



Saúde & Transformação Social / Health &

Social Change

E-ISSN: 2178-7085

rodrigomoretti@ccs.ufsc.br

Universidade Federal de Santa Catarina

Brasil

Rochel de Camargo Jr, Kenneth

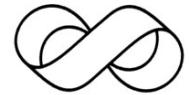
O rei está nu, mas segue impávido: os abusos da bibliometria na avaliação da ciência
Saúde & Transformação Social / Health & Social Change, vol. 1, núm. 1, 2010, pp. 3-8

Universidade Federal de Santa Catarina

Santa Catarina, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=265319560003>

- ▶ Como citar este artigo
- ▶ Número completo
- ▶ Mais artigos
- ▶ Home da revista no Redalyc



Pesquisa Qualitativa: teoria e metodologia

O rei está nu, mas segue impávido: os abusos da bibliometria na avaliação da Ciência

The king is naked but remains undaunted – the abuse of bibliometrics in the evaluation of Science

Kenneth Rochel de Camargo Jr¹

¹ Professor Adjunto da Universidade Estadual do Rio de Janeiro

RESUMO – Apesar de pesadamente criticados, indicadores bibliométricos baseados na frequência de citações continuam sendo amplamente utilizados. Este texto apresenta de forma resumida as principais críticas a este tipo de indicador, apontando para sua inadequação e insuficiência para a avaliação da ciência, propõe algumas razões possíveis para a persistência no seu uso, encerrando-se com algumas reflexões visando à sua superação.

Palavras-chave: Indicadores de Produção Científica; Indicadores Bibliométricos; Avaliação; Ciência.

ABSTRACT – Despite being heavily criticized, bibliometric indicators based on citation frequencies are still widely used. This text presents a summary the main criticisms of this kind of indicator, pointing to their inadequacy and insufficiency for scientific evaluation, proposing possible reasons for the persistency of its use, closing with some reflections geared to supersede them.

Keywords: Scientific Publication Indicators; Bibliometric Indicators; Evaluation; Science.

No clássico *Genesis and Development of a Scientific Fact*, há uma seção com o título “A tenacidade de sistemas de opinião e a harmonia das ilusões; pontos de vista como estruturas autônomas, permeadas por um estilo”^{1:27-38}. Esta seção é aberta pelo seguinte parágrafo: “Uma vez que um sistema de opiniões estruturalmente completo e fechado consistindo de muitos detalhes e relações tenha sido formado, ele oferece persistente resistência a qualquer coisa que o contradiga”^{1:27}.

Neste ensaio, pretendo apresentar sob este ângulo a persistência, a meu ver indevida, da ênfase em indicadores bibliométricos como critério para a avaliação da qualidade da produção científica. No que diz respeito à persistência, acredito dispensar qualquer demonstração; a simples consulta aos modos de operação das principais agências de fomento à pesquisa no país evidenciará fatalmente a recorrente referência à expressão “fator de impacto”, por exemplo. O que restaria argumentar é quanto à razão de considerar tal uso indevido.

Gostaria de deixar claro, contudo, que a crítica aqui apresentada não é de forma alguma uma recusa a processos avaliativos, em particular no que diz respeito à exigência de um certo padrão de publicação por parte de pesquisadores. Publicar, a

partir de sua própria etimologia, é parte fundamental da Ciência, e isto é mais relevante ainda num país como o Brasil onde a quase totalidade da pesquisa é financiada com recursos públicos. Que pesquisadores devem ter um volume razoável de publicações em veículos de boa qualidade é uma tese plenamente defensável; o problema consiste em definir o que é “razoável” e “boa qualidade”.

