View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Apollo

Big Data for Healthcare in China: a Review of the State-of-art

Shanshan ZHANG

Peking University School of Stomatology, P.R. China

Abstract:

This paper aimed to review the status of implementing big data analysis of healthcare in China. Big data refers to very large datasets with complex structures, with high volume of data mass, high velocity of data flow and high variety of data types. Health related big data is created either through traditional and mobile internet, smartphones and social media or in medical settings from hospital or clinical information systems. Big data in healthcare can be used for individual disease diagnoses, disease prognoses, disease prevention and prediction, also can be used to design a tailored health and disease management program based on their lifestyle. A literature review was conducted to identify studies on big data in healthcare in China from 2000 to 2015. We used different combinations of keywords to search Medline and EMBASE via Ovid platform, CNKI and Wanfang database for literatures in English and Chinese, complemented with searches in Google Scholar and Baiduxueshu for relevant articles. This paper reviewed studies describing the application of big data in healthcare in China and discussed issues related to storage and further use of big data.

Key Words: Healthcare, Big data, Medicine

1. INTRODUCTION

Big data refers to very large datasets with complex structures, with high volume of data mass, high velocity of data flow and high variety of data types. Health related big data is collected from multiple sources (Wyber et al., 2015). In China, Internet plus healthcare is currently a hot topic from governmental policy-making to individual entrepreneurship. Despite a few matured programmes from big companies such as IBM and Intel, most of the digital construction for healthcare big data are still undergoing in institutions and industry. Challenges remains in this fast growing area, human resources in medical informatics and the data security were the major concerns that need years of cultivation and exploration (Chen and Xu, 2015). This paper aimed to review the status of implementing big data analysis of healthcare in China and to further discuss the challenges.

2. METHODS

A literature review was conducted to identify studies on big data in healthcare in China from 2000 to 2015. We used different combinations of keywords to search Medline and EMBASE via Ovid platform, CNKI and Wanfang database for literatures in English and Chinese, complemented with searches in Google Scholar and Baiduxueshu for relevant articles. Grey literatures for relevant articles and references of selected articles were also checked. The application, benefits, risks and future opportunities of big data in health care in China was summarized. Recommendations were made for the use of big data in the delivery of healthcare services in China.

3. DEFINITION

Big data refers to very large datasets with complex structures, with high volume of data mass, high velocity of data flow and high variety of data types. Healthcare-related big data is collected from multiple sources, including data created in medical settings from hospital or clinical information systems (HIS or CIS), such as electronic medical record (including text and video/audio files), radiographic images, surveillance and medical device streaming data; data from clinic and hospital charging system; data collected through traditional and mobile internet, smartphones and social media; and data from pharmaceutical company and bio science, human genome and drug discovery. At population level, health data also included information from vital statistics registries (Wang and Krishnan, 2014).

4. APPLICATIONS

Big data in healthcare can be used for individual disease diagnoses, disease prognoses, disease prevention and prediction, as well as design a tailored health and disease management program based on their lifestyle. The application of big data in healthcare consists the 3 major subdomains: human genome project for precise medicine; clinical information analysis for clinical decision making and hospital charging system for health economic evaluation; and mobile data for public health monitor (Cai et al., 2013).

4.1 Human genome project for precise medicine

The most famous example in clinical science of big data application is personalized medical care. In January 2015, US President Obama promoted 'Precision Medicine Initiative' project, aimed to mapping human genome of 1 million US citizens, find precise defect in the primary cause of a disease at genetic level for a group of people, and ultimately guide to develop new generation of drug that precisely to a subset of molecular problems shared by patient groups with a given disease. Precise medicine has evolved from personalized medicine in that our focus in disease treatment and prevention of each individual. Precision medicine initiative will enlarge US national bio-bank database to provide a mega-database for future medical research (Kroll, 2015).

Experts from China also called for changes from evidence-based medicine to precision medicine, which means to create a large gene database, enable researchers to analyze patient's information for better understanding the cause and etiology of disease, and to develop target drugs for particular gene defects. China National Genebank (CNG) was constructed by BGI Shenzheng from 2011, it was granted by National Development and Reform Commission, Ministry of Finance, Ministry of Industry and Information Technology, and National Health and Family Planning Commission. CNG comprises biological bank (bioresources and biobank), bioinformatics bank (database and data center cloud) and network/platform consortium, able to conduct bio-resource specimen collecting, omics data acquiring, omic data analyzing, industrialization and application. In healthcare, CNG covers digital health, molecular diagnoses and precise medicine. CNG provides storage for genetic information, bio-information index and cloud computing (Yang and Sun, 2013). While the preparation work for precise medicine started, it still needs years of research to further transfer the benefit to real population.

