

GIỚI THIỆU PHƯƠNG PHÁP PHÂN TÍCH BAYES

Phần 2: Diễn giải kết quả nghiên cứu

Nguyễn Văn Tuấn*

Tóm tắt

Kết quả nghiên cứu lâm sàng thường được tóm lược bằng trị số P và khoảng tin cậy 95%. Nhưng trị số P rất dễ dẫn đến hiểu lầm và kết luận sai. Trị số P cũng không phản ánh được tầm ảnh hưởng của can thiệp. Một kết quả với trị số $P < 0,05$ thường được xem là có ý nghĩa thống kê, nhưng kết quả đó không nói lên có ý nghĩa lâm sàng hay không. Trong vài năm gần đây, xu hướng ứng dụng phương pháp Bayes trong việc diễn giải kết quả nghiên cứu lâm sàng. Bài viết này đi qua triết lý đằng sau kiểm định thống kê, kiểm định giả thuyết, và trình bày qui trình phân tích và đánh giá kết quả nghiên cứu qua phương pháp Bayes với hi vọng sẽ giúp cho bác sĩ và nhà nghiên cứu có một cách diễn giải gần với thực tế lâm sàng hơn là phương pháp cổ điển.

Summary

Results of clinical research are usually presented in terms of P values and confidence intervals. However, P values are often misinterpreted and could lead to wrong conclusion. P values do not tell us whether an effect size is clinically relevant. A result with $P < 0.05$ is understood as "statistically significant", but such a result may not have clinical significance. In recent years, the Bayesian methodology has been proposed as a calculus of evidence and has increasingly been adopted in clinical research. In this article, I briefly review the Fisher's method significance testing, Neyman-Pearson's method of hypothesis testing, and then present procedures of Bayesian inference via two examples of a clinical trial results. I hope that the ideas and methods presented here will help clinicians and researchers alike to have a better interpretation of results from clinical research.

Phương pháp Bayes

Trong bài trước, tôi đã trình bày cách ứng dụng phương pháp Bayes trong việc diễn giải kết quả chẩn đoán lâm sàng. Tiếp tục bài trước, bài này sẽ trình bày phương pháp phân tích Bayes để đánh giá và diễn giải kết quả của một nghiên cứu. Tôi bắt đầu bằng hai trường hợp thực tế:

Trường hợp 1

Bổ sung vitamin D + calci (CaD) và nhồi máu cơ tim. Một phân tích dữ liệu của công trình nghiên cứu WHI mới công bố trên tập san BMJ⁽¹⁾ cho thấy phụ nữ dùng bổ sung calci và vitamin D có nguy cơ nhồi máu cơ tim tăng (Bảng 1). Tỷ số nguy cơ là 1,22 và khoảng tin cậy 95% dao động từ 1,0 đến 1,5. Bệnh nhân đang dùng bổ sung calci và vitamin D rất quan tâm khi nghe tin này và muốn được bác sĩ tư vấn. Bác sĩ phải nói gì với bệnh nhân?

Bảng 1. Kết quả nghiên cứu WHI: tần số nhồi máu cơ tim (MI) trong nhóm dùng calci và vitamin D (CaD) và nhóm giả dược

Nhóm can thiệp	MI	Không MI	Tổng số
CaD	209 (2,48%)	8220	8429
Giả dược	168 (2,03%)	8121	8289

Trường hợp 2

Strontium ranelate và gãy xương đốt sống. Một công trình nghiên cứu lâm sàng đối chứng ngẫu nhiên về hiệu quả của một loại thuốc điều trị loãng xương (strontium ranelate - SR)⁽²⁾ trong đó 1442 bệnh nhân loãng xương được chia thành 2 nhóm: nhóm 1 gồm 719 người được điều trị với SR, và nhóm giả dược gồm 723 người. Sau 3 năm điều trị và theo dõi, kết quả cho thấy số ca gãy xương đốt sống trong nhóm điều trị là 46 (6,4%) và trong nhóm giả dược là 88 (12,2%). Kết quả này có ý nghĩa lâm sàng hay không?

Bảng 2. Kết quả nghiên cứu hiệu quả của strontium ranelate trong điều trị loãng xương ở phụ nữ sau mãn kinh

Nhóm can thiệp	Số bệnh nhân gãy xương đốt sống	Số bệnh nhân không gãy xương	Tổng số
Strontium ranelate	46 (6.4%)	673	719
Giả dược	88 (12.2%)	635	723

Đứng trước một kết quả nghiên cứu, người thầy thuốc muốn hỏi: có phải mối liên hệ này là thật, hoặc có phải ảnh hưởng của một thuật can thiệp có ý nghĩa lâm sàng? Đó là những câu hỏi thực tế, chẳng những liên quan đến người thầy thuốc, mà còn rất liên quan đến bệnh nhân. Để có câu trả lời chính xác và đáng tin cậy là điều không dễ. Không dễ vì phương tiện phân tích hiện nay mà phần lớn bác sĩ học qua đều không thể ứng dụng để trả lời hai câu hỏi vừa nêu trên. Để hiểu vấn đề từ gốc, chúng ta cần phải hiểu qua triết lý khoa học và trường phái trong phương pháp khoa học (scientific method).

* Viện Nghiên cứu Y khoa Garvan, Sydney, Australia

Kiểm định thống kê và kiểm định giả thuyết

Phương pháp phân tích hiện nay xuất phát từ triết lý phản nghiệm (falsificationism) do Karl Popper đề xuất từ đầu thập niên thế kỉ 20.⁽³⁾ Theo triết lý phản nghiệm, chúng ta không bao giờ chứng minh được một giả thuyết; chúng ta chỉ có thể bác bỏ giả thuyết. Chịu ảnh hưởng bởi triết lý phản nghiệm, Ronald A. Fisher, một nhà di truyền học và thống kê học xuất sắc, đề xuất phương pháp kiểm định thống kê (test of significance) để kiểm tra khả năng của một giả thuyết.⁽⁴⁾ Quy trình kiểm định thống kê có thể mô tả một cách đơn giản như sau:

- Đề ra một giả thuyết vô hiệu H_0 (null hypothesis). Chẳng hạn như trong nghiên cứu 1, giả thuyết vô hiệu là không có mối liên quan giữa CaD và nhồi máu cơ tim (tức tỉ lệ mắc bệnh nhồi máu cơ tim trong nhóm CaD không khác so với nhóm giả dược);
- Thu thập dữ liệu D . Dữ liệu ở đây chính là hiệu số về tỉ lệ mắc bệnh nhồi máu cơ tim giữa nhóm CaD và nhóm giả dược, hoặc cũng có thể là tỉ số nguy cơ và khoảng tin cậy 95%;
- Tính xác suất D với điều kiện H_0 là đúng; nói cách khác, ước tính $P(D | H_0)$.

