

Os desafios do uso de *big data* na avaliação em saúde

The challenges of using big data in health evaluation

Luiz Claudio Santos Thuler

Pesquisador Sênior do Instituto Nacional de Câncer, INCA e Professor Associado da Universidade Federal do Estado do Rio de Janeiro, UNIRIO, Rio de Janeiro, Brasil

Zulmira M. A. Hartz

Professora Catedrática Convidada, GHM, Instituto de Higiene e Medicina Tropical, Universidade NOVA de Lisboa, Portugal

Resumo

Um considerável aumento na quantidade de dados disponíveis na área da saúde tem sido observado nos últimos anos. Essa proliferação os torna crescentemente inacessíveis, gerando um desafio que é como fazer esses dados terem sentido. Neste ensaio sobre os desafios do uso de *big data* em avaliação em saúde os autores iniciam com uma abordagem histórica do crescimento da quantidade de dados observado nas últimas décadas, definem o que é *big data*, descrevem suas características e comentam sobre as principais dificuldades para a sua utilização pelos avaliadores na área da saúde. Destacam ainda as suas limitações, enquanto apontam seu uso potencial nas pesquisas não avaliativas, nos estudos de avaliabilidade, nas avaliações normativas e nas pesquisas avaliativas. Concluem chamando a atenção para a necessidade de integrar as duas abordagens: aquela baseada em dados apresentada recentemente pela Ciência dos Dados e a baseada em hipóteses proposta pelo Método Científico.

Palavras-chave:

Big Data em saúde, avaliação em saúde, estudos de avaliação.

Abstract

In recent years there has been a huge increase in the amount of data available in the health area. This proliferation makes them increasingly inaccessible, creating a challenge that is how to make that data meaningful. In this essay on the challenges of using big data in health evaluation, the authors begin with a historical approach about the growth of the amount of data observed in the last decades, define what is big data, describe its characteristics and comment on the main difficulties for use by health evaluators. They also highlight their limitations, while pointing out their potential use in non-evaluative research, evaluability assessment studies, normative evaluations and evaluative research. They conclude by calling attention to the need to integrate the two approaches: one recently presented based on data as stated by the Science of Data and the other based on hypotheses as endorsed by the Scientific Method.

Key words:

Big data in health, health evaluation, evaluation studies.

Introdução

Nos últimos anos, tem-se observado um considerável aumento na quantidade de dados disponíveis na área da saúde; vivemos um verdadeiro dilúvio de dados. Entretanto, a proliferação de dados traz um grande desafio que é como fazer esses dados terem sentido [1]. Isso faz com que os questionamentos de T. S. Eliot formulados há 85 anos continuem extremamente atuais: “Onde está a sabedoria que perdemos no conhecimento? Onde está o conhecimento que perdemos na informação?” [2].

À medida que os dados se tornam mais abundantes, o principal problema não é mais encontrá-los, mas identificar aqueles que são relevantes para tomadores de decisão [3]. De acordo com Saracci (2018) a produção atual diária de dados tem sido estimada em 2,5 exabytes (10^{18} bytes). A cada um ou dois anos mais dados sejam produzidos do que em todos os anos anteriores da história da humanidade. Projeta-se que o volume acumulado de dados pode chegar a 44 zettabytes (10^{21} bytes) em 2020, mas os dados de saúde representam apenas uma pequena fração desse enorme volume gerado pelas mais diversas atividades da sociedade [2].

Se 70 anos atrás os computadores tornaram os dados legíveis, há 30 os tornaram acessíveis por intermédio da Internet e rastreáveis por meio de motores ou ferramentas de busca. O desafio presente é aprender a usar o “computador” para triturar esses dados, obtendo uma maneira totalmente nova de entender o mundo [4]. A atual disponibilidade de grande volume de dados, e os avanços tecnológicos, incluindo as ferramentas estatísticas disponíveis, traz oportunidades e desafios sem precedentes para as investigações em saúde.

Metodologia

Trata-se de um ensaio teórico-reflexivo, baseado em pesquisa bibliográfica, que tem como objetivo refletir sobre os desafios do uso de *big data* (BD) nas avaliações em saúde face ao crescimento na quantidade de dados disponíveis, destacando suas limitações e usos potenciais.

