trị nào của thông số. Chính khả năng này làm cho phương pháp Bayes trở nên hấp dẫn trong cái nhìn của các nhà khoa học thực nghiệm.

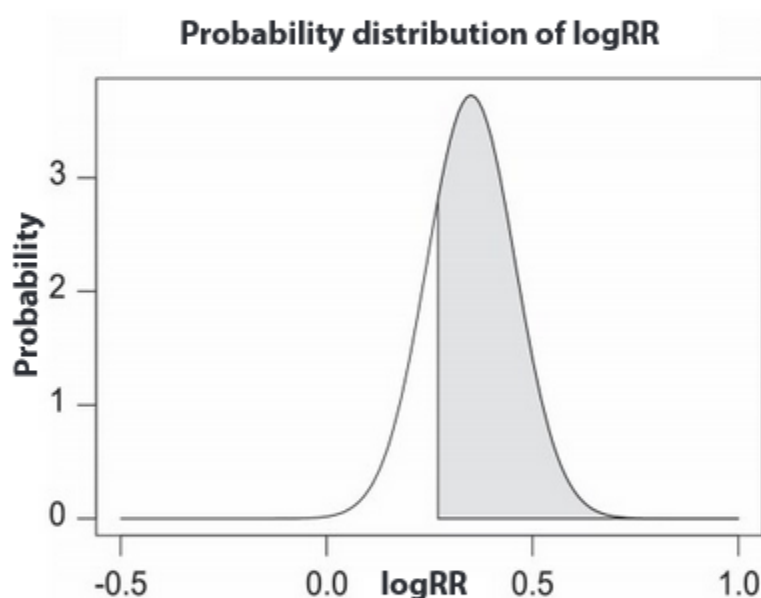
Chẳng hạn như một nghiên cứu theo dõi 23.000 đàn ông người Thụy Điển (Jonsson et al 2002) báo cáo rằng nguy cơ tử vong ở những người có BMI ≥ 30 (tức béo phì) cao hơn những người không béo phì 1,42 lần (với khoảng tin cậy 95% dao động từ 1,15 đến 1,75). Nói theo ngôn ngữ thống kê, $RR = 1,42$ (KTC95: 1,15 đến 1,75). Chúng ta muốn đặt câu hỏi xác suất mà $RR > 1,30$ là bao nhiêu. Phương pháp Bayes có thể trả lời câu hỏi này.

Giả định rằng $\log(RR)$ tuân theo luật bố chuẩn, và xác suất tiền định cho mỗi giá trị của $\log(RR)$ là tương đương nhau (còn gọi là *uniform prior distribution*), chúng ta có thể ước tính số trung bình của $\log(RR)$:

- $\log RR = \log(1,42) = 0,351$

- và độ lệch chuẩn:
- $s = \log(1,75 / 1,15) / 3,92 = 0,107$

Xác suất $P(RR > 1,30)$ có nghĩa là $P(\log RR > 0,2624)$, và kết quả là 0,79. Như vậy, với kết quả trình bày và nếu chúng ta không có bất cứ thông tin tiên định nào về phân bố RR, chúng ta có thể nói rằng xác suất mà $RR > 1,3$ là gần 80% (xem hình 1).



Như chúng ta thấy, diện tích dưới đường phân bố chuẩn chính là xác suất hậu định. Do đó, chúng ta có thể ước tính xác suất của bất cứ giá trị của thông số qua phân bố hậu định, và đó là một điểm rất có ích của phương pháp Bayes trong thực tế.

* * *

Nói tóm lại, các phương pháp phân tích thống kê có thể xếp vào hai nhóm: trường phái ước tính và trường phái kiểm định giả thuyết. Hai trường phái hay 2 nhóm phương pháp này thường bổ sung cho nhau trong thực tế. Chẳng hạn như nếu khoảng tin cậy 95% không bao gồm thông số vô hiệu thì chúng ta có thể đoán rằng trị số $P < 0,05$. Nhưng khoảng tin cậy 95% cung cấp nhiều thông tin hơn trị số P . Khoảng tin cậy 95% cho chúng ta biết giá trị khả dĩ của thông số (mức độ ảnh hưởng), và qua đó chúng ta có vài ý niệm về tầm quan trọng của vấn đề. Trị số P chỉ cho chúng ta biết thông số có hay không có ý nghĩa thống kê, chứ không nói lên tầm quan trọng của ảnh hưởng. Một nghiên cứu với cỡ mẫu rất lớn có thể cho ra trị số $P < 0,05$ (tức có ý nghĩa thống kê) nhưng chẳng có ý nghĩa gì trong thực tế. Ngược lại, một nghiên cứu không có đủ số đối tượng có thể cho ra kết quả không có ý nghĩa thống kê ($P > 0,05$) nhưng mức độ ảnh hưởng có thể có ý nghĩa thực tế. Và lại, ngưỡng để định nghĩa có hay không có ý nghĩa thống kê, như 0,05, là hoàn toàn mang tính tùy tiện chứ hoàn toàn không dựa trên một cơ sở khoa học nào cả. Do đó, không nên quá lệ thuộc vào trị số P để kết luận. Để kết luận về một vấn đề, nhà khoa học phải đặt kết quả phân tích trong bối cảnh chứ không chỉ đơn thuần dựa vào các chỉ số thống kê.

Ngoài hai trường phái trên, còn có một trường phái “mới” càng ngày càng phổ biến: đó là trường phái Bayes. Trường phái Bayes diễn giải xác suất hoàn toàn khác với trường phái tần số (như ước tính và kiểm định giả thuyết). Theo cách hiểu phổ biến (và cổ điển), xác suất một biến cố được định nghĩa là tần số mà biến cố đó xảy ra về lâu dài. Vào thế kỉ 18, Bá tước Bouffon ném đồng xu 4040 lần, và ông đếm rằng có 2048 lần,

bề mặt của đồng xu xuất hiện, và xác suất được hiểu là 50,67%. Cách hiểu xác suất theo trường phái tần số có nhiều vấn đề và thường bị phê bình. Một cách hiểu xác suất khác theo trường phái Bayes là cách hiểu chủ quan (subjective probability). Theo cách hiểu Bayes, xác suất phản ánh mức độ tin tưởng (degree of belief). Xác suất ông Bill Clinton sẽ thắng cử là bao nhiêu? Trường phái xác suất theo tần số không thể trả lời câu hỏi này, vì không nước nào tổ chức hàng trăm cuộc bầu cử tổng thống cho ông Clinton. Trong trường hợp này, người theo trường phái Bayes có thể phát biểu: xác suất ông Clinton đắc cử là x phần trăm, và xác suất đó phản ánh mức độ tin tưởng của người phát biểu.

Do đó, trường phái Bayes cung cấp cho chúng ta phương tiện để đặt câu hỏi rất thiết thực. Chẳng hạn như chúng ta có thể đặt câu hỏi “xác suất mà thuốc giảm nguy cơ tử vong ít nhất là 30% là bao nhiêu”. Dĩ nhiên, câu trả lời phụ thuộc vào phân bố tiên định và dữ liệu thực tế. Với suy nghĩ và xem xét cẩn thận chứng cứ, câu trả lời có thể cho chúng ta những câu trả lời đáng tin cậy. Chính vì thế, tôi thường hay nói rằng các phương pháp phân tích thống kê là phương tiện giúp cho chúng ta có suy nghĩ mới hấp dẫn hơn và thiết thực hơn.

TRỊ SỐ P TRONG NGHIÊN CỨU KHOA HỌC

Bất cứ ai làm nghiên cứu khoa học cũng đều “đối diện” với trị số P. (Viết hoa, chữ nghiêng theo qui ước chung). Trị số P có lẽ là con số có ảnh hưởng lớn nhất đến hoạt động khoa học trong lịch sử cận đại. Nhưng rất nhiều người trong giới khoa học không hiểu hay hiểu sai ý nghĩa của trị số P. Chương này sẽ đi qua lịch sử và ý nghĩa thật của trị số P trong nghiên cứu khoa học.

1. Lịch sử của trị số P

Lịch sử sử dụng khái niệm xác suất trong kiểm định giả thuyết khoa học có thể bắt nguồn từ thế kỉ 18. Tài liệu xưa nhất về trị số P xuất hiện vào năm 1710, trong một bài báo của bác sĩ John Arbuthnot công bố trên một tập san của Royal Society (giống như viện hàn lâm khoa học của Anh). Trong bài báo đó, Arbuthnot trình bày dữ liệu thu thập từ tên của những trẻ sơ sinh lúc rửa tội thuộc thành phố London trong thời gian 1629 - 1710, ông quan sát rằng trong bất cứ năm nào, số bé trai đều nhiều hơn số bé gái. Ông lí giải rằng nếu kết quả sinh trai gái hoàn toàn ngẫu nhiên, thì mỗi năm sẽ có 50% bé trai và 50% bé gái (hay rất gần với 50%). Nói theo ngôn ngữ thời nay, đây chính là giả thuyết vô hiệu. Ông lí giải thêm rằng nếu giả thuyết vô hiệu là đúng thì xác suất ông quan sát suốt

thời gian 82 năm là $5^{82} = 2 \times 10^{-25}$, tức là rất rất thấp (1). Nhưng trong thực tế thì năm nào số bé trai cũng nhiều hơn số bé gái, và vì xác suất vô hiệu xảy ra quá thấp, nên ông nghĩ rằng những gì ông quan sát không phải ngẫu nhiên. Nếu không phải ngẫu nhiên thì ắt phải có một yếu tố nào đó giải thích, và ông cho rằng đó là do ‘nghệ thuật chứ không phải ngẫu nhiên’ (“it follows that it is art, not Chance, that governs”). Nghệ thuật ở đây theo ý ông là do... thần thánh (hay Thượng đế) đã ưu tiên cho nam giới hơn là nữ giới!

Lịch sử khoa học thống kê hiện đại có lẽ khởi xướng từ 1900. Trong những năm đầu thế kỉ 20, Karl Pearson (được xem là “hoàng tử” của khoa học thống kê) phát kiến phương pháp kiểm định Kì bình phương. Ông là người đầu tiên lập bộ môn thống kê học ở Anh. Ông dùng phương pháp Kì bình phương để kiểm định dữ liệu từ 26.306 lượt ném súc sắc, và ghi nhận số lần con số 5 và 6 xuất hiện trong thực tế so với số lần xảy ra dựa trên luật phân bố xác suất đồng dạng (uniform distribution); kết quả tính toán ông có xác suất 0,000016. Dựa vào kết quả này, Pearson kết luận rằng đồng súc sắc lệch về con số cao (2).

Cũng vào năm 1900, công trình nghiên cứu của một tu sĩ người Áo là Gregor Mendel tuy không dùng kiểm định thống kê, nhưng kết quả đó đã tạo nên một trào lưu dùng thống kê học để kiểm định giả thuyết di truyền. Chính từ công trình này và những đóng góp của giới thống kê học mà sau này đã cho ra đời một bộ môn mới về khoa học di truyền (2).

Nhưng phải đến 1925 thì thống kê học mới thật sự phát triển, và trị số P bắt đầu “khởi sắc” từ đó. Năm 1925, nhà thống kê học và di truyền học nổi tiếng Ronald A. Fisher xuất bản

cuốn sách “*Statistical Methods for Research Workers*” (SMRW), được giới khoa học đánh giá là một sự kiện mang tính dấu ấn (landmark) trong lịch sử khoa học, chứ không phải chỉ trong thống kê học. Cuốn sách hay đúng hơn là công trình của Fisher đã quảng bá việc sử dụng trị số P trong nghiên cứu khoa học và suy luận khoa học.

Kể từ đó đến nay, trị số P đã được sử dụng trong tất cả các chuyên ngành khoa học. Từ khoa học thực nghiệm đến khoa học cơ bản, từ nghiên cứu văn học đến xã hội học, tất cả đều, không nhiều thì ít, dùng trị số P để đánh giá một sự phát hiện hay một mối liên quan. Trong y sinh học, có thể nói rằng tất cả các khám phá từ gen đến thuốc đều dựa vào trị số P. Bất cứ khám phá nào có trị số $P < 0,05$ được xem là “có ý nghĩa thống kê”, còn $P > 0,05$ thì thường được xem là một kết quả ‘âm tính’. Ngưỡng 0,05 đã đi vào khoa học như là một ‘huyền thoại’, và nó ảnh hưởng đến hầu hết hoạt động nghiên cứu khoa học.

Nhưng rất tiếc là rất nhiều người hiểu lầm hay hiểu không đúng về ý nghĩa của trị số P, và dẫn đến nhiều kết luận sai trong nghiên cứu khoa học. Sử dụng sai hay lệ thuộc vào trị số P trong khoa học được xem là một trong những nguyên nhân dẫn đến tình trạng bất lập trong khoa học (irreproducibility), tức là rất nhiều kết quả nghiên cứu không được tái lập trong khoa học (3-5). Sự bất lập trong khoa học là một tình trạng khủng hoảng trong khoa học hiện nay. Do đó, chúng ta cần phải điểm qua ý nghĩa thật của trị số P. Tình trạng hiểu sai và ứng dụng sai trị số P nghiêm trọng đến nỗi Hiệp hội Thống kê học Hoa Kỳ (ASA) phải ra một tuyên bố về ý nghĩa và mục đích của trị số P (6).

2. Trị số P là một chỉ số ngạc nhiên

Để hiểu ý nghĩa thật của trị số P, tôi muốn giới thiệu đến các bạn em học sinh Hồ Nguyễn Đức Tài và hoa hậu Đặng Thu Thảo. Đức Tài là một học sinh trung học (năm 2013) ở Sài Gòn. Lúc đó, tuy mới 18 tuổi nhưng chiều cao của em là 204 cm. Báo chí giật cái tít rất hấp dẫn “*Choáng với chiều cao ‘cực khủng’ của nam sinh lớp 12*”. Câu hỏi đặt ra là một thiếu niên 18 tuổi với chiều cao 2,04 m như thế có đáng ngạc nhiên không?

Đặng Thu Thảo là một sinh viên 21 tuổi ở Cần Thơ, đăng quang Hoa hậu Việt Nam năm 2012. Thu Thảo có chiều cao là 173 cm, và 3 tham số cơ thể 83-60-90 cm. Ở đây chúng ta chưa quan tâm đến 3 tham số đó, mà chỉ quan tâm đến chiều cao. Câu hỏi đặt ra là chiều cao của một thanh nữ tuổi 21 như em ấy có đáng ngạc nhiên không?

Khi nói đến “ngạc nhiên”, chúng ta ngầm ý phải so sánh với một **khung tham chiếu**. Nếu so sánh chiều cao với chiều cao của người phương Tây thì có lẽ chiều cao của hai em Đức Tài và Thu Thảo chưa hẳn là đáng ngạc nhiên. Nhưng nếu so sánh với chiều cao của người Việt thì có thể hai em ấy có chiều cao đáng chú ý. Do đó, ngạc nhiên hay không còn tùy thuộc vào khung tham chiếu.

Khung tham chiếu ở đây có nghĩa là chúng ta phải biết hết chiều cao của người Việt, nam và nữ, trong độ tuổi 18-20. Chắc chắn trong cộng đồng người Việt, có người có chiều cao thấp, nhưng cũng có người có chiều cao cao, thậm chí rất cao. Chúng ta không có điều kiện để đo chiều cao cho hàng triệu thanh niên Việt Nam. Nhưng chúng ta cũng không cần làm việc đó, vì đã có thống kê giúp chúng ta. Trong thống kê, chúng ta có thể dùng luật phân bố (law of distribution) để biết được giá trị khả dĩ của chiều cao của một quần thể.

Luật phân bố là một khái niệm không dễ hiểu với người mới học thống kê vì cách mà sách và tác giả của sách giáo khoa giải thích quá phức tạp và phi thực tế, nhưng thật ra nó rất đơn giản. Một cách ngắn gọn: một luật phân bố trong thống kê học chỉ là một phương trình để thể hiện tất cả các giá trị khả dĩ. Mà, phương trình thì phải có tham số. Do đó, giá trị khả dĩ của phân bố phụ thuộc vào tham số. Nói cách khác, với tham số, chúng ta có thể biết được trong quần thể có bao nhiêu người có chiều cao 100 cm, bao nhiêu người cao 102 cm, 103 cm, 104 cm, 105 cm, v.v. Chúng ta cũng có thể trả lời câu hỏi bao nhiêu thiếu nữ có chiều cao trên 173 cm, bao nhiêu thanh niên có chiều cao trên 204 cm.

Qui luật tự nhiên và tiến hoá rất kì diệu. Tuyệt đại đa số các đặc tính cơ thể con người đều có thể mô tả bằng luật phân bố hết sức đơn giản. Ví dụ như chiều cao và cân nặng đều có thể mô tả bằng luật phân bố bình thường - còn gọi là “normal distribution”. Luật phân bố bình thường do nhà toán học thiên tài [thiên tài thật sự] người Đức Friedrich Gauss (tên đúng là Johann Carl Friedrich Gauss) phát hiện vào thế kỉ 19. Luật phân bố bình thường chỉ có hai tham số: trung bình (μ) và độ lệch chuẩn (σ). Trong câu chuyện chiều cao, luật phân bố này phát biểu như sau: trong quần thể có chiều cao trung bình μ và độ lệch chuẩn σ , xác suất mà chúng ta tìm được người có chiều cao x trong quần thể đó là:

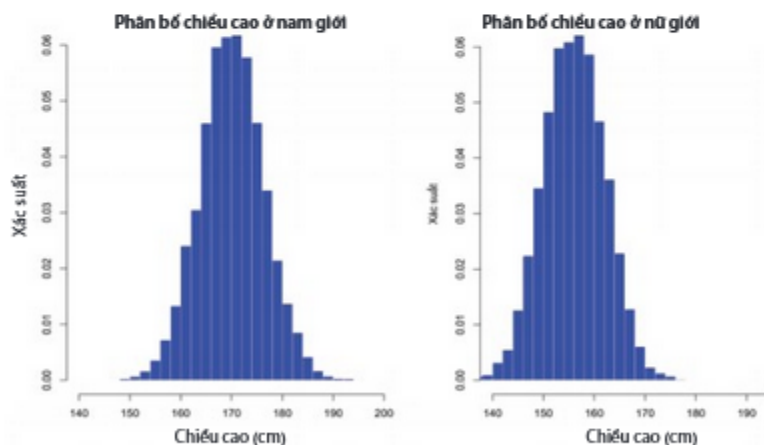
$$f(x) = \frac{1}{k} \exp \left[-\frac{(x - \mu)^2}{\sigma^2} \right]$$

trong đó $k = \sigma \sqrt{2\pi}$. Chú ý σ^2 chính là phương sai. Phương trình này tiếng Anh gọi là “probability density function”, tức hàm số mật độ xác suất. Những người học xác suất được dạy

rằng xác suất chiều cao bằng giá trị x cụ thể (ví dụ như $x = 100$ cm) là 0, nhưng trong thực tế thì không phải vậy (và đây chính là sự bất đồng giữa lý thuyết với thực tế, làm cho xác suất khó hiểu). Dựa vào phương trình trên, nếu chúng ta vẽ một đường biểu diễn mà trục hoành là chiều cao tính bằng cm, và trục tung là giá trị của $f(x)$, chúng ta sẽ có một hình chuông. Chiều cao có giá trị $f(x)$ cao nhất chính là số trung bình μ . Mức độ lệch hai bên số trung bình chính là độ lệch chuẩn s .

Vậy chiều cao của nam và nữ thanh niên Việt Nam là bao nhiêu? Chúng tôi đã làm một nghiên cứu qui mô (khoảng 4000 người) ở Sài Gòn và đo chiều cao cùng cân nặng một cách cẩn thận. Khi giới hạn những người trong độ tuổi 18-30, chúng tôi có kết quả như sau:

- Nam giới: chiều cao trung bình là $m = 170$ cm, độ lệch chuẩn $s = 6,3$ cm.
- Nữ giới: chiều cao trung bình là $m = 160$ cm, độ lệch chuẩn $s = 6,0$ cm.



Hình 1: Phân bố chiều cao ở nam giới (với trung bình 170 cm và độ lệch chuẩn 6,3 cm) và nữ giới (trung bình 156 cm và độ lệch chuẩn 6,3 cm).

Dựa vào kết quả trên, chúng ta có thể mô phỏng phân bố chiều cao của thanh niên Việt Nam như trong biểu đồ trên. Ở nam giới, chúng ta thấy đa số thanh niên có chiều cao trong khoảng 140 cm đến 195 cm, và trung bình [dĩ nhiên] là 170 cm. Ở nữ giới, chiều cao dao động trong khoảng 130 cm đến khoảng 190 cm, và trung bình là 160 cm. Ở mỗi biểu đồ, chúng ta biết rằng diện tích dưới đường biểu diễn là 1. Do đó, lợi dụng vào thực tế này, chúng ta có thể tính xác suất cho bất cứ ngưỡng nào. Chẳng hạn như chúng ta muốn biết có bao nhiêu nam thanh niên có chiều cao bằng hoặc hơn 204 cm (như em Đức Tài), và ở nữ, bao nhiêu người có chiều cao bằng hay cao hơn Thu Thảo, tức 173 cm?

Câu trả lời là tính diện tích dưới đường biểu diễn. Nhìn vào biểu đồ trên, chúng ta thấy xác suất người có chiều cao cỡ Đức Tài và Thu Thảo là rất ít. Xác suất nam thanh niên Việt Nam có chiều cao 204 cm hay cao hơn là 0,0000000339. Nói cách khác, trong số 100 triệu người, chỉ có 3 người có chiều cao như thế (≥ 204 cm). Đối với nữ, nếu tính chính xác thì xác suất những người có chiều cao bằng hoặc cao hơn Thu Thảo là 0,01513014. Nói cách khác, chỉ có khoảng 1,15% nữ có chiều cao 173 cm hay cao hơn.

Hai giá trị trên chính là trị số P

Dựa vào kết quả trên, chúng ta có thể nói rằng chiều cao của Đức Tài là một sự ngạc nhiên, vì hiếm người như thế. Còn chiều cao của Thu Thảo thì không đáng ngạc nhiên mấy, do có đến 1,15% thanh nữ cũng có chiều cao như cô ấy hay thậm chí cao hơn chiều cao của cô ấy. Đó chính là lí do tại sao tôi gọi P là chỉ số hay thước đo của sự ngạc nhiên.

Trị số P trong khoa học

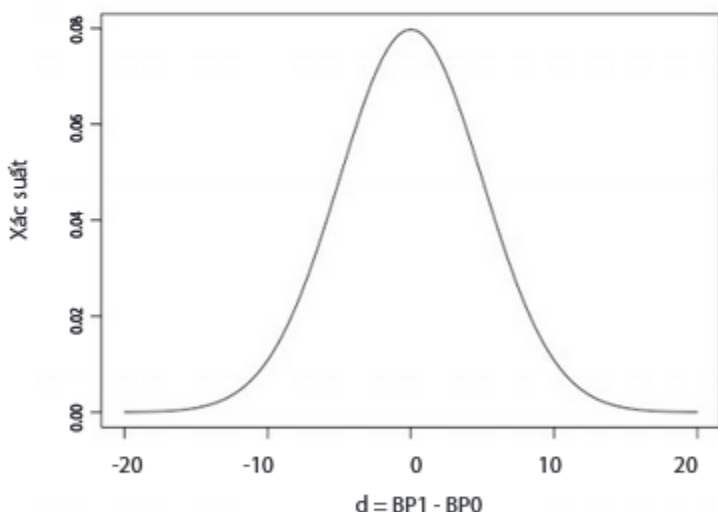
Sự ngạc nhiên cũng có thể áp dụng cho một nghiên cứu khoa học. Sau khi hoàn tất nghiên cứu, thí nghiệm, chúng ta có dữ liệu. Câu hỏi đặt ra là dữ liệu này có đáng ngạc nhiên không? Ngạc nhiên ở đây có nghĩa là kết quả đó nằm ngoài sự kì vọng. Kì vọng của ai? Kì vọng của luật phân bố. Mà, luật phân bố thì phải có tham số. Trong nghiên cứu khoa học, tham số của một luật phân bố được xác định qua giả thuyết khoa học. Như vậy, chúng ta thấy P là một xác suất có điều kiện, chứ không phải là một “con số trơn” vô điều kiện.

Câu “*tham số của một luật phân bố phải được xác định qua giả thuyết khoa học*” cần phải giải thích thêm. Nếu chúng ta nghiên cứu hiệu quả của một loại thuốc điều trị cao huyết áp, thì có lẽ thiết kế đơn giản nhất là chọn một nhóm bệnh nhân gồm 50 người chẳng hạn; đo huyết áp trước khi điều trị (tạm kí hiệu là BP0); sau đó cho họ uống thuốc một thời gian, rồi đo lại huyết áp (gọi tắt là BP1). Giả thuyết khoa học là thuốc sẽ giảm huyết áp cho bệnh nhân, nhưng chúng ta không biết giảm bao nhiêu mmHg. Do đó, theo triết lí phản nghiệm (falsificationism), chúng ta phải bắt đầu từ giả thuyết vô hiệu (mà tiếng Anh gọi là *null hypothesis*) Giả thuyết vô hiệu là thuốc không có hiệu quả.

“Không có hiệu quả” có thể diễn giải bằng ngôn ngữ thống kê học. Nói theo ngôn ngữ thống kê là: tính trung bình, huyết áp trước và sau can thiệp là 0. Nói cách khác, nếu chúng ta tính hiệu số huyết áp trước và sau can thiệp mỗi bệnh nhân, và gọi đó là d (tức là $d = BP1 - BP0$) thì giả thuyết vô hiệu phát biểu rằng $d = 0$. Chúng ta cần phải tìm phân bố của d , tức là các giá trị khả dĩ của d . Thường thì huyết áp tuân theo luật phân bố bình thường,

cho nên chúng ta cũng có thể nói rằng d tuân theo luật phân bố bình thường. Nhưng phân bố bình thường, như trình bày trên, cần 2 tham số: trung bình (μ) và độ lệch chuẩn (σ).

Chúng ta mới biết một tham số là trung bình $\mu = 0$, còn σ thì sao? Qua nghiên cứu trên 50 bệnh nhân chúng ta biết độ lệch chuẩn mẫu $s = 5$ mmHg. Chúng ta có thể vẽ biểu đồ phân bố của d như biểu đồ dưới đây. Biểu đồ này có nghĩa là nếu giả thuyết thuốc vô hiệu là đúng thì chúng ta vẫn có thể quan sát một số bệnh nhân có huyết áp giảm đến 20 mmHg, nhưng cũng có một số bệnh nhân với huyết áp tăng đến 20 mmHg. Phân bố này chính là “khung tham chiếu” mà tôi đề cập ở trên.



Giả dụ như kết quả nghiên cứu cho thấy sau 3 tháng điều trị, huyết áp của 50 bệnh nhân giảm đến 15 mmHg, và độ lệch chuẩn là 5 mmHg. Câu hỏi đặt ra là mức độ giảm 15 mmHg này có đáng ngạc nhiên, có phải là ngẫu nhiên hay không ngẫu

nhiên. Nếu ngẫu nhiên thì nó phải nằm trong giá trị kì vọng của phân bố. Chúng ta có thể tính xác suất giảm 15 mmHg hay cao hơn nếu giả thuyết vô hiệu là đúng bằng R như sau: $1 - \text{pnorm}(15, \text{mean}=0, \text{sd}=5)$, và kết quả là 0,001349898. Đây chính là trị số P . Nói cách khác, nếu thuốc vô dụng, thì xác suất mà chúng ta quan sát huyết áp giảm 15 mmHg hay cao hơn là 0,0013. Bởi vì xác suất này quá thấp, nên chúng ta phải kết luận rằng giả thuyết thuốc vô dụng có thể không đúng. Chúng ta có chứng cứ để bác bỏ giả thuyết vô hiệu.

Có thể xem cách chúng ta dùng trị số P trong khoa học như là cách bồi thẩm đoàn trong toà án luận tội. Trong luật pháp phương Tây, bị cáo vô tội cho đến khi bị kết tội (innocent until proven guilty). Cũng giống như nghiên cứu khoa học, chúng ta bắt đầu với giả thuyết vô hiệu rằng không có mối liên quan. Dựa vào chứng cứ, bồi thẩm đoàn có thể kết tội hay không kết tội. Không kết tội không có nghĩa là bị cáo không có tội, mà vì chưa đủ chứng cứ. Cũng giống như trong khoa học, khi $P > 0,05$, chúng ta chưa kết luận được là có liên quan, vì chứng cứ chưa thuyết phục, chứ không phải là không có liên quan.

3. Định nghĩa trị số P

Sau những ví dụ trên, chúng ta đã sẵn sàng định nghĩa trị số P . Trị số P là một phần trong lí thuyết Fisher gọi là “Test of Significance” (tôi tạm dịch là “Kiểm định ý nghĩa thống kê” hay viết tắt là “Kiểm định thống kê”). Qui trình kiểm định thống kê, theo Fisher, được tiến hành 3 bước:

Bước 1: phát biểu một giả thuyết vô hiệu (null hypothesis, gọi là H_0). Lí do giả thuyết vô hiệu là vì chúng ta không thể nào chứng minh được giả thuyết chính, mà chỉ bắt đầu với giả thuyết đơn giản nhất là vô hiệu.

Bước 2: làm thí nghiệm hay nghiên cứu thu thập dữ liệu D. Tính toán một kiểm định thống kê, như t, Ki bình phương, F, v.v.

Bước 3: tính xác suất D xảy ra nếu H_0 là đúng. Nói cách khác, tính $\Pr(D | H_0)$. Đó chính là trị số P . Một cách rõ ràng hơn, $P = \Pr(D | H_0)$.

Dựa trên qui trình đó, nếu trị số P là thấp thì chúng ta bác bỏ giả thuyết vô hiệu. Bởi vì chỉ có hai giả thuyết (vô hiệu và giả thuyết chính), do đó bác bỏ H_0 cũng có nghĩa là gián tiếp chấp nhận giả thuyết chính? Không hẳn. Fisher viết rất rõ rằng kiểm định thống kê không có bao hàm tiêu chuẩn để chấp nhận một giả thuyết (ông để trong ngoặc kép), vì chấp nhận hay bác bỏ giả thuyết phụ thuộc vào bối cảnh nghiên cứu:

"A test of significance contains no criterion for 'accepting' a hypothesis. According to circumstances it may or may not influence its acceptability."

Đó là cách viết khó hiểu, vì Fisher là người quen cách viết "nước đôi" nên làm cho thể hệ sau phải vất vả tìm hiểu ông thật sự muốn nói gì. Theo lí giải của Fisher, định nghĩa đúng của trị số P là như sau:

"the probability of the observed result, plus more extreme results, if the null hypothesis were true."

Tức là xác suất kết quả quan sát được hay những kết quả cao/thấp hơn, nếu giả thuyết vô hiệu là đúng. Giải thích này rõ ràng chỉ ra rằng trị số P là một xác suất có điều kiện, $\Pr(D | H_0)$, chứ không phải là $\Pr(H_0)$ mà rất rất nhiều người lầm tưởng hay hiểu lầm. Nói cách khác, trị số P là một thước đo về bằng chứng khoa học.

4. Tại sao 0,05?

Nhưng vấn đề đặt ra là ngưỡng của trị số P bao nhiêu để bác bỏ giả thuyết vô hiệu. Fisher đưa ra đề nghị ngưỡng $P = 0,05$, và không ngờ đề nghị này đã gây một ảnh hưởng vô cùng lớn trong lịch sử khoa học, bởi vì từ đó giới khoa học thuộc tất cả các lĩnh vực và khắp thế giới dùng ngưỡng đó để tuyên bố về khám phá khoa học.

Nhưng tại sao Fisher chọn 0,05 làm ngưỡng để phân định giữa nhiều và không nhiều? Đây là một câu hỏi làm nhiều người suy luận và tốn khá nhiều giấy mực. Có truyền thuyết cho rằng một hôm ông tắm trong bồn nhưng vẫn suy tư về ngưỡng của trị số P , nghĩ hoài không ra, thì khi xả nước bồn đột nhiên ông thấy năm ngón chân của mình trên mặt nước làm ông lạnh, và thế là ông chọn trị số $P = 0,05$ làm ngưỡng để tuyên bố “có ý nghĩa thống kê” (7). Không rõ đó là chuyện thật hay chỉ là một “huyền thoại” khoa học, nhưng câu chuyện thú vị cũng giống như sự ra đời của đơn vị “feet” ở Anh.

Một cách chính thức, trong cuốn sách SMRW lần tái bản 13, trang 44, Fisher đề nghị con số 0,05 và ông giải thích như sau:

“Giá trị $P = 0,05$, hay 1 trong 20, là tương đương với 1,96 hoặc gần bằng 2 [độ lệch chuẩn từ trung bình của biến số theo phân bố chuẩn]; đó là một ngưỡng thích hợp để đánh giá độ lệch đó được xem là có ý nghĩa hay không có ý nghĩa. Khi mức độ lệch cao hơn 2 độ lệch chuẩn, chúng ta có thể xem là có ý nghĩa. Dùng tiêu chuẩn này, chúng ta sẽ sai chỉ 1 lần trong số 22 lần thử nghiệm, ngay cả nếu thống kê là chỉ số dẫn đường duy nhất.”

Ở một chỗ khác, Fisher nhận xét rằng 1/20 có thể xem là cái ngưỡng để tạm xác định kết quả là do ngẫu nhiên (tình cờ) hay do can thiệp. Ông thêm rằng nếu 1/20 chưa đủ cao

để bác bỏ, chúng ta có thể lấy 1/50 (2%), hay 1/100 (1%). Tuy nhiên, ông cho biết cá nhân ông thích con số 5%, vì một dữ liệu khoa học được xem là “công nhận” (established) chỉ khi nào nó được đúc kết từ những thí nghiệm được thiết kế tốt.

Nhưng điều đáng chú ý là ông (Fisher) không quá giáo điều về ngưỡng 0,05 như giới khoa học ngày nay. Trong lần tái bản cuốn sách SMRW năm 1956, ông viết:

“[...] không nhà khoa học nào có một ngưỡng trị số P cố định từ năm này sang năm khác và trong tất cả tình huống để bác bỏ những giả thuyết; nhà khoa học phải chú ý đến từng trường hợp cụ thể dựa trên chứng cứ và ý tưởng” (8).

Thế nhưng trong thực tế, giới khoa học vẫn cứ lấy ngưỡng $P < 0,05$ để tuyên bố ý nghĩa thống kê. Thông thường, một phát hiện có trị số $P < 0,05$ được xem là “có ý nghĩa thống kê” hay nói theo tiếng Anh là “statistically significant”. Một phát hiện có ý nghĩa thống kê cũng đồng nghĩa với có một tờ “giấy thông hành” để công bố kết quả nghiên cứu. Nếu kết quả nghiên cứu có trị số $P < 0,05$ thì cơ may được công bố rất cao. Nhưng nếu nghiên cứu cho ra kết quả với $P > 0,05$ thì xem như là thất bại, và rất khó công bố kết quả trên các tạp san quốc tế. Mà, công bố quốc tế có liên quan đến sự nghiệp của một nhà khoa học. Do đó, trị số P có vai trò rất quan trọng không chỉ trong khoa học mà còn cho sự nghiệp của một nhà khoa học.

Như nói ở phần mở đầu, trị số P là một sáng kiến có ảnh hưởng rất lớn đến khoa học. Nhiều người lầm tưởng rằng trị số P là sản phẩm tri thức của thế kỉ 20, nhưng như chúng ta thấy nó đã xuất hiện từ đầu thế kỉ 18. Đến thế kỉ 19 và đầu thế kỉ 20, Karl Pearson có công mở đường cho ứng dụng của

trị số P trong kiểm định giả thuyết. Nhưng trị số P chỉ nở rộ và tìm được ứng dụng trong khoa học từ sau 1925, thời điểm mà cuốn sách “*Statistical Methods for Research Workers*” của Ronald Fisher được xuất bản. Hầu như tất cả các khám phá gen, khám phá trong vật lý, hóa học, sinh học, thuốc điều trị, chính sách, v.v. đều không ít thì nhiều dựa trên trị số $P < 0,05$. Theo trang simplystatistics.org thì cho đến nay trị số này đã có hơn 3 triệu trích dẫn (9)! Tôi thì nghĩ con số có thể còn cao hơn. Qua trị số P và các phương pháp kiểm định thống kê mà chúng ta sử dụng ngày nay, Fisher trở thành nhà khoa học có ảnh hưởng lớn nhất trong lịch sử khoa học, lớn hơn bất cứ ai, kể cả Einstein và những tên tuổi lừng danh cấp đó. Đã có nhà khoa học lấy cảm hứng từ trị số P và sáng tác một bài nhạc rất hay (10).

5. Những hiểu lầm về trị số P trong khoa học

Trị số P có thể xem là “số thông hành”, một loại bảo kê cho khả năng công bố quốc tế. Trị số P cũng là thước đo của khám phá khoa học. Nhưng rất tiếc là trị số P bị diễn giải sai rất nhiều. Diễn giải sai là vì người ta hiểu sai ý nghĩa của nó. Ngạc nhiên thay, ngay cả những người học và làm về thống kê cũng hiểu sai ý nghĩa của trị số P . Trong phần này tôi sẽ liệt kê 10 hiểu lầm phổ biến nhất. Những hiểu lầm này thật ra đã được nhiều nhà khoa học nêu lên trong quá khứ, đặc biệt là Greenland và đồng nghiệp nêu lên sau tuyên bố của ASA (11).

Hiểu lầm 1: Trị số P là xác suất của một giả thuyết khoa học

Một cách hiểu rất phổ biến là trị số P là xác suất giả thuyết vô hiệu. Chẳng hạn như nhà khoa học phát hiện mối liên quan

giữa hút thuốc lá và ung thư với trị số $P = 0,04$, họ diễn giải rằng xác suất *không* có mối liên quan là 4%. Suy ra, xác suất có mối liên quan giữa hút thuốc lá và ung thư phổi là 96%. Nhưng cách hiểu này sai. Trị số P không nói gì về xác suất của một giả thuyết khoa học. Trị số P chỉ giúp chúng ta bác bỏ giả thuyết vô hiệu, chứ không chứng minh giả thuyết nào cả.

Hiểu lầm 2: Trị số P là xác suất dữ liệu xảy ra một cách ngẫu nhiên

Đây là một hiểu lầm mang tính khái niệm cơ bản. Theo cách hiểu này, khi một kết quả phân tích (ví dụ như giảm huyết áp 10 mmHg) với trị số $P = 0,03$, thì nhà nghiên cứu hiểu rằng xác suất mà kết quả đó xảy ra do yếu tố ngẫu nhiên là 3%; suy ra, xác suất kết quả giảm 10 mmHg thật ra là 97%. Nhưng cách hiểu này cũng sai, bởi vì trị số P không nói lên kết quả nghiên cứu là ngẫu nhiên hay mang tính hệ thống.

Hiểu lầm 3: Trị số P là xác suất phát hiện sai

Đây là một hiểu lầm do suy nghĩ chưa cẩn thận. Một hiểu lầm phổ biến khác cho rằng P là xác suất phát hiện sai. Ví dụ như nhà khoa học tính toán hệ số tương quan, và có kết quả $r = 0,25$, với $P = 0,01$, họ hiểu rằng xác suất kết quả này sai là 1%. Suy ra, xác suất kết quả đúng là 99%. Nhưng cách hiểu này hoàn toàn sai, vì trị số P không cho chúng ta biết là kết quả nghiên cứu đúng hay sai.

Hiểu lầm 4: Trị số P là α

Khi thiết kế một nghiên cứu khoa học (nhất là nghiên cứu lâm sàng), nhà khoa học phải xác định ngưỡng α và β của nghiên cứu. Nói nôm na, α là dương tính giả (tức là xác suất mà nghiên cứu cho ra kết quả dương tính nhưng trong

thực tế chẳng có liên quan gì). Còn *beta* là âm tính giả (tức là xác suất mà nghiên cứu cho ra kết quả âm tính, nhưng trong thực tế thì có liên quan). Do đó, có nhiều nhà khoa học hiểu rằng trị số *P* chính là *alpha*, nhưng cách hiểu đó sai. Sai vì hiểu lầm khái niệm kiểm định giả thuyết (test of hypothesis). Trị số *P* là kết quả của kiểm định thống kê (test of significance), chứ không phải kiểm định giả thuyết.

Hiểu lầm 5: Trị số P càng thấp, mức độ ảnh hưởng càng cao

Đây là một hiểu lầm rất tai hại. Nhiều nhà nghiên cứu so sánh trị số *P* để đánh giá mức độ ảnh hưởng. Chẳng hạn như họ tìm trong y văn và thấy ảnh hưởng của thuốc statin trong một nghiên cứu có trị số $P = 0,01$, còn nghiên cứu của họ có trị số $P = 0,001$, họ suy luận rằng mức độ ảnh hưởng họ quan sát cao hơn mức độ ảnh hưởng báo cáo trong y văn. Nhưng cách hiểu này sai, vì trị số *P* không phản ánh mức độ ảnh hưởng, so sánh giữa hai hay nhiều trị số *P* là không có ý nghĩa gì cả.

Hiểu lầm 6: Nô lệ vào trị số P

Nhiều nhà nghiên cứu nghĩ (hay hiểu) rằng khi kết quả nghiên cứu có trị số $P > 0,05$ là xem như không có khác biệt hay không có ảnh hưởng. Họ chỉ chấp nhận kết quả với trị số $P < 0,05$. Nói cách khác, bất cứ kết quả nào có $P < 0,05$ là có ý nghĩa, còn $P > 0,05$ là không có ý nghĩa hoặc không đáng bàn. Tiêu biểu cho cách suy nghĩ này là một bài báo trên JAMA (tập san y khoa hàng số 1 trên thế giới) kết luận: “a low-fat dietary pattern did not result in a statistically significant reduction in invasive breast cancer risk,” nhưng khi nhìn vào kết quả thì thấy tỉ số nguy cơ là 0,91 (khoảng tin cậy 95%: 0,83 đến 1,01) và trị số $P = 0,07$. Đây là một kết

luận... bậy bạ. Trị số P có độ dao động khá cao trong một quần thể, nên kết luận bằng cách lệ thuộc vào trị số P là một sai lầm nghiêm trọng.

Hiểu lầm 7: Trị số $P > 0,05$ là một kết quả “âm tính”

Cũng giống như hiểu lầm #5, nhiều nhà nghiên cứu có cách diễn giải theo nghĩa nhị phân: trắng - đen, có - không. Khi thấy kết quả P cao hơn 0,05, nhà nghiên cứu thường có xu hướng cho rằng kết quả là “âm tính” - negative, là nghiên cứu đã... thất bại. Ví dụ tiêu biểu cho trường hợp này là cách viết “*The study failed to detect an effect*”; chữ *fail* ở đây được dùng không đúng. Chẳng hạn như một nghiên cứu công bố trên tập san *New England Journal of Medicine* (được xem là kinh thánh y khoa) báo cáo tỉ số odds là 2,2 (khoảng tin cậy 95%: 0,83 đến 6,2), trị số $P = 0,13$, và tác giả viết “*No difference in death rates*”! Câu kết luận này sai, bởi vì trong thực tế là có khác biệt (*difference*), nhưng sự khác biệt đó không có ý nghĩa thống kê.

Trị số P không nói lên kết quả là dương tính hay âm tính, càng không cho chúng ta biết là thành công hay thất bại. Nhiều nghiên cứu báo cáo kết quả không có ý nghĩa thống kê (tức $P > 0,05$) thường có độ nhạy (power) thấp, và do đó, không thể cho ra câu trả lời dứt khoát. Nhà nghiên cứu có thể không ‘chứng minh’ hai nhóm khác nhau, nhưng nhà nghiên cứu cũng không thể bác bỏ giả thuyết rằng hai nhóm có thể khác nhau. Người ta có câu “*Absence of proof is not proof of absence*” hay “*Absence of evidence is not evidence of absence*” (không có bằng chứng không có nghĩa là bằng chứng không có). Những nghiên cứu có độ nhạy đầy đủ (tức trên 80%) mà cho ra một kết quả không có ý nghĩa thống kê thì mới có thể xem là một kết quả âm tính.

Hiểu lầm 8: Lẫn lộn giữa ý nghĩa thống kê (statistical significance) và ý nghĩa lâm sàng / ý nghĩa thực tế (clinical significance / practical significance)

Như đề cập trên, nhiều nhà nghiên cứu diễn giải một kết quả có ý nghĩa thống kê ($P < 0,05$) như là khẳng định có mối liên hệ sinh học hay có ý nghĩa lâm sàng. Thật ra, trị số P không có giá trị sinh học, và cũng không thể diễn giải như là có ý nghĩa sinh học hay ý nghĩa lâm sàng. Trong lâm sàng và sinh học, mức độ ảnh hưởng (effect size), mức độ khác biệt giữa hai hay nhiều nhóm mới là điều quan trọng. Có ý nghĩa thống kê là một điều kiện cần, nhưng chưa đủ để kết luận mối liên hệ hay ảnh hưởng là có thật.

Hiểu lầm 9: Diễn giải trị số P trong bối cảnh nhiều giả thuyết

Tình huống tiêu biểu trong thực tế là một nghiên cứu kiểm định nhiều giả thuyết. Chẳng hạn như một nghiên cứu so sánh 50 biến số giữa hai nhóm bệnh nhân, tức nhà khoa học có 50 trị số P . Giả dụ như trong 50 trị số P đó, chỉ có một trị số $P = 0,046$, so sánh về chiều dài của ngón chân cái. Nhà khoa học kết luận rằng liệu pháp điều trị làm tăng chiều dài ngón chân cái của bệnh nhân! Nhưng cách hiểu hay diễn giải này sai, bởi vì nhà khoa học kiểm định quá nhiều giả thuyết, và cái ngưỡng alpha 0,05 không còn áp dụng nữa. Cái phát hiện có ý nghĩa thống kê rất có thể chỉ là ngẫu nhiên.

Hiểu lầm 10: Khoảng tin cậy 95% là xác suất của kết quả

Đây cũng là một hiểu lầm rất phổ biến trong khoa học. Tiêu biểu cho cách hiểu này là nhà nghiên cứu phân tích dữ liệu và có kết quả thuốc bisph giảm nguy cơ tử vong với relative risk 0,75, khoảng tin cậy 95% 0,35 đến 0,97; nhà nghiên cứu

diễn giải rằng thuốc giảm nguy cơ tử vong 25%, và xác suất 95% là mức độ giảm dao động từ 3% đến 65%. Nhưng về mặt lí thuyết cách hiểu này sai. Khoảng tin cậy 95% không phải là xác suất 95%. Để tính được xác suất 95% đó, phải dùng phương pháp Bayes.

Trị số P là một con số hay gặp nhất trong nghiên cứu khoa học. Trong một bài phân tích tuyệt vời trên JAMA, các tác giả dùng máy tính “đọc” các bài báo và bản tóm tắt (*abstracts*) trong Pubmed (công bố trong thời gian 1990-2015), họ nhận dạng ra gần 4,6 triệu trị số P trong số 1,6 triệu *abstracts*, và 3,4 triệu trị số P trong số 385 ngàn bài báo. Điều thú vị là trị số P có vẻ càng ngày càng thấp hơn, tức là có nhiều nghiên cứu có kết quả có ý nghĩa thống kê nhiều hơn (12).

Trên đây là 10 hiểu lầm (cũng có thể xem là sai sót) trong cách hiểu về ý nghĩa của trị số P. Những cách hiểu sai này gây tác hại rất lớn đến nghiên cứu khoa học. Rất nhiều thuốc vô hiệu quả, rất nhiều liệu pháp điều trị vô dụng, rất nhiều khám phá khoa học sai, rất nhiều kết luận gây hoang mang trong công chúng, rất nhiều đầu tư sai lầm, v.v. có nguồn gốc từ hiểu sai trị số P. Chính vì hiểu sai rồi đi đến kết luận sai, nên đại đa số (có thể hơn 90%) các phát hiện khoa học báo cáo trên các tập san quốc tế là sai lầm (3).

Vấn đề sau cùng và có lẽ rất quan trọng là ngưỡng trị số P. Ở trên, tôi đề cập rằng từ 1925 đến nay, giới khoa học dùng $P < 0,05$ để tuyên bố một phát hiện có ý nghĩa thống kê. Nhưng ít ai biết rằng ngưỡng 0,05 có thể dẫn đến nhiều phát hiện dương tính giả (tức phát hiện sai). Một tính toán đơn giản cho thấy trong số 100 phát hiện với trị số $P = 0,05$, thì có 30 phát hiện sai hay dương tính giả (13). Do đó, đã có đề nghị cần phải

có một ngưỡng mới như $P < 0,005$ hay $P < 0,001$ để tuyên bố có ý nghĩa thống kê (14, 15).

Tóm lại, trị số P có thể xem là một chỉ số của sự ngạc nhiên. Trị số P càng thấp (như 0,01 hay 0,001) thì sự ngạc nhiên càng tăng. Trị số P không phản ánh giả thuyết vô hiệu hay giả thuyết chính là đúng hay sai. Một kết quả có ý nghĩa thống kê ($P < 0,005$ hay $P < 0,001$) không có nghĩa là chúng ta đã thật sự phát hiện một mối liên quan, mà chỉ là một tín hiệu để chúng ta tiếp tục nghiên cứu để ‘chứng minh’ mối liên quan đó là thật.

Tham khảo:

- (1) Arbuthnot J. An Argument for Divine Providence taken from the constant Regularity of the Births of Both Sexes. Royal Society 1710.
Dĩ nhiên cách tính của Arbuthnot chưa thỏa đáng, vì ông chỉ quan tâm đến số bé sinh ra theo năm. Nếu chúng ta xem xét đến thứ tự trai/gái xảy ra thì kết luận có thể đã khác.
- (2) Hacking I. Logic of Statistical Inference. Cambridge University Press, Cambridge 1965.
- (3) Ioannidis JAP. Why most published research findings are false. PLoS Medicine 2005;e124.
- (4) Ioannidis JAP. Why Most Discovered True Associations are Inflated (with discussion), *Epidemiology* 2008;19, 640–658.
- (5) Ioannidis JP. Contradicted and Initially Stronger Effects in Highly Cited Clinical Research. *JAMA* 2005; 294, 218–228.
- (6) Wasserstein RL, et al. The ASA's Statement on p -Values: Context, Process, and Purpose. *The American Statistician* 2016;70:129-133.
- (7) Chia KS. Significant-it is: an obsession with the p value. *Scan J Work Environ Health* 1997;23:152–4

- (8) Fisher RA. *Statistical methods and scientific inference*. Edinburgh: Oliver & Boyd, 1956.
- (9) <https://simplystatistics.org/2012/03/07/r-a-fisher-is-the-most-influential-scientist-ever/>
- (10) <https://www.youtube.com/watch?v=yy4nsEvKh2E>
- (11) Greenland S, et al. Statistical tests, *P* values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* 2016; 31: 337–350.
- (12) Chavalarias D, et al. Evolution of Reporting *P* Values in the Biomedical Literature, 1990–2015. *JAMA* 2016; 315(11):1141–1148.
- (13) Ioannidis JAP. The Proposal to Lower *P* Value Thresholds to 0.005. *JAMA* March 22, 2018.
- (14) Benjamin DJ, Berger JO, Johnson VE, et al. Redefine statistical significance. *Nat Hum Behav*. 2018;2:6–10.
- (13) Colquhoun D. An investigation of the false discovery rate and the misinterpretation of *p*-values. *R. Soc. open sci.* 1: 140216
- (14) Ioannidis JAP. The Proposal to Lower *P* Value Thresholds to 0.005. *JAMA* March 22, 2018.
- (15) Benjamin DJ, Berger JO, Johnson VE, et al. Redefine statistical significance. *Nat Hum Behav*. 2018;2:6–10.

Đã có rất nhiều bài viết về trị số *P*, nhưng những bài sau đây nằm trong một phần thư mục của tôi mà các bạn có thể đọc thêm:

- Blume J, Peipert JF. What your statistician never told you about *p* values. *J Am Assoc Gynecol Laparosc* 2003;10:439–4
- Cohen J. The Earth is round ($p < 0.05$). *Am Psychol* 1994; 49:997–1003
- Feinstein AR. *P* values and confidence intervals. Two sides of the same unsatisfactory coin. *J Clin Epidemiol* 1998;51:355–60

- Gardner MJ, Altman D. Confidence intervals rather than *P* values: estimation rather than hypothesis testing. *Br Med J* 1986;292:746–50
- Hubbard R, Lindsay RM. Why *P* values are not a useful measure of evidence in statistical significance testing. *Theory Psychol* 2008;18:69–88
- Goodman S. Commentary: the *P* value, devalued. *Int J Epidemiol* 2003;32:699–702
- Goodman S. A comment on replication, *p*-values and evidence. *Stat Med* 1992;11:875–9
- Sterne J, Smith GD. Sifting the evidence: what's wrong with significance tests? *BMJ* 2001;322:226–31

DIỄN GIẢI KẾT QUẢ NGHIÊN CỨU KHOA HỌC

Một trong những khó khăn của người mới bắt đầu làm nghiên cứu khoa học là đọc và diễn giải một bài báo khoa học. Đối diện với một kết quả có ý nghĩa thống kê, câu hỏi kế tiếp là gì, và có thể tin vào kết quả này hay không, đó là những câu hỏi gai góc mà không phải lúc nào cũng có câu hỏi chính xác và đúng. Chương này sẽ đi qua ba yếu tố quan trọng trong việc diễn giải kết quả nghiên cứu: yếu tố nhiễu, yếu tố bias, và yếu tố ngẫu nhiên.

Nhưng để bàn về 3 yếu tố đó, tôi muốn bàn qua về mục đích chung của các nghiên cứu khoa học. Nhìn chung và một cách tổng quát, nghiên cứu khoa học thường có 3 nhóm mục đích: phân loại, liên quan, và tiên lượng. Nhóm mục tiêu thứ nhất mang tính mô tả, như khám phá gen, phát hiện sinh vật mới, và nhà khoa học thường phân nhóm các phát hiện mới này. Nhóm mục tiêu thứ hai mang tính *association*, tức là tìm những qui luật chung về mối liên quan giữa các yếu tố. Chẳng hạn như nghiên cứu về mối liên quan giữa gan FTO và bệnh tiểu đường, nhà nghiên cứu qua phân tích dữ liệu có thể phát hiện những qui luật chung về sự khác biệt giữa các biến thể gen liên quan đến tỉ trọng mỡ trong cơ thể. Nhóm mục tiêu thứ ba là tiên lượng, hay *prediction*. Rất nhiều nghiên cứu khoa học có mục tiêu tìm các yếu tố quan trọng để phát triển mô hình

tiên lượng. Những nghiên cứu về ung thư và một số bệnh mãn tính trong thực tế là nhằm phát hiện bệnh sớm hơn dựa vào “hồ sơ” môi trường và gen của mỗi cá nhân. Ba nhóm mục tiêu này định hình nhiều nghiên cứu khoa học trong thực tế.

Một cách nhìn khác là nguyên nhân và hệ quả. Thật vậy, đa số các nhà khoa học là những người đi tìm mối liên quan giữa nguyên nhân (cause) và hệ quả (effect). Chữ *cause* ở đây cần phải hiểu theo nghĩa rộng, bao gồm những yếu tố nói theo ngôn ngữ thống kê học là *independent variables* hay biến độc lập. Còn *effect* ở đây có thể hiểu là *outcome* trong y khoa. Nhưng một cách dễ hiểu và bao quát hơn là thay vì nói nguyên nhân và hệ quả, chúng ta có thể nói đến *exposure* và *outcome*. Như vậy, *exposure* ở đây phải hiểu là biến độc lập, là biến tiên lượng, là yếu tố nguy cơ; còn *outcome* là bệnh lí hay một đặc tính lâm sàng mà nghiên cứu quan tâm. Nếu nghiên cứu có mục tiêu tìm mối liên quan giữa uống cà phê và ung thư tụy, thì *exposure* ở đây là cà phê, và *outcome* là ung thư tụy. Đa số nghiên cứu y học là tìm hay xác định mối liên quan giữa *exposure* (E) và *outcome* (O).