4.2 Clinical decision support system

Clinically, large amount of data were being collected everyday. Electronic medical record, radiography record from PACS and laboratory tests results were generated everyday for patients who visited hospitals and clinics. Clinical decision support system using these data can improve the precision of clinical diagnose and treatment planning, ultimately to improve medical staff's efficiency and patients' health(Yu et al., 2014).

Traditional evidence based medicine has one vital inherent flaw – purely rely on the ability of human brain. Traditional evidence based medicine requires physicians and surgeons to collect patient's clinical manifestation and use examination results to make the most likely disease diagnoses based on their own medical knowledge and clinical experience. Although doctors had years of professional medical trainings, it is nearly impossible for one to memorize all medical knowledge for disease diagnoses and treatment, let alone to keep their knowledge updated at all time. For this limitation of human brain, medical professionals learned to separate medical knowledge into many disciplines, so-called the secondary or third grade discipline of clinical medicine, and choose to focus on enriching their knowledge and skills into a narrower but more detailed sub-domain of clinical medicine, such as gynecology, cardiology, and dermatology etc. This approach was naturally developed and it helped to improve the accuracy of diagnosis and treatment planning. However, some argues that for those uncommon diseases, even an expert may lack of relevant clinical experience, then lead to a false diagnosis or treatment difficulty. Another problem is that a patient should be seen as an integrated whole person, not a part of human body (You, 2015).

The advance technology of computer sciences makes it possible to derive more accurate medical diagnosis and treatment planning with collected clinical information and updated medical knowledge. Clinical decision support system can learn mass literature and correct itself constantly, to provide the most appropriate diagnosis and the best treatment plan. A system such as IBM Watson for healthcare was initially developed for triage, with years of development, IBM Watson has established a medical literature and expert database, it can incorporate clinical, pathological and genetic characteristics to propose standardized clinical pathways and individualized treatment recommendation. Some feature of IBM healthcare solutions has been introduced to many hospitals in China and are benefiting Chinese population (e-healthcare, 2014). It cannot only improve the efficiency and the quality of care; it can also reduce adverse reactions and treatment errors. A clinical research in US children hospital showed that the clinical decision support system could avoid 40% of adverse drug events. NICE from the United Kingdom has already conducted cost-effective research and showed the cost-effectiveness advantage of the IBM Watson decision support system. Technically, the clinical diagnosis and treatment recommendation can be provided by clinical decision support system with better success rate and lower cost compared to human physicians, because the amount of information computer can processed and the learning nature of medical practice (Song and Ryu, 2015).

The first step of building clinical decision support system is to transform traditional knowledge and records into a digitally useable format. In China, the state has adopted a number of policies to accelerate the process of medical informatization. In the national ‘Twelfth Five-Year plan’, the government emphasized the need to establish a nationally unified electronic health records and electronic medical records, to regulate the registration for drug, equipment, services and insurance information, strengthen the regional information platform construction, and promote health information resources sharing (TheStateCouncil, 2012) .

Big data also plays an important role in medical and clinical research. Major research institute centers and funding agencies have made large investments in this area. Clinical big data can be used to determine causality, effect or association between risk factors and disease of interest.

The advantage of medical big data compared to large datasets from other disciplines is that clinical big data are often collected based on protocols and relatively structured. Vast amounts of data can be collected through clinical study and the analysis can inform our understanding of diseases (Yoo et al., 2014).

In addition to clinical information system, hospital management information system also contain cost, length of stay and other management information, monitoring these information can help to control insurance fraud, monitor disease pattern and seeking healthcare behavior. For example, Su et al. conducted a research based on China Health and Nutrition Survey 2009 to evaluate the influence of medical insurance of residents' medical consumption and health conditions (Su et al., 2013). This information can help to design national health insurance level, to ultimately achieve affordable healthcare for all and improve the efficiency of medical resource allocation.

