

Métricas de impacto e avaliação da ciência¹

Dr Lilian N. Calò²



Introdução

A avaliação da ciência refere-se à avaliação sistemática do mérito com base no tempo e nos recursos financeiros e humanos utilizados para atingir um objetivo. O processo de avaliação da pesquisa, que consiste na análise de dados e relatórios, é um processo rigoroso e sistemático que envolve a coleta de dados sobre organizações, processos, projetos, serviços e recursos. A avaliação da pesquisa tem por objetivo melhorar a tomada de decisão e levar a aplicações práticas [1].

O objetivo é, portanto, avaliar os resultados da pesquisa científica para saber o que é relevante e o que não é, subsidiar decisões sobre financiamento de projetos e traduzir esta produção científica em programas e políticas públicas para toda a sociedade. [2]

Buckeridge [3] propõe uma definição de impacto científico baseada na ideia de impacto físico. "Impacto é a capacidade de uma descoberta ou um conjunto de descobertas mudar a estrutura das redes de interação entre os modos de pensar dos seres humanos e das sociedades locais ou globais, gerando deformações (mudanças) no mundo físico. O impacto no campo intelectual causa esta "desordem" no cérebro de uma ou mais pessoas. Uma nova ideia interfere na maneira como nosso cérebro entende e interpreta os fenômenos."

As medidas ou métricas do impacto da ciência tornaram-se, nas últimas décadas, a força motriz do meio acadêmico. Pesquisadores, periódicos científicos, bases de dados e *publishers*, instituições de pesquisa e agências de fomento têm seu trabalho regido pelo impacto da pesquisa que produzem, publicam, indexam, promovem e financiam.

Há autores, porém, que discordam veementemente da ideia de métrica e avaliação da ciência para mapear as contribuições da inovação para a sociedade:

"A noção de que as contribuições para a melhoria da sociedade por meio da inovação tecnológica ou social sempre podem ser mapeadas e medidas está errada. Da mesma forma, a noção de que o principal ou único propósito das universidades é impulsionar o crescimento econômico por meio da inovação, de forma que possa ser medida com indicadores quantitativos, está errada. Claramente, a ciência contribuiu enormemente para a modernização da sociedade e para as grandes melhorias nos padrões de vida na Europa e na

¹ Tradução do original em espanhol publicado em julho/2022 na *Revista Peruana de Medicina Experimental y Salud Pública*. Disponível em: <https://doi.org/10.17843/rpmpesp.2022.392.11171> (conteúdo sob licença CC-BY 4.0).

² Coordenadora de Comunicação Científica na BIREME/OPAS/OMS.

América do Norte nos últimos duzentos anos, incluindo o desenvolvimento de uma economia e de uma sociedade com um impacto menos prejudicial à saúde e ao meio ambiente. É hora de parar de julgá-la por métricas que obviamente não fazem justiça ao seu sucesso e, acima de tudo, parar de governá-la com base no que estas métricas mostram [...] ou de quem afirma que 'se você não pode medir, você não pode melhorar' está equivocado, ou a ciência não precisa melhorar, ou meios alternativos mais precisos de avaliar a ciência devem ser desenvolvidos. Ou talvez, todos os três" [4] (tradução livre do original em inglês).

No entanto, as instituições acadêmicas estão condicionadas a operar sob uma série de métricas de avaliação que regem as políticas de carreira, como contratações, promoções, premiações e distinções, além de obter recursos financeiros para pesquisas, influenciar o mercado de editores e periódicos, os lotados rankings universitários.

Este artigo descreve os principais tipos de métricas utilizadas para avaliar a produção científica, suas características, potencialidades e limitações.

Métricas baseadas em citações

Os indicadores bibliométricos são amplamente utilizados para a avaliação da produção científica, apesar da inexistência de uma relação inequívoca entre citações e qualidade, impacto ou mérito científico. Além disso, quando se considera a inovação, característica inerente à pesquisa científica, a relação é ainda mais desconexa [5]. Além disso, existem estudos que analisam as complexidades da prática de citação [6-8], demonstrando o quão pouco pode ser assumido sobre os verdadeiros motivos da citação ao analisar o artigo finalizado. Tudo isso impacta principalmente na atribuição de relevância aos artigos com base exclusivamente nas citações recebidas e, conseqüentemente, nos modelos de avaliação da ciência em geral.