Đối diện với một nghiên cứu mà kết quả cho thấy có mối liên quan giữa E và O, người đọc hay nhà khoa học làm gì để đánh giá mức độ khả tín của kết quả này? Một cách đơn giản nhất là xem xét đến mô hình nghiên cứu. Nếu mô hình nghiên cứu là RCT (randomized controlled trial) thì vấn đề đơn giản. Đối với nghiên cứu RCT so sánh hai nhóm (như nhóm dùng thuốc thật và nhóm dùng giả dược) được thiết kế tốt, tất cả các yếu tố liên quan đến O đều bằng nhau giữa hai nhóm bệnh nhân, nên khác biệt về O giữa hai nhóm chỉ có thể là do can thiệp. Nhưng đối với các nghiên cứu quan sát, nhà nghiên cứu không thể kiểm soát tất cả các yếu

tổ nguy cơ, nên có sự mất cân đối giữa hai nhóm so sánh (mắc bệnh và không mắc bệnh), nên vấn đề trở nên phức tạp hơn nghiên cứu RCT. Để diễn giải kết quả nghiên cứu quan sát một cách hợp lý, nhà nghiên cứu cần phải xem xét đến 3 yếu tố chính:

- Yếu tố ngẫu nhiên (chance hay random error)
- Yếu tố bias
- Yếu tố nhiễu (confounding effect)

Chỉ khi nào loại bỏ ba yếu tố này thì mối liên quan mới có thể xem là mối liên hệ nhân quả.

1. Yếu tố ngẫu nhiên

Bất cứ kết quả nghiên cứu nào cũng có yếu tố ngẫu nhiên. Nếu chúng ta quan sát mối liên quan giữa hút thuốc lá và ung thư phổi, thì câu hỏi đầu tiên đặt ra là: có phải đó là một quan sát ngẫu nhiên? Để trả lời câu hỏi này, các nhà nghiên cứu phải dùng phương pháp kiểm định thống kê (còn gọi là “test of significance”) và trị số P . Nhưng vì trị số P thường hay bị hiểu sai, nên gần đây Hiệp hội Thống kê học Hoa Kỳ (American Statistical Association) mới ra một tuyên cáo về cách diễn giải trị số P (1).

1.1 Trị số P

Phương pháp kiểm định thống kê bắt đầu bằng một giả thuyết vô hiệu (null hypothesis), và đặt kết quả phân tích trong bối cảnh giả thuyết vô hiệu là đúng. Giả thuyết vô hiệu là một phát biểu nghịch đảo. Nếu giả thuyết là hút thuốc lá làm tăng nguy cơ ung thư phổi, thì giả thuyết vô hiệu phát biểu rằng nguy cơ bị ung thư phổi ở người hút thuốc lá và người

không hút thuốc lá là như nhau. Do đó, phương pháp kiểm định thống kê được thực hiện như sau:

- Phát biểu giả thuyết vô hiệu H_0 (không có khác biệt hay không có mối liên hệ);
- Làm nghiên cứu và thu thập dữ liệu (và gọi dữ liệu là D);
- Tính xác suất D xảy ra nếu H_0 là đúng. Đây chính là trị số P ;
- Nếu $P \leq 0,05$, bác bỏ giả thuyết H_0 ; nếu $P > 0,05$, chấp nhận giả thuyết H_0 .

Theo qui trình kiểm định thống kê như trên, trị số P có nghĩa là xác suất quan sát được dữ liệu D (hay dữ liệu cao/thấp hơn D) nếu giả thuyết vô hiệu là đúng. Nói theo ngôn ngữ xác suất ý nghĩa thật của trị số P là:

$$\Pr(\text{Data} \mid H_0)$$

Trị số P không cho chúng ta biết xác suất của giả thuyết H_0 là bao nhiêu, đúng hay sai. Nhưng nhiều người hiểu lầm rằng trị số P là xác suất của giả thuyết vô hiệu! Không đúng. Trị số P là một xác suất có điều kiện (như công thức trên), và nó phản ánh mức độ khả dĩ của dữ liệu (data) NẾU giả thuyết vô hiệu là đúng.

1.2 Vấn đề kiểm định nhiều giả thuyết

Một trong những vấn đề quan trọng trong diễn giải trị số P là vấn đề kiểm định nhiều giả thuyết. Một nghiên cứu thường kiểm định nhiều giả thuyết, chứ không phải đơn thuần chỉ một giả thuyết. Khi kiểm định nhiều giả thuyết thì xác suất tìm ra những mối liên hệ một cách ngẫu nhiên càng tăng cao.

Khi chúng ta kiểm định 1 giả thuyết, thường chúng ta chấp nhận sai sót 5% (còn gọi là sai sót loại I - type I error, còn gọi là α). Sai sót loại I cũng giống như dương tính giả, tức kết quả có ý nghĩa thống kê ($P < 0,05$) nhưng trong thực tế thì chẳng có mối liên hệ gì cả. Cũng có thể ví von sai sót loại I như kết quả chẩn đoán, tức là một cá nhân có kết quả xét nghiệm dương tính nhưng cá nhân đó thật ra không mắc bệnh. Xin nhắc lại, mỗi lần kiểm định giả thuyết, chúng ta chấp nhận sai sót 5%. Nói cách khác, xác suất đúng là 95% hay 0,95. Do đó, kiểm định 2 giả thuyết thì xác suất đúng cả hai là $0,95^2 = 0,9025$; khi kiểm định 10 giả thuyết thì xác suất đúng cả 10 là $0,95^{10} = 0,5987$, tức khoảng 60%. Nói cách khác, khi kiểm định 10 giả thuyết, chúng ta có xác suất 40% tìm ra một kết quả có ý nghĩa thống kê, và kết quả này có thể chỉ là ngẫu nhiên. Nói chung, xác suất tìm ra một kết quả có ý nghĩa thống kê trong k test là $1 - (1 - \alpha)^k$. Khi $k = 100$ thì xác suất gần như 100% là chúng ta sẽ tìm ra một kết quả có ý nghĩa thống kê, nhưng kết quả này rất có thể là do yếu tố ngẫu nhiên.

Một vấn đề khác liên quan đến việc kiểm định nhiều giả thuyết là vấn đề “tra tấn dữ liệu” (data torture), còn gọi là “fishing expedition”. Trong một nghiên cứu, thông thường nhà nghiên cứu kiểm định nhiều giả thuyết chính, nhưng ngoài giả thuyết còn phát sinh vấn đề chủ quan của nhà khoa học. Trong nhiều trường hợp, nhà khoa học muốn khai thác dữ liệu càng nhiều càng tốt (theo ý họ), chẳng hạn như phân tích khác biệt giữa nhóm can thiệp và nhóm chứng theo độ tuổi, theo giới tính, theo thành phần kinh tế, hay nói chung theo những biến mà nhà khoa học nghĩ đến. Tình trạng “lang thang” với giả thuyết này có tên là “subgroup analysis” hay được ví von là “data torture”, tức

tra tấn dữ liệu cho đến khi có kết quả với trị số $P < 0,05$. Trên lý thuyết, các nghiên cứu với nhiều biến phân nhóm thì sẽ có vô vàn so sánh, và trong điều kiện đó chúng ta biết rằng xác suất tìm được một kết quả có ý nghĩa thống kê là gần như 100%, nhưng dĩ nhiên kết quả đó sai. Điều này nói lên rằng khi đọc một bài báo khoa học mà tác giả kiểm định quá nhiều giả thuyết khoa học và có kết quả có ý nghĩa thống kê (ví dụ như phát hiện mối liên hệ có ý nghĩa thống kê ở nhóm phụ nữ 45 - 54 tuổi sống ở thành phố và không hút thuốc không uống rượu) thì cần phải cẩn thận, vì rất có thể đó chỉ là một kết quả ngẫu nhiên.

1.3 Trị số P và cỡ mẫu

Trị số P trong thực tế là một kết quả hỗn hợp của hai thông tin: mức độ cao thấp của mối liên quan và độ chính xác của mối liên quan. Độ chính xác (precision) phụ thuộc vào cỡ mẫu; nghiên cứu với cỡ mẫu thấp thì độ chính xác thấp, nghiên cứu có cỡ mẫu cao thì độ chính xác cao. Nói cách khác, trị số P phản ánh cùng một lúc cả hai thông tin: mức độ liên quan và cỡ mẫu.

Hai câu hỏi hay vấn đề đặt ra là: (i) nếu nghiên cứu có cỡ mẫu nhỏ (ví dụ $n = 10$) và có kết quả với trị số P cũng nhỏ (ví dụ $P = 0,01$); (ii) nếu nghiên cứu có cỡ mẫu lớn (ví dụ $n = 10000$) và trị số P cũng nhỏ (ví dụ $P = 0,001$), thì kết quả có đáng tin cậy không?

Chúng ta biết rằng nghiên cứu với cỡ mẫu nhỏ thì độ ảnh hưởng phải lớn mới có thể cho ra trị số P nhỏ. Nhưng một nghiên cứu với cỡ mẫu lớn hay rất lớn và dù độ ảnh hưởng thấp thì trị số P vẫn có thể nhỏ hay rất nhỏ. Như vậy, một nghiên cứu nhỏ mà cho ra kết quả với trị số P nhỏ thì đáng

tin cậy? Không hẳn như thế. Câu chuyện liên quan đến James Heckman là một ví dụ thú vị về câu hỏi này:

James Heckman là một nhà kinh tế (Giải Nobel kinh tế, 2000) lí giải rằng Nhà nước nên cung cấp các dịch vụ “preschool” cho trẻ em. Bằng chứng cho lí giải này là những nghiên cứu nhỏ vào thập niên 1970. Trong một bài báo trên New York Times, ông phê bình những người phê bình ông về những nghiên cứu đó vì cỡ mẫu nhỏ. Ông lí giải rằng nghiên cứu với cỡ mẫu nhỏ mà cho ra kết quả với trị số P nhỏ thì bằng chứng càng mạnh (chứ không phải yếu):

*“Also holding back progress are those who claim that Perry and ABC are experiments with samples too small to accurately predict widespread impact and return on investment. This is a nonsensical argument. Their relatively small sample sizes actually speak for — not against — the strength of their findings. **Dramatic differences between treatment and control-group outcomes are usually not found in small sample experiments**, yet the differences in Perry and ABC are big and consistent in rigorous analyses of these data.”*

Nhưng Heckman sai. Trong thực tế, những nghiên cứu với cỡ mẫu nhỏ có thể cho ra những kết quả rất lạc quan (trị số P thấp hay rất thấp, kiểu $< 0,001$). Lí do là vì nghiên cứu với cỡ mẫu nhỏ thường bị sai số mẫu cao (sampling error) và có khi kết quả chỉ do vài con số ngoại vi.

Trong bảng dưới đây có 5 nghiên cứu với cỡ mẫu khác nhau; chỉ có 2 nghiên cứu mà tỉ số nguy cơ có ý nghĩa thống kê. Nhưng nghiên cứu A và C có kết quả đáng tin cậy hơn nghiên cứu B, D và E.

Nghiên cứu	Cỡ mẫu	Tỉ số nguy cơ	Trị số P
A	2500	1,4	0,02
B	500	1,7	0,10
C	2000	1,6	0,04
D	250	1,8	0,30
E	1000	1,6	0,06

Nói chung, nghiên cứu với cỡ mẫu thấp ít khi nào có được trị số P thấp; nghiên cứu càng lớn càng có xu hướng tìm ra những kết quả với trị số P rất thấp. Do đó, một nghiên cứu với cỡ mẫu hàng triệu đối tượng mà có kết quả với trị số $P = 0,07$ thì kết quả đó không có gì quá ấn tượng!

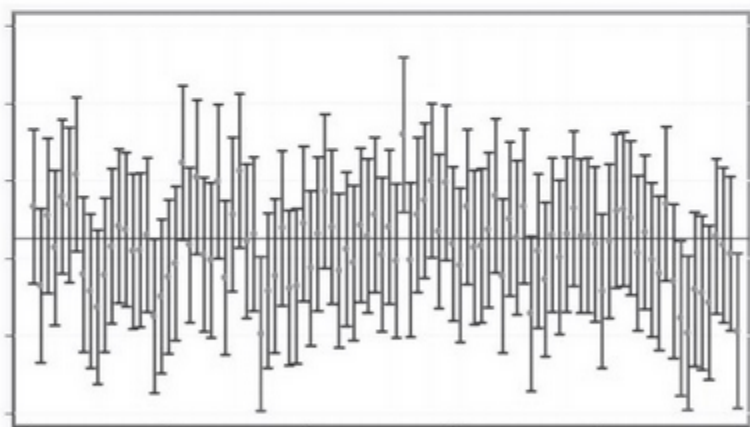
1.4 Diễn giải khoảng tin cậy 95%

Những minh hoạ trên cũng có nghĩa là nếu một nghiên cứu chỉ được báo cáo bằng trị số P , thì chúng ta không thể nào xác định trị số P thấp là do mức độ liên quan lớn hay số cỡ mẫu lớn. Trong thực tế, một nghiên cứu với cỡ mẫu rất cao (hàng triệu đối tượng), nhưng độ ảnh hưởng rất thấp, thì trị số P vẫn có thể rất thấp. *Trị số P do đó không phản ánh mức độ liên quan, không nói lên mức độ ảnh hưởng thấp hay cao.*

Vì trị số P không phản ánh độ ảnh hưởng, nên giới nghiên cứu khoa học thường quan tâm đến khoảng tin cậy 95% (tiếng Anh là “95% confidence interval”). Khoảng tin cậy 95% có thể tính toán cho bất cứ chỉ số thống kê nào, như tỉ số odds, tỉ số nguy cơ, hệ số tương quan, v.v. Khoảng tin cậy 95% phản ánh sự bất định của một mối liên quan, nhưng đồng thời cũng cung

cấp thêm thông tin về sự khả dĩ của mối liên quan. Chẳng hạn như mối liên quan giữa gen A và bệnh tiểu đường có tỉ số odds là 5,2 (với KTC95 là 3,2 đến 7,2), chúng ta có thể hiểu rằng nếu nghiên cứu này được lặp lại rất nhiều lần, và mỗi lần với cùng cỡ mẫu, thì 95% các nghiên cứu đó sẽ có khoảng tin cậy 95% dao động trong khoảng 3,2 đến 7,2.

Nhưng trong thực tế thì có rất rất nhiều nhà khoa học hiểu sai KTC95. Họ thường hiểu rằng KTC95 là xác suất 95% là mức độ ảnh hưởng dao động trong khoảng đó. Chẳng hạn như với ví dụ trên, nhiều nhà khoa học hiểu rằng xác suất 95% là tỉ số odds dao động trong khoảng 3,2 đến 7,2. Nhưng cách hiểu này sai. Khoảng tin cậy 95% không phản ánh xác suất của một sự ảnh hưởng.



Source: <http://sites.nicholas.duke.edu/statsreview/ci/>

Biểu đồ trên có thể dùng để giải thích ý nghĩa của khoảng tin cậy 95%. Có 100 nghiên cứu; mỗi nghiên cứu lấy mẫu từ quần thể, và ước tính KTC95 (dấu chấm màu xanh là trung

binh, đường gạch đỏ ngang là tham số thật). Trong 100 nghiên cứu đó, có 5 nghiên cứu mà KTC95 lệch ra khỏi tham số thật; 95 nghiên cứu còn lại có KTC95 bao trùm tham số thật. Do đó, KTC95% có nghĩa là nếu nghiên cứu được lặp lại nhiều lần và mỗi lần có cùng cỡ mẫu, thì 95% các KTC95 sẽ dao động trong khoảng 3,2 đến 7,2 (ví dụ trên).

Khoảng tin cậy 95% còn cung cấp cho chúng ta một thông tin về sự bất định của độ ảnh hưởng. Bảng dưới đây trình bày kết quả nghiên cứu với khoảng tin cậy 95% cho tỉ số nguy cơ (risk ratio) cho 5 nghiên cứu với cỡ mẫu khác nhau. Chúng ta chú ý thấy nghiên cứu A và C có khoảng tin cậy 95% hẹp so với các nghiên cứu khác (B, D và E), và điều này có nghĩa là kết quả nghiên cứu A và C ổn định hơn các nghiên cứu khác.

Nghiên cứu	Cỡ mẫu	Tỉ số nguy cơ và khoảng tin cậy 95%	Trị số P
A	2500	1,4 (1,2 - 1,7)	0,02
B	500	1,7 (0,7 - 3,1)	0,10
C	2000	1,6 (1,2 - 2,1)	0,04
D	250	1,8 (0,6 - 3,9)	0,30
E	1000	1,6 (0,9 - 2,5)	0,06

1.5 Diễn giải khoảng tin cậy 95% và ý nghĩa thực tế

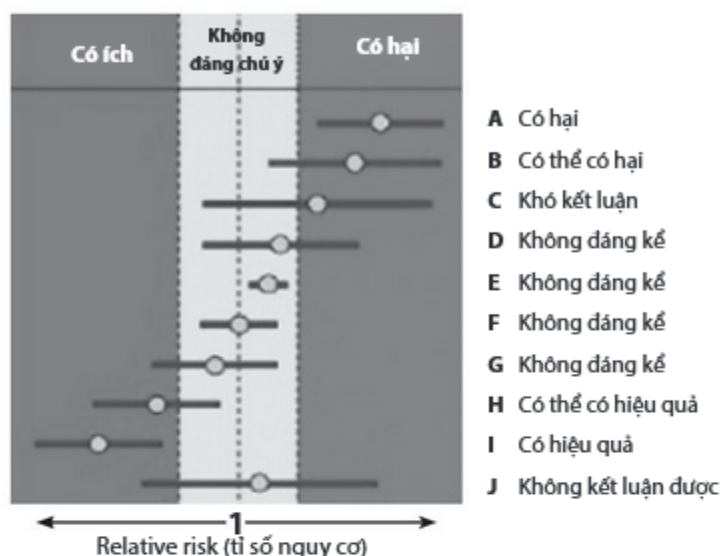
Diễn giải kết quả nghiên cứu cần phải đặt trong ý nghĩa lâm sàng. Một kết quả có thể có ý nghĩa thống kê nhưng chẳng có ý nghĩa lâm sàng. Chẳng hạn như tỉ số nguy cơ gãy xương tay 1,05 có thể có trị số $P < 0,001$, nhưng mức độ tăng 5% nguy cơ thì không thể xem là có ý nghĩa lâm sàng. Do đó, trước khi

đọc và diễn giải KTC95, nhà nghiên cứu cần phải xác định ngưỡng ảnh hưởng được xem là có ý nghĩa lâm sàng.

Chẳng hạn như nếu chúng ta xác định rằng nguy cơ đột quỵ tăng 10% trở lên được xem là có hại và giảm 10% trở lên được xem là có lợi, thì bất cứ yếu tố nguy cơ nào có tỉ số nguy cơ (RR) trong khoảng 1,01 đến 1,09 hay 0,91 đến 0,99 thì có thể xem là không có ý nghĩa lâm sàng. Nhưng nếu một yếu tố nguy cơ có RR từ 1,20 đến 1,70 thì rõ ràng là có hại. Tương tự, một can thiệp giảm nguy cơ đột quỵ với KTC95 RR từ (ví dụ) 0,5 đến 0,7 được xem là có lợi và có ý nghĩa lâm sàng.

Nhưng cũng có tình huống KTC95 dao động quá lớn, và kết quả này không thể cho chúng ta kết luận gì đáng tin cậy. Chẳng hạn như nếu tỉ số nguy cơ RR dao động từ 0,2 đến 3,5 thì kết quả này có thể phản ánh cỡ mẫu của nghiên cứu còn nhỏ. Biểu đồ dưới đây cung cấp vài hướng dẫn để diễn giải một kết quả phân tích thống kê qua tỉ số nguy cơ. Vùng màu vàng có nghĩa là không có ý nghĩa lâm sàng; vùng màu xanh là có ích; và vùng màu đỏ là có hại.

Nghiên cứu A cho thấy KTC95 đều nằm trong vùng màu đỏ thì đó là một kết quả mang tính xác định (vừa có ý nghĩa lâm sàng, vừa có ý nghĩa thống kê). Nghiên cứu B tuy có ý nghĩa thống kê (vì phần dưới của KTC95 nằm ngoài đường tham chiếu), nhưng vì KTC95 bao trùm vùng không có ý nghĩa lâm sàng, nên chúng ta có thể kết luận rằng yếu tố này “có thể có hại”. Nghiên cứu C thuộc vào nhóm “khó kết luận” vì KTC95 quá rộng và cũng không có ý nghĩa thống kê.



Nghiên cứu D tuy có KTC95 không rộng nhưng không có ý nghĩa thống kê và KTC95 cũng bao trùm vùng không có ý nghĩa lâm sàng, nên chúng ta có thể kết luận rằng yếu tố này có độ ảnh hưởng không đáng kể. Nghiên cứu E có KTC95 rất hẹp và có ý nghĩa thống kê, nhưng vì tất cả KTC95 nằm trong vùng không có ý nghĩa lâm sàng, nên chúng ta cũng có thể kết luận độ ảnh hưởng không đáng quan tâm. Cách diễn giải tương tự có thể áp dụng cho nghiên cứu F, G, H, I và J.

1.6 Bayes Factors

Bất cứ nghiên cứu khoa học nghiêm chỉnh nào cũng dựa trên một giả thuyết. Nhà khoa học rất muốn biết kết quả nghiên cứu có phù hợp với giả thuyết đặt ra hay không. Chẳng hạn như nếu chúng ta nghiên cứu về ảnh hưởng của thuốc điều trị loãng xương thì giả thuyết đặt ra có thể là tỉ lệ gãy xương ở nhóm điều trị thấp hơn tỉ lệ gãy xương ở nhóm không

điều trị (nhóm chứng). Giả dụ như nếu chúng ta có kết quả với trị số $P = 0,045$. Câu hỏi đặt ra là trị số P đó có phản ảnh giả thuyết chúng ta là đúng hay sai. Chúng ta biết rằng trị số $P = 0,045$ không có nghĩa rằng xác suất giả thuyết vô hiệu (thuốc không có hiệu quả) là 0,045. Hoàn toàn không có ý nghĩa đó. Trị số $P = 0,045$ có nghĩa là nếu thuốc không có hiệu quả, thì xác suất mà chúng ta quan sát được kết quả hiện tại là 4,5%. Như vậy, trị số P không trả lời câu hỏi xác suất giả thuyết thuốc có hiệu quả là bao nhiêu.

Để ước tính xác suất của một giả thuyết, chúng ta cần đến một phương pháp hoàn toàn khác: phương pháp Bayes. Đối với phương pháp Bayes, xác suất không phải là một tần số mà phản ảnh “niềm tin” về một hiện tượng. Theo trường phái tần số, câu nói “xác suất ông Clinton đắc cử là 0,9” là vô nghĩa vì đó là một hiện tượng đơn lẻ (single event), nhưng đối với trường phái Bayes thì câu nói đó chấp nhận được vì nó phản ảnh độ tin tưởng của một cá nhân.

Để ước tính xác suất của một giả thuyết, phương pháp Bayes dựa vào hai thông tin: xác suất tiền định (prior probability) và số liệu thực tế (data) để ước tính xác suất hậu định (posterior probability). Định lí Bayes phát biểu rằng xác suất hậu định của một giả thuyết, $P(H | \text{data})$, là tích số của xác suất tiền định của giả thuyết đó, $P(H)$, và xác suất dữ liệu nếu giả thuyết là đúng, $P(\text{data} | H)$:

$$P(H | \text{data}) = P(H) \times P(\text{data} | H)$$

Như vậy, xác suất giả thuyết là đúng sau khi đã có dữ liệu phụ thuộc một phần vào giả thuyết tiền định, $P(H)$.

Nhưng trong thực tế, chúng ta không biết hay rất khó xác định xác suất giả thuyết tiền định. Do đó, một cách khác

để đánh giá sự khả tín của chứng cứ là tính hệ số Bayes, còn gọi là Bayes Factor (BF). BF là tỉ số của hai xác suất liên quan đến dữ liệu. Tử số là xác suất dữ liệu nếu giả thuyết vô hiệu là đúng, và mẫu số là xác suất dữ liệu nếu giả thuyết chính là đúng:

$$BF = \frac{P(\text{data} | H1)}{P(\text{data} | H0)}$$

Như vậy, BF là thước đo về chứng cứ. Nếu BF = 1, chứng cứ (dữ liệu) không nghiêng về giả thuyết nào. Nếu BF < 1 thì dữ liệu nghiêng về giả thuyết vô hiệu. Ngược lại, nếu BF > 1 thì dữ liệu nghiêng về giả thuyết chính. BF nhìn chung là một thước đo tương đối khách quan vì nó không phụ thuộc vào xác suất của một giả thuyết nào. Một hướng dẫn chung về cách diễn giải BF là theo bảng dưới đây (tôi để nguyên tiếng Anh):

Bayes Factor	Diễn giải
>100	Decisive evidence for H1
30 đến 100	Very strong evidence for H1
10 đến 30	Strong evidence for H1
3 đến 10	Substantial evidence for H1
1 đến 3	Anecdotal evidence for H1
1	No evidence
1/3 đến 1	Anecdotal evidence for H0
1/10 đến 1/3	Substantial evidence for H0
1/30 đến 1/10	Strong evidence for H0
1/100 đến 1/30	Very strong evidence for H0
<1/100	Decisive evidence for H0

Hoá ra, dưới một số điều kiện phổ quát, BF có liên quan mật thiết với trị số P. Nói cách khác, nếu chúng ta có trị số P thì chúng ta cũng có thể ước tính BF. Trong một bài báo quan trọng công bố vào năm 1987 (2) và sau đó (3), Sellke và đồng nghiệp chỉ ra rằng BF tối thiểu có thể ước tính từ trị số P như sau:

$$\min BF = \frac{1}{-e \times P \times \log(P)}$$

hoặc $\min BF = \exp(-0,5z^2)$. Trong đó, $e = 2,718$ là hằng số Euler, và z là chỉ số z . Chẳng hạn như một nghiên cứu với kết quả $P = 0,045$, chúng ta có thể ước tính BF như sau:

$$\min BF = \frac{1}{-2,718 \times 0,045 \times \log(0,045)} = 2,64$$

Như vậy, kết quả BF cho thấy bằng chứng cho giả thuyết H_1 chỉ thuộc loại “anecdotal”, tức rất yếu (dù $P = 0,045$ là có ý nghĩa thống kê).

Một mối liên hệ khác nữa cũng thú vị không kém là mối liên hệ giữa $\min BF$ và xác suất hậu định của giả thuyết nghiên cứu. Hoá ra, mối liên hệ này rất đơn giản:

$$\min post.p = \frac{1}{1 + \frac{1-q}{q \times \min BF}}$$

trong đó, q là xác suất tiên định của giả thuyết. Nếu chúng ta bắt đầu bằng $q = 0,5$ (tức 50%) thì với $\min BF = 2,64$, xác suất hậu định tối thiểu của giả thuyết là:

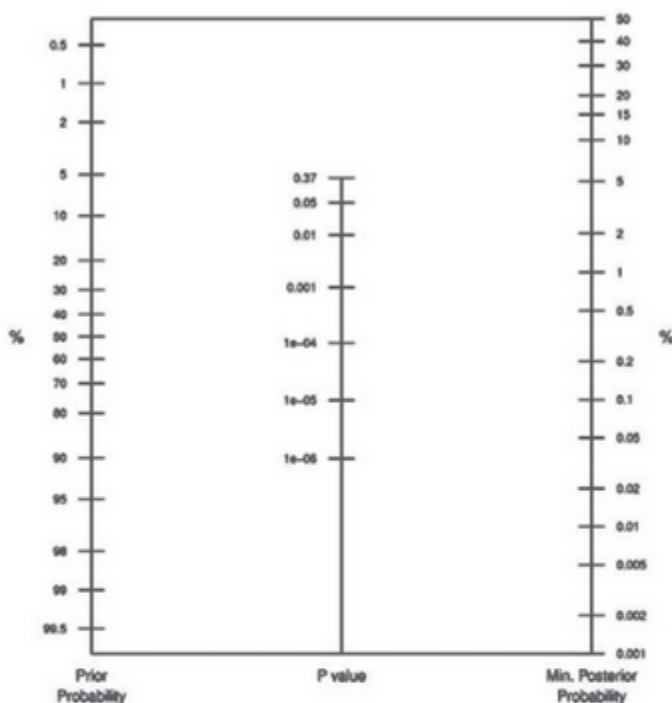
$$\min post.p = \frac{1}{1 + \frac{1-0,5}{0,5 \times 2,64}} = 0,72$$

Nói cách khác, cho dù $P = 0,045$ (tức có ý nghĩa thống kê), và với xác suất tiền định là 0,5, thì xác suất hậu định của giả thuyết chỉ 0,72. Nhưng nếu chúng ta không chắc chắn (trong trường hợp kiểm định nhiều giả thuyết), q có thể chỉ 5% ($q = 0,05$) thì xác suất hậu định chỉ 0,12. Bảng dưới đây trình bày xác suất hậu định cho các trị số $P = 0,05; 0,01; 0,001; 0,0001$ với xác suất tiền định 0,5.

**Xác suất hậu định tối thiểu của giả thuyết nghiên cứu
cho các trị số P và xác suất tiền định**

Xác suất tiền định (q)	$P=0,05$ ($\min BF = 2,46$)	$P=0,01$ ($\min BF = \sim 8,0$)	$P=0,001$ ($\min BF = 53,3$)	$P=0,0001$ ($\min BF = 399,5$)
0,01	0,024	0,165	0,913	1,000
0,05	0,114	0,508	0,982	1,000
0,10	0,214	0,686	0,991	1,000
0,20	0,380	0,831	0,996	1,000
0,30	0,513	0,894	0,998	1,000
0,40	0,621	0,929	0,999	1,000
0,50	0,711	0,952	0,999	1,000

Trong thực tế, chúng ta có thể dùng nomogram do L. Held xây dựng để ước tính xác suất hậu định cho giả thuyết nghiên cứu (4). Cách dùng nomogram này rất đơn giản. Đầu tiên, đánh dấu vào trục số 1 về giả thuyết tiền định; sau đó đánh dấu vào cột trị số P ; và vẽ một đường thẳng từ hai điểm đó đến trục thứ 3 và điểm giao chéo đó là xác suất hậu định của giả thuyết.



2. Yếu tố bias

Bias là chữ có nghĩa rất rộng trong nghiên cứu khoa học, và khó dịch sang tiếng Việt, nên tôi để nguyên chữ tiếng Anh. Trong nghiên cứu khoa học, bias được định nghĩa là một sai sót có hệ thống trong việc thiết kế và thực hiện nghiên cứu, và những sai sót này dẫn đến kết quả nghiên cứu bị lệch so với sự thật. Chẳng hạn như nếu trong thực tế (sự thật), người mắc bệnh tiểu đường có nguy cơ bị gãy xương tăng gấp 2 lần, nhưng trong nghiên cứu quan sát thì tỉ số nguy cơ là 4,5 lần, thì kết quả này được xem là *biased* (tính từ). Một ví dụ khác, nếu trong thực tế có mối liên quan giữa hút thuốc lá và ung

thư phổi, nhưng nghiên cứu không phát hiện mối liên quan này, thì kết quả nghiên cứu cũng được xem là *biased*. Như vậy, bias ở đây có nghĩa là giá trị quan sát (observed value) không bằng với giá trị thật (true value).

Nói theo ngôn ngữ thống kê học, bias còn có nghĩa là giá trị của ước số (estimate) không bằng với giá trị của tham số (parameter). Chúng ta biết rằng parameter là giá trị của quần thể, còn estimate là giá trị của một mẫu, và chúng ta dùng estimate để ước tính giá trị khả dĩ của parameter. Nếu giá trị trung bình về chiều cao của 1000 phụ nữ là 165 cm, và nếu chúng ta biết rằng chiều cao trung bình của phụ nữ Việt Nam là 166 cm, thì sự khác biệt 1 cm đó được xem là *bias*. Bias trong thống kê học, do đó, là độ lệch giữa giá trị trung bình quan sát và giá trị trung bình trong quần thể.

Bởi vì kết quả của một nghiên cứu thường dựa vào một nhóm đối tượng, nên bias là một yếu tố có thể giải thích tại sao chúng ta quan sát mối liên hệ. Điều này cũng có nghĩa là khi diễn giải kết quả có ý nghĩa thống kê, chúng ta phải tự hỏi: *kết quả này có phải là do bias?* Có rất nhiều loại bias, nhưng tựu trung lại, có 3 loại bias chính cần phải quan tâm đến là *selection bias*, *information bias*, và *recall bias*. Selection bias có nghĩa là nghiên cứu chọn sai đối tượng. Information bias liên quan đến sai sót trong việc đo lường và xếp loại thông tin. Recall bias có nghĩa là những sai sót liên quan đến việc thu thập dữ liệu, mà thường đối tượng không nhớ những thông tin cá nhân một cách chính xác. Tất cả những bias này đều có thể dẫn đến kết luận sai của nghiên cứu.

2.1 “Selection bias” và câu chuyện cà phê và ung thư tuyến tụy

Brian MacMahon là một nhà dịch tễ học người Anh rất nổi tiếng, từng làm chủ nhiệm khoa dịch tễ học của Đại học

Harvard trong một thời gian dài (1958 - 1988). Năm 1981, ông công bố một công trình nghiên cứu bệnh chứng về mối liên hệ giữa uống cà phê và ung thư tuyến tụy. Kết quả được công bố trên *New England Journal of Medicine* (5), nên thu hút rất nhiều sự chú ý của công chúng và các đồng nghiệp trong chuyên ngành. Trong nghiên cứu này, ông phỏng vấn 389 bệnh nhân ung thư tuyến tụy, và 644 người trong nhóm chứng (không bị ung thư tuyến tụy). Những thông tin thu thập là hút thuốc lá, dùng bia rượu, trà, và cà phê. Ông báo cáo rằng có mối liên quan yếu giữa hút thuốc lá và nguy cơ ung thư, nhưng người dùng cà phê có nguy cơ ung thư gia tăng cao:

"There was a weak positive association between pancreatic cancer and cigarette smoking, but we found no association with use of cigars, pipe tobacco, alcoholic beverages, or tea. A strong association between coffee consumption and pancreatic cancer was evident in both sexes. [...] after adjustment for cigarette smoking, the relative risk associated with drinking up to two cups of coffee per day was 1.8 (95% confidence limits, 1.0 to 3.0), and that with three or more cups per day was 2.7 (1.6 to 4.7)."

Với kết quả này và với "niềm tin", ông khẳng định rằng đó là một mối liên quan thật. Và, từ niềm tin này, ông ngưng uống cà phê, và chỉ uống trà mà thôi!

Nhưng sau khi công trình nghiên cứu được công bố, nhiều nhóm nghiên cứu khác cũng làm nghiên cứu như thế, nhưng kết quả của họ không nhất quán với kết quả của MacMahon. Các nhóm nghiên cứu này không tìm thấy mối liên quan giữa uống cà phê và ung thư tuyến tụy.

Câu hỏi đặt ra là tại sao có sự khác biệt giữa kết quả của MacMahon và các nhóm sau đó? Sau khi phân tích kỹ qui trình tuyển chọn đối tượng nghiên cứu, các chuyên gia đi đến kết

luận rằng MacMahon đã chọn... sai đối tượng nghiên cứu trong nhóm chứng. Hóa ra, Giáo sư MacMahon chọn các đối tượng trong nhóm chứng từ bệnh nhân của các bác sĩ chuyên về các bệnh tiêu hóa. (Những bác sĩ này cũng là người chẩn đoán ung thư tuyến tụy.) Đúng về mặt nguyên tắc, MacMahon làm đúng, vì nhóm chứng nên chọn từ nguồn của nhóm bệnh. Nhưng vì các bệnh nhân của các bác sĩ chuyên khoa tiêu hóa thường [dĩ nhiên] mắc bệnh tiêu hóa, và họ không dùng cả phê hay được khuyến cáo không dùng cả phê.

	Nhóm bệnh	Nhóm chứng
Dùng cả phê	a	b
Không dùng cả phê	c	d

Do đó, nếu đặt trong bối cảnh nghiên cứu với bảng số liệu 2 x 2 (2 dòng và 2 cột) như trên, thì MacMahon có được d (số người không dùng cả phê trong nhóm chứng) nhiều hơn c. Do đó, khi tính odds $odds1 = a/c$ và $odds2 = b/d$ thì rõ ràng là $odds1$ cao hơn nhiều so với $odds2$. Do đó, kết quả của MacMahon thực chất thể hiện một sự bias. Bias này có tên là “selection bias”.

2.2 Selection bias: câu chuyện HRT và bệnh tim mạch

Năm 1991, Stampfer và đồng nghiệp công bố một phân tích công trình nghiên cứu Nurses' Health Study (NHS) cho thấy thay thế hormone (HRT) có lợi cho sức khỏe phụ nữ sau mãn kinh (6). Công trình NHS theo dõi 48.470 phụ nữ sau mãn kinh, tuổi từ 30 đến 63, những người này không có tiền sử bệnh tim mạch. Trong thời gian 10 năm theo dõi, những người phụ nữ dùng HRT có nguy cơ mắc bệnh xơ vữa động

mạch (coronary heart disease, CHD) thấp hơn nhóm không dùng HRT (tỉ số nguy cơ [RR] 0,56, khoảng tin cậy 95% dao động từ 0,40 đến 0,80). Sau đó, một phân tích tổng hợp do Grady et al (7) thực hiện cho thấy HRT giảm CHD với RR 0,65. Tác giả kết luận rằng *“current estrogen use is associated with a reduction in the incidence of coronary heart disease as well as in mortality from cardiovascular disease.”* Nhưng đây là những kết quả của nghiên cứu quan sát.

Đến cuối thập niên 1990s, một công trình nghiên cứu RCT có tên là HERS (Heart and Estrogen-progestin Replacement Study) trên tạp san JAMA (8). Kết quả của HERS cho thấy HRT không làm giảm CHD, nhưng làm tăng nguy cơ đột quỵ và ung thư vú. Năm 2002, một nghiên cứu RCT khác có tên là WHI (Women Health Initiative Study) cũng được công bố trên JAMA. Kết quả của WHI cũng cho thấy HRT làm tăng CHD, tăng nguy cơ đột quỵ, tăng nguy cơ ung thư vú, nhưng giảm nguy cơ gãy cổ xương đùi (9).

Tỉ số nguy cơ liên quan đến HRT và các bệnh lí mãn tính

Outcome	HERS	WHI
CHD	0,99 (0,80 - 1,22)	1,29 (1,02 - 1,65)
Đột quỵ	1,23 (0,89 - 1,70)	1,41 (1,07 - 1,85)
Pulmonary embolism	2,79 (0,89 - 8,73)	2,13 (1,39 - 3,25)
Ung thư vú	1,30 (0,77 - 2,19)	1,26 (1,00 - 1,59)
Hip fracture	1,10 (0,49 - 2,50)	0,66 (0,45 - 0,98)
Tử vong	1,08 (0,84 - 1,38)	0,98 (0,82 - 1,18)
Global index	NA	1,15 (1,03 - 1,28)

Tại sao có sự khác biệt lớn về ảnh hưởng của HRT giữa nghiên cứu quan sát và nghiên cứu RCT? Lí do chính là do nghiên cứu quan sát không thể ngẫu nhiên hóa các đối tượng nghiên cứu. Do đó, hai nhóm (nhóm dùng HRT và nhóm không dùng HRT) rất khác nhau về các yếu tố quan trọng có thể có liên quan đến bệnh tim mạch. Chẳng hạn như nhóm dùng HRT hóa ra là những phụ nữ thuộc thành phần kinh tế cao, học thức cao, thu nhập cao, hay luyện tập thể dục, ít hút thuốc lá, ít dùng bia rượu, họ hay quan tâm đến sức khỏe, hay nói chung là khỏe mạnh hơn nhóm không dùng HRT. Khi các yếu tố này được phân tích một cách đúng đắn, các nhà nghiên cứu không tìm thấy sự khác biệt về nguy cơ bệnh tim mạch giữa hai nhóm! Do đó, kết quả các nghiên cứu quan sát là bias, và bias này có tên là “*healthy bias*”. Healthy bias là một dạng của selection bias.

2.3 “Recall bias” và câu chuyện dị tật bẩm sinh

Một trong những ca nghiên cứu nổi tiếng hay dùng làm minh họa cho vấn đề recall bias là nghiên cứu bệnh chứng về yếu tố nguy cơ liên quan đến dị tật bẩm sinh. Hầu hết các nghiên cứu đều tìm ra một yếu tố nào đó. Nhưng mỗi liên quan mà các nghiên cứu này phát hiện có thể là do bias trong việc thu thập thông tin. Các cha mẹ có con dị tật bẩm sinh thường có động cơ tìm hiểu các yếu tố nguy cơ, và họ thường hay nhìn lại quá khứ phơi nhiễm của mình. Nói cách khác nhìn theo bảng số liệu dưới đây, nhà nghiên cứu sẽ có số a (số người có yếu tố nguy cơ trong nhóm bệnh) cao hơn nhóm chứng. Do đó, $odds1 = a/c$ cao hơn $odds2 = b/d$.

	Nhóm bệnh	Nhóm chứng
Phơi nhiễm	a	b
Không phơi nhiễm	c	d

Điều này cũng có nghĩa là tỉ số odds (odds ratio) mà nhà nghiên cứu quan sát có thể chỉ là do thông tin thu thập dẫn đến sự mất cân đối về yếu tố nguy cơ giữa hai nhóm. Tình trạng này còn gọi là *differential bias*, *reporting bias*, hay *rumination bias*.

Bias là một vấn đề lớn trong các nghiên cứu quan sát. Nhiều nhà nghiên cứu đã thống kê được 256 loại bias, nhưng trong thực tế tôi nghĩ con số còn cao hơn, nhất là trong thời đại “Dữ liệu Lớn”. Một trong những sai sót “sáng chói” trong kết luận liên quan đến dữ liệu Lớn là nghiên cứu của Google trong việc dựa vào các tweets để biết sự chuyển biến của trận bão Sandy (2016). Tính theo số tweets, Google thấy rất nhiều tweets từ Manhattan và rất ít từ các vùng như Breezy Point, Coney Island, và Rockaway, và họ kết luận rằng Manhattan đang bị cơn bão Sandy. Nhưng thật ra đây là một dạng bias về thông tin, vì những vùng bị ảnh hưởng bởi cơn bão mất điện nên người dân không dùng điện thoại di động được. Ngoài ra, Manhattan là vùng nhà giàu nên số người có điện thoại di động cũng cao hơn các vùng khác. Bài học ở đây là nếu chỉ dựa vào con số mà không xem xét đến bối cảnh thì rất dễ dẫn đến sai lầm mà không có dữ liệu, dù lớn hay nhỏ, có thể giải quyết. Trong bài này, tôi chỉ bàn đến 3 loại bias phổ biến trong nghiên cứu quan sát, nhưng cũng đủ để cảnh báo tất cả chúng ta phải cẩn thận trong việc diễn giải một kết quả có ý nghĩa thống kê.

3. Yếu tố nhiễu (confounding effects)

Chữ *confounder* xuất phát từ tiếng Latin *confundere* có nghĩa là “pha trộn vào nhau”. Do đó, định nghĩa của *confounding* theo Rothman là ảnh hưởng của yếu tố phơi nhiễm bị lẫn lộn với ảnh hưởng của một yếu tố khác và dẫn đến bias. Trong

thực tế, một yếu tố được xem là confounder (tạm kí hiệu là Z), nếu yếu tố đó hội đủ 3 điều kiện sau đây:

- Z có liên quan với outcome (Y);
- Z có liên quan đến exposure (X); và
- Z không nằm trong cơ chế liên quan giữa X và Y.

3.1 Yếu tố nhiễu: dẫn nhập

Giả dụ như chúng ta nghiên cứu về mối liên quan giữa bệnh loãng xương (osteoporosis, OS) và thoái hóa khớp (osteoarthritis, OA) bằng mô hình nghiên cứu đoàn hệ. Giả thuyết đặt ra là loãng xương là yếu tố nguy cơ gây bệnh thoái hóa khớp. Giả dụ như qua 5 năm theo dõi một nhóm gồm 1000 bệnh nhân loãng xương và 1000 người không loãng xương, ghi nhận kết quả như sau:

	OA	Không OA	Tỉ lệ OA
Loãng xương	380	620	0,38
Không loãng xương	110	890	0,11

Kết quả trên cho thấy loãng xương có liên quan đến tăng nguy cơ OA. Tỉ số nguy cơ là $0,38 / 0,11 = 3,5$, có nghĩa là nguy cơ bị OA ở bệnh nhân loãng xương cao gấp 3,5 lần so với nguy cơ OA ở nhóm không loãng xương.

Nhưng kết quả này có thể là do chúng ta chưa xem xét đến yếu tố độ tuổi. Chúng ta biết rằng người bị loãng xương thường là cao tuổi (trên 60). Chúng ta cũng biết rằng những bệnh nhân thoái hóa khớp cũng là bệnh nhân cao tuổi. Do đó, một cách kiểm tra có sự ảnh hưởng của yếu tố nhiễu (độ tuổi)

hay không, chúng ta thử phân tích theo hai nhóm tuổi: nhóm dưới 45-79 và nhóm 80 tuổi trở lên.

Nhóm 45-79 tuổi

	OA	Không OA	Tỉ lệ OA
Loãng xương	20	80	0,20
Không loãng xương	90	810	0,01

Nhóm 80+ tuổi

	OA	Không OA	Tỉ lệ OA
Loãng xương	360	540	0,40
Không loãng xương	20	80	0,20

Chúng ta thấy ở nhóm 45-79 tuổi, tỉ số nguy cơ $RR = 0,20 / 0,10 = 2$. Ở nhóm 80+ tuổi, tỉ số nguy cơ cũng bằng 2 ($0,4 / 0,2$). Như vậy, ở đây tỉ số nguy cơ thật là 2,0, chứ không phải 3,5 như phân tích ở trên. Lí do của sự khác biệt về kết quả là vì mối liên quan giữa loãng xương và thoái hóa khớp chịu sự ảnh hưởng của yếu tố độ tuổi.

3.2 Nhiều bởi chỉ định (confounding by indication)

Trong các nghiên cứu quan sát về mối liên quan giữa dùng thuốc và bệnh lí, yếu tố nhiễu có tên là “confounding by indication” (yếu tố nhiễu vì chỉ định) là một cơn... ác mộng. Một trong những trường hợp tiêu biểu là mối liên quan giữa thuốc chống trầm cảm và sa sút trí tuệ (cognitive deficit). Kết quả của nghiên cứu cho thấy những người dùng thuốc chống trầm cảm có nguy cơ sa sút trí tuệ tăng cao. Nhưng đây là một kết quả có thể do nhiễu vì chỉ định. Lí do đơn giản là vì người dùng thuốc này chắc chắn có tiền sử trầm cảm hay đang bị

trầm cảm, và trầm cảm thì có liên quan đến sa sút trí tuệ; do đó, mối liên quan giữa thuốc chống trầm cảm và sa sút trí tuệ có thể là ảo.

Một nghiên cứu nổi tiếng công bố trên *New England Journal of Medicine* về aspirin và nhồi máu cơ tim (MI) là một ca tiêu biểu cho yếu tố nhiễu vì chỉ định (10). Nghiên cứu thoát đầu được thiết kế theo mô hình RCT, với hai nhóm bệnh nhân: nhóm aspirin được cho uống aspirin (325 g/ngày) và nhóm giả dược. Kết quả cho thấy bệnh nhân nhóm aspirin có nguy cơ MI giảm rõ rệt so với nhóm chứng, và do đó, ủy ban y đức quyết định ngưng công trình nghiên cứu.

Nhưng sau khi công trình nghiên cứu bị ra lệnh dừng lại, tất cả bệnh nhân của 2 nhóm aspirin và nhóm chứng được cho phép uống aspirin, và họ được theo dõi 7 năm. Kết quả của nghiên cứu quan sát này được công bố trên *Arch Int Medicine* vào năm 2000 (11). Kết quả cho thấy aspirin có liên quan đến giảm nguy cơ MI 38% (tỉ số nguy cơ RR 0,72; khoảng tin cậy 95%, 0,55 đến 0,95), không có liên quan đến đột quỵ, nhưng giảm nguy cơ tử vong (RR 0,64; khoảng tin cậy 95%, 0,54 đến 0,77). Bảng dưới đây so sánh kết quả RCT và nghiên cứu quan sát:

Tỉ số nguy cơ (risk ratio) liên quan đến nhồi máu cơ tim, đột quỵ và tử vong

Outcome	RCT	Nghiên cứu quan sát
Nhồi máu cơ tim	0,56 (0,45 - 0,70)	0,72 (0,55 - 0,95)
Đột quỵ	2,14 (0,96 - 4,77)	1,02 (0,74 - 1,39)
Tử vong	0,96 (0,60 - 1,54)	0,64 (0,54 - 0,77)

Nhưng kết quả còn thú vị hơn trên. Những bệnh nhân uống aspirin lại là những người từng có tiền sử gia đình với MI, cao huyết áp, dùng bổ sung vitamin E, và vài yếu tố khác như cân nặng, bia rượu, và luyện tập thể dục. Người uống aspirin do đó thường cao tuổi hơn, hay hút thuốc lá, dùng bia rượu, quá cân và béo phì, và đã có nguy cơ MI cao. Chính vì những yếu tố nguy cơ này mà họ chọn dùng aspirin. Do đó, mối liên quan mà nhà nghiên cứu quan sát giữa aspirin và MI thật ra chỉ là một ảnh hưởng không hẳn của aspirin mà là do các yếu tố nguy cơ.

3.3 Nghịch lí Simpson (Simpson Paradox)

Một trong những hình thức của ảnh hưởng nghịch đảo là “Nghịch lí Simpson”, được nhà thống kê học Edward Simpson mô tả lần đầu vào năm 1951 (tuy nhiên hiện tượng này đã được ghi nhận trước đó khá lâu) (12). Nghịch lí Simpson có thể định nghĩa như là một sự ảnh hưởng không nằm trong kì vọng khi tổng hợp nhiều nhóm nhỏ. Nếu chúng ta quan sát ảnh hưởng của một can thiệp trong 3 nhóm A, B, C một cách nhất quán, Nghịch lí Simpson tiên đoán rằng nếu chúng ta tổng hợp dữ liệu của 3 nhóm thành một quần thể, thì ảnh hưởng của can thiệp trong quần thể có thể ngược lại với ảnh hưởng trong từng nhóm!

Bảng dưới đây là một dữ liệu thật trích từ một nghiên cứu dịch tễ học. Các nhà nghiên cứu quan sát tỉ lệ tử vong ở nam và nữ của hai nhóm bệnh nhân phân nhóm theo giới tính. Kết quả cho thấy ở nam, tỉ lệ tử vong ở nhóm can thiệp thấp hơn nhóm chứng ở cả nam (40% so với 50%) và nữ (5% so với 15%).

Nhóm	Nam		Nữ	
	Tử vong	Sống	Tử vong	Sống
Can thiệp	80 (40%)	120	5 (5%)	95
Chứng	20 (50%)	20	39 (15%)	221

Tuy nhiên, khi các nhà nghiên cứu tổng hợp nam và nữ thành một nhóm như bảng số liệu dưới đây:

Nhóm	Nam + Nữ	
	Tử vong	Sống
Can thiệp	85 (28%)	215
Chứng	59 (20%)	241

thì tỉ lệ tử vong trong nhóm can thiệp lại *cao hơn* nhóm chứng (28% so với 20%).

Hiện tượng này xảy ra, vì sự phân bố nhóm (hay yếu tố nguy cơ) mất cân đối giữa các nhóm đối tượng. Chẳng hạn như chúng ta thấy ở nam, số đối tượng trong nhóm can thiệp cao gấp 5 lần nhóm chứng, còn ở nữ, số đối tượng thuộc nhóm can thiệp thì rất thấp (100) so với nhóm chứng (260). Trong trường hợp này, phân tích bằng cách gộp 2 nhóm nam và nữ có thể cho ra kết quả sai. Tuy nhiên, nếu nhà nghiên cứu có kinh nghiệm, họ có thể sử dụng phương pháp phân tích thống kê để hiệu chỉnh cho sự mất cân đối nhóm.

Hiện tượng Nghịch lí Simpson thường hay thấy trong nghiên cứu quan sát vì cách phân nhóm một cách “tự nhiên”. Điều này dẫn đến sự mất cân bằng trong phân nhóm và kết quả chung có thể bị nhiễu / sai. Do đó, diễn giải kết quả của nghiên cứu quan sát, nhất là kết quả tổng hợp, cần phải hết sức cẩn thận vì có thể chịu sự chi phối của Nghịch lí Simpson.

3.4 “Nghịch lý môi sinh” (Ecologic fallacy)

Émile Durkheim (1858 - 1917) là một nhà xã hội học người Pháp rất nổi tiếng với những công trình nghiên cứu về mối liên quan giữa tín ngưỡng và tự tử. Trong cuốn sách nổi tiếng “*Le Suicide*” xuất bản năm 1897, ông trình bày số liệu cho thấy tỉ lệ tự tử ở những địa phương có nhiều tín đồ đạo Tin Lành (Protestants) cao hơn những địa phương có nhiều tín đồ Công Giáo (Catholics). Ông kết luận rằng do nền tảng xã hội và sự kiểm soát xã hội ở những địa phương có đông tín đồ Công Giáo, nên nguy cơ tự tử ở người Công Giáo thường thấp.

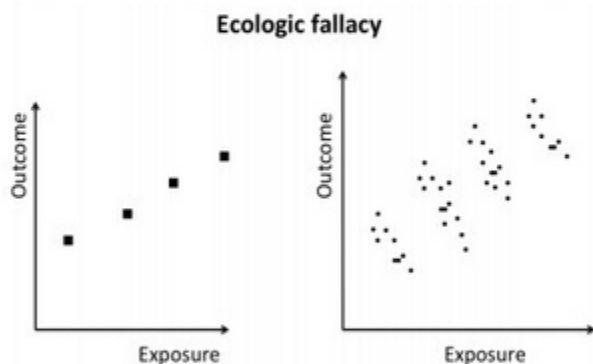
Nhưng đó là một sai sót quan trọng. Sai sót này có tên là nghịch lý môi sinh hay *ecologic fallacy*. Đây là một nghịch lý rất phổ biến trong khoa học xã hội và kinh tế học. Không có địa phương nào là có 100% tín đồ Tin Lành hay 100% tín đồ Công Giáo. Những địa phương có nhiều tín đồ Công Giáo thì cũng có tín đồ Tin Lành nhưng họ là thiểu số. Rất có thể vì cộng đồng Tin Lành là thiểu số trong cộng đồng Công Giáo đa số làm cho tín đồ Tin Lành tự tử nhiều hơn.

Nhưng sai lầm lớn nhất của Durkheim là đơn vị phân tích của ông là địa phương, chứ không phải cá nhân. Vấn đề chúng ta quan tâm là tín đồ Tin Lành có nguy cơ tự tử cao hơn tín đồ Công Giáo. Nhưng Durkheim lại dùng dữ liệu của địa phương, chứ không phải dữ liệu ở cấp độ cá nhân. Dữ liệu cấp địa phương hay cộng đồng không thể dùng để suy luận cho mối liên hệ ở cấp cá nhân.

Để hiểu vấn đề hơn, chúng ta có thể xem qua 2 biểu đồ dưới đây. Biểu đồ bên trái cho thấy có mối liên quan dương tính giữa exposure và outcome; khi giá trị của exposure càng cao thì giá trị outcome cũng càng cao. Dựa vào kết quả này,

nhà nghiên cứu có lẽ kết luận rằng mối liên hệ giữa exposure và outcome là khá chặt chẽ.