Outro aspecto que merece ser considerado é que, ao menos no que diz respeito à área de Saúde Coletiva, houve progresso considerável nos processos avaliativos em uso, em especial na avaliação de programas pela CAPES. Passou-se de um sistema baseado unicamente no fator de impacto para um sistema misto, com base em indicadores calculados por entidades públicas de modo transparente, e não apenas por uma empresa privada, além de adotar-se

Autor correspondente:

Kenneth Rochel de Camargo Jr

Departamento de Planejamento e Administração

Centro Biomédico

Universidade do Estado do Rio de Janeiro

Rua São Francisco Xavier, 524 – 7o. andar – Bloco D – Maracanã

Rio de Janeiro (RJ) – CEP 20559-900

Fone: (21) 25877303

Email: kencamargo@gmail.com

uma base de referência com abrangência muito maior que a base adotada pela Thomson Scientific para o cálculo de seu fator de impacto. A adoção de uma sistemática de avaliação de livros, por um lado, bem como a modificação do sistema de “patamares” para a produção científica para outro baseado em escores representam um avanço considerável no que diz respeito ao tema deste artigo. A sistemática atual permite que se trabalhe com o conjunto da produção dos pesquisadores, e não apenas no subconjunto restrito de revistas que ultrapassam uma dada classificaçãoⁱ. Ainda assim, há muito o que fazer.

As críticas ao uso de indicadores bibliométricos como sinônimo de qualidade de publicação não são novas. O próprio Eugene Garfield, inventor do fator de impacto e um dos grandes beneficiários de sua adoção, fez vários alertas sobre sua utilização indiscriminada. Por exemplo, “embora avaliações de periódicos sejam importantes, a avaliação do corpo docente é um exercício muito mais importante que afeta carreiras individuais. Números de impacto não deveriam ser utilizados como substitutos a não ser em situações excepcionais.”^{2:413} Mais recentemente, em artigo que incidentalmente inclui um elogio significativo à plataforma Lattes, Lane afirma categoricamente que “métricas em uso disseminado, desde o recente e *fashion* índice Hirsch ao cinquentenário índice de citações, são de uso limitado. Suas falhas bem conhecidas incluem o favorecimento a pesquisadores mais velhos, a captura de poucos aspectos do trabalho dos cientistas e a agregação de ciência verificada e desacreditada”^{3:488}.

Problemas conceituais

Uma primeira questão a ser discutida é o próprio evento *citação*. Todos os indicadores bibliométricos partem deste fato básico da escrita acadêmica, atribuindo-lhe um valor intrínseco, indiferenciado e cardinal, isto é, a citação é uma marca de qualidade, todas as citações se equivalem e cada citação agrega valor ao artigo citado (duas citações confeririam o dobro de valor que apenas uma). Estes são pressupostos que carecem, e muito, de discussão. Em um artigo carregado de ironia (que inclui até mesmo uma receita para pesquisadores aumentarem seus próprios índices de citações, ao qual retornarei mais tarde), Lalöe e Mosseri lembram que quando um autor faz uma citação, este não é um ato destinado à bibliometria; trata-se de “processo relativo,

fortemente contextual”; uma citação pode ser feita, por questões de conveniência, a um artigo de revisão e não à fonte original; uma citação pode mesmo ser feita de um artigo considerado errado, a fim de corrigir seus erros; as citações de artigos que descrevem métodos ou aparelhagens podem ser privilegiadas em contextos nos quais a repetibilidade de um experimento seja relevante, e assim por diante^{4:23-4}. A citação de artigos “errados” também é abordada por Todd & Ladle, que ainda chamam a atenção para a ocorrência de citações equivocadas, que se referem a um determinado texto erroneamente, atribuindo ao mesmo ideias ou fatos que este não contém⁵ – fenômeno com o qual efetivamente me deparei ao reconstruir a trajetória de construção da AIDS como categoria diagnóstica; em mais de uma ocasião identifiquei citações que atribuíam a certos autores afirmações que estes não haviam feito, ao menos não no texto citado⁶. Um estudo que fez a análise de conteúdo de vários textos (da área de psicologia) identificou, entre outros problemas, que um quarto das citações eram irrissórias (ou seja, não tinham função definida dentro do próprio texto), até 40% das mesmas eram relativas a livros ou capítulos e 10% eram autorreferências, o que mostra claramente o quanto o cômputo de citações de artigos é limitado⁷.

