

ĐÁNH GIÁ TẬP SAN KHOA HỌC: CHỈ SỐ EIGENFACTOR

NGUYỄN VĂN TUẤN

Đại học New South Wales, Úc

Mặc dù bị phê phán không ít nhưng hiện nay chỉ số ảnh hưởng (Impact Factor - IF) vẫn được nhiều trường đại học và tổ chức nghiên cứu sử dụng để đánh giá chất lượng một tập san khoa học. Gần đây, có một phát kiến mới trong đánh giá tầm ảnh hưởng của một tập san, đó là chỉ số eigenfactor (viết tắt là EF). Bài viết giới thiệu về chỉ số này và một số ý kiến của tác giả để các nhà quản lý và nghiên cứu cùng tham khảo.

Đánh giá kết quả hoạt động khoa học và công nghệ là việc làm cần thiết đối với cả hai giới quản lý khoa học và nghiên cứu khoa học. Đối với những người làm công tác quản lý khoa học, họ muốn biết tiền tài trợ cho nghiên cứu có xứng với “đồng tiền bát gạo”, có đem lại hiệu quả hay không. Dĩ nhiên, “hiệu quả” ở đây có thể đo lường bằng nhiều tiêu chí: sản phẩm được ứng dụng ngay trong thực tế, là bằng sáng chế, hay bài báo khoa học. Phần lớn các nghiên cứu cơ bản, sản phẩm chính vẫn là những bài báo khoa học. Nhưng trong hàng vạn tập san khoa học, các nhà quản lý nếu không phải là người trong chuyên ngành hay thậm chí không phải là người làm khoa học thì làm sao có thể biết tập san nào có chất lượng cao hay thấp.

Chỉ số ảnh hưởng

Chất lượng tập san có thể hiểu là tầm ảnh hưởng, và một trong những cách đánh giá tầm ảnh hưởng của tập san là số lần trích dẫn. Một tập san công bố nhiều công trình khoa học mà không ai trích dẫn thì đó là tín hiệu cho thấy tập san có chất lượng thấp, và tầm ảnh hưởng cũng thấp. Một nhận định tương tự cũng có thể áp dụng cho một bài báo khoa học. Một trong những chỉ số được tính toán từ số lần trích dẫn là *chỉ số ảnh hưởng* (Impact Factor - IF). Chỉ số IF do Eugene Garfield “sáng chế” ra từ giữa thập niên 1950, và gây tác động lớn đến khoa học sau đó. Trước đây (và cho đến nay), giới quản lý khoa học vẫn hay dựa vào chỉ số này để đánh giá chất lượng của một tập san khoa học. Ở Trung Quốc và nhiều trường đại học phương Tây, người ta thường cho nhà khoa học có bài công bố trên những tập san có IF cao. Tác giả nào có bài trên tập san có IF càng cao, thì tiền thưởng càng lớn.

Có thể giải thích IF qua một ví dụ như sau: nếu năm 2000 và 2001 tập san A công bố 100 bài báo, và trong năm 2002, 100 bài này được trích dẫn 2000 lần, thì IF năm 2002 của tập san A được tính là $2000/100 = 20$. Chỉ số IF rất khác biệt giữa các lĩnh vực khoa học, với những tập san ngành khoa học xã

hội thường có IF thấp hơn các tập san khoa học tự nhiên và y sinh học. Do đó, IF chỉ có ý nghĩa trong mỗi chuyên ngành, chứ rất khó so sánh giữa các ngành. Tuy nhiên, đối với các tập san khoa học tổng quát như *Science* hay *Nature* thì IF được xem như là tiêu chuẩn vàng để các tập san khác so sánh.

Vấn đề lớn nhất của IF là chỉ số này chỉ phản ánh tầm ảnh hưởng trong vòng 2 năm. Nhưng một số nghiên cứu gần đây cho thấy thời gian 2 năm có lẽ quá ngắn để đánh giá tầm ảnh hưởng; phần lớn các ngành khoa học, thời gian cần thiết để đánh giá là 5 năm. Nếu sau 5 năm mà vẫn chẳng có ai trích dẫn công trình nghiên cứu thì đó có lẽ là tín hiệu cho thấy nghiên cứu đó chẳng có ảnh hưởng gì (dĩ nhiên, không loại trừ vài trường hợp hiếm hoi có công trình 20 năm sau mới được ghi nhận).