Nhưng thật ra 4 điểm trong biểu đồ này là số trung bình của 4 địa phương (xem biểu đồ bên phải). Biểu đồ bên phải cho thấy ở **mỗi địa phương**, mối liên hệ giữa exposure và outcome là nghịch đảo! Ở mỗi địa phương, khi giá trị của exposure tăng thì giá trị của outcome giảm. Như vậy, mối liên hệ thật giữa exposure và outcome là nghịch đảo. Biểu đồ bên trái “che giấu” mối liên hệ thật, vì nó dựa vào số liệu tổng kết cho từng địa phương, và khi tính hệ số tương quan trên cơ sở địa phương rất dễ dẫn đến sai lầm.



3.5 Hiệu ứng hồi qui về số trung bình

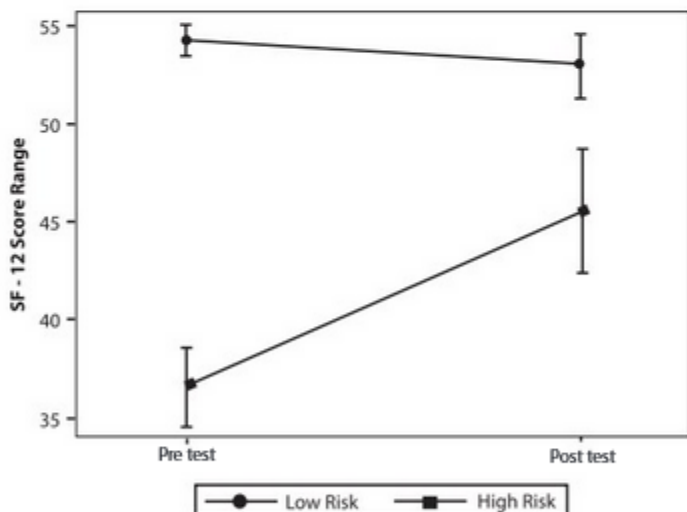
Một trong những wisdom của người Việt tôi thấy rất hay là câu nói “*Không ai giàu ba họ, chẳng ai khó ba đời.*” Nhìn trong thực tế, chúng ta thấy câu này rất ứng nghiệm với những nhà giàu nứt vách như gia đình của Công tử Bạc Liêu. Khi ông giàu có tột đỉnh, ông dám chơi nông bằng cách đốt tiền để nấu chè (theo huyền thoại, chứ có thật hay không cũng chẳng rõ), nhưng vài đời sau thì con cháu của ông sống trong cảnh túng quẫn. Ngược lại, cũng có những gia đình

xuất thân rất nghèo vài thế hệ, nhưng sau đó thì có người làm giàu nứt vách. Chúng ta thấy sau 1975, có những gia đình xưa kia rất nghèo khó, chủ yếu làm nghề nông hay theo du kích, nhưng sau đó thì giàu có nhất nhì nước. Nhưng nếu có dịp định lượng, tôi nghĩ con cháu của những gia đình rất giàu hay những gia đình rất nghèo tính trung bình là những người có xu hướng có thu nhập trung bình. Đây cũng là một hiện tượng hay thấy trong khoa học mà tôi tạm dịch là “*Hiệu ứng hồi qui về số trung bình*”.

Tiếng Anh gọi hiệu ứng này là “*regression toward the mean effect*”, và nó có nguồn gốc từ nhà khoa học lừng danh Francis Galton. Trong bài báo “Regression towards Mediocrity in Hereditary Stature,” Galton quan sát thấy có mối tương quan giữa chiều cao của con và chiều cao trung bình của cha và mẹ (ông gọi là mid-parent height). Khi chiều cao của cha và mẹ ở số trung bình trong quần thể, thì chiều cao của con bằng chiều cao trung bình đó (68,2 in); khi chiều cao trung bình của cha mẹ cao hơn số trung bình thì chiều cao của con có xu hướng thấp hơn cha mẹ và quay về số trung bình; và khi chiều cao của cha mẹ thấp hơn trung bình thì chiều cao của con cao hơn cha mẹ nhưng cũng có xu hướng quay về số trung bình quần thể. Galton gọi qui luật hay hiện tượng này là “regression toward the mean” (RTM).

Qui luật hồi qui về số trung bình (RTM) còn thấy trong hầu hết các thí nghiệm khoa học tự nhiên và khoa học xã hội. Trong các nghiên cứu trước - sau (before-after study), bệnh nhân được đo lường hai lần: trước khi điều trị (hay can thiệp) và tạm kí hiệu là X_0 , và sau khi điều trị, tạm gọi là X_1 . Do đó, mỗi bệnh nhân, chúng ta có thể tính hiệu quả bằng hiệu số $D = X_1 - X_0$. Hiệu ứng RTM tiên đoán rằng D sẽ liên quan nghịch đảo với X_0 .

Giả dụ rằng chúng ta làm một nghiên cứu can thiệp một nhóm bệnh nhân bằng tập thể dục và outcome là mật độ xương (MĐX). Mật độ xương của một nhóm bệnh nhân được đo trước và sau khi tập thể dục. Hiệu ứng RTM cho biết những bệnh nhân có MĐX cao hay rất cao sẽ có mức độ giảm cao sau khi can thiệp. Nhưng RTM cũng nói rằng những bệnh nhân có MĐX thấp hay rất thấp, thì sau khi can thiệp MĐX sẽ tăng. Xu hướng tăng và giảm đó quay về số trung bình. Do đó, khi chúng ta quan sát MĐX tăng thì điều đó không hẳn là do tập thể dục mà có thể là do hiệu ứng RTM. Tương tự, khi chúng ta quan sát những người có MĐX cao lúc ban đầu, sau khi tập thể dục, MĐX của họ suy giảm, thì điều đó không hẳn có nghĩa là tập thể dục làm giảm sức mạnh của xương. Đó chỉ là do hiệu ứng của RTM.



Ví dụ về hiệu ứng RTM: Nhóm với score ban đầu thấp sau khi can thiệp thì score tăng; nhóm có score cao lúc ban đầu, sau khi can thiệp có xu hướng giảm.

Nguồn: <https://bmcmredsmethodol.biomedcentral.com/articles/10.1186/1471-2288-13-119>

Hiệu ứng RTM có thể thấy trong rất nhiều bối cảnh. Từ nghiên cứu giáo dục (điểm học sinh) đến nghiên cứu tâm lý, tất cả đều chịu sự tác động của hiệu ứng RTM. Những học sinh thoát đầu có điểm thấp và sau này có điểm cao, hay những học sinh ban đầu có điểm cao sau này có điểm thấp, có thể là do hiệu ứng RTM. Trong phẫu thuật, những bệnh nhân nặng nhất thường có hiệu quả cao nhất, nhưng đó có thể là do hiệu ứng RTM chứ không phải do phẫu thuật.

Để giảm thiểu ảnh hưởng của hiệu ứng RTM, các nhà nghiên cứu có kinh nghiệm thường thiết kế nghiên cứu có nhóm chứng (tức không can thiệp). Đó cũng chính là lí do tại sao các nghiên cứu RCT thường có nhóm chứng. Không có nhóm chứng, những “hiệu quả” mà chúng ta quan sát rất khó diễn giải, và có lẽ chẳng có hiệu quả gì cả!

4. Chín điều kiện để đánh giá một mối liên hệ nhân quả

Bradford Hill là một nhà thống kê học nổi tiếng trong thế giới y khoa. Ông là người đầu tiên thực hiện công trình nghiên cứu RCT về thuốc kháng sinh và bệnh lao phổi. Năm 1965, ông công bố một luận văn (bài báo) được nhiều trích dẫn sau đó (13). Trong bài báo, ông đề ra 9 tiêu chí, hay nói đúng hơn là 9 điều kiện, để đi đến một kết luận về nguyên nhân và hệ quả (tiếng Anh):

- Strength of association
- Consistency
- Specificity
- Temporality
- Biological gradient
- Plausibility

- Coherence
- Experiment
- Analogy

4.1 Mức độ liên quan (strength of association)

Những mối liên hệ nhân quả thật sự thường có mức độ liên quan cao. Mức độ liên quan ở đây thể hiện qua các chỉ số thống kê như tỉ số nguy cơ, tỉ số odds, hiệu số, hay mức độ ảnh hưởng nói chung (effect size). Chẳng hạn như trong mối liên hệ giữa loãng xương và gãy cổ xương đùi có tỉ số nguy cơ bằng 3 hoặc cao hơn (bệnh nhân loãng xương có nguy cơ gãy xương cao gấp 3 lần so với người không bị loãng xương). Đó là một mối liên hệ khá cao, và là một tín hiệu của mối liên hệ nhân quả. Nhưng cũng có những mối liên hệ thấp với tỉ số odds cỡ 1,05 hay thậm chí 1,5, và những kết quả này cần phải đặt trong vòng nghi ngờ.

4.2 Tính nhất quán (consistency)

Một mối liên hệ nhân quả thật sự thường được quan sát ở nhiều đối tượng khác nhau, môi trường khác nhau, địa phương khác nhau, và ở nhiều thời điểm khác nhau. Chẳng hạn như mối liên hệ giữa hút thuốc lá và ung thư phổi, hay giữa loãng xương và gãy xương đã được lặp lại ở nhiều đối tượng tại nhiều quần thể ở gần như bất cứ thời điểm nào. Do đó, các mối liên hệ này rất có thể là liên hệ nhân quả, chứ không phải ngẫu nhiên hay do một yếu tố khác.

4.3 Tính đặc hiệu (specificity)

Khái niệm đặc hiệu được phát kiến vào thời các bệnh truyền nhiễm còn phổ biến. Khái niệm đặc hiệu có nghĩa là một nguyên nhân chỉ dẫn đến một bệnh (như vi trùng lao *Mycobacterium tuberculosis* chỉ là nguyên nhân của bệnh lao). Nhưng khái niệm này không còn thích hợp cho thời đại của các bệnh mãn tính

phức tạp vốn chịu ảnh hưởng của nhiều yếu tố. Chẳng hạn như hút thuốc lá không chỉ là nguyên nhân của ung thư phổi, mà còn ảnh hưởng đến nhiều bệnh mãn tính khác như xơ vữa động mạch và tiểu đường. Do đó, tuy đặc hiệu là một điều kiện cho mối liên hệ nhân quả, nhưng không phải là điều kiện bắt buộc.

4.4 Tính thời gian (temporality)

Đây là điều kiện liên quan đến thời gian. Điều kiện này có nghĩa là yếu tố nguy cơ hay nguyên nhân phải xảy ra trước khi bệnh xảy ra. Chẳng hạn như nếu loãng xương là nguyên nhân của gãy xương, thì loãng xương phải xảy ra trước biến cố gãy xương. Điều này cũng có nghĩa là các nghiên cứu đoàn hệ và nghiên cứu RCT mới có thể phát biểu về các mối liên hệ nguyên nhân và hệ quả.

4.5 Mối liên hệ liều lượng và ảnh hưởng (biologic gradient)

Theo Hill, một mối liên hệ nhân quả thường thể hiện qua sự liên hệ mang tính liều lượng (dose - response relationship). Chẳng hạn như mối liên hệ giữa hút thuốc lá và ung thư phổi thì điều kiện này có nghĩa là người hút thuốc lá càng nhiều thì nguy cơ mắc bệnh ung thư càng cao. Tương tự, người bị suy giảm mật độ xương càng nhiều thì nguy cơ bị gãy xương càng tăng cao. Trong thực tế cũng có những mối liên hệ không tuân theo mối tương quan tuyến tính, mà có thể là phi tuyến tính, tức là chỉ ở một ngưỡng nào đó của yếu tố phơi nhiễm thì nguy cơ bệnh mới tăng nhanh. Nếu một mối liên hệ không tuân theo qui luật dose - response thì rất có thể đó không phải là mối liên hệ nhân quả.

4.6 Cơ sở sinh học (plausibility)

Một mối liên hệ nhân quả phải có cơ sở sinh học. Chẳng hạn như mối liên hệ giữa mật độ xương và gãy xương có thể giải thích rằng vì mật độ xương có liên quan trực tiếp với sức mạnh

của xương, và do đó người có mật độ xương thấp thì xương rất yếu, khó có thể chịu được một ngoại lực. Tuy nhiên, trong thực tế, có rất nhiều mối liên hệ không thể giải thích bằng cơ sở sinh học, bởi vì đơn giản là chúng ta chưa có mô hình sinh học để giải thích. Do đó, cơ sở sinh học cũng không phải là điều kiện nhất thiết cho suy luận về mối liên hệ nhân quả.

4.7 Tính khúc chiết (coherence)

Điều kiện khúc chiết có nghĩa là mối liên hệ nhân quả thường không mâu thuẫn với những hiểu biết về quá trình phát sinh của bệnh. Chẳng hạn như mối liên hệ giữa hút thuốc lá và ung thư phổi là khúc chiết vì nó nhất quán với nhiều quan sát khác như (i) khi tỉ lệ người hút thuốc lá càng tăng theo thời gian thì số ca ung thư cũng tăng theo; (ii) những nước có nhiều người hút thuốc lá cũng là những nước có nhiều ca ung thư phổi; (iii) bằng chứng từ histology cho thấy bronchial epithelium của người hút thuốc lá khác với người không hút thuốc lá; và (iv) bằng chứng từ nghiên cứu trên chuột cũng cho thấy chuột phơi nhiễm khói thuốc lá có nguy cơ bị ung thư phổi cao. Tất cả những chứng cứ này nhất quán với nhau, và đó là tín hiệu cho thấy mối liên hệ giữa hút thuốc lá và ung thư phổi là mối liên hệ nhân quả.

4.8 Chứng cứ từ thí nghiệm (experiment)

Theo John Stuart Mill, nếu A là nguyên nhân của B, thì nếu tất cả các yếu tố khác cố định, một sự thay đổi A sẽ dẫn đến thay đổi B. Và, thí nghiệm có kiểm soát (controlled experiment) là một phương pháp để kiểm định giả thuyết về mối liên hệ nhân quả giữa A và B. Trong mối liên hệ giữa loãng xương và gãy xương, nếu can thiệp tăng mật độ xương dẫn đến giảm nguy cơ gãy xương, thì đó là một tín hiệu cho thấy mối liên hệ nhân quả.

4.9 Sự tương đương (analogy)

Theo Hill, nhà nghiên cứu phải dùng các ví von giữa các mối liên quan để giải thích (thuyết phục) một mối liên hệ nhân quả. Chẳng hạn như mối liên hệ giữa thuốc thalidomide và rubell, chúng ta có thể chấp nhận một chứng cứ tương tự giữa một loại thuốc với một bệnh truyền nhiễm ở phụ nữ đang mang thai. Tuy nhiên, trong thực tế ít ai nghĩ đến điều kiện này vì nó vừa định tính vừa chủ quan, thiếu tính khoa học!

* * *

Như đề cập ở phần đầu, nếu nghiên cứu được thiết kế theo mô hình RCT thì cách diễn giải kết quả rất đơn giản. Do các yếu tố nhiễu và bias đều được kiểm soát chặt chẽ qua ngẫu nhiên hoá, nên sự khác biệt giữa hai nhóm nghiên cứu (ví dụ như nhóm can thiệp và nhóm chứng) có thể xem như là một mối liên hệ nhân quả. Nhưng đối với phần lớn các nghiên cứu quan sát (như mô hình nghiên cứu bệnh chứng, cắt ngang, đoàn hệ) thì chúng ta không thể kiểm soát hết tất cả các yếu tố nhiễu và bias, do đó vấn đề diễn giải kết quả trở nên phức tạp.

Khi đối diện với một kết quả nghiên cứu có ý nghĩa thống kê ($P < 0,05$), nhà nghiên cứu phải suy nghĩ đến 3 cách giải thích. Thứ nhất, kết quả đó có thể là ngẫu nhiên, và nghiên cứu chỉ “may mắn” có được, chứ không phản ánh mối quan hệ trong thực tế. Thứ hai, kết quả đó có thể là do sự ảnh hưởng của các yếu tố bias, kể cả bias trong việc chọn đối tượng nghiên cứu, bias trong đo lường và thu thập thông tin. Thứ ba, kết quả đó cũng có thể do các yếu tố nhiễu. Các yếu tố nhiễu có thể là

có một biến thứ ba mà nhà nghiên cứu không biết hay không kiểm soát được, có thể là bias vì chỉ định (đặc biệt trong các nghiên cứu về thuốc), có thể là nghịch lý Simpson, nghịch lý môi sinh, hay hiệu ứng hồi qui về số trung bình.

Khi 3 yếu tố ngẫu nhiên, bias và nhiễu được loại bỏ thì vấn đề kế tiếp là câu hỏi: đó có phải là một mối liên hệ nhân quả. Tôi đã trình bày 9 tiêu chuẩn (hay đúng ra là điều kiện) của Bradford Hill để diễn giải một mối liên hệ nhân quả. Không phải bất cứ mối liên hệ nhân quả nào cũng đáp ứng 9 điều kiện này, nhưng đó là những điểm để tham khảo trong việc diễn giải một kết quả nghiên cứu khoa học.

Tham khảo và đọc thêm:

- (1) Ronald L. Wasserstein và Nicole A. Lazar. The ASA's Statement on p -Values: Context, Process, and Purpose. The American Statistician 2016;70:129-133.
- (2) Berger JO, Sellke T: Testing a point null hypothesis: Irreconcilability of P values and evidence (with discussion). J Am Stat Assoc. 1987, 82: 112-139.
- (3) Sellke T, Bayarri MJ, Berger JO: Calibration of p Values for Testing Precise Null Hypotheses. Am Stat. 2001, 55: 62-71.
- (4) L Held. A nomogram for P values. BMC Medical Research Methodology 2010;10:21.
- (5) MacMahon B, et al. Coffee and cancer of the pancreas. N Engl J Med. 1981 Mar 12;304(11):630-3.
- (6) Stampfer M et al. Postmenopausal estrogen therapy and cardiovascular disease. Ten-year follow-up from the nurses' health study. N Engl J Med. 1991 Sep 12;325(11):756-62.

- (7) Grady D, et al. Hormone therapy to prevent disease and prolong life in postmenopausal women. *Ann Intern Med.* 1992 Dec 15;117(12):1016-37.
- (8) Hulley S, et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/Progestin Replacement Study (HERS) Research Group. *JAMA* 1998;280:605-13.
- (9) Rossouw JE, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 2002;288:321-33.
- (10) Final report on the aspirin component of the ongoing Physicians' Health Study. Steering Committee of the Physicians' Health Study Research Group. *N Engl J Med.* 1989 Jul 20;321(3):129-35.
- (11) Cook NR, Hebert PR, Manson JE, Buring JE, Hennekens CH. Self-selected posttrial aspirin use and subsequent cardiovascular disease and mortality in the physicians' health study. *Arch Intern Med.* 2000;160(7):921-8.
- (12) Simpson EH. "The Interpretation of Interaction in Contingency Tables". *Journal of the Royal Statistical Society, Series B* 1951;13: 238-241.
- (13) Hill AB. The environment and disease: association or causation? *Proc R Soc Med.* 1965;58:295-300.

NHỮNG NGUYÊN TẮC TRONG VIỆC SOẠN MỘT BÀI BÁO KHOA HỌC

Sau khi phân tích dữ liệu, bước kế tiếp trong qui trình nghiên cứu khoa học là xuất bản. Xuất bản có nghĩa là công bố kết quả nghiên cứu trên một tập san khoa học có bình duyệt (peer reviewed journals). Chú ý rằng chúng ta bàn về tập san (journal) chứ không phải tạp chí (magazine). Chương này sẽ bàn về ý nghĩa, qui trình, và cách viết một bài báo khoa học.

Đây là tập tài liệu tôi soạn để dùng trong các khóa học (workshop) về cách viết và công bố bài báo khoa học cho các bạn đồng nghiệp Việt Nam. Các khóa học này thật ra là xuất phát từ các 'workshop' bên Úc và Thái Lan, và sau này ở Việt Nam. Các workshop này thường có khoảng 8 bài giảng, xoay quanh phương pháp đặt tựa đề; cách viết phần dẫn nhập, phương pháp, kết quả, và bàn luận; văn phong tiếng Anh trong khoa học.

Tại sao phải công bố báo cáo khoa học?

Trong hoạt động nghiên cứu khoa học, các bài báo khoa học đóng một vai trò rất quan trọng. Nó không chỉ là một bản báo cáo về một công trình nghiên cứu, mà còn là một đóng góp cho kho tàng tri thức của thế giới. Khoa học tiến bộ cũng nhờ một phần lớn vào thông tin từ những bài báo khoa học, bởi vì qua chúng mà các nhà khoa học có dịp trao đổi, chia sẻ và học hỏi kinh nghiệm lẫn nhau.

Một công trình nghiên cứu thường được tài trợ từ các cơ quan nhà nước, và số tiền này là do dân chúng đóng góp. Khi tiến hành nghiên cứu, nhà nghiên cứu có khi phải nhờ đến sự tham gia của tình nguyện viên hay của bệnh nhân. Trong trường hợp đó, tình nguyện viên và bệnh nhân phải bỏ thì giờ, tạm bỏ qua công ăn việc làm để tự nguyện cung cấp thông tin và có khi hi sinh một phần da máu cho nhà nghiên cứu. Nếu một công trình nghiên cứu đã hoàn tất mà kết quả không được công bố, thì công trình nghiên cứu đó có thể xem là có vấn đề về y đức và đạo đức khoa học, và nhà nghiên cứu có thể xem như chưa hoàn thành nghĩa vụ của mình với sự đóng góp của quần chúng. Do đó, báo cáo khoa học trên các diễn đàn khoa học quốc tế là một **nghĩa vụ** của nhà nghiên cứu, là một cách gián tiếp cảm ơn sự đóng góp của bệnh nhân và giúp đỡ của dân chúng qua sự quản lý của nhà nước.

Đối với cá nhân nhà khoa học, báo cáo khoa học trên các tập san khoa học quốc tế là một “currency” (đơn vị tiền tệ). Đó là những viên gạch xây dựng sự nghiệp của giới khoa bảng. Tại các đại học Tây phương, số lượng và chất lượng bài báo khoa học là tiêu chuẩn số một trong việc xét để bạt lên chức giảng sư hay giáo sư. Vì thế công bố báo cáo khoa học, đối với giới khoa bảng Tây phương, là một việc làm ưu tiên hàng đầu của họ. Chính vì thế mà các đại học Tây phương có cái văn hóa gọi là “publish or perish” (xuất bản hay là tiêu tan). Nếu trong vòng 1 hay 2 năm mà nhà khoa bảng không có một bài báo nào đăng trên các tập san khoa học quốc tế, ban giám hiệu sẽ mời vị đó trả lời câu hỏi “tại sao”. Nếu có lí do chính đáng thì còn giữ chức vụ; nếu không có lí do chính đáng thì có nguy cơ mất chức như bõn.

Nói tóm lại, báo cáo khoa học trên các diễn đàn khoa học quốc tế (không chỉ ở trong nước) là một việc làm chính yếu,

một nghĩa vụ, và một điều kiện để tồn tại của một nhà khoa học. Nhưng từ lúc tiến hành nghiên cứu, thu thập dữ kiện đến lúc có báo cáo là một quá trình gian nan. Một công việc còn gian nan hơn nữa là làm sao đảm bảo báo cáo được đăng trên một tập san khoa học có uy tín trên thế giới. Vì thế, các nhà khoa học cần phải đặc biệt chú ý đến việc soạn thảo một báo cáo khoa học sao cho đạt tiêu chuẩn quốc tế. Bài viết này mách bảo một cách thân mật những “mẹo” và kỹ năng để đạt tiêu chuẩn đó.

1. Cách đặt tựa đề bài báo khoa học

Tựa đề bài báo khoa học (hay bất cứ sáng tác nào) là một yếu tố quan trọng. Phần lớn độc giả đọc tựa đề đầu tiên trước khi đọc nội dung bài báo. Có một thống kê của tập san y khoa *JAMA* cho biết cứ 500 người đọc tựa đề thì chỉ có một người đọc toàn văn bài báo, và điều này cho thấy tựa đề là một yếu tố để thu hút người đọc. Ngoài ra, tựa đề còn được dùng như là một văn bản tham khảo ngắn cho thư viện, cơ sở dữ liệu tham khảo, thư mục khoa học, báo cáo, v.v. Do đó, chúng ta có nhiều lý do để quan tâm đến tựa đề, và phần này sẽ hướng dẫn các bạn cách đặt tựa đề cho khoa học.

1.1 Ba loại và hai dạng tựa đề

Giới nghiên cứu khoa học phân biệt tựa đề dựa trên nội dung và cấu trúc. Dựa trên nội dung, tựa đề có thể chia làm 3 loại:

Loại 1 là tựa đề mang nội dung mô tả (descriptive): Ví dụ như “*Genetics and the individualized assessment of fracture*” là loại mô tả, trung dung, không phát biểu một thông điệp nào cả.

Loại 2 là tựa đề mang nội dung tuyên ngôn (declarative), tức tựa đề mang tính khẳng định. Chẳng hạn như tựa đề

“Mutation of PRX gene is associated with increased risk of death” nói thẳng rằng đột biến gen PRX có liên quan đến tăng nguy cơ tử vong.

Loại 3 là tựa đề mang nội dung hỏi (interrogative): Đây là tựa đề thường có tính gợi sự tò mò của độc giả. Ví dụ *“Is estrogen bad for bone?”* là một cách hỏi và tác giả có lẽ hi vọng rằng độc giả sẽ đọc bài báo để biết câu trả lời.

Theo nhiều nghiên cứu về tựa đề trước đây thì tựa đề mang tính trung dung (mô tả) được trích dẫn nhiều hơn hai loại tựa đề kia. Tựa đề hỏi câu hỏi thì được nhiều người tải về đọc, nhưng họ không trích dẫn nhiều. Tựa đề loại phát biểu (tuyên ngôn) cũng được nhiều trích dẫn, nhưng theo nhiều người có kinh nghiệm thì loại tựa đề này dễ làm đồng nghiệp cảm thấy không thoải mái, vì trong khoa học ít khi nào có câu trả lời dứt khoát, do đó tựa đề kiểu “tôi biết chân lý” là khó chấp nhận.

Trong mỗi loại tựa đề trên, có hai dạng tựa đề. Loại thứ nhất là chỉ có một câu văn, và loại thứ hai là tựa đề 2 phần. Tiêu biểu cho tựa đề một câu là *“Network analysis implicates alpha-synuclein (Snca) in the regulation of ovariectomy-induced bone loss”*.

Loại tựa đề hai phần, với khoảng cách là dấu hai chấm. Với tựa đề 2 phần, tác giả có khoảng trống để đưa vào tựa đề một điểm nhấn, thường là phương pháp. Ví dụ như *“Improved survival in homozygous sickle cell disease: Lessons from a cohort study”*, thì phần sau là nhấn mạnh đến dạng nghiên cứu, làm thông tin bổ sung cho phần đầu.

1.2 Năm nguyên tắc đặt tựa đề

Đặt tựa đề cũng là một khoa học, và điều này không phải là nói theo kiểu mỉa mai, mà là một thực tế. Trong quá khứ

đã có rất nhiều nghiên cứu về mối liên quan giữa tựa đề, số chữ trong tựa đề, và tần số trích dẫn của bài báo. Qua những nghiên cứu này, các nhà khoa học có kinh nghiệm đã đi đến vài qui luật (hay đúng hơn là “nguyên tắc») về cách đặt tựa đề sao cho tối ưu. Năm nguyên tắc sau đây cần phải chú ý khi đặt bút viết tựa đề.

Nguyên tắc 1: Một tựa đề tốt nên chuyển tải được những nét chính của nghiên cứu với số chữ ít nhất.

Nét chính của nghiên cứu có thể bao gồm phát hiện chính, phương pháp hay cách tiếp cận mới. Câu hỏi đặt ra là bao nhiêu chữ là vừa? Một nghiên cứu trên tập san y khoa lừng danh *Lancet* cho thấy số chữ trong tựa đề dao động trong khoảng 4 đến 34 chữ. Vẫn theo phân tích này, những bài báo được nhiều trích dẫn có số chữ trung bình trong tựa đề là 18, còn những bài báo có số trích dẫn thấp thì số chữ trong tựa đề là 9. Do đó, theo như kết quả nghiên cứu này, tôi nghĩ số chữ tối ưu trong tựa đề là từ 10 đến 18 chữ.

Như nói trên, tựa đề bài báo nên có một thông điệp, và nếu được, có thêm một yếu tố mới. Những chữ mới có thể là *new*, *innovative*, *novel*, v.v. Ví dụ như bài báo có tựa đề dưới đây báo cáo một mô hình mới để tiên lượng bệnh tiểu đường:

An innovative prognostic model for predicting diabetes risk in the Thai population.

Dĩ nhiên, không phải tựa đề nào cũng cần cái mới (nhiều cái mới quá làm người đọc có khi “dị ứng”). Có khi tựa đề nói lên một thông điệp về cơ chế, ví dụ như bài báo kinh điển của John Vane trên tập san *Nature* (1971):

Inhibition of prostaglandin synthesis as a mechanism of action of aspirin-like drugs.

Cần tránh những tựa đề quá dài và quá “tham vọng”, vì muốn chuyển tải nhiều thông điệp. Những tựa đề như dưới đây (chỉ là ví dụ):

The effect of calcium antagonist felodipine on blood pressure, heart rate, working capacity, plasma renin activity, plasma angiotensin II, urinary catecholamines and aldosterone in patients with essential hypertension.

rất khó hiểu. Một cách đơn giản hóa tựa đề trên là suy nghĩ đến cái thông điệp chính (ảnh hưởng của thuốc calcium antagonist felodipine), còn những chỉ số lâm sàng khác thì có lẽ không cần. Do đó, tựa đề đơn giản hơn là:

Essential hypertension: effect of calcium antagonist felodipine.

Nguyên tắc 2: Bắt đầu tựa đề với từ quan trọng.

Trong thế giới hiện đại, người ta không có nhiều thì giờ để đọc. Người ta không phải tiêu ra 20 giây để biết cái gì quan trọng trong tựa đề. Do đó, một nguyên tắc quan trọng là nên dùng những chữ quan trọng ngay trong phần đầu của tựa đề. Nếu bài báo nhấn mạnh đến yếu tố hút thuốc lá thì có thể nên viết tựa đề:

Smoking is associated with post-fracture mortality.

Nếu yếu tố di truyền là điểm nhấn của bài báo, thì tựa đề có thể là:

Genetics and the individualized assessment of fracture.

Nguyên tắc 3: (Nếu có thể) Tựa đề nên có một điểm mạnh về phương pháp.

Trong y khoa, có nhiều mô hình nghiên cứu, nhưng mô hình randomized controlled trial, hay có khi meta-analysis

được xem là có giá trị cao. Tác giả có thể “khoe” điểm mạnh đó trong tựa đề. Chúng ta có thể xem qua 3 tựa đề dưới đây:

- Zinc supplementation for growth
- Zinc supplementation for growth in preterm infants
- Zinc supplementation for growth in preterm infants: a randomized controlled trial

Tựa đề thứ nhất rất chung chung, nên độc giả khó biết nội dung gì trong nghiên cứu. Tựa đề thứ hai thì tốt hơn tựa đề đầu, vì độc giả biết được đối tượng nghiên cứu. Nhưng tựa đề thứ ba là tốt nhất vì nói lên được điểm quan trọng là công trình nghiên cứu lâm sàng đối chứng ngẫu nhiên.

Do đó, khi đặt tựa đề, tác giả cần phải suy nghĩ và tự hỏi 3 câu hỏi:

- Điểm mới của nghiên cứu là gì?
- Phát hiện chính là gì?
- Có điểm gì “mạnh” về phương pháp.

Chẳng hạn như nghiên cứu ước tính sự ảnh hưởng của yếu tố di truyền mà dùng mô hình nghiên cứu sinh đôi là một thể mạnh, cần phải “quảng cáo”. Do đó, tựa đề có thể là:

Genetic determinants of bone mass: a twin study.

Nguyên tắc 4: Dùng những “từ khóa” (keywords) trong tựa đề.

Như nói trong phần mở đầu, tựa đề bài báo thường được sử dụng trong các thư mục, đặc biệt là Pubmed. Mà, thư mục thì thường dựa vào từ khóa; cho nên tựa đề phải có từ khóa. Còn số từ khóa bao nhiêu có lẽ không quan trọng; quan trọng là điểm nhấn của bài báo. Ví dụ như nếu bài báo dùng siêu âm để đánh giá gãy xương, tôi phải suy nghĩ hai từ khóa quan trọng là “fracture” và “ultrasound”, và tôi có thể đặt tựa đề:

Fracture risk assessment: the role of ultrasound.

Nhưng trong tựa đề này, điểm nhấn là gãy xương; nếu tôi đổi ý và muốn chuyển tải ý tưởng dùng siêu âm thì tựa đề có thể đổi thành:

Ultrasonography as a novel tool for fracture risk assessment.

Nguyên tắc 5: Tựa đề cần phải có thông tin, hay nói theo tiếng Anh là “informative”.

Một tựa đề có thông tin có nghĩa là hàm chứa một thông điệp hay một điểm trong bài báo. Ví dụ như nếu nghiên cứu của tôi tìm ra mối liên quan giữa số tế bào trong máu và mật xương, tôi có thể viết:

Bone density loss is associated with blood cell counts

Nhưng nếu tôi viết “*Bone density loss and blood cell counts*” thì tựa đề này được xem là ít thông tin. Người đọc không biết tôi tìm ra điều gì trong nghiên cứu. Hay như tựa đề “*Postmenopausal osteoporosis*” thì cũng là thiếu thông tin, vì người đọc không biết tác giả muốn nói điều gì. Tựa đề thiếu thông tin rất dễ bị đọc giả bỏ qua.

1.3 Một số điều cần tránh khi đặt tựa đề

Trên đây là 5 nguyên tắc trong việc đặt tựa đề. Trong thực tế, nhiều tác giả có lẽ chưa qua huấn luyện kỹ, nên vẫn phạm phải vài sai lầm trong cách đặt tựa đề. Chỉ cần một vòng xem qua những tựa đề bài báo trên Pubmed, chúng ta dễ dàng thấy những tựa đề mà đọc qua thì thấy rất bí hiểm, vì hình như tác giả chỉ viết cho họ đọc, chứ không phải chia sẻ thông tin với độc giả. Dưới đây là 4 điều cần phải tránh khi suy nghĩ về đặt tựa đề: không dùng biệt ngữ, không viết tắt, tránh chữ chung chung, và tránh chi tiết linh tinh.

(a) *Không dùng những “biệt ngữ” (jargon).* Chúng ta cần nhiều người đọc và hiểu. Do đó, cần phải tránh những tựa đề với những biệt ngữ quá khó hiểu với người ngoài ngành.

(b) *Không dùng chữ viết tắt.* Đây là một điểm cần phải lưu ý, vì thường chúng ta quen với ngành của tác giả và những thuật ngữ chuyên ngành, nên hay dùng chữ viết tắt mà người ngoài ngành có thể hiểu khác. Chẳng hạn như người trong chuyên ngành loãng xương ai cũng biết BMD là gì (bone mineral density), nhưng người ngoài ngành có thể hiểu là “bone mass density”. Dĩ nhiên, những chữ viết tắt nổi tiếng như DNA thì có thể chấp nhận được.

(c) *Không dùng chữ chung chung.* Những tựa đề dùng chữ chung chung như “Calcium supplementation for growth” là không tốt. Đọc tựa đề này, chúng ta không biết “growth” (tăng trưởng) ở đây là tăng trưởng cái gì hay ở đối tượng nào.

(d) *Không để những chữ hay chi tiết không cần thiết.* Những chữ như “Study” hay địa điểm và đặc điểm của đối tượng nghiên cứu là thừa, và do đó không cần thiết. Không cần phải viết:

A study of association between statin and bone loss in women aged 60-90 years in District 1, Ho Chi Minh City, Vietnam.

mà chỉ cần đơn giản hơn, có thể là:

Protective effect of statin on bone loss: a prospective study

hay một cách nhấn khác:

Bone protective effect of statin in elderly women.

là đủ và chuyển tải được thông điệp chính.

1.4 Tình trạng ở Việt Nam

Nhìn qua tựa đề của những luận văn ngành y ở Việt Nam chúng ta dễ dàng thấy chưa đáp ứng các nguyên tắc chung mà tôi trình bày trên. Những tựa đề luận văn ở Việt Nam có xu hướng chung như sau:

- Thường bắt đầu bằng động từ như “Nghiên cứu”, “đánh giá”;
- Thường có địa điểm cụ thể trong tựa đề, như “Nghiên cứu tỷ lệ và nguyên nhân tử vong tại huyện Ba Vì - Hà Nội”;
- Thường có thời gian cụ thể, ví dụ như “Nghiên cứu tỷ lệ và nguyên nhân tử vong tại huyện Ba Vì - Hà Nội năm 2006-2008”; và
- Thường mang tính phát biểu trung dung, không nói lên thông điệp chính.

Tuy nhiên, bài báo khoa học thì không có xu hướng bắt đầu bằng động từ “nghiên cứu”. Phần lớn những bài báo này cũng không nói lên được thông điệp chính của nghiên cứu, và hình như mang tính mô tả là chính.

Tóm lại, đặt tựa đề một bài báo khoa học đòi hỏi phải có suy nghĩ cẩn thận và đầu tư thời gian. Năm nguyên tắc trong việc đặt tựa đề là ngắn gọn (10-18 chữ), tận dụng từ khóa, bắt đầu với từ quan trọng, điểm nhấn về phương pháp, và giàu thông tin. Nên chọn tựa đề hai phần, và tựa đề mang tính mô tả hay phát biểu, vì những tựa đề loại này tính trung bình thu hút độc giả và có nhiều trích dẫn.

2. Cách viết phần Dẫn nhập

Đối với đa số nghiên cứu sinh, viết phần *Dẫn nhập* (Introduction) của một bài báo khoa học là một việc... đau đầu. Viết câu gì trước có khi tốn cả giờ đồng hồ, thậm chí cả ngày, mà nghĩ không ra. Nhưng viết được câu đầu rồi, thì nội dung phần Dẫn nhập phải như thế nào để thu hút người đọc lại là một vấn đề khác lớn hơn và khó hơn. Trong phần này tôi sẽ chia sẻ cùng các bạn cách viết phần Dẫn nhập, kèm theo những câu văn cụ thể để các bạn có thể mô phỏng theo.

Một bài báo khoa học thường được cấu trúc theo công thức IMRaD. **I** là *Introduction*, tức là dẫn nhập. **M** là *Methods*, tức phương pháp. **R** là *Results*, kết quả. **D** là *Discussion*, bàn luận. Phần Dẫn nhập là quan trọng, bởi vì nó dàn dựng bối cảnh cho công trình nghiên cứu. Người đọc đánh giá tầm quan trọng của nghiên cứu qua phần Dẫn nhập. Do đó, trong phần này, bạn phải trả lời cho được câu hỏi “Tại sao tôi thực hiện nghiên cứu này?”

Để trả lời câu hỏi đó (tại sao có nghiên cứu này), bạn cần phải nói theo ngôn ngữ trong ngành là tạo cho mình một “lãnh địa», viết tắt là **CaRS** (creating a research space). Để tạo một lãnh địa hay để biện minh cho nghiên cứu, người viết cần phải chia phần Dẫn nhập thành 3 phần nhỏ:

- Phần thứ nhất là tạo ra một tình thế;
- Phần thứ hai là tạo ra một cái “niche”, hay nói đúng hơn là một vấn đề; và
- Phần thứ ba là chiếm cái niche đó.

2.1 Tạo tình thế

Để tạo tình thế, người viết phải phát biểu cho được vấn đề lớn hiện nay là gì, tầm quan trọng của vấn đề, và tình trạng

kiến thức hiện nay ra sao. Chẳng hạn như nếu tác giả nghiên cứu về gen loãng xương, tác giả phải vào bài nói về loãng xương và hệ quả của nó lớn như thế nào bằng cách dùng các con số thống kê về prevalence và incidence. Để nói về tầm quan trọng của bệnh, tác giả có thể “đoạ” người đọc rằng bệnh này có thể làm cho bệnh nhân giảm tuổi thọ như ung thư vậy. (Nếu nói gãy xương làm giảm tuổi thọ, có khi người ta không “sợ»; phải so sánh với ung thư thì may ra người ta mới sợ). Sau đó, tác giả sẽ viết về những kiến thức hiện hành, như những gen đã được khám phá và vai trò của chúng trong việc giúp chúng ta hiểu hơn về bệnh lí loãng xương.

2.2 Tạo ra cái ‘*niche*’ (lãnh địa)

Đến phần thứ hai của Dẫn nhập, tác giả phải tạo ra một lãnh địa. Ở đây, tác giả phải nói chỉ ra khoảng trống của kiến thức. Phần trên là chỉ ra kiến thức hiện hành, còn phần này là chỉ ra phần cần phải làm thêm. Cái gọi là “khoảng trống tri thức” có thể là hiểu biết hiện nay chưa đầy đủ, là những quan điểm còn đối nghịch và trong vòng tranh cãi, là những câu hỏi mới nảy sinh.

Quay lại ví dụ về nghiên cứu gen loãng xương, tác giả có thể biện luận rằng những gen trước đây được khám phá chỉ giải thích 5% sự khác biệt giữa các bệnh nhân về xu hướng mắc bệnh. Do đó, vấn đề đặt ra là còn gen nào khác, và đây chính là vấn đề cần phải quan tâm. Nếu là một công trình phân tích tổng hợp (meta-analysis), tác giả có thể biện minh rằng câu hỏi beta-blockers có ảnh hưởng gì đến gãy xương hay không vẫn còn trong vòng tranh cãi, vì các nghiên cứu trước đây dựa trên cỡ mẫu hơi nhỏ, và do đó, một phân tích tổng hợp có thể góp phần giải quyết những tranh cãi đó. Nói tóm

lại, đây là phần quan trọng, bởi vì tác giả phải thuyết phục rằng nghiên cứu có lí do tồn tại và có ý nghĩa.

2.3 Chiếm cái *niche*

Sau khi đã tạo ra một lãnh địa, bước kế tiếp là phải “chiếm” lấy lãnh địa đó. Cách chiếm lãnh địa ở đây phát biểu cho được giả thuyết khoa học, và từ giả thuyết, tác giả phải phát biểu mục tiêu nghiên cứu. Mỗi giả thuyết có thể có hơn một mục tiêu nghiên cứu. Trong khoa học xã hội, người ta còn viết về cấu trúc của bài báo trong phần này, thậm chí có người tóm lược những phát hiện chính trong phần “chiếm lãnh địa” nghiên cứu. Tuy nhiên, nói gì thì nói, phần sau cùng của phần Dẫn nhập phải phát biểu cho được những mục tiêu nghiên cứu; không phát biểu được mục tiêu nghiên cứu là rất dở, phản ánh sự thất bại của tác giả.

2.4 Công thức 3C

Cách viết trong phần Dẫn nhập là phải từng bước một và dùng công thức “3C”. Từng bước ở đây có nghĩa là đi từ cái đã biết đến cái chưa biết, và sau cùng là mục tiêu nghiên cứu. Phần này chủ yếu dùng động từ thì quá khứ.

Công thức 3C (**Citation + Critique + Constructive**), tức là có trích dẫn + phê bình + và tính xây dựng. Một phát biểu quan trọng mang tính ý tưởng hay có hàm chứa con số (dữ liệu) cần phải có citation, tức là trích dẫn. Trích dẫn từ những bài báo mà tác giả từ nhóm có uy tín và được công bố trên tập san có uy tín. Tuyệt đối tránh trích dẫn những bài báo từ các tập san dỏm (*predatory journals*). Tác giả không cần phải điểm qua những thông tin hay y văn quá cổ xưa, mà chỉ tập trung vào những nghiên cứu trong vòng 5 năm qua, hay cao

lầm là 10 năm. Nếu được, tránh trích dẫn những nguồn chưa qua bình duyệt (như các tập san bên Việt Nam).

Trong phần tạo ra cái *niche* hay lãnh địa, có khi tác giả phải đi qua các nghiên cứu trước và có khi phê bình. Nhưng ở đây cần phải học cách phê bình có văn hoá, chứ không phải cách phê bình thô lỗ. Phê bình có văn hoá có nghĩa là có chứng cứ từ các nguồn tập san đáng tin cậy, và tác giả không bao giờ nói người khác là “wrong” hay sai. Cần phải tỏ ra, nếu phê bình, tính xây dựng chứ không chỉ trích và đánh phá. Cách nhẹ nhàng nhất là chỉ ra một cách diễn giải khác cho tác giả gốc, hay chỉ ra những nghiên cứu sau này không thể lặp lại kết quả của nghiên cứu trước. Đây là phần rất tế nhị, nên tác giả cần phải tỏ ra là người có văn hoá khoa học. Tuyệt đối không lên giọng giảng dạy người khác trong phần Dẫn nhập.

Đối với các bài báo y khoa, phần Dẫn nhập không nên viết quá dài. Viết dài quá người ta nghĩ tác giả “nhiều chuyện”; viết ít quá người ta nghĩ tác giả thiếu ý tưởng hay lười biếng đọc! Thường chỉ 2 trang A4 (double space) là đủ. Thông thường, người có kinh nghiệm viết phần Dẫn nhập sau khi đã viết phần Kết quả và Bàn luận.

Tóm lại, phần Dẫn nhập phải trả lời câu hỏi “*Tại sao tôi thực hiện nghiên cứu này?*” Phần này viết theo mô hình CaRS. Theo mô hình này và để trả lời câu hỏi đó, tác giả phải thuyết phục vấn đề tác giả quan tâm là quan trọng, nhưng vẫn còn vài khoảng trống về kiến thức (tạo một lãnh địa), và mục tiêu nghiên cứu (chiếm lấy lãnh địa). Nên nhớ rằng phần Dẫn nhập đi từ cái chung (bối cảnh nghiên cứu) đến cái riêng (mục tiêu nghiên cứu); còn phần Bàn luận (sẽ bàn sau) đi từ cái riêng đến cái chung.

2.5 Ví dụ cách viết phần Dẫn nhập

Dưới đây là một nghiên cứu nhằm xác định tiêu chuẩn chẩn đoán béo phì cho người Á châu.

[Vào đầu, tác giả phát biểu về tầm quan trọng của béo phì] *Although obesity is recognized as a global public health problem, the extent of obesity is a matter of contention, due largely to a lack of consensus regarding definition.*

[Giải thích vấn đề chẩn đoán hiện nay] *Clinically, obesity is defined as a condition characterized by excessive body fat to the extent that it is harmful to well being and health (1). Currently, the operational definition of obesity is based on body mass index (BMI). According to the World Health Organization criteria, any individual whose BMI is greater than or equal to 30 kg/m² is considered obese (2).*

[Nhưng đó là cách làm không tốt] *Although BMI is widely used in the diagnosis of obesity, it has been criticized, because it does not distinguish between fat mass, muscle mass, bone and vital organs (3-8).*

[Tại sao mỡ quan trọng?] *It has been argued that a better classification of obesity should be based on percent body fat (PBF), in which any woman whose PBF >35% and any man whose PBF >25% is considered obese (9).*

[Kiến thức hiện hành] *Using the relationship between BMI and PBF, it has been suggested that in Asian populations, a BMI greater or equal to 25 should be classified as obese (10), because a BMI of 25 kg/m² is assumed to correspond to about 25% and 35% body fat for Asian men and women, respectively (9).*

[Tạo cho mình một lãnh địa] *This classification is based on the assumption that for a given BMI, Asians have greater PBF than*

Caucasians (11, 12). However, a close examination of the data on which this assumption is based on (12) reveals little difference in PBF between Chinese in New York and Caucasian women.

[Mục tiêu nghiên cứu] *In this paper, we examine the validity of this assumption by comparing PBF between White American women of European ancestry and Vietnamese women living in Vietnam.*

Cuối cùng tích hợp lại thì sẽ có phần Dẫn nhập như sau:

Although obesity is recognized as a global public health problem, the extent of obesity is a matter of contention, due largely to a lack of consensus regarding definition. Clinically, obesity is defined as a condition characterized by excessive body fat to the extent that it is harmful to well being and health (1). Currently, the operational definition of obesity is based on body mass index (BMI). According to the World Health Organization criteria, any individual whose BMI is greater than or equal to 30 kg/m² is considered obese (2). Although BMI is widely used in the diagnosis of obesity, it has been criticized, because it does not distinguish between fat mass, muscle mass, bone and vital organs (3-8).

It has been argued that a better classification of obesity should be based on percent body fat (PBF), in which any woman whose PBF >35% and any man whose PBF >25% is considered obese (9). Using the relationship between BMI and PBF, it has been suggested that in Asian populations, a BMI greater or equal to 25 should be classified as obese (10), because a BMI of 25 kg/m² is assumed to correspond to about 25% and 35% body fat for Asian men and women, respectively (9). This classification is based on the assumption that for a given BMI, Asians have greater PBF than Caucasians (11, 12). However, a close examination of the data on which this assumption is based on (12) reveals little difference in PBF between Chinese in New York and Caucasian

women. In this paper, we examine the validity of this assumption by comparing PBF between White American women of European ancestry and Vietnamese women living in Vietnam.

3. Cách viết phần Phương pháp

Không nói ra thì có lẽ ai cũng biết rằng phương pháp là phần quan trọng nhất của một bài báo khoa học. Phần này quan trọng là vì gần 70% các bài báo bị từ chối công bố là do khiếm khuyết hay cách mô tả phần phương pháp. Nếu phần kết quả được ví von là trái tim, phần bàn luận là bộ não, thì phương pháp là tay chân và xương của bài báo; xương có vững thì bài báo mới đứng vững được. Do đó, cần phải quan tâm đúng mức đến cách viết phần phương pháp nghiên cứu.

Trong các phần của một bài báo khoa học, phần phương pháp có lẽ là dễ viết nhất. Lí do đơn giản là trước khi một nghiên cứu được thực hiện, hầu hết những chi tiết về thiết kế, chọn đối tượng, đo lường, và phân tích đều đã được trình bày trong một đề cương. Do đó, khi soạn bài báo khoa học, tác giả chỉ cần lấy những chi tiết trong đề cương để chỉnh sửa cho phù hợp với tình hình thực tế. Dĩ nhiên, vẫn phải có thêm các chi tiết phát sinh trong quá trình làm nghiên cứu, và nhất là phần phân tích dữ liệu thì cần phải bổ sung cho từng bài báo riêng lẻ.

3.1 Bốn nguyên tắc viết phần phương pháp

Khi viết phần phương pháp, tác giả cần phải biết đến 4 nguyên tắc chính như sau:

Nguyên tắc 1: Hợp lý nội tại và ngoại tại

Hai khía cạnh quan trọng nhất của một nghiên cứu khoa học là tính hợp lý nội tại (internal validity) và hợp lý ngoại tại

(external validity). Một cách ngắn gọn, **hợp lí nội tại** đề cập đến các phương pháp đo lường sử dụng trong nghiên cứu. Phương pháp đo lường có độ chính xác (accuracy) và độ tin cậy cao (reliability) là yếu tố cần thiết để nâng cao giá trị khoa học của một nghiên cứu. **Hợp lí ngoại tại** (còn gọi là khái quát hoá) thường đề cập đến các tiêu chuẩn tuyển chọn đối tượng cho nghiên cứu. Những đặc điểm của đối tượng nghiên cứu có liên quan đến tính khái quát hóa của kết quả nghiên cứu cho một quần thể lớn hơn. Tuy nhiên, trong thực tế hợp lí nội tại quan trọng hơn hợp lí ngoại tại. Nếu một nghiên cứu thiếu tính hợp lí nội tại thì kết quả cũng không có ý nghĩa hợp lí ngoại tại. Đảm bảo tính hợp lí nội tại là rất quan trọng trong bất cứ công trình nghiên cứu y học nào.

Nguyên tắc 2: Đảm bảo tính tái lập

Một khía cạnh khác phân biệt khoa học thật với khoa học dỏm là tính tái lập (reproducibility). Nếu một nghiên cứu độc lập dùng phương pháp và đối tượng của một nghiên cứu trước thì kết quả cũng sẽ giống nhau; nếu kết quả không giống nhau, thì nghiên cứu trước có thể có vấn đề. Nhưng một vấn nạn ngày nay là có quá nhiều nghiên cứu không mang tính tái lập, và do đó khoa học tính trở thành một câu hỏi lớn đối với khoa học hiện đại. Điều này có ý nghĩa lớn đến cách viết phần phương pháp. Nói cách khác, tác giả phải mô tả phần phương pháp vừa đủ chi tiết sao cho người khác có thể lặp lại qui trình để có cùng kết quả.

Nguyên tắc 3: Chứng cứ

Văn chương khoa học là văn chương chứng cứ. Đối với phần phương pháp, điều này có nghĩa là các phương pháp

chưa phổ biến hay những biến số mới cần phải kèm theo tài liệu tham khảo. Một số phương pháp phân tích thống kê tương đối mới (ví dụ như bootstrap, phương pháp Bayesian Model Averaging) cần phải có tài liệu tham khảo gốc để độc giả có thể tham khảo, nếu cần thiết. Chứng cứ ở đây có khi còn có nghĩa là tác giả phải cung cấp những mã máy tính để người khác có thể lặp lại phân tích trong bài báo khoa học.

Nguyên tắc 4: Chính xác

Về văn phong trong phần phương pháp, cũng như các phần khác là chính xác, ngắn gọn, và trong sáng. Chẳng hạn như khi mô tả phương pháp phân tích dữ liệu, tác giả phải nói rõ biến phụ thuộc là gì, biến tiên lượng và biến nhiễu là gì, cùng những thao tác trước khi đưa vào mô hình phân tích (như hoán chuyển dữ liệu, xử lý giá trị trống). Phần phương pháp phải cung cấp một cách chi tiết những gì tác giả đã làm và làm như thế nào trong công trình nghiên cứu. Ở đây, tác giả phải cẩn thận quân bình giữa hai nhu cầu: súc tích (vì không thể mô tả tất cả các kỹ thuật với những chi tiết chi li) và đầy đủ (tác giả phải trình bày đầy đủ thông tin sao cho người đọc biết được những gì đã làm). Đạt được sự cân đối giữa súc tích và đầy đủ là một thách thức của người viết, và có thể của cả biên tập và nhà xuất bản.

Có thể bài báo đề ra một phương pháp mới, và trong trường hợp đó, tác giả cần phải chú ý những tên gọi (và ý tưởng) xuất hiện nhiều lần trong bài báo. Tác giả nên suy nghĩ kỹ về những tên gọi này: phải dùng chữ ngắn gọn mà dễ hiểu. Nên gọi phương pháp điều trị là gì? Phải sử dụng từ gì để mô tả chỉ tiêu của nghiên cứu? Kinh nghiệm người viết bài này cho thấy trước khi viết cần phải liệt kê ra danh sách những từ hay sử dụng trong bài báo. Không có gì lộn lộn và khó chịu

cho người đọc hơn là dùng nhiều từ khác nhau để gọi một hiện tượng!