Problemas metodológicos

Outra classe de problemas está ligada ao próprio processo de elaboração dos indicadores, desde a definição do escopo de sua abrangência até a manipulação numérica utilizada para chegar ao indicador. O editor da publicação online PLoS/Medicine apresentou uma crítica contundente desse processo, que deveria ao menos nos fazer refletir melhor sobre a colocação de uma parte substantiva da avaliação nas mãos de uma empresa privada: “Um problema com esse jogo, deixando de lado a sua ética, é que as regras não são claras – editores podem, por exemplo, tentar persuadir a Thomson Scientific a reduzir o denominador [do cálculo do fator de impacto], mas a companhia se recusa a tornar público seu processo para escolher os tipos de artigos ‘citatíveis’. A Thomson Scientific, o único árbitro do jogo do fator de impacto, é parte da Thomson Corporation, uma organização com fins lucrativos que responde primariamente a seus acionistas. Ela não tem nenhuma obrigação de prestar contas a quaisquer dos participantes interessados que mais se importam com o fator de impacto – os autores e leitores da pesquisa científica. (...) nós percebemos que a Thomson Scientific não tem qualquer processo explícito para decidir que artigos além dos relativos a

ⁱ Ver o documento da área em http://www.capes.gov.br/images/stories/download/avaliacao/SAU_COL15out2009.pdf <Acesso em 13.07.2009>

pesquisa original ela considera citável. Nós concluímos que a ciência é correntemente avaliada por um processo que é ele mesmo não científico, subjetivo e secreto.”⁸ (ênfase minha).

Acrescente-se a isso, como já descrito anteriormente, o fato das bases de dados incluírem apenas artigos no seu processamento, deixando livros de lado^{4:24}. Outro problema é a representatividade destas bases com relação à produção mundial, observando-se com frequência a sub-representação de certas línguas ou países; padrões de citação variam também de disciplina para disciplina, o que dificulta a utilização de índices baseados em citação para comparação entre disciplinas diferentes^{9:5} – isto é particularmente relevante para áreas interdisciplinares, como é o caso da Saúde Coletiva. De um modo geral, língua, localização geográfica e escopo disciplinar determinam a priori uma base potencial de leitores, por fatores que nada tem a ver com a qualidade da publicação, mas que certamente influenciam muito a quantidade de citações. Adler, Ewing & Taylor chamam a atenção para estes fatores de confundimento, e também para a precariedade de um índice baseado exclusivamente em uma única medida de tendência central, a média^{9:1}.

A variabilidade no número de autores por artigo também tem repercussões; a contribuição de um autor, para fins de cálculo dos indicadores, é a mesma, não importando quantos coautores constam na referência. Lalöe e Mosseri comentam que “poderia parecer elementar calcular índices G' e H', estabelecidos sobre a base do número de citações dividido pelo número de autores”, o que nunca é feito^{4:24}. Em trabalho anterior¹⁰, apresentamos evidências de diferenças significativas no número mediano de coautores nos artigos das diversas subáreas da Saúde Coletiva, o que evidentemente traz repercussões para processos de avaliação.

Todd & Ladle lembram ainda que a contagem de citações é sujeita não apenas a erros como a desvios propositais (o que é chamado “*gaming the system*”, utilizar o próprio sistema para burlá-lo, inflando artificialmente o número de citações). Erros e distorções propositais não são mutualmente excludentes, o que aumenta ainda mais o potencial de distorção da medida final^{5:16}.