O primeiro indicador bibliométrico de que se tem notícia é o Fator de Impacto (FI®), criado em 1972 por Eugene Garfield [9] para avaliar periódicos, com a publicação do *Science Citation Index* do *Institute for Scientific Information* (ISI).

Para calcular o FI, divide-se o número de citações recebidas pelo periódico em determinado intervalo de tempo (três ou cinco anos) pelo número de artigos publicados neste mesmo intervalo. Para contabilizar as citações, é utilizada a base de dados Web of Science (WoS), (que desde 2016 pertence à Clarivate Analytics), e assim são contabilizadas apenas as citações de periódicos indexados nesta base de dados, que indexa, até o momento, aproximadamente 13.000 periódicos.

Algumas considerações sobre o cálculo do FI devem ser observadas. Se trata de um valor médio por periódico e não por artigo; nem tudo o que é publicado no periódico é contabilizado como artigo (denominador do quociente), mas as citações de um artigo que não são contabilizadas no denominador podem ser consideradas no numerador; existem dispositivos que são usados pelos editores para aumentar o FI dos periódicos. Além disso, a base de dados que dá acesso aos FIs dos periódicos, o *Journal Citation Reports* (JCR), parte integrante da WoS, é acessível por assinatura.

O FI permaneceu como o principal (e único) índice de impacto dos periódicos desde sua criação por Garfield, em 1972, até 2008, quando foi lançado o SCImago Journal Rank (SJR), medido na base de dados Scopus, do publisher multinacional Elsevier. A questão sobre o FI é que ele foi usado mais do que como um índice para classificar os periódicos. Por ser um índice de fácil cálculo, seu uso para avaliar pesquisadores, instituições, programas de pós-graduação e qualquer outra avaliação da produção científica que pudesse se beneficiar de uma avaliação qualitativa ou mais abrangente, muitas vezes ficou reduzido a uma lista de publicações associadas a um FI.

Em 2012, um grupo de editores e *publishers* de periódicos acadêmicos reunidos na *Annual Meeting of the American Society of Cell Biology* em San Francisco, EUA, redigiu um documento que ficou conhecido como a *San Francisco Declaration on Research Assessment* [10], o qual recomenda que

métricas baseadas em citações, como o FI, não sejam utilizadas para avaliar pesquisadores em situações de contratação, promoção ou decisões de concessão de financiamento para projetos de pesquisa. Em números atuais (janeiro de 2023), mais de 22.000 pessoas e organizações de 158 países assinaram a Declaração de São Francisco.

Em 2014, o Manifesto de Leiden [11], originado na 19ª Conferência Internacional sobre Indicadores em Ciência e Tecnologia, em 2014, em Leiden, Holanda, orienta o uso de métricas de avaliação científica na Europa. Até o momento, o Manifesto foi traduzido para 25 idiomas, adotado por instituições e reconhecido também por publishers de todo o mundo.

Em 2004, o publisher multinacional Elsevier lança a base de dados Scopus, disponível online por assinatura. Em 2007, o pesquisador espanhol Felix Moya-Anegón lançou o SJR, índice de impacto criado como alternativa ao FI. É calculado de forma semelhante ao FI, ou seja, citações por artigo, e também é um indicador médio por periódico, com a diferença de que o cálculo reflete o prestígio do periódico que cita [12]. Para isso, é utilizado o algoritmo *PageRank*, o mesmo do Google para ordenar as páginas mais visitadas em uma busca. Além disso, é um indicador independente de tamanho e seus valores ordenam os periódicos por “prestígio médio por artigo”. Apesar do Scopus ser uma base de acesso por assunto, o SCImago Journal Rank [13] está disponível em acesso aberto.

Nos anos seguintes, a Scopus lança novos índices para a família de indicadores Elsevier: *Source Normalized Impact per Paper* (SNIP), CiteScore metrics e o índice h para periódicos, que possuem diferentes características e aplicações, conforme descrito na página web da Elsevier. [14].

