

Utilizing big data to build personalized technology and system of diagnosis and treatment in traditional Chinese medicine

Baoyan Liu

China Academy of Chinese Medical Sciences, Beijing 100700, China

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2014

Advent of the big data era

In September 2008, a special issue was published by *Nature* devoted to big data storage, management, and analysis, and it signified the coming of the era of big data [1,2]. Later, more reports and special issues on big data were published by *Science* [3].

In China, big data is becoming a hot topic. Headed by Guojie Li, an academician from the Chinese Academy of Engineering, a task force on big data was created in October 2012 by the China Computer Federation in Dalian, China [4]. In March 2013, the National Natural Science Foundation of China held the 89th Shuangqing Forum on “Challenging the Scientific Issues of Big Data Technology and Application” in Shanghai.

In 2009, a few weeks before the influenza A (H1N1) outbreak, an engineer from Google published a paper that stunned the public. In the article, the author identified the combinations of the 45 most common search terms and fed them into a specific mathematical model. The author compared a list of 50 million most common terms searched by Americans, such as “drugs for cough and fever,” with the U.S. Centers for Disease Control and Prevention (CDC) data on the spread of seasonal influenza from 2003 to 2008. The prediction of a flu pandemic was strongly (97%) correlated with that revealed by CDC figures. The article predicted the flu not only at the national level but also at the regional and even state level. Furthermore, the prediction was near real-time, unlike the figures of the CDC, which were unavailable until one or two weeks after the flu broke out [5,6]. Many other instances signal the emergence of big data. The use of big data slowly triggers a huge transformation in our times.

What are big data then? No consensus on its definition has been achieved so far. According to “Mining the Big Data,” the annual ACM SIGKDD (Association for Computing Machinery’s Special Interest Group on Knowledge Discovery and Data Mining) conference held in August 2012, big data should be larger than what 10 servers can store in terms of storage capacity. Data similar to that must be processed online as one computer’s memory is unable to handle the data. From the perspective of data analysis, some experts argue that big data is so complicated that most methods available are unable to process them and frameworks, such as Hadoop, should be used to handle them. National Science Foundation summarizes the definition of big data as “large, diverse, complex, longitudinal, and/or distributed data sets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future” [7]. However, in their book on big data, Mayer-Schonberger and Cukier (2013) proposed that, from the perspective of exploitation, the tasks of big data could only be accomplished on the basis of massive data rather than on small data. Therefore, big data is a resource for humans to obtain new knowledge and create new value, as well as a way for markets, organizational structures, and relations between governments and citizenry to transform [8].

Therefore, big data is not only huge in volume and complex in structure but also dynamic and rich in useful value. Today, information resources with a potential nature of big data can be found everywhere. Most of these resources are stored in the form of paper or analog information. With the wide use of digitization, networks, and Internet of Things (IOT), these data can be largely activated once digitized and transformed into something with considerable innovative value. In the same vein, data can grow into big data that can boost traditional Chinese medicine (TCM) development and enhance health services generated. These data come from

people's daily lives and monitoring of their health conditions because of the use of IOT and from medical records, rich literature on ancient TCM, and modern scientific research on TCM.

In 1855, Mathew Fontaine Maury, then superintendent of the Depot of Charts and Instruments of the U.S. Navy, made the most out of huge piles of old logbooks from past voyages of Navy captains, which had big data potential, by drawing navigational charts that caused enormous changes in ocean voyages. The logbooks were the captains' sailing records and had not been exploited before. After he found these logbooks in the depot, Maury, together with his 20 data-processing assistants, extracted data from the ragged and musty logbooks and tabulated them. As soon as the data had been aggregated, the map of the whole Atlantic was divided into five segments in terms of longitude and latitude. Monthly temperatures, wind speeds, and directions were noted on the map. Information from these logbooks contributed to the creation of useful charts for marine navigation.

To continuously improve the accuracy of sea routes, Maury demanded all seafarers using his charts to document their voyage experiences using the standard tables especially made for logbooks. The charts were adjusted and improved using the data derived from feedback. The charts enabled the cutting of long voyages by a third and the reduction of voyaging risks. Later, not only the U.S. Navy but also merchant vessels used the charts [9].