Problemas do construto

Um último, e talvez mais grave, problema diz respeito ao próprio *significado* do indicador. Como lembram Bollen et al., “Não dispomos de um padrão ouro,

universalmente aceito, de impacto para calibrar por comparação quaisquer novas medidas. Na verdade, não temos nem ao menos uma definição utilizável da própria noção de ‘impacto científico’, a não ser que revertamos à tautologia de definí-la como o número de citações recebido por uma publicação”^{11:2}. No artigo citado, esses autores analisaram (utilizando o método de componentes principais) 39 diferentes indicadores bibliométricos – a simples plethora de indicadores já deveria fazer soar alarmes – alguns baseados em citação e outros em uso (por exemplo, frequência de *downloads* de um portal). Além de identificar que o conjunto de medidas de uso é mais consistente internamente que as baseadas em citação^{11:9}, os autores observaram que, ao ranquear todas as medidas em termos de sua correlação umas com as outras, o fator de impacto e o SJR (calculado pelo consórcio Scimago) foram classificados em 34º e 38º lugares. Segundo esses autores, “esses resultados deveriam levar aqueles que consideram o fator de impacto como ‘padrão ouro’ do impacto científico a ter mais cautela (...) [esses indicadores] expressam um aspecto algo particular do impacto científico que pode ou não estar no núcleo da noção de ‘impacto’ científico”^{11:10}. O mesmo ponto é reafirmado por praticamente todos os autores consultados, com destaque particular para o texto de Adler, Ewing e Taylor⁹, apresentado como informe oficial de relevantes instituições das áreas de matemática e estatística. Entre outras afirmações relevantes destes autores, destaco as seguintes: “Embora os números pareçam ser ‘objetivos’, sua objetividade pode ser ilusória. O significado de uma citação pode ser ainda mais subjetivo do que a revisão por pares. Porque esta subjetividade é menos evidente para as citações, aqueles que utilizam dados de citação são menos propensos a entender as suas limitações.”^{9:1} “Aqueles que promovem as estatísticas de citações como medida predominante da qualidade da pesquisa não respondem à pergunta essencial: o que significam as citações? Eles juntam grandes quantidades de dados sobre o número de citações, processam os dados a fim de obter estatísticas e, em seguida afirmam que o processo de avaliação resultante é ‘objetivo’. No entanto, é a *interpretação* das estatísticas que leva à avaliação e a interpretação baseia-se no *significado* das citações, o que é bastante subjetivo”⁹ (ênfases dos autores). Ou seja, os indicadores disponíveis são num certo sentido números sem maior significado, um construto matemático sem correspondência relevante com os processos efetivos de produção científica e menos ainda de estabelecimento de sua qualidade.

A harmonia das ilusões

E, no entanto, esse uso, desacreditado por matemáticos e pesquisadores de todas as estirpes, persiste. O aparente contra-senso expresso pela utilização de medidas desacreditadas como padrão ouro numa área tão sensível como a avaliação da ciência parece evocar o comportamento de um personagem de anedota, que, embriagado, rastejava ao pé de um poste de iluminação pública em frente à sua casa. Indagado por um vizinho sobre o que fazia, respondeu que procurava sua chave, para poder abrir a porta. O vizinho pergunta se a chave havia sido perdida ali, ao que o protagonista de nossa estória responde que não, que havia perdido na entrada de sua casa. Surpreso, o vizinho faz a pergunta óbvia, porque então procurava a chave naquele local. A resposta: "por que aqui tem mais luz".

A conveniência e facilidade de produção de um número oculta o fato de que a busca por indicadores de citação que seriam inherentemente mais precisos, por substituir julgamentos complexos por números simples, é um grave equívoco. Esta crença, apontada como não justificada por Bollen et al.¹¹, Gingras e Adler¹⁹, Ewing & Taylor⁹ e Browman & Stergiou¹², entre outros, é comparada por Lalöe e Mosseri a uma crença irracional, como a dedicada à astrologia ou a numerologia, que guardam semelhança com um procedimento científico mas jamais passariam pelos rigorosos procedimentos de uma verdadeira avaliação científica^{4:24}.