Em resposta aos lançamentos da Elsevier, em 2007 a WoS lançou os índices Eigenfactor® e Article Influence®, desenvolvidos por Carl Bergstrom e Jevin Westen na Universidade de Washington [15]. Ambos os índices utilizam o algoritmo *PageRank* do Google e ademais levam em consideração a importância das citações recebidas (de acordo com o prestígio do periódico citante). O *Eigenfactor* e o *Article Influence* ajustam-se para diferentes padrões de citação, permitindo comparar o desempenho de periódicos de diferentes disciplinas e eliminar as autocitações. Os índices são independentes de seus valores numéricos, ao contrário do IF. Além de estar disponíveis no site do JCR-WoS-Clarivate Analytics (acesso por assinatura), os dois índices estão disponíveis em uma página de acesso aberto. [15] Obviamente, apenas os periódicos incluídos no JCR têm valores de *Eigenfactor* e *Article Influence* atribuídos. Todavia, é notável que o cálculo preciso e extremamente elegante destes índices não tenha sido utilizado em nenhum sistema de avaliação de periódicos de instituições, rankings universitários ou programas de pós-graduação, que eu tenha conhecimento. Sua complexidade pode parecer difícil de interpretar aos usuários, embora permita análises mais precisas.

Em 2005, o físico J.E. Hirsch idealiza um cálculo [16] para quantificar a produtividade científica de um pesquisador ou instituição. O índice h é definido como o número de publicações com número de citações $\geq h$. Hirsch argumenta que seu índice é preferível a outros critérios de número único comumente usados para avaliar a produção científica de um pesquisador. Como o índice h favorece pesquisadores com maior senioridade científica, para permitir comparações entre cientistas de diferentes idades, é comum o uso do índice h5 ou h10. Nesses casos, são contabilizadas as publicações (e citações) dos últimos 5 ou 10 anos, respectivamente.

Existem várias maneiras de obter o índice h de um pesquisador. No WoS, por meio do recurso *Citation Report*, e no Google Scholar, por meio do perfil do autor. Em geral, o índice h calculado pelo Google Scholar é maior do que aquele aferido na WoS, que contabiliza apenas as publicações indexadas naquela base de dados.

A base de dados de pesquisa Dimensions [17] da Digital Science foi lançada em 2016 para pesquisa e consulta. Em 2018, a DS relançou uma versão estendida do Dimensions, uma plataforma de pesquisa acadêmica comercial que permite pesquisar periódicos, conjuntos de dados, auxílios à pesquisa, patentes e ensaios clínicos. A versão gratuita da plataforma permite apenas a busca de publicações e

conjuntos de dados. Estudos publicados em 2021 concluíram que o Dimensions fornece cobertura temporal e de fonte de publicação mais ampla do que Scopus e WoS na maioria das áreas temáticas, e que o Dimensions está mais próximo do Google Scholar em sua cobertura.

Uma das principais diferenças dos índices bibliométricos do Dimensions em relação ao WoS e Scopus é que o Dimensions apresenta métricas relacionadas aos documentos e não aos periódicos, como os índices FI e SJR, por exemplo. As métricas apresentadas no Dimensions Badge que se veem em certas publicações referem-se às citações recebidas pelos artigos individualmente.

Medidas de uso e download

Um dos principais desafios na utilização de indicadores de uso e download para medir o impacto de artigos, em alternativa às citações ou menções na web (Altmétrias), é a multiplicidade de plataformas de publishers em que os artigos estão disponíveis e a dificuldade em agregar as contagens de download para obter uma visão do total.

Para usar as medidas de download como um índice de "impacto" de um artigo, supõe-se que, quando um usuário visualiza o texto completo em HTML de um artigo ou baixa o PDF, se trata de uma indicação do interesse do leitor pelo artigo e como consequência, uma medida de impacto.

Uma das vantagens do uso de medidas de download em relação às citações está nos intervalos de tempo. Enquanto as citações são contadas em intervalos de 2 a 5 anos, é possível começar a contar downloads logo após a publicação online e obter métricas consistentes após apenas alguns meses.

A análise de uso e download pode ser muito útil para monitorar o desempenho de periódicos indexados em bases de dados. Por exemplo, é possível avaliar de um ano para o outro se um periódico registrou aumento ou queda no número de downloads de seus artigos, e comparar estes dados com as citações recebidas ou a evolução do índice Altmétrico de seus artigos.

É importante adotar normas de boas práticas ao registrar o uso e o download de artigos. O Código de Práticas COUNTER [18] permite que os provedores de conteúdo produzam dados de uso consistentes, comparáveis e confiáveis para seu conteúdo online. De acordo com as normas COUNTER, robôs e registros duplicados são excluídos quando um usuário acessa o mesmo artigo várias vezes na mesma seção, por exemplo, acessa o resumo, depois o HTML e logo a seguir faz o download do PDF do mesmo artigo.