Vấn đề thứ hai của IF là không xem xét đến “uy danh” của tập san trích dẫn. Để hiểu vấn đề, chúng ta thử so sánh 2 tập san như sau: Tập san X công bố 50 bài, và nhận được 90 lần trích dẫn từ những tập san danh tiếng (như *Science*, *Nature*, *PNAS*, *Cell*) và 10 lần trích dẫn từ những tập san của quốc gia đang phát triển T; Tập san Y cũng công bố 50 bài, và nhận được 10 lần trích dẫn từ những tập san danh tiếng và 90 lần trích dẫn từ những tập san của quốc gia T. Theo cách tính chỉ số IF thì hai tập san trên có IF như nhau ($100/50 = 2$). Nhưng nếu xét kỹ, chúng ta thấy tập san X chắc phải có uy danh cao hơn tập san Y, vì X được nhiều tập san danh tiếng trích dẫn hơn tập san Y.

Vấn đề thứ ba là IF dễ bị lạm dụng. Một minh chứng dễ hiểu nhất là nếu tập san công bố 100 bài, và trong số đó 95 bài chẳng ai trích dẫn, chỉ có 5 bài được trích dẫn tất cả là 500 lần, thì $IF = 5$. Điều này cho thấy IF dễ bị “lệch” bởi các bài báo nổi tiếng, và do đó, không phản ánh đúng uy tín của tập san. Biết được nhược điểm này, một số tập san cho công bố các bài tổng quan (review) vì những bài này thường được trích dẫn nhiều (dù biết rằng đó là trò không mấy hay ho, nhưng hoàn toàn “hợp pháp”!).

Gần đây, một nhóm nhà khoa học có tiếng đã ra tuyên bố DORA, đề nghị không dùng IF trong việc đánh giá khoa học. Tôi là một trong những người có tham gia thảo luận trước khi DORA ra đời, nhưng cũng như nhiều người khác, tôi không tham gia ký tên vì không đồng ý với vài điểm trong đó. Tuyên bố DORA tuy được đánh giá cao nhưng ít ai làm theo! Lý do là vì DORA không chỉ ra được cái gì mới, và

quan trọng hơn là họ chẳng đề xuất một cách đánh giá nào cụ thể và khả thi. Những gì DORA chỉ ra là khiếm khuyết thì giới khoa học đã biết 30 năm qua. Có người nói rằng, DORA chỉ đắm vào... không khí.

Nhưng trong thực tế, và nhìn chung, phải nói rằng IF đúng là có phản ánh phần nào đó chất lượng tập san. Sự thật không thể bác bỏ là trong mỗi chuyên ngành, tập san có IF càng cao thì cũng chính là tập san được các chuyên gia trong ngành đánh giá cao. Trong một nghiên cứu thú vị, các nhà nghiên cứu hỏi 264 bác sĩ nội khoa (internal medicine) đánh giá chất lượng tập san, rồi sau đó tính hệ số tương quan với IF. Kết quả cho thấy hệ số tương quan giữa điểm đánh giá và IF là 0,82, tức rất cao. Như vậy, rõ ràng IF phản ánh chất lượng tập san theo đánh giá của người trong ngành.

Eigenfactor

Để khắc phục những điểm yếu của IF, hai nhà khoa học Mỹ Jevin West và Carl Bergstrom (Đại học Washington) đề xuất một chỉ số mới để đánh giá tầm ảnh hưởng của một tập san khoa học. Họ gọi đó là chỉ số *EF*. Chỉ số EF vẫn dựa vào số lần trích dẫn, nhưng là số lần trích dẫn trong 5 năm (chứ không phải 2 năm như IF). Như đề cập ở trên, thời gian 5 năm là lý tưởng để đánh giá tầm ảnh hưởng của một công trình khoa học. Hệ số tương quan giữa số lần trích dẫn trong vòng 5 năm và tổng số lần trích dẫn là khoảng 0,81-0,91 cho ngành y sinh học; 0,85 cho ngành hoá học; 0,75 cho ngành toán học; 0,87 cho ngành vật lý và 0,79 cho ngành khoa học xã hội.

Ngoài ra, EF còn cho trọng số của những tập san trích dẫn, và do đó, khắc phục một điểm yếu quan trọng của IF. Triết lý của EF cũng giống như triết lý: *nói cho tôi biết bạn của bạn là ai thì tôi sẽ cho biết bạn là người như thế nào*. Nếu tập san công bố những bài báo được các tập san danh tiếng trích dẫn thì đó là một tín hiệu cho thấy tập san thuộc vào đẳng cấp cao. Jevin West và Carl Bergstrom còn đề xuất một chỉ số khác có tên là *Article Influence (AI)*. Chỉ số AI đo lường tầm ảnh hưởng của những bài báo trong tập san. AI được tính bằng cách lấy EF chia cho số bài báo tập san công bố và chuẩn hóa sao cho AI trung bình là 1. Nói cách khác, nếu một tập san có AI bằng 0,1 thì tập san đó có tầm ảnh hưởng thấp hơn trung bình (là 1), nhưng nếu tập san có AI bằng 1,2 thì tầm ảnh hưởng của tập san đó cao hơn trung bình 20%.