3.2 Viết gì trong phần phương pháp?

Trong phần phương pháp, tác giả phải trả lời cho được câu hỏi *tác giả đã làm gì* - *What did you do?* Để trả lời câu hỏi này, tác giả phải cung cấp các thông tin thiết yếu sau đây (tôi tập trung vào các nghiên cứu lâm sàng trong y khoa):

- Mô hình nghiên cứu (study design), địa điểm, và thời gian
- Quy trình nghiên cứu
- Đối tượng nghiên cứu (participants)
- Phương pháp thu thập dữ liệu
- Định nghĩa biến outcome và các biến phân tích
- Cỡ mẫu
- Phân tích dữ liệu

Khi viết phần phương pháp, tác giả có kinh nghiệm thường dùng những tiêu đề nhỏ cho các thông tin trên. Cách cấu trúc như thế giúp cho độc giả dễ theo dõi và đối chiếu với phần kết quả. Do đó, một điều rất quan trọng là **tất cả những gì trình bày trong phần kết quả thì cần phải được đề cập trong phần phương pháp**. *Chẳng hạn như nếu tác giả đề cập đến số người béo phì hay loãng xương trong phần kết quả, thì cần phải định nghĩa rõ thế nào là béo phì và loãng xương trong phần phương pháp.*

Mô hình nghiên cứu

Thông thường đây là câu văn đầu tiên mô tả thể loại hay mô hình nghiên cứu. Ở Việt Nam, nhiều người rất bận tâm với

những chữ để mô tả nghiên cứu, nhưng trong thực tế, chỉ có vài mô hình phổ biến như sau:

- Nghiên cứu cắt ngang hay “thiết diện” (cross-sectional);
- Nghiên cứu bệnh chứng (case-control);
- Nghiên cứu đoàn hệ (cohort hay prospective study);
- Nghiên cứu lâm sàng đối chứng ngẫu nhiên (randomized controlled trial).

Nếu công trình nghiên cứu được thiết kế như là một nghiên cứu cắt ngang, tác giả có thể viết:

“The study was designed as a cross-sectional investigation.”

Nhưng đó là cách viết quá đơn giản (rất phổ biến ở Việt Nam). Không nên. Một đoạn văn phải có ít nhất là 3 câu văn. Viết quá ngắn, độc giả sẽ nghĩ tác giả không có ý tưởng hay không biết viết! Do đó, có thể thêm vài thông tin cần thiết như số bệnh nhân và cách chọn:

“The study was designed as a cross-sectional investigation, in which 210 women were randomly sampled by the cluster sampling scheme. The women aged between 20 and 85 years. The study setting was in Saigon, Vietnam.”

Cũng có những nghiên cứu không nằm trong 4 mô hình phổ biến trên đây, và tác giả phải tìm thuật ngữ thích hợp để mô tả. Chẳng hạn như một nghiên cứu về kháng thuốc trong bệnh viện, không thuộc bất cứ mô hình nào trong 4 mô hình trên, tác giả có thể viết:

“The study was conducted at ABC Hospital in Ho Chi Minh City, Vietnam.”

Nhưng cần phải nói thêm về bệnh viện đó:

"The Hospital is a major tertiary teaching hospital of Ho Chi Minh City, providing care to residents of South Vietnam."

Thậm chí một câu về lịch sử (để độc giả nước ngoài biết rằng đây là một bệnh viện có bề dày lịch sử, và cũng là một cách quảng bá):

"The Hospital has a history of more than 100 years of existence, and currently has 1500 beds, but it is also constantly overloaded with patients."

và vài thông tin về qui mô của ICU:

"The Hospital has 7 intensive care units (ICU), including internal medicine, surgery, coronary events, cardiology, stroke, and neonatal care. The medical ICU (MICU) has 20 beds, with average admission being 5-7 new patients per day (i.e., about 2,500 patients a years). The surgical ICU (SICU) has 65 beds and admits about 19,000 patients a year. The stroke ICU has 6 beds and admits patients from emergency room and stable medical ICU patients. The present study had been conducted in ICUs between November, 2014 and September, 2015."

Một nghiên cứu khác được thực hiện ở Bệnh viện Cần Thơ (ít độc giả quốc tế biết), nên cần phải có một đoạn văn hơi dài:

"The study was conducted at the Can Tho National Hospital (for development cohort) and the Can Tho General Hospital (for validation cohort), Vietnam. The Can Tho National Hospital is a tertiary teaching hospital that serves 17 million residents in the Mekong Delta region; its ED on average admits 75 non-surgical patients per day. The Can Tho General Hospital has 400 beds, providing care to the residents of Can Tho City. On average, the Hospital's ED admits 50 non-surgical patients per day."

Đối tượng tham gia

Đối với những công trình nghiên cứu quan trọng và qui mô lớn, tác giả mô tả phần đối tượng tham gia nghiên cứu rất cẩn thận. Các chi tiết về tuổi, giới tính, tiêu chuẩn chọn và tiêu chuẩn loại trừ đều được mô tả kỹ, vì đây là phần có liên quan đến hợp lý ngoại tại. Ví dụ như công trình về thuốc denosumab (Cummings et al NEJM 2009), tác giả mô tả những phụ nữ được tuyển chọn vào nghiên cứu như sau:

"Women between the ages of 60 and 90 years with a bone mineral density T score of less than -2.5 at the lumbar spine or total hip were eligible for inclusion. Women were excluded if they had conditions that influence bone metabolism or had taken oral bisphosphonates for more than 3 years. If they had taken bisphosphonates for less than 3 years, they were eligible after 12 months without treatment. Women were also excluded if they had used intravenous bisphosphonates, fluoride, or strontium for osteoporosis within the past 5 years; or parathyroid hormone or its derivatives, corticosteroids, systemic hormone-replacement therapy, selective estrogen-receptor modulators, or tibolone, calcitonin, or calcitriol within 6 weeks before study enrollment."

Nghiên cứu ở Cần Thơ mô tả các tiêu chuẩn tuyển chọn và loại trừ nghiên cứu một cách chi tiết:

"We had enrolled medical patients (non-trauma and non-surgical) from the two emergency departments between 13 March 2013 and 31 March 2014. The inclusion criteria were: all patients aged 16 years and older and who could give informed consent. Patients were excluded from the study if they had one of the following conditions: acute coronary syndrome, burns, cardiac arrest before admitting to the hospital or which occurred in the ED with failure of cardiopulmonary resuscitation, snakebite,

insect bite or sting, poisoning (drugs, alcohol, intoxication, paraquat, insecticides, rodenticides, corrosive substances). We also excluded patients with burns, cardiac arrest with failure of cardiopulmonary resuscitation because these patients were deemed to be at high risk of mortality. Women in labor or patients dead on arrival were also excluded from the study. The study protocol and procedure were approved by the Can Tho National Hospital ethics committee."

và một câu văn về y đức:

"All patients gave written informed consent. Patients could withdraw from the study at any time without giving reasons."

Địa điểm và bối cảnh nghiên cứu (setting).

Có nghiên cứu cung cấp thông tin về địa điểm mà công trình nghiên cứu được thực hiện, hay nơi mà dữ liệu được thu thập, bởi vì địa điểm có thể ảnh hưởng đến tính hợp lý ngoại tại của kết quả nghiên cứu. Chẳng hạn như khi làm nghiên cứu về vitamin D, tác giả cần phải cung cấp thông tin về thành phố mà mình thực hiện công trình nghiên cứu:

"The study was designed as a cross-sectional investigation, in which the setting was Ho Chi Minh City (formerly Saigon). The City is located at 10°45'N, 106°40'E in the southeastern region of Vietnam. The City is in the tropic and close to the sea; therefore it has a tropical climate, with an average humidity of 75%. There are only two distinct seasons: the rainy season, with an average rainfall of about 1,800 millimetres annually (about 150 rainy days per year), usually begins in May and ends in late November; the dry season lasts from December to April. The average temperature is 28°C (82°F), the highest temperature sometimes reaches 39°C (102°F) around noon in late April, while the lowest may fall below 16°C (61°F) in the early mornings of late December."

Quy trình nghiên cứu (Procedures).

Trong phần này, tác giả phải tóm lược từng bước nghiên cứu, kể cả những chỉ dẫn cho đối tượng nghiên cứu như thế nào. Việc phân nhóm trong nghiên cứu, chi tiết về can thiệp hay điều trị (nếu có). Nếu công trình có liên quan đến ngẫu nhiên hóa, tác giả cần phải mô tả cụ thể quy trình ngẫu nhiên hóa (randomization) như thế nào, kỹ thuật gì đã được sử dụng để đảm bảo các nhóm cân đối, v.v. Ví dụ:

"All patients who met inclusion and exclusion criteria were invited to participate in the study. Upon giving the written informed consent, data collection was conducted by a trained research worker by using a structured questionnaire. The questionnaire collected data concerning demographic data, medical history, physiological data, haematological and biochemical tests (see Supplementary Table S1 online). After 30 days from admission, the research worker would contact the patient or relative or guardian to obtain information on survival status. Patients who had stayed in hospital for more than 30 days were considered "censored."

Đối với nghiên cứu có xét nghiệm, tác giả phải mô tả chi tiết về cách thu thập sinh phẩm và xử lý ban đầu. Ví dụ:

"All specimens were collected at the bed-side, and then transferred to the microbiology laboratory immediately for inoculation on proper culture media within 2 hours. Sample collection was collected from all patients by bedside bronchoscopy procedure. A 600 mm length and 50 mm view depth fiberscope (PortaView LF-TP; Olympus Optical Co. Ltd, Tokyo, Japan) was attached to a video camera (Olympus, Tokyo, Japan) and focus with white balance was manipulated until optimum view was achieved. The fiberscope was lubricated with sterile xylocaine jelly 2% (AstraZeneca, Sweden). Midazolam and fentanyl and/

or xusamethonium were used in sedation. The fiberscope was then introduced until 2 cm above the carina, where the tube was railroaded into the trachea under vision. Between 2 and 5 ml of bronchoalveolar lavage (BAL) fluid was collected. Samples were then transferred to the microbiology department within 30 minutes after collection for analysis."

Ngoài ra, tác giả phải mô tả cẩn thận kỹ thuật đo lường được sử dụng trong nghiên cứu, như tên của máy, model gì, software phiên bản nào, và nơi sản xuất. Cần phải mô tả điều kiện (nhiệt độ, ánh sáng) trong khi đo lường, cũng như các hệ số về độ tin cậy và độ chính xác của kỹ thuật đo lường.

Ví dụ: "Blood pressure (diastolic phase 5) while patient was sitting and had rested for at least five minutes was measured by a trained nurse with a Copal UA-251 or a Takeda UA-751 electronic auscultatory blood pressure reading machine (Andrew Stephens, Brighouse, West Yorkshire) or with a Hawksley random zero sphygmomanometer (Hawksley, Lancing, Sussex) in patients with atrial fibrillation. The first reading was discarded and the mean of the next three consecutive readings with a coefficient of variation below 15% was used in the study, with additional readings if required."

Định nghĩa chỉ tiêu lâm sàng

Một công trình nghiên cứu lâm sàng phải có một *outcome*, mà tôi tạm dịch là "chỉ tiêu lâm sàng", là một đại lượng dùng làm thước đo của một thuật can thiệp. Do đó, tác giả cần phải định nghĩa rõ ràng chỉ tiêu lâm sàng của công trình nghiên cứu là gì, và nhất là phương pháp đo lường (như vừa đề cập) ra sao. Thông thường, một nghiên cứu có 2 chỉ tiêu lâm sàng mà tiếng Anh gọi là "primary endpoint" (chỉ tiêu chính) và "secondary endpoint" (chỉ tiêu phụ).

Ví dụ: “The primary outcome of the study was mortality within 30 days since the day of admission to the hospital. Mortality was defined as (1) death in hospital from any cause; (2) family-initiated discharge and death either on the way home or within 24 hrs after discharge; (3) doctor-initiated discharge and death at home. It should be noted that in Vietnamese culture, when a patient was in the end stage of disease, the patient or family often requests for discharge from hospital because they prefer to pass away at home.”

Nên nhớ rằng ở phần này tác giả chỉ mô tả những biến có liên quan đến bài báo, chứ không phải mô tả tất cả những biến đã được thu thập trong công trình nghiên cứu. Chẳng hạn như nếu bài báo chỉ nói về mật độ xương, thì tác giả không cần phải nói đến gãy xương (vì hai biến này rất khác nhau). Nguyên tắc là: chỉ mô tả những gì có liên quan đến phần kết quả.

Cỡ mẫu

Cỡ mẫu là một yếu tố rất quan trọng trong một nghiên cứu lâm sàng. Thông thường, các nghiên cứu lâm sàng đối chứng ngẫu nhiên (RCT) phải có một câu văn mô tả cách tính cỡ mẫu. Không phải viết ra những công thức tính toán (như hay thấy trong nhiều bài báo ở Việt Nam), mà là những giả định đằng sau cách tính. Điều này quan trọng, vì qua giả định, người đọc có thể đánh giá khả năng mà công trình nghiên cứu có thể giải quyết câu hỏi đặt ra trong phần dẫn nhập.

Ví dụ: “We consider that the incidence of symptomatic deep venous thrombosis or pulmonary embolism or death would be 4% in the placebo group and 1.5% in the ardeparin sodium group. Based on 0.9 power to detect a significant difference ($p < 0.05$,

two-sided), 976 patients were required for each study group. To compensate for nonevaluable patients, we planned to enroll 1000 patients in each group."

Ngẫu nhiên hóa

Trong các công trình nghiên cứu lâm sàng đối chứng ngẫu nhiên (randomized controlled trial hay RCT), bệnh nhân thường được phân nhóm một cách ngẫu nhiên. Có nhiều cách phân nhóm bằng máy tính và thuật toán, cho nên tác giả có trách nhiệm phải mô tả rõ phương pháp phân nhóm để người đọc có thể đánh giá chất lượng của nghiên cứu. Nếu cách phân nhóm có hiệu quả thì kết quả thường cho thấy các nhóm rất tương đương về các đặc tính lâm sàng. Một ví dụ về cách mô tả phương pháp phân nhóm có thể thấy trong đoạn văn sau đây:

"Women had an equal probability of assignment to the groups. The randomization code was developed using a computer random number generator to select random permuted blocks. The block lengths were 4, 8, and 10 varied randomly."

Mật hóa (còn gọi là *Blinding*). Trong các công trình RCT, có khi cả bác sĩ điều trị và bệnh nhân đều không biết bệnh nhân mình (hay mình) nằm trong nhóm nào của nghiên cứu. Đây là một biện pháp nhằm tăng tính khách quan khi đánh giá hiệu quả của can thiệp. Cũng như ngẫu nhiên hóa có thể thực hiện bằng nhiều thuật toán, cách mật hóa cũng có thể thực hiện bằng nhiều "thủ thuật". Cách mô tả thủ thuật đó có thể tìm thấy trong đoạn văn sau đây:

"All study personnel and participants were blinded to treatment assignment for the duration of the study. Only the study statisticians and the data monitoring committee saw unblinded data but none had any contact with study participants."

Phân tích dữ liệu

Thiết kế và phân tích các nghiên cứu lâm sàng đều cần đến các phương pháp thống kê. Do đó, phần này tuy là phần cuối trong phần phương pháp của bài báo khoa học, nhưng có thể đóng vai trò rất quan trọng. Rất nhiều bài báo và công trình nghiên cứu rất tốt nhưng vì phân tích sai nên đành phải từ chối. Con số bài báo bị từ chối vì phân tích sai có khi lên đến 50% (như với tập san JAMA chẳng hạn). Do đó, trong phần phân tích, tác giả phải phát biểu cho được biến phụ thuộc (hay endpoints hoặc outcome) là gì, biến độc lập (hay risk factors hoặc covariates) là gì, và định nghĩa rõ ràng các biến này được xử lý ra sao. Nếu số liệu đã qua hoán chuyển thì tác giả phải giải thích tại sao. Vì có nhiều phương pháp phân tích số liệu và kiểm định giả thuyết, nên tác giả còn phải giải thích tại sao đã chọn phương pháp A mà không là phương pháp B. Đôi khi tác giả cũng phải nói ra đã dùng phần mềm nào cho phân tích.

Ví dụ về cách viết đoạn văn này như sau:

"All data analysis was carried out according to a pre-established analysis plan. Proportions were compared by using Chi-squared tests with continuity correction or Fisher's exact test when appropriate. Multivariate analyses were conducted with logistic regression. The durations of episodes and signs of disease were compared by using proportional hazards regression. Mean serum retinol concentrations were compared by t-test and analysis of covariance... Two-sided significance tests were used throughout. The analysis was performed with the R statistical language."

Đối với một nghiên cứu RCT, phần này có khi cần phải mô tả chi tiết hơn (Cummings et al NEJM 2009):

"Analyses of efficacy were based on the intention-to-treat principle. To adjust for multiplicity and maintain the overall significance

level at 0.05, the primary end point of new vertebral fracture was required to achieve significance before the next end points in the sequence (nonvertebral fracture and hip fracture) could be tested. Analyses regarding vertebral fractures included all subjects who had at least one follow-up radiograph.

The effect of treatment on the risk of new vertebral fracture was analyzed with the use of a logistic-regression model with adjustment for age strata. An age-stratified Cox proportional-hazards model was used to compare the two study groups for the secondary end points. Score tests were used to calculate P values in each model.^{14, 15} Subjects who were lost to follow-up or withdrew before having a fracture event had their last known fracture status carried forward. Radiographically defined vertebral fractures were analyzed by cumulative incidence and secondary end points by time-to-event analysis with the use of Kaplan–Meier methods. The absolute risk reduction between study groups was computed as the difference in incidence at 36 months for the primary end point and the difference in the Kaplan–Meier estimates at 36 months for the secondary end points with the use of a weighted average across the age strata. Analyses of changes in bone mineral density included all subjects who had at least one follow-up measurement at or before the time point under consideration. Missing values were imputed by carrying forward the last observation.”

Nói chung, phần Phương pháp thường dài gấp 2 hay 3 lần phần Dẫn nhập. Sẽ không có vấn đề gì nếu tác giả mô tả phần Phương pháp một cách chi tiết, vì nếu tập san thấy không cần thiết thì họ sẽ cắt bỏ hay đưa vào phần phụ chú (appendix). Nhưng sẽ là vấn đề nếu tác giả cố tình mô tả phần Phương pháp một cách mù mờ và vắn tắt, bởi vì người duyệt

bài sẽ nghĩ tác giả hoặc là muốn giấu chi tiết, hoặc là thiếu thành thật! Mô tả chi tiết ở mức độ sao cho đồng nghiệp có thể lặp lại nghiên cứu và có kết quả tương đương.

Kinh nghiệm của tôi cho thấy những sai sót phổ biến trong cách viết phần phương pháp là:

- Viết như là những chỉ dẫn (giống như viết powerpoint);
- Nhầm lẫn giữa phương pháp và kết quả;
- Bỏ qua phần giải thích phương pháp;
- Những biến số đề cập trong kết quả mà không mô tả trong phần phương pháp;
- Trình bày những thông tin không liên quan đến bài báo hay phần kết quả.

Xin nhắc lại rằng hơn 70% bài báo khoa học bị từ chối là do phương pháp không đúng hay mô tả không đầy đủ. Vì thế, tác giả cần phải hết sức thận trọng trong phần mô tả Phương pháp nghiên cứu. Viết về công trình nghiên cứu được thực hiện như thế nào cũng quan trọng như thực hiện nghiên cứu. Những điểm cần nói lên cho được trong phần phương pháp là:

- Đã làm những gì
- Làm như thế nào
- Dữ liệu được thu thập ra sao
- Dữ liệu được phân tích như thế nào

Những thông tin và chi tiết trong phần Phương pháp cũng là một cách phát biểu về khả năng và trình độ kỹ thuật - công nghệ của tác giả. Người trong ngành chỉ cần đọc qua vài đoạn văn là có thể biết được giá trị khoa học của công trình nghiên cứu. Do đó, có thể nói rằng phần Phương pháp cũng

là một cách phát biểu về uy tín của tác giả, và sự hiểu ngầm này đòi hỏi tác giả phải hết sức thận trọng trong việc cung cấp những thông tin và chi tiết một cách chuyên nghiệp.

4. Cách viết phần kết quả

Nếu phần Dẫn nhập được xem là bộ mặt của một nghiên cứu, thì phần Kết quả là trái tim của một công trình nghiên cứu. Nhưng viết phần Kết quả có khi là một thách thức đáng kể cho những người mới vào nghiên cứu, vì không biết viết cái gì trước và cái gì sau, hay viết sao cho thuyết phục. Phần này sẽ hướng dẫn các bạn vài điểm quan trọng trong việc viết phần kết quả.

Thông thường, một nghiên cứu cho ra rất nhiều dữ liệu, và quyết định tập trung vào dữ liệu nào hay thông điệp gì để gửi đến độc giả là một quyết định không dễ dàng. Nhiều khi tác giả phải suy nghĩ rất nhiều và tham khảo các nghiên cứu trước. Cách tốt nhất là phác họa một danh sách các bảng số liệu, biểu đồ, hình ảnh thí nghiệm. Các bảng biểu này dựa trên các bài báo trong cùng chủ đề đã được công bố trước đây. Để viết phần kết quả tốt (thường chỉ 2-3 trang A4), các nguyên tắc các bạn cần phải lưu ý:

1. Phân biệt dữ liệu (data) và thông tin (information)
2. Bám sát mục tiêu
3. Dùng thì quá khứ
4. Dùng chữ khách quan
5. Tiết kiệm chữ
6. Dung lượng thông tin
7. Dùng câu văn dẫn nhập
8. Cách viết từng phần kết quả

Tôi sẽ bàn cụ thể từng nguyên tắc dưới đây:

4.1 Dữ liệu không phải là thông tin

Viết phần kết quả có nghĩa là biến những bảng số liệu, biểu đồ và hình ảnh đó... biết nói. *Data* hay dữ liệu là những gì chúng ta đo lường và thu thập được. Còn *information* hay thông tin là ý nghĩa của *data*. Những bảng số liệu, biểu đồ và hình ảnh là *data*, và nhiệm vụ của tác giả là biến *data* đó thành *information*. Chẳng hạn như nếu số liệu về nguy cơ mắc bệnh của nhóm A và B là 0,05 và 0,10, thì đó là *data*, nhưng khi tác giả viết "*The risk of disease in group B was two-fold higher than the risk in group A*" thì đó là *information*. Điều này có nghĩa rất quan trọng là các bạn không lặp lại những con số trong bảng biểu, hay có lặp lại thì cũng phải tối thiểu và chỉ khi nào cần thiết.

Như nói trên, trước khi ngồi xuống viết phần kết quả, tác giả phải có sẵn những dữ liệu trước mặt. Cần có sẵn các bảng số liệu, các biểu đồ, các hình ảnh từ thí nghiệm, và những bài báo trước đây. Trong khoa học, không có bảng biểu thì rất khó viết được gì có ý nghĩa. Nhà văn có thể tưởng tượng và hư cấu câu chuyện để viết, nhà khoa học thì không thể viết từ tưởng tượng mà phải viết ra "câu chuyện" từ dữ liệu thật.

4.2 Bám sát mục tiêu

Cũng như việc mô tả phần phương pháp nghiên cứu, mô tả kết quả nghiên cứu nhất định phải bám sát mục tiêu nghiên cứu. Nếu mục tiêu nghiên cứu là A, B, C, thì kết quả phải trả lời A, B và C. Điều này có nghĩa là tập trung cao độ vào việc mô tả những dữ liệu liên quan đến mục tiêu, không nên để mất tập trung và sa đà vào những tiểu tiết. Chẳng hạn như tránh sa đà vào mô tả chi tiết về phân bố thống kê của một biến số như độ tuổi, vì làm như thế sẽ sao lãng "bức tranh"

lớn hơn của công trình nghiên cứu. Tập trung vào những dữ liệu quan trọng, những dữ liệu giúp tác giả có lí do để viết ra một qui luật hay một thông điệp cho bài báo khoa học.

4.3 Sử dụng thì quá khứ

Phần kết quả chủ yếu dùng thì quá khứ đơn giản (simple past tense). Lí do là các bạn mô tả những gì đã xảy ra. Nhưng thỉnh thoảng tôi vẫn thấy nhiều tác giả dùng thì hiện tại trong các bản thảo, và đây là một sai sót rất căn bản. Chẳng hạn như để mô tả số bệnh nhân tác giả đã tuyển được, có thể viết *“Ultimately, we RECRUITED 250 patients into the study. The average age of the patients WAS 50 years”* (không nên viết *“We have recruited”*, và càng không viết *“the average ... is”*).

Tuy nhiên, khi đề cập đến một bảng số liệu hay biểu đồ trong bài báo thì tác giả phải dùng thì hiện tại. Ví dụ:

“Figure 1 SHOWS the distribution of fasting plasma glucose among the 250 patients”. Nhưng khi mô tả ý nghĩa trong biểu đồ thì lại dùng thì quá khứ: *“The distribution of FPG WAS slightly skewed to the lower values.”*

Trong vài năm gần đây, các tập san khoa học khuyến khích tác giả dùng thể văn tích cực (active voice) thay vì thể văn vô danh. Lí do là thể văn tích cực giúp cho độc giả dễ hiểu hơn và lãnh hội thông tin nhanh hơn là thể văn impersonal. Thử so sánh hai cách viết:

“Greater insulin levels were found to be associated with greater bone density”

và

“We found that greater insulins were positively associated with bone mineral density in younger patients.”

4.4 Dùng chữ khách quan

Văn phong trong phần kết quả không nên cảm tính. Những chữ cảm tính thỉnh thoảng xuất hiện trong phần kết quả là “*highly significant*”, “*very large*”, “*extremely*”, v.v. Những trạng từ đó được xem là cảm tính, và cần phải tránh trong phần kết quả (nhưng có thể dùng trong phần Bàn luận). Do đó, thay vì viết “*The different was highly significant*”, thì nên viết “*The difference was statistically significant (P = 0,001)*”. Con số P có *highly* hay không thì do người đọc cảm nhận, tác giả không nên “nhét chữ” vào miệng người đọc.

Ví dụ trên để khuyến cáo rằng trong phần kết quả, tác giả nên dùng cách viết khách quan. Thay vì viết “*The ANOVA showed that...*” thì nên viết “*We found that*”, vì phương pháp thống kê không có chỉ ra cái gì cả; viết như thế có khi bị hiểu là trốn tránh trách nhiệm.

Trong y khoa (và tôi nhấn mạnh rằng trong y khoa, sinh học) thì phần kết quả là phải trung dung. Trung dung ở đây hiểu theo nghĩa “có sao viết vậy”, không được áp đặt ý nghĩa theo cách hiểu của tác giả vào độc giả. Hãy xem câu văn sau đây:

“Over 60% of patients on the treatment group exhibited a reduction in inflammatory response, illustrating the seriousness of...”

Mệnh đề “*illustrating the seriousness of*” được hiểu là nhét chữ vào miệng người đọc. Cần phải tránh. Chỉ nên viết:

“Sixty-two per cent of patients on the treatment group exhibited a reduction in inflammatory response.”

4.5 Tiết kiệm chữ

Bất cứ bài báo khoa học nào cũng giới hạn số chữ. Do đó, tác giả cần phải hết sức tiết kiệm chữ (nếu được) và đi thẳng vào vấn đề. Những cách viết [phổ biến] như

"As shown in Figure 2, antiresorptive therapies inhibited all bone turnover markers"

thì có thể viết ngắn hơn bằng cách sử dụng ngoặc kép:

"Antiresorptive therapies inhibited all bone turnover markers (Figure 2)."

4.6 Viết với dung lượng thông tin (informative)

Chữ *informative* ở đây có nghĩa là câu văn phải có thông tin. Chẳng hạn như câu *"There was a pronounced increase in the blood levels of vitamin D"* thì không có nhiều thông tin, vì chữ *"pronounced increase"* ai muốn hiểu sao thì hiểu. Nhưng nếu viết rằng *"The average increase in vitamin D levels was 20 mmol/L"* thì được xem là *informative*.

Văn phong trong phần kết quả phải giàu thông tin. Giàu thông tin ở đây cũng có nghĩa là đầy đủ. Không chỉ mô tả con số trung bình, mà còn phải thêm một chỉ số phản ánh độ dao động (như độ lệch chuẩn). Do đó, câu

"Systolic blood pressure was 122.85 ± 12.44 "

là khó chấp nhận được, bởi vì chẳng ai biết những con số đó có nghĩa là gì. Ngoài ra, cũng chẳng có đơn vị thì làm sao người đọc có thể hiểu được. Cũng chẳng có bối cảnh, tức không biết con số đề cập đến nhóm nào. Do đó, câu văn trên có thể sửa lại thành:

“Systolic blood pressure in men (mean $\pm$ SD) was 122.85 ± 12.44 mm Hg”

Thỉnh thoảng, có những bài báo mà tác giả [có lẽ do thiếu kinh nghiệm] mô tả kết quả theo cách dài dòng, và hệ quả là người đọc khó theo dõi. Ví dụ dưới đây là một ca tiêu biểu:

“We followed the subjects for an average of 2 months (range 1 to 6 months); follow-up was 98% complete. Using an intention-to-treat analysis, adjusting for baseline differences in the groups in age, serum LDH level, and history of alcohol use, we found ...”

Người đọc phải đọc 40 chữ để biết tác giả sắp tìm ra cái gì! Đây là cách viết rất Á châu! Nên vào ngay chủ đề

“We found that This effect was present even after adjusting for baseline differences in the groups in age, serum LDH level, and history of alcohol use in the ITT analysis.”

Một cách viết rất dở khác là tập trung vào thống kê như:

“Factor analysis disclosed that eigenvalues were significant for the cognitive and social domains”

thì chẳng có thông tin gì trong câu văn.

4.7 Nếu kết quả phức tạp, có thể dùng một câu dẫn nhập

Thỉnh thoảng, kết quả nghiên cứu rất phức tạp và không dễ hiểu đối với người đọc. Trong trường hợp này, tác giả có thể viết một câu dẫn nhập mang tính giới thiệu để người đọc nắm được vấn đề. Ví dụ:

[Giải thích biểu đồ] *To assess whether the observed relationships between BMD at various sites and body composition were attributable to genetic or environmental factors, multivariate genetic model-fitting analysis (Figure 1) was performed.*

[Giải thích con số có nghĩa gì để giúp người đọc hiểu] *Squared standardised path coefficients (Table 2) can be interpreted as estimates of heritability of specific and decomposed in terms of the portion in common with and independent of other genetic factors. Off-diagonal elements of this analysis were small relative to diagonal elements, which indicate that the majority of heritability of each variable trait is due to specific genetic factors.*

4.8 Cách viết từng phần kết quả

Ở đây, tôi chỉ bàn về cách viết phần kết quả của một nghiên cứu y học. Nhưng cách viết bài báo ngành sinh học, khoa học thực nghiệm, hay khoa học xã hội cũng tương tự như thế. Trong một bài báo y học, phần kết quả thường được chia thành 2 phần chính: phần vào đầu viết về những đặc điểm chính của “đối tượng” nghiên cứu (nhất là nghiên cứu lâm sàng), và phần hai là nhằm trả lời từng câu hỏi nghiên cứu đã được đặt ra trong phần Dẫn nhập. Kinh nghiệm của tôi cho thấy cách viết phần kết quả có thể phân thành 3 “nhóm văn chương” như sau:

- Mô tả kết quả đơn giản;
- So sánh; và
- Mô tả mối liên hệ.

(a) Cách viết kết quả mô tả (*descriptive results*)

Trong phần này, tác giả thường cung cấp cho người đọc biết nghiên cứu có bao nhiêu người (hay động vật), và kèm theo những đặc điểm chính của họ. Những chỉ số thường sử dụng là trung bình, độ lệch chuẩn, tỉ lệ (%), v.v. Thường, những đặc điểm này được gói gọn trong bảng số 1 (gọi là “*table one data*”). Nhưng vì một nghiên cứu có thể thu thập rất nhiều biến

số, nên không thể nào mô tả tất cả bằng chữ được. Điều này có nghĩa là tác giả phải suy nghĩ kĩ mình muốn nói gì. Điều gì muốn nói phải là điều nổi bật nhất, nhưng không lặp lại dữ liệu trong bảng số liệu.

Table 1.
Anthropometric characteristics and lifestyle factors
of study participants

Variable	Men	Women	P-value
Number of participants	355	862	
Age (yr)	43.7 (18.8)	48.6 (16.4)	<0.0001
Number of participants by age group			
18–29	107	134	
30–39	46	95	
40–49	54	209	
50–59	74	212	
60+	74	212	
Weight (kg)	62.0 (9.5)	52.3 (7.7)	<0.0001
Height (cm)	165.1 (6.7)	153.3 (5.6)	<0.0001
Body mass index (kg/m ²)	22.7 (3.0)	22.3 (3.1)	0.013
Whole body fat mass (kg)	15.0 (5.1)	18.2 (5.0)	<0.0001
Percent body fat (%)	24.2 (5.8)	34.7 (5.2)	<0.0001
Whole body lean mass (kg)	43.8 (5.8)	32.0 (4.0)	<0.0001
Whole body bone density (g/cm ³)	1.06 (0.10)	0.98 (0.11)	<0.0001
Current smokers (n; %)	159 (44.8)	7 (0.8)	<0.0001
Alcohol users (n; %)	182 (51.3)	29 (3.4)	<0.0001

Note: Values are mean and standard deviation (in brackets), or actual number of participants and gender-specific percentage (in brackets).

doi:10.1371/journal.pone.0127198.t001

Chẳng hạn như nhìn vào bảng số liệu trên đây mô tả vài đặc điểm lâm sàng chính cho nam và nữ của một nghiên cứu. Nếu là tác giả, bạn sẽ viết như thế nào? Chúng ta chú ý thấy những khác biệt về độ tuổi, chiều cao, trọng lượng giữa nam

và nữ chẳng có gì đáng chú ý hay thú vị, và do đó không đáng để viết ra. Vì đây là một nghiên cứu về béo phì, nên có lẽ điều đáng nói ra là có bao nhiêu người được xem là béo phì theo tiêu chuẩn Á châu ($BMI > 27,5$) mà không có trong bảng số liệu. Cuối cùng, chiến lược để mô tả bằng số liệu này là: (i) báo cho người đọc biết có bao nhiêu đối tượng tham gia nghiên cứu và phân bố nam nữ, theo độ tuổi ra sao; và (ii) tần số béo phì giữa nam và nữ. Nghĩ như thế, có lẽ cách viết cụ thể là:

[Vào đầu]: *"We randomly recruited 335 men and 865 women aged 20 years and above who met all the inclusion and exclusion criteria. The average (SD) of age for men and women was ~44 (19) and 49 (16), respectively. Approximately 42% of men and 50% of women aged 50 years and older, and the difference was statistically significant ($P < 0.0001$)."*

[Mô tả tiếp về BMI] *On average, the average (SD) BMI was 22.7 (3) for men and 22.3 (3.1) for women. [Nếu viết đến đó thì ... chán quá. Sau khi tính toán thêm, chúng ta biết rằng có 15% nam và 17% nữ được xem là béo phì theo tiêu chuẩn Á châu] Using the criteria of $BMI > 27.5$, approximately 15% of men ($n = 53$) and 17% of women ($n = 146$) were classified as "obese".*

[Viết tiếp về tỉ trọng mỡ] *As expected, men had lower percent body fat (PBF) but greater lean body mass than women. The difference in PBF between men and women was almost 2 SDs.*

[Nhấn mạnh đến tỉ lệ hút thuốc lá giữa nam và nữ] *Almost 45% of men and 1% of women self-reported as past and current smokers (Table 1).*

Chú ý trong cách viết trên là phải cụ thể và "informative". Đặc biệt khi viết về tỉ lệ hay phần trăm, phải kèm theo số n thì mới đúng chuẩn. Con số 50% có thể là 1/2 nhưng cũng có thể

là 100/200; do đó, phải kèm theo n thì mới được xem là “khoa học”. Chúng ta có thể viết

“During 6 years of follow-up, 46 patients died in the hospital.”

mới xem qua thì có vẻ chẳng có gì sai, nhưng nếu là tác giả có kinh nghiệm thì nên thêm phần trăm:

“During the follow-up period of 6 years, 46 (23%) of the admitted patients died in the hospital.”

(b) Cách viết so sánh

Một dạng kết quả khác thường mang tính so sánh. Cách viết về so sánh trong khoa học cũng cần phải học, bởi vì trong thực tế rất nhiều tác giả thường viết không đạt hay không đúng. Một phần là do cách chúng ta quen với cách viết tiếng Việt, vốn không phù hợp với cách viết so sánh bằng tiếng Anh. Chẳng hạn như chúng ta có thể nói *“Chiều cao của A cao hơn B”* bằng tiếng Việt, nhưng viết như thế bằng tiếng Anh, *“The body height of A is higher than B”* thì không đúng.

Khi viết so sánh, điều rất quan trọng là phải có nhóm tham chiếu (reference group). Nếu nói *“Nhóm A cao hơn 5%”* thì không thể chấp nhận được, vì không biết 5% là so sánh với nhóm nào hay thời điểm nào. Do đó, khi so sánh, cần phải viết rõ là so sánh với nhóm nào. Ngoài ra, khi mô tả so sánh, nếu được cần thêm khoảng tin cậy 95% hoặc trị số P để người đọc có thể tự mình đánh giá tầm quan trọng. Ví dụ, bệnh nhân với marker chu chuyển xương thấp đáp ứng thuốc cao gấp 2 lần so với bệnh nhân có marker chu chuyển xương bình thường, có thể viết:

“Patients with lower baseline bone remodeling were twice as likely to respond to antiresorptive therapy.”

thì cũng có thể xem là ổn về văn phạm, nhưng chưa “khoa học”, vì (i) thiếu nhóm so sánh, và (ii) thiếu con số về bất định hay

khoảng tin cậy 95%. Nếu khoảng tin cậy 95% là 1,2 đến 3,3, tác giả có thể viết:

“Compared with patients with normal remodeling markers, patients with lower baseline bone remodeling were twice (95% confidence interval, 1.2 – 3.3) as likely to respond to antiresorptive therapy.”

thì ổn hơn và rõ ràng hơn. Tương tự, khi so sánh liên quan đến tỉ lệ sống sót, cách viết tương tự cũng có thể dùng:

“During an average of 6 years of follow-up, survival was 9% (95% confidence interval: 4% – 14%; $P < 0.001$) greater among subjects treated with novastatin than in the control group (Figure 1).”

Mô tả một bảng số liệu hay một biểu đồ đòi hỏi suy nghĩ, giống như... xem tranh. Xem tranh đòi hỏi chúng ta phải nắm được “cái thần”, cái sắc và thông điệp của tác giả. Tương tự, đọc một bảng số liệu, chúng ta phải tìm cái thông điệp trong đó và viết xuống. Có khi cả ngày, chúng ta chỉ viết được vài dòng!

Table 2:

Characteristics of Interns and Residents the Day after Being on Call

Characteristics	Interns (n = 40)	Residents (n = 40)
Alert in rounds	40%	60%
Mood		
Happy	20%	40%
Neutral	30%	30%
Sad	50%	30%
Sleep, mean hours per night	4.2	6.8

Bảng số liệu trên đây so sánh tình trạng làm việc của hai nhóm bác sĩ (interns và residents). Chúng ta thấy chỉ có 40% bác sĩ intern là tỉnh táo trong các clinical rounds, và tỉ lệ này ở bác sĩ resident là 60%. Còn về mood, thì không nói ra chúng ta cũng đoán được số bác sĩ intern “buồn” cao hơn nhóm bác sĩ đàn anh resident. Có lẽ đó không phải điều đáng nói, vì người ta có thể đọc và thấy. Nhưng chú ý rằng nhóm bác sĩ intern ngủ chỉ có 4,2 giờ, so với nhóm đàn anh là 6,8 giờ. Vậy thì câu hỏi đặt ra là viết gì để nói lên cái điểm chính của kết quả nghiên cứu này? Có lẽ chúng ta sẽ viết:

“Results of analysis indicated that only 40% of interns were alert in clinical rounds, and this proportion was significantly lower than that of residents.”

Nhưng đọc lại câu trên thì thấy có vài vấn đề. Những chữ vào để “Results of analysis indicated that” là thừa. Dĩ nhiên là phân tích rồi, nói làm gì! Con số thì chỉ lặp lại những gì trong bảng số liệu. Câu văn hơi dài cho chỉ 2 con số là hơi... nhiều chuyện. Nghĩ lại và viết lại:

“Interns are 0.67 times (or 40% / 60%) as likely to be alert compared with residents”

Nhưng con số 0,67 rất khó cảm nhận. Một cách khác là chuyển giá trị tham chiếu interns / residents thành residents / interns. Với cách chuyển hóa đó, chúng ta sẽ có một thông điệp “ấn tượng” hơn:

“Compared with interns, residents are 1.5 times as likely to be alert (60%/40% = 1.5).”

Khi viết về so sánh, nên tận dụng các chữ sau đây: *Compared with, in contrast, more, greater, bigger, higher, less,*

fewer, lower, similar to, relative risk, odds ratio, twice as likely as, related to, associated with, correlated with.

(c) Cách viết về mối liên quan

Trong các nghiên cứu lâm sàng thường có những phân tích liên quan dùng các mô hình thống kê như hồi qui logistic. Mô tả kết quả phân tích trong các bảng này nhiều khi cũng là một vấn đề nan giải. Chẳng hạn như bảng số 3 sau đây báo cáo về các yếu tố có liên quan đến thoái hóa khớp (osteoarthritis) (3). Những yếu tố nguy cơ bao gồm giới tính, tuổi, BMI, đau khi ngồi chồm hổm, sưng phồng ở khớp xương, lạo xạo xương. Vấn đề là viết gì, vì không thể mô tả từng yếu tố một thì rất... buồn chán.

Table 3:
Risk factors for osteoarthritis of the knee and its
relationship to self-reported pain

Factor	Prevalence ratio (95% confidence interval)	P-value
Gender (women)	1,15 (0,71 – 1,85)	0,565
Age (+5yr)	1,56 (1,43 – 1,71)	<0,0001
BMI (+1)	1,14 (1,07 – 1,23)	0,0001
Pain when squatting (yes)	2,19 (1,42 – 3,39)	0,0004
Body enlargement (yes)	3,54 (1,57 – 8,01)	0,0002
Crepitus (yes)	1,81 (1,18 – 2,79)	0,007

Quyết định viết gì có lẽ dựa trên tiêu chuẩn về “*effect size*”, tức mức độ ảnh hưởng. Trong bảng số liệu 3 *effect size* là chỉ số “*prevalence ratio*” (giống như *risk ratio*, nhưng *prevalence*

ratio dùng cho nghiên cứu cắt ngang). Chúng ta chú ý thấy *effect size* lớn nhất là yếu tố sưng khớp xương, kể đến là đau khi ngồi xổm. Còn các yếu tố như độ tuổi và BMI thì không ngạc nhiên đối với người trong chuyên ngành. Do đó, có lẽ chúng ta sẽ nhấn mạnh đến hai yếu tố quan trọng kia. Cách viết có lẽ là:

"As expected, advancing age and greater BMI were significantly associated with a greater risk of OA of the knee (Table 3). Apart from age and BMI, three knee-related complaints were also independently associated with increased risk of OA of the knee: pain when squatting (PR 2.19; 95% CI 1.42–3.39), bony enlargement (PR 3.54; 95% CI 1.57–8.01), and crepitus (PR 1.81; 95% CI 1.18–2.79)."

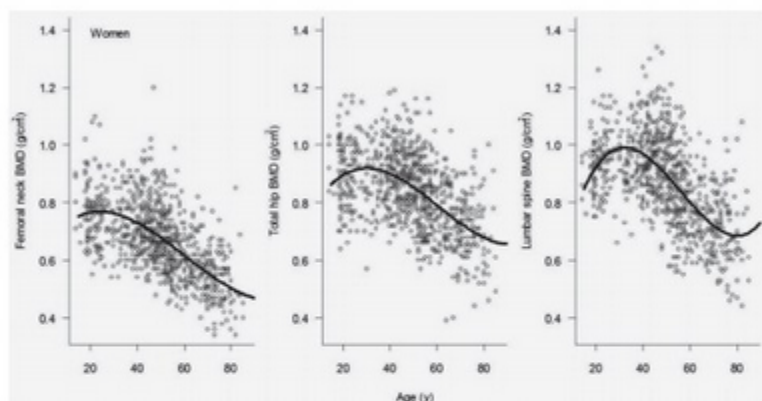
Để cung cấp thêm thông tin, chúng ta nên viết về chỉ số AUC của mô hình:

"The area under the receiver operating characteristic curve (ROC) was 0.83. Gender was not an independent predictor of knee OA when the above factors were considered in the model."

Với cách viết trên, người đọc không cần nhìn bảng số liệu mà vẫn biết được "câu chuyện" chúng ta muốn nói!

Một dạng mô tả mối liên quan khác nằm trong phân tích hồi qui tuyến tính. Chẳng hạn như nhìn vào biểu đồ dưới đây (4) về mối liên quan giữa mật độ xương đo ở ba vị trí (cổ xương đùi, xương đùi, và xương cột sống), chúng ta thấy gì và nên viết gì. Như tôi từng nói (và xin nói lại), nhìn vào một biểu đồ như nhìn vào bức tranh, cần phải có thời gian tĩnh tâm suy nghĩ và nắm bắt cái thông điệp chính. Cái khác giữa nhìn tranh và nhìn biểu đồ có lẽ là ở yếu tố tác giả. Tranh là sản phẩm của người khác, còn biểu đồ là sản phẩm của mình. Là sản phẩm của

mình, mình đáng lí ra phải biết nói điều gì. Nhưng trong thực tế thì chúng ta thỉnh thoảng... bí chữ. Nhìn thấy hình ảnh và xu hướng đó, nhưng viết gì có khi không dễ chút nào.



Trong trường hợp biểu đồ vừa đề cập, chúng ta chú ý thấy cái mô hình chung của các mối liên quan là mô hình đa thức bậc 3. Đặc điểm của mô hình này là giá trị BMD tăng và đạt mức độ cao nhất ở một độ tuổi, và sau đó thì suy giảm. Chú ý rằng BMD đạt đỉnh ở tuổi 20-30. Chúng ta viết:

“The relationship between BMD and age was best described by the third-degree polynomial regression model (Figure).”

[Giải thích tiếp mô thức bậc 3 có nghĩa là gì]: *“The relationship was characterized by three phases, namely, BMD increased between the ages of 18 and 25, followed by a steady period (aged between 25 and 45), and then gradually declined after the age of 45.”*

[Thêm chi tiết không có trong biểu đồ]: *“The age-related decrease in BMD in women was greater than that in men. For example, compared with lumbar spine BMD among women aged*

between 20-30 years, lumbar spine BMD among women aged 70+ years was decreased by 27%; however, in men, the corresponding rate of decrease was ~15%."

Như các bạn thấy, người đọc không cần xem biểu đồ vẫn có thể hiểu xu hướng chung về mối liên quan giữa BMD và độ tuổi.

Ở trên, tôi đã bàn về những gì bạn nên làm trong khi viết phần kết quả. Để kết thúc phần này, tôi nên nói những gì các bạn *không* nên viết. Những điều sau đây là cần phải tránh khi viết phần kết quả:

- **Không lặp lại dữ liệu trong bảng số liệu hay biểu đồ.** Đây là một cách viết rất phổ biến ở những người chưa có kinh nghiệm. Không có lí do gì để lặp lại tất cả các dữ liệu trong bảng số liệu hay biểu đồ. Nếu cần lặp lại thì chỉ chú tâm đến những dữ liệu chính, nhưng dùng cách viết khác. Chẳng hạn như nếu *risk ratio* (RR) là 1,5 với khoảng tin cậy 95% 1,1 - 2,2, thì cách viết có thể là *"the allele was associated with a 50% increase in the risk of CAD."*
- **Không báo cáo những dữ liệu vụn vặt, chi tiết không cần thiết.** Những con số về đặc điểm bệnh nhân không cần phải liệt kê tất cả. Những con số phần trăm không cần phải chính xác đến 2 hay 2 con số lẻ (rất dở và không logic). Cần phải loại bỏ những data không quan trọng, vì chúng có thể làm cho người đọc sao lãng cái data chính yếu.
- **Không báo cáo một danh sách dài các số liệu mà không nói lên điểm gì.** Đây cũng là một cách viết rất phổ biến trong các luận án và bài báo trong nước. Ý đồ liệt kê các số liệu là tốt, nhưng rất tiếc là nó thừa. Như nói ở trên,

liệt kê data không là chưa đủ, bởi vì nhiệm vụ của tác giả là biến *data* thành *information*.

- **Không dùng bảng số liệu hay biểu đồ cho số liệu đơn giản.** Có nhiều bạn dùng biểu đồ bánh chỉ để mô tả con số 45% nam và 55% nữ! Cách trình bày này rất kém tính chuyên nghiệp và cần phải tránh.

Nên nhớ rằng một phần kết quả của một bài báo khoa học được ví von là “trái tim” của công trình nghiên cứu, và điều này đòi hỏi tác giả phải chăm sóc trái tim cho tốt. Chăm sóc qua cách trình bày bằng biểu và hình ảnh, và quan trọng là cách viết.

5. Cách viết phần Bàn luận

Phần Bàn luận có liên quan một cách logic với phần Dẫn nhập và Kết quả. Phần Dẫn nhập nhằm trả lời câu hỏi “*Tại sao tôi thực hiện nghiên cứu này?*” Phần Kết quả thì trả lời câu hỏi “*Tôi đã tìm thấy những gì?*” Còn phần Bàn luận thì nhằm trả lời câu hỏi “*Những kết quả này có ý nghĩa gì?*” Do đó, nếu nói phần Kết quả là trái tim của bài báo, thì phần Bàn luận là bộ não, là “tri thức” của bài báo. Người đọc có thể biết tác giả thuộc đẳng cấp tri thức nào qua đọc phần Bàn luận. Điều này đòi hỏi tác giả phải am hiểu vấn đề và cấu trúc những bàn luận sao cho logic. Nên nhớ rằng phần Dẫn nhập, tác giả đi từ cái chung (bối cảnh) đến cái riêng (mục tiêu nghiên cứu), còn phần Bàn luận thì đi từ cái riêng (phát hiện) đến cái chung (qui luật).

Phần Bàn luận rất khó viết, vì nó không có một cấu trúc cố định và sách hướng dẫn cụ thể về cách viết bài báo khoa học. Mỗi tác giả cho ra một lời khuyên không nhất quán. Có lẽ cái “công thức” tôi trình bày ở đây cũng... chẳng giống ai,

nhưng tôi thấy công thức này rất có hiệu quả. Vài điểm cần lưu ý khi viết phần Bàn luận:

Thứ nhất tác giả phải tỏ ra là người có suy nghĩ kĩ về vấn đề. Lúc nào cũng nhìn một vấn đề qua nhiều khía cạnh, chứ không bao giờ nhìn vấn đề một chiều. Phải tỏ ra mình là người khách quan, và dựa vào chứng cứ chứ không phải với thái độ “trust me” (tin tôi đi).

Thứ hai là phải nói lên được một câu chuyện. Mỗi một bài báo khoa học là một câu chuyện. Câu chuyện dựa trên những “nhân vật” và sự kiện trong phần kết quả. Nhân vật ở đây là những biến số, những yếu tố nguy cơ, những “outcome”. Còn sự kiện là những mối tương quan giữa các biến số. Để nói lên một câu chuyện từ dữ liệu đòi hỏi tác giả phải lên kịch bản và phác họa những ý chính. Sau đó sẽ dùng chữ nghĩa để mô tả câu chuyện sao cho thuyết phục.

Thứ ba là tránh lặp lại kết quả và những con số. Phần bàn luận là phần “tri thức”, tức là phần mà tác giả phải vận dụng một loại ngôn ngữ cấp cao (high-level language), chứ không phải dùng các thuật ngữ. Những con số là chứng cứ và đã được báo cáo trong phần Kết quả, còn phần Bàn luận là có mục đích biến những con số đó “biết nói”. Nói cách khác, phần Bàn luận là chuyển hóa *data* từ Kết quả sang *information* và *knowledge* để độc giả có thể thấu hiểu vấn đề hơn.

Thứ tư là độ dài vừa phải. Trong nghiên cứu y khoa, phần Bàn luận chỉ dao động trong khoảng 3 đến 5 trang A4. Nếu viết ngắn quá, người ta có ấn tượng tác giả không có ý tưởng. Nhưng nếu viết dài quá thì người đọc có thể nghĩ rằng tác giả “nhiều chuyện”, dữ liệu thì ít mà nói thì nhiều.

Tìm chất liệu

Cách viết phần Bàn luận hiệu quả nhất là đọc những nghiên cứu trước. Tác giả nên tìm những bài báo quan trọng về chủ đề mình quan tâm, những bài báo được công bố trên những tập san lớn, của các tác giả có tiếng trong chuyên ngành, in ra và đọc. Đọc phần bàn luận xem các tác giả viết gì, họ quan tâm đến vấn đề gì, cách dùng thuật ngữ ra sao. Những điểm các bạn cần lưu ý và viết xuống khi đọc các bài báo này là:

- Cái thông điệp chính là gì?
- Cơ chế mà tác giả đề nghị để giải thích phát hiện của họ là gì?
- Phương pháp nghiên cứu có điểm gì sáng và chưa sáng?
- Kết luận có dựa vào dữ liệu hay không?

Đó là cách tác giả có thể làm quen với ý văn và những ý tưởng trong chuyên ngành của mình. Từ đó, có thể hình thành những ý có thể trình bày trong phần bàn luận của bài báo.

Viết gì trong phần Bàn luận

Như nói trên, phần Bàn luận là phần khó viết nhất, vì nó không có một cấu trúc nhất định và cố định. Đây chính là lí do tại sao nhiều bài báo có phần Bàn luận đọc rất... chán. Tôi có một “công thức” viết Bàn luận, và đã thử nghiệm qua hàng trăm bài báo, nên biết cấu trúc này có thể giúp nhiều bạn. Công thức viết Bàn luận của tôi rất đơn giản, bao gồm 6 phần:

- **Phần 1:** tóm tắt lí do nghiên cứu và những phát hiện chính;
- **Phần 2:** so sánh kết quả với các nghiên cứu trước;
- **Phần 3:** giải thích cơ chế các mối liên quan;

- **Phần 4:** bàn về ứng dụng hay khái quát hóa;
- **Phần 5:** bàn về những điểm mạnh và điểm yếu của nghiên cứu;
- **Phần 6** là kết luận.

5.1 Tóm tắt

Đây là đoạn văn (chỉ 1 đoạn văn) mở đầu cho phần Bàn luận. Viết một câu rất ngắn để nói về bối cảnh của vấn đề và một câu về khoảng trống tri thức. Câu kế tiếp là nhắc lại những phát hiện chính để lấp vào khoảng trống tri thức. Nên nhớ, và xin nhắc lại rằng, ở đây chỉ dùng chữ chứ không dùng số; nếu có dùng số thì cũng phải rất đơn giản và dùng rất rất ít.

Một ví dụ phần 1 của Bàn luận là đoạn văn dưới đây. Các bạn sẽ thấy, vào đầu, tác giả viết về bối cảnh chung, trong trường hợp này nói rằng ai cũng biết BMD là yếu tố quan trọng để tiên lượng gãy xương:

“There has been little doubt that BMD measured at various sites is one of the best measureable determinants of fracture risk [28-30].”

Sau đó là câu văn nói về vấn đề. Vấn đề là BMD chịu sự tác động của gene, hormone, dinh dưỡng, v.v. Câu kế tiếp là viết về mục tiêu:

“BMD is, in turn, regulated by genetic, hormonal, dietary and mechanical factors. The present study addressed a small part of this complex system by using the classical twin design.”

Kế đến là những câu về phát hiện chính:

“We found that (i) both lean mass and fat mass were associated with areal BMD; however, fat mass alone appeared to have an independent effect on BMD/height ratios and volumetric BMD; (ii) both lean mass and fat mass as well as BMD were under strong

genetic influence and (iii) the association between fat mass (and lean mass) and BMD were mainly mediated through environmental influences.”

Như các bạn thấy cách viết trên theo cái motif người ta gọi là “linear logic”. Mở đầu là nói về BMD, kế đến là yếu tố ảnh hưởng đến BMD, sau đó là vấn đề mà nghiên cứu giải quyết, và sau cùng là những phát hiện chính. Cái điểm nhấn ở đây là BMD, nên nó được nhắc nhở đến nhiều lần trong đoạn văn mở đầu.

