Views & Comments

Engineering 2015, 1(3): 277–279
DOI 10.15302/J-ENG-2015075

Big Data for Precision Medicine

By Daniel Richard Leff and Guang-Zhong Yang*

ABSTRACT This article focuses on the potential impact of big data analysis to improve health, prevent and detect disease at an earlier stage, and personalize interventions. The role that big data analytics may have in interrogating the patient electronic health record toward improved clinical decision support is discussed. We examine developments in pharmacogenetics that have increased our appreciation of the reasons why patients respond differently to chemotherapy. We also assess the expansion of online health communications and the way in which this data may be capitalized on in order to detect public health threats and control or contain epidemics. Finally, we describe how a new generation of wearable and implantable body sensors may improve wellbeing, streamline management of chronic diseases, and improve the quality of surgical implants.

KEYWORDS big data, biosensors, body-sensing networks, implantable sensors, clinical decision support systems, pharmacogenetics, mHealth

1 Introduction

The complexity, diversity, and rich context of data being generated in healthcare are driving the development of big data for health [1]. Volume, velocity, variety, veracity, variability, and value are the "V's" of big data, and these are encapsulated in the inherent challenges of biomedical and health informatics. Effective ways of tackling these challenges would pave the way for more intelligent healthcare systems focused on prevention, early detection, and personalized treatments.

2 Big data for precision medicine

The electronic patient health record (EHR) is a source of big data containing information regarding socio-demographics, medical conditions, genetics, and treatments; yet the human ability to process this data without effective decision support is finite. For healthcare, the goal is to provide a continually learning infrastructure with real-time knowledge production and to develop a system that is preventative, predictive, and participatory [2]. In order to achieve this goal, computer models are required to help clinicians organize the data, recognize patterns, interpret results, and set thresholds for actions. Examples of big data analytics for new knowledge generation, improved clinical care, and streamlined public health surveillance are already apparent. For example, the EHR has been successfully mined for post-market surveillance of medications and improved pharmacovigilance.

In the United Kingdom, the National Health Service intends to be paperless by 2018. The EHR will provide an integral resource for future clinical decision support systems (CDSSs) that may overcome human limitations in data comprehension and multitasking. CDSSs are already being developed to assess and improve protocol adherence [3], for medication reminders [4], to improve screening [5], and to predict hospital readmission [6]. Certain conditions have multiple treatment options and CDSSs are being developed to help clinicians optimize strategies. For example, problems associated with anterior cruciate ligament (ACL) may be treated by physiotherapy, medicine, or surgery. In this regard, hierarchical learning for EHR data, along with information regarding occupational, recreational, and musculoskeletal data, has been used to classify treatment options [7].

It is expected that big data for health can play an important role in pharmacogenetics and stratified healthcare. Patients with a similar cancer subtype often respond differently when challenged with the same chemotherapeutics. For example, *CYP2D6* is a polymorphic gene associated with response to Tamoxifen [8], *BRAF* mutations (*Y472C*) have been linked to Dasatinib response in non-small cell cancer of the lung [9], and multiple gene signatures have been recently associated with the response of rectal cancer to chemoradiotherapy [10]. Genomic instability is believed to be responsible for the observed diversity of drug response. Recent efforts have focused on exposing the complex interplay of genomics and chemotherapeutic sensitivity, resistance, and toxicity [11–13]. For example, the Cancer Genome Atlas research network has launched the Pan-Cancer project [11] to analyze multiple tumor types and molecular aberrations in cancer types, and to enable scientists to discover new aberrations. Similarly, several projects such as the Cancer Cell Line Encyclopedia [12] and the Genomics of Drug Sensitivity in Cancer [13] are generating large genomic databases to specifically interrogate links between genomic biomarkers and drug sensitivity in hundreds of cancer cell lines. As evidence of the ability to leverage large pharmacogenetic databases to predict drug sensitivity, recent data suggests that computational algorithms for predicting drugs for individual cell lines

The Hamlyn Centre, South Kensington Campus, Imperial College London, London SW7 2AZ, UK
*Correspondence author. E-mail: g.z.yang@imperial.ac.uk

can be improved based on genomic profiles and drug-response data [14]. Future work will involve testing the results of such algorithmic predictions on tumor response and toxicity in patients undergoing chemotherapy.

