14
Phân tích tổng hợp

Một vấn đề khoa học cần đến nhiều nghiên cứu. Một nghiên cứu riêng lẻ không thể giải quyết hay cung cấp câu trả lời dứt khoát cho một vấn đề khoa học. Nhu cầu lặp lại nghiên cứu trong điều kiện khác nhau rất quan trọng trong hoạt động khoa học. Trong nghiên cứu khoa học nói chung và y học nói riêng, nhiều khi chúng ta cần phải xem xét nhiều kết quả nghiên cứu từ nhiều nguồn khác nhau để giải quyết một vấn đề cụ thể.

14.1 Nhu cầu cho phân tích tổng hợp

Trong những năm gần đây, trong nghiên cứu khoa học xuất hiện khá nhiều nghiên cứu dưới danh mục "meta-analysis", tạm dịch là "phân tích tổng hợp". Vậy phân tích tổng hợp là gì, mục đích, và cách tiến hành ra sao là những câu hỏi mà rất nhiều bạn đọc muốn biết. Chương này tôi sẽ mô tả vài khái niệm và cách tiến hành một phân tích tổng hợp, với hi vọng bạn đọc có thể tự mình làm một phân tích mà không cần đến các phần mềm đắt tiền.

Nguồn gốc và ý tưởng tổng hợp dữ liệu khởi đầu từ thế kỉ 17. Thời đó, các nhà thiên văn học nghĩ rằng cần phải hệ thống hóa dữ liệu từ nhiều nguồn để có thể đi đến một quyết định chính xác và hợp lí hơn các nghiên cứu riêng lẻ. Nhưng phương pháp phân tích tổng hợp hiện đại phải nói là bắt đầu từ hơn nửa thế kỉ trước trong ngành tâm lí học. Năm 1952, nhà tâm lí học Hans J. Eysenck tuyên bố rằng tâm lí trị liệu (psychotherapy) chẳng có hiệu quả gì cả. Hơn hai mươi năm sau, năm 1976, Gene V. Glass, một nhà tâm lí học người Mĩ, muốn chứng minh rằng Eysenck sai, nên ông tìm cách thu thập dữ liệu của hơn 375 nghiên cứu về tâm lí trị liệu trong quá khứ, và tiến hành tổng hợp chúng bằng một phương pháp mà ông đặt tên là "meta-analysis" [1]. Qua phương pháp phân tích này, Glass tuyên bố rằng tâm lí trị liệu có hiệu quả và giúp ích cho bệnh nhân.

Phân tích tổng hợp – hay meta-analysis – từ đó được các bộ môn khoa học khác, nhất là y học, ứng dụng để giải quyết các vấn đề như hiệu quả của thuốc trong việc điều trị bệnh nhân. Cho đến nay, các phương pháp phân tích tổng hợp đã phát triển một bước dài, và trở thành một phương pháp chuẩn để thẩm định các vấn đề gai góc, các vấn đề mà sự nhất trí giữa các nhà khoa học vẫn chưa đạt được. Có người xem phân tích tổng hợp có thể cung cấp một câu trả lời sau cùng cho một câu hỏi y học. Tuy phát biểu này quá lạc quan, nhưng

phân tích tổng hợp là một phương pháp rất có ích cho chúng ta giải quyết những vấn đề còn trong vòng tranh cãi. Phân tích tổng hợp cũng có thể giúp cho chúng ta nhận ra những lĩnh vực nào cần phải nghiên cứu thêm hay cần thêm bằng chứng.

Kết quả của mỗi nghiên cứu đơn lẻ thường được đánh giá hoặc là "tích cực" (tức là, chẳng hạn như, thuật điều trị có hiệu quả), hoặc là "tiêu cực" (tức là thuật điều trị không có hiệu quả), và sự đánh giá này dựa vào trị số P. Thuật ngữ tiếng Anh gọi qui trình đó là "significance testing" – thử nghiệm ý nghĩa thống kê. Nhưng ý nghĩa thống kê tùy thuộc vào số mẫu được chọn trong nghiên cứu, và một kết quả "tiêu cực" không có nghĩa là giả thiết của nghiên cứu sai, mà có thể đó là tín hiệu cho thấy số lượng mẫu chưa đầy đủ để đi đến một kết luận đáng tin cậy. Cái logic của phân tích tổng hợp, do đó, là chuyển hướng từ **significance testing** sang ước tính **effect size** - mức độ ảnh hưởng. Câu trả lời mà phân tích tổng hợp muốn đưa ra không chỉ đơn giản là có hay không có ý nghĩa thống kê (significant hay insignificant) mà là mức độ ảnh hưởng bao nhiêu, có đáng để chúng ta quan tâm, có thích hợp để chúng ta ứng dụng vào thực tế hay không.

14.2 Fixed-effects và Random-effects

Hai thuật ngữ mà bạn đọc thường gặp trong các phân tích tổng hợp là fixed-effects (tạm dịch là **ảnh hưởng bất biến**) và random-effects (**ảnh hưởng biến thiên**). Để hiểu hai thuật ngữ này chúng ta sẽ xem xét một ví dụ tương đối đơn giản. Hãy tưởng tượng chúng ta muốn ước tính chiều cao của người Việt Nam trong độ tuổi trưởng thành (18 tuổi trở lên). Chúng ta có thể tiến hành 100 nghiên cứu tại nhiều địa điểm khác nhau trên toàn quốc; mỗi nghiên cứu chọn mẫu (samples) một cách ngẫu nhiên từ 10 người đến vài chục ngàn người; và cứ mỗi nghiên cứu chúng ta tính toán chiều cao trung bình. Như vậy, chúng ta có 100 số trung bình, và chắc chắn những con số này không giống nhau: một số nghiên cứu có chiều cao trung bình thấp, cao hay … trung bình. Phân tích tổng hợp là nhằm mục đích sử dụng 100 số trung bình đó để ước tính chiều cao cho toàn thể người Việt. Có hai cách để ước tính: fixed-effects meta-analysis (phân tích tổng hợp ảnh hưởng bất biến) và random-effects meta-analysis (phân tích tổng hợp ảnh hưởng bất biến) [2].

Phân tích tổng hợp ảnh hưởng bất biến xem sự khác biệt giữa 100 con số trung bình đó là do các yếu tố ngẫu nhiên liên quan đến mỗi nghiên cứu (còn gọi là within-study variance) gây nên. Cái giả định đằng sau cách nhận thức này là: nếu 100 nghiên cứu đó đều được tiến hành giống nhau (như có cùng số lượng đối tượng, cùng độ tuổi, cùng tỉ lệ giới tính, cùng chế độ dinh dưỡng, v.v…) thì sẽ không có sự khác biệt giữa các số trung bình.

Nếu chúng ta gọi số trung bình của 100 nghiên cứu đó là $x_1, x_2, \ldots, x_{100}$, quan điểm của phân tích tổng hợp ảnh hưởng bất biến cho rằng mỗi x_i là một biến số gồm hai phần: một phần phản ánh số trung của toàn bộ quần thể dân số (tạm gọi là *M*), và phần còn lại (khác biệt giữa x_i và *M* là một biến số e_i. Nói cách khác:

$$x_1 = M + e_1$$
$$x_2 = M + e_2$$
$$\ldots.$$
$$x_{100} = M + e_{100}$$

Hay nói chung là:

$$x_i = M + e_i$$

Tất nhiên e_i có thể <0 hay >0. Nếu *M* và e_i độc lập với nhau (tức không có tương quan gì với nhau) thì phương sai của x_i (gọi là $\text{var}[x_i]$) có thể viết như sau:

$$\text{var}[x_i] = \text{var}[M] + \text{var}[e_i] = 0 + s_e^2$$

Chú ý var[*M*] = 0 vì *M* là một hằng số bất biến, s_e^2 là phương sai của e_i. Mục đích của phân tích tổng hợp là ước tính *M* và s_e^2.

Phân tích tổng hợp ảnh hưởng biến thiên xem mức độ khác biệt (còn gọi là variance hay phương sai) giữa các số trung bình là do hai nhóm yếu tố gây nên: các yếu tố liên quan đến mỗi nghiên cứu (within-study variance) và các yếu tố giữa các nghiên cứu (between-study variance). Các yếu tố khác biệt giữa các nghiên cứu như địa điểm, độ tuổi, giới tính, dinh dưỡng, v.v… cần phải được xem xét và phân tích. Nói cách khác, phân tích tổng hợp ảnh hưởng biến thiên đi xa hơn phân tích tổng hợp ảnh hưởng bất biến một bước bằng cách xem xét đến những khác biệt giữa các nghiên cứu. Do đó, kết quả từ phân tích tổng hợp ảnh hưởng biến thiên thường "bảo thủ" hơn các phân tích tổng hợp ảnh hưởng bất biến.