$P(D | H_0)$ còn được biết đến như là trị số P . Fisher còn đề nghị nếu $P < 0,05$ thì có thể xem đó là một kết quả có ý nghĩa thống kê (statistically significant) và có thể bác bỏ giả thuyết H_0 . Phương pháp vừa mô tả được công bố trong một loạt bài báo khoa học và sau này thành hai cuốn sách giáo khoa có ảnh hưởng cực kỳ lớn (chỉ sau Kinh Thánh), đó là cuốn *Statistical Methods for Research Workers* (1925) và *The Design of Experiments* (1935a).

Phương pháp kiểm định thống kê như Fisher đề xuất là một cách suy luận một chiều. Trong thực tế, khoa học không chỉ dựa vào một giả thuyết vô hiệu, mà còn giả thuyết đảo (alternative hypothesis). Không hài lòng với phương pháp này, Jerzy Neyman và Egon Pearson đề xuất một phương pháp khác có tên là kiểm định giả thuyết (test of hypothesis).⁽⁵⁾ Cần nói thêm rằng Jerzy Neyman là một nhà toán học gốc Ba Lan rất nổi tiếng sang tị nạn chính trị ở Anh. Neyman nghiên cứu với Egon Pearson (là con trai của Karl Pearson, người phát triển lý thuyết Chi bình phương). Sau này vì sự hẹp hòi của Fisher, Neyman phải một lần nữa bỏ Anh sang tị nạn ở Mỹ và trở thành giáo sư nổi tiếng của Đại học California tại Berkeley. Phương pháp kiểm định giả thuyết là một cải tiến lý thuyết kiểm định

thống kê của Fisher. Phương pháp này có thể tóm lược qua các bước như sau:

- Phát biểu hai giả thuyết (chứ không phải một như Fisher): giả thuyết vô hiệu (H_0) và giả thuyết đảo (H_A). Giả thuyết đảo trong nghiên cứu 1 là có sự khác biệt về tỉ lệ mắc bệnh nhồi máu cơ tim giữa nhóm CaD và giả dược;
- Xác định giá trị α và β . Giá trị α là sai sót loại I (type I error), tức là xác suất bác bỏ H_0 khi H_0 là đúng. Giá trị β là sai sót loại II (type II error) là xác suất bác bỏ H_A khi H_A là đúng. Có thể xem α là tỉ lệ dương tính giả và β là tỉ lệ âm tính giả như trong chẩn đoán y khoa;
- Thu thập dữ liệu D và tính toán test thống kê (như z , t , F , Chi bình phương, v.v.);
- Nếu giá trị của các test nằm trong vùng của H_A (như z cao hơn 1,96), bác bỏ giả thuyết H_0 ; nếu không thì chấp nhận H_0 và bác bỏ H_A .

Trong kiểm định giả thuyết, không có trị số P . Đến đây, cần phải nhấn mạnh rằng rất nhiều người (kể cả các nhà thống kê học) hiểu lầm rằng trị số P là sai sót loại I (tức α). Theo lý thuyết vừa mô tả trên, sai sót I không phải là trị số P . Chính Fisher từng chỉ ra rằng trị số P không thể diễn giải theo quan niệm tần số mẫu (sampling interpretation) như α . Trị số $P = 0,05$ không có nghĩa là 1 trong 20 nghiên cứu sẽ bác bỏ giả thuyết vô hiệu. Kiểm định thống kê chỉ có thể cung cấp một chứng cứ khách quan để không tin vào giả thuyết vô hiệu, chứ nó không có ý nghĩa lâu dài như trường phái tần số.

Fisher phản đối quyết liệt lý thuyết kiểm định giả thuyết của Neyman và Pearson. Fisher chế nhạo rằng Neyman và Pearson là những nhà toán học tiêu biểu, chẳng biết nghiên cứu khoa học là gì (Fisher không xem toán là một bộ môn khoa học). Ngược lại, Neyman và Pearson cho rằng phương pháp của Fisher là không phản ánh được thực tế, vì khoa học lúc nào cũng có hơn một giả thuyết. Sự tranh cãi giữa Fisher và Neyman-Pearson kéo dài hơn 10 năm với hàng chục bài báo khoa học để lại cho khoa học một di sản rất phong phú.

Trở trở thay, mô hình nghiên cứu khoa học mà chúng ta sử dụng ngày nay là một sự kết hợp hai lý thuyết kiểm định thống kê của Fisher và kiểm định giả thuyết của Neyman-Pearson. Thật vậy, ngày nay một công trình nghiên cứu khoa học được tiến hành theo các trình tự như sau:

- Đề xuất giả thuyết vô hiệu (H_0) và giả thuyết đảo (H_A);
- Xác định giá trị α và β , và qua đó xác định cỡ mẫu;

- Thu thập dữ liệu D và tính toán test thống kê (như z , t , F , Chi bình phương, v.v.);
- Tính toán trị số P . Nếu $P < \alpha$, bác bỏ giả thuyết H_0 ; nếu không thì chấp nhận H_0 .

Vấn đề trị số P

Trị số P từ đó đóng một vai trò rất quan trọng trong nghiên cứu khoa học. Có thể nói không ngoa rằng trị số P đã trở thành một loại “giấy thông hành” để công bố kết quả nghiên cứu. Một kết quả với trị số $P < 0,05$ thường có khả năng được công bố cao hơn một kết quả với trị số $P > 0,05$. Đã có không biết bao nhiêu nhà nghiên cứu trên thế giới suốt gần 100 năm qua hiểu lầm ý nghĩa của trị số P và do đó đã có những quyết định sai lầm. Một trong những hiểu lầm phổ biến nhất là cho rằng trị số P là xác suất giả thuyết vô hiệu H_0 ; theo đó nếu một kết quả có trị số $P = 0,05$, người ta hiểu rằng xác suất giả thuyết vô hiệu đúng (tức không có khác biệt giữa 2 nhóm) là 5%. Nhưng đó là một hiểu lầm. Trong thực tế, có đến 12 hiểu lầm về trị số P mà tôi trích dẫn dưới đây để tham khảo (Bảng 3) ⁽⁶⁾.

Bảng 3. Mười hai ngộ nhận về trị số P

1. Nếu trị số $P = 0,05$, giả thuyết vô hiệu chỉ có 5% khả năng là đúng.
2. Một kết quả không có ý nghĩa thống kê ($P > 0,05$) có nghĩa là không có khác biệt giữa các nhóm.
3. Một kết quả có ý nghĩa thống kê là có ý nghĩa lâm sàng.
4. Nghiên cứu với trị số P đối nghịch với 0,05 là mâu thuẫn.
5. Hai nghiên cứu với trị số P giống nhau có nghĩa là chứng cứ tương đương nhau.
6. $P < 0,05$ có nghĩa là chúng ta đã quan sát dữ liệu chỉ xảy ra dưới 5% trong điều kiện giả thuyết vô hiệu.
7. $P = 0,05$ và $P < 0,05$ có ý nghĩa như nhau.
8. Trị số P phải viết theo bất đẳng thức (ví dụ: “ $P < 0,02$ ” khi $P = 0,015$).
9. $P = 0,05$ có nghĩa là nếu chúng ta bác bỏ giả thuyết vô hiệu, thì xác suất sai sót loại I chỉ 5%.
10. Với $P = 0,05$ để xác định ý nghĩa thống kê, khả năng sai sót loại I sẽ là 5%.
11. Chúng ta dùng trị số P một chiều (one-sided P value) khi chúng ta không quan tâm đến kết quả của một chiều hướng, hoặc khác biệt chiều hướng đó không thể xảy ra.
12. Một kết luận khoa học nên dựa vào trị số P có ý nghĩa hay không.