Resultados e discussões

O termo BD foi empregado pela primeira vez em

1997, pelos cientistas da NASA Michael Cox e David Ellsworth, em alusão a conjuntos de dados muito grandes [5]. Segundo o *Cambridge Dictionary* BD pode ser definido como “conjuntos muito grandes de dados produzidos por pessoas que usam a Internet e que só podem ser armazenados, compreendidos e usados com a ajuda de ferramentas e métodos especiais” [6]. Na prática, os BD têm sido associados: 1) à disrupção no processamento tradicional de dados com desafios relacionados à quantidade de dados, seu processamento em tempo real, diversidade de fontes, e baixo custo; 2) à tomada de decisão orientada por dados; 3) ao armazenamento e processamento contínuo de grandes volumes de dados; 4) a um movimento cultural por meio do qual descobre-se como a humanidade interage com o mundo; 5) e, por fim, à ausência de necessidade de interpretação humana, com decisões baseadas em algoritmos [7].

De acordo com Forss e Norén [8] há quatro categorias principais de BD: 1) dados originados ativamente: dados não estruturados armazenados intencionalmente; 2) dados originados passivamente: dados não estruturados armazenados involuntariamente; 3) dados originados por algoritmos (ou seja, comunicação máquina a máquina): dados não estruturados que são ativamente e intencionalmente armazenados por entidades não humanas e 4) estatísticas públicas (ou dados abertos): geralmente dados estruturados que são ativamente inseridos em bancos de dados por atores e indivíduos específicos. Nesta última categoria encontram-se dados tradicionalmente utilizados na área de avaliação, quer primários quer secundários. Os dados primários referem-se a dados coletados para responder a uma questão de pesquisa específica usando um instrumento próprio (por exemplo, inquéritos e entrevistas). Já os dados secundários referem-se a dados que foram coletados para um propósito diferente da pergunta de pesquisa em questão. Os dados secundários podem vir da vigilância de rotina em Saúde Pública, inquéritos de saúde de base populacional, dados de alta hospitalar, administração e faturamento de serviços de saúde, ou de outros projetos de pesquisa [9].

As pesquisas com BD podem utilizar desde dados moleculares (Ex: uso de dados de expressão gênica para previsão de desfechos clínicos), ou teciduais (Ex: uso de dados de ressonância magnética para predições clínicas), até aqueles provenientes de pacientes (Ex: predição da readmissão em Unidade de Terapia Intensiva e da taxa de mortalidade de pa-

cientes com certas características), ou de populações (Ex: dados coletados por médicos, clínicas, hospitais ou mídia social) [10]. O desenvolvimento de novos modelos para avaliação ou estratificação riscos, integrando dados de atenção à saúde formal e informal e resultados de pesquisas, poderia ajudar a definir planos de ação mais adequados para os pacientes [11]. No Brasil, Chiavegatto Filho [12] destaca o relacionamento de bancos de dados mantidos pelo Ministério da Saúde, como o Sistema de Informações sobre Mortalidade (SIM) e o Sistema de Informações sobre Nascidos Vivos (Sinasc). A esses podem ser acrescentados, por exemplo, o Sistema de Informação Hospitalar (SIH) e o Sistema de Informação Ambulatorial (SIA), bem como bancos de dados de inquéritos e pesquisas disponibilizados no sítio do Datasus [13]. Em 2001 os BD foram inicialmente caracterizados por três letras Vs, referentes ao alto Volume, alta Velocidade e alta Variedade de informações [14].

Somente em 2012 foi proposta uma quarta letra V, referente a Valor, uma vez que será a combinação das dimensões que irá conferir valor aos BD. Em 2013 já eram listadas 10 características, acrescentando-se a validade, veracidade, volatilidade, viabilidade, visualização e variabilidade. Mais recentemente foram identificados 42 Vs [15] (Figura 1).