Sem dúvida nenhuma, a conveniência e suposta "objetividade" de indicadores numéricos ajudam a explicar sua persistência frente à crítica persistente, consistente e por autores com efetiva autoridade epistêmica para fazê-las. Mas seria isso razão suficiente? A própria necessidade de indicadores é questionada de modo contundente por Browman e Stergiou, ao afirmarem que "se você necessita de uma métrica para saber quem são as melhores pessoas no seu campo, ou quem são os melhores e mais promissores novatos, então nós afirmamos que, por definição, você não é um *expert* no assunto"^{12:2:3}. A insistência no uso de tais indicadores é ela mesma uma questão complexa, e reduzi-la a um único aspecto seria um equívoco da mesma ordem do que se critica quanto à sua adoção cega. De todo modo, não é possível deixar de lado os interesses econômicos que movimentam o que denominei em outro texto de "indústria do conhecimento"¹³; em um mercado editorial controlado por poucas empresas, a concorrência entre diferentes periódicos necessita de marcadores de distinção como parte de seu *marketing*. Da mesma forma, não é possível ignorar as

relações de poder que presidem a adoção de indicadores que privilegiam alguns e enfraquecem outros; sistemas de avaliação se prestam tanto a refletir quanto a reforçar tais relações^{14,12}.

Seja qual for o motivo da continuidade das avaliações centradas em indicadores de citação, o fato é que as mesmas geram importantes consequências para pesquisadores, programas de pós-graduação, instituições de ensino e pesquisa. Apenas como exemplo, um dado artigo cita que artigos resultantes de pesquisas laboratoriais são publicados em revistas com fatores de impactos geralmente maiores do que os oriundos de pesquisas clínicas, levando as universidades a seletivamente privilegiarem os primeiros artigos nas avaliações da pesquisa, o que por sua vez está causando uma redução no financiamento da investigação clínica^{15:564}. Numa situação tal como a descrita, de competição selvagem por recursos cada vez mais escassos, é quase que inevitável a adoção de estratégias utilitaristas que procurem o máximo rendimento, ainda que paradoxalmente às custas da qualidade que se supunha avaliar. É nesse contexto que as sarcásticas "recomendações" de Lalöe e Mosseri se encerram com a seguinte sugestão: "não dêem muita importância à missão fundamental dos pesquisadores, a produção de conhecimento, em particular ao redigir seus artigos; é a comunicação que vem em primeiro lugar"^{4:24}.

Um possível efeito particularmente perverso é a redução da própria diversidade editorial garantida pela existência de múltiplas publicações nas diversas áreas. Editores de diversas revistas europeias (de filosofia, ciências sociais e humanas, *science studies*, entre outros) reagiram fortemente à tentativa recente de imposição de um ranking sobre as mesmas, elaborando um editorial que foi publicado simultaneamente em mais de quarenta periódicos. Entre outras considerações, os editores destas revistas afirmaram que "a qualidade dos periódicos não pode ser separadas de seus conteúdos e seus processos de revisão. Excelentes pesquisas podem ser publicadas em qualquer lugar e em qualquer idioma. Um trabalho realmente inovador pode mais provavelmente aparecer de fontes marginais, dissidentes ou inesperadas, ao invés do *mainstream* bem estabelecido e entrincheirado. Nossas publicações são diversas, heterogêneas e distintas. Algumas são destinadas a um público amplo, geral e internacional, outras são mais especializadas em seus conteúdos e audiência implícita. Seu alcance e leitores não dizem nada sobre a qualidade do seu conteúdo intelectual."^{16:3} Ou seja, uma comunidade científica ativa necessita de um amplo "ecossistema" de

publicação, para que a heterodoxia não esterilize a indispensável diversidade de pensamento. Isso significa a necessidade de valorizar diversas publicações para não constranger autores a procurarem apenas um pequeno grupo de veículos, por um lado, nem condenar à extinção as publicações excluídas do seletivo círculo de vencedoras.

