Journal of Gastroenterology and Hepatology
JGHF
doi:10.1111/jgh.14606

EDITORIAL

The emerging role of big data in gastroenterology and hepatology

Clinical medicine as well as basic biomedical research are huge generators of biomedical data. In the United States, the healthcare system is expected to reach the zettabyte (10^{21}) scale from electronic health records, scientific instruments, clinical decision support systems, or even research articles in medical journals.[1] In Hong Kong, the Hospital Authority, the sole public healthcare provider for a population of over seven millions, has thousands of types of clinical data (e.g. diagnosis, procedures, and laboratory tests and results) on line.[2] These data will provide unprecedented opportunities for biomedical research and even guidance to clinical practice. New sources of health-related data such as social media and networks, wearable device equipped with biosensing technology recording real-time digital data will add further to this ocean of health-related data.[3] How could we use these data properly and efficiently to the benefit of the patient and the healthcare system? At the end of the day, who is going to benefit from this new capability? Who might be the loser?

There are at least six main use cases of big data in healthcare systems—namely, (i) efficient use of expensive treatments in patients care (to ensure cost-effective usage of expensive drugs, equipment, and services), (ii) reduce hospital admissions and readmission (to monitor specific parameters of patients at home or after discharge), (iii) patient triage (to direct patients of different risk levels to necessary treatments), (iv) monitoring functional decompensation of patients (to keep surveillance of multiple physiological parameters with wearable devices and telemetry), (v) optimization of treatments (in drug dosing, device adjustment, etc.), and[4] (vi) Monitoring adverse events related to treatments (to predict occurrence of unwanted clinical outcomes using genetic/genomic and clinical parameters).

One of the important goals of these use cases is to reduce cost and at the same time improve patient outcomes.[5] In the field of gastroenterology and hepatology, big data has already been used to facilitate the research on less common but important clinical events related to therapy, for example, cancers risk[6] and fracture risk[7] in patients receiving nucleos(t)ide analogues; clinical outcomes including progression to cirrhosis and liver cancer after hepatitis B surface antigen seroclearance[8,9] in patients with chronic hepatitis B; effectiveness and risk of chemoprevention such as using low-dose aspirin[10–12], just to name a few. All these studies require clinical data from huge cohorts of subjects with long study period.

One of the major challenges of performing analytics of big data in clinical medicine would be patient data security as health data is very personal and untowards leakage of such information can cause serious consequences. Therefore, healthcare providers and health authorities of different jurisdictions have always been very cautious in opening up such data sources. The famous National Health Service (NHS) story does have a few important lessons for everyone. In 2015, the Royal Free NHS Foundation Trust in the United Kingdom (UK) gave 1.6 million patient records to Google DeepMind to create a healthcare app for the diagnosis and alert of acute kidney injury. The data transfer between NHS & Google was found not properly complying the Data Protection Act by informing patients the use of their data which had included sensitive patient data (e.g. HIV status, mental health history, and abortions). This was subsequently ruled illegal and reported in the news that "Royal Free breached UK data law in 1.6m patient deal with Google's DeepMind."[13] After that Google DeepMind made an open statement, "We underestimated the complexity of the NHS and of the rules around patient data, as well as the potential fears about a well-known tech company working in health."[14] This is, if anything, only the tip of an iceberg and could be just be the end-of-the-beginning of a huge debate and controversy. A similar scenario can happen when patients blood chemistry are being monitored for their liver enzymes and hepatitis marker to predict acute liver failure, or using aspirin or other prescriptions that may trigger sensitive issues (e.g. drug-related side effects).

There are also other aspects warrant special attention and deliberation when we employ big data in medical research and clinical decision making. This includes legal compliance, accuracy, intellectual property, cyber security, and transparency, just to name a few.[15] This list is not exhaustive. In addition, there may be sector specific risks of which disciplines need to be aware of. The World Health Organization urges that future policies must take into account the distinct challenges posed by big data on top of the potential benefits.[16]

Big-data approaches have resulted in many new opportunities both in research and clinical care. Clinical data are now almost ubiquitously available, yet analytics is the main point where everything convey. The future of big data will involve linking multiple data sources—social media, mobile devices, cloud computing, etc. Who will be the winners and losers in this big wave (Fig. 1)?

Patients will win as they will have more choices and (hopefully) more efficient and timely care offered through big data analytics. On the other hand, they may have to pay higher premium in health insurance for more options, and before that, users may have to consent for their health-related and lifestyle data to share this program. Healthcare providers will win as the efficiencies of healthcare delivery may improve and cost (hopefully) be reduced. But, let us face it, some mundane job might be lost, replaced by smart machines and artificial intelligence. Healthcare authority and insurance company may win, because they can deploy their resources in a more cost-effective way. New policies can be developed to benefit more patients and/or insurance clients. However, healthcare authorities must face an increasing demand from unhealthy as well as healthy people. As predictive algorithms find their way into future health care, there will be tough decisions on when to intervene (before it is too late). Finally, researchers are most likely to be benefit from this new era of conducting research. Exciting findings and publications will be generated. At the same time, research workers must learn how to distinguish "noise" from