One of the major contributors of dramatic change in healthcare is the upsurge in communication instigated by social media. One recent estimate suggests over one billion healthcare tweets, reflecting the enormity of dialogue between healthcare providers, patients, organizations, and third parties. Approximately 20% of patients with chronic healthcare conditions such as diabetes, cardiovascular disease, and cancer, go online to actively seek others and share experiences of related conditions; such actions are creating patient communities on social media sites such as Twitter and Facebook. Social media is providing new avenues for investigators to enroll patients in research, and for patients to engage in sharing their health data. For example, in TuAnalyze [15], a joint initiative between TuDiabetes and the Boston Children's Hospital, diabetics can monitor, evaluate, and share their results while actively participating in research on diabetes. Arguably one of the most interesting applications of big data analytics is in the ability to predict and track major outbreaks in order to improve public healthcare resources and the dissemination of healthcare messages to victims using social media. Predictions of serious healthcare emergencies such as exacerbations of asthma can be better predicted in models that combine social media analysis with environmental data. Unlike conventional models that base predictions with a two-week lag, Ram et al. [16] were able to create accurate prediction models of the volume of daily emergency-department visits for acute asthma (volume defined as low, moderate, or high) using Twitter activity, Google searches, and air-quality data [16]. Similarly, for major public epidemics such as Ebola, data can be collected and analyzed to support early warning systems for epidemic trends and to deliver messages for healthcare education interventions [17]. For example, Odlum et al. [17] demonstrated that analyzing tweet activity around Ebola virus detection (EVD) captured progressive increases in the number of tweets discussing EVD case identification in Nigeria occurring at least three days prior to the news alert and seven days before the official Centre for Disease Control warnings. Finally, several investigators are exploiting the potential of social media toward behavioral change and improvements in healthcare, including targeted interventions for developing countries for the purpose of global health promotion [18].

Another important driver for big data is the widespread deployment of sensing technologies. There has been increasing interest in wearable and implantable sensing owing to improvements in technologies, including sensors with enhanced wireless communications that have increased bandwidth and improved microelectronics. As a result of these developments, continuous, multimodal, and context-aware sensing is now feasible [19]. Simultaneously, advances in sensor miniaturization, embodiment, biofouling mitigation, and microelectronic fabrication schemes have improved the versatility and reliability of implantable biosensors. Episodic monitoring is being replaced, therefore, with continuous sensing and parallel improvements in integrated care, toward personalized and stratified healthcare. Patients at high risk of critical events, such as arrhythmias, following myocardial infarction may benefit from continuous monitoring of blood pressure, pulse, and cardiac rhythm such that arrhythmias can be detected in near real time and signals sent to a smartphone for ulterior processing [20]. Similarly, patients with chronic health conditions such as diabetes may benefit from implantable microsystems for continuous monitoring [21], toward improved long-term glucose control. Neuromodulation, such as with implantable spinal stimulators, is having an impact in the management of chronic back pain [22]. Implantable sensors that monitor the axial load on an individual subject's spine [23] may lead to personalized orthopedic prostheses. In future, it is hoped that sensors applied in close proximity to the site of operative intervention, which monitor local hemodynamics or tissue ionic content, will enable wound infections to be detected and treated at a sub-clinical stage, preventing fulminant sepsis. White blood cells and neutrophil counts could be continually monitored in patients undergoing chemotherapy cycles such that the earliest sign of neutropenia could be coupled with granulocyte stimulation to mitigate sepsis.

3 Conclusions

In conclusion, the sources and computational techniques for big data are rapidly increasing and the possible uses for improving health and well-being are manifold. For healthcare, the goal is to provide a continually learning infrastructure with real-time knowledge production and to develop a system that is preventative, predictive, and participatory. Big data analysis clearly has tremendous potential to improve healthcare and transform the health of populations. However, successfully exploiting this potential will depend on solving challenges associated with data privacy, security, ownership, and governance.

Compliance with ethics guidelines

Daniel Richard Leff and Guang-Zhong Yang declare that they have no conflict of interest or financial conflicts to disclose.