4.3 Epidemiology of individual lifestyle and behavior pattern

Digital epidemiology, also referred to as digital disease detection, is motivated by the same objectives as traditional epidemiology, but digital disease detection focuses on electronic data sources that merged with the advent of information technology. It collects the vast amount of health related information generated by Internet and mobile devices and utilizes global real-time data for public health surveillance, monitoring and control. It can be use for early detection of disease outbreaks, and assessment of health behavior and attitudes and facilitate hospital aftercare. This fast-growing field has changed the ways in which epidemiologic information is obtained, analyzed and disseminated, which is likely to result in great social benefits. The most recent sample worldwide is the Ebola virus outbreaks in West Africa. Reports of emerging outbreak were detected by digital surveillance channels in advance of official reports (Vayena et al., 2015).

The Chinese Center of Disease Control and Prevention has collaborated with Baidu to carry out big data analysis collaboration and will use Baidu data and engine technology to construct the first influenza forecast system in China. This will benefit billions of people by early detect and control of communicable disease.

Health cloud platform construction in China focuses on the prevention of non-communicable chronic disease and monitoring the health status of population. Many devices and applications installed in smartphones now can collect personal health information such as pulse, blood pressure and blood sugar level, enable the real-time monitoring for essential physical status and the recommendation on the healthy diet, exercise and medication (Teng et al., 2014).

5. CHALLENGES

Big data in healthcare has not only changed the way we process data, but also the attitude toward healthcare information, the logic behind the decision-making and the organization structure of healthcare services. Firstly, we do not rely on only a small amount of sample data to reveal causality, neither attached to the accuracy of the structured data. Secondly, the drivers of clinical and public health decision-making come from data analysis based on comprehensive medical knowledge structures rather than from the intuition. This ensures the effectiveness and efficiency of clinical and public health decision-making. Thirdly, the core-competitiveness of future healthcare organizations has changed to how they hold, process and explore insights from big data (Chen and Xu, 2015). Traditional healthcare organization is facing the challenges into decentralized, fragmented healthcare service deliver bodies, which equipped with powerful healthcare databases and the ability to translate the value of healthcare big data into practical guidance towards individual health benefit.

Although the big data construction has already started, challenges remains to the full-scale implementation of big data in China. We will discuss the challenges in the following 5 aspects: Data volume and quality, type of data, value of data, data storage and process, ethical issues and practical challenges.

5.1 Data volume and quality

Medical big data can be obtained from many aspects, genetic data, cost data, radiographic information, laboratory test results, and from social media or mobile software. In the next 10 years, data volume will increase to 35 zettabyes by 2020. Large volume of data will bring the problem of data storage and data retrieval. Another challenge is how to handling missing data of large datasets and selection bias, which are important for data analysis and interpretation. With the large data volume and variety of data quality, one must consider all the aspects of dataset, including collection, curation, extraction, integration, interpretation, imputation and selection of appropriate statistical methods, during processing and analysis (Chen and Xu, 2015).

5.2 Type of data

Data analysis requires structured data, however, more than 80% of existing data are unstructured and most newly collected data remain unstructured. Electronic medical record and genetic information can be structured if planned ahead, whereas patient oral communication, hand-write medical record were still unstructured, as well as most of the personal health behavior data generated from social media and mobile devices. How to organize these data into a useful format requires policy-makers foresight and medical staffs' cooperation (Yu et al., 2014).

5.3 Value of data

Big data facilitates health-system focused approaches, rather than disease-specific programmes. The big data approach fits horizontal programme better than a vertical programme and could potentially improve the control and treatment of all human disease. Therefore the value of big data exists in the horizontal programme. However, most of the current big data programme are still driven by disease specific interests(Gao and Sang, 2013).

5.4 The speed of real data analysis

Big data emphasizes the real time effect of data analysis, it requires the system process the data at all time real time and provide search results at any time. This may burden the existing healthcare system. Big data approach can amplify the existing difficulties associated with health-care delivery in settings with scarce resources. It may be impossible for frontline health workers to extend their work to the non-essential collection of data. This dilemma requires policy-makers to ease the ways of data collection, not view this approach as a distraction for care delivery (Gao et al., 2013).

5.5 Ethical challenges

The collection of information from individuals is fraught with ethical, regulatory and technological issues. It is difficult to differentiate commercial versus public health uses of data. Given the complexity of the field, the protection of individuals and populations must move to emphasize appropriate use, risk assessment and risk minimization. The anonymization of data must be robust, monitored and enforced (Zhou and Li, 2014). Moreover, the data security needs to be particularly high in settings, there is an increasing concern on maintaining data privacy and security while garnering the benefits (Chen, 2013).