5.2 So sánh với nghiên cứu trước

Thông thường kết quả nghiên cứu của chúng ta có thể nhất quán hay không nhất quán với các nghiên cứu trước. Nếu nhất quán thì cái giá trị khoa học của nghiên cứu có thể đáng tin cậy hơn. Nhưng nếu không nhất quán thì chúng ta cần phải giải thích tại sao:

“Our findings are consistent with those of Kay and Rugen, given the wide confidence intervals in both studies.”

hay có thể dùng chữ *confirm*:

“Our study confirms the familial influence on bone density with estimates of heritability for the lumbar spine, femoral neck and total body BMD of 78%, 76% and 79%, respectively, comparable with previous estimates [12-16]. However, the present study also indicates that a common source of genetic ...”

Cách viết về không nhất quán thì hơi tế nhị và phải giải thích tại sao. Những lí do về sự khác biệt giữa kết quả hiện tại và các kết quả nghiên cứu trước đây có thể nhiều, thậm chí rất nhiều, nhưng tựu trung lại là do:

- Nghiên cứu hiện tại là đúng, nghiên cứu trước sai. Nhưng cũng có thể nghiên cứu hiện tại là sai, nghiên cứu trước đúng.

Tuy nhiên, rất khó để nói đúng hay sai ở đây, bởi vì khái niệm đúng/sai trong khoa học tùy thuộc vào nhiều yếu tố khác mà tôi bàn dưới đây.

- Phương pháp đo lường. Nhiều nghiên cứu sử dụng phương pháp đo lường khác nhau, và do đó kết quả cũng khác nhau. Ví dụ như trong lĩnh vực loãng xương, có labo dùng phương pháp DXA để đo mật độ xương, nhưng cũng có labo dùng siêu âm hay một công nghệ khác. Ngay cả hai nhóm dùng phương pháp DXA để đo, thì kết quả vẫn có thể khác nhau, do hãng sản xuất dùng công nghệ khác nhau (như máy Hologic và Lunar chẳng hạn). Do đó, tác giả cần phải tìm hiểu và đọc kỹ những nghiên cứu trước để biết tại sao kết quả của mình khác họ.

- Thiết kế thí nghiệm hay mô hình nghiên cứu. Cho dù hai nghiên cứu dùng công nghệ đo lường như nhau, nhưng cách thiết kế nghiên cứu cũng ảnh hưởng đến kết quả. Có thể nghiên cứu trước thiết kế theo mô hình RCT, còn nghiên cứu hiện tại là cắt ngang. Có thể nghiên cứu trước làm thí nghiệm trên dòng tế bào A, còn nghiên cứu này làm trên dòng tế bào X, và do đó kết quả cũng khác nhau.

- Đối tượng nghiên cứu. Đối tượng nghiên cứu ở đây là tế bào, động vật, con người, v.v. Trong nghiên cứu lâm sàng, chúng ta thu thập dữ liệu trên người, và những đặc điểm cá nhân (như tuổi tác, giới tính, bệnh trạng) có thể rất khác nhau giữa các nghiên cứu. Trước đây khi chúng tôi nghiên cứu về gen loãng xương ở người Việt, kết quả nghiên cứu của chúng tôi đi ngược lại với kết quả công bố trước đây ở phương Tây. Sau khi tìm hiểu và so sánh thì mới phát hiện cái tần số allele của gen ở người Việt và người da trắng rất khác nhau, và do đó nó làm đảo lộn mối liên hệ. Điều thú vị là kết quả của chúng

tôi sau này được một nhóm ở Hàn Quốc lặp lại. Điều này cho thấy đối tượng nghiên cứu có ảnh hưởng rất lớn đến sự khác biệt về kết quả giữa các nghiên cứu.

- Phương pháp phân tích dữ liệu cũng có thể là nguyên nhân dẫn đến sự khác biệt giữa các nghiên cứu. Nhưng nói chung sự khác biệt về phương pháp phân tích chỉ mang tính định lượng chứ ít khi nào làm thay đổi kết luận. Tuy nhiên, trong vài trường hợp hiếm, phương pháp phân tích có thể làm thay đổi toàn bộ kết quả. Chẳng hạn như một nghiên cứu về di truyền học về bệnh tự kỉ mới công bố hồi tháng 3/2017 trên *Nature Genetics*, khi các đồng nghiệp ở Anh phân tích lại thì thấy tác giả bài báo gốc đã sai lầm nghiêm trọng, và do đó kết luận cũng có thể sai (<http://biorxiv.org/content/early/2017/03/12/115964>).

Sau đây là một cách viết về so sánh kết quả. Bài báo này dùng siêu âm để đánh giá khả năng gãy xương:

"Our results are agreeable with previous observations [25-30], in which BUA measurement was significantly lower in individuals with fracture. In women, BUA measurement remained significantly associated with fracture risk, suggesting that BUA at calcaneus is an independent predictor of fracture risk as reported by other authors [18, 32]. However, in men, the combination of BUA and FNBMD did not enhance the predictive value of fracture, which is also consistent with that reported by the MrOS [42]. This indicated that BUA provides little or no additional prognostic information in fracture risk assessment in men. The results observed in men, compared with that of women is likely related to the higher bone mass and lower fracture risk in men."

Có khi tác giả không so sánh được với nghiên cứu trước, bởi vì nghiên cứu của chúng ta là đầu tiên trên thế giới. Trong trường hợp đó, chúng ta có thể viết:

"It is difficult to compare the present study's finding with other studies because of different study designs and clinical outcomes. However, our results confirm the proof of principle study that a profiling of up to 50 genetic variants could improve the fracture prediction"

5.3 Giải thích ý nghĩa của những phát hiện

Mục 3 của phần Bàn luận là giải thích cơ chế của mối liên quan hay ý nghĩa của những phát hiện được báo cáo trong mục 2. Trong phần này, tác giả cần phải chú ý đến 2 điểm:

- Những kết quả này có ý nghĩa gì?
- Đặt ý nghĩa của kết quả trong bối cảnh chung.

Đây là mục quan trọng của phần Bàn luận, bởi vì qua đó mà độc giả có được "tri thức" từ công trình nghiên cứu. Cái "nào" của bài báo chính là ở đây, và do đó, tác giả phải hết sức thận trọng trong cách viết và phải tỏ ra mình có suy nghĩ cẩn thận.

Giả dụ như tác giả tìm ra mối liên hệ giữa yếu tố A và bệnh X (như mối liên hệ giữa thuốc điều trị tiểu đường rosiglitazone và bệnh tim mạch), thì vấn đề đặt ra là phải thuyết phục người đọc rằng phát hiện này là "thật". Thật ở đây có nghĩa là mối liên hệ nhân quả, chứ không phải ngẫu nhiên hay do các yếu tố nhiễu khác. Để thuyết phục người đọc đó là mối liên hệ thật, tác giả cần phải xem xét đến 5 tình huống có thể giải thích cho một mối liên hệ: ngẫu nhiên, bias, confounding, do hiện tượng của bệnh, và sau cùng là do liên hệ nhân quả.

(a) *Kết quả này có phải là do yếu tố ngẫu nhiên?* Để loại bỏ yếu tố này, tác giả có thể dùng trị số P để loại bỏ. Nên nhớ rằng

một kết quả với trị số $P = 0,05$ thì vẫn có thể sai sót, và sai sót (dương tính giả) có thể lên đến 30%. Do đó, nếu kết quả có $P < 0,001$ thì mới “mạnh mẽ” nói, còn chưa đạt được mức độ đó thì nên khiêm tốn và dè dặt khi phát biểu. Dĩ nhiên, không ai muốn nói kết quả của mình là... ngẫu nhiên, nhưng chúng ta cũng phải thành thật để cập đến yếu tố ngẫu nhiên.

(b) *Kết quả này có thể là do bias?* Khái niệm bias rất khó giải thích, nhưng nói chung có thể hiểu bias là sự lệch giữa những gì chúng ta quan sát (kết quả) và sự thật. Có hàng chục, thậm chí hàng trăm bias, mà chúng ta phải để ý. Bias về chọn đối tượng, bias trong đo lường, bias trong cách chọn biến số, bias trong phân tích, v.v. tất cả đều có thể góp phần vào kết quả chúng ta quan sát. Nếu chẳng hạn như chúng ta quan sát thấy tỉ số odds (OR - odds ratio) 6,5, thì chúng ta phải suy nghĩ tại sao kết quả quá cao như vậy. Rất có thể vì nghiên cứu làm trong bệnh viện, và nhóm chứng được chọn chưa thích hợp, nên dẫn đến *bias*. Chẳng hạn như yếu tố nguy cơ trong nhóm chứng ở quần thể là 1000/5000, nhưng trong nghiên cứu bệnh chứng là 20/30 thì rõ ràng là có vấn đề. Nhiệm vụ của nhà nghiên cứu là phải chỉ ra những bias này; không chỉ ra thì người đọc có thể nghĩ tác giả thiên vị.

(c) *Kết quả này có thể phản ánh hiện tượng “effect – cause”?* Trong các nghiên cứu quan sát (tức không phải thí nghiệm và can thiệp), có nhiều khi chúng ta quan sát một mối liên hệ, nhưng đó là mối liên hệ theo kiểu bệnh dẫn đến yếu tố nguy cơ, chứ không phải ngược lại. Ví dụ như nếu chúng ta quan sát qua một nghiên cứu cắt ngang rằng bệnh nhân lao có nồng độ vitamin D thấp, chúng ta khó kết luận rằng vitamin D là yếu tố nguy cơ của bệnh lao phổi, bởi vì bệnh nhân lao có thể ít ra tiếp xúc với ánh nắng mặt trời và do đó họ có nồng độ vitamin

D thấp. Do đó, tác giả cần phải xem xét đến tình huống này (bệnh --> yếu tố nguy cơ).

(d) *Kết quả này có thể là do yếu tố nhiễu (confounding)?*
 Một dạng bias khác có tên là “confounding effect”, có nghĩa là mối liên hệ chúng ta quan sát được thực chất là do một yếu tố khác mà nghiên cứu chưa xem xét đến. Chẳng hạn như chúng ta quan sát rằng người mắc bệnh thoái hóa khớp có mật độ xương thấp hơn người không bị thoái hóa khớp. Đây là một mối liên hệ đáng ngờ, bởi vì người bị thoái hóa khớp thường là béo phì, và béo phì thì thường có mật độ xương cao. Do đó, mối liên hệ chúng ta quan sát [ngược lại với kiến thức sinh học] có thể là do yếu tố nhiễu. Rất có thể những người mắc bệnh thoái hóa khớp là cao tuổi hơn người không bị thoái hóa khớp, và cao tuổi cũng liên quan đến mật độ xương thấp. Trong trường hợp này, tuổi là một confounder bởi vì tuổi vừa liên quan với bệnh và liên quan với yếu tố nguy cơ (mật độ xương). Nếu nghiên cứu chưa điều chỉnh cho độ tuổi thì có thể xem là chưa đạt.

(e) *Nếu tất cả 4 lí giải trên bị loại bỏ thì kết quả này là do hiện tượng nhân quả “cause – effect”.* Như vậy, chúng ta thấy khi viết phần Bàn luận, tác giả phải tỏ ra là người suy nghĩ đa chiều và khách quan, chứ không được suy nghĩ theo kiểu một chiều. Cách suy nghĩ này cũng giống như chúng ta xem xét một tài liệu trên mạng, trước khi tin nó, chúng ta phải tìm cách bác bỏ nó bằng những giải thích khả dĩ. Nói theo ngôn ngữ thông thường là cho dù chúng ta thấy mối liên hệ rõ ràng đó, nhưng chúng ta phải tìm cách bác bỏ nó (vì có nhiều yếu tố chúng ta chưa biết hay chưa xem xét đến), và chỉ khi nào chúng ta bác bỏ tất cả những lí do khả dĩ thì mới tạm chấp nhận đó là mối liên hệ thật.

Sau khi bác bỏ các yếu tố khả dĩ, phần còn lại là mối liên hệ nhân quả. Nhưng tác giả không dừng ở đó, mà phải giải thích tại sao có mối liên hệ này. Trong đoạn văn bản về mối liên quan giữa thuốc chống tiểu đường rosiglitazone và nguy cơ bệnh tim mạch (Nissen, NEJM), tác giả viết như sau. Đầu tiên, tác giả nói rằng cái cơ chế chưa rõ ràng, nhưng ngay sau đó họ dùng dữ liệu trong quá khứ cho thấy rosiglitazone tăng LDL và họ suy luận rằng đó có thể là cơ chế thuốc tăng nguy cơ bệnh tim mạch:

"The potential mechanism(s) for cardiovascular (CV) harm from rosiglitazone use (and the differences from pioglitazone use) remains to be elucidated, but there are several reasonable hypotheses. Rosiglitazone therapy increased low-density lipoprotein cholesterol (LDL-C) levels as much as 23% in trials, leading to approval [21]. [...] Although the FDA has not established a level of increase in LDL-C that is presumed to cause harm, a drug that increases LDL-C levels would reasonably be expected to increase CV adverse events. Interestingly, the lipid effects of the 2 marketed thiazolidinediones, pioglitazone and rosiglitazone, are markedly different".

Trong nhiều trường hợp, chúng ta không giải thích được tại sao, và phải viết như thế. Chẳng hạn như trong các nghiên cứu di truyền học, chúng ta có thể phát hiện nhiều gen, nhưng chúng toàn nằm ở vùng "non-function" (tức không có chức năng) nên chúng ta không biết cơ chế của chúng là gì, và phải nói rõ là ... không biết. Ví dụ như sự ảnh hưởng của [ví dụ] gen HAL đến xương chưa rõ ràng, chúng ta có thể viết: *"The exact mechanism by which HAL gene is associated with bone mass is not known."* Nhưng chẳng lẽ viết một câu ngắn như thế thì chúng tỏ chúng ta chẳng có ý tưởng gì! Vấn đề là phải đặt ra

một giả thuyết về gen HAL và xương để giải thích. Do đó, câu kế tiếp có thể là *“However, a hypothesis can be postulated to account for the association. The BB genotype of HAL gene is known to be associated with greater bone resorption levels [12], and this could be the mechanism of the association between HAL and bone mass.”* Nói tóm lại, nếu không biết thì nói... không biết, nhưng cũng phải kèm theo một giả thuyết để chứng tỏ tác giả là người chịu khó suy nghĩ, chứ không phải lười biếng suy nghĩ.

5.4 Khái quát hóa kết quả nghiên cứu

Một trong những đoạn bàn luận cần phải có là chỉ ra ý nghĩa của kết quả nghiên cứu, nhất là khái quát hóa. Một cách để viết phần này là chỉ ra ứng dụng (nếu được) của kết quả nghiên cứu. Trong ví dụ dưới đây, tác giả bàn về ý nghĩa của mối liên quan giữa thiếu vitamin D ở bệnh nhân lao phổi trong lâm sàng:

Vào đầu viết một câu văn ‘tuyên ngôn’, và ngay sau đó là một chút lịch sử để người đọc an tâm:

“The finding of high prevalence of vitamin D insufficiency in TB patients has a number of clinical implications. Vitamin D in the form of cod liver oil and sunlight exposure was once a therapy for tuberculosis prior to the Robert Koch’s discovery of the etiology of this disease.”

Câu kế tiếp là bàn về việc bổ sung vitamin D cho bệnh nhân:

“The association between vitamin D insufficiency and the risk of tuberculosis suggests that supplementation of vitamin D may help prevent and reduce the severity of tuberculosis.”

và thêm bằng chứng về hiệu quả của bổ sung:

“Indeed, a recent randomized controlled trial has shown that the severity of TB at the end of treatment was less for patients with normal vitamin D status at baseline than for those with vitamin D insufficiency, without adverse effects. However, the vitamin D dose used in the intervention (100,000 IU) is probably too low to warrant a clinical effect.”

và một câu kết:

“These results taken together suggest that low vitamin D status in TB patients, whether cause or effect, might be an important determinant of treatment outcome and comorbidities.”

Một nghiên cứu khác về nguy cơ tử vong ở bệnh nhân gãy cổ xương đùi, tác giả phải đặt kết quả vào một bối cảnh lớn và rộng hơn. Câu hỏi là nếu khái quát hóa kết quả này cho cả thế giới thì mỗi năm có bao nhiêu người chết vì gãy cổ xương đùi. Cái ý đó được tác giả Nissen diễn giải như sau:

“The relevance of study’s finding can be appreciated within the context of the world-wide burden of hip fracture. It has been estimated that in the year 2000, approximately 1.5 million hip fractures occurred world-wide in women and men, aged 60 or older (men 422,000 and women 1.1 million) [25]. With the upper confidence intervals of our estimates of the number needed to harm, we estimated that worldwide, an extra of 120,000 women and 105,000 men would have died within the first-year following a hip fracture. The additional significance of these estimates is that although fracture in men accounts for only a-third of the incidence of hip fracture, the mortality from men accounted for half of excess deaths following a hip fracture.”

5.5 Bàn về điểm yếu và điểm mạnh của nghiên cứu

Bất cứ nghiên cứu nào cũng có những điểm mạnh và điểm yếu, thậm chí khiếm khuyết. Nhiệm vụ của tác giả là phải chỉ ra

những điểm mạnh và yếu này để người đọc có thể cảm nhận và rút kinh nghiệm. Đây cũng là cách để tác giả tự nhìn lại nghiên cứu của mình và chứng tỏ tính khách quan của tác giả.

Thông thường, cách viết đoạn này là viết những điểm mạnh trước, và sau đó là đề cập đến những điểm yếu. Ví dụ như đoạn sau đây tác giả viết về điểm mạnh:

"The present findings must be interpreted in the context of a number of potential limitations. An important strength of this study is that the results are derived from a long-term population based study. The study included both men and women, and was able to compare the relative survival between sexes. Furthermore, by applying relative survival techniques and by accounting for the expected mortality in the background population of a similar age, sex and calendar period, we have addressed the proportion of mortality attributable to osteoporotic hip fracture."

sau đó là những điểm yếu:

"However, the sample size (e.g., the number of hip fracture cases) in the study was relatively modest, which limited our ability to examine factors with low effect sizes. Importantly we lack information of the smoking and prior fracture status of the background population and therefore the expected mortality in these groups."

Thỉnh thoảng cũng nên tỏ ra "critical" với chính kết quả của mình. Chẳng hạn như có thể viết thẳng rằng *"Our results could also be interpreted as showing that therapy with calcium is ineffective in preventing osteoporosis. Although bone density was increased, fracture rates were no different than in the control group."* Nếu kết quả với mức độ ảnh hưởng thấp cho dù có ý nghĩa thống kê, tác giả cũng có thể viết *"Though the differences we observed were statistically significant, they were generally small and may not be clinically meaningful."*

5.6 Kết luận

Đây là đoạn văn cuối cùng của phần Bàn luận và là đoạn văn như một nốt nhạc “finale”. Hãy tưởng tượng bạn đang nhìn vào mắt độc giả và nói một câu để độc giả tin là bạn nói thật và nói cái câu mà tiếng Anh gọi là “Take-home message”. Cái thông điệp đem về nhà đó có thể làm thay đổi niềm tin, thay đổi cách làm, hay thay đổi nhận thức của người đọc. Đó là những điểm cần lưu ý để viết câu kết luận.

Đoạn văn dưới đây là kết luận về mối liên hệ giữa mật độ xương và thành phần cơ thể. Chú ý câu sau cùng là một thông điệp về lâm sàng:

“In conclusion, these data indicate that the clinically relevant association between volumetric BMD and body composition is mediated only through fat mass. Furthermore, lean mass and fat mass, as with These data also suggest that modulation of environmental factors could translate to clinically relevant changes in BMD and presumably fracture risk.”

Trong phần kết luận của một bài báo khoa học, tác giả phải có cái “conviction” (lập trường), chứ không nên như là một “wimp”. Tránh những câu văn vô duyên và làm mất thì giờ của người đọc như:

- Additional research is needed.
- Our findings may have clinical significance.
- Further studies to confirm these findings would be helpful.
- The meaning of these results is uncertain.
- In conclusion, we conclude that we found what we just said we found.

Tránh những kết luận mang tính suy đoán, võ đoán; những kết luận mang tính áp đặt. Đặc biệt không bao giờ viết kết luận mà không dựa vào chứng cứ. Cũng không nên đề nghị bất cứ cái gì trong tương lai, bởi vì đó chưa hẳn là nội dung của bài báo. Ở Việt Nam người ta có thói quen (hay qui định?) là phần cuối bài báo phải có đề nghị can thiệp hay chính sách, nhưng trong thế giới khoa học thì những đề nghị và khuyến cáo không dựa trên bằng chứng can thiệp là không chấp nhận được.

Nói tóm lại, phần Bàn luận là cái não của bài báo khoa học. Công thức viết bàn luận là 6 điểm như trình bày trên: (i) giới thiệu lí do nghiên cứu và tóm tắt kết quả chính; (ii) so sánh với nghiên cứu trước; (iii) giải thích cơ chế; (iv) giải thích ý nghĩa và khái quát hóa; (v) bàn về những điểm mạnh và những điểm yếu; và (vi) kết luận. Công thức đó đã được tôi thử nghiệm nhiều lần với nhiều nghiên cứu sinh, và có thể nói là lần nào cũng thành công. Tôi hi vọng các bạn có dịp ứng dụng công thức trên cho bài báo của mình và hi vọng rằng kết quả cũng sẽ tốt.

6. Cách viết Abstract

Tựa đề và bản tóm tắt (abstract) là hai phần chiếu đầu tiên đập vào mắt của người đọc. Đây cũng là phần mà đại đa số người đọc đọc trước khi quyết định có nên đọc tiếp hay không. Tất nhiên, tựa đề và bản tóm tắt là hai phần được đưa vào danh mục của thư viện điện tử. Do đó, nhà nghiên cứu cần phải để tâm suy nghĩ cẩn thận khi soạn hai phần này sao cho thu hút sự chú ý của người đọc. Nhiều người nói rằng *abstract* là văn bản viết trước, nhưng trong thực tế thì đây là phần sau cùng của một bài báo khoa học.

Các bạn thử đọc qua abstract sau đây:

"This paper presents and assesses a framework for an engineering capstone design program. We explain how student preparation, project selection, and instructor mentorship are the three key elements that must be addressed before the capstone experience is ready for the students. Next, we describe a way to administer and execute the capstone design experience including design workshops and lead engineers. We describe the importance in assessing the capstone design experience and report recent assessment results of our framework. We comment specifically on what students thought were the most important aspects of their experience in engineering capstone design and provide quantitative insight into what parts of the framework are most important."

Có nhiều điều cần học từ bản tóm tắt này. Học cái không nên viết.

(a) Bản tóm tắt này bắt đầu bằng một mục tiêu, nhưng ngay sau đó thì tác giả bắt đầu lan man từ các chi tiết mang tính kĩ thuật, và đến cuối cùng thì độc giả vẫn không nắm được cái thông điệp chính của bài báo.

(b) Cách viết thì dùng thể văn cá nhân, kiểu *"We explain"*, *"we discuss"*, *"we comment"*, v.v. tuy không có gì sai, nhưng dùng những *I, We* nhiều quá làm cho độc giả có thể cảm thấy khó chịu.

(c) Toàn bộ bản abstract không hề có một kết quả gì để độc giả lãnh hội. Bất cứ một bài báo nào cũng có vài kết quả, hoặc là định lượng hoặc là định tính, nhưng bài này thì không thấy trình bày trong bản tóm tắt (*"Next, we describe... We comment specifically on what students thought were the most important aspects of their experience in engineering capstone design..."*).

Tác giả chỉ nói về cơ cấu bài báo! Nói theo dân nhà nghề là “chỉ có súp mà không có thịt.”

Nhìn chung, tôi thấy bản tóm tắt này viết chưa đạt. Đọc xong bản abstract, người đọc không thấy mình lãnh hội được thông tin gì mới, cũng chẳng có một thông tin gì cụ thể về phương pháp, và quan trọng nhất là không có thông điệp chính. Những thiếu sót này cần phải tránh khi viết bản abstract.

Chữ *Abstract* trong tiếng Anh thật ra có nguồn từ tiếng Latin, với *ab* có nghĩa là phía ngoài, và *trahere* có nghĩa là rút ra. Do đó, *abstract* có thể hiểu nôm na là rút những nét chính ra ngoài. Trong thực tế, abstract hay bản tóm tắt, như tên gọi, là một văn bản lược giản những nội dung chính của một bài báo. Do đó, bản tóm lược phải là một văn bản độc lập, hiểu theo nghĩa người đọc không cần đọc toàn bộ bài báo mà vẫn nắm được mục tiêu nghiên cứu, phương pháp, kết quả chính, và kết luận của công trình nghiên cứu.

Do đó, bản tóm tắt đóng vai trò quan trọng. Thật vậy, cứ 500 người đọc bản tóm tắt thì chỉ có 1 người đọc toàn văn bài báo. Tổng biên tập tập san JAMA từng nói rằng bản tóm tắt là một văn bản rất quan trọng, nhưng thường được viết rất dở. Tất cả những thông tin trên phải được viết trong vòng 250 đến 300 từ (tùy theo qui định của tập san). Viết bản tóm lược cho tốt, do đó, là một thách thức đáng kể. Cái gì cũng cần phải có nguyên tắc, và soạn bản tóm tắt cũng không phải là một ngoại lệ. Dưới đây tôi trình bày 5 nguyên tắc để viết bản tóm tắt: cái mới, tầm quan trọng, chứng cứ, ngắn gọn, và trong sáng (4 nguyên tắc đầu là do Bs Petrache và Diette của tập san Thoracic đề xướng).

Nguyên tắc 1: cái mới

Tâm lý độc giả khoa học thường hay chú ý đến cái mới. Khi đọc một bài báo họ tự hỏi có cái gì mới trong bài báo

này. Do đó, tác giả cần phải nghĩ đến điểm gì mới của bài báo, và điểm mới này phải được viết ra ngay trong bản tóm tắt. Có thể bắt đầu ở phần bối cảnh (background), với một câu văn như:

“Trabecular bone score has emerged as an important predictor of fragility fracture, but factors underlying the individual differences in TBS have not been explored. In this study, we sought to determine the genetic contribution to the variation of TBS in the general population.”

Một ví dụ khác:

“It has been widely assumed that for a given BMI, Asians have higher percent body fat (PBF) than whites, and that the BMI threshold for defining obesity in Asians should be lower than the threshold for whites. This study sought to test this assumption by comparing the PBF between US white and Vietnamese women.”

Trong câu trên, chúng ta thấy tác giả chỉ ra ngay cái khoảng trống tri thức về TBS ngay từ câu đầu tiên. Câu thứ hai chính là mục tiêu của nghiên cứu. Chỉ cần ngắn gọn như thế là có thể thu hút người đọc.

Nguyên tắc 2: tầm quan trọng

Những độc giả y khoa thường quan tâm đến cái tiếng Anh gọi là “significance” hay tầm quan trọng của nghiên cứu. Họ hỏi “Nghiên cứu này có làm thay đổi nhận thức của tôi, hay có làm thay đổi thực hành lâm sàng?” Do đó, tác giả cần phải nghĩ đến cái gì là quan trọng trong công trình nghiên cứu. Cái quan trọng có thể liên quan đến - nhưng không nhất thiết phải là - cái mới. Ví dụ như bài báo về khám phá 56 gen có liên quan đến loãng xương (Nat Genet 2012), các tác giả viết như sau:

“These findings shed light on the genetic architecture and pathophysiological mechanisms underlying BMD variation and fracture susceptibility.”

hay như bài báo về kháng thuốc ở bệnh nhân ICU, tác giả viết:

“These data indicated that critically ill patients on ventilator in Vietnam were at disturbingly high risk of antimicrobial resistance”.

Cách viết ngắn gọn đó nói lên cái tầm ảnh hưởng của phát hiện được trình bày trong abstract.

Nguyên tắc 3: chứng cứ

Một bản abstract tốt là phải có thông tin định lượng, và kết luận dựa trên thông tin đó. Nếu kết quả cho thấy mối liên quan có ý nghĩa lâm sàng mà kết luận cho rằng không có ý nghĩa lâm sàng là rất không nên. Chẳng hạn như bài báo gần đây trên JAMA về ảnh hưởng của calcium và vitamin D (CaD) trên ung thư, phần kết quả tác giả ghi nhận rằng bổ sung CaD giảm nguy cơ ung thư 30% (Lappe et al, JAMA 2017), nhưng khi kết luận thì họ viết:

“supplementation with vitamin D3 and calcium compared with placebo did not result in a significantly lower risk of all-type cancer at 4 years.”

Lại có nhiều bản abstract được viết ra với lời kết luận thiếu chứng cứ hay không hợp với chứng cứ. Đó là một điều... đại kị. Chẳng hạn như nghiên cứu phát hiện mối liên hệ giữa hút thuốc lá và ung thư, thì tác giả lại kết luận rằng nếu cấm hút thuốc lá thì sẽ ngăn ngừa ung thư. Lại có những kiến nghị theo sau mỗi bài báo. Nếu nghiên cứu chỉ phát hiện mối liên quan,

tác giả chỉ có thể bàn về mối liên quan, chứ không thể nói gì về ngăn ngừa do chưa có chứng cứ về can thiệp.

Nguyên tắc 4: ngắn gọn

Bản abstract là một văn bản rất... hà tiện. Số chữ có giới hạn (thường 100 đến 300 chữ), nhưng phải phản ánh nội dung của một công trình nghiên cứu có khi kéo dài đến 10 năm. Do đó, cần phải tập cách viết ngắn gọn, sao cho không có chữ nào là thừa và cũng không có chữ nào thiếu. Nghĩ đến bản tóm tắt như là một chapeau của giới báo chí, tức chỉ vài chữ mà độc giả nắm được nội dung chính của bài báo.

Nguyên tắc 5: trong sáng

Bản tóm tắt là một văn bản khoa học, mà khoa học thì đòi hỏi một văn phong trong sáng, trong sáng, và... trong sáng. Trong sáng ở đây phải hiểu là không để độc giả có thể nghĩ ra một cách diễn giải khác. Để đạt được yêu cầu này, tác giả phải tập cách viết câu văn ngắn (không quá dài), cấu trúc đơn giản, và có định lượng. Ví dụ như các câu văn dưới đây thì người ta không thể hiểu khác:

"The study involved 1765 patients in the development cohort and 1728 in the validation cohort. The main outcome was mortality up to 30 days after admission."

nhưng nếu viết

"The study involved two cohorts of patients from two independent hospitals. The main outcome was death after ICU admission"

thì không thể xem là trong sáng được.

Để đáp ứng 5 nguyên tắc trên, tác giả cần phải lưu ý 5 điểm dưới đây:

6.1 Abstract là một văn bản độc lập với bài báo

Bản tóm tắt tuy là một phần của bài báo khoa học, nhưng nó là một văn bản độc lập. Nếu các bạn nộp bài báo khoa học cho tập san, người ta thường yêu cầu nộp văn bản abstract riêng trong hệ thống online. Khi nói “độc lập” tôi cũng có ý nói rằng abstract là một bài báo ngắn. Nói ví von một chút thì abstract là một bài thơ của bài báo. Chúng ta biết rằng thơ cần phải được viết một cách cô đọng, và đó chính là cách viết phần abstract.

6.2 Abstract là một tổng quan câu chuyện chính

Vì là một văn bản độc lập, nên bản abstract phải là tổng quan của “câu chuyện”. Câu chuyện ở đây là toàn bộ công trình nghiên cứu. Phải viết sao cho người đọc chỉ cần đọc qua bản abstract là biết được câu chuyện mà tác giả muốn chuyển tải đến người đọc.

6.3 Mỗi phần trong bài báo cần phải được đề cập vài nét chính

Vì là tổng quan, cho nên mỗi phần trong bài báo như dẫn nhập, phương pháp, kết quả, bàn luận và kết luận phải hiện diện trong abstract. Dĩ nhiên, vì là “bài thơ”, nên mỗi phần như thế chỉ hiện diện rất ngắn gọn. Có lẽ phần kết quả là phần dài nhất của abstract, nhưng cũng chỉ có thể đưa vào những phát hiện chính, chứ không thể nào viết vài dòng được.

6.4 Một abstract cần/nên có những thông tin về kết quả và kết luận

Rất nhiều abstract, nhất là theo cách viết “truyền thống” ở trong nước, thường chỉ có những câu văn mang tính “hứa hẹn”. Nhưng cách viết đó không còn thích hợp nữa, nhất là trong thời đại người ta bị quá tải về thông tin. Do đó, một bản abstract phải

có thông tin (information content), và thông tin cần thiết nhất là kết quả nghiên cứu. Nếu có thông tin định lượng thì càng tốt. Ví dụ như một nghiên cứu với mục tiêu chính là đánh giá hiệu quả của một thuốc điều trị, thì thông tin cần phải có trong bản abstract là các chỉ số như tỉ số nguy cơ (relative risk) hay risk difference. Nếu nghiên cứu có mục tiêu đánh giá mối liên quan thì cần phải có kết quả về tỉ số odds để độc giả có thể nắm được kết quả chính. Ngoài ra, bản abstract nhất định phải có một câu văn kết luận. Nếu còn dư “đất” thì thêm một câu văn về ý nghĩa của phát hiện vừa trình bày trong phần kết quả.

6.5 Giới hạn trong vòng 100 - 300 chữ

Các tập san y sinh học thường giới hạn số chữ trong bản abstract, và số chữ cho phép dao động trong khoảng 100 đến 300 chữ, kể cả khoảng trống. Ngày nay, có một số tập san còn yêu cầu tác giả phải có phần “min-abstract” mà tác giả phải viết trong vòng 200 mẫu tự để nói về nghiên cứu của mình. Mini-abstract có thể xem như là một chapeau của giới báo chí. Tóm tắt một bài báo 20-30 trang (có khi 50 trang) xuống còn 300 chữ hay 200 mẫu tự là một thách thức lớn, nhưng rất hào hứng. Nó luyện tác giả phải chọn chữ cho thích hợp và tập trung suy nghĩ về bài báo của mình. Do đó, cần phải đầu tư (có khi cả ngày suy nghĩ) chỉ để viết 300 chữ!

Có 2 loại tóm tắt: không có tiêu đề và có tiêu đề. Loại tóm tắt không có tiêu đề là một đoạn văn duy nhất tóm tắt công trình nghiên cứu. Loại tóm tắt có tiêu đề - như tên gọi - là bao gồm nhiều đoạn văn theo các tiêu đề sau đây: *Background*, *Aims*, *Methods*, *Outcome Measurements*, *Results*, và *Conclusions*. Tuy nhiên, dù là có hay không có tiêu đề, thì một bản tóm tắt phải chuyển tải cho được những thông tin quan trọng sau đây:

Câu hỏi và mục đích của nghiên cứu. Phần này phải mô tả bằng 2 câu văn. Câu văn thứ nhất mô tả vấn đề mà tác giả quan tâm là gì, và tình trạng tri thức hiện tại ra sao. Câu văn thứ hai mô tả mục đích nghiên cứu một cách gọn nhưng phải rõ ràng. Ví dụ:

“Antimicrobial resistance has emerged as a major concern in developing countries. The present study sought to define the pattern of antimicrobial resistance in ICU patients with ventilator-associated pneumonia.” Câu đầu tiên là nói về vấn đề lớn, câu thứ hai là mục tiêu.

Phương pháp nghiên cứu. Cần phải mô tả công trình nghiên cứu được thiết kế theo mô hình gì, đối tượng tham gia nghiên cứu đến từ đâu và đặc điểm của đối tượng, phương pháp đo lường, yếu tố nguy cơ (risk factors), chỉ tiêu lâm sàng (clinical outcome). Phần này có thể viết trong vòng 4-5 câu văn. Ví dụ:

“Between November 2014 and September 2015, we enrolled 220 patients (average age ~71 yr) who were admitted to ICU in a major tertiary hospital in Ho Chi Minh City, Vietnam. Data concerning demographic characteristics and clinical history were collected from each patient. The Bauer–Kirby disk diffusion method was used to detect the antimicrobial susceptibility.”

Một ví dụ khác:

“The study involved 1765 patients in the development cohort and 1728 in the validation cohort. The main outcome was mortality up to 30 days after admission. Potential risk factors included clinical characteristics, vital signs, and routine haematological and biochemistry tests. The Bayesian Model Averaging method within the Cox’s regression model was used to identify independent risk factors for mortality.”

Hai ví dụ trên cho thấy trong phần phương pháp, tác giả phải chỉ ra cho được số lượng bệnh nhân, mô hình nghiên cứu, và những đo lường chính. Có khi phương pháp phân tích đóng vai trò quan trọng trong bài báo, và do đó, cần phải đề cập đến (như trong ví dụ 2).

Kết quả. Trong phần này, tác giả trình bày những kết quả chính của nghiên cứu, kể cả những số liệu có thể lấy làm điểm thiết yếu của nghiên cứu. Nên nhớ rằng kết quả này phải được trình bày sao cho trả lời câu hỏi nghiên cứu đặt ra từ câu văn đầu tiên.

Ví dụ như nghiên cứu về kháng thuốc, vì mục tiêu để ra là xác định qui mô kháng thuốc, nên phần kết quả phải bám sát theo mục tiêu này. Mở đầu phần kết quả là đi thẳng vào kết quả quan trọng:

"Antimicrobial resistance was commonly found in ceftriaxone (88%), ceftazidime (80%), ciprofloxacin (77%), cefepime (75%), levofloxacin (72%). Overall, the rate of antimicrobial resistance to any drug was 93% (n = 153/164), with the majority (87%) being resistant to at least 2 drugs. The three commonly isolated microorganisms were Acinetobacter (n = 75), Klebsiella (n = 39), and Pseudomonas aeruginosa (n = 29). Acinetobacter baumannii were virtually resistant to ceftazidime, ceftriaxone, piperacilin, imipenem, meropenem, ertapenem, ciprofloxacin and levofloxacin. High rates (>70%) of ceftriaxone and ceftazidime-resistant Klebsiella were also observed."

Chú ý trong phần kết quả trên, vì quá nhiều kết quả, nên chỉ có thể tập trung vào những điểm chính mà thôi. Những con số quan trọng để trong ngoặc kép. Có khi kết quả phân tích rất nhiều, nên chỉ có thể nhắc đến một hay 2 điểm mà thôi:

“We identified 56 loci (32 novel) associated with BMD at genome-wide significant level ($P < 5 \times 10^{-8}$). Several of these factors cluster within the RANK-RANKL-OPG, mesenchymal-stem-cell differentiation, endochondral ossification and the Wnt signaling pathways. However, we also discovered loci containing genes not known to play a role in bone biology.”

Chú ý trong đoạn văn trên, tác giả không cần đề cập đến các chi tiết về bệnh nhân (như tuổi, giới tính) mà đi thẳng vào kết quả phân tích.

Kết luận. Một hoặc 2 câu văn kết luận và ý nghĩa của kết quả nghiên cứu. Có thể nói phần lớn độc giả chú tâm vào câu văn này trước khi họ đọc các phần khác, cho nên tác giả cần phải chọn câu chữ sao cho “thuyết phục” và thu hút được sự chú ý của độc giả trong 2 câu văn quan trọng này.

Ví dụ như công trình về phát triển mô hình tiên lượng tử vong ở bệnh nhân cấp cứu, tác giả chỉ cần một câu đơn giản nhưng “punchy” là đủ:

“We conclude that the risk of mortality among ED patients could be accurately predicted by using common clinical signs and biochemical tests.”

Ví dụ cụ thể về cách viết bản tóm tắt. Sau đây là một bản tóm lược tiêu biểu có tiêu đề. Bài báo này trình bày một công trình nghiên cứu về mối liên hệ giữa các thành phần cơ thể (mỡ, cơ, xương) ở một nhóm phụ nữ Việt Nam sau mãn kinh (LT Ho-Pham, et al. BMC Musculoskeletal Disorders 2010, 11:59). Bản tóm lược có 4 tiêu đề: dẫn nhập, phương pháp, kết quả và kết luận. Phần dẫn nhập chỉ tóm gọn trong 2 câu văn, với câu đầu nêu vấn đề vẫn còn trong vòng tranh cãi, và câu 2 phát biểu về giả thuyết và mục đích của nghiên cứu. Phần

phương pháp mô tả số phụ nữ tham gia, độ tuổi, nơi nghiên cứu, phương pháp đo lường, và phương pháp phân tích. Phần kết quả đi thẳng vào kết quả chính với những con số cụ thể. Đương nhiên, những con số này sẽ được lặp lại chi tiết hơn trong bài báo. Phần kết luận chỉ một câu văn có tính cách trả lời câu hỏi nghiên cứu.

"Background. The relative contribution of lean and fat to the determination of bone mineral density (BMD) in postmenopausal women is a contentious issue. The present study was undertaken to test the hypothesis that lean mass is a better determinant of BMD than fat mass.

Methods. This cross-sectional study involved 210 postmenopausal women of Vietnamese background, aged between 50 and 85 years, who were randomly sampled from various districts in Ho Chi Minh City (Vietnam). Whole body scans, femoral neck, and lumbar spine BMD were measured by DXA (QDR 4500, Hologic Inc., Waltham, MA). Lean mass (LM) and fat mass (FM) were derived from the whole body scan. Furthermore, lean mass index (LMi) and fat mass index (FMi) were calculated as ratio of LM or FM to body height in metre squared (m^2).

Results. In multiple linear regression analysis, both LM and FM were independent and significant predictors of BMD at the spine and femoral neck. Age, lean mass and fat mass collectively explained 33% variance of lumbar spine and 38% variance of femoral neck BMD. Replacing LM and FM by LMi and LMi did not alter the result. In both analyses, the influence of LM or LMi was greater than FM and FMi. Simulation analysis suggested that a study with 1000 individuals has a 78% chance of finding the significant effects of both LM and FM, and a 22% chance of finding LM alone significant, and zero chance of finding the effect of fat mass alone.

Conclusions. These data suggest that both lean mass and fat mass are important determinants of BMD. For a given body size -- measured either by lean mass or height -- women with greater fat mass have greater BMD."

Bản tóm lược dưới đây là một abstract tiêu biểu không có tiêu đề (LT Ho-Pham, et al. Similarity in percent body fat between white and Vietnamese women: implication for a universal definition of obesity. Obesity 2010; 18:1242-6). Toàn bộ bản tóm lược chỉ là một đoạn văn. Nhưng nếu chú ý kỹ sẽ thấy những thông tin được trình bày trong abstract tuân thủ theo cấu trúc IMRAD. Phần dẫn nhập gồm 2 câu văn: câu đầu tiên nêu vấn đề nghiên cứu; câu thứ hai phát biểu mục đích nghiên cứu. Các câu kế tiếp mô tả phương pháp nghiên cứu, kết quả, và kết luận (xem phần viết đậm).

[Viết về bối cảnh nghiên cứu] *It has been widely assumed that for a given BMI, Asians have higher percent body fat (PBF) than whites, and that the BMI threshold for defining obesity in Asians should be lower than the threshold for whites.*

[Mục tiêu nghiên cứu] *This study sought to test this assumption by comparing the PBF between US white and Vietnamese women.*

[Phương pháp] *The study was designed as a comparative cross-sectional investigation. In the first study, 210 Vietnamese women ages between 50 and 85 were randomly selected from various districts in Ho Chi Minh City (Vietnam). In the second study, 419 women of the same age range were randomly selected from the Rancho Bernardo Study (San Diego, CA).*

[Kết quả] *In both studies, lean mass (LM) and fat mass (FM) were measured by dual-energy X-ray absorptiometry (DXA) (QDR 4500; Hologic). PBF was derived as FM over body weight. Compared with Vietnamese women, white women had much*

more FM (24.8 +/- 8.1 kg vs. 18.8 +/- 4.9 kg; $P = 0.30$, 19% of US white women and 5% of Vietnamese women were classified as obese. Approximately 54% of US white women and 53% of Vietnamese women had their PBF >35% ($P = 0.80$).

[Kết luận] Although white women had greater BMI, body weight, and FM than Vietnamese women, their PBF was virtually identical. Further research is required to derive a more appropriate BMI threshold for defining obesity for Asian women.

Chú ý trong bản tóm tắt, tác giả không cần viết nhiều chi tiết về phương pháp (ngoại trừ đó là bài báo về phương pháp). Không nên viết tắt những thuật ngữ mà người đọc ngoài ngành không am hiểu. Thông thường bản tóm lược được viết sau khi đã hoàn tất bài báo. Kinh nghiệm cá nhân cho thấy có khi tốn đến cả ngày chỉ để viết một bản tóm lược với 200 chữ. Abstract có thể xem như là một bài thơ, tức là tác giả phải chọn từ ngữ rất cẩn thận để phản ánh cô đọng những điều mình muốn chuyển tải đến cộng đồng khoa học.

Báo cáo khoa học: khổ hạnh

Mỗi bài báo khoa học là một công trình khổ hạnh. “Khổ hạnh” ở đây phải được hiểu theo nghĩa vừa đau khổ, vừa hạnh phúc. Đau khổ trong quá trình chuẩn bị và viết thành một bài báo, và hạnh phúc khi nhìn thấy bài báo được công bố trên một tập san có nhiều đồng nghiệp đọc và chia sẻ. Để đạt kết quả sau cùng này, tác giả phải phấn đấu làm sao để giữ sự cân bằng giữa tính trong sáng và [nhưng] nội dung phải đầy đủ. Bài báo phải làm sao hấp dẫn người đọc và để người đọc “nhập cuộc”. Bài báo phải được viết bằng một văn phong cực kì súc tích, nhưng phải đầy đủ. Đó là những yêu cầu rất khó mà không phải tác giả nào cũng đạt được.

Bài báo khoa học, nói cho cùng, là một bản báo cáo những gì nhà nghiên cứu đã làm và tại sao làm, kết quả ra sao và có ý nghĩa gì, và những gì cần làm tiếp. Người đọc bản báo cáo là những đồng nghiệp, kể cả chuyên gia và nghiên cứu sinh, trên toàn thế giới. Chính những người đọc này là những người “phản biện” thích hợp nhất và có lẽ công bằng nhất. Vì thế, công bố những gì đã nghiên cứu là một hình thức “thử lửa” tốt nhất cho nghiên cứu sinh và nhà khoa học, bởi vì qua đó mà đồng nghiệp trong và ngoài nước có thể thẩm định chất lượng của công trình nghiên cứu và luận án của nghiên cứu sinh.

Nếu công trình nghiên cứu được đồng nghiệp đánh giá cao, thì bài báo đó còn thể hiện một sự đóng góp vào kho tàng của tri thức nhân loại. Hiện nay, sự có mặt của khoa học Việt Nam trên trường quốc tế còn rất hạn chế, nếu không muốn nói là con số gần số không. Do đó, việc công bố bài báo khoa học còn là một hình thức thực tế nhất để đóng góp vào việc nâng cao năng suất khoa học của Việt Nam.

Báo cáo khoa học trên các tập san quốc tế còn là một cơ hội để nghiên cứu sinh làm quen với quá trình hội nhập quốc tế. Nhiều người nghĩ rằng sau khi xong luận án tiến sĩ, có mảnh bằng trong tay là xong. Thật ra, hoàn tất luận án và được mang danh xưng “tiến sĩ” chỉ mới là bước đầu trong sự nghiệp khoa bảng, bởi vì nghiên cứu sinh còn phải tiêu ra một thời gian làm nghiên cứu hậu tiến sĩ ở một trường đại học khác. Để có một chân nghiên cứu hậu tiến sĩ là một cuộc cạnh tranh gay gắt. Hàng năm, có hàng chục ngàn tiến sĩ trên khắp thế giới cạnh tranh nhau để xin được vài trăm vị trí nghiên cứu hậu tiến sĩ. Những người thành công thường là những nghiên cứu sinh đã có vài bài báo khoa học trong khi học tiến sĩ. Vì thế, công bố bài báo khoa học trên các tập san

quốc tế còn là một hình thức đảm bảo và nâng cao khả năng cạnh tranh của nghiên cứu sinh trên trường quốc tế.

Công bố kết quả nghiên cứu trên các diễn đàn khoa học quốc tế không phải là một thách thức lớn, nếu công trình nghiên cứu tiến sĩ có chất lượng xứng đáng với học vị. Nói cách khác, nếu đề án nghiên cứu được thụ ý cẩn thận, với những ý tưởng mới và có ích và phương pháp thực hiện đạt chuẩn khoa học, thì việc công bố kết quả không phải là một vấn đề gì to tát lắm. Nếu đề án nghiên cứu không có giá trị khoa học hay không xứng đáng với một luận án tiến sĩ thì nghiên cứu sinh cũng chẳng học được gì sau ba hay bốn năm theo học.

TIẾNG ANH TRONG BÀI BÁO KHOA HỌC

Tôi phục vụ trong ban biên tập của một số tạp san y khoa trong vai trò thành viên ban biên tập và phó biên tập (Associate Editor). Một trong những đặc lợi của công việc biên tập là học. Học từ ý tưởng, phương pháp, đến tiếng Anh. Học từ những người có kinh nghiệm hơn mình. Học từ những sai sót của người khác và của chính mình. Trong chương này tôi sẽ trình bày một nguyên tắc viết tiếng Anh mà tôi tạm đặt là I D E A, và sau đó tôi sẽ liệt kê những sai sót phổ biến về tiếng Anh trong bài báo khoa học.

1. Văn phong khoa học: nguyên tắc IDEA

Viết văn khoa học vừa là một nghệ thuật nhưng cũng là một khoa học. Tính nghệ thuật trong văn chương khoa học liên quan đến việc chọn chữ, “điệu văn” (tone), và duyên dáng. Tính khoa học trong văn chương khoa học liên quan đến cấu trúc nội dung, từ đoạn văn đến câu văn, phải sao cho trong sáng. Để đạt được yêu cầu trong sáng, tôi đề nghị tuân theo 4 nguyên tắc có thể gọi tắt là IDEA: Idea (ý tưởng) + Data (dữ liệu) + English (tiếng Anh) + Art (nghệ thuật). Bốn nguyên tắc này có thể tóm lược như sau:

- (a) Idea: có ý tưởng, biết mình viết cái gì, đạt mục tiêu gì;
- (b) Data: có sẵn dữ liệu, tức là phải đọc các nghiên cứu trước, đọc rất nhiều;

(c) English: có kỹ năng tiếng Anh, tức là cách chọn chữ, cấu trúc câu văn; và

(d) Art: phải có một chút nghệ thuật, tức phải tập khiếu thẩm mĩ trong văn chương.

Bốn nguyên tắc trên đòi hỏi tác giả phải đọc. Đọc trước hết là để thu thập dữ liệu, nhưng sau đó là học cách người trước mình viết và diễn giải ra sao. Không có gì hiệu quả hơn là học từ thực tế. Có thể nói không ngoa rằng sự nghiệp học tập của một sinh viên hay nghiên cứu sinh là... đọc. Do đó, không phải ngẫu nhiên mà ngày xưa người ta gọi đi học đại học là “read” (chẳng hạn như câu “*She read biology at the University of Cambridge*”).

1.1 Ý tưởng: lên “kịch bản”

Bước quan trọng nhất trong bài báo khoa học là dàn bài, vì bước này quyết định cái khung của bài báo. Dàn bài ở đây hiểu theo nghĩa là lên “kịch bản” cho bài báo hay một phần của bài báo. Điều rất quan trọng là dàn bài phải nhắm đến mục tiêu của nghiên cứu. Chẳng hạn như mục tiêu nghiên cứu là xây dựng hồ sơ gen (genetic profiling) để tiên lượng loãng xương, tác giả phải nghĩ cách lên kịch bản trong phần dẫn nhập sao cho cuối cùng độc giả thấy mục tiêu xây dựng hồ sơ gen là cần thiết. Để chỉ ra mục tiêu là cần thiết, tác giả phải nói cho độc giả biết tầm quan trọng của vấn đề, những nghiên cứu trong quá khứ về vấn đề ra sao, khoảng trống tri thức là gì, và sau cùng là mục tiêu. Có lẽ tác giả nghĩ đến một kịch bản như sau:

1. Loãng xương là bệnh quan trọng vì nhiều người mắc;
2. Nhưng vì bệnh lí diễn ra một cách âm thầm, nên cần phải nhận dạng người có nguy cơ cao (trước khi quá muộn);

3. Nguy cơ của bệnh thì chịu nhiều tác động, từ yếu tố lâm sàng đến di truyền;
4. Đã có nhiều nghiên cứu về di truyền và đã phát hiện nhiều gen, nhưng ảnh hưởng của mỗi gen còn thấp. Có cách nào hay hơn để sử dụng thông tin gen cho tiên lượng bệnh? (Ở đây tác giả phải đang nghĩ đến khoảng trống tri thức);
5. Nhưng có vài cách sử dụng gen, và một trong những cách đó là xây dựng hồ sơ gen, tổng hợp thông tin từ nhiều gen và tạo ra “chữ kí gen” cho mỗi cá nhân. Giả thuyết là chữ kí này sẽ giúp tiên lượng bệnh tốt hơn. Sau đó là phát biểu mục tiêu của nghiên cứu.

Chú ý là các ý tưởng trên được dàn xếp theo logic đường thẳng, từ $A \rightarrow B$, $B \rightarrow C$, v.v. Như vậy, sau khi lên kịch bản như thế (có lẽ trên bàn cà phê!) phần dẫn nhập có lẽ sẽ có 5 đoạn văn, vì mỗi đoạn văn chỉ nói lên 1 ý tưởng.

1.2 Data: tìm chất liệu và viết

Bước kế tiếp là tìm chất liệu (tức là data - dữ liệu) cho mỗi ý hay mỗi đoạn văn. Chẳng hạn như trong đoạn văn đầu về tầm quan trọng của bệnh, tác giả cần phải nghĩ mình viết gì để nói lên cái ý đó. Người có kinh nghiệm có thể sẽ viết về qui mô của bệnh trong cộng đồng; về ảnh hưởng của bệnh đến tử vong hay các biến chứng; về chi phí điều trị và do đó gánh nặng kinh tế cho cá nhân; về chất lượng cuộc sống. Sau khi đã biết mình viết gì, tác giả phải tìm trong y văn các dữ liệu cần thiết để nói lên các điểm vừa kể. Giả dụ như sau khi đã có dữ liệu, tác giả phải nghĩ đến cách viết. Cách viết một đoạn văn như chúng ta đã bàn là gồm 3 phần: câu văn chủ đề, các câu văn yểm trợ, và câu văn nhấn mạnh. Câu văn chủ đề có lẽ là:

“Osteoporosis is recognized as a public health burden, because the prevalence of the disease is high in the general population, and it is associated with an increased risk of death.”

Câu văn này mở đầu cho các câu văn sau. Vì câu mở đầu nói đến *prevalence* và tử vong, nên các câu sau phải “yểm trợ” cho hai khía cạnh đó. Có lẽ các câu tiếp theo là:

“Among Caucasian individuals aged 50 years and older, approximately one third of women and one tenth of men have osteoporosis (X). Emerging evidence suggests that the prevalence of osteoporosis appears to increase with time (X), particularly in Asian populations (X). More importantly, osteoporosis is associated with reduced life expectancy (X), and the reduction is more pronounced in men than in women (X). For instance, a large scale study in the United States found that the risk of mortality among men with osteoporosis was increased by 3-fold compared with the background population of the same age; however, the increased risk in women was 1.8-fold.”

Chú ý những (X) là tài liệu tham khảo. Nếu có thêm dữ liệu thì viết nữa để cung cấp cho độc giả những thông tin liên quan. Nếu không thì có thể kết thúc đoạn văn với câu kết. Thật ra, đây là câu khó viết, vì viết không khéo sẽ lặp lại ý chính của câu văn chủ đề. Suy nghĩ xem phải nhấn mạnh điều gì... Có lẽ nên nhấn mạnh đến tương lai khi tình trạng lão hoá gia tăng thì qui mô và tác động của bệnh cũng tăng. Quyết định như thế, tác giả có thể viết:

“Given that the population aging has taken place in many parts of the world, it is expected that the burden and impact of osteoporosis is likely to increase in the future.”