References

1. J. Andreu-Perez, C. C. Poon, R. D. Merrifield, S. T. Wong, G. Z. Yang. Big data for health. *IEEE J. Biomed. Health Inform.*, 2015, 19(4): 1193–1208
2. L. Hood, M. Flores. A personal view on systems medicine and the emergence of proactive P4 medicine: Predictive, preventive, personalized and participatory. *New Biotechnol.*, 2012, 29(6): 613–624
3. J. G. Klann, V. Anand, S. M. Downs. Patient-tailored prioritization for a pediatric care decision support system through machine learning. *J. Am. Med. Inform. Assoc.*, 2013, 20(e2): e267–e274
4. B. G. Nair, S. F. Newman, G. N. Peterson, W. Y. Wu, H. A. Schwid. Feedback mechanisms including real-time electronic alerts to achieve near 100% timely prophylactic antibiotic administration in surgical cases. *Anesth. Analg.*, 2010, 111(5): 1293–1300

5. K. B. Wagholikar, et al. Clinical decision support with automated text processing for cervical cancer screening. *J. Am. Med. Inform. Assoc.*, 2012, 19(5): 833–839
6. J. Futoma, J. Morris, J. Lucas. A comparison of models for predicting early hospital readmissions. *J. Biomed. Inform.*, 2015, 56: 229–238
7. K. Mei, J. Peng, L. Gao, N. N. Zheng, J. Fan. Hierarchical classification of large-scale patient records for automatic treatment stratification. *IEEE J. Biomed. Health Inform.*, 2015, 19(4): 1234–1245
8. M. P. Goetz, et al. The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. *Breast Cancer Res. Tr.*, 2007, 101(1): 113–121
9. B. Sen, et al. Kinase-impaired BRAF mutations in lung cancer confer sensitivity to dasatinib. *Sci. Transl. Med.*, 2012, 4(136): 136ra70
10. T. Watanabe, T. Kobunai, T. Akiyoshi, K. Matsuda, S. Ishihara, K. Nozawa. Prediction of response to preoperative chemoradiotherapy in rectal cancer by using reverse transcriptase polymerase chain reaction analysis of four genes. *Dis. Colon Rectum*, 2014, 57(1): 23–31
11. Cancer Genome Atlas Research Network; J. N. Weinstein, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.*, 2013, 45(10): 1113–1120
12. J. Barretina, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 2012, 483(7391): 603–607
13. M. J. Garnett, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature*, 2012, 483(7391): 570–575
14. J. Sheng, F. Li, S. T. Wong. Optimal drug prediction from personal genomics profiles. *IEEE J. Biomed. Health Inform.*, 2015, 19(4): 1264–1270
15. Anon. TuAnalyze is here! 2010-05-19. http://www.tudiabetes.org/forum/topics/tuanalyze-is-here
16. S. Ram, W. Zhang, M. Williams, Y. Pengetnze. Predicting asthma-related emergency department visits using big data. *IEEE J. Biomed. Health Inform.*, 2015, 19(4): 1216–1223
17. M. Odlum, S. Yoon. What can we learn about the Ebola outbreak from tweets? *Am. J. Infect. Control*, 2015, 43(6): 563–571
18. S. Bahkali, N. Alkharjy, M. Alowairdy, M. Househ, O. Da'ar, K. Alsurimi. A social media campaign to promote breastfeeding among Saudi women: A web-based survey study. *Stud. Health Technol. Inform.*, 2015, 213: 247–250
19. G. Z. Yang. *Body Sensor Networks*. 2nd ed. London: Springer-Verlag, 2014
20. F. Rincón, P. R. Grassi, N. Khaled, D. Atienza, D. Sciuto. Automated real-time atrial fibrillation detection on a wearable wireless sensor platform. In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Piscataway, NJ: IEEE Service Center, 2012: 2472–2745
21. M. M. Ahmadi, G. A. Jullien. A wireless-implantable microsystem for continuous blood glucose monitoring. *IEEE Trans. Biomed. Circuits Syst.*, 2009, 3(3): 169–180
22. J. Padwal, M. M. Georgy, B. A. Georgy. Spinal cord stimulators in an outpatient interventional neuroradiology practice. *J. Neurointerv. Surg.*, 2014, 6(9): 708–711
23. M. K. Moore, S. Fulop, M. Tabib-Azar, D. J. Hart. Piezoresistive pressure sensors in the measurement of intervertebral disc hydrostatic pressure. *Spine J.*, 2009, 9(12): 1030–1034