Para pensar o futuro

Apesar de tudo que foi dito anteriormente, persiste o desafio de avaliar de forma imparcial, justa e consequente a produção científica de pesquisadores, tarefa hercúlea frente à enorme quantidade de dados a avaliar. Só na área de Saúde Coletiva, são centenas de pesquisadores, milhares de revistas e outros tantos artigos a cada período de referência, configurando a impossibilidade de avaliar-se individualmente cada publicação de cada autor. Os índices bibliométricos apresentam-se desta forma como recurso conveniente para rapidamente considerar um grande volume de dados. Mas parte do problema talvez esteja justamente em considerar inevitável o grande volume de dados. Apesar do crescimento do número de pesquisadores é razoável supor-se que a adoção de políticas do tipo *publish or perish* tenham ainda maior responsabilidade no grande aumento da produção. É possível neste contexto que o aumento do número de artigos não tenha necessariamente como contrapartida um aumento, ou mesmo manutenção, da qualidade geral dos mesmos. Ao menos um trabalho já mostrou correlação entre este tipo de política e a presença de um viés de publicação de resultados positivos¹⁷. É razoável portanto supor que a adoção de critérios de avaliação que estimulam a produção de quantidades cada vez maiores de títulos publicados contribua para a deterioração da relação sinal/ruído na comunicação científica. A adoção de medidas jocosamente denominadas de “subir o sarrado” (isto é, exigir a cada ciclo de avaliação um quantitativo mais elevado de publicações) é, portanto, contraprodutiva a médio e longo prazo, se o que se deseja é ciência de qualidade, e não simplesmente número de títulos publicados.

O que me parece claro é a necessidade de realização de pesquisas interdisciplinares sobre avaliação da ciência, incluindo não apenas especialistas em bibliometria e estatísticos, mas também economistas, cientistas políticos e filósofos da ciência, entre outros, além de *experts* dos diversos campos do conhecimento. É necessário que saímos da atual situação de adoção de critérios *ad-hoc* para a avaliação; como na maior parte das vezes o que se

busca são critérios de *diferenciação*, e não exatamente de *avaliação*, com frequência limites arbitrários são predeterminados, como o estabelecimento de percentuais máximos para estratos superiores para fins de ranqueamento.

Talvez seja chegado o momento de ousarmos um pouco mais na adoção de critérios de avaliação mais complexos e menos apoiados em “números mágicos”, mas nem por isso menos rigorosos, para evitar ao menos alguns dos problemas criados pelos próprios processos de avaliação. Isto é particularmente relevante para a Saúde Coletiva, campo complexo em todas as suas dimensões, que inclui de forma interdisciplinar aspectos de produção de conhecimento e intervenção, tendo como denominador comum um objeto igualmente complexo, a saúde dos coletivos humanos.

Apenas a título de exemplo e como contribuição inicial para o debate, consideremos a questão do ranqueamento de revistas. À parte a negação mais radical do mesmo, como o editorial já citado de Fox e Andersen et al.¹⁶, que terminava por solicitar a exclusão das revistas signatárias da lista que se propunha (embora não conste que os mesmos editores também tenham feito solicitação semelhante à Thomson Scientific ou ao consórcio Scimago), é possível pensar em lógicas de ranqueamento de publicações que não se prendam exclusivamente a indicadores numéricos. Decerto isto é impossível para o conjunto total de publicações dos pesquisadores de uma determinada área, mas pode-se, por exemplo, aplicar uma avaliação qualitativa para as principais revistas da mesma área; no caso da Saúde Coletiva, em que pese sua interdisciplinariedade, temos claramente um conjunto de revistas, identificável até pela concentração de artigos publicados nas mesmas, que aborda boa parte do espectro temático da área. Essas revistas, não mais que uma dúzia, poderiam ser avaliadas qualitativamente, sendo as demais enquadradas por critérios bibliométricos. Esse critério teria a vantagem adicional de não ser uma barreira tão difícil para novos periódicos quanto é a utilização exclusiva de índices de citação.