Although Google's prediction of the influenza outbreak was more than 100 years behind Maury's creation of navigational charts, the ways and speed to store, manage, and use data have been revolutionarily changed. These examples, on the basis of the "inheritance" of either the search terms or logbooks available, turned all information into data to produce innovative outcomes that either then loggers of the logbooks or contemporary Internet users could not imagine.

Clearly, big data is a resource and tool for inheritance and innovation. The process of harnessing big data for inheritance and innovation is distinct from that of exacting data from small samples, which is studied through causality analysis. However, the former may rely on complete samples from multiple sources to determine correlations for better results. Big data assist in decision making and prediction of the future. The inheritance and innovation produced under big data circumstances are beyond the constraint of causality thinking. With the view of correlative thinking, more works or activities should be directed toward the improvement of service quality and capacity and to specific issues in practice. These works will greatly advance and accelerate the development of the society.

Creation of a TCM navigational chart based on big data

TCM has been developing for over a millennium. As human-centered medicine with significant clinical effects, TCM is based on the correlation of the human body's dynamic condition with interventions to form a theoretical system. This correlation created the ways and methods for its development. Inheritance and innovation are the major developing patterns of TCM. That is, medical knowledge derived from and fed back to clinical practice is the route to progress and advancement, and eminent physicians serve as vehicles and representations. The numerous medical works document the rich experiences and understanding of medicine of TCM practitioners. Aside from the "logbooks" that veritably record TCM development courses, these medical works are important resources for TCM theory and new methods and techniques.

Over the course of history, by continuously summarizing and refining the "logbooks," scholars and TCM practitioners have discovered much hidden knowledge behind diagnosis and treatment in TCM and drawn various "navigational charts" that guide that journey of TCM. Similar to *Huangdi's Inner Classic*, which was the earliest medical work systematically explaining TCM theory, and *the Treatise on Cold Damage and Miscellaneous Diseases*, which pioneered the diagnosis method of pattern differentiation, the "logbooks" are the "navigational charts" produced by ancient wise people who diligently studied ancient medical experiences. These wise people widely collected classic medical formulas passed down in history and sorted out the available TCM "logbooks" by integrating them with clinical experience. Over the course of TCM development, the charts have facilitated TCM to navigate its way through history and guided a great number of TCM "sailors" to save countless lives and greatly contribute to human health.

To sketch a navigational chart, reliable information is essential. Novel thinking and related methods and technology are required to collect valuable data from the countless TCM works and to aggregate them to produce new knowledge. TCM navigational charts were created by copying or abstracting from massive amounts of medical works or notes taken during clinical practices to generate valuable data. This process depended on the data processors' wit and great knowledge to filter and transform data into new knowledge. As the process was constrained by the way ancient medical works were stored, the amount of data individuals could possess, the individuals' capability to read, and their ability to efficiently extract, merge, and process data, valuable information in most TCM "logbooks" is yet to be

exhausted. Therefore, buried in a huge mound of unclassified data, knowledge has not been excavated yet, and a large part of the foundation of TCM has not yet been activated or exploited.

TCM data resources have been greatly enriched. Aside from the numerous ancient TCM works, a great amount of contemporary literature, including that on conventional medicine and biology, and dynamically updated multimedia clinical records exist. These resources form significant *de facto* big data. Technologies applied to digital information, network communication, IOT, data mining, and data storage and transferring provide strong technological support for TCM navigational chart sketching. Through the collaboration of scientists from multiple disciplines, a professional team covering TCM big data generation, management, and utilization should be set up, and a corresponding technical platform should be built. Based on abundant TCM literature and clinical data resources, and combined with research advancements in conventional medicine and biology, a brand new concept of TCM navigational chart driven by clinical issues, guided by data, and underlain by mathematical models should be created. This new TCM navigational chart concept will activate the inheritance-innovation outcomes derived from thousands of years of TCM practice experience; greatly enhance the ability of individualized diagnosis and treatment with pattern differentiation in terms of TCM theory, human health condition prediction, and illnesses prevention and treatment; and improve the service capacity for human health. Navigating with a chart underpinned by big data, TCM should transform from traditional experienced-based medicine to evidence-based medicine. Therefore, a complete transformation of TCM will ensue.

Initiation of the creation of a contemporary TCM navigational chart

In recent years, with information technology constantly developing and advancing, China Academy of Chinese Medical Science (CACMS), as a state-level research entity, has digitized ancient TCM works, contemporary literature, and clinical records, and has laid an important foundation for drawing a contemporary TCM navigational chart. The construction project has formed the digitization, datafication, and knowledge banking of ancient TCM works.