Trị số P đã bị phê bình ngay từ lúc ra đời trên 70 năm qua. Nhưng vì phương pháp kiểm định giả thuyết và kiểm định thống kê tương đối đơn giản, và cũng phục vụ cho khoa học khá tốt, nên phương pháp này vẫn còn được ứng dụng rộng rãi cho đến ngày nay. Có thể liệt kê ra 4 khiếm khuyết chính của trị số P như sau:

Thứ nhất, trị số P không phản ánh sự khả dĩ của một giả thuyết khoa học. Trị số P là một xác suất có

điều kiện. Như trình bày trên, P là xác suất dữ liệu quan sát (và chưa quan sát) xảy ra nếu giả thuyết vô hiệu là đúng. Nhân mạnh: nếu. Do đó, trị số P không nói gì về giả thuyết H_0 hay H_A cả, mà là xác suất của dữ liệu với điều kiện H_0 là đúng. Sự khó hiểu của trị số P vì nó xuất phát từ phương pháp phản chứng. Có thể minh họa logic của phản chứng (tức trị số P) như sau:

- Mệnh đề 1: Nếu giả thuyết vô hiệu đúng, thì sự kiện này không thể xảy ra;
- Mệnh đề 2: Sự kiện xảy ra;
- Mệnh đề 3 (kết luận): Giả thuyết vô hiệu không thể đúng.

Nếu cách lập luận trên khó hiểu, chúng ta thử xem một ví dụ cụ thể như sau:

- Nếu ông Tuấn bị cao huyết áp, thì ông không thể có triệu chứng rụng tóc (hai hiện tượng sinh học này không liên quan với nhau, ít ra là theo kiến thức y khoa hiện nay);
- Ông Tuấn bị rụng tóc;
- Do đó, ông Tuấn không thể bị cao huyết áp.

Trị số P , do đó, gián tiếp phản ánh xác suất của mệnh đề 3. Và đó cũng chính là một khiếm khuyết quan trọng của trị số P , bởi vì nó ước tính mức độ khả dĩ của dữ liệu, chứ không nói cho chúng ta biết mức độ khả dĩ của một giả thuyết. Điều này làm cho việc suy luận dựa vào trị số P rất xa rời với thực tế, xa rời với khoa học thực nghiệm. Trong khoa học thực nghiệm, điều mà nhà nghiên cứu muốn biết là với dữ liệu mà họ có được, xác suất của giả thuyết chính là bao nhiêu, chứ họ không muốn biết nếu giả thuyết đảo là sự thật thì xác suất của dữ liệu là bao nhiêu. Nói cách khác và dùng kí hiệu mô tả trên, nhà nghiên cứu muốn biết $P(H_A | D)$, chứ không muốn biết $P(D | H_0)$ hay $P(D | H_A)$.

Sự vô lý ở đây còn là những dữ liệu chưa quan sát được. Nói cách khác trị số P phản ánh một phần những dữ liệu quan sát được và một phần chưa quan sát được. Chính vì thế mà trị số P rất u “mong manh”. Chẳng hạn như trong nghiên cứu 1, chỉ cần thay đổi 2 ca nhồi máu cơ tim là kết quả khác (trị số $P > 0,05$), tức không có ý nghĩa thống kê!

Thứ hai, cỡ mẫu có ảnh hưởng lớn đến trị số P . Có thể nói rằng trị số P là một thước đo của cỡ mẫu. Với một cỡ mẫu lớn (hay rất lớn), bất cứ một khác biệt nào -- dù thấp đến đâu -- vẫn có thể trở nên có ý nghĩa thống kê, tức $P < 0,05$. Chẳng hạn như một nghiên cứu với khoảng 20.000 bệnh nhân trong mỗi nhóm, dù sự khác biệt dưới 0,01% vẫn có thể có ý nghĩa thống kê! Tuy nhiên, vấn đề mà chúng ta quan tâm là ý nghĩa lâm sàng chứ không

phải ý nghĩa thống kê. Một công trình nghiên cứu có ý nghĩa phải được thiết kế để phát hiện một ảnh hưởng có ý nghĩa lâm sàng. Có ý nghĩa thống kê là điều kiện cần, nhưng chưa đủ để phát biểu về ý nghĩa lâm sàng. Do đó, trị số P không cung cấp thông tin về mức độ ảnh hưởng lâm sàng.

Thứ ba, khi kiểm định nhiều giả thuyết và dựa vào tiêu chuẩn $P < 0,05$ dễ dẫn đến sai lầm. Theo mô hình kiểm định giả thuyết, cứ mỗi lần so sánh (hay kiểm định một giả thuyết) chúng ta chấp nhận một sai sót α và đúng $1-\alpha$. Nói cách khác, nếu chúng ta kiểm định k giả thuyết thì xác suất có ít nhất một kết quả có ý nghĩa ($P < \alpha$) là $1 - (1 - \alpha)^k$. Chẳng hạn như với 10 kiểm định giả thuyết và nếu $\alpha = 0,05$, thì xác suất có ít nhất 1 kết quả có ý nghĩa thống kê hoàn toàn do ngẫu nhiên là $1 - (1 - 0,05)^{10} = 0,40$ hay 40%. Khi số lần so sánh trong một nghiên cứu là 100, thì chắc chắn sẽ có ít nhất là một khác biệt có ý nghĩa thống kê nhưng hoàn toàn do yếu tố ngẫu nhiên.

Thứ tư, khoảng tin cậy 95% của phân tích theo trường phái tần số cũng khó diễn giải. Vì những khiếm khuyết hiển nhiên của trị số P nên khoảng tin cậy được đề nghị như là một chỉ số khác để bổ sung của trị số P . Khoảng tin cậy 95% (KTC 95%) rõ ràng là một cải tiến so với trị số P , vì nó cung cấp cho chúng ta thông tin về mức độ ảnh hưởng (magnitude of effect hay effect size). Nhưng KTC 95% vẫn không giải quyết được vấn đề cơ bản của kiểm định giả thuyết khoa học, vì nếu trị số $P < 0,05$ thì KTC 95% không bao gồm giá trị vô hiệu, và nếu trị số $P > 0,05$ thì KTC 95% bao gồm giá trị vô hiệu. Ngoài ra, có quá nhiều người hiểu lầm rằng KTC 95% là một xác suất của thông số, như nếu KTC 95% dao động từ 1,1 đến 1,5 thì người ta hiểu rằng xác suất thông số trong khoảng 1,1 đến 1,5 là 95%. Nhưng đó là một hiểu lầm, bởi vì KTC 95% không có ý nghĩa đó. Để có thể diễn giải theo ý nghĩa đó đòi hỏi phải dùng đến phương pháp Bayes.