Formas mais inovadoras de avaliação já estão em uso em outros países. Lawrence apresenta algumas alternativas exequíveis, aplicadas em avaliações no Canadá, EUA e Reino Unido, baseadas em uma abordagem qualitativa. Um dos exemplos que propõe parece particularmente interessante: “Tudo o que os solicitantes teriam que fazer, caso escolhessem essa opção, seria a lista de, digamos, três dos que eles acham que são os seus melhores artigos dos últimos cinco anos e descrever sucintamente quem foi

responsável pelo trabalho representado nos mesmos. Uma vantagem de ter um pequeno número de trabalhos para avaliar, além da óbvia de ter menos a ler, é que os autores seriam incentivados a escrever um menor número de artigos que signifiquem e transmitam algo importante, ao invés de um grande número de trabalhos a serem contados.”^{18:11}

Em suma, indicadores bibliométricos, ainda que tendo utilidade limitada, são um recurso pobre e perigoso para a avaliação da qualidade da produção científica. O conforto ilusório de uma medida tido como “objetiva” não dá conta das complexas dimensões da tarefa. O debate sobre formas de avaliação deve permanecer aberto e os sistemas institucionais devem ser continuamente repensados e aperfeiçoados. O que não podemos nos permitir é simplesmente continuar repetindo práticas que criticamos, e cujas consequências nefastas antecipamos. Precisamos ter a clareza e a coragem de buscar o que desejamos onde pode efetivamente ser encontrado, e não onde é mais fácil procurar.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Fleck L. *Genesis and development of a scientific fact*. Chicago: The Chicago University Press; 1979.
2. Garfield E. *Fortnightly Review: How can impact factors be improved?* BMJ 1996; 313:411-413.
3. Lane J. Let's make science metrics more scientific. *Nature* 2010; 464(25):488-9.
4. Laloë F, Mosseri R. L'évaluation bibliométrique des chercheurs: même pas juste... même pas fausse! *Reflets phys* 2009; 13:23-4.
5. Todd PA, Ladle RJ. Hidden dangers of a 'citation culture'. *Ethics Sci Environ Polit* 2008; 8: 13–16.
6. Camargo KR. As ciências da AIDS e a AIDS das ciências. Rio de Janeiro: Relume-Dumará; 1994.
7. Krampen G, Becker R, Wahner U, Montada L. On the validity of citation counting in science evaluation: Content analyses of references and citations in psychological publications. *Scientometrics* 2007; 71(2):191–202.
8. The PLoS Medicine Editors. The impact factor game. *PLoS Medicine* 2006; 3(6):291.
9. Adler R, Ewing J, Taylor P. Citation Statistics - A Report from the International Mathematical Union (IMU) in Cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS). *Statistical Science* 2009; 24(1):1–14.
10. Camargo KR, Medina CC, Caetano R, Rangel MV. Produção intelectual em saúde coletiva: epistemologia e evidências de diferentes tradições. *Rev Saúde Pública* 2010; 44(3):394-8.
11. Bollen J, Sompel HV, Hagberg A, Chute R. A principal component analysis of 39 scientific impact measures. <http://arxiv.org/abs/0902.2183v1>. <Acesso em 12.07.2010>
12. Browman HI, Stergiou KI. Factors and indices are one thing, deciding who is scholarly, why they are scholarly, and the relative value of their scholarship is something else entirely. *Ethics Sci Environ Polit* 2008; 8:1–3.
13. Camargo KR. Public health and the knowledge industry. *Rev Saúde Pública* 2009; 43(6):1078-83.
14. Caponi S, Rebelo F. Sobre juízes e profissões: a avaliação de um campo disciplinar complexo. *Physis: Revista de Saúde Coletiva* 2005; 15(1):59-82.
15. Brown H. How impact factors changed medical publishing—and science. *BMJ* 2007; 334:561-564.
16. Fox R, Andersen H, Ariew R, Bag AK, Barrow – Green J, Benson K, et al. Journals under threat: a joint response from history of science, technology and medicine editors. *Notes Rec R Soc* 2009; 63:1-3.
17. Fanelli D. Do Pressures to Publish Increase Scientists' Bias? An Empirical Support from US States Data. *PLoS ONE* 2010; 5(4): e10271.
18. Lawrence P. Lost in publication: how measurement harms science. *Ethics Sci Environ Polit* 2008; 8: 9–11.
19. Gingras Y. La fièvre de l'évaluation de la recherche. Du mauvais usage de faux indicateurs. http://www.cirst.uqam.ca/Portals/0/docs/note_rech/2008_05.pdf <Acesso em 30.06.2010>