

# O uso do Big Data em pesquisas médicas

## Big Data use in medical research

Thiago Gonçalves dos Santos Martins<sup>1,2</sup>, Ana Luiza Fontes de Azevedo Costa<sup>1,2</sup>,  
Thomaz Gonçalves dos Santos Martins<sup>3</sup>

<sup>1</sup> Universidade Federal de São Paulo, São Paulo, SP, Brasil.

<sup>2</sup> Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, São Paulo, SP, Brasil.

<sup>3</sup> Universidade Estácio de Sá, Rio de Janeiro, RJ, Brasil.

**DOI: 10.1590/S1679-45082018ED4087**

Atualmente, a velocidade e o volume de dados criados a partir de sites da internet, sensores eletrônicos e celulares estão na ordem de exabytes (1 bilhão de gigabytes) a cada 2 dias. Trata-se da quantidade correspondente ao produzido no período desde o início da civilização até 2003. Este número impressionante tende a dobrar a cada 40 meses.<sup>(1)</sup>

O enorme conjunto de dados, que é grande o suficiente para ser considerado não gerenciável para a interpretação humana sem auxílio de um processo computadorizado e/ou analítico de dados, é conhecido como Big Data. Apesar do volume e da velocidade quase em tempo real do processamento de dados, é necessário trabalhar a qualidade deles, para que gerem informações úteis.

Médicos atuam em pesquisas de aprendizado de máquinas (*machine learning*), que é a elaboração de algoritmos que respondam e se adaptem automaticamente aos dados, sem necessidade de intervenção humana contínua, visando ao desenvolvimento de inteligência artificial que ajude a tomar decisões, considerando este grande volume de dados.<sup>(2,3)</sup> Estes profissionais podem ajudar na programação destas máquinas, a fim de garantir confiabilidade no padrão de decisão. Este processo nos faz refletir sobre a formação médica, buscando novas habilidades e formas de atuar que permitam selecionar adequadamente as informações, para facilitar a tomada de decisões na prática clínica.<sup>(4,5)</sup>

Pesquisas internacionais multicêntricas se desenvolvem de forma mais simples, envolvendo um número crescente de participantes com menor custo. Para tal, disporemos de uma imensa base de dados armazenada em memórias de diversos equipamentos eletrônicos, interligados com a rede, facilitando também o acesso a prontuários eletrônicos. A coleta de dados não poderá ser feita sem o devido consentimento, o que nos obriga a reforçar o debate ético sobre a necessidade da adequação da legislação a esta nova realidade.<sup>(6,7)</sup>

Um dos desafios futuros será o processo de aquisição da autorização para o uso dos dados já automaticamente coletados por todos os equipamentos de uso diário ligados à internet. Para proteger a privacidade do paciente, serão necessárias novas técnicas de criptografia, de modo a garantir o sigilo dos dados. Nesse tipo de pesquisa, com milhares de testes de hipótese ocorrendo simultaneamente, a possibilidade de uma associação aleatória deve ser considerada como significativo fator de risco.

### Como citar este artigo:

Martins TG, Costa AL, Martins TG.

O uso do Big Data em pesquisas médicas.

einstein (São Paulo). 2018;16(3):eED4087.

<https://doi.org/10.1590/S1679-45082018ED4087>

### Autor correspondente:

Thiago Gonçalves dos Santos Martins

Rua Botucatu, 821, 2º andar – Vila Clementino

CEP: 04023-062 – São Paulo, SP, Brasil

Tel.: (11) 6456-2182

E-mail: thiagogsmartins@yahoo.com.br

### Data de submissão:

15/4/2017

### Data de aceite:

29/6/2018

### Copyright 2018



Esta obra está licenciada sob  
uma Licença *Creative Commons*  
Atribuição 4.0 Internacional.

Saber lidar com amostras enviesadas sempre ocupou boa parte do tempo dos médicos.<sup>(8)</sup> Manejar um volume enorme de dados gerados na Medicina (o Big Data) pode revolucionar o poder decisório do médico e seu conhecimento sobre muitas doenças. Porém, saber coletar, selecionar e analisar os dados obtidos a partir da divulgação de resultados em tempo real envolverá um novo aprendizado na Medicina, para evitar que esta grande quantidade de dados analisados sem uma profunda compreensão de seu contexto produza apenas um big noise (o ruído é aquilo que atrapalha a comunicação entre o emissor e o receptor, na teoria da comunicação), demonstrando a importância da validação cruzada dos dados achados, utilizando outros conjuntos de dados para confirmação da reprodutibilidade e avaliação da possibilidade de generalização.<sup>(9)</sup>

O Big Data nos oferece a oportunidade de construir uma grande base de dados brasileira, que poderá ser útil para desenvolver, avaliar e melhorar continuamente as diretrizes de práticas clínicas, além de servir como fonte de dados para diversos estudos multicêntricos, tanto nacionais quanto internacionais. Com isso, haverá um proeminente ganho de tempo, dinheiro, vidas e conhecimento.

O futuro da Medicina volta-se para o desenvolvimento de sensores para monitorização de funções vitais e novas moléculas marcadoras de doenças combinadas com supercomputadores, capazes de processar um

imenso volume de dados, criando um sistema global de apoio ao diagnóstico médico.<sup>(10)</sup>

### INFORMAÇÃO DOS AUTORES

Martins TG: <https://orcid.org/0000-0002-3878-8564>

Costa AL: <https://orcid.org/0000-0002-6925-4207>

Martins TG: <https://orcid.org/0000-0001-5440-8149>

### REFERÊNCIAS

1. Varian HR. Big data: new tricks for econometrics. *J Econ Perspect*. 2014;28(2):3-28.
2. Story JD. A direct approach to false discovery rates. *J R Stat Soc B*. 2002;64(3):479-98.
3. McKenna J. Big data: big promise. *Eur Heart J*. 2017;38(7):470-1.
4. Rumsfeld JS, Joynt KE, Maddox TM. Big data analytics to improve cardiovascular care: promise and challenges. *Nat Rev Cardiol*. 2016;13(6):350-9. Review.
5. Simpaio AF, Ahumada LM, Rehman MA. Big data and visual analytics in anaesthesia and health care. *Br J Anaesth*. 2015;115(3):350-6. Review.
6. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature*. 2009;457(7232):1012-4.
7. Grimmer J. We are all social scientists now: how big data, machine learning, and casual inference work together. *PS: Political Sci Politics*. 2015;48(1):80-3.
8. Câneo PK, Rodina JM. Prontuário eletrônico do paciente: conhecendo as experiências de sua implantação. *J Health Inform*. 2014;6(2):67-71.
9. Gabriel SE, Normand SL. Getting the methods right — the foundation of patient-centered outcomes research. *N Engl J Med*. 2012;367(9):787-90.
10. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol*. 2005;58(4):323-37.