Nói tóm lại, phương pháp phân tích dựa vào trị số P có nhiều vấn đề về logic. Có thể nói phương pháp này đã gây tác hại đến khoa học, và đã đến lúc cần phải khắc phục những khiếm khuyết đó. John Nelder, cựu chủ tịch Hội Thống kê học Hoàng gia Anh tuyên bố: “Trước mắt chúng ta, việc quan trọng nhất là phải đánh đổ văn hóa trị số P vốn đã bám rễ sâu một cách đáng sợ trong nhiều lĩnh vực khoa học thuần túy và khoa học ứng dụng, và công nghệ” (“The most important task before us (...) is to demolish the P -value culture, which has taken root to a frightening extent in many areas of both pure and applied science, and technology”).⁽⁷⁾

Phương pháp Bayes

Trước những khiếm khuyết đó, các nhà khoa học càng ngày càng nhận ra những khiếm khuyết trong suy luận dựa vào trị số P . Có người đề nghị rằng cần phải tẩy chay trị số P bởi vì phương pháp này dẫn đến quá nhiều sai lầm trong những kết luận khoa học. Có khi một phát hiện quan trọng nhưng chỉ vì $P = 0,051$ và không được công bố! Ngược lại, có quá nhiều phát hiện có ý nghĩa thống kê, nhưng thật ra là do “đương tính giả”. Do đó, có thể nói rằng phần lớn (có thể là 90%) những phát hiện trong y khoa trong thời gian qua có thể sai lầm⁽⁸⁾!

Một trường phái (hay cũng có thể nói là phương pháp) suy luận tốt hơn là dựa vào lý thuyết Bayes. Lý thuyết Bayes dựa trên cơ sở định lý Bayes do linh mục Thomas Bayes đề xuất vào năm 1763 trong một bài báo khoa học quan trọng.⁽⁹⁾ Định lý Bayes có thể phát biểu như sau: gọi H_A là giả thuyết thuộc có hiệu quả, D là dữ liệu quan sát được, xác suất H_A đúng với dữ liệu D là:

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

Trong đó, $P(H_A)$ là xác suất H_A đúng (còn gọi là xác suất tiền định – prior probability); $P(D | H_A)$ là xác suất dữ liệu quan sát được nếu H_A đúng; và $P(D)$ là xác suất dữ liệu xảy ra. Do đó, phương pháp phân tích Bayes có 3 thông tin:

- Xác suất tiền định cho một giả thuyết, $P(H_A)$;
- Dữ liệu thực tế, có khi còn gọi là likelihood, $P(D)$;
- Xác suất hậu định, tức $P(H_A | D)$.

Xác suất tiền định

Trước khi thực hiện một công trình nghiên cứu, chúng ta thường đã có vài ý kiến về khả năng của một giả thuyết. “Khả năng” có thể mô tả bằng xác suất. Chẳng hạn như đối với mối liên hệ giữa CaD và nhồi máu cơ tim (nghiên cứu 1), thiếu những lý do sinh học nào để tin rằng có một mối liên hệ như thế; cho nên chúng ta có thể phát biểu sơ khởi rằng xác suất có mối liên hệ là rất thấp, như $\pi = 0,05$ (5%). Trong nghiên cứu 2 (strontium ranelate) và gãy xương đốt sống, chúng ta có nhiều nghiên cứu cơ bản cho thấy SR có thể tăng mật độ xương và do đó có thể giảm nguy cơ gãy xương. Nhưng để có lý do thực hiện một nghiên cứu lâm sàng chúng ta phải chấp nhận khả năng 50/50 (equipoise), tức xác suất có hiệu quả là $\pi = 0,5$.

Trong thực tế, xác suất tiền định có thể “phát biểu” bằng các luật phân phối xác suất. Bởi vì dữ liệu của cả hai nghiên cứu có thể mô tả bằng tỉ số

nguy cơ (relative risk – RR). Tỷ số nguy cơ có thể dao động, chẳng hạn như từ 0,2 (thuốc giảm 80% nguy cơ) đến 10 (thuốc tăng nguy cơ 5 lần). Nhưng chúng ta biết rằng khả năng thuốc giảm 90% nguy cơ hay tăng nguy cơ gấp 5 lần là rất thấp, nếu không muốn nói là không thể có. Vì thế có thể thể hiện xác suất tiền định bằng luật phân phối chuẩn.

Trong phân phối chuẩn, thay vì thể hiện bằng RR, chúng ta thể hiện bằng logarit RR (ký hiệu logRR) tiện hơn. Do đó, thay vì phát biểu RR = 1, chúng ta phát biểu logRR = 0 (vì log của 1 là 0); thay vì RR = 0,2, chúng ta viết logRR = -1,61; và thay vì RR = 5, chúng ta viết logRR = 1,61. Phân phối chuẩn của logRR được xác định bởi 2 thông số: trung bình và phương sai. Gọi số trung bình của logRR là m_0 và phương sai của logRR là v_0 . Với giả thuyết vô hiệu, chúng ta giả định rằng không có sự khác biệt giữa hai nhóm (như CaD và giả dược, hay giữa SR và giả dược), do đó:

$$m_0 = 0$$

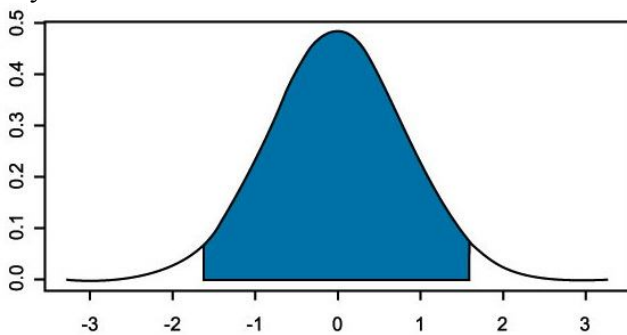
Nhưng chúng ta nghĩ rằng logRR có thể dao động từ -2,3 đến 1,61, với khả năng logRR < -2,3 là 2,5%, và khả năng logRR > 1,61 là 2,5%. Do đó, độ lệch chuẩn tiền định (standard deviation, ký hiệu s_0) có thể ước tính bằng:

$$s_0 = (1,61 - (-1,61)) / 3,92 = 0,821$$

Vì thế, phương sai (độ lệch chuẩn bình phương) bằng:

$$v_0 = (0,997)^2 = 0,675$$

Có thể thể hiện phân bố này bằng Biểu đồ 1 sau đây:



Biểu đồ 1. Phân bố xác suất tiền định thể hiện “niềm tin”, với giả thuyết khởi đầu là trung bình RR = 1 (tức logRR = 0) và độ lệch chuẩn 0,821. Diện tích tô màu xanh là khoảng tin cậy 95%, thể hiện xác suất RR < 0,2 là 2,5% và xác suất RR > 5 cũng 2,5%.

