MESSAGE FROM THE PRESIDENT

Big data and medicine

Leon Rajah

FCS(SA)Orth
President: SA Orthopaedic Association

'We are on the verge of a digital revolution across every aspect of this sector, from the lab bench to the patient's bed side'.[1,2]

It is predicted that three emerging technologies will drive the next wave of medical innovation:

1. **Internet of things**: For example, wearable devices can track measures such as walking speed, balance and movement. Such real-time data provides a better understanding of disease progression and impact of treatment.
2. **Artificial intelligence (AI) and machine learning**: These technologies will revolutionise the way we interrogate data.
3. **Emerging data platforms**: These will allow unprecedented computing power and advances in data management systems for analysis and insight generation referred to as big data.

Big data refers to the analysis of massive amounts of data points to gain novel insight; and its key characteristics may be understood by considering data, method and purpose.[3]

Data

Big data is a massive shift in the ability to collect and analyse data quickly and cheaply. In the future we will routinely collect and analyse massive data sets from a larger number of individuals relevant to a phenomenon; and when possible, analyse all data collected, rather than just data samples.

With big data, volume may be traded off against quality. The 'unreasonable effectiveness of data' maintains that heterogeneous sources for data of limited quality may be better if one generates a huge amount of it, compared to only a small amount of data at high quality.[3] Using comprehensive data leads us to ask a further question: when do we stop collecting data and what do we do with 'new' data? Big data suggests tentativeness; learning is a summary of what is known of a dynamic phenomenon and necessitates re-evaluation at regular intervals.

Method

A big data approach requires the use of AI and its application to machine learning.[4-6] AI refers to the ability of a machine to perform cognitive functions usually associated with the human mind (perception, reasoning, learning and problem solving). Machine learning is the application of AI to massive data sets using complex self-learning algorithms to detect patterns, make predictions and generate hypotheses. The potential of AI is enormous: In 2017, Google's Alpha Zero Program Self-Learning AI chess programme taught itself chess with no human instruction, and after only 8 hours beat the then reigning 2016 World Computer Chess Champion Stockfish 8.

Purpose

There are two distinctive features of big data analysis. The first is the inductive nature of big data systems – analysis of a massive number of data points to identify patterns that prompt hypothesis generation; this in contrast to the conventional research method of using data to validate a human hypothesis. The second is that big data approaches are correlational. Big data does not demonstrate causality, is agnostic to cause and has been criticised for lacking causal explanatory value.

This is not a new argument: in 1847 the hygienist Semmelweist identified that hand washing with chlorine in maternity wards dramatically decreased mortality rates.[3] Most of his colleagues rejected the findings (he inferred an incorrect underlying cause) and resisted hand washing with chlorine, causing the unnecessary deaths of tens of thousands. Big data insights are going to raise similar issues in the future: What is sufficient evidence to act? How high is the burden of proof?

The approaches need not be exclusionary. In a recent study of Alzheimer's disease, millions of variables were measured following DNA and RNA sequencing in different brain regions. Conclusions were reached by allowing the data to speak to a likely driver of disease. The data analysis identified the immune system and microglial cells as a key driver of disease (as opposed to traditional concepts relating to tangles or plaques).[7] This raised possible novel therapies, which may be evaluated using hypothesis-testing in prospective randomised controlled trials. Big data may herald a change to a more staged discovery process – with correlational results and ensuing causal inquest.

The usefulness of the big data approach in health care remains disputed, however – does it provide a future with novel insights or does it create more noise that drowns out true signals? Jacofsky refers to these as: a lack of data set reliability and clarity; a preponderance of unstructured data; ineffective and inaccurate measures transposed to manage the behaviour of providers and income from payer claims or coders; a lack of intersystem reliability and inconsistent value of output from a system (analogous to a calculator providing a different answer to the same calculation).[8] Big data can impose the same challenges as small data; and adding more data without physicians to control and standardise definitions will most often not solve but merely magnify the problem.

Conclusion

Big data enthusiasts propose that medicine has changed to an information science.[9] Popular literature declares the physical examination of a patient redundant.[10] In our prime directive – *'the only interest to be served is the interest of the patient'* – is embedded that human spirit to defend the integrity of clinical practice, thought and innovation; and posit clinical medicine as integral to a defence against a future dominated by digital dictatorship, financial oligarchy and human redundancy.[11,12]

References

1 *3 things that will change medicine in 2018*. Vasant Narasimhan, CEO, Global Head, Drug Development; Novartis AG; World Economic Forum Annual Meeting on Future of Health and Healthcare; 24 Jan 2018.

2 *The Economist* editorial: A revolution in health care is coming Welcome to Doctor You; 1st February 2018.

3 Mayer-Schonberger & Ingelsson. Big Data and medicine: a big deal? *Journal of Internal Medicine*, 2018;**283**:418-29.

4 Hawking S. *Brief Answers to the Big Questions*. John Murray Publishers, Great Britain, 2018.

5 *An executive's guide to AI*. McKinsey Analytics, McKinsey & Company, 2018.

6 Domingos P. Our digital double: AI will serve our species, not control it. *Scientific America* special issue: The Science of being Human; September 2018.

7 *The role of big data in medicine*: Interview with Dr Eric Schadt of the Icahn Institute for Genomics and Multiscale Biology at New York's Mount Sinai Health System by Sastry Chilukuri for McKinsey and Company; November 2018.

8 Jacofsky DJ. Instructional Review: The myths of 'big data' in health care. *Bone Joint J* 2017;99-B:1571-76.

9 Krumholz HM. Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. *Health Aff* (Millwood). 2014 July;**33**(7):1163-70. doi:10.1377/hlthaff.2014.0053

10 Kraft D. 'Connected' and high-tech: your medical future. *National Geographic* special issue: The Future of Medicine; January 2019.

11 Varoufakis Y. *Talking to my daughter about the economy a brief history of capitalism*. Penguin Random House Publishers; United Kingdom 2017.

12 Harari YN. *21 Lessons for the 21st Century*. Penguin Random House Publishers, United Kingdom 2018.