Os BD se constituem numa importante oportunidade de expansão da capacidade de gerar novos conhecimentos, contribuindo para a sua disseminação, e ao trazer a medicina de precisão para a prática clínica podem contribuir para a transformação da assistência à saúde [16]. Na medicina de precisão as recomendações terapêuticas devem ser personalizadas com base em pesquisas com elevada validade externa, o que pode ser obtido com grande número de pacientes, o que geralmente implica na realização de estudos multicêntricos, ou por meio do relacionamento de bases de dados públicas existentes, provenientes,

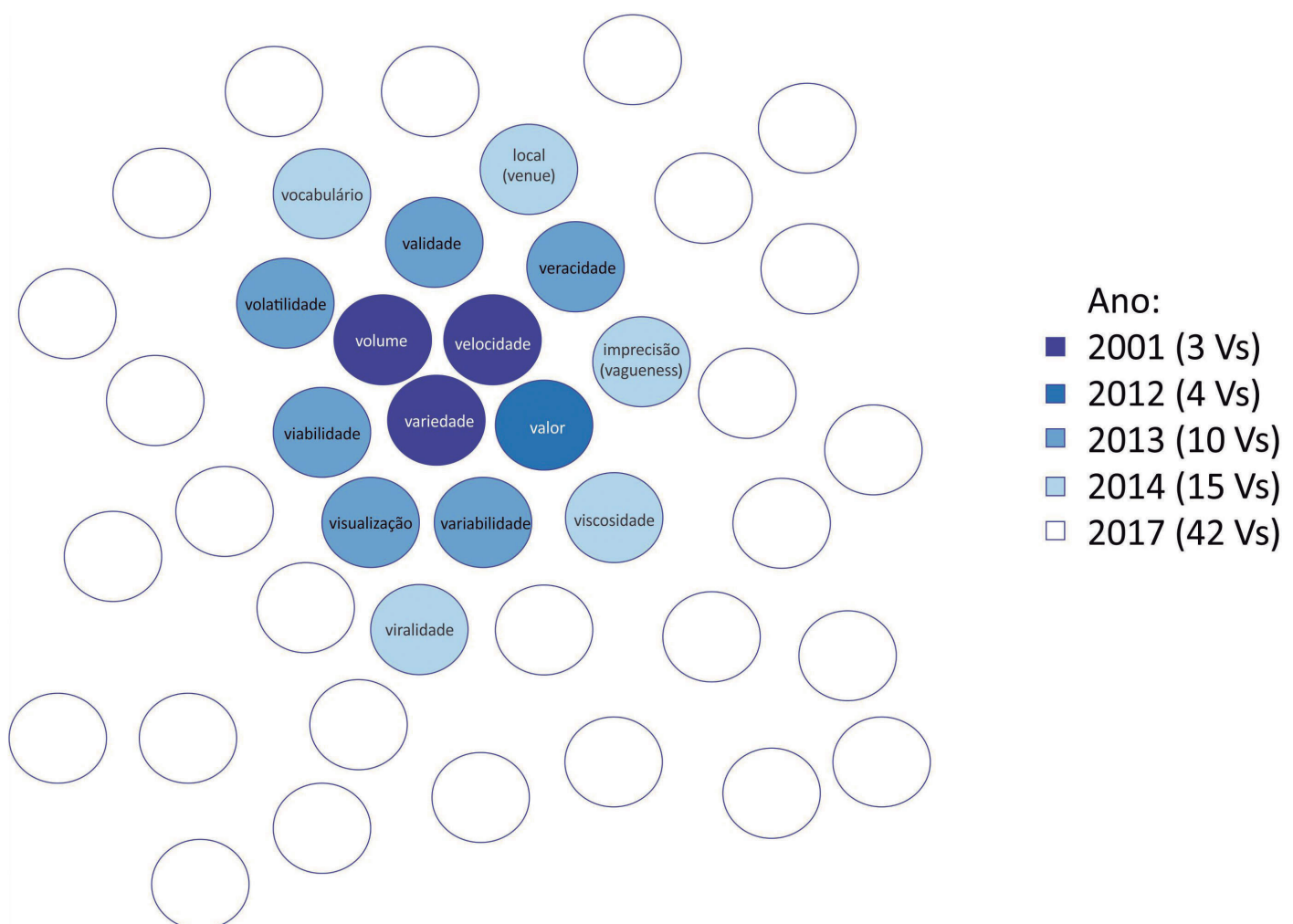


Figura 1 - Características dos *big data*

por exemplo, de prontuários eletrônicos integrados [12]. Como destacam Cano et al. (2017) um potencial único dos BD é a criação de uma estrutura digital de saúde, por meio da integração de múltiplas fontes heterogêneas de informação, gerando novos conhecimentos que impactem nos resultados da assistência médica e na experiência do paciente [11].

