

Elliot Naidus¹, Leo Anthony Celi¹

***Big Data* em saúde: estamos perto?**

Big Data in healthcare: are we close to it?

1. Beth Israel Deaconess Medical Center -
Massachusetts, United States.

INTRODUÇÃO

A tradução da pesquisa médica em diretrizes para a prática clínica não é tarefa fácil. O número de artigos biomédicos publicados tem crescido muito,⁽¹⁾ mas a forma como os médicos adaptam esses artigos à prática não é direta. Além disso, a validade da pesquisa biomédica tem sido recentemente questionada.⁽²⁾ A ocorrência de vieses em publicações com ênfase em descobertas sensacionais em oposição à reprodutibilidade, a não aceitação de estudos negativos, e a pressão acadêmica por publicar têm, em conjunto, contribuído para a falta de confiabilidade da pesquisa biomédica. Uma das consequências é o fenômeno de “pêndulo médico”, tratamentos e ferramentas diagnósticas que são considerados benéficos em uma década, para mais tarde se comprovarem destituídos de valor, ou pior, danosos. Um exemplo na terapia intensiva é o uso do cateter arterial pulmonar, amplamente difundido na década de 1980 e início da década de 1990, mas que, mais tarde, estudos observacionais retrospectivos sugeriram ser destituído de benefícios e possivelmente danoso,⁽³⁾ achados confirmados por estudos prospectivos randomizados subsequentes.^(4,5) Enquanto estudos clínicos são melhores para inferir causalidade, não são apropriados para demonstrar efeitos de pequena monta, o que é típico da maioria das intervenções administradas em terapia intensiva a um grupo heterogêneo de pacientes. Além do mais, os estudos clínicos tipicamente excluem importantes subgrupos (pacientes mais idosos e indivíduos com comorbidades, por exemplo), de forma que os achados não são generalizáveis para o mundo real.

Em razão das limitações das pesquisas clínicas, inclusive seus custos, muitas diretrizes são apoiadas por evidência de baixa qualidade.⁽⁶⁾ Um levantamento dos boletins práticos do *American College of Obstetricians and Gynecologists* demonstrou que apenas 29% das recomendações tinham nível A, isto é, eram “baseadas em evidências com boa e consistente validade científica”,⁽⁷⁾ enquanto uma avaliação das diretrizes de prática clínica do *American College of Cardiology* e da *American Heart Association* identificou que apenas 314 das 2.711 recomendações (11%) se baseavam em evidência de alta qualidade.⁽⁸⁾

Para piorar, estas diretrizes são frequentemente adotadas em países subdesenvolvidos e em desenvolvimento, onde o financiamento para pesquisas é limitado, inclusive no Brasil.

A digitalização dos dados de atendimento à saúde pode proporcionar uma oportunidade para desenvolver diretrizes práticas localmente relevantes nos países subdesenvolvidos e em desenvolvimento, em vez de adotar as diretrizes de outros países. Os dados digitais proliferam em diferentes formatos no campo da saúde, não apenas em razão da adoção de prontuários eletrônicos, mas também do

Conflitos de interesse: Nenhum.

Submetido em 4 de fevereiro de 2016
Aceito em 11 de fevereiro de 2016

Autor correspondente:

Elliot Naidus
Beth Israel Deaconess Medical Center
330 Brookline Avenue, Boston, MA, USA, 02215
E-mail: enaidus@bidmc.harvard.edu

Editor responsável: Jorge Ibrain Figueira Salluh

DOI: 10.5935/0103-507X.20160008

crescimento das tecnologias sem fio para monitoramento ambulatorial. Como desenvolver pesquisas clínicas para informar diretrizes práticas pode ser muito dispendioso para países subdesenvolvidos ou em desenvolvimento, dados digitais de saúde oferecem oportunidade para a condução de pesquisa localmente relevante. A literatura médica tem demonstrado que estudos observacionais rigorosos apresentam boa correspondência com pesquisas clínicas, em termos de estimativa dos riscos e dimensionamento do efeito.⁽⁹⁻¹¹⁾

Big Data como solução

Conceitualmente, o termo “*Big Data*” inclui conjuntos de dados que são grandes a ponto de serem considerados não gerenciáveis para interpretação humana sem auxílio de um processo computadorizado e/ou analítico de dados. Enquanto é um desafio para as técnicas estatísticas tradicionais, em razão do nível de granularidade e resolução, a tecnologia *Big Data* demanda novas metodologias de inferência causal, para modelar exposições que variam com o tempo e outras covariáveis. Um dos usos de *Big Data* em medicina é a aplicação de técnicas de aprendizado da máquina para prever a probabilidade de eventos com base em correntes contínuas de dados. Por exemplo, o *Google* emprega um método automatizado para analisar a influenza em relação a buscas para traçar a movimentação da epidemia. Enquanto os dados do *Google* têm elevada correlação com a estatística de casos do *Center for Disease Control* (CDC), seu método tem uma vantagem, em termos de tempo exigido para uma análise em tempo real, assim demonstrando um mecanismo possivelmente melhor para prever e acompanhar epidemias.⁽¹²⁾ Durante o período mais intenso da epidemia de Ebola em Serra Leoa, a tecnologia móvel foi muito utilizada para coletar grandes quantidades de dados nas pequenas vilas. Ferramentas analíticas de dados em tempo real ajudaram nos esforços de quarentena, até levar à contenção da epidemia.⁽¹³⁾

A era do *Big Data* e a técnicas analíticas da próxima geração estão entre nós. Tanto grandes conjuntos de dados, como relevantes técnicas de aprendizado da máquina estão disponíveis há anos, apenas encontrando lentamente o caminho para os domínios da medicina clínica.