Quan điểm của phân tích tổng hợp ảnh hưởng biến thiên cho rằng mỗi nghiên cứu có một giá trị trung bình cá biệt phải ước tính, gọi là m_i. Do đó, x_i là một biến số gồm hai phần: một phần phản ánh số trung bình của quần thể

mà mẫu được chọn (m_i , chú ý ở đây có chỉ từ *i* để chỉ một nghiên cứu riêng lẻ *i*), và phần còn lại (khác biệt giữa x_i và m_i là một biến số e_i . Ngoài ra, phân tích tổng hợp ảnh hưởng biến thiên còn phát biểu rằng m_i dao động chung quanh số tổng trung bình M bằng một biến ngẫu nhiên ε_i . Nói cách khác:

$$x_i = m_i + e_i$$

Trong đó:

$$m_i = M + \varepsilon_i$$

Do đó:

$$x_i = M + \varepsilon_i + e_i$$

Và phương sai của x_i bây giờ có hai thành phần:

$$\text{var}[x_i] = \text{var}[M] + \text{var}[\varepsilon_i] + \text{var}[e_i] = 0 + s_\varepsilon^2 + s_e^2$$

Như ta thấy qua công thức này, s_ε^2 phản ánh độ dao động giữa các nghiên cứu (between-study variation), còn s_e^2 phản ánh độ dao động trong mỗi nghiên cứu (within-study variation). Mục đích của phân tích tổng hợp ảnh hưởng biến thiên là ước tính *M,* s_e^2 và s_ε^2.

Nói tóm lại, Phân tích tổng hợp ảnh hưởng bất biến và Phân tích tổng hợp ảnh hưởng biến thiên chỉ khác nhau ở phương sai. Trong khi phân tích tổng hợp bất biến xem s_ε^2 = 0, thì phân tích tổng hợp biến thiên đặt yêu cầu phải ước tính s_ε^2. Tất nhiên, nếu s_ε^2 = 0 thì kết quả của hai phân tích này giống nhau. Trong bài này tôi sẽ tập trung vào cách phân tích tổng hợp ảnh hưởng bất biến.

14.3 Qui trình của một phân tích tổng hợp

Cũng như bất cứ nghiên cứu nào, một phân tích tổng hợp được tiến hành qua các công đoạn như: thu thập dữ liệu, kiểm tra dữ liệu, phân tích dữ liệu, và kiểm tra kết quả phân tích.

- Bước thứ nhất: sử dụng hệ thống thư viện y khoa PubMed hay một hệ thống thư viện khoa học của chuyên ngành để tìm những bài báo liên quan đến vấn đề cần nghiên cứu. Bởi vì có nhiều nghiên cứu, vì lí do nào đó (như kết quả "tiêu cực" chẳng hạn), không được công bố, cho nên nhà nghiên cứu có khi cũng cần phải thu thập các nghiên cứu đó. Việc làm này tuy nói thì dễ, nhưng trong thực tế không dễ dàng chút nào!

- Bước thứ hai: rà soát xem trong số các nghiên cứu được truy tìm đó, có bao nhiêu đạt các tiêu chuẩn đã được đề ra. Các tiêu chuẩn này có thể là đối tượng bệnh nhân, tình trạng bệnh, độ tuổi, giới tính, tiêu chí, v.v… Chẳng hạn như trong số hàng trăm nghiên cứu về ảnh hưởng của viatmin D đến loãng xương, có thể chỉ vài chục nghiên cứu đạt tiêu chuẩn như đối tượng phải là phụ nữ sau thời mãn kinh, mật độ xương thấp, phải là nghiên cứu lâm sàng đối chứng ngẫu nhiên (randomized controlled clinical trials - RCT), tiêu chí phải là gãy xương đùi, v.v… (Những tiêu chuẩn này phải được đề ra trước khi tiến hành nghiên cứu).

- Bước thứ ba: chiết số liệu và dữ kiện (data extraction). Sau khi đã xác định được đối tượng nghiên cứu, bước kế tiếp là phải lên kế hoạch chiết số liệu từ các nghiên cứu đó. Chẳng hạn như nếu là các nghiên cứu RCT, chúng ta phải tìm cho được số liệu cho hai nhóm can thiệp và đối chứng. Có khi các số liệu này không được công bố hay trình bày trong bài báo, và trong trường hợp đó, nhà nghiên cứu phải trực tiếp liên lạc với tác giả để tìm số liệu. Một bảng tóm lược kết quả nghiên cứu có thể tương tự như **Bảng 1** dưới đây.

- Bước thứ tư: tiến hành phân tích thống kê. Trong bước này, mục đích là ước tính mức độ ảnh hưởng chung cho tất cả nghiên cứu và độ dao động của ảnh hưởng đó. Phần dưới đây sẽ giải thích cụ thể cách làm.

- Bước thứ năm: xem xét các kết quả phân tích, và tính toán thêm một số chỉ tiêu khác để đánh giá độ tin cậy của kết quả phân tích.

Cũng như phân tích thống kê cho từng nghiên cứu riêng lẻ tùy thuộc vào loại tiêu chí (như là biến số liên tục – continuous variables – hay biến số nhị phân – dichotomous variables), phương pháp phân tích tổng hợp cũng tùy thuộc vào các tiêu chí của nghiên cứu. Chúng ta sẽ lần lược mô tả hai phương pháp chính cho hai loại biến số liên tục và nhị phân.

14.4 Phân tích tổng hợp ảnh hưởng bất biến cho một tiêu chí liên tục (Fixed-effects meta-analysis for a continuous outcome).

14.4.1 Phân tích tổng hợp bằng tính toán "thủ công"

Ví dụ 1. Thời gian nằm viện để điều trị ở các bệnh nhân đột quị là một tiêu chí quan trọng trong việc vạch định chính sách tài chính. Các nhà nghiên cứu muốn biết sự khác biệt về thời gian nằm viện giữa hai nhóm bệnh viện chuyên khoa và bệnh viện đa khoa. Các nhà nghiên cứu ra soát và thu thập số liệu từ 9 nghiên cứu như sau (xem **Bảng 1**). Một số nghiên cứu cho thấy thời gian nằm viện trong các bệnh viện chuyên khoa ngắn hơn các bệnh viện đa khoa (như nghiên cứu 1, 2, 3, 4, 5, 8), một số nghiên cứu khác cho thấy ngược lại (như nghiên cứu 7 và 9). Vấn đề đặt ra là các số liệu này có phù hợp với giả thiết bệnh nhân các bệnh viện chuyên khoa thường có thời gian nằm viện ngắn hơn các bệnh viện đa khoa hay không. Chúng ta có thể trả lời câu hỏi này qua các bước sau đây:

Bước 1: tóm lược dữ liệu trong một bảng thống kê như sau:

Bảng 1. Thời gian nằm bệnh viện của các bệnh nhân đột quị trong hai nhóm bệnh viện chuyên khoa và đa khoa

Nghiên cứu (*i*)	Bệnh viện chuyên khoa			Bệnh viện đa khoa		
	N_{1i}	LOS_{1i}	SD_{1i}	N_{2i}	LOS_{2i}	SD_{2i}
1	155	55	47	156	75	64
2	31	27	7	32	29	4
3	75	64	17	71	119	29
4	18	66	20	18	137	48
5	8	14	8	13	18	11
6	57	19	7	52	18	4
7	34	52	45	33	41	34
8	110	21	16	183	31	27
9	60	30	27	52	23	20
Tổng cộng	548			610		

Chú thích: Trong bảng này, *i* là chỉ số chỉ mỗi nghiên cứu, *i*=1,2,…,9. N_1 và N_2 là số bệnh nhân nghiên cứu cho từng nhóm bệnh viện; LOS_1 và LOS_2 (length of stay): thời gian trung bình nằm viện (tính bằng ngày); SD_1 và SD_2: độ lệch chuẩn (standard deviation) của thời gian nằm viện.

Bước 2: ước tính mức độ khác biệt trung bình và phương sai (variance) cho từng nghiên cứu. Mỗi nghiên cứu ước tính một độ ảnh hưởng, hay nói chính xác hơn là khác biệt về thời gian nằm viện kí hiệu, và chúng ta sẽ đặt kí hiệu là d_i. Chỉ số ảnh hưởng này chỉ đơn giản là:

$$d_i = LOS_{1i} - LOS_{2i}$$

Phương sai của d_i (tôi sẽ kí hiệu là s_i^2) được ước tính bằng một công thức chuẩn dựa vào độ lệch chuẩn và số đối tượng trong từng nghiên cứu. Với mỗi nghiên cứu i (i = 1, 2, 3, …, 9), chúng ta có:

$$s_i^2 = \frac{(N_{1i}-1)SD_{1i}^2 + (N_{2i}-1)SD_{2i}^2}{N_{1i}+N_{2i}-2}\left(\frac{1}{N_{1i}}+\frac{1}{N_{2i}}\right)$$