De acordo com Kurtz e Bollen [19], *“ainda existem desafios consideráveis no que diz respeito à padronização do registro e agregação de medidas de uso. Na situação atual, os dados de uso são registrados em uma infinidade de formatos diferentes, cada um representando uma permutação diferente de interfaces de registro, campos de dados, semântica de dados e normalização de dados”*.

Portanto, as medidas de uso e download de artigos não são grandezas que têm significado isoladamente. Assim, são analisados por comparação entre, por exemplo, periódicos da mesma área do conhecimento, artigos individuais comparados a outros, influência do idioma ou do ano de publicação etc. A situação mais próxima do ideal ocorre quando se analisam periódicos de uma mesma plataforma ou de um determinado publisher, pois assim se eliminam muitas das variáveis enumeradas por Kurtz e Bollen [19].

Por exemplo, a plataforma SciELO [20] proporciona aos usuários dados sobre o uso de mais de 1.400 periódicos da rede. Por meio da ferramenta denominada SciELO Sushi API, é possível obter informes de acesso a um determinado artigo, periódico ou coleção. Os resultados obtidos podem ser observados por país de origem dos acessos, ano de publicação dos documentos ou idioma do documento acessado, entre outros; também é possível informar parâmetros para escolha do período a ser analisado.

Redes sociais para medir impacto científico

As redes sociais oferecem novas possibilidades na comunicação científica, criando formas de divulgação de conteúdo que agilizam o processo de publicação e avaliação, conectando pesquisadores, editores, estudantes, instituições acadêmicas, agências de fomento e a sociedade em geral.

Um estudo mostra que *“menos da metade dos artigos científicos publicados são citados uma ou mais vezes, ou seja, quando discutimos a citação como referência para o uso do artigo, invariavelmente deixamos de fora pelo menos metade das pesquisas que são feitas em o mundo”* [21]. Isso não significa que o impacto destes artigos na comunidade científica seja nulo, pelo contrário. As publicações são lidas, baixadas, compartilhadas e citadas por meio de redes sociais, blogs, canais de notícias, políticas públicas e outras formas de presença online, coletadas e mensuradas em índices como o *Altmetric – alternative metric*.

Altmetric é um serviço pago fornecido pela Digital Science para grupos de periódicos ou periódicos individuais, e indica o impacto de um artigo com base em sua divulgação nas redes sociais. Este indicador é atualizado diariamente e atribui diferentes pontuações para cada canal de comunicação [22].

A velocidade com que os artigos recém-publicados são compartilhados na web é um dos pontos fortes da altmetria em comparação com as métricas baseadas em citações, que são contadas dois a três anos após a publicação. Além disso, estudos indicam que artigos com alta presença nas redes sociais são mais divulgados e recebem mais citações. No entanto, é importante considerar a presença de artigos em outro idioma além do inglês em índices como o Altmetric. Estudos recentes [23] mostram que dos 140.000 artigos publicados no período 2015-2018 em periódicos da América Latina e Caribe em português, espanhol e inglês, apenas 13% foram mencionados na web social. Dessa fração, 57% das menções foram para artigos em inglês, 24% em espanhol e 18% em português.

Deve-se considerar também que a maioria dos desenvolvimentos e aplicações para a web, principalmente os de aplicação acadêmica, são desenvolvidos por pesquisadores para publicações em língua inglesa, razão pela qual em nações não anglófonas existem vieses no monitoramento de material na web por tais aplicativos.

Considerações finais

As métricas para avaliação de pesquisa evoluem, mudam, novas metodologias surgem e se discutem formas de aperfeiçoar os mecanismos existentes. Um tema que não abordamos neste artigo é a avaliação de projetos de pesquisa. Não porque não seja importante, pelo contrário, é muito importante e ocuparia muitas páginas. Instituições de pesquisa e agências de fomento em todo o mundo estão discutindo a maneira mais eficiente de conduzir a avaliação por pares de propostas financiamento à pesquisa [24], se é válido abrir os informes de revisão, como está sendo feito com a revisão de artigos, de acordo com as práticas de ciência aberta. De qualquer forma, é uma questão muito sensível, pois se um artigo for rejeitado para publicação, pode influenciar a carreira de um pesquisador. A avaliação de um projeto, porém, tem influência mais direta em uma área do conhecimento, pois se a pesquisa não for financiada, pode ser que ela nunca se concretize. As métricas devem ser consideradas com muita seriedade para avaliar a ciência, para não deter a própria ciência.