Segundo Salerno et al., é necessária uma perspectiva equilibrada que proporcione salvaguardas para os indivíduos, mas que também melhore a saúde da população. De forma semelhante às demais pesquisas envolvendo seres humanos, esses autores propõem que o uso de BD suscita preocupações com questões éticas, que não devem ser negligenciadas, tais como a proteção da privacidade, da confidencialidade, e dos danos individuais; a segurança e a integridade de dados; a definição de situações nas quais há necessidade de consentimento livre e esclarecido; e o cuidado com falsas inferências [9].

Avaliação em saúde e big data

A avaliação é inerente à atividade humana, uma vez que, para que possamos nos posicionar e agir socialmente, submetemos as situações, objetos, processos e práticas cotidianas a algum julgamento [17].

Embora a *United Nations Global Pulse* tenha reconhecido o potencial dos BD como fonte complementar a dados tradicionais para o monitoramento e avaliação de programas, a demanda por BD nessas áreas ainda é reconhecidamente fraca [8, 18]. Os BD não

têm sido muito usados por avaliadores, sobretudo, pela falta de transparência sobre como são gerados; faltam informações básicas sobre a representatividade das populações estudadas, critérios de inclusão, e vieses prováveis; além disso, muitas vezes as equipes de avaliação não dispõem das competências necessárias para localizar e usar os BD [8]. Ainda que existam enormes possibilidades científicas em BD, Lazer et al. (2014) destacam que a quantidade de dados não nos autoriza a ignorar aspectos básicos da teoria das medidas, validade de construto, confiabilidade e dependência entre as fontes de dados [19].

De acordo com Dias & Duarte (2015) a utilização de BD para melhor informar as tomadas de decisão pode gerar grandes economias aos sistemas de saúde, otimizando a inovação, melhorando a eficiência de pesquisas e ensaios clínicos e construindo novas ferramentas para médicos, consumidores, seguradoras e reguladores para cumprir a promessa de abordagens mais individualizadas [16].

Entretanto, conforme destacam Brousselle et al. (2011) as tomadas de decisão só em parte se apoiam em pesquisas ou em avaliações [17]. Neste cenário, destaca-se o potencial do uso de BD tanto nas pesquisas não avaliativas, como nas avaliações normativas e nas pesquisas avaliativas. Do mesmo modo, os estudos de avaliabilidade, que devem anteceder avaliações mais complexas a fim de determinar o quanto o programa é avaliável e qual seria o propósito e o foco da avaliação [20], poderiam se beneficiar dos BD disponíveis, favorecendo a racionalização de recursos e maximizando a abordagem avaliativa (Figura 2).

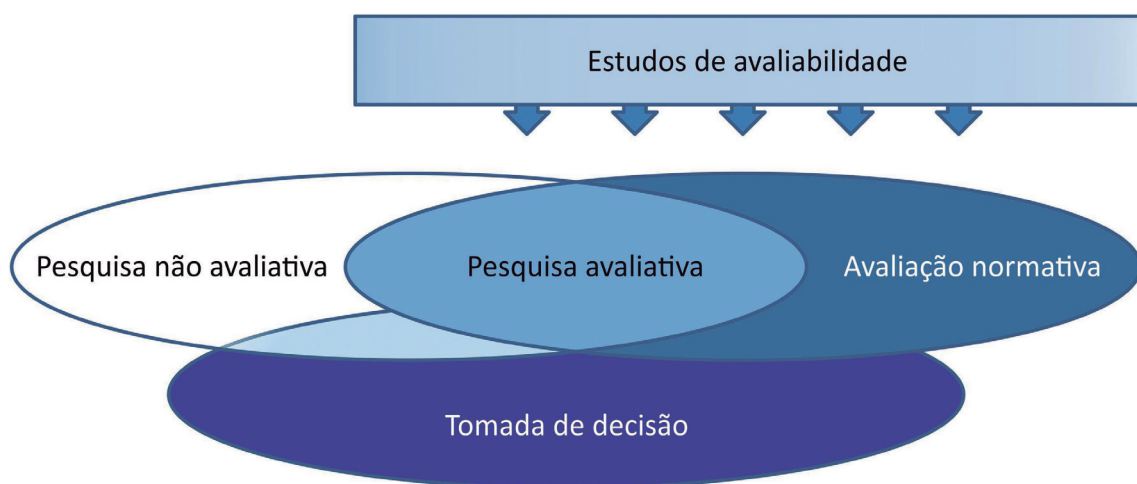


Figura 2 - Usos potenciais dos Big Data em pesquisa e avaliação para a tomada de decisão
Fonte: adaptado de Brousselle et al., 2011.