Tổng kết lại, tác giả có một đoạn văn mở đầu:

“Osteoporosis is recognized as a public health burden, because the prevalence of the disease is high in the general population, and it is associated with an increased risk of death. Among Caucasian individuals aged 50 years and older, approximately one third of women and one tenth of men have osteoporosis (X). Emerging evidence suggests that the prevalence of osteoporosis appears to increase with time (X), particularly in Asian populations (X). More importantly, osteoporosis is associated with reduced life expectancy (X), and the reduction is more pronounced in men than in women (X). For instance, a large scale study in the United States found that the risk of mortality among men with osteoporosis was increased by 3-fold compared with the background population of the same age; however, the increased risk in women was 1.8-fold. Given that the population aging has taken place in many parts of the world, it is expected that the burden and impact of osteoporosis is likely to increase in the future.”

1.3 English: tiếng Anh

Tiếng Anh được xem là một ngôn ngữ của nghiên cứu khoa học. Không chỉ đa số thuật ngữ mới trong khoa học được “sáng chế” qua tiếng Anh, mà đa số các tập san khoa học trên thế giới đều dùng tiếng Anh như là một phương tiện chuyển tải thông tin. Thật vậy, hơn 90% các tập san khoa học quốc tế dùng tiếng Anh. Ngay cả ở những nước mà tiếng Anh không phải là bản ngữ, người ta cũng dùng Anh để công bố kết quả nghiên cứu. Do đó, tất cả các tác giả khoa học cần phải học cách viết tiếng Anh đúng chuẩn mực khoa học.

Nhưng viết văn khoa học bằng tiếng Anh không phải là điều dễ dàng. Trong thực tế, ngay cả những nghiên cứu sinh và nhà khoa học dùng tiếng Anh như là tiếng mẹ đẻ ở các nước Âu Mỹ vẫn gặp trở ngại trong việc soạn bài báo khoa học bằng tiếng

Anh. Khó khăn về tiếng Anh đối với các nhà khoa học ở các nước đang phát triển càng nặng nề hơn. Trong thực tế, khoảng phân nửa những bài báo khoa học từ các nước đang phát triển bị trả lại do tiếng Anh không đạt; trong số này có khoảng 30% là bị từ chối vì lí do tiếng Anh có quá nhiều sai sót. Do đó, ở các nước Âu Mỹ, nghiên cứu sinh thường được gửi đi học những workshop chuyên hướng dẫn về cách soạn bài báo khoa học.

Viết văn khoa học bằng tiếng Anh, cũng như viết luận văn, phải tuân thủ theo một số nguyên tắc. Những nguyên tắc này sẽ được bàn chi tiết và cụ thể trong phần kế tiếp; ở đây, tôi chỉ muốn bàn qua 4 điểm liên quan đến việc dùng tiếng Anh trong bài báo khoa học liên quan đến cấu trúc một đoạn văn, văn phạm, cách dùng chữ chính xác, và văn phong tốt.

Một trong những sai sót rất phổ biến trong các bài báo khoa học là **cấu trúc đoạn văn**. Nguyên tắc căn bản nhất là *mỗi đoạn văn chỉ nói lên một ý tưởng*. Ý tưởng đó phải được “tuyên bố” trong câu văn đầu tiên. Các câu văn tiếp theo có chức năng cung cấp chứng cứ hay dữ liệu để làm cơ sở cho câu văn tuyên ngôn. Câu văn cuối cùng có chức năng nhấn mạnh cái ý tưởng nêu ra lúc ban đầu. Ấy thế mà rất nhiều nhà khoa học quên công thức cơ bản này.

Riêng ở Việt Nam, nhiều nhà khoa học có cách viết mà tôi gọi là “cách viết powerpoint”. Cách viết powerpoint là cách viết ngắn và hay dùng gạch đầu dòng. Ngắn đến độ chỉ có một hay 2 câu văn, mà không giải thích thêm. Nhiều tác giả dùng dấu gạch đầu dòng như là một danh sách. Tiêu biểu cho cách viết này là như sau:

“Phương pháp thu thập dữ liệu:

- Mời bệnh nhân tham gia nghiên cứu.
- Tiến hành phỏng vấn theo bảng thu thập dữ liệu.
- Chẩn đoán giai đoạn bệnh.
- Sau đó phân tích gen.

Nhưng đây là một cách viết không thích hợp cho văn bản khoa học như luận án hay bài báo khoa học. Đó không phải là kiểu viết điện tín (telegraphic), nhưng gần với cách viết đó. Người ta biện minh rằng đó là cách viết ngắn gọn, mà khoa học thì cần phải viết ngắn gọn. Văn phong khoa học là phải ngắn gọn, nhưng không ngắn đến nỗi đoạn văn (paragraph) chỉ có 1 câu văn. Cần phải tuyệt đối tránh cách viết powerpoint.

Thỉnh thoảng được đọc đề cương nghiên cứu và bài báo khoa học của các đồng nghiệp bên nhà, tôi phát hiện ra một thói quen mà có lẽ các bạn ấy không chú ý: dùng dấu "...", tiếng Anh gọi là "ellipsis":

"Các nguyên nhân khác: do nhiễm trùng, methyldopa, phenytoin..."

hay

"Phần lớn trường hợp SGM có các triệu chứng cơ năng mơ hồ, mệt mỏi, biếng ăn..."

Các câu văn như trên có chức năng như là liệt kê một danh sách các sự vật. Nhưng cách dùng "..." cho người đọc cảm tưởng như tác giả đã hết ý hay lười biếng. Hết ý nghĩ ra một sự vật khác tiếp theo. Hoặc cũng có thể tác giả lười biếng không tìm hiểu thêm còn có sự vật nào, hay làm biếng viết ra hết danh sách. Dù bất cứ lí do gì thì cách dùng "..." trong văn phong khoa học là đại kỵ. Cần phải tuyệt đối tránh dấu "..." khi viết bài báo hay luận án.

Về **văn phạm**, tiếng Anh là một ngôn ngữ phức tạp, nhưng trong khoa học thì không quá phức tạp. Khi viết bài báo khoa học, tác giả cần phải chú ý đến cách dùng thì (tense), sự nhất quán giữa chủ đề và động từ, và cú pháp tốt. Những mối liên hệ, hiện tượng được xem là phổ quát (như “*smoking causes lung cancer*”) thì cần dùng thì hiện tại. Khi đề cập đến dữ liệu báo cáo trong bài báo, thì hiện tại cũng được dùng (ví dụ như “*Figure 1 shows that*”). Tuy nhiên, khi báo cáo kết quả nghiên cứu thì dùng thì quá khứ (ví dụ: *We found that a family history of stroke was associated with an increased risk of stroke*”).

Một trong những sai sót hay được phát hiện trong các bài báo khoa học là sự thiếu nhất quán giữa chủ từ và động từ. Chẳng hạn như câu “*The effect of genetic variants on bone density were observed in men*” là sai, vì chủ từ ở đây là “the effect” chứ không phải “genetic variants”, và do đó động từ “were” là không đúng. Mặc dù sai sót này rất sơ đẳng, nhưng ngạc nhiên thay rất nhiều bản thảo bài báo khoa học phạm phải sai sót này.

Văn phong khoa học phải **chính xác**. Tính chính xác ở đây phải hiểu theo nghĩa là câu văn hay chữ không thể hiểu theo nghĩa khác mà tác giả muốn chuyển tải. Thay vì viết “*The majority of participants*” tác giả nên dùng chữ định lượng như “*Ninety per-cent of participants*.”

Về **văn phong và chọn từ**, tác giả nên chọn cách viết tích cực và dùng từ mang tính định lượng. Nguyên tắc gần như bất di bất dịch là dùng chữ đơn giản, tránh dùng chữ có nguồn gốc Latin và Hi Lạp. Ngoài ra, cũng tránh dùng những chữ mang tính màu mè (như *cutting edge, food for thought*), những chữ lấp lửng (*basically, in a sense, in fact*), và nhất là tuyệt đối tránh những ví von vô duyên.

1.4 Arts: nghệ thuật

Cứ mỗi lần viết xong một đoạn văn, tác giả cần phải đọc lại và đọc lại và đọc lại. Sau khi viết xong một đoạn văn, tác giả cần phải đọc lại nhiều lần và chỉnh sửa. Có 5 câu hỏi tác giả cần phải tự đặt ra là:

- Các câu văn đã nói lên cái ý chính chưa?
- Có câu hay chữ nào thừa?
- Có chữ nào khó hiểu không?
- Có câu nào cấu trúc sai văn phạm không?
- Cái tone của các câu văn có ổn chưa?

Chẳng hạn như câu *“Osteoporosis is recognized as a public health burden, because the prevalence of the disease is high in the general population, and it is associated with an increased risk of death”* có dài quá chẳng? Có nên tách nó thành hai câu riêng? Chữ *“recognized”* (ghi nhận) có ổn chưa? Ai ghi nhận? Nếu chỉ viết chung chung như thế thì có thể bị xem là... lười biếng suy nghĩ. Có lẽ nên dùng chữ khác, tích cực hơn. Chúng ta thử làm tác giả tách ra hai câu riêng:

“Osteoporosis is increasingly becoming an important public health problem. The importance of the disease is characterized by its high prevalence and serious clinical consequences, including mortality.”

Có lẽ hai câu này rõ ràng hơn và dễ hiểu hơn. Chữ *“clinical consequence”* mở cửa cho tác giả viết thêm về tác động lâm sàng của bệnh trong các câu văn sau. Vậy thì đoạn văn cần phải viết lại như sau:

“Osteoporosis is increasingly becoming an important public health problem. The importance of the disease is characterized by its high prevalence and serious clinical consequences, including

mortality. Among Caucasian individuals aged 50 years and older, approximately one third of women and one tenth of men have osteoporosis (X). Emerging evidence suggests that the prevalence of osteoporosis appears to increase with time (X), particularly in Asian populations (X). Moreover, osteoporotic patients are associated with comorbidities such as osteoarthritis, sarcopenia, and diabetes. Patients with osteoporosis have a higher risk of sustaining a fracture. More importantly, osteoporosis is associated with reduced life expectancy (X), and the reduction is more pronounced in men than in women (X). For instance, a large scale study in the United States found that the risk of mortality among men with osteoporosis was increased by 3-fold compared with the background population of the same age; however, the increased risk in women was 1.8-fold. Given that the population aging has taken place in many parts of the world, it is expected that the burden and impact of osteoporosis is likely to increase in the future.”

Có thể đọc lại lần nữa để tìm xem 5 yếu tố trên đã ổn chưa. Chẳng hạn như chữ “For instance” có lẽ thừa, và nên bỏ. Thậm chí những chữ “a large scale study in the United States found that” cũng không cần thiết. Hay như câu “Given that the population aging has taken place in many parts of the world” có thể viết lại cho gọn hơn như “Because the population aging has taken place worldwide” (gọn hơn, vì bỏ được 6 chữ “in many parts of the world”). Câu “the burden and impact of osteoporosis is” sai văn phạm. Đoạn văn bây giờ là:

“Osteoporosis is increasingly becoming an important public health problem. The importance of the disease is characterized by its high prevalence and serious clinical consequences, including mortality. Among Caucasian individuals aged 50 years and older, approximately one third of women and one tenth of men have osteoporosis (X). Emerging evidence suggests that the prevalence of osteoporosis appears to increase with time (X), particularly

in Asian populations (X). Moreover, osteoporotic patients are associated with comorbidities such as osteoarthritis, sarcopenia, and diabetes. Patients with osteoporosis have a higher risk of sustaining a fracture. More importantly, osteoporosis is associated with reduced life expectancy (X), and the reduction is more pronounced in men than in women (X). The risk of mortality among men with osteoporosis was increased by 3-fold compared with the background population of the same age; however, the increased risk in women was 1.8-fold (X). Because the population aging has taken place worldwide, it is expected that the burden and impact of osteoporosis are likely to increase in the future."

Nhưng câu cuối có cái tone chưa được "đẹp". Có thể viết lại như sau:

"Because the population aging has taken place worldwide, it is expected that osteoporosis is likely to impose a greater burden and impact on the society in the near future."

Có lẽ đoạn văn như thế là tạm ổn. Cứ để đó, tác giả nên viết tiếp đoạn thứ hai, và cái chu kì chỉnh sửa lại bắt đầu cho đến khi đoạn văn hoàn tất. Sau khi viết xong đoạn văn số 6, cũng cần phải đọc lại toàn bộ phần dẫn nhập để xem các đoạn văn đã ăn khớp với nhau và nói lên một "câu chuyện".

Như để cập ở phần đầu, viết văn là một nghệ thuật, nhưng viết văn khoa học lại là một khoa học. Bởi vậy một nhà văn lừng danh, William Zinsser, từng nói rằng viết là suy nghĩ trên trang giấy - *writing is thinking on paper* (còn ngày nay thì viết có nghĩa là suy nghĩ trên... màn hình). Suy nghĩ mù mờ, hời hợt, thì văn cũng như thế; suy nghĩ rõ ràng và logic sẽ phản ánh qua cấu trúc một bài báo. Tôi hi vọng rằng nguyên tắc IDEA trình bày trong phần này sẽ giúp cho các bạn soạn một bài báo khoa học hoàn chỉnh.

2. Tiếng Anh trong khoa học

Trong 4 nguyên tắc IDEA về viết văn khoa học, phần tiếng Anh đóng vai trò rất quan trọng. Tiếng Anh cũng là vấn đề đối với giới khoa học Việt Nam. Sau khoa học, tiếng Anh là một rào cản lớn đối với đồng nghiệp bên nhà, đặc biệt là những người làm trong lĩnh vực đòi hỏi khả năng ngôn ngữ như khoa học xã hội và các ngành khoa học thực nghiệm. Ngay cả những người đã từng theo học ở nước ngoài hơn 10 năm cũng chưa chắc viết được một bài báo khoa học hoàn chỉnh. Cái khó là làm quen với văn phong khoa học. Trong phần này tôi sẽ trình bày vài kinh nghiệm cá nhân, và hi vọng sẽ giúp ít nhiều cho các bạn đang học viết bài báo khoa học.

Good writing = clarity + convincing

Một câu hỏi mà nhiều người, kể cả những người đã thành danh trong khoa học, thắc mắc là: thế nào là một cách viết hay (good writing). Câu trả lời đơn giản nhất chỉ tóm lược trong hai chữ C: clarity và convincing (trọng sáng và thuyết phục). Để đạt được hai yêu cầu đó, một bài báo cần phải chú ý đến 4 yếu tố sau đây:

- Cấu trúc một đoạn văn;
- Văn phạm;
- Tính chính xác; và
- Văn phong.

2.1 Cấu trúc đoạn văn

Nên nhớ rằng mỗi đoạn văn chỉ nói lên 1 ý tưởng. Nhất định không có hơn một ý tưởng trong một đoạn văn. Một đoạn văn thường chỉ có 3 đến 6 câu văn. Câu văn khởi đầu là câu văn tuyên bố ý tưởng. Các câu văn tiếp theo là cung cấp

chúng cứ để yểm trợ cho ý tưởng. Câu văn sau cùng là một nhấn mạnh đến ý tưởng phát biểu lúc ban đầu. Ví dụ, đoạn văn sau đây viết về tầm quan trọng của thuốc thiazolidinedione (Steven Nissen, NEJM):

[Vào đầu, nói về ý tưởng chính] *Thiazolidinedione drugs are widely used to lower blood glucose levels in patients with type 2 diabetes mellitus. [Cung cấp dữ liệu]* *In the United States, three such agents have been introduced: troglitazone, which was removed from the market because of hepatotoxicity, and two currently available agents, rosiglitazone (Avandia, GlaxoSmithKline) and pioglitazone (Actos, Takeda). The thiazolidinediones are agonists for peroxisome-proliferator-activated receptor γ (PPAR- γ). PPAR- γ receptors are ligand-activated nuclear transcription factors that modulate gene expression, lowering blood glucose primarily by increasing insulin sensitivity in peripheral tissues. [Nhấn mạnh]* *Rosiglitazone was introduced in 1999 and is widely used as monotherapy or in fixed-dose combinations with either metformin (Avandamet, GlaxoSmithKline) or glimepiride (Avandaryl, GlaxoSmithKline)."*

Một cách viết khác cũng theo mô hình trên:

"[Câu văn mở đầu mang tính tuyên bố chủ đề] *Most medical research findings are false. [Những câu sau là minh hoạ & chứng cứ]* *The falsity of findings is caused by many factors, including among others, biases, flaws in experimental design, abuse of tests of hypothesis, misuse of statistics, and incorrect interpretation of statistical results. In this article, I will make use of real world examples to illustrate some common errors in conceptual understanding, technical modeling, and interpretation. I will then introduce my own principle of data analysis which can be summarized in five letters: SMART (for simplicity, modeling, adequacy and assumption, reproducibility, and transparency). [Câu văn nhấn mạnh]* *The outcome of this lecture is that*

participants will acquire useful statistical knowledge that helps reduce the prevalence of false findings in medical research."

Sau khi viết một đoạn văn, bạn cần phải ngừng một chút và suy nghĩ. Ngừng để đọc lại đoạn văn cho đầy đủ, và tự hỏi rằng "tôi đã nói hết những gì tôi muốn nói chưa", và sau cùng câu hỏi quan trọng là "tôi đã nói cái ý chính chưa?"

2.2 Chú ý đến văn phạm

Dĩ nhiên, phần lớn những người viết bài báo khoa học đã học qua tiếng Anh và văn phạm tiếng Anh. Tuy nhiên, trong thực tế, vẫn còn rất nhiều bài báo khoa học mà tác giả viết sai văn phạm. Những sai sót này chủ yếu ở 3 khía cạnh:

- Dùng sai thì;
- Thiếu nhất quán giữa chủ từ và động từ; và
- Sai cú pháp.

(a) Về cách dùng thì, cần phải nhớ những tình huống sau đây:

Nếu câu văn đề cập đến một loại kiến thức rất phổ biến, kiến thức phổ thông, và được chấp nhận như là "sự thật", dùng thì hiện tại. Ví dụ: "*Smoking is associated with lung cancer*" có thể dùng thì hiện tại vì đó là một mối liên hệ phổ quát.

Những kiến thức hay kết quả chưa được nhiều người chấp nhận như là sự thật hay chưa được xem là qui luật, dùng thì quá khứ. Ví dụ: *Some recent clinical trials suggested an increased breast cancer risk with long-term use of hormone replacement therapy*

Phần kết quả của bài báo, lúc này cũng dùng thì quá khứ: *The index of heritability was 85%; We found that polymorphisms of the FTO gene.*

Nhưng nếu bạn đề cập đến bảng biểu trong phần kết quả, thì nên dùng thì hiện tại. Ví dụ: *Figure 1 shows the age-related change in fasting plasma glucose.*

Khi đề cập hay trích dẫn đến một bài báo đã công bố, dùng thì quá khứ. Ví dụ: *Nguyen et al showed that a genetic profiling was an independent predictor of fracture risk.* Chú ý rằng một số người có xu hướng dùng thì hiện tại để đề cập đến bài báo đã công bố, nhưng tôi nghĩ đó là cách sử dụng thì không hợp lí.

(b) Vấn đề nhất quán giữa chủ từ và động từ không phải là mới, nhưng ngạc nhiên thay, rất nhiều bản thảo bài báo tôi nhận được thường phạm phải sai sót. Những sai sót này chủ yếu là quên ghi nhận số nhiều và số ít của một số danh từ gốc Latin. Nên nhớ rằng những danh từ như *data*, *criteria*, *addenda*, v.v. là số nhiều, chứ không phải số ít. Do đó, nên viết “*the following criteria were*” (thay vì “*the following criteria was*”). Ngoài ra, sau đây là vài sai sót các bạn cần phải chú ý:

Xác định chủ đề của câu văn là số nhiều hay số ít và chia động từ thích hợp. Chẳng hạn như câu “*The effects of alcohol on enzyme induction was studied in vitro*” là sai, bởi vì chủ từ là “*the effects*”, chứ không phải *enzyme induction*, và do đó câu văn nên viết lại là “*The effects of alcohol on enzyme induction were studied in vitro*”. Tương tự, câu “*The generation of excessively large sets of data were responsible for*” cũng sai.

Khi câu văn có hai chủ đề thì dĩ nhiên là phải dùng chia động từ số nhiều. Ví dụ: *Both the controls and cases were*).

Khi hai chủ đề được nối kết bằng *either/neither*, chia động từ số ít. Ví dụ: *Neither the doctor nor the patient was informed of the drug used.*

Nếu *none* có nghĩa là “không có ai” thì nên dùng động từ số ít. Ví dụ: “*None of the applicants is fully qualified.*”

(c) Cú pháp

Những vấn đề cần chú ý trong phần này là những mệnh đề mang tính bổ nghĩa (modifying phrases), vị trí của trạng từ và giới từ, những động tính từ “động đưa” (participles), và động danh từ (gerund).

Chẳng hạn như câu văn “*We chose a species relevant to the disease called Zanka*” thì có vấn đề về bổ nghĩa. Người đọc không biết *Zanka* là tên của bệnh hay là tên của loại động vật. Thay vì viết như thế, chúng ta có thể viết lại cho rõ ràng hơn: “*We chose a species called Zanka that is relevant to the disease.*” Tương tự, câu văn “*The study involved a cohort of patients in CR Hospital with osteoporosis*” cũng có thể thay thế bằng “*The study involved a cohort of patients with osteoporosis in CR Hospital.*”

Vị trí của trạng từ cũng là một điều đáng bàn, vì tác giả thường đặt trạng từ sai vị trí. Thử xem xét và so sánh hai câu văn sau đây:

*“It is the investigator’s duty to **inform** the patient **fully** before initiating the therapy”*

*“It is the investigator’s duty to **fully inform** the patient before initiating the therapy”*

Câu nào đúng hơn?

Thỉnh thoảng có những tác giả dùng nhiều động-tính-từ động đưa (dangling participles) làm cho câu văn khó hiểu và... yếu điệu. Cách viết này khá phổ biến ở người Á châu, có lẽ một phần là do văn hoá thi vị. Chẳng hạn như câu “*Structured into various sections, the readers of this review can choose the topic of*

primary interest” là một ví dụ tiêu biểu cho kiểu viết đong đưa đó. Thay vì viết như vậy, tác giả có thể dùng cách viết chuẩn và dễ hiểu hơn: *“The review is structured into various sections, which allows the readers to choose the topic of primary interest.”*

2.3 Tính chính xác

Người Á châu chúng ta quen với những cách nói/viết xấp xỉ. Do đó, chúng ta có những chữ như “khoảng”, “cỡ”, “gần”, v.v. Những cách viết đó không có vấn đề gì trong ngôn ngữ hàng ngày, nhưng với văn phong khoa học thì khó chấp nhận. Văn phong khoa học đòi hỏi yếu tố chính xác và định lượng.

Để minh họa cho ý này, chúng ta thử đọc câu:

“Most patients have suffered from an adverse event”

Chữ *most* ở đây chưa thỏa đáng, vì chỉ nói chung là “phần đông”. Nhưng “phần đông” là bao nhiêu thì không ai biết. Trong văn phong báo chí thì cách viết đó có thể chấp nhận được, nhưng văn phong khoa học thì cách viết đó có thể bị phê bình. Thay vì viết như trên, chúng ta có thể cho một con số cụ thể:

“Ninety percent of patients have suffered from an adverse event”

hay:

“Most patients (90%) have suffered from an adverse event”

Do đó, giảm dùng những chữ như *most*, *many*, *much*, *enormous*, *majority*, *etc.* càng nhiều càng tốt, và thay vào đó là những con số cụ thể (nếu được).

2.4 Văn phong (style)

Yếu tố thứ tư (và cũng là yếu tố sau cùng) để có một bài báo khoa học tốt là về văn phong. Khi nói đến văn phong ở

đây, tôi muốn nói đến cách dùng chữ. Dùng chữ khách quan (objective words), chữ mang tính tích cực, và tránh dùng các biệt ngữ của chuyên ngành (jargon).

Cần phải nhớ rằng văn phong bài báo khoa học là nghiêm trang. Đặc tính này có nghĩa là người viết cần phải giảm (hay tốt hơn là tránh) dùng những chữ:

Không có điểm tham chiếu hay giới hạn: *a lot, fairly, long term, quite, really, short term, slightly, somewhat, sort of, very.*

Định tính: *assuredly, beautiful, certainly, disappointing, disturbing, exquisite, fortuitous, hopefully, inconvenient, intriguing, luckily, miraculously, nice, obviously, of course, regrettable, remarkable, sadly, surely, unfortunately.*

Lấp lửng: *alright, basically, in a sense, indeed, in effect, in fact, in terms of, it goes without saying, one of the things, with regard to.*

Màu mè: *agree to disagree, bottom line, brute force, cutting edge, easier said than done, fell through the cracks, few and far between, food for thought, leaps and bounds, no nonsense, okay, quibble, seat of the pants, sketchy, snafu, tad, tidbit, tip of the iceberg.*

Văn phong ở đây cũng còn có nghĩa là cách viết nhấn mạnh sao cho thích hợp. Nguyên tắc chung trong tiếng Anh là bắt đầu mệnh đề với điểm mình muốn nhấn mạnh. Các bạn thử so sánh 4 câu văn sau đây:

1. *Although the treatment is highly effective, it has significant side effects.*
2. *Although the treatment has significant side effects, it is highly effective.*
3. *The treatment has significant side effects, but it is highly effective.*
4. *The treatment is highly effective, and it has significant side effects.*

Bốn câu văn nói lên một ý, nhưng cách nhấn mạnh thì rất khác nhau. Câu đầu tiên muốn nhấn mạnh rằng điều trị có hiệu quả tốt (điểm chính), nhưng có vài tác động phụ (điểm không muốn nhấn mạnh). Câu thứ hai thì lại nhấn mạnh vấn đề tác động phụ, mà xem hiệu quả là điểm kém quan trọng.

Nên nhớ rằng không nên bắt đầu câu văn bằng những chữ không quan trọng. Chẳng hạn như câu:

First let us consider that bisphosphonates reduce the risk of fracture.

thì chữ mệnh đề đầu tiên (*First let us consider that*) có lẽ không cần thiết. Nên viết trực tiếp hơn:

Bisphosphonates reduce the risk of fracture.

Sau cùng, cố gắng dùng văn phong tích cực hơn là tiêu cực. Thay vì viết "*The study failed to show a significant relationship between the genetic variant and hypertension (P = NS)*" (tiêu cực) thì nên viết "*Although the minor allele of the genetic variant was associated with a greater risk of hypertension, the association did not reach a statistical significance (P = 0.07).*"

3. Những sai sót phổ biến về tiếng Anh trong bài báo khoa học

3.1 Lẫn lộn giữa số ít và số nhiều.

Đa số chúng ta đều biết số nhiều của *cell*, *gene* là *cells*, *genes*. Nhưng có những danh từ thỉnh thoảng làm người viết lẫn lộn do những bài báo từ Mỹ và Anh có cách viết khác nhau. Các bài báo Mỹ thường xem *data* là số ít, nhưng đúng ra đó là số nhiều của *datum*. Nói chung, những danh từ gốc Latin với *um* thì số nhiều là *a*: medium thành *media*, datum

thành data, bacterium thành bacteria, optimum thành optima, symposium thành symposia, equilibrium thành equilibria, v.v.

Ngoài ra, những *is, us* thành *eses, i*: hypothesis số nhiều là hypotheses, thesis thành theses, axis thành axes, locus thành loci, fungus thành fungi, v.v. Điều này có nghĩa là khi chia động từ, các bạn nên nhớ phân biệt giữa “The data were analyzed by ...”, chứ không phải “The data was analyzed by”

3.2 Chia động từ sai vì ‘quên’ chủ từ.

Bất cứ ai học tiếng Anh cũng đều biết chia động từ phải phù hợp với chủ từ số nhiều hay số ít. Nhưng những câu văn phức hợp thì có khi làm cho người viết “quên” chủ từ chính là gì, và dẫn đến chia động từ sai. Chẳng hạn như câu:

“The effects of bone loss on the risk of CVD disease was studied in a sample of ...”

thì cách chia động từ sai, bởi vì chủ từ chính là “the effects”. Vì thế, đúng ra, câu văn trên nên viết là:

“The effects of bone loss on the risk of CVD disease were studied in a sample of ...”

Cũng cần chú ý đến *or/either/neither*, chỉ chia động từ theo chủ đề gần nhất. Ví dụ như

“Either the doctor or the patients was aware of the presence of AND”

thì không đúng, vì chủ từ gần nhất là *patients*. Do đó, phải viết câu trên là:

“Either the doctor or the patients were aware of the presence of AND”

Tuy nhiên, câu văn với *none* thì thường chia động từ theo số ít. Ví dụ:

"None of the applicants is fully qualified."

3.3 Những mệnh đề “đong đưa”.

Cẩn thận với vị trí của trạng từ và động từ để không làm thay đổi ý nghĩa của câu văn. Một ví dụ tiêu biểu là *"The study involved a sample of children in a local hospital with type I diabetes"* dễ bị hiểu lầm hơn là *"The study involved a sample of children with type I diabetes in a local hospital."* Một ví dụ khác cũng vui vui: *"We selected a scientist with good expertise in the field called Tom Smith"* dễ bị hiểu lầm là lĩnh vực nghiên cứu tên là Tom Smith, nhưng viết lại cho rõ ràng hơn thì *"We selected a scientist called Tom Smith who has good expertise in the field."*

3.4 Dùng chữ ‘bình dân’ hay văn nói.

Rất thường xuyên, tôi hay thấy những bản thảo có cách viết theo văn nói, thường không thích hợp cho văn khoa học. Chẳng hạn như:

- *We wondered if ...*
- *We did just that ...*
- *We asked participants whether they had ...*
- *There were a couple of samples in the study ...*

Đó là những cách viết được xem là unprofessional, thiếu tính chuyên nghiệp. Nên tránh những cách viết và những chữ thiếu tính chuyên nghiệp.

3.5 Dùng sai while/since/as liên quan đến thời gian.

Những chữ như *while*, *since*, *as* dùng để chỉ mối liên hệ mang tính thời gian. Không dùng những chữ này để hàm ý “mặc dù” (*while*) hay “bởi vì” (*as*) hay “do vậy” (*since*).

Đúng: *While the patient was being examined in the clinic, the patient was also undergone ...*

Sai: *While fracture is associated with lower bone density, it can also be related to ...*

Đúng: *Whereas/Although fracture is associated with lower bone density, it can also be related to ...*

3.6 Viết hoa không đúng cách và không cần thiết.

Nhiều bản thảo từ Việt Nam và Trung Quốc thường có những cách viết hoa rất ... tùy tiện. Chẳng hạn như người ta viết “the Drug has effect” hay “the drug Beta-blocker”, v.v. Những chữ đó không có lí do gì phải viết hoa.

3.7 Lẫn lộn giữa *which* và *that*.

Đây là một trong những lẫn lộn rất phổ biến ở những người mà tiếng Anh không phải là tiếng mẹ đẻ. Tuy nhiên, phân biệt và dùng đúng không phải là điều dễ dàng. Qui luật chung là:

- (a) “*that*” được sử dụng trong ý nghĩa giới hạn và mang tính nhận dạng;
- (b) “*which*” thì không có ý nghĩa giới hạn, nhưng có nghĩa bổ nghĩa.

Ví dụ về giới hạn ý nghĩa và nhận dạng:

*“the drug **that** is used for reducing bone resorption is bisphosphonate”*

không dùng which trong câu trên, vì câu văn mang tính nhận dạng danh tính của thuốc. Nhưng

“the receptor, which is present in bone cells, is slightly different from ...”

thì đúng. Chú ý dấu phẩy trước which và mệnh đề bổ nghĩa “present in bone cells”.

3.8 “an” và “a” trước những chữ viết tắt

Chú ý rằng “A university”, chứ không phải “An university”; do đó, bất cứ chữ viết tắt nào với U đều dùng “a”. Chẳng hạn như “a UNT component”, nhưng “an MMC-modulated reaction”. Nói chung, các chữ viết tắt với phụ âm đi đầu như SDS, SOS, LUT, NIA, v.v. đều dùng “an”, chứ không phải “a”.

3.9 Lẫn lộn giữa “et al.” và “etc.”

Chữ viết tắt “et al” có gốc Latin, với “et” có nghĩa là “và”, và “al” là từ “alii” có nghĩa là “cái khác”. Do đó, et al có thể hiểu là “và khác nữa”, như “Nguyen et al” có thể hiểu là “Nguyen và cộng sự khác”.

Chữ “etc” dĩ nhiên có nguồn gốc Latin “et cetera”, có nghĩa là “và những gì giống thế”. Nó có nghĩa giống như “vân vân” hay “...” trong cách viết tiếng Việt. Nhưng chữ này, “etc” KHÔNG bao giờ dùng trong văn phong khoa học.

3.10 Lẫn lộn giữa “e.g” và “i.e”

“i.e.” trong văn bản khoa học có gốc từ tiếng Latin “id est”, có nghĩa “có nghĩa là” (hay tiếng Anh ‘that is’).

“e.g.” cũng xuất phát từ tiếng Latin “exemplia gratii”, có nghĩa là “ví dụ như” (hay tiếng Anh ‘for example’).

Chúng ta có thể viết “*the predictors (i.e., risk factors) of this model were ...*”, nhưng không thể viết “*the predictors (e.g., risk factors) of this model were ...*”.

Tuy nhiên, chúng ta có thể viết “*Lifestyle factors (e.g., smoking, alcohol intakes) were considered ...*”, nhưng không thể viết “*Lifestyle factors (i.e., smoking, alcohol intakes) were considered ...*”.

3.11 Không có khoảng trống giữa số và đơn vị đo lường.

Bài báo khoa học yêu cầu phải có khoảng trống (space) giữa số và đơn vị đo lường. Vài ví dụ cụ thể như sau:

(a) Phải viết là 0,61 cm, chứ không phải là 0,61g/cm.

(b) Ngay cả đơn vị đo lường nhiệt độ cũng phải có khoảng trống: cách viết đúng là 37 °C, chứ không phải 37°C.

(c) Đơn vị li tâm (centrifuge) cần phải có “x” như chúng ta biết, nhưng phải có khoảng trống. Cách viết đúng là 10.000 x g, chứ không phải 10.000xg.

(d) Ngoại lệ: số phần trăm thì không cần khoảng trống: chúng ta có thể viết 35% thay vì 35 %.

3.12 Khoảng trống cho sai số chuẩn và dấu ±

Phải có khoảng trống trước và sau dấu cộng trừ ±. Nên viết 45 ± 11 , chứ không nên viết 45 ± 11 . Tương tự, số cỡ mẫu cũng phải có khoảng trống trước và sau dấu bằng: viết $n = 5$, chứ không nên $n=5$.

3.13 Dấu nối (“-” hay hyphen) giữa số và đơn vị, nếu dùng nó như là một mệnh đề tính từ.

Tất cả các tính từ ghép với con số cần phải có dấu gạch nối. Nên viết “40-mm rod”, chứ không phải “40 mm rod”. Ở đây,

40-mm có ý nghĩa tính từ. Ngoài ra, chúng ta viết “*Two-sided P value*” hay “*Four-step procedure*” chứ không phải “*Two sided P value*” hay “*Four step procedure*”. Dĩ nhiên, chúng ta có thể viết “*A procedure with 4 steps*” mà không có vấn đề gì về văn phạm.

Dấu gạch nối cũng có thể dùng cho những tiền tố ngữ như “*Multi-component model*”, “*Bi-directional tool*”. Trong thực tế, tôi thấy nhiều người (nhất là bên Mỹ) bỏ dấu gạch nối trong các chữ trên.

3.14 Dấu nối cho các chữ ghép.

Các chữ ghép (compound word) ở đây bao gồm những chữ ghép giữa danh từ và động từ (thường là thì quá khứ). Chẳng hạn như muốn nói phản ứng do thuốc gây ra, chúng ta có thể viết “*Reaction attributable to drug*” (hơi dài), nhưng một cách viết gọn hơn là “*Drug-induced reaction*”, và chú ý có dấu nối. Tương tự, chúng ta hay thấy các chữ như “*time-dependent variable*” hay “*time-variant model*”, v.v.

Dĩ nhiên, khi bắt đầu câu văn, không ai viết “22 participants”, mà phải là “Twenty-two participants”.

3.15 Dùng dấu phẩy trong các mệnh đề ngoặc kép

Những mệnh đề ngoặc kép (parenthetical phrases) là những mệnh đề gồm nhiều chữ, chúng không cần dấu phẩy. Trước đây thì cần, nhưng ngày nay, các tập san không thích dùng dấu phẩy nữa.

Ví dụ: chúng ta có thể viết “*the protein kinase inhibitor genistein was related to*”, chứ không cần phải viết theo kiểu cổ điển “*the protein kinase inhibitor, genistein, was related to*.” Cả hai cách viết đều đúng, chỉ có cách đầu được các tập san khoa học ưa chuộng hơn cách viết sau.

Nhưng có những mệnh đề theo ý nghĩa “and” và “but” thì cần dấu phẩy. Ví dụ như chúng ta nên viết:

“The placebo group, contrary to our hypothesis, exhibited no effect”,

“The participants were largely drawn from B and, as a result, biased toward ...”

chứ không nên viết:

“The placebo group contrary to our hypothesis exhibited no effect”

“The participants were largely drawn from B and as a result biased toward ...”

3.16 Quá nhiều số lẻ thập phân

Một trong những sai sót hay thấy nhất trong các bản thảo bài báo khoa học là tác giả trình bày con số với nhiều số lẻ không cần thiết. Chẳng hạn như chúng ta biết rằng máy DXA đo mật độ xương chính xác 2 số lẻ (như 0,71, 0,90 g/cm²), do đó báo cáo 0,711 hay 0,899 là không cần thiết. Đó có thể xem là một sai sót. Tương tự, huyết áp 150 mmHg có ý nghĩa, chứ 150,65 mmHg thì chẳng có ý nghĩa gì, mà chỉ là một ‘nô lệ’ con số!

Tương tự, con số phần trăm chỉ cần chính xác 1, hay cao lắm là 2 số lẻ. Nên viết 90,4%, chứ không nên viết 90,415%.

3.17 Lẫn lộn những chữ có cùng (hay giống) cách phát âm

Chú ý những chữ có phát âm giống giống nhau như:

effect / affect

here / hear

its / it's

whose / who's

Trên đây chỉ là những sai sót phổ biến trong bài báo khoa học. Trong thực tế thì còn nhiều sai sót khác nữa. Những sai sót này thường hay thấy trong những bản thảo của các nhà khoa học ngoài các nước nói tiếng Anh. Nhưng ngay cả những nhà khoa học trong các nước nói tiếng Anh vẫn vấp phải những sai sót trên. Do đó, vấn đề không hẳn là văn phạm tiếng Anh, mà là làm quen với văn phong khoa học. Bất cứ ai, kể cả tôi, cũng đều có sai sót, thậm chí sai sót ngay trong những bài báo đã công bố. Do đó, biết được những sai sót này là để giúp cho bản thảo kế tiếp sẽ hoàn chỉnh hơn. Chúc các bạn may mắn!

* * *

Tóm lại, một bài báo khoa học được xem là tốt, ngoài nội dung tốt, thì cách viết cũng phải tốt. Cách viết tốt là viết rõ ràng và thuyết phục. Để đạt hai yêu cầu đó (rõ ràng và thuyết phục), người viết phải chú ý đến 4 điểm chính: cấu trúc đoạn văn một cách logic; đúng văn phạm; dùng chữ chính xác và định lượng; và văn phong khách quan.

Francis Darwin (con trai của Charles Darwin) từng nói rằng trong khoa học, công trạng thuộc về người thuyết phục thế giới, chứ không phải thuộc về người đề ra ý tưởng đầu tiên (*"In science, the credit goes to the man who convinces the world, not to the man to whom the idea first occurs"*). Do đó, có nội dung tốt là quan trọng, nhưng cách viết một cách thuyết phục cũng quan trọng không kém. Làm khoa học là phải học viết và nói. Tôi có thể nói câu đó cả chục lần để nhấn mạnh rằng viết và nói là hai kỹ năng giúp nhà khoa học thành công hay thất bại.

Tham khảo và đọc thêm:

1. Robert B. Taylor. The Clinician's Guide to Medical Writing. Srpinge 2005. Đây là cuốn quan trọng nhất và là 'gối đầu giường' cho bất cứ ai làm trong y học.
2. Martin Volk and Sven Lautenbach. Writing scientific papers - 10 biggest mistakes from a reviewer's perspective. Chia sẻ kinh nghiệm sai sót mà chính tác giả trải qua.
http://www.iemss.org/sites/iemss2012/images/WS_J2_Volk_Lautenbach_pdf.pdf
3. Sydney University: Scientific Writing. Một hướng dẫn rất cụ thể và chi tiết.
http://sydney.edu.au/science/biology/learning/generic_skills/scientific-writing.pdf
4. Myron L. Toews. The "Grammar Hammer": Common Mistakes in Scientific Writing. Đây là một bài nói chuyện mà tác giả trình bày những hướng dẫn (và sai sót) trong cách trình bày những qui ước về tiếng Anh trong khoa học.
https://medschool.creighton.edu/fileadmin/user/medicine/Departments/Biomedical_Sciences/PowerPoint/Grammar_Hammer-_CU_BioMedSci-_Oct_2012.pptx
5. Jorge Faber. Writing scientific manuscripts: most common mistakes. Dental Press J. Orthod. vol.22 no.5 Maringá Sept./ Oct. 2017. Một chia sẻ kinh nghiệm về những sai sót phổ biến ở người không nói tiếng Anh như tiếng mẹ đẻ.
http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2176-94512017000500113
6. Mariel A. Marlow. Writing scientific articles like a native English speaker: top ten tips for Portuguese speakers. Clinics vol.69 no.3 São Paulo Mar. 2014. Một chia sẻ khác về cách viết bài báo khoa học.
http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1807-59322014000300153

7. Và sau cùng là sách của tôi *“Từ nghiên cứu đến công bố Kỹ năng mềm cho nhà khoa học”* (Nhà xuất bản Tổng hợp TP.HCM) có hướng dẫn về cách soạn bài báo khoa học, cách trình bày powerpoint, cách viết đề cương nghiên cứu. Sách tổng hợp và chia sẻ những kinh nghiệm trong việc xuất bản bài báo khoa học.

<https://nxbhcm.com.vn/34/tu-nghien-cuu-den-cong-bo-ky-nang-mem-cho-nha-khoa-hoc-1187>

NHỮNG BÀI HỌC VỀ VIẾT BÀI BÁO KHOA HỌC

Bất cứ ai làm nghiên cứu khoa học cũng đều có những trải nghiệm khó quên về viết bài báo khoa học, nhất là những trải nghiệm cá nhân lần đầu viết. Trong chương này, tôi chia sẻ cùng các bạn kinh nghiệm soạn và công bố bài báo khoa học, và những bài học của mỗi lần trải nghiệm. Tôi cũng liệt kê một số lý do bài báo khoa học bị từ chối để các bạn chuẩn bị. Tôi sẽ chia sẻ trong vai trò của một tác giả và người biên tập.

Những bài học rút ra từ người học trò

Bài báo đầu tiên tôi viết đã gần 30 năm trước đây và lần trải nghiệm đó thật khó quên. Tôi có hai người thầy hướng dẫn, nhưng người hướng dẫn phụ trong thực tế lại là người... hướng dẫn chính. Lý do là giáo sư hướng dẫn chính rất bận rộn, mỗi tháng chỉ gặp riêng khoảng vài phút, nhưng hàng tuần thì vẫn gặp trong các buổi họp labo. Do đó, công việc ‘cầm tay chỉ việc’ lại do một giáo sư trẻ hơn phụ trách. Vị giáo sư đó cũng bận với công việc bên bệnh viện và đại học, nên ông cũng chỉ gặp tôi mỗi tuần khoảng 10 phút để trao đổi và kiểm tra tiến độ công việc. Sự hướng dẫn của giáo sư mang tính định hướng, chứ không đi vào chi tiết. Người đi vào chi tiết là một nhà khoa học cấp thấp hơn, như hậu tiến sĩ chẳng hạn. Sau này tôi mới hiểu

bài học thứ nhất là làm nghiên cứu sinh chủ yếu là tự học dưới sự hướng dẫn của giáo sư; do đó, năng lực cũng như kỹ năng tự học rất quan trọng.

Trước khi soạn bài báo đầu tiên, giáo sư hướng dẫn nói với tôi rằng đây là một nghiên cứu quan trọng, và ý định là sẽ công bố trên một tập san danh giá. Vấn đề là bắt đầu từ đâu? Ông khuyên tôi nên tìm đọc những bài trước đây trên tập san đó, và cứ thế mà viết theo. Thời đó, các tập san khoa học xuất bản trên báo giấy, chứ không có phiên bản trực tuyến, và tôi phải vào thư viện để sao lại những bài báo liên quan. *Bài học 2 mà tôi học được là học từ những bài báo trước đã công bố trên tập san mà mình có ý định sẽ công bố.*

Sau hơn 8 tuần lễ miệt mài viết, soạn hình ảnh, chỉnh sửa, và lại chỉnh sửa, cuối cùng tôi cũng có một bản thảo dài hơn 30 trang. Tôi rất tự hào với thành quả của mình, và hẹn giờ để trình bày cho giáo sư hướng dẫn xem và góp ý. Giáo sư hướng dẫn nhìn qua bản thảo của tôi, ông đột nhiên quăng vào thùng rác bên cạnh, và nhìn tôi dò xét. Lúc đó, tôi rất ngạc nhiên, có chút tức tối nhưng cũng lúng túng, không biết làm gì và nói gì. Tôi trầm nghĩ chẳng lẽ tiếng Anh của mình tệ như thế?

Ông nhìn tôi và mỉm cười, nhặt bản thảo lên, rồi giải thích những sai sót và những chỗ chưa đúng với chuẩn mực khoa học. Vừa giải thích, ông vừa sửa (bằng bút mực đỏ), gạch bỏ chỗ này, viết thêm chỗ kia, ghi chú trên hình ảnh, v.v. Và, kết quả là một bản thảo đầy màu đỏ! Nó không còn là văn của tôi nữa. Tôi còn nhớ ông giáo sư phê bình văn viết của tôi nặng chất thơ quá và nhiều chữ *rhetoric* quá (có lẽ tôi xuất thân từ văn hoá Việt Nam và quen với những ví von). Lí do đơn giản là vì văn phong khoa học có những đặc thù và thuật ngữ mà chỉ có người trong

chuyên ngành mới am hiểu. *Bài học 3 là tiếng Anh mà chúng ta học từ trường hay qua các kì thi như TOEFL, IELTS tuy có ích trong giao tiếp nhưng giúp rất ít vào việc viết bài báo khoa học.*

Văn phong khoa học không thể ví von và màu mè, mà phải chính xác, đơn giản và có khi hình tượng. Theo Giáo sư Steven Pinker, một phần ba của bộ não con người được giành cho thị giác, và sự thật này có nghĩa là văn chương khoa học phải làm cho độc giả ‘thấy’ mục tiêu cụ thể của bài báo, và làm được điều này sẽ có hiệu ứng rất tích cực cho bài báo. Nhiều thí nghiệm đã cho thấy rằng độc giả hiểu và nhớ các thông tin tốt hơn khi thông tin được diễn tả bằng một ngôn ngữ mà họ có thể thấy bằng hình ảnh. Ngoài ra, mỗi bài báo là một câu chuyện, và cách viết kể chuyện giúp người đọc lĩnh hội rất nhanh. Ngày nay, cách viết kể chuyện hay đàm thoại cũng được rất nhiều tập san khoa học khuyến cáo các tác giả nên viết. *Bài học 4 mà tôi học được là viết văn khoa học phải nhắm đến kể một câu chuyện có đầu, có đuôi, và văn phong giàu hình ảnh.*

Người mới bước vào khoa học như tôi thời đó chưa có tầm nhìn lớn, nên rất quan tâm (thậm chí ‘khoe’) đến phần phương pháp và kĩ thuật, nhưng người có kinh nghiệm thì quan tâm đến thông điệp chính của bài báo. Câu hỏi mà giáo sư hướng dẫn hay giầy vò là tại sao bài báo này quan trọng, và cái thông điệp chính là gì. Có khi phải bỏ ra cả tuần chỉ nghĩ ra cách viết thông điệp chính. Thông điệp đó phải được viết bằng văn phong đơn giản. Giáo sư hướng dẫn thường hay chỉ tay xuống đường phố và nói với tôi rằng ‘phải viết sao cho người lái taxi vẫn hiểu được.’ Thông điệp chính phải nhất quán với dữ liệu, không được phát biểu những gì đi ra ngoài phạm vi của dữ liệu. *Bài học 5 là nội dung và thông điệp bài báo rất quan trọng, nên cần phải đầu tư thời gian để viết cho thuyết phục.*

Dữ liệu ở đây không phải là số liệu, mà còn là hình ảnh và y văn trước đây. Dữ liệu phải được trình bày bằng một hình thức trang nhã. Trang nhã từ cách chọn màu, chọn phông chữ, kích thước chữ, khoảng cách giữa các dòng chữ, v.v. đều phải đảm bảo đúng như yêu cầu của tập san. Thời đó, những hình ảnh chưa được số hóa, nên phải in trên giấy bóng, mà phải kèm theo chú thích sao cho người đọc có thể hiểu được một cách dễ dàng. Giáo sư hướng dẫn tôi thường hay nói câu tương đương với cách nói của người Việt chúng ta, ‘người đẹp vì lụa’, có nghĩa là một nội dung hay và quan trọng cần phải được cho ‘mặc’ một bộ đồ đẹp và trang nhã. *Bài học 6 là hình thức trình bày dữ liệu cũng rất quan trọng.*

Nhưng bản thảo chưa dừng ở đó, mà còn phải qua nhiều chỉnh sửa từ những người đồng tác giả nữa. Chỉ riêng với giáo sư trực tiếp hướng dẫn, bản thảo đã phải qua ít nhất 5 lần chỉnh sửa. Giáo sư hướng dẫn chính tuy là người ít gặp nhất, nhưng lại là người chỉnh sửa nhiều nhất. Ông nổi tiếng là người rất thông minh (học cử nhân năm 15-16 tuổi), khó tính, và rất *elite*, và ông kì vọng những người chung quanh phải như ông! Ông chỉnh sửa thêm, yêu cầu phân tích lại một số câu hỏi, và ‘ác mộng’ là vấn đề trình bày hình ảnh sao cho có chất lượng tốt. Ông cũng chính là người quyết định ai là tác giả đầu (tôi), ai là tác giả thứ hai (giáo sư hướng dẫn phụ), và ai là tác giả sau cùng (chính là ông). Sếp có thể không phải là người trực tiếp hướng dẫn, nhưng lại là người có tiếng nói sau cùng; đây chính là người chịu trách nhiệm chính trước công chúng, là tác giả liên lạc (*correspondence author*). Lúc này tôi mới học được *bài học 7 là thư tự tác giả rất quan trọng, vì nó cho biết ai là ‘sếp’ trong công trình nghiên cứu.*

Sau giáo sư hướng dẫn chính phê chuẩn bản thảo, nó được gửi cho một tập san danh tiếng qua đường bưu điện bảo đảm. Thời đó, chưa có hệ thống nộp bài báo trực tuyến như bây giờ, nên tất cả bản thảo (4 bản sao) đều phải gửi qua bưu điện hỏa tốc. Khoảng 6 tháng sau, chúng tôi nhận được thư của ban biên tập, kèm theo những bình luận của 3 chuyên gia nặc danh trong chuyên ngành. Đọc qua 8 trang bình duyệt của họ, cảm giác đầu tiên của tôi là choáng váng. Bình luận mang tính khen có, nhưng chê thì nhiều hơn. Khổ nỗi những gì họ chê xem ra... rất có lí. Có chuyên gia đòi hỏi thêm dữ liệu mà chúng tôi không thể nào có được. Hai giáo sư hướng dẫn xem qua bình duyệt và quyết định có lẽ bài báo không có cơ hội ở tập san này (đó là tập san *New England Journal of Medicine* -- *NEJM*). *Bài học 8 là chuẩn bị tinh thần bị từ chối và bị chê, vì từ chối là một qui luật trong xuất bản khoa học.*

Những những ý kiến của các chuyên gia bình duyệt rất có ích để chỉnh sửa tiếp. Chúng tôi nộp cho một tập san khác tuy cũng có uy tín cao, nhưng thấp hơn tập san *NEJM*. Lại chờ thêm 4 tháng để nhận được thư bình duyệt, nhưng lần này thì khả quan và 'thân thiện' hơn lần trước. Ban biên tập cho chúng tôi 3 tháng trả lời các bình luận, và nộp lại bản thảo. Khi ban biên tập cho cơ hội trả lời có nghĩa là chúng tôi có... cơ may. Lại thêm một thời gian quây quần trong chu trình bình duyệt lần thứ hai - trả lời lần 2 - bình duyệt lần 3 - trả lời lần 3, v.v. tất cả tốn gần 2 tháng nữa thì bài báo mới được chấp nhận cho công bố. Như vậy, tính từ lúc hoàn thiện bản thảo đến lúc công bố mất khoảng 1 năm. Đó là thời gian trung bình, vì trong thực tế, có những bài báo may mắn hơn và có thời gian công bố ngắn hơn, nhưng cũng có những bài báo cần đến 2 năm để có nơi công bố. *Bài học 9 là mỗi lần bị từ chối là một cơ hội để*

cải thiện cho lần sau, tập san có uy tín càng cao thì yêu cầu về phẩm chất càng cao và họ từ chối càng nhiều.

Trả lời bình duyệt là một nghệ thuật nhưng cũng là một khoa học. Nghệ thuật là đặt vấn đề và những phê bình của các chuyên gia trong bối cảnh. Có khi cùng một dữ liệu, nhưng có chuyên gia khen, mà cũng có chuyên gia chê, và nghệ thuật ở đây là dung hoà hai quan điểm trái chiều. Tính khoa học là trả lời bình duyệt bằng chứng cứ và dữ liệu khoa học, chứ không cảm tính và tuyệt đối không được chỉ trích cá nhân. Không bao giờ bảo chuyên gia bình duyệt nên đọc bài báo này hay cuốn sách kia, vì đó là một cách trả lời rất ư mất lịch sự và thiếu tôn trọng, và bài báo rất dễ bị từ chối. *Bài học 10 là trả lời các ý kiến bình duyệt một cách đầy đủ, lịch sự, và có chứng cứ.*

Công bố được một vài bài báo khoa học là một thành tích, nhưng chưa thể thành thạo trong cách viết bài báo khoa học. Sau bài báo đầu tiên, tôi lao vào nghiên cứu và viết một loạt bài báo khoa học liên quan. Tuy nhiên, cứ mỗi lần tôi viết xong thì hai giáo sư hướng dẫn lại... tiếp tục sửa. Họ sửa nhiều đến nỗi đến bài thứ 5 mà tôi vẫn thấy mình hình như chẳng viết được bài nào hoàn chỉnh! Nhưng thú thật tôi thấy bài nào hai giáo sư sửa đều hợp lí và... hay. So sánh bản thảo của mình và bản thảo sau khi đã được chỉnh sửa, tôi thấy rõ ràng là mình vẫn chưa học hết kĩ năng bài báo khoa học. Thật ra, kĩ năng tiếng Anh chỉ là một vấn đề; vấn đề lớn hơn là bức tranh chung và chiến lược viết, và chiến lược thì đòi hỏi người giàu kinh nghiệm trong chuyên ngành. Phải đến bài thứ 10 và năm thứ 4 tôi mới cảm thấy mình đủ tự tin để tự mình soạn một bài báo khoa học. Do đó, *đối với một người mà tiếng Anh không phải là tiếng mẹ đẻ, cho dù đã xong luận án tiến sĩ, thì vẫn chưa đủ khả năng để tự mình viết bài báo khoa học hoàn chỉnh.*

Cơ chế bình duyệt và vận hành của ban biên tập

Theo thời gian, tôi có cơ hội đóng góp nhiều hơn cho chuyên ngành qua việc tham gia bình duyệt và biên tập. Tôi trở thành chuyên gia bình duyệt cho các tập san, và sau này được giao cho trách nhiệm cao hơn là phó biên tập (Associate Editor) cho một số tập san, trong đó có tập san *JBMR*, được xem là quan trọng nhất và uy tín nhất trong chuyên ngành xương. Tập san *JBMR* nhận khoảng 1000 bài báo mỗi năm, nhưng đa số (khoảng 80%) là bị từ chối, và chỉ chấp nhận cho công bố khoảng 15-20%. Qua biên tập và quyết định chấp nhận hay từ chối bài báo, tôi nhận ra rằng các tập san khoa học là nơi có chế độ kiểm duyệt khắt khe nhất! Thật ra, người trong khoa học không gọi đó là cơ chế “kiểm duyệt”, mà là “gác cổng khoa học.”