Chẳng hạn như với nghiên cứu 1, chúng ta có:

$$d_1 = 75 - 55 = 20$$

và phương sai của d_1:

$$s_1^2 = \frac{(155-1)(47)^2 + (156-1)(64)^2}{155+156-2}\left(\frac{1}{155}+\frac{1}{156}\right) = 40.59$$

hay độ lệch chuẩn: $s_1 = \sqrt{40.59} = 6.37$

Với độ lệch chuẩn s_i chúng ta có thể ước tính khoảng tin cậy 95% (95% confidence interval hay 95%CI) cho d_i bằng lí thuyết phân phối chuẩn (Normal distribution). Cần nhắc lại rằng, nếu một biến số tuân theo định luật phân phối chuẩn thì 95% các giá trị của biến số sẽ nằm trong khoảng ±1,96 lần độ lệch chuẩn. Do đó, khoảng tin cậy 95% cho mức độ khác biệt của nghiên cứu 1 là:

$$d_i - 1.96*s_i = 20 - 1.96*6.37 = 7.71 \text{ ngày}$$

đến

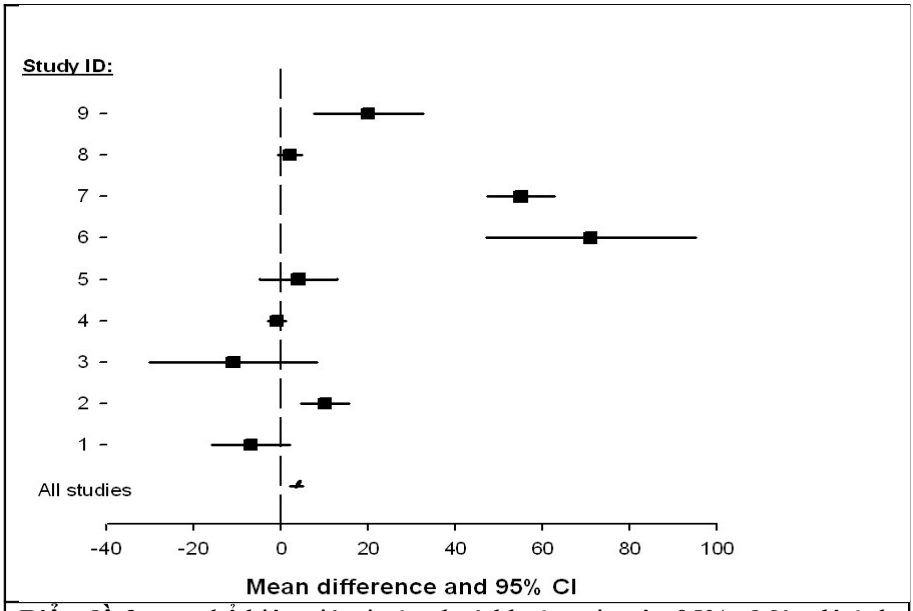
$$d_i + 1.96*s_i = 20 + 1.96*6.37 = 32.49 \text{ ngày}$$

Tiếp tục tính như thế cho các nghiên cứu khác, chúng ta sẽ có thêm bốn cột trong bảng sau đây:

Bảng 1a. Độ khác biệt về thời gian giữa hai nhóm và khoảng tin cậy 95%

Nghiên cứu (*i*)	d_i	s_i^2	s_i	d_i-1.96*s_i	d_i+1.96*s_i
1	20	40.6	6.37	7.51	32.49
2	2	2.0	1.43	-0.80	4.80
3	55	15.3	3.91	47.34	62.66
4	71	150.2	12.26	46.98	95.02
5	4	20.2	4.49	-4.81	12.81
6	-1	1.2	1.11	-3.17	1.17
7	-11	95.4	9.77	-30.14	8.14
8	10	8.0	2.83	4.45	15.55
9	-7	20.7	4.55	-15.92	1.92

Đến đây chúng ta có thể thể hiện mức độ ảnh hưởng d_i và khoảng tin cậy 95% trong một biểu đồ có tên là "**forest plot**" như sau:



Biểu đồ forest thể hiện giá trị của d_i và khoảng tin cậy 95%. Mức độ ảnh hưởng d_i ghi nhận từ nghiên cứu 5, 7 và 9 được xem là không có ý nghĩa thống kê, vì khoảng tin cậy 95% vượt qua cột mốc 0.

Bước 3: **ước tính "trọng số" (weight) cho mỗi nghiên cứu.** Trọng số (W_i) thực ra chỉ là số đảo của phương sai s_i^2 ,

$$W_i = 1/s_i^2$$

Chẳng hạn như với nghiên cứu 1, chúng ta có: $W_1 = \frac{1}{40.59} = 0.0246$

Và chúng ta có thêm một cột mới cho bảng trên như sau:

Bảng 1b. Trọng số (weight) cho từng nghiên cứu

Nghiên cứu	d_i	s_i^2	W_i
1	20	40.6	0.0246
2	2	2.0	0.4886
3	55	15.3	0.0654
4	71	150.2	0.0067
5	4	20.2	0.0495
6	-1	1.2	0.8173
7	-11	95.4	0.0105
8	10	8.0	0.1245
9	-7	20.7	0.0483
Tổng số			1.6354

Bước 4: **ước tính trị số trung bình của *d* cho tất cả các nghiên cứu.** Chúng ta có thể đơn giản tính trung bình *d* bằng cách cộng tất cả d_i và chia cho 9, nhưng cách tính như thế không khách quan, bởi vì mỗi giá trị d_i có một phương sai và trọng số (W_i) cá biệt. Chẳng hạn như nghiên cứu 4, vì phương sai cao nhất (150.2), chứng tỏ rằng nghiên cứu này có số đối tượng ít hay độ dao động rất cao, và độ dao động cao có nghĩa là chúng ta không đặt "niềm tin cậy" vào đó cao được. Chính vì thế mà trọng số cho nghiên cứu này rất thấp, chỉ 0.0067. Ngược lại, nghiên cứu 6 có trọng số cao vì độ dao động thấp (phương sai thấp) và ước tính ảnh hưởng của nghiên cứu này có "trọng lượng" hơn các nghiên khác trong nhóm.

Do đó, để tính trung bình *d* cho tổng số nghiên cứu, chúng ta phải xem xét đến trọng số W_i. Với mỗi d_i và W_i chúng ta có thể tính trị số **trung bình trọng số** (**weighted mean**) theo phương pháp chuẩn như sau:

$$d = \frac{\sum_{i=1}^{9} W_i d_i}{\sum_{i=1}^{9} W_i}$$

Bất cứ một ước tính thống kê (estimate) nào cũng phải có một phương sai. Và trong trường hợp *d,* phương sai (kí hiệu là s_d^2) chỉ đơn giản là số đảo của tổng trọng số W_i:

$$s_d^2 = \frac{1}{\sum_{i=1}^{9} W_i}$$

Sai số chuẩn (standard error, SE) của *d*, do đó là: $SE(d) = s_d$. Theo lí thuyết phân phối chuẩn (Normal distribution), khoảng tin cậy 95% (95% confidence interval, 95%CI) có thể được ước tính như sau:

$$95\%\text{CI của } d = d \pm 1.96\left(s_d\right)$$

Để tính *d* chúng ta cần thêm một cột nữa: đó là cột $W_i d_i$. Chẳng hạn như với nghiên cứu 1, chúng ta có $W_1 d_1 = 0{,}0246 \times 20 = 0{,}4928$. Tiếp tục như thế, chúng ta có thêm một cột.

Bảng 1c. Tính toán trị số trung bình

Nghiên cứu	d_i	s_i^2	W_i	$W_i d_i$
1	20	40.6	0.0246	0.4928
2	2	2.0	0.4886	0.9771
3	55	15.3	0.0654	3.5993
4	71	150.2	0.0067	0.4726
5	4	20.2	0.0495	0.1981
6	-1	1.2	0.8173	-0.8173
7	-11	95.4	0.0105	-0.1153
8	10	8.0	0.1245	1.2450
9	-7	20.7	0.0483	-0.3383
Tổng số			**1.6354**	**5.7140**

Sau đó, cộng tất cả W_i và $W_i d_i$ (trong hàng "Tổng số" của bảng trên). Như vậy, trị số trung bình trọng số của *d* là:

$$d = \frac{\sum_{i=1}^{9} W_i d_i}{\sum_{i=1}^{9} W_i} = \frac{0.4928 + 0.9771 + \ldots - 0.3383}{0.0246 + 0.4886 + \ldots + 0.0483} = \frac{5.7140}{1.6354} = 3.49.$$

Và phương sai của *d* là: $s_d^2 = \frac{1}{1.6345} = 0.61$.

Nói cách khác, sai số chuẩn (standard error) của *d* là:

$$s_d = \sqrt{0.61} = 0.782.$$

Khoảng tin cậy 95% (95% confidence interval hay 95%CI) có thể được ước tính như sau:

$$3.49 \pm 1{,}96*0.782 = 1.96 \text{ đến } 5.02.$$

Đến đây, chúng ta có thể nói rằng, tính trung bình, thời gian nằm viện tại các bệnh viện đa khoa dài hơn các bệnh viện chuyên khoa 3.49 ngày và 95% khoảng tin cậy là từ 1.96 ngày đến 5.02 ngày.