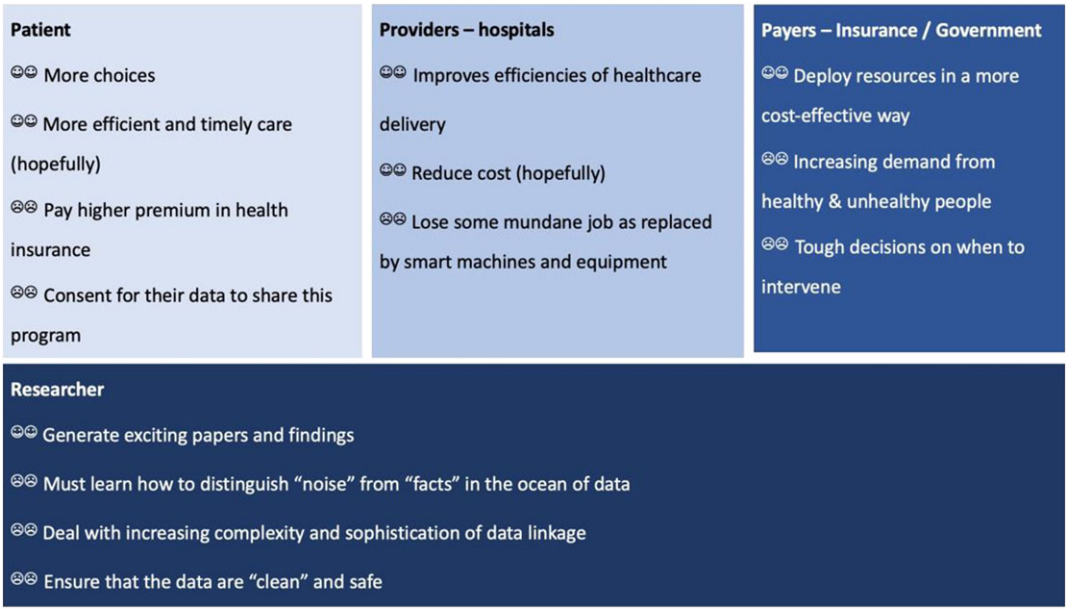

Figure 1 Winners and losers of big data in medicine. [Color figure can be viewed at wileyonlinelibrary.com]

"facts" in the ocean of data. They have to ensure that the data are "clean" and safe. They have to learn to deal with increasing complexity and sophistication of data linkage. All in all, we are entering a new age in gastroenterology and hepatology.

The Editors
Grace Lai-Hung Wong*,† and Joseph Jao-Yiu Sung*,†

**Institute of Digestive Disease, †Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong SAR, China*

References

1 Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf. Sci. Syst.* 2014; **2**: 3.
2 Webpage: http://www3.ha.org.hk/data/DCL/ProjectDataCatalogue/. Accessed on 25 December 2018.
3 Brodie MA, Pliner EM, Ho A *et al*. Big data vs accurate data in health research: large-scale physical activity monitoring, smartphones, wearable devices and risk of unconscious bias. *Med. Hypotheses* 2018; **119**: 32–6.
4 Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff (Millwood).* 2014; **33**: 1123–31.
5 Rumsfeld JS, Joynt KE, Maddox TM. Big data analytics to improve cardiovascular care: promise and challenges. *Nat. Rev. Cardiol.* 2016; **13**: 350–9.
6 Wong GL, Tse YK, Wong VW, Yip TC, Tsoi KK, Chan HL. Long-term safety of oral nucleos(t)ide analogs for patients with chronic hepatitis B: a cohort study of 53,500 subjects. *Hepatology* 2015; **62**: 684–93.
7 Wong GL, Tse YK, Yip TC, Chan HL, Tsoi KK, Wong VW. Long-term use of oral nucleos(t)ide analogues for chronic hepatitis B does not increase cancer risk—a cohort study of 44 494 subjects. *Aliment. Pharmacol. Ther.* 2017; **45**: 1213–24.
8 Yip TC, Chan HL, Wong VW, Tse YK, Lam KL, Wong GL. Impact of age and gender on risk of hepatocellular carcinoma after hepatitis B surface antigen seroclearance. *J. Hepatol.* 2017; **67**: 902–8.
9 Yip TC, Wong GL, Chan HL *et al*. HBsAg seroclearance further reduces hepatocellular carcinoma risk after complete viral suppression with nucleos(t)ide analogues. *J. Hepatol.* 2018 [Epub ahead of print].
10 Tsoi KK, Chan FC, Hirai HW, Sung JJ. Risk of gastrointestinal bleeding and benefit from colorectal cancer reduction from long-term use of low-dose aspirin: a retrospective study of 612 509 patients. *J. Gastroenterol. Hepatol.* 2018; **33**: 1728–36.
11 Tsoi KK, Ho JM, Chan FC, Sung JJ. Long-term use of low-dose aspirin for cancer prevention: a 10-year population cohort study in Hong Kong. *Int. J. Cancer* 2018 [Epub ahead of print].
12 Wu CY, Chen YJ, Ho HJ *et al*. Association between nucleoside analogues and risk of hepatitis B virus-related hepatocellular carcinoma recurrence following liver resection. *JAMA* 2012; **308**: 1906–14.
13 Webpage: https://www.theguardian.com/technology/2017/jul/03/google-deepmind-16m-patient-royal-free-deal-data-protection-act. Accessed on 25 December 2018.
14 Webpage: https://www.cbronline.com/ehealth/google-deepmind-nhs-medical-trial-got-wrong-need-better/. Accessed on 25 December 2018.
15 Baird P. Going big data, with caution. *Biomed. Instrum. Technol.* 2018; **52**: 39–40.
16 Vayena E, Dzenowagis J, Brownstein JS, Sheikh A. Policy implications of big data in the health sector. *Bull. World Health Organ.* 2018; **96**: 1–76.