Dữ liệu thực tế

Dữ liệu thực tế được trình bày trong Bảng 1 và 2. Tuy nhiên, vì chúng ta làm việc với tỷ số nguy cơ qua đơn vị logarit, nên dữ liệu phải được trình bày lại dưới hình thức logRR và khoảng tin cậy 95%.

Để tiện theo dõi, tôi trình bày lại dữ liệu trong bảng số liệu dưới đây:

Nhóm can thiệp	Bệnh	Không bệnh	Tổng số
Bối cảnh chung			
Nhóm can thiệp	a	b	a+b
Nhóm không can thiệp	c	d	c+d
Nghiên cứu 1			
CaD	209 (2,48%)	8220	8429
Giả dược	168 (2,03%)	8121	8289
Nghiên cứu 2			
Strontium ranelate	46 (6,4%)	673	719
Giả dược	88 (12,2%)	635	723

Trước hết chúng ta cần phải ước tính RR, và sau đó là logRR. Gọi tỷ lệ mắc bệnh trong nhóm điều trị là p_1 và nhóm giả dược là p_0 , chúng ta có thể ước tính tỷ số nguy cơ RR như sau:

$$RR = \frac{p_1}{p_0}$$

Do đó, logRR (sẽ ký hiệu là m_1) có thể ước tính đơn giản là: $m_1 = \log(RR)$

Theo lý thuyết thống kê, phương sai của logRR (ký hiệu v_1) là:

$$v_1 = \frac{1/a}{1/(a+b)} + \frac{1/c}{1/(c+d)}$$

Cố nhiên, độ lệch chuẩn cũng có thể tính từ v_1 như sau: $s_1 = \sqrt{v_1}$

Dựa vào những công thức trên, chúng ta có thể ước tính m_1 và v_1 cho hai nghiên cứu trên như sau:

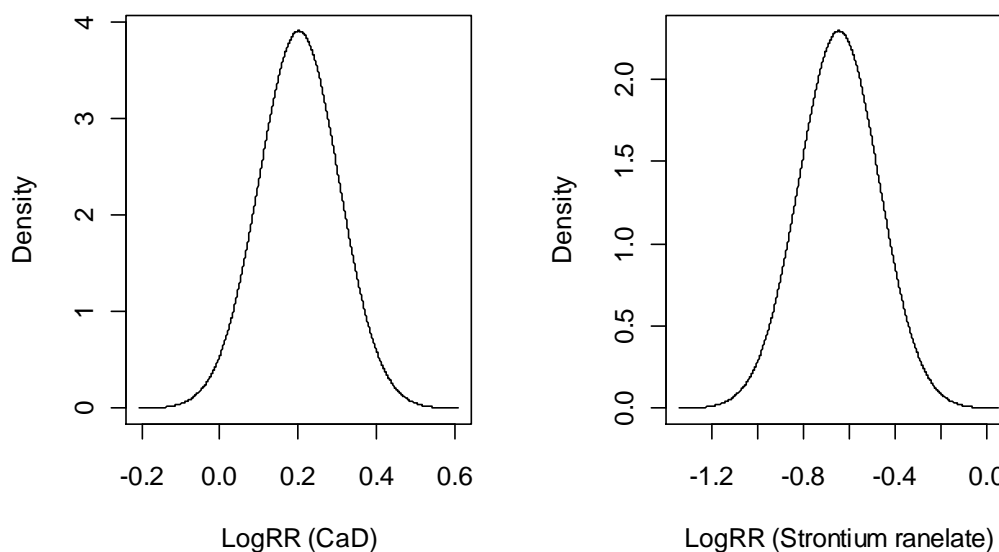
Nghiên cứu	Chỉ tiêu lâm sàng	m_1	v_1	s_1	KTC 95%
1. CaD và giả dược	Nhồi máu cơ tim	0,2016	0,0104	0,102	0,0008, 0,402
2. SR và giả dược	Gãy xương đốt sống	-0,6431	0,0303	0,174	-0,984, -0,302

Có thể thể hiện phân bố của dữ liệu thực tế của hai nghiên cứu trên qua hai Biểu đồ 2.

Từ kết quả trên, có thể hoán chuyển ngược lại đơn vị RR và khoảng tin cậy 95%. Khoảng tin cậy 95% của m_1 có thể ước tính đơn giản bằng lý thuyết phân phối chuẩn: tức là lấy số trung bình cộng hoặc trừ cho 1,96 lần độ lệch chuẩn:

$$KTC\ 95\% = m_1 \pm 1,96s_1$$

Kết quả trình bày trong cột sau cùng của bảng số liệu trên. Cố nhiên, bởi vì đơn vị logarit chỉ dùng cho tính toán, và trong thực tế chúng ta cần hoán chuyển sang đơn vị thường (tức RR) để có thể hiểu được. Do đó, có thể lấy hàm số mũ của các kết quả trên như sau (ví dụ cho nghiên cứu 1):



Biểu đồ 2. Phân bố logRR cho nghiên cứu 1 (trái, CaD và nhồi máu cơ tim) và nghiên cứu 2 (bên phải, strontium ranelate) và gãy xương đốt sống.

- RR (nhồi máu cơ tim) = $e^{0,2016} = 1,22$
- và khoảng tin cậy 95%: $e^{0,008}$ đến $e^{0,401} = 1,00$ đến $1,50$

Kết quả hoán chuyển có thể xem bảng dưới đây.

Nghiên cứu	Kết cục	RR	KTC 95%
1. CaD và giả dược	Nhồi máu cơ tim	1,22	1,00, 1,50
2. SR và giả dược	Gãy xương đốt sống	0,52	0,37, 0,74

Cần nói thêm rằng, đối với phân tích cổ điển, qui trình phân tích ngừng ở đây. Có lẽ nhà nghiên cứu kết luận rằng CaD tăng nguy cơ nhồi máu cơ tim 22% và ảnh hưởng này dao động trong khoảng 0% đến 50% ($P = 0,05$). Tương tự, trong nghiên cứu 2, strontium ranelate giảm nguy cơ gãy xương đốt sống 48%, nhưng mức độ ảnh hưởng có thể giảm từ 28% đến 63% ($P < 0,001$). Cả hai kết quả đều có ý nghĩa thống kê.