Bước 5: **ước tính chỉ số đồng nhất (homogeneity) và bất đồng nhất (heterogeneity) giữa các nghiên cứu [3].** Trong thực tế, đây là chỉ số đo lường độ khác biệt giữa mỗi nghiên cứu và trị số trung bình trọng số. Chỉ số đồng nhất (index of homogeneity) được tính theo công thức sau đây:

$$Q = \sum_{i=1}^{k} W_i (d_i - d)^2$$

Ở đây, *k* là số nghiên cứu (trong ví dụ trên *k* = 9). Theo lí thuyết xác suất, Q có độ phân phối theo luật *Chi-square* với **bậc tự do** (**degrees of freedom** – **df**) là *k*-1 (tức là χ^2_{k-1}). Nói cách khác, nếu Q lớn hơn χ^2_{k-1} thì đó là tín hiệu cho thấy sự bất đồng nhất giữa các nghiên cứu "có ý nghĩa thống kê" (significant).

Nhiều nghiên cứu trong thời gian qua chỉ ra rằng *Q* thường không phát hiện được sự bất đồng nhất một cách nhất quán, cho nên ngày nay ít ai dùng chỉ số này trong phân tích tổng hợp. Một chỉ số khác thay thế *Q* có tên là **index of heterogeneity** (I^2), tạm dịch là **chỉ số bất đồng nhất**, nhưng sẽ giữ cách viết I^2. Chỉ số này được định nghĩa như sau:

$$I^2 = \frac{Q - (k-1)}{Q}$$

I^2 có giá trị từ âm đến 1. Nếu $I^2 < 0$, thì chúng ta sẽ cho nó là 0; nếu I^2 gần bằng 1 thì đó là dấu hiệu cho thấy có sự bất đồng nhất giữa các nghiên cứu.

Trong ví dụ trên, để ước tính Q và I^2, chúng ta cần tính $W_i(d_i - d)^2$ cho từng nghiên cứu. Chẳng hạn như, với nghiên cứu 1:

$$W_i(d_i - d)^2 = 0{,}0246*(20 - 3.49)^2 = 6{,}7129$$

Bảng 1d. Tính toán các chỉ số đồng nhất và bất đồng nhất

Nghiên cứu	d_i	s_i^2	W_i	$W_i d_i$	$W_i(d_i - d)^2$
1	20	40.6	0.0246	0.4928	6.7129
2	2	2.0	0.4886	0.9771	1.0903
3	55	15.3	0.0654	3.5993	173.6080
4	71	150.2	0.0067	0.4726	30.3356
5	4	20.2	0.0495	0.1981	0.0127
6	-1	1.2	0.8173	-0.8173	16.5054
7	-11	95.4	0.0105	-0.1153	2.2026
8	10	8.0	0.1245	1.2450	5.2701
9	-7	20.7	0.0483	-0.3383	5.3215
Tổng số			**1.6354**	**5.7140**	**241.05**

Sau khi đã ước tính $W_i(d_i - d)^2$ cho từng nghiên cứu, chúng ta cộng lại số này (xem cột sau cùng) và đó chính là Q :

$$Q = \sum_{i=1}^{k} W_i(d_i - d)^2 = 241.05$$

Từ đó, I^2 có thể ước tính như sau:

$$I^2 = \frac{241.05 - 8}{241.05} = 0.966$$

Chỉ số bất đồng nhất I^2 rất cao, cho thấy độ dao động về d_i giữa các nghiên cứu rất cao. Điều này chúng ta có thể thấy được chỉ qua nhìn vào cột số 2 trong bảng thống kê trên.

Bước 6: **đánh giá khả năng publication bias [4].** Publication bias (tạm dịch: **trong thiên vị**) là một khái niệm tương đối mới có thể giải thích bằng tình huống thực tế sau đây. Chúng ta biết rằng khi một nghiên cứu cho ra kết quả "negative" (kết quả tiêu cực, tức là không phát hiện một ảnh hưởng hay một mối liên hệ có ý nghĩa thống kê) công trình nghiên cứu đó rất khó có cơ hội được công bố trên các tập san, bởi vì giới chủ bút tập san nói chung không thích

in những bài như thế. Ngược lại, một nghiên cứu với một kết quả "tích cực" (tức có ý nghĩa thống kê) thì nghiên cứu có khả năng xuất hiện trên các tập san khoa học cao hơn là các nghiên cứu với kết quả "tiêu cực". Thế nhưng phần lớn những phân tích tổng hợp lại dựa vào các kết quả đã công bố trên các tập san khoa học. Do đó, ước tính của một phân tích tổng hợp có khả năng thiếu khách quan, vì chưa xem xét đầy đủ đến các nghiên cứu tiêu cực chưa bao giờ công bố.

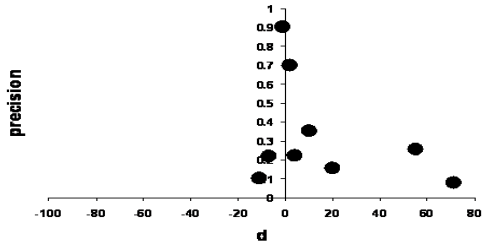
Một số nhà nghiên cứu đề nghị dùng biểu đồ funnel (còn gọi là **funnel plot**) để kiểm tra khả năng publication bias. Biểu đồ funnel được thể hiện bằng cách vẽ **độ chính xác – precision** (trục tung, y-axis) với ước tính mức độ ảnh hưởng cho từng nghiên cứu. Ở đây precision được định nghĩa là số đảo của sai số chuẩn (standard error):

$$\text{precision} = \frac{1}{s_{di}}$$

Nói cách khác, biểu đồ funnel biểu diễn *precision* với d_i. Chẳng hạn như với nghiên cứu 1, chúng ta có: precision = $1/\sqrt{40,6} = 0,157$. Tính cho từng nghiên cứu, chúng ta có dùng bảng thống kê sau để vẽ biểu đồ funnel như sau:

Bảng 1e. Ước tính publication bias

Nghiên cứu	d_i	s_i^2	$1/s_i$
1	20	40.6	0.1570
2	2	2.0	0.6990
3	55	15.3	0.2558
4	71	150.2	0.0816
5	4	20.2	0.2225
6	-1	1.2	0.9041
7	-11	95.4	0.1024
8	10	8.0	0.3528
9	-7	20.7	0.2198



Biểu đồ funnel (biểu đồ phễu): trục tung là precision và trục hoành là *d*. Biểu đồ này cho thấy phần lớn các nghiên cứu có kết quả thời gian nằm viện trong các bệnh viện đa khoa thường lâu hơn các bệnh viện chuyên khoa.

Cái logic đằng sau biểu đồ funnel là nếu các công trình nghiên cứu lớn (tức có độ precision cao) có khả năng được công bố cao, thì số lượng nghiên cứu với kết quả tích cực sẽ nhiều hơn số lượng nghiên cứu nhỏ hay với kết quả tiêu cực trong các tập san. Và nếu điều này xảy ra, thì biểu đồ funnel sẽ thể hiện một sự thiếu cân đối (asymmetry). Nói cách khác, sự thiếu cân đối của một biểu đồ funnel là dấu hiệu cho thấy có vấn đề về publication bias. Nhưng vấn đề đặt ra là publication bias đó có ý nghĩa thống kê hay không? Biểu đồ funnel không thể trả lời câu hỏi này, chúng ta cần đến các phương pháp phân tích định lượng nghiêm chỉnh hơn.

Kiểm định Egger

Vài năm gần đây có ý kiến cho rằng biểu đồ funnel rất khó diễn dịch, và có thể gây nên ngộ nhận về publication bias [5-6]. Thật vậy, một số tập san y học có chính sách khuyến khích các nhà nghiên cứu tìm một phương pháp khác để đánh giá *publication bias* thay vì dùng biểu đồ funnel.

Một trong những phương pháp đó là kiểm định Egger (còn gọi là **Egger's test**). Với phương pháp này, chúng ta mô hình rằng *SND* = *a* + *b* x *precision,* trong đó *SND* được ước tính bằng cách lấy *d* chia cho sai số chuẩn của d, tức là: $SND_i = \frac{d_i}{s_{di}}$, *a* và *b* là hai thông số phải ước tính từ mô hình hồi qui đường thẳng đó. Ở đây, *a* cung cấp cho chúng ta một ước số về tình trạng thiếu cân đối của biểu đồ funnel: *a*>0 có nghĩa là xu hướng nghiên cứu càng có qui mô lớn càng có ước số về độ ảnh hưởng với sự chính xác cao.

Trong ví dụ trên, chúng ta có thể dùng một phần mềm phân tích thống kê (như SAS hay R) để ước tính *a* và *b* như sau:

$$SND_i = 4.20 + -4.17084*\text{precision}_i$$

Kết quả ước số $a = 4.20$ tuy là >0 nhưng không có ý nghĩa thống kê, cho nên ở đây bằng chứng cho thấy không có sự publication bias.