Hiện nay, chỉ số EF và AI càng ngày càng phổ biến. Ngay cả Thomson ISI cũng sử dụng hai chỉ số này để xếp hạng các tập san khoa học. Vừa qua, chúng tôi có dịp thảo luận với một số đồng nghiệp từ một số trường đại học của Mỹ, Úc cùng cả đại diện của ISI, và biết rằng, trong tương lai gần, có lẽ ISI sẽ đưa EF và AI chính thức làm chỉ số đánh giá uy tín và tầm ảnh hưởng của tập san khoa học và cũng là một cách thay thế chỉ số IF.

Các nhà khoa học phát kiến chỉ số EF và AI còn lập ra một website (<http://eigenfactor.org>) để tất cả chúng ta có thể tra cứu từng tập san. Chúng tôi thử tra cứu một số tập san trong nhóm nội tiết thì thấy 10 tập san hàng đầu như sau:

Tập san	EF	AI
1. Journal of Clinical Endocrinology and Metabolism	0,128	2,00
2. Diabetes Care	0,114	2,44
3. Diabetes	0,104	2,92
4. Endocrinology	0,086	1,50
5. Cell Metabolism	0,071	8,61
6. Diabetologia	0,059	2,20
7. Free Radical Biology & Medicine	0,053	1,58
8. Obesity	0,053	1,59
9. Journal of Bone and Mineral Research	0,048	2,25
10. Am J Physiol Endo Metab	0,048	1,59

Đối với một số tập san được giới khoa học đánh giá rất cao thì hai chỉ số EF và AI như sau:

Tập san	EF	AI
Science	1,412 (100)	17,52 (100)
Nature	1,655 (100)	20,37 (100)
PNAS	1,601 (100)	4,90 (100)
Nature Genetics	0,329 (100)	17,58 (100)
JAMA	0,286 (100)	13,11 (100)
Lancet	0,361 (100)	13,61 (100)
New England Journal of Medicine	0,664 (100)	21,30 (100)
Annals of Internal Medicine	0,121 (99)	7,91 (99)
Bone	0,040 (96)	1,42 (89)
Osteoporosis International	0,029 (94)	1,50 (90)
Archives of Internal Medicine	0,107 (99)	4,76 (99)

Ghi chú: số trong ngoặc là thứ hạng tính trên phần trăm. Chẳng hạn như Science có EF là 1,412, đứng hàng top 100 các tập san về khoa học tổng quát, còn Bone có EF 0,04, đứng hàng top 96 trong lĩnh vực loãng xương

Qua những kết quả này, có thể nói, các chỉ số EF và AI rất nhất quán với quan điểm trên của giới khoa học.

Tuy chỉ số EF đang được giới nghiên cứu về đánh giá khoa học chú ý, nhưng tôi vẫn thấy không thoải mái với chỉ số này. Lý do đơn giản là vì rất khó kiểm tra giả định đằng sau cách tính toán. Ngoài ra, phương pháp tính toán cũng rất phức tạp, mà có người cho là “black box” (hộp đen), nên rất khó hiểu chính xác chỉ số EF có nghĩa thật là gì. Chúng ta có thể hiểu ý nghĩa chung và có thể so sánh qua EF, nhưng chúng ta không rõ nó có ý nghĩa thật là gì. Tuy nhiên, một phân tích gần đây cho thấy, chỉ số EF có tương quan rất cao với IF (hệ số tương quan là 0,93). Do đó, dù EF có cải tiến về mặt lý thuyết, nhưng trong thực tế thì chẳng khác mấy so với IF. Trong trường hợp này, có thể nói rằng, chỉ số IF (dù bị chỉ trích và phê phán rất nhiều) vẫn có giá trị thực tế của nó và rất dễ hiểu! Có lẽ phương pháp Eigenfactor cung cấp cho chúng ta một cách đánh giá khách quan hơn là dựa vào cả hai IF và EF để xếp hạng tập san khoa học

Tài liệu tham khảo

1. Có thể tìm IF của các tập san tại: <http://www.impactfactorsearch.com>
2. Bergstrom C.T, West J.D, Wiseman M.A. The Eigenfactor™ Metrics. Journal of Neuroscience 2008;28: 11433-1434.
3. Fersht A. The most influential journals: Impact Factor and Eigenfactor. PNAS 2009; 106:6883-4.
4. Davis P.M. Eigenfactor: Does the principle of repeated improvement result in better estimates than raw citation counts? J am soc info sci tech 2008;59:2186-2188.
5. Moving from impact to influence: measurement and the changing role of medical journals. Eur Heart J 2012 33 (23) 2892-2896.
6. Sipido K.R. What does the future hold for cardiovascular research? Cardiovasc Res 2013 97 (1) 1-3.
7. Wang J. Citation time window choice for impact evaluation. Scientometrics 2013;94:851-872.