5.6 Practical challenges

There are many practical challenges that we cannot ignore. First of all, frontline medical staffs and healthcare workers may lack of motivation of collect electronic data with less incentives and difficulties in the initial learning period. Secondly, health care is a huge industry with complicated structured and entangled relations. Hospitals, district health centers, insurance companies, food and drug administration, medical devices monitoring center, public health surveillance datasets are like isolated data islands which can not be easily linked (He et al., 2014). Thirdly, lack of standardized information collection system increases the difficulty in data sharing and gathering. Last but not least, the shortage of health information technology human resources impede the fast moving of the informatization process (Cai et al., 2013).

6. CONCLUSION

This paper reviewed studies describing the application of big data in healthcare in China and discussed issues related to data analysis and further use of big data. Big data is the new era of the medicine and becoming a common feature of biological, clinical and public health studies. Medical big data have grown incredibly in size due to use of modern technologies for collection and recording of data. However, we still faces multiple challenges for the use of big data and the data it self has limitations. Policy makers’ foresight, researchers’ interdisplinary skills and general public’s support are essential to make medical big data beneficial to improve health of the whole human being.

REFERENCE

[1] Cai J, Zhang T, Zong W, 'Challenges and considerations of the big data of medicine', *Chinese Journal of Health Informatics and Management*, 10-4 (2013), 292-295.

[2] Chen H, Xu W, '大数据视角下医疗行业发展的新思维', *Modern Management Science* 4(2015),70-72.

[3] Chen L, 'Overview of Medical data privacy protection', *China Digital Medicine,* 8-11 (2013), 95-98.

[4] e-healthcare, '深圳市儿童医院携手 IBM 实现智慧医院跨越式发展', *e-healthcare,* 11 (2014).

[5] Gao H, Sang Z, 'Big data lifecycle and governance in medical industry', *Journal of Medical Intelligence,* 34-9 (2013), 7-11.

[6] Gao H, Xiao L, Xu D, Sang Z, 'Medical Data Mining Platform Based on Cloud Computing', *Journal of Medical Intelligence,* 34-5 (2013), 7-12.

[7] He P, Zhang J, Zhao R, Du N, 'Building and Application of Regional Clinical Pathway Application service System', *China Digital Medicine,* 9-2 (2014), 26-29.

[8] Kroll D, 'Obama's Precision Medicine Initiative: Paying for precision drugs is the challenge', (2015).

[9] Song TM, Ryu S, 'Big data analysis framework for healthcare and social sectors in Korea', *Healthcare informatics research,* 21-1 (2015), 3-9.

[10] Su C, Li Q, Wang D, 'The Influence of Different Basic Medical Insurance on Chinese Residents' Medical Consuption', *Research on Economics and Management*, 10-23-30 (2013).

[11] Teng Q, Fan X, He C, Li Y, Lu D, 'Characteritics ming of large-scale physiological signal and early warning of major disease', *Journal of Network New Media,* 3-1 (2014), 50-54.

[12] The State Council, '十二五期间深化医药卫生体制改革规划暨实施方案', *The Bulletin of the State Council* (2012).

[13] Vayena E, Salathe M, Madoff LC, Brownstein JS, 'Ethical challenges of big data in public health', *PLoS computational biology,* 11-2 (2015), e1003904.

[14] Wang W, Krishnan E, 'Big data and clinicians: a review on the state of the science', *JMIR medical informatics,* 2-1 (2014), e1.

[15] Wyber R, Vaillancourt S, Perry W, Mannava P, Folaranmi T, Celi LA, 'Big data in global health: improving health in low- and middle-income countries', *Bulletin of the World Health Organization,* 93-3 (2015), 203-208.

[16] Yang B, Sun Y, 3-Million Genomes Project (2013).

[17] Yoo C, Ramirez L, Liuzzi J, 'Big data analysis using modern statistical and machine learning methods in medicine', *International neurourology journal*, 18-2 (2014), 50-57.

[18] You S, 'Embracing medical innovation in the era of big data', *Chinese Journal of Gastrointestinal Surgery,* 18-1 (2015), 1-5.

[19] Yu G, Bao X, Huang X, Liu H, Xu B, Yu N *et al.,* 'Medical and Health Big Data: types, characteristics and relevant issues', *Journal of Medical Intelligence,* 35-6 (2014), 9-12.

[20] Zhou D, Li H, 'The research and design of big-data security platform in regional medical field', *Information Technology & Standardization,* 8-25-29 (2014).