Tuy nhiên, như đã thấy trong thực tế, kiểm định Egger này cũng chỉ là một cách thể hiện biểu đồ funnel mà thôi, chứ cũng không có thay đổi gì lớn. Có một cách đánh giá publication bias, cho đến nay, được xem là đáng tin cậy nhất: đó là phương pháp phân tích hồi qui đường thẳng (linear regression) giữa d_i và tổng số mẫu (N_i). Nói cách khác, chúng ta tìm *a* và *b* trong mô hình [7]:

$$d_i = a + b*N_i$$

Nếu không có publication bias thì giá trị của *b* sẽ rất gần với 0 hay không có ý nghĩa thống kê. Nếu trị số *b* khác với 0 thì đó là một tín hiệu của publication bias. Trong ví dụ vừa nêu với dữ liệu sau đây,

Nghiên cứu	d_i	N_i
1	20	311
2	2	63
3	55	146
4	71	36
5	4	21
6	-1	109
7	-11	67
8	10	293
9	-7	112

chúng ta có phương trình:

$$d_i = 16.0 - 0.0009*/N_i$$

và quả thật giá trị của *b* quá thấp (cũng như không có ý nghĩa thống kê), cho nên đến đây chúng ta có thể kết luận rằng không có vấn đề publication bias trong nghiên cứu vừa đề cập đến.

Nói tóm lại, qua phân tích tổng hợp này, chúng ta có bằng chứng đáng tin cậy để kết luận rằng thời gian nằm viện của bệnh nhân trong các bệnh viện đa khoa dài hơn các bệnh viện chuyên khoa khoảng 3 ngày rưỡi, hoặc trong

95% trường hợp thời gian khác biệt khoảng từ 2 ngày đến 5 ngày. Kết quả này cũng cho thấy không có thiên vị xuất bản (publication bias) trong phân tích.

14.4.2 Phân tích tổng hợp bằng R

R có hai package được viết và thiết kế cho phân tích tổng hợp. Package được sử dụng khá thông dụng là `meta`. Bạn đọc có thể tải miễn phí từ trang web của R (trong phần packages): http://cran.R-project.org.

Để phân tích tổng hợp bằng R chúng ta phải nhập package `meta` vào môi trường vận hành của R (với điều kiện, tất nhiên, là bạn đọc đã tải và cài đặt `meta` vào R).

```
> library(meta)
```

Sau đó, chúng ta sẽ nhập số liệu trong ví dụ 1 vào R biến như sau:

- Nhập dữ liệu cho từng cột trong **Bảng 1** và cho vào một dataframe gọi là `los`:

```
> n1 <-   c(155,31,75,18,8,57,34,110,60)
> los1 <- c(55,27,64,66,14,19,52,21,30)
> sd1 <-  c(47,7,17,20,8,7,45,16,27)

> n2 <-   c(156,32,71,18,13,52,33,183,52)
> los2 <- c(75,29,119,137,18,18,41,31,23)
> sd2 <-  c(64,4,29,48,11,4,34,27,20)

> los <- data.frame(n1,los1,sd1,n2,los2,sd2)
```

- Sử dụng hàm `metacont` (dùng để phân tích các biến liên tục – do đó cont=continuous variable) và cho kết quả vào đối tượng `res`:

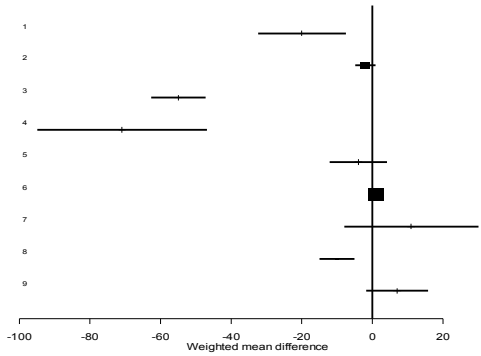
```
> res <- metacont(n1,los1,sd1,n2,los2,sd2,data=los)
> res
> res
  WMD                 95%-CI %W(fixed) %W(random)
1 -20  [-32.4744;  -7.5256]      1.44      10.69
2  -2  [ -4.8271;   0.8271]     28.11      12.67
3 -55  [-62.7656; -47.2344]      3.73      11.89
4 -71  [-95.0223; -46.9777]      0.39       7.39
5  -4  [-12.1539;   4.1539]      3.38      11.80
6   1  [ -1.1176;   3.1176]     50.11      12.72
7  11  [ -8.0620;  30.0620]      0.62       8.76
8 -10  [-14.9237;  -5.0763]      9.27      12.41
9   7  [ -1.7306;  15.7306]      2.95      11.67
```

```
Number of trials combined: 9

                        WMD     95%-CI           z       p.value
Fixed effects model  -3.464  [ -4.96; -1.96]  -4.53  <0.0001
Random effects model -13.98  [-24.03; -3.93]  -2.73   0.0064

Quantifying heterogeneity:
tau^2 = 205.4094; H = 5.46 [4.54; 6.58];
I^2 = 96.7% [95.2%; 97.7%]

Test of heterogeneity:
      Q d.f.  p.value
 238.92    8 < 0.0001

Method: Inverse variance method
```

`meta` cung cấp cho chúng ta hai kết quả: một kết quả dựa vào mô hình fixed-effects và một dựa vào mô hình random-effects. Như thấy qua kết quả trên, mức độ khác biệt giữa hai mô hình khá lớn, nhưng kết quả chung thì giống nhau, tức kết quả của cả hai mô hình đều có ý nghĩa thống kê.

Ngoài ra, chúng ta cũng có thể sử dụng hàm `plot` để thể hiện kết quả trên bằng biểu đồ forest như sau:

```
> plot(res, lwd=3)
```



14.5. Phân tích tổng hợp ảnh hưởng bất biến cho một tiêu chí nhị phân (Fixed-effects meta-analysis for a dichotomous outcome).

Trong phần trên, tôi vừa mô tả những bước chính trong một phân tích tổng hợp những nghiên cứu mà tiêu chí là một biến liên tục (continuous variable). Đối với các biến liên tục, trị số trung bình và độ lệch chuẩn là hai chỉ số thống kê thường được sử dụng để tóm lược. Nhưng hai chỉ số này không thể ứng dụng cho những tiêu chí mang tính thể loại hay thứ bậc như tử vong, gãy xương, v.v… vì những tiêu chí này chỉ có hai giá trị: hoặc là có, hoặc là không. Một người hoặc là còn sống hay chết, bị gãy xương hay không gãy xương, mắc bệnh suy tim hay không mắc bệnh suy tim, v.v… Đối với những biến này, chúng ta cần một phương pháp phân tích khác với phương pháp dành cho các biến liên tục.

14.5.1 Mô hình phân tích

Đối với những tiêu chí nhị phân (chỉ có hai giá trị), chỉ số thống kê tương đương với trị số trung bình là **tỉ lệ** hay **proportion**, có thể tính phần trăm); và chỉ số tương đương với độ lệch chuẩn là sai số chuẩn (**standard error**). Chẳng hạn như nếu một nghiên cứu theo dõi 25 bệnh nhân trong một thời gian, và trong thời gian đó có 5 bệnh nhân mắc bệnh, thì tỉ lệ (kí hiệu là p) đơn giản là: $p = 5/25 = 0,20$ (hay 20%). Theo lí thuyết xác suất, phương sai của p (kí hiệu là var[p]) là:

$$\text{var}[p] = p(1-p)/n = 0,2*(1 - 0,8)/25 = 0,0064.$$

Theo đó, sai số chuẩn của p (kí hiệu SE[p]) là:

$$SE[p] = \sqrt{\text{var}[p]} = \sqrt{0,0064} = 0,08.$$

Chúng ta còn có thể ước tính khoảng tin cậy 95% của tỉ lệ như sau:

$$p \pm 1,96 \times SE[p] = 0,2 \pm 1,96 \times 0,08 = 0,04 \text{ đến } 0,36.$$

Vì cách tính của các tiêu chí nhị phân khá đặc thù, cho nên phương pháp phân tích tổng hợp các nghiên cứu với biến nhị phân cũng khác. Để minh họa cách phân tích tổng hợp dạng này, chúng ta sẽ lấy một ví dụ (phỏng theo một nghiên cứu có thật).

Ví dụ 2: Beta-blocker (viết tắt là BB) là một loại thuốc có chức năng điều trị và phòng chống cao huyết áp. Có giả thiết cho rằng BB cũng có thể phòng chống bệnh suy tim, hay ít ra là làm giảm nguy cơ suy tim. Để thử nghiệm giả thiết này, hàng loạt nghiên cứu lâm sàng đối chứng ngẫu nhiên đã được tiến hành trong thời gian 20 năm qua. Mỗi nghiên cứu có 2 nhóm bệnh nhân: nhóm được điều trị bằng BB, và một nhóm không được điều trị (còn gọi là placebo hay giả dược). Trong thời gian 2 năm theo dõi, các nhà nghiên cứu xem xét tần số tử vong cho từng nhóm. **Bảng 2** sau đây tóm lược 13 nghiên cứu trong quá khứ:

Bảng 2. Beta-blocker và bệnh suy tim (congestive heart failure)

Nghiên cứu *(i)*	Beta-blocker		Placebo	
	N_1	Tử vong (d_1)	N_2	Tử vong (d_2)
1	25	5	25	6
2	9	1	16	2
3	194	23	189	21
4	25	1	25	2
5	105	4	34	2
6	320	53	321	67
7	33	3	16	2
8	261	12	84	13
9	133	6	145	11
10	232	2	134	5
11	1327	156	1320	228
12	1990	145	2001	217
13	214	8	212	17
Tổng cộng	4879	420	4516	612

N: số bệnh nhân nghiên cứu; Tử vong: số bệnh nhân chết trong thời gian theo dõi.