Referências

1. Eyre-Walker A, Stoletzki N. The assessment of science: The relative merits of post-publication review, the impact factor, and the number of citations. *PLoS Biology*. 2013; 11(10):e1001675. <https://doi.org/10.1371/journal.pbio.1001675>

2. de los Ríos R. La promoción de la investigación en salud pública: búsqueda del equilibrio entre pertinencia y excelencia. Rev Panam Salud Publica. 1999; 5(4-5):309-15. <https://doi.org/10.1590/s1020-49891999000400021>
3. Buckeridge, M.S. Ciência de Alto Impacto. [Internet] En 2º Escola de Pesquisadores do Campus USP de São Carlos, 2018. [citado el 25 de Mayo del 2022] Disponível em: <https://escritacientifica.sc.usp.br/2a-escola-de-pesquisadores-do-campus-usp-de-sao-carlos/>
4. Hallonsten O. Stop evaluating science: A historical-sociological argument. Soc Sci Inform. 2021; 60:7-26. <https://doi.org/10.1177/0539018421992204>
5. Stephan P, Veugelers R, Wang J. Reviewers are blinkered by bibliometrics. Nature. 2017; 544(7651):411-412. <https://doi.org/10.1038/544411a>
6. Erikson MG, Erlandson P. A taxonomy of motives to cite. Soc. Stud. Sci. 2014; 44(4):625-637. <https://doi.org/10.1177/0306312714522871>
7. Merton RK. The normative structure of science. In: Merton RK, ed. The Sociology of Science: Theoretical and Empirical Investigations. Chicago, IL: University of Chicago Press, 1973. pp. 267–278.
8. Cozzens SE. What do citations count? The rhetoric-first model. Scientometrics. 1989; 15(5–6): 437–47. <https://doi.org/10.1007/BF02017064>
9. Garfield E. Citation analysis as a tool in journal evaluation. Science. 1972; 178:471-479. <https://doi.org/10.1126/science.178.4060.471>
10. DORA. 2012. [Internet] The San Francisco Declaration on Research Assessment. Disponível em: <https://sfdora.org/>
11. El Manifiesto de Leiden sobre indicadores de investigación. 2015 [Internet] Disponível em: <https://www2.ingenio.upv.es/es/manifiesto>
12. González-Pereira B, Guerrero-Bote VP, Moya-Aneón F. A new approach to the metric of journals scientific prestige: The SJR indicator. J. Informetr. 2010; 4(3):379–391. <http://dx.doi.org/10.1016/j.joi.2010.03.002>
13. SCImago Journal Rank. [Internet]. Disponível em: <http://www.scimagojr.com/>
14. Elsevier. [Internet] Measuring a journal's impact.2022. Disponível em : <https://www.elsevier.com/authors/tools-and-resources/measuring-a-journals-impact>
15. Eigenfactor. [Internet] About Eigenfactor. Disponível em: <http://www.eigenfactor.org/about.php>
16. Hirsch JE. An index to quantify an individual's scientific research output. Proc. Natl. Acad. Sci. USA. 2005; 102(46):16569-72. <https://doi.org/10.1073/pnas.0507655102>
17. Dimensions. [Internet]. Disponível em: <https://www.dimensions.ai/>
18. COUNTER. 2022. [Internet] Counter Code of Practice Release 5.0.2. Disponível em: <https://cop5.projectcounter.org/en/5.0.2/>
19. Kurtz MJ, Bollen J. Usage Bibliometrics. Annual Review of Information Science and Technology. 2010; 44, 3-64. <https://doi.org/10.1002/aris.2010.1440440108>
20. SciELO. [Internet]. Disponível em: <https://scielo.org>
21. Scientific Electronic Library Online. Entrevista con Atila Iamarino [Internet]. SciELO en Perspectiva, 2013. Disponível em: <https://blog.scielo.org/es/2013/11/29/entrevista-con-atila-iamarino/>
22. Altmetric. How is the Altmetric Attention Score calculated? [Internet] Modified on 21 Sept 2021. Disponível em: <https://help.altmetric.com/support/solutions/articles/6000233311-how-is-the-altmetric-attention-score-calculated>
23. Spatti AC, Cintra PR, Bin A, Araújo RF. Métricas alternativas para avaliação da produção científica latino-americana: um estudo da rede SciELO. Inf. Inf. 2021; 26(2):596-624. <http://doi.org/10.5433/1981-8920.2021v26n2p596>
24. Global State of Peer Review [Internet]. Publons. 2018. Disponível em: <https://publons.com/community/gspr>