Xác suất hậu định

Nhưng nếu chịu khó suy nghĩ, chúng ta sẽ không hài lòng với kết luận đó. Trong trường hợp CaD và nhồi máu cơ tim, một tỉ số nguy cơ 1,22 có thể nói là một sự ảnh hưởng rất thấp, nhất là trong điều kiện khoảng tin cậy 95% có thể là không có ảnh hưởng (phần hạ của khoảng tin cậy 95% bằng 1). Nói cách khác, kết quả này quá “mong manh”. Chỉ cần thay 2 ca nhồi máu cơ tim trong nhóm chứng là kết quả sẽ không có ý nghĩa thống kê. Chúng ta hỏi:

có phải kết quả trên do ngẫu nhiên, hay là một kết quả dương tính giả.

Vấn đề thứ hai là kết quả trên có ý nghĩa lâm sàng hay không? Nên nhớ rằng kết quả trên dựa vào giả thuyết vô hiệu là nguy cơ mắc bệnh của hai nhóm bằng nhau. Tức là dù có khác biệt 0,1% vẫn được xem là có khác biệt. Nhưng trong thực tế 0,1% khác biệt chẳng có ý nghĩa lâm sàng nào cả. Do vậy, để trả lời câu hỏi này, chúng ta cần phải xác định ngưỡng tỉ số nguy cơ để có thể cho là có ý nghĩa lâm sàng. Theo y văn trong lĩnh vực tim mạch, phần lớn những nghiên cứu lâm sàng được thiết kế để chứng minh can thiệp giảm nguy cơ từ 15% trở lên. Một ngưỡng tương tự cũng thường hay thấy trong các nghiên cứu loãng xương. Do đó, chúng ta sẽ đề ra ngưỡng 15% để đánh giá tầm quan trọng lâm sàng. Nói cách khác, chúng ta có một tiêu chí mới:

- Nếu CaD (hay bất cứ can thiệp nào) tăng nguy cơ nhồi máu cơ tim cao hơn 15% (tức tỉ số nguy cơ cao hơn 1,15 hay $\log RR > 0,14$) sẽ được xem là có hại;
- Tương tự, nếu can thiệp nào giảm nguy cơ gãy xương đốt sống trên 15% (tức tỉ số nguy cơ thấp hơn 0,85, hay $\log RR < -0,16$) thì sẽ được xem là thật sự có ý nghĩa lâm sàng.

Phân tích Bayes nhằm trả lời những câu hỏi trên. Để trả lời câu hỏi trên, chúng ta cần phải ước tính xác suất hậu định (posterior probability) của một giả

thuyết. Thông số mà chúng ta quan tâm là tỉ số nguy cơ RR. Nhưng như đề cập trên, để cho tiện lợi trong tính toán, chúng ta sẽ dựa vào đơn vị logRR. Gọi số trung bình của logRR (hậu định) là m và phương sai của logRR là v . Định lý Bayes phát biểu rằng:

$$N(m, v) = N(m_0, v_0) \times N(m_1, v_1)$$

Trong đó N là viết tắt của phân phối chuẩn (normal distribution), và những thông số trong ngoặc là số trung bình và phương sai. Công thức trên hàm ý nói rằng phân phối logRR hậu định bằng phân phối logRR tiền định nhân cho phân bố logRR của dữ liệu thực tế. Sau vài thao tác toán, có thể chứng minh rằng:

$$m = \frac{\frac{m_0}{v_0} + \frac{m_1}{v_1}}{\frac{1}{v_0} + \frac{1}{v_1}} \quad \text{và} \quad v = \frac{1}{\frac{1}{v_0} + \frac{1}{v_1}}$$

Công thức trên cho thấy m chính là một trọng số trung bình (weighted average) của logRR tiền định và logRR thực tế. Trọng số ở đây là số đảo phương sai ($1/v_0$ và $1/v_1$). Phương sai nhỏ có trọng số lớn, và ngược lại, phương sai lớn (tức dữ liệu có dao động lớn) có trọng số nhỏ. Vì xác suất hậu định tuân theo luật phân phối chuẩn, cho nên 2 thông số trên có nghĩa là hàm phân phối chuẩn của xác suất hậu định là:

$$f(x) = \frac{1}{\sqrt{2\pi v}} \exp\left[-\frac{(x-m)^2}{2v}\right]$$

Trong đó, x là số logRR khả dĩ. Hàm $f(x)$ cung cấp một số ước tính rất có ích. Chẳng hạn như chúng ta có thể đặt câu hỏi: xác suất mà CaD tăng nguy cơ nhồi máu cơ tim trên 15% là bao nhiêu? Như nói trên, tăng nguy cơ trên 15% có nghĩa là $P(\log RR > 0.14)$. Do đó, đáp số cho câu trả lời là:

$$\Pr(\log RR > 0.14) = \int_{0.14}^{\infty} f(x)$$

Tuy công thức có vẻ phức tạp, nhưng thật ra, tất cả những tính toán có thể thực hiện bằng phần mềm Excel hay R một cách rất dễ dàng.

Quay lại với hai nghiên cứu trên, chúng ta có thể ước tính hai thông số hậu định m và v cho nghiên cứu 1 như sau:

$$m = \frac{\frac{0}{0.675} + \frac{0.2016}{0.0104}}{\frac{1}{0.675} + \frac{1}{0.0104}} = 0.197 \quad \text{và}$$

$$v = \frac{1}{\frac{1}{0.675} + \frac{1}{0.0104}} = 0.010$$

Một tính toán tương tự cho nghiên cứu 2 sẽ cho ra kết quả $m = -0,624$ và $v = 0,029$. Chúng ta có thể tóm lược 3 thông tin vừa phân tích trong bảng dưới đây:

Nghiên cứu	Tiền định		Dữ liệu		Hậu định	
	m_0	v_0	m_1	v_1	m	v
Đơn vị logarit						
1. CaD và giả được	0	0,675	0,2016	0,0104	0,197	0,010
2. SR và giả được	0	0,675	-0,6431	0,0303	-0,615	0,029

Kết quả trên cho thấy, logRR hậu định không thay đổi nhiều so với logRR thực tế. Lý do là vì chúng ta bắt đầu bằng thông tin tiền định với logRR = 0. Ngoài ra, vì phương sai của thông tin tiền định cao (phản ánh sự bất định về kiến thức của chúng ta), nên trọng số của thông tin tiền định không cao.