Như chúng ta thấy, một số nghiên cứu có số mẫu khá nhỏ, lại có những nghiên cứu với số mẫu gần 4000 người! Câu hỏi đặt ra là tổng hợp các nghiên cứu này, kết quả có nhất quán hay phù hợp với giả thiết BB làm giảm nguy cơ suy tim hay không? Để trả lời câu hỏi này, chúng ta tiến hành những bước sau đây:

Bước 1: ước tính mức độ ảnh hưởng cho từng nghiên cứu. Mỗi nghiên cứu có hai tỉ lệ: một cho nhóm BB và một cho nhóm placebo (giả dược). Gọi hai tỉ lệ này là p_1 và p_2, chỉ số để đánh giá mức độ ảnh hưởng của thuốc BB là tỉ số nguy cơ tương đối (relative risk – RR), và RR có thể được ước tính như sau:

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

Chẳng hạn như, trong nghiên cứu 1, chúng ta có: $p_1 = \frac{5}{25} = 0{,}20$ và $p_2 = \frac{8}{25} = 0{,}24$. Như vậy tỉ số nguy cơ cho nghiên cứu 1 là: $RR = \frac{0{,}20}{0{,}24} = 0{,}833$. Tính toán tương tự cho các nghiên cứu còn lại, chúng ta sẽ có một bảng như sau:

Bảng 2a. Ước tính tỉ lệ tử vong và tỉ số nguy cơ tương đối

Nghiên cứu (*i*)	Tỉ lệ tử vong nhóm BB (p_1)	Tỉ lệ tử vong nhóm placebo (p_2)	Tỉ số nguy cơ *(RR)*
1	0.200	0.240	0.833
2	0.111	0.125	0.889
3	0.119	0.111	1.067
4	0.040	0.080	0.500
5	0.038	0.059	0.648
6	0.166	0.209	0.794
7	0.091	0.125	0.727
8	0.046	0.155	0.297
9	0.045	0.076	0.595
10	0.009	0.037	0.231
11	0.118	0.173	0.681
12	0.073	0.108	0.672
13	0.037	0.080	0.466

Bước 2: biến đổi RR thành đơn vị logarithm và tính phương sai, sai số chuẩn. Mỗi ước số thống kê, như có lần nói, đều có một luật phân phối, và luật phân phối có thể phản ánh bằng phân sai (hay sai số chuẩn). Cách tính phương sai của RR khá phức tạp, cho nên chúng ta sẽ tính bằng một phương pháp gián tiếp. Theo phương pháp này, chúng ta sẽ biến đổi RR thành log[RR] (chú ý "log" ở đây có nghĩa là loga tự nhiên, tức là log_e hay có khi còn viết tắt là *ln* – natural logarithm) , và sau đó sẽ tính phương sai của log[RR].

Nếu N_1 và N_2 là lần lược tổng số mẫu của nhóm 1 và nhóm 2; và d_1 và d_2 là số tử vong của nhóm 1 và nhóm 2 của một nghiên cứu, thì phương sai của log[RR] có thể ước tính bằng công thức sau đây:

$$\text{Var[logRR]} = \frac{1}{d_1} - \frac{1}{N_1 - d_1} + \frac{1}{d_2} - \frac{1}{N_2 - d_2}$$

Và sai số chuẩn của log[RR] là:

$$\text{SE[logRR]} = \sqrt{\frac{1}{d_1} - \frac{1}{N_1 - d_1} + \frac{1}{d_2} - \frac{1}{N_2 - d_2}}$$

Trong ví dụ trên, với nghiên cứu 1, chúng ta có:

$$\text{Log[RR]} = \log_e(0.833) = -0.182$$

Với phương sai:

$$\text{var}[\log RR] = \frac{1}{5} - \frac{1}{25-5} + \frac{1}{6} - \frac{1}{25-6} = 0.264$$

Và sai số chuẩn:

$$SE[\log RR] = \sqrt{0.264} = 0.514$$

Dựa vào luật phân phối chuẩn, chúng ta cũng có thể tính toán khoảng tin cậy 95% của RR cho từng nghiên cứu bằng cách biến đổi ngược lại theo đơn vị RR. Chẳng hạn như với nghiên cứu 1, chúng ta có khoảng tin cậy 95% của log[RR] là:

$$\log RR \pm 1.96*SE[\log RR] = -0.182 \pm 1.96*0.514 = -1.19 \text{ đến } 0.82$$

hay biến đổi thành đơn vị nguyên thủy của RR là:

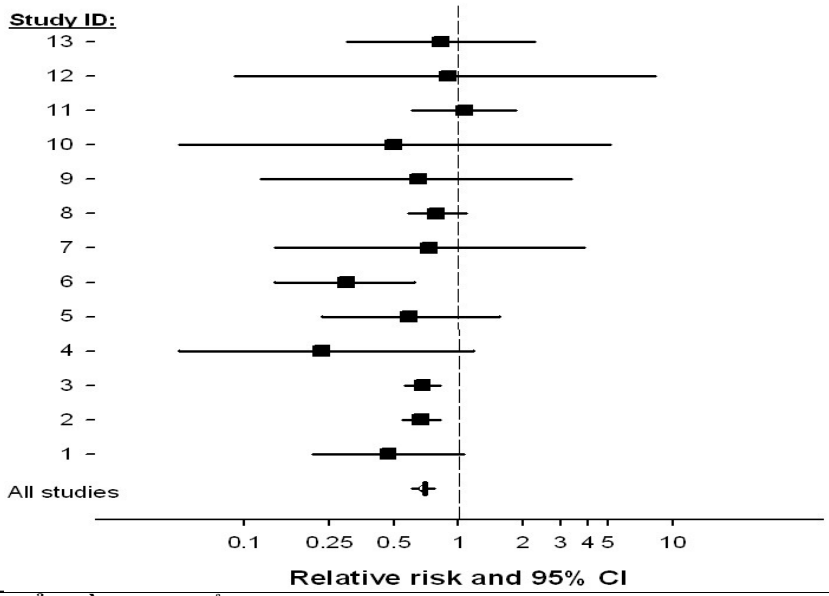
$$\exp(-1.19) = 0.30 \text{ đến } \exp(0.82) = 2.28$$

Tính toán tương tự cho các nghiên cứu khác, chúng ta có thêm một bảng mới như sau:

Bảng 2b. Ước tính tỉ số nguy cơ tương đối, phương sai, sai số chuẩn và khoảng tin cậy 95% cho từng nghiên cứu

Nghiên cứu (*i*)	Tỉ số nguy cơ *(RR)*	Log[RR]	Var[logRR]	SE[logRR]	Phần thấp 95%CI của RR	Phần cao 95% CI của RR
1	0.200	-0.182	0.264	0.514	0.30	2.28
2	0.111	-0.118	1.304	1.142	0.09	8.33
3	0.119	0.065	0.079	0.282	0.61	1.85
4	0.040	-0.693	1.415	1.189	0.05	5.15
5	0.038	-0.434	0.709	0.842	0.12	3.37
6	0.166	-0.231	0.026	0.162	0.58	1.09
7	0.091	-0.318	0.729	0.854	0.14	3.87
8	0.046	-1.214	0.142	0.377	0.14	0.62
9	0.045	-0.520	0.242	0.492	0.23	1.56
10	0.009	-1.465	0.688	0.829	0.05	1.17
11	0.118	-0.385	0.009	0.095	0.56	0.82
12	0.073	-0.398	0.010	0.102	0.55	0.82
13	0.037	-0.763	0.174	0.417	0.21	1.06

Chúng ta có thể thể hiện RR và khoảng tin cậy 95% bằng biểu đồ forest như sau:



Biểu đồ forest thể hiện giá trị của *RR* và khoảng tin cậy 95%. Các ước tính khoảng tin cậy 95%CI của RR vượt qua cột mốc 1 được xem là không có ý nghĩa thống kê.