Big Data como problema

Tyler Vigen publicou um livro famoso que mostra correlações espúrias, relatando tendências disparatadas como a correlação da taxa de divórcio na região do Maine com o consumo *per capita* de margarina, ou o dispêndio dos Estados Unidos em ciências, espaço e tecnologia com a ocorrência de suicídios por enforcamento e estrangulamento.⁽¹⁴⁾ *Big Data*, analisados sem uma profunda compreensão do contexto, levam ao risco de produzir apenas “*big noise*”, isto é, muito barulho. Não é possível enfatizar em excesso a importância da validação cruzada dos achados, tanto interna quanto externamente, utilizando outros conjuntos de dados para confirmar a reprodutibilidade e avaliar a possibilidade de generalização. Tornar os conjuntos de dados acessíveis para investigadores externos e incentivar o ecossistema de pesquisa colaborativa podem, assim esperamos, ajudar a lidar com o quebra-cabeça representado pelas pesquisas não confiáveis.

CONCLUSÃO

A digitalização dos dados relativos à saúde vem se tornando um fenômeno global à medida que computadores, sensores e tecnologia sem fio passam a ser universais. Foi demonstrado que estudos observacionais produzem estimativas de efeito e risco que se correlacionam bem com pesquisas clínicas. Os assim chamados *Big Data* oferecem uma oportunidade aos países subdesenvolvidos e em desenvolvimento para construir sua própria base de conhecimento, a partir da qual podem desenvolver, avaliar continuamente e melhorar as diretrizes de prática clínica específicas para suas populações. Novas metodologias de inferência causal podem melhorar ainda mais o campo dos estudos observacionais. Para evitar as armadilhas de transformar nossos *Big Data* em “*big noise*”, é essencial transformar o processo de pesquisa, para que se torne mais aberto, autocrítico e colaborativo.

AGRADECIMENTOS

Leo Anthony Celi recebe apoio financeiro do *National Institute of Health* por meio da bolsa R01 EB017205-01A1.

REFERÊNCIAS

- Bornmann L, Mutz R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *J Assoc Inf Sci Technol*. 2015;66(11):2215-22.
- Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124.
- Connors AF Jr, Speroff T, Dawson NV, Thomas C, Harrell FE Jr, Wagner D, et al. The effectiveness of right heart catheterization in the initial care of critically ill patients. SUPPORT Investigators. *JAMA*. 1996;276(11):889-97.
- Sandham JD, Hull RD, Brant RF, Knox L, Pineo GF, Doig CJ, Laporta DP, Viner S, Passerini L, Devitt H, Kirby A, Jacka M; Canadian Critical Care ClinicalTrials Group. A randomized, controlled trial of the use of pulmonary-artery catheters in high-risk surgical patients. *N Eng J Med*. 2003;348(1):5-14.

5. Harvey S, Harrison DA, Singer M, Ashcroft J, Jones CM, Elbourne D, Brampton W, Williams D, Young D, Rowan K; PAC-Man study collaboration. Assessment of the clinical effectiveness of pulmonary artery catheters in management of patients in intensive care (PAC-Man): a randomised controlled trial. *Lancet*. 2005;366(9484):472-7.
6. Graham R, Mancher M, Solman DM, Greenfield S, Steinberg E, editors. *Clinical practice guidelines we can trust*. Washington, DC: National Academies Press; 2011.
7. Chauhan SP, Berghella V, Sanderson M, Magann EF, Morrison JC. American College of Obstetricians and Gynecologists practice bulletins: an overview. *Am J Obstet Gynecol*. 2006;194(6):1564-72; discussion 1072-5. Review.
8. Tricoci P, Allen JM, Kramer JM, Califf RM, Smith SC Jr. Scientific evidence underlying the ACC/AHA clinical practice guidelines. *JAMA*. 2009;301(8):831-41.
9. Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev*; 2014;4:MR000034.
10. Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA*. 2001;286(7):821-30.
11. Kitsios GD, Dahabreh IJ, Callahan S, Paulus JK, Campagna AC, Dargin JM. Can we trust observational studies using propensity scores in the critical care literature? A systematic comparison with randomized clinical trials. *Crit Care Med*. 2015;43(9):1870-9.
12. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature*. 2008;457(7232):1012-4.
13. Fallah MP, Nyenswah T, Dahn B, Thomas P, Harris TP, Freeman S. *Public Health Emergencies: Informatics in tracking the Ebola Virus Disease outbreak in Liberia*. In: *Global Health Informatics to Improve Quality of Care*. Cambridge: MIT Press; 2016.
14. Vigen T. *Spurious correlations*. New York: Hachette Books; 2015.