Dựa vào kết quả trên, chúng ta có thể ước tính xác suất logRR > 0,14 cho nghiên cứu 1 và logRR < -0,16 cho nghiên cứu 2. Kết quả là:

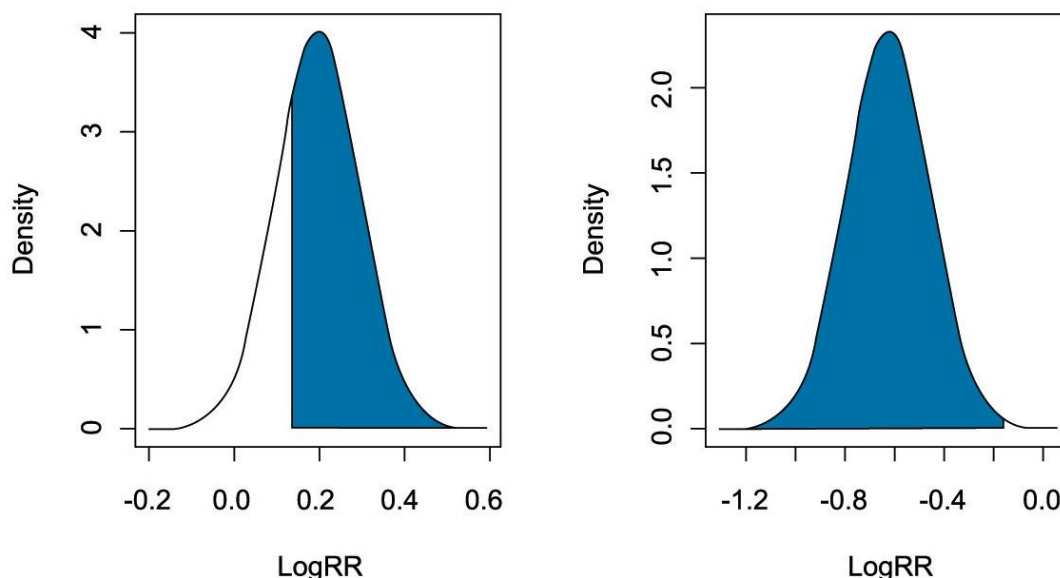
- Nghiên cứu 1: $P(\log RR > 0,14) = 0,725$
- Nghiên cứu 2: $P(\log RR < -0,16) = 0,996$

Nói cách khác, xác suất CaD tăng nguy cơ nhồi máu cơ tim trên 15% là 72,5%, và xác suất SR giảm nguy cơ gây xương đốt sống trên 15% là 99,6%. Cần nói thêm rằng, chỉ khi nào xác suất có hiệu quả lâm sàng trên 90% hoặc trên 95% mới có thể xem là đáng chú ý. Do đó, trong hai trường hợp trên, ảnh hưởng của CaD đến nguy cơ nhồi máu cơ tim không đáng chú ý, nhưng ảnh hưởng của SR đến nguy cơ gãy xương đốt sống là đáng chú ý. Có thể thể hiện hai xác suất trên đây qua Biểu đồ 3 (trang kế bên):

Xin nhắc lại, những kết quả trên là tính trên đơn vị logarit. Để dễ hiểu hơn, chúng ta cần phải hoán chuyển sang đơn vị RR bằng cách lấy hàm số mũ của logRR và khoảng tin cậy 95%.

Với nghiên cứu 1, logRR = 0,197, phương sai $v = 0,01$, nên độ lệch chuẩn là $s = \sqrt{0.01} = 0.1$. Từ đó, khoảng tin cậy 95% của logRR là $0,197 \pm 1,96 \times 0,10 = 0,001$ đến $0,393$. Nói cách khác, RR = $e^{0,197} = 1,22$, và khoảng tin cậy 95% là $e^{0,001} = 1,0$ đến $e^{0,393} = 1,48$. Kết quả hoán chuyển được trình bày trong bảng số liệu sau đây:

Nghiên cứu	m	s	RR	Khoảng tin cậy 95%
1. CaD và giả được	0,197	0,10	1,22	1,00 đến 1,48
2. SR và giả được	-0,615	0,17	0,54	0,39 đến 0,75



Biểu đồ 3. Phân bố logRR cho nghiên cứu 1 (trái, CaD và nhồi máu cơ tim) và nghiên cứu 2 (bên phải, strontium ranelate và gãy xương đốt sống. Diện tích màu xanh là xác suất có ý nghĩa lâm sàng (tức xác suất RR > 1,15 hay logRR > 0,14 cho nghiên cứu 1, và RR < 0,85 hay logRR < -0,16 cho nghiên cứu 2).

Những kết quả trên là bằng chứng để chúng ta có thể phát biểu rằng tỉ số nguy cơ nhồi máu cơ tim liên quan đến CaD là 1,22, và xác suất 95% của tỉ số nguy cơ là từ 1,0 đến 1,48. Tương tự, chúng ta cũng có thể kết luận rằng tỉ số nguy cơ gãy xương đốt sống liên quan đến strontium ranelate là 0,54 và xác suất 95% là tỉ số này sẽ dao động trong khoảng 0,39 đến 0,75. Tuy nhiên, chỉ có ảnh hưởng của strontium ranelate đến nguy cơ gãy xương đốt sống là có ý nghĩa lâm sàng, còn ảnh hưởng của CaD đến nguy cơ nhồi máu cơ tim thì vẫn còn bất định, chưa thể kết luận một cách dứt khoát như nhóm tác giả viết.

Xem xét kỹ kết quả phân tích trên, chúng ta có thể thấy rằng thông tin tiên định đóng vai trò quan trọng. Đó cũng là một khía cạnh của phương pháp Bayes bị phê bình khá nhiều. Những phê phán có thể tóm lược trong hai điểm chính: thông tin tiên định đến từ đâu, và thông tin tiên định có thể không khách quan. Cả hai phê phán đều hợp lý, nhưng cũng... không hợp lý. Hợp lý là vì việc xác định thông tin tiên định rất khó. Khó hơn là làm sao để định lượng khả năng của một giả thuyết. Tuy nhiên không hợp lý là vì trong thực tế, chúng ta có thể có khá nhiều thông tin trước khi thực hiện một nghiên cứu lâm sàng, và do đó, không thể nói rằng không có thông tin. Và lại, nếu không có thông tin về khả năng của một giả thuyết, chúng ta có thể khởi đầu bằng xác suất 50/50, tức hoàn toàn khách quan. Thực ra, nếu một nghiên cứu lâm sàng (như đánh giá hiệu quả thuốc) mà khả năng thuốc có hiệu quả

hơn 50% là không có lý do gì để có nhóm chứng! Do đó, tất cả các nghiên cứu lâm sàng đối chứng ngẫu nhiên phải bắt đầu bằng xác suất tiên định là 50%, và đó cũng chính là một thông tin cho phân tích Bayes. Chính vì thế mà có người cho rằng không có lý do gì dữ liệu nghiên cứu lâm sàng không sử dụng phương pháp phân tích Bayes.