Quy trình và ‘số phận’ của một bản thảo bài báo khoa học nằm trong sự tương tác giữa ban biên tập và chuyên gia bình duyệt. Nói là ‘ban biên tập’, nhưng người trực tiếp phụ trách chính là phó biên tập. Theo quy trình chuẩn này, khi bản thảo bài báo được gửi đến tập san thì hàng loạt việc làm sẽ được khởi động:

Trước hết, một nhân viên phụ tá của tập san sẽ kiểm tra bản thảo có đạt yêu cầu về cơ cấu, hình thức, và các qui định của tập san liên quan đến tài liệu tham khảo và các thông tin quan trọng. Mỗi tập san có những qui định riêng, và tác giả phải tuân thủ theo. Ngoài ra, phụ tá biên tập còn kiểm tra những câu văn có trùng hợp với các bài báo trước đây để đảm bảo đạo văn không xảy ra, hay ‘tự đạo văn’ ở mức độ có thể chấp nhận được. Bản thảo bài báo không đáp ứng các yêu cầu và qui định kỹ thuật này sẽ được trả lại cho tác giả để chỉnh sửa và nộp lại.

Sau khi phần kiểm tra kỹ thuật và qui định xong, bản thảo sẽ được tổng biên tập giao cho một phó biên tập phụ trách. Một tập san có 3-5 phó biên tập, và mỗi người phụ trách một mảng chuyên môn hẹp mà phó biên tập có tiếng. Phó biên tập đọc qua bản thảo và sẽ quyết định từ chối ngay hoặc gửi ra ngoài cho các chuyên gia bình duyệt. Nếu từ chối ngay, thì tác giả sẽ được thông báo trong vòng 1 tuần bằng một lá thư tôi hay nói đùa là 'lá thư đau lòng' (vì câu đầu tiên là 'tôi đau lòng báo tin rằng ...'). Nếu quyết định bản thảo có thể gửi ra ngoài bình duyệt, thì 2-3 chuyên gia có tiếng trong chuyên ngành sẽ được mời bình duyệt.

Các chuyên gia bình duyệt sẽ đọc và đánh giá bài báo, viết báo cáo gửi cho phó biên tập. Báo cáo của mỗi chuyên gia tập trung vào ý tưởng, phương pháp, và cách diễn giải kết quả. Họ cũng kèm theo một khuyến cáo: (a) chấp nhận cho công bố không cần sửa; (b) cần sửa chút ít; (c) cần chỉnh sửa nhiều hay viết lại; hay (d) từ chối. Chỉ một trong 3 chuyên gia bình duyệt đề nghị từ chối thì khả năng bài báo sẽ bị từ chối lên rất cao. Có nhiều trường hợp bài báo phải trải qua 3 lần bình duyệt, và tốn rất nhiều thời gian (trên 12 tháng) nhưng cuối cùng có khi vẫn bị từ chối!

Nếu bài báo được chấp nhận, tác giả sẽ nhận được thư của tổng biên tập thông báo. Sau khi xong phần khoa học, bản thảo sẽ được gửi cho bộ phận sản xuất và biên tập của nhà xuất bản. Biên tập viên của nhà xuất bản sẽ xem xét tất cả những bảng biểu và văn bản trong bài báo. Có khi họ đề nghị chỉnh sửa cách viết cho gọn hơn, rõ ràng hơn, vì nhiệm vụ của họ là tiết kiệm chữ (và qua đó tiết kiệm mực in). Sau khi tác giả đồng ý với những chỉnh sửa của biên tập viên, bài báo sẽ được công bố lập tức trên mạng, và sau đó vài tháng trên

báo giấy. Dĩ nhiên, ngày nay có nhiều tập san chỉ công bố trực tuyến chứ không có báo giấy.

Đa số bài báo bị từ chối công bố

Như đề cập trên, đa số những bản thảo gửi đến ban biên tập bị từ chối. Tỷ lệ từ chối dao động rất lớn giữa các tập san. Những tập san có ảnh hưởng lớn như *New England Journal of Medicine, Science, Nature, Cell* từ chối khoảng 90-95% bản thảo. Nên nhớ rằng những bản thảo được gửi đến những tập san này đã rất chọn lọc. Những tập san đa có uy tín cao trong y khoa như *JAMA, BMJ, Nature Medicine, Annals of Internal Medicine, Journal of Clinical Investigation* có tỉ lệ từ chối bản thảo dao động trong khoảng 80 đến 90%. Những tập san hàng đầu trong chuyên ngành như nội tiết, xương khớp, ung thư, tim mạch, v.v. thì tỉ lệ từ chối cũng khoảng 70-80%. Tuy nhiên, những tập san có hệ số ảnh hưởng thấp có xu hướng chấp nhận bản thảo trên 50%, có khi lên đến 80%. Đây chính là lí do tại sao công bố trên những tập san có uy tín cao là một dấu ấn về đẳng cấp của nhà khoa học.

Bài báo khoa học bị từ chối ở 3 giai đoạn: ban biên tập, bình duyệt, và tái bình duyệt. Tỷ lệ từ chối rất khác nhau giữa 3 giai đoạn. Một nghiên cứu trên tập san *British Medical Journal* (một trong những tập san y khoa hàng đầu thế giới), trong giai đoạn 1 (tức ban biên tập), tỉ lệ từ chối khoảng 50% những bài báo gửi đến, mà không gửi ra ngoài để bình duyệt. Giai đoạn 2, sau khi gửi đi cho các chuyên gia bình duyệt, thì vẫn bị từ chối khoảng 45%. Ngay cả ở giai đoạn 3 là tái bình duyệt, xác suất bị từ chối là khoảng 5%.

Một trong những câu hỏi đặt ra là có sự 'kì thị' hay thành kiến với các tác giả từ các nước đang phát triển (như Việt Nam

chẳng hạn)? Bằng chứng khoa học cũng cho thấy quả thật có một sự kì thị trong tiềm thức đối với các tác giả từ các nước này. Trong một thí nghiệm thú vị, các tác giả chọn một số bài báo của các nhóm nghiên cứu nổi tiếng từ Mỹ, nhưng đối nhóm nghiên cứu thành những trung tâm ở các nước nghèo (Phi châu), và gửi các bài báo cho các nhà khoa học phương Tây. Kết quả cho thấy các bài báo không được đánh giá cao và cho điểm thấp hơn những bài báo có địa chỉ từ Mỹ. Đây là một trong nhiều chứng cứ cho thấy quả thật có sự “kì thị địa lí” trong khoa học.

Kinh nghiệm cá nhân của tôi cũng nhất quán với chứng cứ khoa học. Có những bài báo của đồng nghiệp từ trong nước được tôi giúp thiết kế nghiên cứu và biên tập, và đã được góp ý từ các đồng nghiệp có kinh nghiệm cao từ Mỹ, nhưng khi gửi cho tập san vẫn bị chê là... viết sai tiếng Anh! Dĩ nhiên, khi yêu cầu chỉ ra sai chỗ nào thì chuyên gia bình duyệt không nói. Ngược lại, có bản thảo từ Úc của nghiên cứu sinh tôi, với tiếng Anh có vài chỗ sai về văn phạm và ngữ vựng (cổ ý), nhưng tất cả chuyên gia bình duyệt không phản nài!

Tuy có sự kì thị trong tiềm thức, nhưng tỉ lệ từ chối không hẳn khác nhau giữa các quốc gia. Chẳng hạn như thống kê của các tập san y khoa lớn như *New England Journal of Medicine* và *JAMA* cho thấy tỉ lệ từ chối bài báo từ Mỹ và ngoài Mỹ không khác nhau. Tuy nhiên, đối với các tập san chuyên ngành thì có sự khác biệt lớn, chẳng hạn như tập san *Circulation Research* (chuyên về tim mạch), tỉ lệ từ chối chung là 85%, không khác mấy so với tỉ lệ từ chối các bài báo từ Hàn Quốc (88%), nhưng bài báo từ Trung Quốc có tỉ lệ từ chối lên đến 99% vì chất lượng quá kém và tiếng Anh chưa đạt. Một phân tích thú vị

khác của tập san *American Journal of Roentgenology* cho thấy trong thời gian 2003 – 2005, tập san này chấp nhận 72% những bài báo từ Mỹ, nhưng tỉ lệ này cho các nước ngoài Mỹ là 60%. Những chứng cứ thực tế trên đây cho thấy bài báo từ các nước nghèo hay kém phát triển thường có ‘nguy cơ’ bị từ chối cao hơn là những bài báo từ các nước tiên tiến.

Tại sao bài báo bị từ chối?

Lí do 1: Bài báo không thích hợp cho tập san. Đây là lí do ban biên tập từ chối rất nhanh. Tập san khoa học có đẳng cấp riêng trong chuyên ngành, và đẳng cấp này có thể phân biệt qua *hệ số ảnh hưởng* (impact factor, viết tắt là IF). Tập san có IF cao cũng có nghĩa là có ảnh hưởng lớn, và những tập san này chỉ công bố những công trình quan trọng. Nếu công trình nghiên cứu không phải thuộc vào loại “đột phá” thì không nên gửi cho các tập san như *Science* và *Nature*, mà nên xem xét đến các tập san chuyên ngành.

Lí do 2: Thiếu cái mới. Có thể nói rằng ban biên tập của bất cứ tập san nào cũng thích cái mới trong một bài báo. Cái mới có thể là mới về phương pháp, mới về cách tiếp cận, kết quả mới, cách diễn giải mới, v.v. Do đó, những công trình nghiên cứu không có cái mới, không có yếu tố ‘ngạc nhiên’, không làm cho người đọc hào hứng, thì khó có cơ hội công bố trên các tập san có tiếng. Những nghiên cứu mà câu trả lời hay kết quả chẳng ảnh hưởng gì đến chuyên ngành, chẳng gây tác động gì đến chính sách công hay thực hành lâm sàng cũng khó có cơ hội được công bố. Tầm quan trọng và thiếu cái mới trong công trình nghiên cứu là lí do hàng đầu (80% bài báo bị từ chối với lí do này).

Lí do 3: Phương pháp nghiên cứu. Phương pháp nghiên cứu có thể ví von như là bộ xương của bài báo; bộ xương cứng thì cơ thể mới có thể đứng vững. Một nghiên cứu trên các nhà khoa học dành giải Nobel và tổng biên tập cho thấy gần 3/4 bài báo khoa học bị từ chối vì lí do phương pháp không đạt. Phương pháp ở đây bao gồm những khiếm khuyết về cách tiếp cận, thiết kế nghiên cứu (như thiếu nhóm chứng, phương pháp đo lường chưa đạt, qui trình thực hiện thiếu tính hệ thống, phương pháp phân tích dữ liệu sai hay quá kém, cỡ mẫu không đủ). Trong giới biên tập người ta thường nói chỉ cần nhìn phần phương pháp là có thể đánh giá đẳng cấp của tác giả.

Lí do 4: Cách trình bày dữ liệu. Nội dung bài báo quan trọng nhưng nếu trình bày không tốt cũng dễ bị từ chối. Có 3 nguyên nhân chính dẫn đến bài báo bị từ chối đăng: trình bày dữ liệu không đầy đủ (32%), có mâu thuẫn giữa các dữ liệu trình bày (25%) và không cung cấp đầy đủ chi tiết về phương pháp nghiên cứu (25%).

Lí do 5: Tiếng Anh. Một nghiên cứu ở Đài Loan cho thấy khoảng 1/4 những bài báo trong ngành y học bị từ chối là do tiếng Anh không đạt. Nhiều tập san có thể thông cảm nếu văn phong tiếng Anh có sai sót có thể chỉnh sửa được, nhưng họ không thể chấp nhận bài báo có quá nhiều sai sót về tiếng Anh. Về cách viết, các nhà khoa học không ưa cách viết sử dụng từ ngữ hoa mỹ và sáo rỗng (ít thông tin), hoặc cách viết dùng những từ “đao to búa lớn” mà không có ý nghĩa cụ thể. Ngoài ra, diễn đạt ý tưởng không khúc chiết (21%) và câu văn thừa (11%) cũng là những nguyên nhân bị từ chối.

Trên đây là những kinh nghiệm và bài học của một người từng làm nghiên cứu sinh, và những trải nghiệm

của một người trong vai trò biên tập của vài tập san khoa học. Những kinh nghiệm này thật ra chẳng có gì mới, bởi vì những ai từng qua thời làm nghiên cứu sinh ở nước ngoài đều ít nhiều trải qua, chỉ có khác nhau là môi trường khoa học mà thôi. Nội dung và thông điệp của bài báo rất quan trọng, nhưng tiếng Anh cũng đóng vai trò quyết định, nhất là đối với những tác giả từ các nước đang phát triển. Nhưng tiếng Anh là một rào cản lớn của rất nhiều nhà khoa học Việt Nam. Không nên kì vọng rằng một nghiên cứu sinh mới xong chương trình tiến sĩ hay hậu tiến sĩ ở nước ngoài là có ngay khả năng độc lập nghiên cứu và soạn bài báo khoa học. Những kĩ năng này đòi hỏi thời gian và tương tác với những người có kinh nghiệm tốt.

Ở nước ta, hiện nay các nghiên cứu sinh và giáo sư chịu áp lực công bố khoa học. Theo qui định mới, một nghiên cứu sinh phải công bố ít nhất 2 bài báo khoa học để được bảo vệ luận án. Tôi nghĩ qui định này hợp lí (và đã có nhiều người, trong đó có tôi, kêu gọi trước đây). Tuy nhiên, vì tiếng Anh là rào cản lớn cho đa số nghiên cứu sinh và ngay cả người hướng dẫn luận án, nên yêu cầu 2 bài báo trong thực tế là một yêu cầu không thấp. Tôi nghĩ cần phải công bằng hơn bằng cách huấn luyện kĩ năng soạn bài báo khoa học cho tất cả các nghiên cứu sinh trước khi họ bắt tay vào nghiên cứu.

VẤN NẠN TẬP SAN KHOA HỌC “DORM”

Một trong những vấn đề lớn của công bố khoa học hiện nay là sự ra đời các tập san mạo danh khoa học. Nhiều nhà khoa học coi đây là sự “khủng hoảng” đang làm tổn hại đến tính liêm chính của nghiên cứu khoa học toàn cầu. Một số nhà khoa học Việt Nam cũng đã trở thành nạn nhân của kỹ nghệ công bố khoa học “dorm”. Trong bài viết, tác giả phân tích về sự ra đời, phát triển của những “tập san dorm” và các đặc điểm để nhận dạng. Qua đó góp phần vào việc nâng cao chất lượng công tác quản lý khoa học.

Tập san khoa học chính thống

“Sản phẩm” của nghiên cứu khoa học thường là các bài báo được công bố trên các tập san có bình duyệt (peer-reviewed journals) hay sẽ gọi ngắn là ‘tập san khoa học’. Các tập san này phải phân biệt với các tạp chí không có bình duyệt (magazines). Các tập san khoa học thường do các hiệp hội chuyên ngành điều hành, nhưng do một nhà xuất bản quản lí. Hội đồng biên tập của các tập san này thường là những nhà khoa học nổi tiếng trong chuyên ngành, và họ có nhiệm kì thường là 5 năm. Tập san khoa học có truyền thống từ thế kỉ 17. Có thể xem tập san *Philosophical Transactions of the Royal Society* (xuất bản lần đầu năm 1665) của Anh là tập san khoa học đầu tiên trên thế giới.

Bên cạnh các tập san của các hiệp hội chuyên môn, còn có các tập san do các tập đoàn xuất bản khoa học điều hành và quản lý. Các tập san này thường không có hội đồng biên tập, nhưng các bài báo thì do các nhà khoa học bên ngoài bình duyệt. Tiêu biểu cho nhóm này là các tập san trong ba tập đoàn xuất bản *Nature*, *Springer*, và *Elsevier*. (Chú ý, *Nature* sau này đã được *Springer* mua lại).

Hai nhóm tập san trên được xem là ‘chính thống’. Tính chính thống ở đây hiểu theo nghĩa được giới khoa học thế giới công nhận hoặc do các hiệp hội chuyên môn điều hành và vận hành theo qui trình chuẩn. Theo qui trình chuẩn, các bài báo phải qua bình duyệt bởi các chuyên gia và kiểm tra trước khi công bố. Tập san có tầm ảnh hưởng càng cao (như *New England Journal of Medicine*, *Nature*, *Science*, *Cell*) thì tỉ lệ từ chối bài báo càng cao, có khi lên đến 95%. Các tập san chuyên ngành khác thường có tỉ lệ từ chối dao động trong khoảng 10% đến 30%.

Bốn vai trò của tập san khoa học

Tập san khoa học đóng vai trò rất quan trọng trong nghiên cứu khoa học và duy trì tính liên chính của khoa học. Vai trò thứ nhất là diễn đàn để chuyển tải và chia sẻ thông tin khoa học, và qua đó vừa duy trì vừa phát triển khoa học. Tập san khoa học được xem là một diễn đàn, nơi mà các nhà khoa học công bố kết quả nghiên cứu, chia sẻ ý tưởng và phương pháp, từ thế hệ này sang thế hệ tiếp nối. Gerard Piel (người đầu tiên xuất bản tạp chí *Scientific American*) từng nói một câu chí lí rằng *nền khoa học sẽ chết nếu không có công bố* (“*Without publication, science is dead*”).

Vai trò thứ hai là cung cấp chứng từ khoa học. Các bài báo khoa học và kết quả nghiên cứu được trình bày trong đó được

xem là chứng từ khoa học để phục vụ cho các chính sách công, hay trong y khoa có thể dùng làm chứng cứ cho thực hành y học. Trong thực tế, các bài báo trên các tập san lớn như *New England Journal of Medicine* và *Lancet* có tác động làm thay đổi thực hành lâm sàng và chính sách y tế.

Vai trò thứ ba là người 'gác cổng', đảm bảo phẩm chất khoa học, không để cho những thông tin rác rưởi hay kém phẩm chất xuất hiện trên trang giấy. Tất cả các bài báo đều phải qua bình duyệt bởi các chuyên gia trong chuyên ngành để đảm bảo phẩm chất nghiên cứu đúng với chuẩn mực mà người trong chuyên ngành chấp nhận.

Vai trò thứ tư là 'lăng xê' các nhà khoa học hoặc nhóm nghiên cứu khoa học đến cộng đồng khoa học thế giới. Thật vậy, qua các vai trò diễn đàn và người gác cổng, các tập san khoa học có thể giới thiệu những nhà khoa học có triển vọng hay tạo điều kiện cho các nghiên cứu sinh mới vào chuyên ngành có dịp tiếp xúc với các đồng nghiệp trước đây. Phần lớn các giải thưởng danh giá đều được giới thiệu qua các công trình trên các tập san khoa học.

Trên thế giới ngày nay có khoảng 100.000 tập san như thế, và con số vẫn còn gia tăng. Tuy nhiên, trong số đó, chỉ có khoảng 12.000 đến 20.000 được đưa vào danh mục nổi tiếng như Clarivate Analytics (xưa kia là ISI của Thomson) và Scopus của tập đoàn xuất bản Elsevier. Trong số này, các tập san trong danh mục Clarivate Analytics được xem là có uy tín và tầm ảnh hưởng cao. Do đó, các nhà khoa học có uy tín và thuộc các trường đại học có tiếng chỉ công bố bài báo trên các tập san trong danh mục Clarivate Analytics. Ngay cả khi đánh giá năng suất khoa học của một quốc gia, người ta cũng thường dựa vào cơ sở dữ liệu của Clarivate Analytics.

Bối cảnh ra đời của tập san mở

Xuất bản khoa học là một kỹ nghệ làm tiền rất tuyệt vời vì lợi nhuận rất cao. Mô thức xuất bản khoa học đã tồn tại gần 400 năm nhưng vẫn không thay đổi. Theo mô thức này, nhà khoa học xin tài trợ, làm nghiên cứu, soạn bản thảo, và nộp cho một tập san. Tập san khoa học gửi bản thảo ra ngoài để bình duyệt, và dựa trên kết quả bình duyệt, ban biên tập quyết định chấp thuận hay từ chối bản thảo. Trong cái qui trình đó, nhà xuất bản hoàn toàn không làm gì cả! Họ không trả tiền cho ban biên tập và các chuyên gia bình duyệt, vì tất cả thành viên ban biên tập đều làm việc thiện nguyện. Khi bài báo được chấp nhận cho công bố, tác giả phải trả phí cho nhà xuất bản. Và, còn vô lý hơn nữa, bản quyền của bài báo thuộc về nhà xuất bản chứ không thuộc về tác giả! Điều này có nghĩa là ngay cả khi tác giả muốn in lại một bảng số liệu hay biểu đồ, tác giả phải xin phép nhà xuất bản, và trong vài trường hợp, phải tốn tiền để sử dụng các bảng biểu đó! Tỷ lệ lợi nhuận trên đầu tư của kỹ nghệ xuất bản khoa học còn cao hơn cả các tập đoàn như IBM và Apple.

Hầu như bất cứ ai cũng thấy sự vô lý của mô thức xuất bản khoa học truyền thống. Vào cuối thập niên 1990, một nhóm nhà khoa học Mỹ đứng ra kêu gọi thiết lập một mô thức công bố mới mà họ gọi là Open Access (Mở). Theo mô thức mới, nhà khoa học chỉ cần trả ấn phí cho nhà xuất bản chính thống, bản quyền bài báo đó thuộc về nhà khoa học, và bất cứ ai trên thế giới có thể truy cập được.

Một trong những nhóm thành công nhất là Public Library of Science, với hàng loạt tập san nổi tiếng như *PLoS Medicine*, *PLoS Biology*, *PLoS ONE*, v.v. Hàng loạt các nhà xuất bản khác

như *BioMedCentral* cũng theo mô thức của PLoS và rất thành công (sau này BioMedCentral bị Springer mua lại). Ngay cả các nhà xuất bản truyền thống cũng có lựa chọn Open Access cho tác giả. Có thể nói rằng mô thức xuất bản Mở đã làm một cuộc cách mạng trong xuất bản khoa học hiện đại.

Các tập san Mở như *PLoS ONE* (của Public Library of Science) và *Scientific Reports* (của tập đoàn Springer-Nature) mỗi năm công bố trên 30.000 bài báo khoa học. Những tập san này cũng có thu nhập và lợi nhuận rất cao. Nếu tính trung bình mỗi bài báo các nhóm xuất bản này lấy ấn phí 1.500 USD, thì mỗi năm chỉ một tập san có thể có thu nhập trên 45 triệu USD! Nhà xuất bản, như nói trên, không cần có nhiều nhân sự và chỉ cần đầu tư cho máy tính và công nghệ thông tin là có được một cơ sở xuất bản khoa học.

Kĩ nghệ xuất bản khoa học “dòm”

Chính vì siêu lợi nhuận từ mô thức xuất bản trực tuyến mà nhiều nhóm và nhiều cá nhân trên thế giới đã thành lập nhà xuất bản và tập san khoa học. Bên cạnh những nhà xuất bản chính thống, còn xuất hiện rất nhiều các cơ sở kinh doanh xuất bản dưới danh nghĩa “xuất bản khoa học” với những danh xưng cũng rất khoa học như “*Journal*”, “*Archive*”, “*Trends*”, “*Proceedings*”, v.v. Tất cả những trạm xuất bản này đều tự xưng là tập san Mở và tất cả đều in bài trực tuyến. Những tập san này không thuộc một nhà xuất bản có tiếng nào, cũng chẳng thuộc một hiệp hội chuyên môn nào điều hành. Bất cứ ai, chỉ cần có máy tính nối mạng và với kĩ thuật làm website, đều có thể tạo ra một hay nhiều ‘tập san khoa học’.

Nói thẳng ra, đây là những tập san “dòm”, đúng theo nghĩa của nó, tức là không có tính chất học thuật, mà chỉ là các cơ sở

làm tiến. Các trạm xuất bản này càng ngày càng biến hoá như vi khuẩn biến hoá, nên có khi rất khó phân biệt thật và giả.

Tập san khoa học dỏm, tiếng Anh là “predatory journal”, và thuật ngữ này do ông Jeffrey Beall (thủ thư của Đại học Colorado ở Denver) dùng lần đầu vào năm 2010, sau khi ông chú ý có nhiều email mời ông tham gia vào ban biên tập của những “tập san” mà ông chưa bao giờ nghe đến. Sau đó, ông Beall lập một trang blog nổi tiếng với danh sách các “nhà xuất bản” và “tập san” mà ông xếp vào nhóm *predatory* hoặc nghi ngờ thuộc vào nhóm *predatory*. Danh sách này đã giúp cho rất nhiều trường đại học và viện nghiên cứu trên thế giới phân biệt giữa thật và dỏm. Nhưng vào đầu năm 2017, ông Beall phải đóng cửa trang blog, vì những đe dọa liên tục từ các trạm xuất bản dỏm trên thế giới.

Kỹ nghệ xuất bản khoa học dỏm đã và đang làm vẩn đục khoa học, vì qui mô rất lớn của kỹ nghệ này. Trong một bài phân tích mới công bố trên *BMC Medicine*, các tác giả Phần Lan ước tính rằng hiện nay thế giới có gần 12.000 tập san dỏm, xuất bản bởi 996 “nhà xuất bản”. Con số này tương đương với số tập san khoa học chính thống.

Khoảng 27% những tập san dỏm này có trụ sở ở Ấn Độ, kể đến là các nước Á châu khác (12%), Bắc Mỹ (18%), Phi châu (5.5%), và không thể xác định địa điểm (27%). Cần nói thêm rằng mặc dù ‘văn phòng’ của những trạm này có thể chỉ là một nhà để xe hay một căn phòng nhỏ ở những nước vừa kể, nhưng họ rất ma mãnh vì khi gửi email mời tác giả nộp bài báo họ thường lấy địa chỉ văn phòng ở Mỹ hoặc các nước như Canada, Úc và Âu châu!

Chỉ trong năm 2014, các trạm xuất bản dỏm này đã công bố 420.000 bài báo khoa học. Đây là một sự tăng trưởng rất

nhANH, vì năm 2010 họ chỉ công bố được 53.000 bài. Khoảng 38% những bài báo thuộc vào nhóm “đa ngành”, kể đến là khoảng 23% thuộc vào ngành kĩ thuật (engineering), 17% là những bài liên quan đến y sinh học, 12% là kinh tế và khoa học xã hội. Đa số những tác giả của những bài báo này cũng tập trung ở Á châu. Gần 35% tác giả công bố trên các tập san dỏm là từ Ấn Độ. Phần còn lại là 26% từ các nước Á châu khác (ngoài Ấn Độ), Phi châu (16%), Bắc Mỹ (9%), Âu châu (9%).

Thời gian cần thiết để đăng một bài báo tính trung bình là khoảng 2 tháng, nhưng có không ít bài chỉ tốn 1 tuần. Các tập san chính thống có bình duyệt nghiêm chỉnh thì thời gian (nếu bình duyệt êm xuôi) từ lúc nộp đến công bố thường là 6 tháng đến 12 tháng; nếu bình duyệt khó khăn thì có thể kéo dài đến 2 năm.

Tính bình quân, chi phí xuất bản mà các trạm xuất bản dỏm lấy từ tác giả là 178 USD đến gần 800 USD. Những trạm xuất bản dỏm có nhiều tập san thường tính phí càng cao. Với ấn phí như thế, hai tác giả ước tính rằng thị trường tập san dỏm trên thế giới có giá trị 74 triệu USD. Còn các tập san Mở chính thống thì có thị trường khoảng 244 triệu USD.

Tiêu chuẩn để phân biệt tập san dỏm và chính thống

Đối với nhiều người chưa quen hay mới làm quen với công bố khoa học, việc phân biệt tập san chính thống và tập san dỏm rất khó khăn. Phần lớn những trạm xuất bản dỏm rất tinh ranh, họ đặt tên tập san nhái theo tập san chính thống để nâng cao tính chính thống của họ. Chẳng hạn như tập san chính thống là *Journal of Biological Chemistry* thì họ nhái là “*Journal of Biological Sciences*”! Ngoài ra, tiếng Anh của họ thì rất kém, sai sót về văn phạm và ngữ vựng rất nhiều. Tuy

nhiên, những ai từng có kinh nghiệm công bố quốc tế thì việc nhận dạng tập san dỏm không quá khó khăn. Tôi nghĩ đến vài nhóm tiêu chí giúp nhận dạng tập san dỏm: đặc điểm nhà xuất bản, ban biên tập, và các khía cạnh “linh tinh” khác.

Thứ nhất, tập san dỏm thường được xuất bản bởi những “nhà xuất bản” đáng ngờ, như chẳng có danh tiếng, không nằm trong hiệp hội xuất bản nào, không có địa chỉ đất (đường phố và thành phố rõ ràng) mà chỉ là trực tuyến, và thường có địa chỉ ở một nước phương Tây nhưng toàn công bố bài ở các nước đang phát triển. Có những nhà xuất bản có địa chỉ ở những nước như Trung Quốc, Ấn Độ, Ả Rập, Phi châu.

Thứ hai là ban biên tập lôm côm. Một tập san nghiêm chỉnh thường (không phải tất cả) là do các hiệp hội chuyên môn điều hành hay bảo trợ. Chẳng hạn như tập san *Osteoporosis International* là do Hội loãng xương quốc tế chủ trương, hay JAMA là của Hiệp hội Y khoa Hoa Kỳ sáng lập và xuất bản. Do đó, ban biên tập là các thành viên của hiệp hội, và họ chỉ phục vụ theo nhiệm kỳ. Ngược lại, các tập san dỏm thường chẳng có hiệp hội nào bảo trợ cả. Họ cũng có ban biên tập, nhưng thành viên ban biên tập là những người “vô danh”, hoặc không có địa chỉ cụ thể, hoặc chưa bao giờ công bố nghiên cứu trên các tập san có uy tín cao. Có tập san dỏm có tổng biên tập cũng là chủ nhà xuất bản! (Bởi vì “nhà xuất bản” chỉ có... 1 người).

Ngoài ra, một số tín hiệu chung cũng có thể nâng cao mức độ nghi ngờ tập san dỏm. Đó là những tín hiệu như:

- Tên tập san chung chung hay nhái theo tập san chính thống;
- Không có trong danh mục ISI, Scopus;
- Tổng biên tập không có thành tích khoa học tốt;

- Ban biên tập lờ mờ, toàn những người vô danh;
- Bài báo kém chất lượng; tiếng Anh sai nhiều.

Nhưng nguy hiểm nhất là những tập san nằm ở biên giới của dỏm và thật. Đây là những tập san không do các nhà xuất bản danh tiếng xuất bản nhưng có trong danh mục ISI do họ biết cách khai thác kẽ hở của tiêu chuẩn ISI. Những tập san loại này cũng có những người có tiếng trong ban biên tập, nhưng họ không công bố trên các tập san đó. Chính vì những đặc điểm này mà những người công bố trên các “tập san mờ mờ tỏ tỏ” có lí do để biện minh cho sự chính thống của tập san và bài báo của họ. Và, đây là những trường hợp làm cho giới quản lí khoa học rất nhức đầu để đối phó với những nguy hiểm.

Vấn nạn ở các nước phát triển

Sự có mặt của tập san dỏm có ý nghĩa quan trọng đến các nước đang phát triển như Việt Nam. Đối với những nước có nền khoa học lâu đời và đã định hình, thì giới khoa học dễ dàng phân biệt dỏm với thật, nên các tập san dỏm khó có thể gây tác động lớn. Nhưng với các nước đang phát triển, và áp lực công bố lớn, thì các tập san dỏm là một sự nan giải cho giới quản lí khoa học, vì họ thiếu kĩ năng và kinh nghiệm để phân biệt dỏm và thật. Trong thực tế, có khá nhiều nhà khoa học Việt Nam đã trở thành nạn nhân của kĩ nghệ xuất bản dỏm. Một số không biết mình là nạn nhân, nhưng một số khác thì có vẻ muốn trở thành “nạn nhân” của kĩ nghệ này. Người viết bài này đã từng đứng ra hoà giải cho một trường hợp ở trong nước: tác giả thì nói tập san là thật và đòi tiền thưởng, còn người quản lí thì nói là tập san dỏm và không cho tiền thưởng.

Do đó, nếu người thiếu kinh nghiệm khoa học đọc lí lịch với danh sách bài báo khoa học trên các tập san tiếng Anh, thì

rất khó có thể biết bài nào là công bố trên tập san chính thống, và bài nào trên tập san dỏm. Không phân biệt được thì dễ bị nhầm lẫn bởi những người làm khoa học dỏm. Trong thực tế tình trạng này đã xảy ra ở Việt Nam. Tôi nghĩ đã đến lúc các cơ quan quản lý khoa học đặt điều kiện về công bố quốc tế cho giới khoa học, với một danh sách những tập san cụ thể cần phải tránh và những tập san được công nhận. Làm được như thế sẽ góp phần giảm tình trạng ô nhiễm khoa học.

Ở trên, tôi đã nói về 4 vai trò của tập san khoa học chính thống, các tập san “dỏm” làm thay đổi ngược lại tất cả 4 vai trò đó:

- Thay vì đóng vai trò là diễn đàn truyền bá kiến thức, kỹ năng và công nghệ, các tập san dỏm chỉ quan tâm đến đồng tiền. Lí do đơn giản là vì họ không chịu trách nhiệm trước bất cứ hiệp hội chuyên khoa nào; họ chỉ là những nhóm kinh doanh thông tin khoa học dỏm.
- Thay vì đóng vai cung cấp chứng từ khoa học nhằm phục vụ cho chính sách công và thực hành khoa học, các tập san dỏm cung cấp những thông tin rác, làm sai lệch và xuyên tạc khoa học. Chẳng hạn như có những nhóm chống vaccine dùng thông tin từ các tập san dỏm để biện minh cho quan điểm [sai trái] rằng vaccine là có hại cho sức khỏe! Có thể nói rằng các tập san dỏm đã làm tác hại nguy hiểm đến khoa học.
- Thay vì làm người ‘gác cổng’ phẩm chất và duy trì chuẩn mực nghiên cứu khoa học, các tập san dỏm xiển dương những nghiên cứu kém chất lượng, thậm chí sai trái. Tuyệt đại đa số - nếu không muốn nói là tất cả - các bài báo trên các tập san dỏm chẳng có giá trị khoa học.

Ngay cả tiếng Anh cũng chẳng có một chuẩn mực nào cả, viết sai cỡ nào cũng được công bố. Các tập san này đã làm vẩn đục môi trường khoa học.

Sau cùng, thay vì giới thiệu những cá nhân xuất sắc, các tập san dỏm lại xiển dương những tác giả mà thành tích khoa học là một câu hỏi lớn. Những bài báo của những tác giả đó làm cho lẫn lộn giữa thật với giả, giữa chính thống và phi chính thống. Có những chuyện hài hước đã xảy ra khi các tác giả viết những điều mà chính họ không biết mình viết cái gì cũng được công bố! Có thể nói rằng với việc xiển dương cái dỏm, các tập san dỏm đã tạo nên một cộng đồng đối lập với khoa học chính thống và gây tác hại khôn lường đến xã hội nói chung.

Tiêu chí nhận dạng tập san dỏm

Nhận dạng tập san khoa học dỏm đã trở thành một nhu cầu quan trọng ở Việt Nam, vì có nhiều đồng nghiệp đã trở thành nạn nhân của những tập san dỏm. Các tiêu chuẩn dưới đây do ông Jeffrey Beall đề xướng, và tôi lược sau đây:

Tiêu chí 1: Biên tập và nhân sự

- Chủ nhà xuất bản cũng là tổng biên tập; hay tổng biên tập không phải là người có tiếng trong chuyên ngành;
- Không có cá nhân nào được liệt kê là tổng biên tập;
- Tập san không có ban biên tập hay nhóm chuyên gia bình duyệt;
- Không có thông tin về nơi công tác (affiliation) của các thành viên trong ban biên tập;
- Không có chứng cứ cho thấy ban biên tập hay tổng biên tập có chuyên môn hay tư cách chuyên môn để “gác

cổng” học thuật;

- Thành viên trong ban biên tập có tên trong 2 (hay hơn) tập san của cùng nhà xuất bản;
- Giả tạo danh sách thành viên ban biên tập, tức thành viên ảo, không có ngoài đời. Thỉnh thoảng có tập san đưa tên các nhà khoa học danh tiếng vào ban biên tập nhưng họ không hề hay biết.

Tiêu chí 2: Cơ sở xuất bản

- Thiếu minh bạch trong thương vụ xuất bản;
- Không có chính sách và qui định về “digital preservation”;
- Phụ thuộc vào ấn phí của tác giả trong việc điều hành nhà xuất bản;
- Khởi đầu với rất nhiều tập san;
- Không cung cấp đầy đủ thông tin, hay dấu điểm thông tin về ấn phí.

Tiêu chí 3: Liền chính

- Tên của tập san không nhất quán với sứ mệnh của tập san;
- Tên của tập san không phản ánh đầy đủ nguồn gốc của nó (ví dụ như tập san với tên “Canadian” hay “Swiss” trong tên chẳng có liên quan gì đến Canada và Thụy Sĩ);
- Tập san giả mạo rằng có impact factor, hay dùng vài chỉ số chưa được công nhận (như số views);
- Nhà xuất bản thường gửi spam email đến các nhà khoa học nhờ bình duyệt, và mời nộp bài;
- Nhà xuất bản nói dối rằng họ có trong danh bạ danh tiếng như ISI và Scopus;

- Nhà xuất bản không có biện pháp ngăn chặn các hành vi gian lận như đạo văn, tự đạo văn, đạo đức, v.v.;
- Nhà xuất bản yêu cầu tác giả chính để cử chuyên gia bình duyệt, và sử dụng các chuyên gia đó mà không xem qua thành tích khoa học của họ.

Tiêu chí 4: Những tín hiệu đáng ngờ

- Công bố những bài báo đã được công bố trên các tạp san khác mà không hề ghi nguồn;
- Dùng ngôn ngữ thậm xưng như “leading publisher” dù nhà xuất bản chỉ mới ra đời;
- Thường có địa chỉ ở các nước như Trung Quốc, Ấn Độ, Phi châu, hoặc có địa chỉ ở một nước phương Tây nhưng công bố bài ở các nước đang phát triển;
- Không biên tập bài báo, hay biên tập rất tối thiểu;
- Công bố những bài báo chẳng có gì là học thuật tính, nhưng luận văn phổ thông hay dành cho đại chúng;
- Có địa chỉ liên lạc “contact us” nhưng chỉ dưới hình thức trực tuyến mà không có địa chỉ email hay địa chỉ bưu điện.

Các cách làm dưới đây có thể xem là chuẩn mực thấp dù chưa hẳn là ở dạng “dỏm”, nhưng tác giả phải chú ý:

- Nhà xuất bản sao chép nguyên văn phần hướng dẫn cho tác giả (“authors guidelines”) của các nhà xuất bản nổi tiếng khác;
- Nhà xuất bản cung cấp không đầy đủ thông tin để liên lạc, như không cho biết tổng hành dinh ở đâu;

- Nhà xuất bản xuất bản những tập san có cái tên rất chung chung (như *Journal of Education*) để thu hút nhiều bản thảo và để tăng thu nhập;
- Nhà xuất bản xuất bản những tập san hỗn hợp giữa 2 hay nhiều lĩnh vực (như *International Journal of Business, Humanities and Technology*);
- Nhà xuất bản đòi tác giả chuyển nhượng bản quyền và giữ tác quyền về nội dung của tập san. Hoặc tập san đòi tác quyền ngay khi nộp bài báo;
- Nhà xuất bản không duy trì website tốt, hay website có những đường link bị gãy, sai tiếng Anh, v.v;
- Nhà xuất bản dùng các hình ảnh từ các website khác một cách bất hợp pháp;
- Nhà xuất bản gửi nhiều thư rác (spam email) đến các nhà khoa học để tìm tác giả, người bình duyệt, và thành viên ban biên tập;
- Nhà xuất bản dùng các email có địa chỉ như gmail.com, yahoo.com, hotmail.com, v.v;
- Nhà xuất bản không có chính sách về Open Access, hay hiểu sai nguyên lý của Open Access;
- Nhà xuất bản không có chính sách về rút lại bài báo, không có chính sách chỉnh sửa khi bài báo có sai sót;
- Nhà xuất bản không có số ISSN và DOI;
- Nhà xuất bản thường dùng những cái tên như Network, Center, Association, Institute, v.v;
- Nhà xuất bản có quá nhiều quảng cáo trên trang web và gây ảnh hưởng đến việc đọc thông tin trên trang web;

- Nhà xuất bản không phải là thành viên của hiệp hội nào để dễ kiểm tra;
- Nhà xuất bản thường hứa công bố nhanh và bình duyệt nhanh;
- Nhà xuất bản tập trung vào tác giả (chứ không phải độc giả) và ẩn phí;
- Họ không quan tâm đến chất lượng, mà chỉ quan tâm đến tiền;
- Nhà xuất bản thật ra chỉ là một cá nhân có thể có kinh nghiệm trong kinh doanh, nhưng không có kinh nghiệm gì trong xuất bản khoa học;
- Nhà xuất bản sao chép tên tập san từ các nhà xuất bản khác;
- Chẳng ai trong ban biên tập từng công bố trên các tập san trong ISI hay Scopus.

Nạn nhân của kỹ nghệ xuất bản dỏm

Câu chuyện bắt đầu từ một bác sĩ thuộc một đại học y khoa lớn ở Thái Lan nhận được thư chúc mừng rằng bài báo khoa học của chị đã được chấp nhận cho công bố. Người thầy của chị bác sĩ là Giáo sư P gửi email hỏi tôi về tính chính danh của tập san mà anh nghi ngờ, vì bài báo mới gửi đi chưa đầy 10 ngày mà đã được chấp nhận, một thời gian kỷ lục! Chỉ cần nhìn qua tên "tập san" vốn thuộc một nhóm xuất bản bên Ấn Độ, tôi có thể nói ngay rằng đó là một tập san "dỏm" mà tiếng Anh gọi là "Predatory Journal". Chị bác sĩ gửi email đến "nhà xuất bản" rút lại bài báo, nhưng họ không chịu và đòi phải trả 500 USD tiền ẩn phí, cho dù bài báo chưa được đăng. Đó là một cách kinh doanh bất chính của những cơ sở chuyên lợi dụng internet để làm tiền từ các nhà khoa học chủ yếu đến từ các nước đang phát triển.

Kĩ nghệ xuất bản dỏm không quan tâm đến nội dung

Trường hợp 1: Hai tác giả Peter Boghossian và James Lindsay viết một bài báo có tựa đề là "The conceptual penis as a social construct" (có lẽ tạm dịch là nhận thức về dương vật là một ý tưởng xã hội). Đó là một bài báo hoàn toàn vô nghĩa, nhưng dùng toàn những từ ngữ 'đao to búa lớn'. Trong phần tóm tắt, họ viết những câu bí hiểm và vô nghĩa như "Đứng trên phương diện cơ thể học, dương vật hiện hữu, nhưng những người đàn ông trước khi phẫu thuật chuyển giới thành phụ nữ cũng có dương vật. Do đó, dương vật như là một cơ phận nam tính là một ý niệm thiếu tính mạch lạc. Chúng tôi lí giải rằng nhận thức về dương vật có thể hiểu tốt hơn không phải là một cơ phận sinh lí mà là một ý niệm xã hội từ đồng hình đến hình tượng mạnh mẽ và độc hại của nam giới." Họ lấy tên ảo là Peter Boyle và Jamies Lindsay, và địa chỉ trung tâm nghiên cứu "Southeast Independent Social Research Group" cũng dỏm luôn. Khi bài báo được nộp cho một tạp san nằm giữa biên giới dỏm và thật, và tạp san chấp nhận cho công bố! Sau khi tạp san công bố, hai tác giả mới lên tiếng rằng bài báo của họ hoàn toàn vô nghĩa vì chính họ cũng không biết họ viết cái gì!

Trường hợp 2: Các tạp san khoa học dỏm thường gửi email rác đến các tác giả để quảng bá và mời đăng bài. Một nhà khoa học máy tính người Úc (Tiến sĩ Peter Vamplew) khó chịu với những đợt tấn công của các tạp san dỏm, nên ông đã viết một "bài báo khoa học" mà chỉ có 7 chữ "Get me off your fucking mailing list" (hãy bỏ tên tôi khỏi cái danh sách email chết tiệt của bạn) được lặp đi lặp lại suốt gần 10 trang giấy. Ông gửi "bài báo" cho một tạp san dỏm International Journal of Advanced Computer Technology, và ngạc nhiên thay chỉ vài ngày sau thì bài báo được chấp nhận cho công bố!

Get me off Your Fucking Mailing List

David Mazières and Eddie Kohler
New York University
University of California, Los Angeles
<http://www.mailavenger.org/>

t your fucking mailing list. Get me off
ing mailing list. Get me off your fu
our fucking mailing list. Get me off ing list. Get me off your fucking r
g mailing list. Get me off your fuck- Get me off your fucking mailing list.
list. Get me off your fucking mail- your fucking mailing list. Get me off

CÁCH ĐỌC BÀI BÁO KHOA HỌC

Sinh viên mới bắt đầu nghiên cứu phải làm quen với cách đọc bài báo khoa học. Những câu hỏi như đọc phần nào trước, và cần có kỹ năng gì để lĩnh hội nội dung của một bài báo khoa học. Tôi nghĩ câu trả lời tùy thuộc vào vai trò của người đọc hay vai trò của biên tập của tập san. Trong chương này tôi muốn chia sẻ kinh nghiệm cá nhân trong vai trò của người đọc để tham khảo.

Trong vai trò người đọc, tôi thường bắt đầu rất nhanh với tựa đề bài báo và sau đó đọc phần abstract (bản tóm tắt). Tựa đề bài báo, nếu được soạn thảo có suy nghĩ (nói theo tiếng Anh là “thoughtful”), thì có thể chuyển tải cái thông điệp chính của bài báo. Nhưng cũng không chắc lắm, vì một số tác giả có xu hướng đặt tựa đề trung dung, không nói lên cái thông điệp nào cả. Cũng có tác giả đặt tựa đề một cách giật gân, không nhất thiết phải tương đồng và nhất quán với dữ liệu thật.

Do đó, sau phần tựa đề, tôi phải đọc qua phần **abstract**, tức bản tóm tắt bài báo. Abstract là một “bức ảnh” của bài báo khoa học, hay cũng có thể xem là một phiên bản thu gọn của bài báo. Nếu abstract viết theo thể có tiêu đề theo cấu trúc (như *background, aim, methods, results, và conclusion*) thì rất dễ theo dõi. Nhưng có nhiều tập san thì yêu cầu tác giả viết abstract theo dạng không có tiêu đề, và trong trường hợp này thì hơi khó theo dõi. Nhưng dù là ở dạng nào thì thời gian đọc abstract cũng chỉ 10 phút là cao nhất. Vì thế, nhất định phải đọc abstract

trước. Vừa đọc vừa ghi ra những con số cụ thể, và so sánh với phần kết luận xem có nhất quán hay không. Có một số tác giả viết phần kết luận hơi “chênh” với dữ liệu thực tế. Chúng ta, người đọc, phải quyết định tin hay không tin vào kết luận dựa trên dữ liệu thật, chứ không phải dựa trên câu chữ của tác giả.

Sau phần abstract, tôi sẽ đọc các **bảng số liệu, biểu đồ, hình ảnh** (nếu có) trong bài báo. Bảng số liệu và biểu đồ là “trái tim” của bài báo khoa học. Tác giả có thể dùng chữ hoa mỹ, vắn vẹo từ ngữ, nhưng không thể nào vắn vẹo được dữ liệu. Nếu tác giả là người có kinh nghiệm tốt, thì mỗi biểu đồ và mỗi bảng số liệu đều nói lên một điểm. Chỉ cần nhìn vào một bảng số liệu hay một biểu đồ được thiết kế tốt, người đọc có thể biết được một thông điệp. Sau khi “lắp ráp” các thông điệp từ các bảng biểu, người đọc đã có một câu chuyện mà tác giả muốn chuyển tải đến người đọc.

Thỉnh thoảng, tôi viết ra những số liệu quan trọng, và nếu cần thiết tôi sẽ tự mình phân tích lại. Một số bài báo trình bày dữ liệu rõ ràng giúp cho người đọc có thể tự mình “đọc” bằng cách phân tích lại. Tuy đại đa số bài báo tác giả phân tích đúng, nhưng cũng có vài trường hợp mà kết quả của tôi không phù hợp với kết quả tác giả báo cáo, và những ca này làm tôi phải suy nghĩ và đặt câu hỏi tại sao. Có thể tác giả hiệu chỉnh mà tôi không có dữ liệu để làm, và như thế tôi có thể biết thêm một cách diễn giải khác. Nhưng nói chung, rất cần viết xuống những dữ liệu quan trọng để... nghiền ngẫm.

Sau khi đọc xong bảng biểu, tôi sẽ chú ý đến chi tiết trong phần **Methods** (phương pháp). Phần phương pháp có thể xem như cơ chế của bài báo khoa học. Trong phần này, tôi chú ý đến phương pháp đo lường nhất, vì giá trị khoa

học của công trình nghiên cứu là ở phần này. Phẩm chất của đo lường chính là yếu tố quyết định tính hợp lý nội tại của bài báo. Sau đó, tôi sẽ xem cách họ chọn đối tượng nghiên cứu theo tiêu chuẩn nào, vì phần này liên quan đến khái niệm “hợp lý ngoại tại”. Tôi không quan tâm mấy đến cái gọi là “tính đại diện”, nhưng tôi rất quan tâm đến tính “thông minh” trong thiết kế nghiên cứu. Một phần khác cũng quan trọng không kém là phương pháp phân tích dữ liệu, vì phần này quyết định một phần cái thông điệp của bài báo. Phương pháp phân tích là gì, biến outcome là gì, biến tiên lượng là gì, có hiệu chỉnh cho yếu tố nhiễu hay không, nếu có thì dùng phương pháp gì, dùng nhu liệu nào cho tính toán (nếu dùng Excel là xem như vứt đi, vì thiếu tính chuyên nghiệp). Đôi khi, chỉ cần đọc qua phần phương pháp và phân tích dữ liệu là biết đẳng cấp của tác giả ở đâu.

Sau cùng, tôi sẽ đọc phần **Discussion** (bàn luận). Thật ra, đọc phần này chỉ để biết quan điểm của tác giả là gì, và tìm những tài liệu tham khảo liên quan. Phần bàn luận phản ánh “tri thức” của tác giả thâm hậu cỡ nào; tác giả có kiến thức tốt viết phần này đọc rất hay, còn tác giả xoàng thì chỉ đọc vài dòng là biết họ chẳng có ý gì. Những bài báo quan trọng từ các labo nghiên cứu có tiếng thường được viết rất hay, và đây cũng là cách để tôi học cách viết, thậm chí học những từ ngữ hay của họ. Chú ý đến ý nghĩa của kết quả, và cách tác giả bàn về cơ chế của ảnh hưởng. Nếu phần bàn luận được viết tốt, phải có vài dòng đề cập đến những điểm mạnh và điểm yếu của nghiên cứu.

Đọc xong bài báo, người đọc phải trả lời 3 câu hỏi sau đây:

Câu hỏi 1: tôi có thể tin vào kết quả của bài báo này? Tin vào thông điệp nào?

Câu hỏi 2: nếu tin, tôi phải làm gì? có cần thay đổi thực hành?

Câu hỏi 3: tôi học được gì từ bài báo này, và nếu tôi phải làm lại, tôi sẽ làm như thế nào?

Câu hỏi thứ nhất có thể đánh giá qua trị số P, qua khoảng tin cậy 95%, và sự nhất quán giữa các kết quả. Nếu các kết quả nghiên cứu đều nói lên một “câu chuyện” nhất quán với nhiều nghiên cứu trước đây thì đó là kết quả đáng tin cậy. Đây chính là tính đại diện trong khoa học.

Câu hỏi thứ hai rất quan trọng, vì có nhiều công trình tuy kết quả tốt, nhưng chẳng gây tác động gì đáng kể. Nếu một can thiệp chỉ giảm nguy cơ gãy xương khoảng 5% (chỉ là ví dụ) thì mức độ ảnh hưởng quá nhỏ, khó có thể thay đổi thực hành lâm sàng (dù có ý nghĩa khoa học).

Câu hỏi thứ ba mang tính khoa học và phương pháp. Nếu phương pháp mà tác giả sử dụng tốt thì câu hỏi đặt ra là có phương pháp nào khác, có cách tiếp cận nào khác tốt hơn, rẻ tiền hơn, mà có thể cho ra kết quả với chất lượng cao. Một câu hỏi khác là nếu với dữ liệu trình bày, chúng ta có thể phân tích lại bằng một phương pháp khác (như Bayes chẳng hạn) và kết quả có làm thay đổi kết luận của tác giả.

Nói tóm lại, đọc một bài báo khoa học vừa là một nghệ thuật mà cũng là một khoa học. Nghệ thuật ở chỗ đặt thông điệp vào bối cảnh chung, cái đẹp của dữ liệu, và tri thức mới. Khoa học là so sánh và tính tái lập của kết quả nghiên cứu. Cũng như “văn ôn võ luyện”, đọc bài báo khoa học là một thói quen cần phải được luyện tập thường xuyên, cố gắng mỗi tuần phải đọc ít nhất một bài báo quan trọng và rút ra ít nhất một thông tin mới.

Một ví dụ tiêu biểu về đọc báo khoa học là bài tôi mới đọc hồi tuần vừa qua. Đó là bài trên JAMA về công trình RCT đánh giá ảnh hưởng của bổ sung calcium và vitamin D (viết tắt CaD) xem có giảm nguy cơ ung thư không (1). Đầu tiên tôi thấy tựa đề nằm trong radar quan tâm của mình, vì tôi đang nghiên cứu về CaD ở bệnh nhân loãng xương. Yếu tố thứ hai tôi quan tâm là bài báo được công bố trên JAMA và từ một nhóm tác giả tôi biết.

Đọc bản abstract, tôi chú ý ngay đến đoạn tác giả kết luận "among healthy postmenopausal older women... supplementation with vitamin D3 and calcium compared with placebo did not result in a significantly lower risk of all-type cancer at 4 years." (Tác giả kết luận rằng bổ sung CaD sau 4 năm không làm giảm đáng kể nguy cơ ung thư ở phụ nữ sau mãn kinh). Nhưng tôi rất nghi ngờ kết luận này. Lí do là vì trong bản abstract, tác giả báo cáo rằng bổ sung CaD giảm nguy cơ ung thư đến 30%. Nhưng vì tỉ số nguy cơ [relative risk] có khoảng tin cậy 95% dao động từ 0,47 đến 1,02, và trị số $P = 0,06$, nên tác giả nói rằng "not significantly"!