Pode-se ainda especular que os BD poderiam, por exemplo, contribuir para tornar mais eficientes as avaliações, uma vez que dados colhidos passivamente podem ser mais relevantes do que dados de pesquisas e entrevistas, ao expor “o que as pessoas fazem” em vez de “o que as pessoas dizem” [8].

Cabe enfatizar que a abordagem avaliativa, alvo deste ensaio, deve principiar por identificar respostas a questões iniciais tais como [17]: 1) Qual é o problema a ser avaliado? 2) Qual é a intervenção e quais são os seus componentes? 3) Qual é o modelo lógico? 4) Qual é o contexto? 5) Quais são as finalidades da avaliação? 6) Qual a perspectiva da avaliação? 7) Qual etapa da intervenção está sendo avaliada? 8) Qual é a viabilidade operacional da intervenção? 9) Há recursos disponíveis para desenvolver a avaliação?

Como enfatizam Potvin e Bisset (2014), a prática da avaliação requer competências metodológicas e técnicas para condução de investigações sistemáticas [21]. Essa prática, assim como um paradigma, exige uma racionalidade organizada, da qual fazem parte cinco dimensões interrelacionadas: teologia, ontologia, epistemologia, metodologia e deontologia.

A dimensão teleológica diz respeito ao que o profissional coloca como meta ou objetivo central da avaliação. Ao definir a intencionalidade, contribui para a formulação da visão, missão, objetivos e metas.

Já a dimensão ontológica tem relação com os objetos que a prática busca regular reproduzir ou transformar. Podem ser indivíduos, classes, atributos ou relacionamentos.

Por sua vez, a dimensão epistemológica tem a ver com o relacionamento entre o profissional e o objeto de sua prática. Refere-se ao conhecimento que te-

mos sobre os programas ou políticas a serem avaliados que nos permitam refletir sobre como alterá-los. Neste contexto, tanto o objeto a ser avaliado quanto a avaliação influenciam um ao outro.

A dimensão metodológica se refere ao tipo de ação que os profissionais conduzem para alcançar seus objetivos. Essa dimensão diz respeito à análise da possibilidade de avaliação, definição da pergunta e do projeto de avaliação e uso do conhecimento.

Finalmente, a dimensão deontológica refere-se à competência técnica do profissional e ao seu compromisso com valores e deveres.

Conclusões

Concluindo, como destacam Petersson, Leeuw e Olejniczak (2017), é essencial que avaliadores invistam na habilidade de combinar diferentes teorias, métodos e perspectivas, aumentando o uso dos BD [22]. O uso dos BD deve ser encorajado nas pesquisas não avaliativas, nos estudos de avaliabilidade, nas avaliações normativas e nas pesquisas avaliativas, cabendo ao avaliador escolher a base de dados que mais se apropria ao seu objeto de estudo. Há muito que aprender com comunidades emergentes, como as dos BD, mas há também muitas visões a serem compartilhadas pelos avaliadores em retribuição. Os BD podem nos dar as respostas, mas o que não podemos perder de vista é a singularidade de elaborar as questões. Cabe, neste cenário, promover pesquisas avaliativas e intervenções apoiadas em dados como propõe a Ciência dos Dados e em hipóteses como preconiza o Método Científico.