Phân tích Bayes đòi hỏi tính toán phức tạp hơn phương pháp phân tích theo trường phái tần số. Có nhiều vấn đề quá phức tạp không/chưa ai có thể giải được! Tuy nhiên, một điều rất may mắn là với sự phát triển của máy tính và thuật toán MCMC (Markov Chain Monte Carlo), hầu như bất cứ vấn đề nào cũng có thể giải đáp được (mà toán học không giải được). Do đó, phương pháp Bayes càng ngày càng được ứng dụng trong nghiên cứu lâm sàng với nhiều thành công ngoạn mục. Người ta tiên đoán rằng phương pháp Bayes sẽ “thống trị” khoa học trong thế kỷ 21, và phương pháp tần số sẽ dần dần lùi vào hậu trường do những khiếm khuyết cơ bản của nó. Kinh nghiệm của cá nhân chúng tôi cho thấy phương pháp Bayes giải quyết được nhiều vấn đề nghiên cứu và cung cấp những câu trả lời mà trước đây còn tranh cãi,⁽¹⁰⁻¹²⁾ nhất là vấn đề cá nhân hóa trong tiên lượng y khoa.⁽¹³⁾

Tóm lại, suy luận khoa học theo phương pháp Bayes là một suy luận dựa vào logic thông thường, mà bất cứ bác sĩ lâm sàng nào hay thậm chí người ngoài khoa học nào cũng nắm được. Đó là một cách tích lũy kiến thức. Những gì chúng ta đã biết (thông

tin tiên định) cộng với những gì chúng ta thu thập được (dữ liệu thực tế) giúp cho chúng ta cũng cố kiến thức tốt hơn (thông tin hậu định). Trong nghiên cứu lâm sàng, chúng ta muốn có câu trả lời cho 2 câu hỏi: (a) với những dữ liệu quan sát được, xác suất có ảnh hưởng thật là bao nhiêu, và (b) xác suất ảnh hưởng có ý nghĩa lâm sàng là bao nhiêu. Chỉ có phương pháp Bayes trả lời được hai câu hỏi này. Phương pháp phân tích theo tần số (frequentist) như trị số P không thể cung cấp câu trả lời cho hai câu hỏi trên. Chính vì thế mà suy luận theo phương pháp Bayes ngày càng được ứng dụng trong tất cả các lĩnh vực khoa học và công nghệ. Hi vọng những mô tả trên đây đã cung cấp một số thông tin và phương pháp có ích trong việc đánh giá tầm quan trọng và ý nghĩa của một nghiên cứu lâm sàng.

Tài liệu tham khảo

- (1) Bolland MJ, Grey A, Avenell A, Gamble GD, Reid IR. Calcium supplements with or without vitamin D and risk of cardiovascular events: reanalysis of the Women's Health Initiative limited access dataset and meta-analysis. *BMJ* 2011;342:d2040.
- (2) Meunier PJ, et al. The Effects of Strontium Ranelate on the Risk of Vertebral Fracture in Women with Postmenopausal Osteoporosis. *N Engl J Med* 2004;350:459-68.
- (3) Senn S. Falsificationism and clinical trials. *Stat Med* 1991;10:1361-74.
- (4) Fisher RA. *Statistical Methods for Research Workers*, Edinburgh: Oliver and Boyd, 1925. Xem thêm bài: *Statistical Tests*. *Nature* 1935;136:474.
- (5) Neyman J, Pearson ES. On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference. Part I. *Biometrika* 1928;20A:175-240.
- (6) Goodman SN. A Dirty Dozen: Twelve P -Value Misconceptions. *Semin Hematol* 2008; 45:135-140.
- (7) Nelder JA. From Statistics to Statistical Science. *J Royal Statist Soc: Series D* 1999; 48(2):257-269.
- (8) Ioannidis JAP. Why most published research findings are false. *PLoS Med* 2005;2(8): e12
- (9) Bayes, Thomas; Price, Mr. (1763). An Essay towards solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London* 53 (0): 370-418. Có thể xem bản gốc của bài báo quan trọng này tại địa chỉ: <http://www.stat.rice.edu/~blairc/seminar/Files/danTalk.pdf>
- (10) Nguyen TV. Interpretation of Randomized Controlled Trials of Fracture Prevention. *IBMS BoneKEy*, August 1, 2009; 6(8): 279 - 294.
- (11) Nguyen ND, Wang CY, Eisman JA, Nguyen TV. On the association between statin and fracture: a Bayesian consideration. *Bone* 2007 Apr;40(4):813-20.
- (12) Nguyen ND, Eisman JA, Nguyen TV. Anti-hip fracture efficacy of biophosphonates: a Bayesian analysis of clinical trials. *J Bone Miner Res* 2006;21(2):340-9.
- (13) Nguyen TV. Individualization of osteoporosis risk. *Osteoporos Int* 2007 Sep;18(9):1153-6.

Chú thích kỹ thuật

Mã R để vẽ biểu đồ cho dữ liệu tiên định

```
par(mfrow=c(1,1))
mean=0; sd=0.821; lb=mean-1.96*sd;
```

```
ub=mean+1.96*sd
x = seq(-4, 4, length=10000)*sd + mean
hx = dnorm(x, mean, sd)
plot(x, hx, type="n", xlab="LogRR",
ylab="Density")
i <- x >= lb & x <= ub
lines(x, hx)
polygon(c(lb,x(i),ub), c(0,hx(i),0),
col="red")
# In terms of RR
mean=0; sd=0.821; lb=mean-1.96*sd;
ub=mean+1.96*sd
x = seq(-4, 4, length=10000)*sd + mean
hx = dnorm(x, mean, sd)
plot(exp(x), hx, type="n", xlab="LogRR",
ylab="Density")
lines(exp(x), hx)
```

Mã R để vẽ biểu đồ cho dữ liệu thực tế

```
par(mfrow=c(1,2))
mean=0.2016; sd=0.102; lb=mean-1.96*sd;
ub=mean+1.96*sd
x = seq(-4,4,length=1000)*sd + mean
hx = dnorm(x,mean,sd)
plot(x, hx, type="n", xlab="LogRR (CaD)",
ylab="Density", main=" ")
lines(x, hx)
```

```
mean=-0.6431; sd=0.174; lb=mean-1.96*sd;
ub=mean+1.96*sd
x = seq(-4,4,length=1000)*sd + mean
hx = dnorm(x,mean,sd)
plot(x, hx, type="n", xlab="LogRR (Strontium
ranelate)", ylab="Density", main=" ")
lines(x, hx)
```

Mã R để vẽ biểu đồ thông tin hậu định.

```
par(mfrow=c(1,2))
# plot for CaD and MI
mean=0.197; sd=sqrt(0.01); lb=mean-1.96*sd;
ub=mean+1.96*sd
crit=log(1.15)
x = seq(-4, 4, length=10000)*sd + mean
hx = dnorm(x, mean, sd)
plot(x, hx, type="n", xlab="LogRR",
ylab="Density")
i = x > crit
lines(x, hx)
polygon(c(crit, x(i), ub), c(0,hx(i),0),
col="red")
# plot for SR and vert fx
mean=-0.615; sd=sqrt(0.029); lb=mean-1.96*sd;
ub=mean+1.96*sd
crit=log(0.85)
x = seq(-4, 4, length=10000)*sd + mean
hx = dnorm(x, mean, sd)
plot(x, hx, type="n", xlab="LogRR",
ylab="Density")
i = x < crit
lines(x, hx)
polygon(c(crit, x(i), crit), c(0, hx(i),0),
col="red")
```