Đây là một kết luận tôi hay gọi là nỏ lệ vào thống kê. Chỉ vì P cao hơn 0,05 chỉ có 1% mà tác giả kết luận rằng không có ảnh hưởng! Trong khi đó số liệu cho thấy rõ ràng rằng bổ sung CaD có thể làm giảm nguy cơ ung thư đến 53%, mặc dù có một xác suất nhỏ là CaD có thể tăng nguy cơ 2%. Nhưng nhìn chung, lợi vẫn cao hơn "hại" rất rất nhiều.

Do đó, bước kế tiếp là tôi đọc các bảng số liệu và biểu đồ. Biểu đồ quan trọng nhất cho thấy quả thật bổ sung CaD giảm nguy cơ ung thư ngay từ năm thứ hai. Nhưng vì con số ca ung thư còn thấp, nên các phân tích thống kê có thể cho ra kết quả thiếu nhất quán. Đọc kĩ hơn nữa trong phần kết quả, tôi phát hiện hai số liệu quan trọng:

- (a) khi giới hạn phân tích trong nhóm người tuân thủ theo protocol thì kết quả có ý nghĩa thống kê ($P = 0,046$), và CaD giảm ung thư khoảng 35%.
- (b) Bệnh nhân ung thư có nồng độ vitamin D3 thấp hơn nhóm không bị ung thư.

Cả hai sự thật này rất nhất quán với giả thuyết về ảnh hưởng của CaD đến ung thư.

Tôi không cần đọc phần bàn luận vì một khi tác giả đã kết luận như thế, thì họ sẽ viết bàn luận theo kiểu định hướng. Và lại, bàn luận là phần mang tính chủ quan, tác giả có quyền suy luận và viết theo ý của họ. Nhưng để trả lời 3 câu hỏi trên, tôi đi đến kết luận:

Câu hỏi 1: tôi có thể tin vào kết quả của bài báo này? Tin. Đây là nghiên cứu RCT, có giá trị khoa học khá cao.

Câu hỏi 2: nếu tin, tôi phải làm gì? Trước mắt thì không làm gì cả. Nhưng rõ ràng tôi có chứng cứ để tin rằng bổ sung CaD là có ích cho người cao tuổi.

Câu hỏi 3: tôi học được gì từ bài báo này? Tôi học cái không nên làm, đó là kết luận chưa nhất quán với kết quả nghiên cứu.

KHOA HỌC VIỆT NAM ĐANG Ở ĐÂU

Theo lí thuyết về phát triển khoa học, nền khoa học của một quốc gia trải qua 4 giai đoạn phát triển: tiến phát triển, xây dựng, củng cố và mở rộng, và quốc tế hóa. Dựa trên những dữ liệu trắc lượng khoa học trong thời gian qua, có thể nói rằng khoa học Việt Nam vẫn còn đang trong giai đoạn thứ hai, tức xây dựng thực lực khoa học.

Nghiên cứu khoa học ở Việt Nam trong thời gian 20 năm qua đã có nhiều chuyển biến tích cực. Số lượng bài báo khoa học công bố trên những tập san trong danh mục ISI đã tăng nhanh trong thời gian gần đây, đạt khoảng 17% mỗi năm. Vào đầu thập niên 1990, số bài báo khoa học của Việt Nam trên các tập san ISI chỉ khoảng 300; đến đầu thập niên 2000, số bài báo công bố quốc tế cũng chỉ ~400; nhưng đến năm 2015 thì số công bố quốc tế đã xấp xỉ 3000 bài. Những dữ liệu này phản ánh *mức độ* hoạt động khoa học đang gia tăng nhanh.

Nhưng một khía cạnh khác của sự tăng trưởng ít khi nào được quan tâm đúng mức là hợp tác quốc tế. Hợp tác quốc tế trong nghiên cứu khoa học là qui luật chung và rất cần thiết cho các nước đang phát triển. Qua hợp tác quốc tế, giới khoa học có thể “gia nhập” mạng nghiên cứu quốc tế, chia sẻ tài

nguyên khoa học, và nâng cao chất lượng nghiên cứu. Vào thập niên 1990, đa số (80%) những công bố quốc tế từ Việt Nam có yếu tố hợp tác quốc tế. Đến nay, con số này có phần giảm nhưng vẫn xấp xỉ 77%. Mức độ hợp tác quốc tế cũng phản ánh quá trình phát triển khoa học của một quốc gia.

Bốn giai đoạn phát triển

Theo cách nhìn của lí thuyết phát triển, nền khoa học của một quốc gia trải qua 4 giai đoạn phát triển (1), và mỗi giai đoạn được định hình bởi một vài chỉ số về trắc lượng khoa học (scientometrics) và hợp tác quốc tế:

Trong giai đoạn đầu, *tiền phát triển*, mức độ hoạt động nghiên cứu khoa học còn thấp, vì nghiên cứu chỉ được thực hiện bởi một nhóm nhỏ nhà khoa học. Trong giai đoạn tiền phát triển, cơ chế tài trợ cho nghiên cứu còn nhiều bất cập và chưa định hướng nghiên cứu rõ ràng. Hai đặc điểm chính của giai đoạn tiền phát triển là số lượng bài báo khoa học thấp và dao động nhiều giữa các năm, và rất ít hợp tác quốc tế trong giai đoạn tiền phát triển.

Giai đoạn 2 được xem là thời gian *xây dựng*. Trong giai đoạn này, số lượng bài báo gia tăng theo thời gian, và hướng nghiên cứu cũng bắt đầu định hình. Các nhà khoa học bắt đầu hội nhập thế giới, qua thiết lập những mối liên hệ với đồng nghiệp nước ngoài, phần lớn là từ các nước đã phát triển. Do đó, mặc dù số lượng bài báo tăng nhanh, nhưng đa số những bài báo này là do hợp tác quốc tế. Trong hợp tác này, vai trò của nhà khoa học địa phương còn khiêm tốn, thường là vai trò thứ phát, vì người chủ trì dự án nghiên cứu thường là người nước ngoài.

Giai đoạn 3 được xem là giai đoạn *củng cố và mở rộng*. Trong thời gian này, quốc gia đã có được cơ sở vật chất cho nghiên cứu khoa học, và năng lực khoa học đã tạm vững vàng. Tài trợ cho nghiên cứu khoa học bắt đầu tăng nhanh và có chương trình rõ ràng hơn. Các tập san khoa học nội địa cũng bắt đầu quốc tế hóa và có xác suất cao được chấp nhận trong danh mục ISI hay Scopus. Số bài báo khoa học tiếp tục tăng nhanh, số bài báo hợp tác quốc tế cũng tăng, nhưng tỉ trọng bài báo hợp tác quốc tế bắt đầu giảm dần

Giai đoạn sau cùng là *quốc tế hóa*. Trong giai đoạn này, năng lực nghiên cứu khoa học được mở rộng, các trung tâm nghiên cứu bắt đầu “trưởng thành” và có thể đóng vai trò lãnh đạo trong các chương trình hợp tác quốc tế. Số bài báo khoa học vẫn tăng, nhưng chất lượng khoa học cũng cải tiến, và nghiên cứu bắt đầu có tác động trên trường quốc tế. Hợp tác quốc tế trong giai đoạn này không cao như giai đoạn xây dựng, và các nhà khoa học nội địa là những người đứng đầu trong các dự án hợp tác khoa học.

Những chuyển biến về công bố quốc tế trong vùng

Trong thời gian 15 năm (2001 - 2015), các nhà khoa học Việt Nam đã công bố được 18.076 bài báo khoa học trên các tập san trong danh mục ISI (2). So sánh với các nước trong vùng trong cùng thời kì, số bài báo khoa học của Việt Nam đã vượt qua Indonesia và Philippines, nhưng vẫn còn thấp hơn so với Thái Lan, Malaysia và Singapore. Thật vậy, số bài báo khoa học của Việt Nam chỉ mới bằng 28% của Thái Lan, 25% của Malaysia, và 15% của Singapore.

Một xu hướng đáng chú ý là Malaysia đã bắt đầu vượt qua Singapore về số lượng bài báo khoa học công bố trên các tập san trong danh mục ISI. Năm 2015, Malaysia công bố được

12.341 bài, thấp hơn Singapore (13.631 bài). Nhưng năm 2016, số công bố quốc tế của Malaysia (14.129) đã bằng, thậm chí cao hơn, Singapore (14.120).

Ở mức độ thấp hơn, Indonesia trước đây có số công bố quốc tế thấp hơn Việt Nam, nhưng trong hai năm qua có xu hướng tăng nhanh và có thể vượt qua Việt Nam. Trong năm 2014, số công bố quốc tế của Indonesia là 1795, thấp hơn Việt Nam (2596). Nhưng đến năm 2016 thì con số công bố quốc tế của Indonesia (3748) đã xấp xỉ bằng Việt Nam (3814). Những xu hướng này cho thấy nghiên cứu khoa học ở Đông Nam Á đang có chuyển biến tích cực và thú vị, và Malaysia có lẽ sẽ là nước có nhiều công bố quốc tế nhất trong vùng.

Các nước khác như Brunei, Cam-pu-chia, Lào và Myanmar cũng có công bố quốc tế, nhưng số lượng còn thấp. Tổng số công bố quốc tế từ 4 nước này chỉ đóng góp ~2% tổng số bài báo khoa học từ ASEAN. Đa số những công bố quốc tế từ 4 nước vừa kể chủ yếu là do hợp tác quốc tế.

Hợp tác quốc tế trong nghiên cứu khoa học

Một khía cạnh quan trọng trong nghiên cứu khoa học là “nội lực”. Một cách để biết được nội lực của khoa học của một quốc gia, người ta hay dùng chỉ số về hợp tác quốc tế trong nghiên cứu khoa học. Ở những nước mà nền khoa học ở vào giai đoạn III và IV thì tỉ lệ hợp tác quốc tế thường ở mức 50%, và phần lớn họ đóng vai trò chủ đạo trong các dự án nghiên cứu.

Đối chiếu lại tình hình ở Việt Nam trong thời gian qua, có một số xu hướng đáng quan tâm. Trong một nghiên cứu trước đây, chúng tôi chỉ ra rằng trong thời gian 1991 - 2010, trong số 10.618 bài báo khoa học từ Việt Nam, thì số bài báo có hợp tác quốc tế chiếm gần 80% (2). Tỉ lệ hợp tác quốc tế trong cùng

thời gian ở các nước như Thái Lan và Malaysia dao động trong khoảng 55 đến 65%.

Trong thời gian 2001 đến 2015, chúng tôi phân tích một lần nữa và tỉ lệ hợp tác quốc tế vẫn chưa thay đổi đáng kể (3). Tính chung, trong số 18.076 bài báo khoa học trên các tập san ISI, 77% công trình là có hợp tác quốc tế. Nói cách khác, tỉ trọng nội lực chỉ 23%. Tỉ lệ hợp tác quốc tế trong các bài báo khoa học của Thái Lan và Malaysia là 56%.

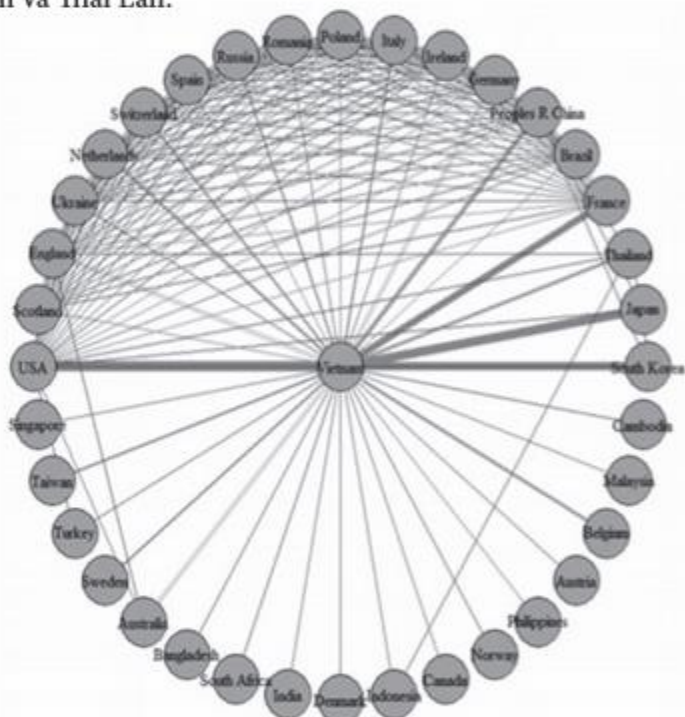
Tuy nhiên, có sự khác biệt về mức độ hợp tác quốc tế giữa các lĩnh vực nghiên cứu (**Bảng 1**). Các lĩnh vực có tỉ lệ hợp tác quốc tế cao nhất (trên 85%) là y tế công cộng, y học lâm sàng, y sinh học, và khoa học xã hội. Nhưng các lĩnh vực như hoá học và kĩ thuật cũng có tỉ lệ hợp tác quốc tế khá cao.

Bảng 1:

Tỉ trọng (%) hợp tác quốc tế trong các bài báo khoa học của Việt Nam (2001 - 2015)

Lĩnh vực nghiên cứu	2001 - 2005	2006 - 2010	2011 - 2015	Tính chung 2001 - 2015
Tất cả các lĩnh vực	79,2	78,7	76,1	77,0
Toán học	40,0	42,5	42,5	41,2
Vật lí	71,7	68,3	68,4	68,8
Hoá học	89,8	83,2	75,5	78,9
Kĩ thuật và công nghệ	76,8	77,2	73,1	74,3
Khoa học trái đất	88,8	91,3	90,5	90,2
Y học lâm sàng	91,8	90,7	89,6	90,6
Y tế công cộng	95,4	92,5	92,2	93,8
Y sinh học	95,7	94,6	85,3	89,4
Khoa học xã hội	78,9	75,7	78,9	90,4
Kinh tế	63,3	70,7	67,5	67,9

Phân tích chi tiết hơn cho thấy các nhà khoa học Việt Nam hợp tác với đồng nghiệp của hơn 150 quốc gia trên thế giới. Trong nhiều trường hợp, nhiều nhà khoa học của 5 nước cùng hợp tác trong một dự án, và có vài trường hợp nghiên cứu vật lý có hơn 100 quốc gia tham gia (kể cả Việt Nam). Nhìn chung, Việt Nam có xu hướng hợp tác với các nước phương Tây (Hình 1). Mười nước Việt Nam có hợp tác nhiều nhất là Mi Nhật, Hàn Quốc, Pháp, Anh, Úc, Đức, Trung Quốc, Hà Lan và Thái Lan.



Hình 1:

Hợp tác quốc tế trong nghiên cứu khoa học của Việt Nam trong thời gian 2001 - 2015. Đường nối càng dày thể hiện số bài báo khoa học với tác giả của hai nước càng cao. Mạng lưới phần trên của hình cho thấy Việt Nam chủ yếu hợp tác nghiên cứu khoa học với các nước phương Tây.

Hợp tác quốc tế có vẻ tăng chất lượng nghiên cứu. Hầu hết trong các lĩnh vực nghiên cứu, bài báo có “yếu tố quốc tế” đều có chỉ số trích dẫn cao hơn bài báo thuần Việt. Chẳng hạn như trong lĩnh vực y học lâm sàng, chỉ số trích dẫn của các bài báo thuần Việt là 7,74, nhưng bài báo có hợp tác quốc tế có chỉ số trích dẫn cao hơn 3 lần. Cần nhắc lại rằng 91% bài báo y học lâm sàng là có hợp tác quốc tế. Ngay cả trong lĩnh vực toán học, bài báo có hợp tác quốc tế vẫn có chỉ số trích dẫn cao gấp 2,4 lần so với bài báo thuần Việt. Đặc biệt trong lĩnh vực kinh tế, các bài báo có hợp tác quốc tế (chiếm 68% tổng số) có chỉ số trích dẫn cao gấp 4,5 lần so với bài báo thuần Việt.

Một xu hướng thú vị là các công trình có hợp tác quốc tế có chỉ số trích dẫn cao và được công bố trên các tạp san có chỉ số tác động cao. Xu hướng này thật ra cũng phù hợp và từng được ghi nhận ở các nước đang phát triển. Khi phân tích chỉ số trích dẫn theo hợp tác quốc tế và vai trò của tác giả, một xu hướng rất thú vị khác xuất hiện. Những công trình có hợp tác quốc tế và tác giả chính là người nước ngoài có chỉ số trích dẫn cao nhất. Kế đến là những công trình hợp tác quốc tế nhưng tác giả chính là người Việt Nam. Những công trình không có hợp tác quốc tế (tức “thuần Việt”) có chỉ số trích dẫn thấp nhất. Xu hướng này hiện hữu gần như trong tất cả các lĩnh vực nghiên cứu.

Tuy nhiên, hợp tác khoa học cũng có khi phải trả một cái giá về tự chủ. Các chuyên gia quản lý khoa học thường xem một tỉ lệ hợp tác cỡ 80% hay cao hơn là “lệ thuộc”. Nếu dùng tiêu chuẩn này, kết quả phân tích trong bài này cho thấy đa số các lĩnh vực nghiên cứu của Việt Nam (như Hoá học, Khoa học Trái Đất, Y học lâm sàng, Y sinh học, Y tế công cộng) có thể xem là đang trong tình trạng lệ thuộc.

Việt Nam đang ở đâu trong quá trình phát triển khoa học?

Những dữ liệu về công bố quốc tế và xu hướng hợp tác quốc tế cung cấp một bức tranh chung về quá trình phát triển khoa học. Các nước như Cam-pu-chia, Lào, Brunei và Miến Điện đang còn trong giai đoạn tiến phát triển, vì số công bố quốc tế còn thấp và chủ yếu là do hợp tác quốc tế.

Giai đoạn I Tiến phát triển	Giai đoạn II Xây dựng	Giai đoạn III Củng cố & mở rộng	Giai đoạn IV Quốc tế hóa
Cam-pu-chia	Việt Nam	Thái Lan	Singapore
Lào	Indonesia	Malaysia	
Miến Điện	Philippines		
Brunei			

Hình 2:

Bốn giai đoạn phát triển khoa học của các nước trong vùng Đông Nam Á dựa trên phân tích trắc lượng khoa học.

Việt Nam, Philippines và Indonesia đã bước ra khỏi giai đoạn tiến phát triển và đang ở trong giai đoạn II, tức giai đoạn xây dựng. Đó là giai đoạn với đặc điểm chính là nghiên cứu khoa học tăng nhanh phản ánh qua số công bố quốc tế, và tỉ trọng hợp tác quốc tế rất cao (trên 70%) và các nhà khoa học nước ngoài đóng vai trò chủ trì.

Dữ liệu trắc lượng khoa học cho thấy Thái Lan và Malaysia đang ở trong giai đoạn củng cố và mở rộng (giai đoạn III). Số bài báo khoa học của hai nước này cao gấp 3-4 lần số bài báo của Việt Nam. Tỉ trọng hợp tác quốc tế của Thái Lan và Mã Lai trong 20 năm qua dao động trong khoảng 50 đến 60%, và đa số công trình là do giới khoa học nội địa chủ trì. Ngoài ra, các đại học của Thái Lan và Malaysia đang thu hút nghiên cứu sinh từ nước ngoài (kể cả từ Việt Nam) và họ mở rộng hợp tác nghiên cứu khoa học sang các nước lân cận như Lào, Cam-pu-chia và Miến Điện.

Trong vùng ASEAN, chỉ có Singapore là ở trong giai đoạn IV, tức quốc tế hóa. Mặc dù số bài báo khoa học năm 2016 của Singapore xấp xỉ con số của Malaysia, nhưng chất lượng nghiên cứu của Singapore cao hơn Malaysia. Khoảng 60% các nghiên cứu khoa học từ Singapore có yếu tố hợp tác quốc tế, nhưng các nhà khoa học Singapore chủ trì trong các dự án này. Singapore không chỉ thu hút nghiên cứu sinh từ nước ngoài, mà còn tích cực tài trợ cho một số dự án quốc tế. Do đó, nền khoa học Singapore đã ở mức độ quốc tế hóa.

Quay lại tình trạng của Việt Nam, một điều đáng quan tâm là thời gian xây dựng đã diễn ra hơn 25 năm. Hơn 25 năm qua, tỉ trọng hợp tác quốc tế của Việt Nam vẫn còn trong mức «lệ thuộc», tức 75-80%. Trong khi đó, ở các nước như Thái Lan, thời gian xây dựng chỉ khoảng 20 năm, hay như Malaysia thời gian thậm chí còn ngắn hơn (~15 năm). Hơn 25 năm, nội lực khoa học Việt Nam vẫn chưa phát triển theo nhịp độ các nước trong vùng. Những kết quả này cũng có thể làm nền tảng để chúng ta phải suy nghĩ lại về chiến lược nghiên cứu khoa học và tổ chức hoạt động khoa học cấp vi mô và vĩ mô để đạt được một sự tăng trưởng tốt hơn và bền vững hơn trong tương lai.

Tham khảo:

- (1) Moed HF and Halevi G. Tracking scientific development and collaborations – The case of 25 Asian countries. Research Trends (2014), September 2014.
- (2) Nguyen TV, Pham LT. Scientific output and its relationship to knowledge economy: an analysis of ASEAN countries. Scientometrics (2011) 89:107–117
- (3) Nguyen TV, Ho-Le TP, Le UV. International collaboration in scientific research in Vietnam: an analysis of patterns and impact. Scientometrics (2017) 110:1035–1051

NHỮNG VẤN ĐỀ THỜI SỰ

Nếu ai hỏi tôi vấn đề lớn nhất của khoa học hiện nay là gì, tôi sẽ nói đó là vấn đề “reproducibility”, mà tôi dịch là “tái lập”. Rất nhiều kết quả nghiên cứu không được lặp lại, một phần là do sai sót, một phần nhỏ là do gian dối. Theo một phân tích công bố trên PLoS Biology thì vấn đề bất lập (irreproducibility) gây lãng phí lên đến 28 tỉ USD mỗi năm, và con số này chỉ tính riêng cho Mỹ (1). Đầu tháng 9 năm 2016, Trường Y Đại học Sydney mời tôi nói một bài về vấn đề tái lập, và chương này tóm tắt những ý trong bài nói chuyện đó để chia sẻ cùng các bạn.

Tình trạng bất lập (irreproducibility)

Khái niệm tái lập rất quan trọng trong khoa học, vì nó là yếu tố định hình khoa học và phân biệt giữa khoa học thật và khoa học dỏm. Dù khái niệm *reproducibility* vẫn còn trong vòng tranh luận, nhưng đại khái giới khoa học đồng ý với **định nghĩa** chung là: khả năng một nhà nghiên cứu có thể lặp lại kết quả nghiên cứu trước đây với cùng một phương pháp và qui trình được sử dụng bởi nghiên cứu gốc (“*The ability of a researcher to duplicate the results of a prior study using the same materials... as were used by the original investigator*”). Định nghĩa này cũng có thể mở rộng cho khái niệm *repeatability*, *replicability*, và *robustness*. Cái khác biệt quan trọng giữa khoa học thật và khoa học dỏm là ở khả năng tái lập.

Ngược lại với tái lập là bất lập. Nếu một nghiên cứu đầu tiên cho ra kết quả rằng gene A có liên quan đến tử vong, và

nếu kết quả nghiên cứu đó được các nhóm khác lặp lại với cùng phương pháp thì phát hiện đó được xem là mang tính khoa học. Nhưng nếu kết quả đó người khác không lặp lại được thì kết quả đó được xem là phi khoa học. Tương tự, nếu một ý tưởng (hay một chủ nghĩa) mà người ta tuyên bố rằng nó mang tính khoa học, thì cái outcome của ý tưởng hay chủ nghĩa đó phải mang tính tái lập; nếu không thì đó là chủ nghĩa không mang tính khoa học hay khoa học dỏm.

Trong một bài luận rất quan trọng và có ảnh hưởng rất lớn trong khoa học, John Ioannidis “chứng minh” rằng đa số các nghiên cứu được công bố là sai. Nguyên văn là *“It can be proven that most claimed research findings are false”* (2). Và, thực tế cho thấy đúng như thế: phần lớn những kết quả nghiên cứu khoa học, từ khoa học tự nhiên đến khoa học xã hội, đều sai hoặc không thể lặp lại một cách độc lập. ***Đây chính là một khủng hoảng trong khoa học mà trước đây ít người để cập đến.***

Chẳng hạn như một bài tổng quan, các tác giả cho thấy trong số 20 nghiên cứu quan sát với nhiều kết quả rất thuyết phục và hứa hẹn, nhưng chỉ có 1 nghiên cứu duy nhất có thể lặp lại những kết quả ban đầu đó. Công ti dược Bayer Healthcare điểm qua 67 nghiên cứu được công bố trên các tạp san số 1 như Science, Cell, Nature, nhưng họ không thể lặp lại 3/4 các kết quả. Ngay cả nghiên cứu lâm sàng đối chứng ngẫu nhiên (RCT) cũng có vấn đề. Trong số 49 RCT nổi tiếng (trong thời gian 1999 - 2003), chỉ có 11 RCT mà kết quả chưa bị “thách thức”, phần còn lại 76% đều không lặp lại được hay kết quả ngược lại với kết quả ban đầu. Nghiên cứu về di truyền học và tiên lượng bệnh còn tệ hại hơn: trong số 7 nghiên cứu rất triển vọng trong ung thư học, khi người ta lặp lại thì 5 kết quả chẳng tốt hơn ngẫu nhiên! Các nghiên cứu về tế bào gốc có lẽ là tệ nhất vì hầu hết kết quả đều không tái lập được.

Tình trạng bất lập này rất nghiêm trọng trong các lĩnh vực “nóng” như di truyền học, miễn dịch học, thần kinh học, và tâm lý học thực nghiệm. *Tình trạng này đã được rất nhiều báo chí đại chúng nêu thành vấn đề hay đúng hơn là vấn nạn khoa học. Từ những tờ nổi tiếng như New York Times, Washington Post, Economist, đến Atlantic, Wall Street Journal, Guardian* đều có những phân tích sâu về tình trạng bất lập trong khoa học. Cá nhân tôi cũng được nhiều nơi mời nói về tình trạng này. Có nhiều nhà báo ác ý mỉa mai rằng giới khoa học chỉ làm tổn tiền công chúng, và đề nghị phải giám sát chặt chẽ hơn việc phân bổ ngân sách cho nghiên cứu.

Tại sao bất lập?

Đây là câu hỏi đắt giá và đòi hỏi nhiều suy nghĩ. Riêng tôi thì qua kinh nghiệm cá nhân, tôi nghĩ đến 7 lý do, và 7 lý do này có thể chia thành 2 nhóm. Nhóm đầu tiên liên quan đến vấn đề *bias*, nhóm thứ hai liên quan đến thống kê. *Bias* ở đây được hiểu là lệch so với sự thật. Nhóm *bias* bao gồm publication bias, experimental bias, và thường cho kết quả dương tính. Nhóm lý do thống kê bao gồm ứng dụng sai phương pháp, yếu tố nhiễu, dùng trị số P tùy tiện, và vấn đề kiểm định nhiều giả thuyết.

1. Bias trong công bố

Đây là một hiện tượng rất phổ biến trong khoa học. Thường thường thì các nghiên cứu có kết quả “dương tính” dễ được công bố hơn các kết quả âm tính. Ngay cả nhà khoa học cũng không thiết tha gì với việc công bố một kết quả không có ý nghĩa thống kê. Tình trạng *bias* này dần dần tích lũy thành một sự lệch lạc trong khoa học, làm cho khoa học càng ngày càng một chiều.

Một phân tích thú vị của Fanelli trên PLoS ONE (2010)

cho thấy hầu hết (khoảng 80-90%) các kết quả công bố đều dương tính. Hiện tượng này xuất hiện trong tất cả các ngành, từ thần kinh học, nghiên cứu lâm sàng, đến sinh thái học, hầu hết kết quả đều nhất quán với giả thuyết của tác giả. Tỷ lệ kết quả dương tính 80-90% phải nói là một ngạc nhiên, vì phương pháp khoa học không thể nào tuyệt vời như thế. Đó chính là một tín hiệu cho thấy lệch lạc trong công bố kết quả.

2. Bias trong thí nghiệm

Rất nhiều kết quả nghiên cứu không đáng tin cậy và thiếu tính chính xác. Một phần là do sai lầm trong thiết kế, chọn sai đối tượng, hay phương pháp đo lường chưa được hoàn hảo. Có nhiều nghiên cứu không ngẫu nhiên hoá nên dẫn đến kết quả sai lệch. Một nghiên cứu trên chuột cho thấy ngay cả giới tính của người tiếp xúc với chuột cũng gây ra kết quả khác vì chuột đực như nhận được người sờ chúng là nam hay nữ, và do đó phản ứng rất khác nhau! Tất cả những yếu tố này dẫn đến lệch lạc trong kết quả nghiên cứu, và khi người khác cố gắng lặp lại kết quả thì thường thất bại.

3. Kết quả dương tính

Hiện tượng “the winner takes it all” rất phổ biến trong khoa học. Nghiên cứu cho ra kết quả dương tính sẽ giúp cho nhà khoa học có cơ hội công bố trên những tập san tốt. Công bố trên các tập san tốt giúp nhà khoa học xin được tài trợ; tài trợ dẫn đến cơ hội được đề bạt. Với chức danh càng cao, việc xin tài trợ càng dễ. Nói tóm lại, kết quả dương tính là một khởi đầu cho vòng tròn khoa học có lợi. Chính vì lý do này mà giới khoa học có xu hướng làm cho kết quả dương tính, hay “suppress” các kết quả âm tính. Xu hướng này dĩ nhiên dẫn đến sự lệch lạc về tri thức khoa học.

4. Ứng dụng sai phương pháp thống kê

Phân tích thống kê rất quan trọng cho các nghiên cứu khoa học thực nghiệm, vì đó là một phần quyết định thành bại của một nghiên cứu. Nhưng trong thực tế thì do thiếu hiểu biết và kinh nghiệm, nên rất rất nhiều nhà khoa học dùng sai phương pháp, hoặc ứng dụng mô hình phân tích sai bối cảnh, hoặc không kiểm tra giả định, nên phạm rất nhiều sai sót trong phân tích dữ liệu. Tỷ lệ sai sót về ứng dụng thống kê thường dao động trong khoảng 25-50% trong các bài báo khoa học.

Một sai sót “ngao mạn” gần đây là công trình nghiên cứu RCT diệt sán lãi ở Phi châu. Một nhóm nhà kinh tế thực hiện một nghiên cứu RCT ở Kenya, và họ báo cáo kết quả cho thấy diệt giun sán có thể giúp học sinh khoẻ mạnh hơn và học tập tốt hơn, một kết quả rất... có lí. Đến khi một nhóm chuyên gia dịch tễ học ở London phân tích lại dữ liệu thì họ phát hiện các nhà kinh tế phân tích sai, chẳng những sai phương pháp mà còn sai về mã hoá, và khi phân tích đúng thì kết quả ngược lại với kết quả đã công bố. Đây là một trường hợp hơi cực đoan, nhưng nó cho thấy ứng dụng sai phương pháp rất dễ dẫn đến kết luận sai.

5. Yếu tố nhiễu

Bất cứ nghiên cứu nào, nhất là nghiên cứu quan sát, cũng có yếu tố nhiễu. Yếu tố nhiễu là một trong những yếu tố làm nhức đầu cho giới khoa học. Những kết quả chúng ta quan sát được chưa chắc đó là thật, mà là do một yếu tố khác là thủ phạm (mà chúng ta hay gọi là “yếu tố nhiễu”). Một ví dụ tiêu biểu là ảnh hưởng của liệu pháp thay thế hormone (HRT) đến sức khoẻ nữ giới sau mãn kinh. Các nghiên cứu quan sát ban đầu đều cho thấy HRT giảm nguy cơ các bệnh tim mạch,

loãng xương, thậm chí giảm tử vong. Nhưng sau khi người ta làm nghiên cứu RCT thì kết quả ngược lại: ngoại trừ loãng xương bệnh lý mà HRT có ảnh hưởng tốt, các bệnh lý khác như tim mạch và ung thư thì HRT làm tăng nguy cơ mắc bệnh! Phải một thời gian dài suy nghiệm, người ta mới phát hiện ra rằng các nghiên cứu ban đầu là do nghiên cứu ở phụ nữ trung lưu, và chính yếu tố thành phần kinh tế này (chứ không phải HRT) giải thích tại sao HRT có hiệu quả như là “thần dược”. Đó là một bài học kinh điển và... nhớ đời.

Có những yếu tố nhiều nhà khoa học có thể hiệu chỉnh trong khi phân tích, nhưng có rất nhiều yếu tố nhiều nhà khoa học không biết được vì không nghĩ đến hay không thu thập thông tin.

6. Sử dụng trị số P tùy tiện

Một trong những trị số bị hiểu lầm và diễn giải sai nhiều nhất trong khoa học phải là trị số P. Trong lý thuyết kiểm định giả thuyết, trị số P có nghĩa là xác suất quan sát dữ liệu (D) nếu giả thuyết vô hiệu là đúng. Nói cách khác, nếu gọi dữ liệu là D và giả thuyết vô hiệu là H_0 , thì trị số $P = \Pr(D | H_0)$. Thế nhưng trong thực tế, rất nhiều người trong giới khoa học lại hiểu ngược lại rằng $P = \Pr(H_0 | D)$, tức là xác suất giả thuyết vô hiệu là đúng nếu quan sát được dữ liệu D. Từ cách hiểu sai này dẫn đến nhiều kết luận sai hay kết luận quá lạc quan.

7. Kiểm định nhiều giả thuyết

Mỗi lần chúng ta kiểm định một giả thuyết khoa học bằng một phương pháp thống kê (như t-test, Ki bình phương) chúng ta chấp nhận một sai sót “dương tính giả” α phần trăm. Alpha thường là 0,05 hay 0,01 (tức 5% hay 1%). Khi chúng ta kiểm định 2 giả thuyết, xác suất chúng ta quan sát một kết quả

có ý nghĩa thống kê hoàn toàn ngẫu nhiên là $1 - (1 - \alpha)^2$. Nếu chúng ta kiểm định 100 giả thuyết với $\alpha = 0,05$ thì xác suất chúng ta ‘phát hiện’ ít nhất một kết quả có ý nghĩa thống kê sẽ là $(1 - 0,95^{100}) = 0,99$. Thông thường, một nghiên cứu có thể kiểm định hàng trăm, thậm chí hàng ngàn giả thuyết. Nếu là nghiên cứu với hàng triệu yếu tố (như di truyền) thì số giả thuyết kiểm định là hàng trăm tỉ. Trong điều kiện rất nhiều giả thuyết như thế, việc phát hiện một hay nhiều kết quả có ý nghĩa thống kê một cách ngẫu nhiên là điều dễ hiểu. Nhưng những “phát hiện” ngẫu nhiên loại này đóng góp vào bức tranh chung về sự bất lập trong khoa học.

Trong một nghiên cứu mang tính mô phỏng (simulation), các nhà nghiên cứu chia 1073 bệnh nhân ra thành 2 nhóm: nhóm điều trị và nhóm chứng. Họ chia nhóm để sao cho hai nhóm này có tỉ lệ tử vong y chang nhau (tức là điều trị không có hiệu quả). Sau đó, họ thực hiện phân tích trên 18 nhóm nhỏ (subgroups) theo giới tính, độ tuổi, và yếu tố nguy cơ. Họ phát hiện rằng điều trị có hiệu quả ở nữ giới trong độ tuổi 35-55 với điều kiện là sống ở vùng nông thôn và không có tiền sử bệnh tim mạch. Dĩ nhiên, đây là một hình thức “tra tấn dữ liệu” (data torture) để cho ra một kết quả hoàn toàn ngẫu nhiên và chẳng có ý nghĩa gì cả. Nhưng có người làm như thế để cho ra một kết quả “dương tính”, và do đó, đóng góp vào kho rác khoa học.

Năm căn bệnh trong khoa học

Ngoài vấn đề tái lập, tác giả John Antonakis, tổng biên tập tạp chí *Leadership Quarterly* (3) còn chỉ ra 5 căn bệnh trong khoa học. Ông đặt tên theo kiểu chơi chữ là *Significosis*, *Neophilias*, *Theorrhea*, *Arigorium*, và *Disjunctivitis*.

Bệnh tìm ý nghĩa thống kê: Significosis

Là cách chơi chữ từ “significance” trong thống kê học, và osis là tiếp vĩ ngữ y khoa có nghĩa là trạng thái rối loạn. Người mắc bệnh này thường có xu hướng đi tìm những kết quả có ý nghĩa thống kê (có nghĩa là “statistically significant”). Những kết quả như thế giúp cho nhà khoa học nâng cao khả năng công bố bài báo. Một kết quả với $P < 0,05$ có khi được xem như là một giấy thông hành cho công bố quốc tế. Nhưng trong thực tế, ít ai biết rằng 1/3 những kết quả như thế ($P < 0,05$) là sai, tức là dương tính giả. Nhìn thấy nó có ý nghĩa thống kê đó, nhưng trong thực tế thì chẳng có ảnh hưởng hay tác động gì cả. Do đó, công bố những bài báo chỉ dựa vào “ý nghĩa thống kê” làm cho nền khoa học mắc bệnh. Đó có thể xem significosis là bệnh “rối loạn thống kê”.

Bệnh tìm cái mới: Neophilia

Cũng là một cách sáng tạo chữ mới từ neo (có nghĩa là “mới») và philia có nghĩa là “yêu thích»). Do đó, neophilia là bệnh yêu thích cái mới. Trong thực tế, rất hiếm khi khoa học tìm ra cái mới, vì trước đó đã có quá nhiều phát hiện. Đại đa số các nghiên cứu khoa học ngày nay chỉ mang tính “incremental», tức là thêm kiến thức một chút so với kiến thức cũ, chứ rất rất hiếm các nghiên cứu dạng “breakthrough” hay đột phá. Nhưng bất cứ nhà khoa học nào cũng muốn tìm cái mới. Cái mới có thể là một phát hiện động vật mới, bệnh lí mới, hay nói chung là kết quả mới. Cái mới giúp cho nhà khoa học nâng cao khả năng bài báo được công bố trên các tạp san khoa học lừng danh. Vì thế, giới khoa học tìm cách nâng tầm nghiên cứu của mình để sao cho nó có cái mới, và cách vạy vẹo đó dẫn đến tình trạng rất nhiều nghiên cứu phát hiện ra “cái mới”, nhưng khi các nhóm nghiên cứu độc lập thực hiện thì không lặp lại được. Có thể nói

rằng hầu hết (có thể 90%) các kết quả nghiên cứu công bố trên các tập san khoa học hiện nay là sai hoặc không tái lập được. Do đó, **Neophilia** có thể xem là bệnh “Ái tân”.

Bệnh hám lí thuyết: Theorrhea

Cũng là một cách chơi chữ hay, vì theoros có nghĩa là suy nghiệm, và tiếp vĩ ngữ orrhea trong y khoa có nghĩa là “chảy”. Do đó, bệnh này có thể hiểu là bệnh “tiêu chảy lí thuyết.” Khoa học chẳng những yêu cái mới, mà còn yêu lí thuyết, giả thuyết. Đọc bài báo khoa học nào cũng thấy tác giả cố gắng để ra một giả thuyết mới, một lí thuyết mới. Bài báo bắt đầu bằng một giả thuyết được đánh giá cao hơn bài báo kiểu “bước đầu đánh giá”. Nhưng trong thực tế, những cái gọi là lí thuyết mới đó chỉ là hoa ngữ thôi, chứ chẳng phải lí thuyết thật sự, và cũng chẳng bao giờ được kiểm định cả. Do đó, sự hiện diện của quá nhiều lí thuyết dỏm làm cho môi trường khoa học bị dơ bẩn, và dẫn đến... tiêu chảy.

Bệnh hám thuật toán: Arigorium

Tôi đoán là lấy từ tên của một lang băm, Jose Arigo, người tự cho rằng ông có khả năng làm phẫu thuật tâm thần. Có lẽ tác giả muốn dùng tên ông này để chỉ những nghiên cứu không có cơ sở lí thuyết vững vàng, mà chỉ dựa trên những kết quả bề ngoài. Đa số các nghiên cứu khoa học xã hội, kể cả kinh tế học, là những nghiên cứu “soft”, hiểu theo nghĩa dữ liệu là không có độ chính xác cao. Khoa học xã hội không có cái xa xỉ với máy móc thiết bị đắt tiền để đo lường hiện tượng tự nhiên; họ phải dựa vào bộ câu hỏi hay những phương tiện bán định tính như thế. Do đó, để nâng tầm “khoa học”, giới khoa học xã hội thường sáng chế ra những phương pháp phân tích thống kê phức tạp. Những phương pháp này dựa trên nhiều giả định, và ít khi nào nói lên

các mối liên hệ nhân quả. Vì không nói lên mối liên hệ nhân quả, nên nhiều phát hiện trong nghiên cứu khoa học chỉ mang tính ngẫu nhiên, như lang băm Arigo chữa bệnh bằng tay vậy.

Bệnh chạy theo lượng: Disjunctivitis

Có lẽ là lấy từ chữ disjunction có nghĩa là rời rạc, và itis là tiếp vĩ ngữ trong y khoa có nghĩa là viêm. Do đó, bệnh này có thể dịch là bệnh “viêm lượng phẩm”. Có thể hiểu là bệnh mất cân đối giữa lượng và phẩm. Vì áp lực công bố, nên nhiều nhà khoa học có xu hướng công bố nhiều bài báo khoa học, nhưng phẩm chất thì rất kém. Có thể nói rằng rất nhiều bài báo khoa học hiện nay là thừa thải, hay nói theo dân gian Việt Nam chúng ta là “không mợ thì chợ vẫn đông”, tức là có hay không có bài báo thì khoa học vẫn chẳng có ảnh hưởng gì. Lí do là những công trình này chỉ lặp lại những gì người khác đã làm mà không có cái gì mới. Nhưng nhà khoa học, đặc biệt là từ các nước đang phát triển, phải công bố càng nhiều càng tốt, kể cả công bố trên tập san dỏm. Công bố để được bảo vệ luận án, được giải thưởng và được đề bạt. Nhưng công bố bất chấp phẩm chất như thế không làm cho khoa học tiến bộ thêm, chẳng tạo ra được kiến thức mới, mà chỉ làm cho khoa học thêm vẩn đục.

Sau khi đã chẩn đoán bệnh, vậy câu hỏi là làm gì để điều trị bệnh? Tác giả đề ra 8 liệu pháp:

1. Nghiên cứu nên có cơ sở lí thuyết tốt;
2. Sáng tạo trong đo lường;
3. Thiết kế nghiên cứu thực tiễn;
4. Làm nghiên cứu định tính có phẩm chất cao;
5. Chú ý đến các yếu tố nhiễu và bias;

6. Minh bạch trong báo cáo và dữ liệu;
7. “Khai báo” những mâu thuẫn về lợi ích; và
8. Thành thật trong khoa học.

Tôi nghĩ đây là những liệu pháp hoàn toàn khả thi. Số là bài báo đang được bình duyệt và xác suất cao sẽ được chấp nhận, nhưng ban biên tập đã viết cho chúng tôi gợi ý là công bố toàn bộ dữ liệu! Các bạn thử tưởng tượng nhóm nghiên cứu bỏ ra cả hai năm trời, mai phục ngày và đêm “bắt” bệnh nhân cho được, với kết quả hơn 4000 bệnh nhân, mà bây giờ họ muốn mình công bố cho họ xem. Cái nguyên lý đăng sau sự minh bạch là tuyệt vời vì nó giúp cho khoa học thêm liêm chính và sẽ làm cho những nhà khoa học ăn gian biến mất, nhưng nguyên tắc này gây nên vấn đề đạo đức khoa học.

Tôi nghĩ những người trong cuộc ngày nay không còn nghi ngờ gì nữa: khoa học đang rơi vào khủng hoảng. Khủng hoảng niềm tin. Người ta càng ngày càng ít tin khoa học vì có quá nhiều kết quả nghiên cứu không được lặp lại, quá nhiều nghiên cứu vô bổ không sản sinh ra kiến thức mới, và hệ quả là làm cho khoa học trở nên sai và lẫn lộn. Nếu điều trị dứt 5 căn bệnh này (bệnh rối loạn thống kê, ái tân, tiêu chảy lý thuyết, Arigo, và bệnh viêm lượng phẩm) thì may ra có thể cứu vãn được khoa học và khôi phục niềm tin ở công chúng.

Tóm lại, tình trạng bất lập hay irreproducibility trong khoa học hiện nay là một khủng hoảng lớn nhất mà khoa học đang phải đối phó. Rất nhiều kết quả nghiên cứu không thể lặp lại được, và đó là tín hiệu cho thấy kết quả đó có thể sai. Đó cũng chính là lý do tại sao chúng ta đọc được những phát hiện trái ngược nhau: hôm thì có nghiên cứu cho thấy uống rượu vang đỏ có lợi cho sức khỏe, hôm thì có tác giả nói rượu vang

đỏ có hại cho sức khỏe. Thử tưởng tượng xã hội có 1000 yếu tố nguy cơ liên quan đến lối sống và 5000 bệnh lí, chúng ta sẽ có rất rất nhiều phát hiện được báo chí quan tâm. Nhưng đại đa số những cái gọi là “phát hiện” đó là sai.

Nguyên nhân sai có thể là tác giả chỉ thích chọn công bố kết quả dương tính, sai lầm trong hệ thống thưởng cho kết quả dương tính, sai lầm trong thí nghiệm và thu thập dữ liệu, và sai lầm trong phân tích và diễn giải kết quả phân tích thống kê. Sự khủng hoảng của bất lập trong khoa học là một thực tế, và thực tế này đòi hỏi người “tiêu thụ” thông tin khoa học (kể cả dĩ nhiên những người ra chính sách công) phải rất cẩn thận và sáng suốt trong việc sử dụng thông tin khoa học.

Tài liệu tham khảo và đọc thêm:

- (1) <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002165>
- (2) <http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124>. Bài này đã được trích dẫn hơn 4000 lần kể từ khi công bố vào năm 2005, và trở thành một trong những bài có ảnh hưởng lớn nhất đến khoa học.
- (3) <http://www.sciencedirect.com/science/article/pii/S104898431730070X>

On doing better science: From thrill of discovery to policy implications

CON ĐƯỜNG TRỞ THÀNH NHÀ KHOA HỌC ĐỘC LẬP

Sự nghiệp nghiên cứu khoa học có những hướng đi đặc thù so với các ngành nghề. Sự nghiệp khoa học thường - nhưng không nhất thiết phải - bắt đầu với văn bằng tiến sĩ. Từ tiến sĩ, đến nghiên cứu sinh hậu tiến sĩ, giáo sư trợ lý (assistant professor), giáo sư dự bị (associate professor), và giáo sư thực thụ (full professor) là một hành trình gian nan. Trong chương này tôi muốn chia sẻ cùng bạn đọc con đường từ sinh viên sau đại học đến một nhà khoa học độc lập và xác suất thành công.

Ở nước ngoài, con đường từ cử nhân đến tiến sĩ, và từ tiến sĩ đến tư cách nhà khoa học độc lập (independent scientist) là một thời gian dài. Sau khi xong tiến sĩ, người tốt nghiệp sẽ cùng với thầy tìm một nơi để làm nghiên cứu hậu tiến sĩ (gọi là *postdoc*). Nghiên cứu hậu tiến sĩ thường phải làm ở một labo khác, thường là nước ngoài. Những nước được giới nghiên cứu sinh ưa chuộng là Mỹ, Anh, Pháp, Đức, Úc, Canada, v.v. chứ ít khi nào ở các nước “thấp” hơn. Thời gian nghiên cứu hậu tiến sĩ phải tốn 2-3 năm. Nhưng không phải chỉ một labo, mà thường là 2 labo, nên tổng cộng là 3-6 năm.

Thời gian nghiên cứu hậu tiến sĩ là giai đoạn định hình sự nghiệp. Một số người (khoảng 30%) sau khi xong hậu tiến sĩ,

họ không muốn theo đuổi sự nghiệp nghiên cứu, khoa bảng, mà chọn làm việc cho kỹ nghệ, như các tập đoàn dược (nếu là ngành y sinh học). Một số chọn làm quản lý trong hệ thống khoa học của chính phủ và kỹ nghệ. Một số chọn giảng dạy trong đại học làm sự nghiệp, số này lên đến 30%. Và, sau cùng là một số lớn chọn ở lại làm nghiên cứu khoa học. Nói là “lớn”, nhưng cũng chỉ 40% mà thôi. Do đó, sau tiến sĩ còn có nhiều con đường khác nhau cho sự nghiệp, và giai đoạn để mình tự quyết định là giai đoạn hậu tiến sĩ.

Nếu chọn ở lại với khoa học, giai đoạn hậu tiến sĩ là thời gian để định hình hướng nghiên cứu. Thông thường sau khi xong tiến sĩ, người tốt nghiệp đã học được kỹ năng về nghiên cứu, đã theo đuổi một vấn đề suốt 3-4 năm trời, nên họ có một số kiến thức cơ bản, chứ chưa đủ để xem là *chuyên gia* (expert). Một người mới tốt nghiệp tiến sĩ về nghiên cứu cơ bản y sinh chỉ mới nắm được vài kỹ thuật và mô hình labo, chứ chưa thể độc lập được, vì họ vẫn phải được sự hướng dẫn của thầy cô và nghiên cứu sinh hậu tiến sĩ. Giai đoạn hậu tiến sĩ chính là thời gian để họ học thêm, học chuyên sâu hơn, và có thể độc lập. Có thể xem là sau hậu tiến sĩ thì mới có thể nói là “xuống núi”.

Trong thời gian nghiên cứu hậu tiến sĩ, người tiến sĩ phải công bố nhiều công trình khoa học. Thông thường, trong các lĩnh vực khoa học thực nghiệm, mỗi năm 1-2 bài là tiêu chuẩn bắt thành văn, nhưng trong thực tế các postdoc thành công thì công bố nhiều hơn con số đó. Để mang tính cạnh tranh, postdoc không chỉ công bố nghiên cứu, mà còn phải công bố trên tập san hàng đầu. Có người còn nói rằng không

có công bố trên tập san cỡ *Nature* hay *Science* thì chưa đủ để cạnh tranh! Ngoài công bố quốc tế, người tiến sĩ trong giai đoạn nghiên cứu hậu tiến sĩ còn phải rèn luyện kỹ năng lãnh đạo, và thường được giao nhiệm vụ chỉ dẫn các nghiên cứu sinh tiến sĩ về kỹ thuật (nhưng chưa được phép hướng dẫn tiến sĩ). Nói chung, giai đoạn nghiên cứu hậu tiến sĩ là thời gian để tự củng cố mình trước khi ra “biển lớn”.

Theo thống kê, chỉ có 10% hậu tiến sĩ là ở lại với nghiên cứu khoa học (dù 80% họ hi vọng được ở lại đại học) (1). Nói cách khác, có đến 90% hậu tiến sĩ bỏ sự nghiệp khoa học. Một phân tích khác của Royal Society (2) cho thấy hành trình từ khi tốt nghiệp tiến sĩ đến một nhà khoa học thực thụ cũng rất gian nan. Theo thống kê của báo cáo này:

- Hơn phân nửa (53%) tiến sĩ ở Anh tìm sự nghiệp ngoài khoa học và kỹ thuật;
- 17% làm nghiên cứu ở các công ty;
- 30% tìm được việc làm ở giai đoạn “early career” (tức là hậu tiến sĩ, giảng viên, assistant professor) trong các đại học;

Chỉ có 3,5% trong số tốt nghiệp tiến sĩ tìm được việc làm nghiên cứu khoa học trong các đại học. Chỉ có 0,45% (tức ~5 người trên 1000) sau này trở thành giáo sư thực thụ (2). Còn số người thành giáo sư thực thụ và có labo độc lập thì còn hiếm hơn nữa.

Cần nói thêm rằng quá trình trên chủ yếu là trong khoa học thực nghiệm như y khoa, sinh học, vật lý, hóa học, v.v. Có thể các ngành khác như kỹ thuật và công nghệ thì quá trình

từ tiến sĩ thành giáo sư khác với các ngành khoa học thực nghiệm.

Như vậy, quá trình từ tốt nghiệp tiến sĩ đến nhà khoa học độc lập là một hành trình rất gian nan và lâu năm. Chỉ có 30% tiến sĩ có vị trí hậu tiến sĩ. Chỉ có 10% hậu tiến sĩ còn tồn tại trong hệ thống nghiên cứu khoa học. Và, chỉ có một số ít hậu tiến sĩ sau này trở thành giáo sư; và trong số này chỉ có một số rất ít đứng đầu labo nghiên cứu. Thời gian từ tiến sĩ đến xong hậu tiến sĩ là từ 3-6 năm, nhưng cũng có thể lâu hơn. Điều này chỉ để nói rằng đào tạo được một nhà khoa học thực thụ cần thời gian và đầu tư nhiều công sức và tài nguyên.

Tham khảo:

- (1) <https://www.ecoom.be/en/research/doctoralcareers>
- (2) http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/publications/2010/4294970126

MỤC LỤC

- Lời nói đầu 3
- Tại sao nghiên cứu khoa học? 7
- Thế nào là nghiên cứu khoa học? 15
- Ý tưởng nghiên cứu đến từ đâu? 24
- Cách tiếp cận câu hỏi nghiên cứu 47
- Thế nào là “cái mới” trong khoa học? 57
- Chọn mô hình nghiên cứu
và những vấn đề liên quan đến thiết kế 66
- Cỡ mẫu và huyền thoại con số 30 85
- Ngộ nhận về tính đại diện
trong nghiên cứu khoa học 98
- Cách thiết kế bộ câu hỏi 109

• Phân tích dữ liệu	125
• Hai trường phái phân tích dữ liệu	151
• Trị số P trong nghiên cứu khoa học	171
• Diễn giải kết quả nghiên cứu khoa học	193
• Những nguyên tắc trong việc soạn một bài báo khoa học	232
• Tiếng Anh trong bài báo khoa học	310
• Những bài học về viết bài báo khoa học	339
• Vấn nạn tập san khoa học "dorm"	352
• Cách đọc bài báo khoa học	368
• Khoa học Việt Nam đang ở đâu	373
• Những vấn đề thời sự	382
• Con đường trở thành nhà khoa học độc lập	394

Bạn muốn dẫn thân vào nghiên cứu khoa học? Bạn muốn hiểu qui trình nghiên cứu khoa học để có thể đánh giá một phát hiện được tường thuật trên các tập san khoa học? Bạn muốn làm nghiên cứu khoa học nhưng không biết ý tưởng đến từ đâu, tiêu chí gì để đánh giá ý tưởng có xứng đáng để theo đuổi, cách thiết kế bộ câu hỏi (questionnaire), cách thiết kế nghiên cứu, cách soạn bài báo khoa học, v.v. là những câu hỏi sẽ được giải đáp qua cuốn sách nhỏ này.

Bạn đọc sẽ được dẫn dắt qua một qui trình nghiên cứu khoa học từ lúc hình thành ý tưởng, thu thập dữ liệu qua thí nghiệm, đến cách soạn bài báo khoa học bằng tiếng Anh, và công bố khoa học. Mỗi qui trình được minh hoạ bằng những ví dụ thực tế rất sinh động cùng những kinh nghiệm và trải nghiệm của chính tác giả. Những trải nghiệm của tác giả từng là một nghiên cứu sinh, một giám đốc labo nghiên cứu, đến biên tập cho các tập san y khoa nổi tiếng sẽ giúp cho bạn đọc cảm thấy nghiên cứu khoa học là một hành trình khám phá thú vị và hào hứng.