Bibliografia

1. Data, data everywhere. *Economist* 27 de fevereiro de 2010. [Consultado em 25 de setembro de 2018]. Disponível em: <http://www.emc.com/collateral/analyst-reports/ar-the-economist-data-data-everywhere.pdf>
2. Saracci R. Epidemiology in wonderland: Big Data and precision medicine. *Eur J Epidemiol* 2018; 33: 245-257.
3. Magalhães J, Hartz Z, Martins MRO. Big Data para a investigação em saúde e a ciência aberta: um contributo para a gestão do conhecimento. *An Inst Hig Med Trop* 2016; Supl. 2: S75-S82.
4. Anderson C. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete 23 de junho de 2008. [Consultado em 17 de setembro de 2018]. Disponível em: <https://www.wired.com/2008/06/pb-theory/>
5. Cox M, Ellsworth D. Application-Controlled Demand Paging for Out-of-Core Visualization. *Proceedings of the 8th Conference on Visualization* 97; 1997. Disponível em: https://www.researchgate.net/publication/3736976_Application-controlled_demand_paging_for_out-of-core_visualization
6. Dicionário Cambridge Inglês-Português. Definição de “big data”. © Cambridge University Press, 2019. [Consultado em 24 de abril de 2019]. Disponível em: <https://dictionary.cambridge.org/pt/>
7. Lefebvre-Naré F, Lemire S, Petersson GJ. What is big data? In: Petersson GJ, Breul JD. In: *Cyber Society, Big Data, and Evaluation: Comparative Policy Evaluation*. Taylor & Francis, NY, USA; 2017.
8. Forss K, Norén J. Using ‘bigdata’ forequity-focused evaluation – understanding and utilizing the dynamics of data ecosystems. In: Petersson GJ, Breul JD. In: *Cyber Society, Big Data, and Evaluation: Comparative Policy Evaluation*. Taylor & Francis, NY, USA; 2017.
9. Salerno J, Knoppers BM, Lee LM, Hlaing WM, Goodman KW. Ethics, Big Data and Computing in Epidemiology and Public Health. *Annals of Epidemiology* 2017; 27(5):297–301.
10. Herland M, Khoshgoftaar TM, Wald R. A review of data mining using big data in health informatics. *Journal of Big Data* 2014; 1:2.
11. Cano I, Tenyi A, Vela E, Miralles F, Roca J. Perspectives on Big Data applications of health information. *Current Opinion in Systems Biology* 2017; 3:36-42.
12. Chiavegatto Filho AD. Uso de big data em saúde no Brasil: perspectivas para um futuro próximo. *Epidemiol. Serv. Saúde* 2015; 24(2):325-332.
13. Ministério da Saúde (BR), DATASUS. Rio de Janeiro: Coordenação-Geral de Disseminação de Informações em Saúde. 2019. Disponível em: <http://datasus.saude.gov.br/informacoes-de-saude>
14. Laney D. 3D Data Management: Controlling Data Volume, Velocity, and Variety. *Application Delivery Strategies* 6 February 2001. 949: 1-3. Disponível em: <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
15. Shafer T. The 42 V’s of Big Data and Data Science. 2017. [Consultado em 23 de abril de 2019]. Disponível em: <https://www.elderresearch.com/company/blog/42-v-of-big-data>
16. Dias JA, Duarte P. Big data opportunities in healthcare. How can medical affairs contribute? *Rev Port Farmacoter* 2015; 7:230-236.
17. Brousseau A, Champagne F, Contandriopoulos A-P, Hartz Z. *L'évaluation: concepts et methods*, 2^e edition, Presses de l'Université de Montréal, Montréal, Canada, 2011.
18. United Nations Global Pulse (2016). *Integrating Big Data Into the Monitoring and Evaluation of Development Programmes*. [Consultado em 13 de setembro de 2018]. Disponível em: http://unglobalpulse.org/sites/default/files/IntegratingBigData_intoMEDP_web_UNGP.
19. Lazer D, Kennedy R, King G, Vespignani A. The parable of Google Flu: traps in big data analysis. *Science* 2014; 343:1203-1205. [Consultado em 19 de setembro de 2018]. Disponível em: <https://gking.harvard.edu/files/gking/files/0314policyforumff.pdf>
20. Thurston W, Potvin L. Evaluability assessment: a tool for incorporating evaluation in social change programmes. *Evaluation* 2003; 9(4):453–469. Disponível em: http://www.stes-apes.med.ulg.ac.be/Documents_electroniques/EVA/EVA-PROG/ELE%20EVA-PROG%207370.pdf
21. Potvin L, Bisset S. Há mais na metodologia do que o método. In Hartz Z, Potvin L, Bodstein R. *Avaliação em promoção da saúde: Uma antologia comentada da parceria entre o Brasil e a Cátedra de abordagens comunitárias e iniquidades em saúde (CACIS), da Universidade de Montreal de 2002 a 2012*. CONASS, Brasília-DF, Brasil, 2014.
22. Petersson GJ, Leeuw F, Olejniczak K. *Cyber Society, Big Data and Evaluation: A Future Perspective*. In: *Cyber Society, Big Data, and Evaluation: Comparative Policy Evaluation*. Taylor & Francis, NY, USA; 2017.

Conflitos de interesses:

Os autores declaram que não existem conflitos de interesses.