Bước 3: ước tính trọng số (weight) cho từng nghiên cứu và RR cho toàn bộ nghiên cứu. Biểu đồ trên cho thấy một số nghiên cứu có độ dao động RR rất lớn (chứng tỏ các nghiên cứu này có số mẫu nhỏ hay ước số RR không ổn định), và ngược lại, một số nghiên cứu lớn có ước số RR ổn định hơn. Trọng số cho mỗi nghiên cứu (W_i – cho vào kí hiệu *i*) để đo lường độ ổn định này là số đảo của phương sai:

$$W_i = \frac{1}{\text{var}[\log RR_i]}$$

Và số trung bình trọng số của log[RR] (kí hiệu là log*wRR*) có thể ước tính từ tổng của tích $W_i \times \log[RR_i]$:

$$\log wRR = \frac{\sum W_i \times \log[RR_i]}{\sum W_i}$$

Với phương sai:

$$\text{Var}[\log wRR] = \frac{1}{\sum W_i}$$

và sai số chuẩn:

$$SE[\log wRR] = \sqrt{\frac{1}{\sum W_i}} \text{ i1}$$

Ngoài ra, khoảng tin cậy 95% có thể ước tính bằng:

$$\log wRR \pm SE[\log wRR]$$

Để tính trung bình trọng số logRR, chúng ta cần một cột $W_i \times \log[RR_i]$. Chẳng hạn như với nghiên cứu 1, chúng ta có:

$$W_1 = \frac{1}{0,264} = 3.79$$

và

$$W_i \times \log[RR_i] = 3.79 \times (-0.182) = -0.69$$

Tính toán tương tự cho các nghiên cứu khác, chúng ta sẽ có một bảng số liệu mới như sau:

Bảng 2c. Ước tính tỉ trọng số (W_i)

Nghiên cứu (i)	Log[RR]	Var[logRR]	W_i	$W_i \times \log[RR_i]$
1	-0.182	0.264	3.79	-0.69
2	-0.118	1.304	0.77	-0.09
3	0.065	0.079	12.61	0.82
4	-0.693	1.415	0.71	-0.49
5	-0.434	0.709	1.41	-0.61
6	-0.231	0.026	38.30	-8.86
7	-0.318	0.729	1.37	-0.44
8	-1.214	0.142	7.03	-8.54
9	-0.520	0.242	4.13	-2.15
10	-1.465	0.688	1.45	-2.13
11	-0.385	0.009	110.78	-42.63
12	-0.398	0.010	96.13	-38.23
13	-0.763	0.174	5.75	-4.39
Tổng số			**284.24**	**-108.42**

Chúng ta có:

$$\sum W_i = 3.79 + 0.77 + \ldots + 5.75 = 284.24$$

$$\sum W_i \times \log[RR_i] = -0.69 - 0.09 + \ldots -4.39 = -108.42$$

Do đó. trung bình trọng số của log[RR] có thể ước tính bằng:

$$\log wRR = \frac{\sum W_i \times \log[RR_i]}{\sum W_i} = \frac{-108,42}{284,24} = -0.38$$

Với phương sai:

$$Var[\log wRR] = \frac{1}{\sum W_i} = \frac{1}{284.24} = 0.0035$$

và sai số chuẩn:

$$SE[\log wRR] = \sqrt{\frac{1}{\sum W_i}} = \sqrt{0.0035} = 0.06$$

Do đó. khoảng tin cậy 95% của log*wRR* có thể ước tính bằng:

$$\log wRR \pm SE[\log wRR] = -0.38 \pm 1.96 \times 0.06 = 0.498 \text{ đến } -0.265$$

Nhưng chúng ta muốn thể hiện bằng đơn vị gốc (tức tỉ số); do đó. các ước số trên phải được biến chuyển về đơn vị gốc:

$$RR = \exp(\log wRR) = \log(-0.38) = 0.68$$

Và khoảng tin cậy 95%:

$$\text{Exp}(-0.498) = 0.61 \text{ đến } \text{Exp}(-0.265) = 0.77.$$

Đến đây chúng ta có thể nói rằng tỉ lệ tử vong trong các bệnh nhân được điều trị bằng BB là 0.68 (hay thấp hơn 32%) so với các bệnh nhân placebo. Ngoài ra. vì khoảng tin cậy 95% không bao gồm 1, chúng ta cũng có thể phát biểu rằng mức độ khác biệt này có ý nghĩa thống kê.

Bước 4: ước tính chỉ số đồng nhất và bất đồng nhất. Như đã nói trong phần (1) liên quan đến phân tích biến liên tục, sau khi đã ước tính tỉ số nguy cơ trung bình. chúng ta cần phải xem xét chỉ số I^2. Để ước tính chỉ số I^2,

chúng ta cần tính $W_i\left(\log RR_i - \log wRR\right)^2$ cho mỗi nghiên cứu. Chẳng hạn như với nghiên cứu 1, chúng ta có:

$$W_i\left(\log RR_i - \log wRR\right)^2 = 3.79\times(-0.182 + 0.38)^2 = 0.1502$$

và tính toán tương tự cho các nghiên cứu khác, chúng ta sẽ có một bản số liệu mới như sau:

Bảng 2d. Ước tính chỉ số heterogeneity (I^2)

Nghiên cứu (i)	Log[RR$_i$]	W_i	$W_i\left(\log RR_i - \log wRR\right)^2$
1	-0.182	3.79	0.1502
2	-0.118	0.77	0.0533
3	0.065	12.61	2.5118
4	-0.693	0.71	0.0687
5	-0.434	1.41	0.0040
6	-0.231	38.30	0.8635
7	-0.318	1.37	0.0054
8	-1.214	7.03	4.8731
9	-0.520	4.13	0.0790
10	-1.465	1.45	1.7074
11	-0.385	110.78	0.0012
12	-0.398	96.13	0.0253
13	-0.763	5.75	0.8382
Tổng số		**284.24**	**11.1811**

Ví dụ 2 có k = 13 nghiên cứu. Do đó.

$$Q = \sum_{i=1}^{k} W_i\left(\log RR_i - \log wRR\right)^2 = 11.1811$$

Và.

$$I^2 = \frac{Q - (k-1)}{Q} = \frac{11.18 - 12}{11.18} = -0.16$$

Vì $I^2 < 0$, nên chúng ta có thể cho $I^2 = 0$. Nói cách khác, mức độ khác biệt về *RR* giữa các nghiên cứu không có ý nghĩa thống kê.

Bước 5: **đánh giá khả năng *publication bias*.** Như đã giải thích trong phần 1f, cách đánh giá khả năng publication bias có ý nghĩa nhất là phân tích hồi qui đường thẳng log[RR] và tổng số mẫu (N):

$$\log[RR_i] = a + b \times N_i$$

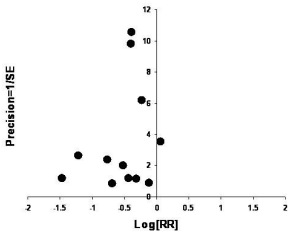
Dựa vào bảng thống kê sau.

Nghiên cứu (*i*)	Log[RR_i]	N_i
1	-0.182	50
2	-0.118	25
3	0.065	383
4	-0.693	50
5	-0.434	139
6	-0.231	641
7	-0.318	49
8	-1.214	345
9	-0.520	278
10	-1.465	366
11	-0.385	2647
12	-0.398	3991
13	-0.763	426

Chúng ta có thể ước tính *a* và *b* như sau:

$$\log[RR_i] = -0.534 + 0.00003 \times N_i$$

Ước tính b = 0.00003 không có ý nghĩa thống kê (p = 0.782). Do đó. chúng ta có thể phát biểu rằng mức độ thiên lệch về xuất bản không đáng kể trong phân tích tổng hợp này.



Biểu đồ funnel cũng cho thấy không có vấn đề publication bias

14.5.2 Phân tích bằng R

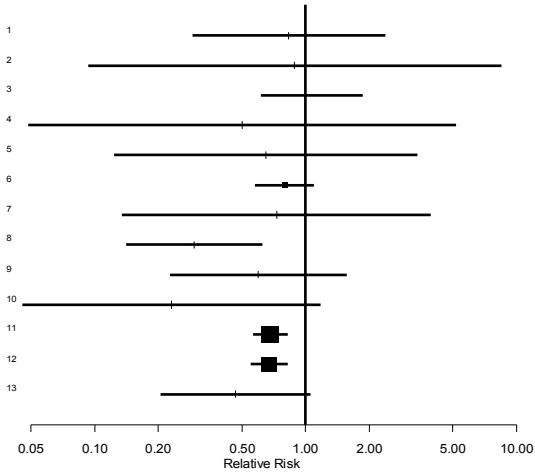
Package `meta` có hàm `metabin` có thể sử dụng để tiến hành phân tích tổng hợp cho các biến nhị phân như số liệu trong ví dụ 2 trên đây. Khởi đầu. chúng ta "nạp" package meta (nếu chưa làm) vào môi trường vận hành. và sau đó thu nhập số liệu vào một data frame:

```
library(meta)

# Số liệu từ ví dụ 2
n1 <- c(25.9.194.25.105.320.33.261.133.232.1327.1990.214)
d1 <- c(5.1.23.1.4.53.3.12.6.2.156.145.8)
n2 <- c(25.16.189.25.34.321.16.84.145.134.1320.2001.212)
d2 <- c(6.2.21.2.2.67.2.13.11.5.228.217.17)

# Tạo một dataframe lấy tên là bb
bb <- data.frame(n1.d1.n2.d2)

# Phân tích bằng hàm metabin và kết quả trong res
> res <- metabin(d1.n1.d2.n2.data=bb.sm="RR".meth="I")
> res
> res
       RR            95%-CI %W(fixed) %W(random)
1  0.8333  [0.2918; 2.3799]      1.26       1.26
2  0.8889  [0.0930; 8.4951]      0.27       0.27
3  1.0670  [0.6116; 1.8617]      4.47       4.47
4  0.5000  [0.0484; 5.1677]      0.25       0.25
5  0.6476  [0.1240; 3.3814]      0.51       0.51
6  0.7935  [0.5731; 1.0986]     13.08      13.08
7  0.7273  [0.1346; 3.9282]      0.49       0.49
8  0.2971  [0.1410; 0.6258]      2.49       2.49
9  0.5947  [0.2262; 1.5632]      1.48       1.48
10 0.2310  [0.0454; 1.1744]      0.52       0.52
11 0.6806  [0.5635; 0.8221]     38.81      38.81
12 0.6719  [0.5496; 0.8214]     34.31      34.31
13 0.4662  [0.2056; 1.0570]      2.07       2.07

Number of trials combined: 13

                         RR            95%-CI       z  p.value
Fixed effects model  0.6821  [0.6064; 0.7672] -6.3741 < 0.0001
Random effects model 0.6821  [0.6064; 0.7672] -6.3741 < 0.0001

Quantifying heterogeneity:
tau^2 = 0; H = 1 [1; 1.45]; I^2 = 0% [0%; 52.6%]

Test of heterogeneity:
  Q d.f.  p.value
 11   12   0.5292
Method: Inverse variance method
```

Kết quả từ mô hình fixed-effects và random-effects một lần nữa cho chúng ta bằng chứng để kết luận rằng beta-blocker có hiệu nghiệm trong việc làm giảm nguy cơ tử vong.

Biểu đồ forest

```
> plot(res. lwd=3)
```



***

Thực ra, trong khoa học nói chung, chúng ta đã có một truyền thống lâu đời về việc duyệt xét bằng chứng nghiên cứu (review), duyệt xét kiến thức hiện hành. Nhưng các duyệt xét như thế thường mang tính định chất (qualitative review). và vì tính định chất. chúng ta khó mà biết chính xác được những khác biệt mang tính định lượng giữa các nghiên cứu. Phân tích tổng hợp cung cấp cho chúng ta một phương tiện định lượng để hệ thống bằng chứng. Với phân tích tổng hợp, chúng ta có cơ hội để:

- Xem xét những nghiên cứu nào đã được tiến hành để giải quyết vấn đề;
- Kết quả của các nghiên cứu đó như thế nào;
- Hệ thống các tiêu chí lâm sàng đáng quan tâm;
- Rà soát những khác biệt về đặc tính giữa các nghiên cứu;
- Cách thức để tổng hợp kết quả; và
- Truyền đạt kết quả một cách khoa học.

Mục đích của phân tích tổng hợp, xin nhắc lại một lần nữa, là ước tính một chỉ số ảnh hưởng trung bình sau khi đã xem xét tất cả kết quả nghiên cứu

hiện hành. Một kết quả chung như thế giúp cho chúng ta đi đến một kết luận chính xác và đáng tin cậy hơn.

Hai ví dụ trên đây hi vọng đã giúp ích cho bạn đọc hiểu được "cơ chế" và ý nghĩa của một phân tích tổng hợp. Hi vọng bạn đọc có thể tự mình làm một phân tích như thế khi đã có dữ liệu. Thực ra, tất cả các tính toán trên có thể thực hiện bằng một phần mềm như Microsoft Excel. Ngoài ra, một số phần mềm chuyên môn khác (như SAS chẳng hạn) cũng có thể tiến hành những phân tích trên. Các phép tính trong phân tích tổng hợp thật đơn giản. Vấn đề của phân tích tổng hợp không phải là tính toán, mà là dữ liệu đằng sau tính toán.

Phân tích tổng hợp cũng không phải là không có những khiếm khuyết. Trong nghiên cứu người ta có câu "rác vào, rác ra", tức là nếu các dữ liệu được sử dụng trong phân tích không có chất lượng cao thì kết quả của phân tích tổng hợp cũng chẳng có giá trị khoa học gì. Do đó, vấn đề quan trọng nhất trong phân tích tổng hợp là chọn lựa dữ liệu và nghiên cứu để phân tích. Vấn đề này cần phải được cân nhắc cực kì cẩn thận để đảm báo tính hợp lí và khoa học của kết quả.

Tài liệu tham khảo và chú thích

[1] Glass GV. Primary. secondary. and meta-analysis of research. Educational Researcher 1976; 5:3-8.

[2] Normand SL. Meta-analysis: formulating. evaluating. combining. and reporting. Stat Med. 1999;18(3):321-59.

[3] Higgins JPT. Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21:1539-1558

[4] Egger M. Davey Smith G. Schneider M. Minder C. Bias in meta-analysis detected by a simple. graphical test. Br Med J 1997;315**:**629–34.

[5] Tang JL. Liu JL. Misleading funnel plot for detection of bias in meta-analysis. J Clin Epidemiol. 2000;53(5):477-84.

[6] Peters JL. Sutton AJ. Jones DR. Abrams KR. Rushton L. Comparison of two methods to detect publication bias in meta-analysis. JAMA. 2006;295(6):676-80.

[7] Macaskill P. Walter SD. Irwig L. A comparison of methods to detect publication bias in meta-analysis. Stat Med. 2001;20:641-654.

Tóm tắt phân tích tổng hợp

Đối với các biến số liên tục	**Đối với các biến số nhị phân**
Nhóm 1 (số mẫu. trung bình. độ lệch chuẩn): n_{1i} . x_{1i} . s_{1i} ; i = 1. 2. 3. …. k	Nhóm 1 (số mẫu. số sự kiện): n_{1i} . x_{1i} ; i = 1. 2. 3. …. k
Nhóm 2 (số mẫu. trung bình. độ lệch chuẩn): n_{2i} . x_{2i} . s_{2i}	Nhóm 2 (số mẫu. số sự kiện): n_{2i} . x_{2i} ; i = 1. 2. 3. …. k
Độ ảnh hưởng (effect size. ES): $d_i = x_{2i} - x_{1i}$	Độ ảnh hưởng (effect size. ES) tính bằng tỉ số nguy cơ RR: $RR_i = \left(\frac{x_{2i}}{n_{2i}}\right) \div \left(\frac{x_{1i}}{n_{1i}}\right)$ Biến chuyển sang logarithm: $\theta_i = \log(RR_i)$
Phương sai của d_i : $s_{di}^2 = \frac{(n_{1i}-1)s_{1i}^2 + (n_{2i}-1)s_{2i}^2}{n_{1i}+n_{2i}-2} \times \left(\frac{1}{n_{1i}} + \frac{1}{n_{2i}}\right)$	Phương sai của θ_i: $s_{\theta i}^2 = \frac{1}{x_{1i}} - \frac{1}{n_{1i}-x_{1i}} + \frac{1}{x_{2i}} - \frac{1}{n_{2i}-x_{2i}}$
Sai số (standard error) của d_i : $s_{di} = \sqrt{s_{di}^2}$	Sai số của θ_i: $s_\theta = \sqrt{\frac{1}{x_{1i}} - \frac{1}{n_{1i}-x_{1i}} + \frac{1}{x_{2i}} - \frac{1}{n_{2i}-x_{2i}}}$
Trọng số: $W_i = \frac{1}{s_{di}^2}$	Trọng số: $W_i = \frac{1}{s_{\theta i}^2}$
Ước số ảnh hưởng chung: $d = \sum_{i=1}^{k} W_i d_i \,/\, \sum_{i=1}^{k} W_i$	Ước số ảnh hưởng chung: $\theta = \sum_{i=1}^{k} W_i \theta_i \,/\, \sum_{i=1}^{k} W_i$
Phương sai của *d:* $s^2 = 1 / \sum_{i=1}^{k} W_i$	Phương sai của *θ:* $s^2 = 1 / \sum_{i=1}^{k} W_i$
Khoảng tin cậy 95%: $d \pm 1{,}96 \times \sqrt{s^2}$	Khoảng tin cậy 95%: $\theta \pm 1{,}96 \times \sqrt{s^2}$
Index of homogeneity: $Q = \sum_{i=1}^{k} W_i (d_i - d)^2$	Index of homogeneity: $Q = \sum_{i=1}^{k} W_i (\theta_i - \theta)^2$
Index of heterogeneity:	Index of heterogeneity:

$I^2 = \frac{Q-(k-1)}{Q}$	$I^2 = \frac{Q-(k-1)}{Q}$
Xem xét publication bias: Phân tích hồi qui tuyến tính: $d_i = a + b*N_i$. (N_i là tổng số mẫu của nghiên cứu i). Xem ý nghĩa thống kê của b.	Xem xét publication bias: Phân tích hồi qui tuyến tính: $\theta_i = a + b*N_i$. (N_i là tổng số mẫu của nghiên cứu i). Xem ý nghĩa thống kê của b.
Ước tính phương sai giữa các nghiên cứu (between-study variance): $\tau^2 = \max\left[0, \frac{Q-(k-1)}{\sum_{i=1}^{k} W_i - \frac{\sum_{i=1}^{k} W_i^2}{\sum_{i=1}^{k} W_i}}\right]$	Ước tính phương sai giữa các nghiên cứu (between-study variance): $\tau^2 = \max\left[0, \frac{Q-(k-1)}{\sum_{i=1}^{k} W_i - \frac{\sum_{i=1}^{k} W_i^2}{\sum_{i=1}^{k} W_i}}\right]$