Headed by Changhua Liu, the CACMS team is composed of more than 100 experts nationwide who specialize in the compilation, emendation, digitization, and datafication of ancient works. Thus, the National Center for TCM Ancient Works Compilation, Emendation, and Digitization was established. Supported by the

State Administration of Traditional Chinese Medicine and R&D Infrastructure and Facility Development Program of the Ministry of Science and Technology of China, 1690 TCM ancient works containing more than 1.39 million pages have been scanned, and 613 works roughly containing 69.57 million characters have been appraised. A part of these works is available for full text and image-text cross-referencing. An image database and full text database for 800 works have been established, 350 of which have been further indexed containing 45 million characters. These accomplishments, which began from 2000, form the “Methods for Computerized Presentation of TCM Ancient Works in Terms of Knowledge Elements,” which resolves the issue of digitization of TCM ancient works [10]. Thus, the study team formulated a series of pamphlets or booklets on technical standards that guide the digitization and knowledge elements indexing of TCM ancient works, such as “Guidelines for Emendation and Compilation of Digital Versions of TCM Ancient Works,” “Guidelines on Elusive Words in TCM Ancient Works,” “Criteria for Image Collection of TCM Ancient Works,” “Specifications of Image Processing Technology of TCM Ancient Works,” “Criteria for Metadata of TCM Ancient Works,” and “Specifications of Indexing Procedure for Knowledge Elements of TCM Ancient Works,” among others. Furthermore, “Platform of Knowledge Processing of TCM Ancient Works,” “Processing System of Descriptor Word List of TCM Ancient Works,” “Knowledge Bank of TCM Ancient Works,” and “Reading System of TCM Ancient Works” were developed to provide technological support for the digitization and datafication of TCM ancient works on a large scale and knowledge services for clinical practice. These works address the critical problems of digitization and datafication of TCM ancient works.

Research on the methods for data mining of TCM ancient works is underway. The aforementioned works have laid a strong foundation and provided powerful tools for the creation of a contemporary navigational chart based on TCM ancient works.

Data sharing platform of contemporary TCM technology is a vital part of the Sharing Platform of National Demographics and Human Health Science Data

Since 1984, a study team from CACMS led by Meng Cui has been building a massive database, formulating the specifications of metadata, and constructing the unified TCM language system. Based on WEB technology, a co-constructed platform of TCM scientific literature, with CACMS as its hub, was built. A large team composed of more than 300 researchers from 35 institutes across the

country came together for the digitization of TCM scientific journals and literature. Over 100 databases have been created relevant to TCM, Chinese medicinal acupuncture, and ancient literature, such as the database of TCM journals (as of 1984), database of disease diagnosis and treatment, database of Chinese medicine, database of TCM formulas, database of ethnomedicine, database of Chinese medicinal pharmaceuticals, and database of various national standards covering TCM patterns, treatment principles, illnesses, medicines, and formulas [11]. All these databases are available in an online TCM database retrieval system with a Chinese version suitable for both simplified and traditional Chinese. Some databases provide English versions. Compact discs are available for all databases as well. The retrieval system can conduct a search on one or multiple databases simultaneously. A flat-table database search can be conducted in a specified database with relevant terms, and cross-database searches and multiple-type searches can be performed in all databases.

To take advantage of these data resources, the study team constructed an aggregation platform for data mining and high-capacity computing platform. Until now, as a vital part of the Sharing Platform of National Demographics and Health Science Data, these databases serve and provide local and international users with scientific data. The databases are powerful tools to sketch the contemporary TCM navigational chart by exploiting modern scientific literature.

Formation of the sharing system of TCM clinical and scientific research information and creation of the technical platform in more than 20 hospitals for clinical data collection and massive data analysis

Since 2001, sponsored by mega programs of the Science and Technology Committee of Beijing City, the CACMS study team has focused on clinical effectiveness assessment methods and started to build a sharing system for TCM scientific research information to resolve issues on real-time digitization of clinical diagnosis and treatment data as well as analysis and utilization of complex and massive clinical data. "Application System of Digital TCM Clinical Terms," "Unified Data Collection System for Structured Clinical Studies," "Clinical Data Preprocessing System," "TCM Clinical Data Warehouse System," "Multidimensional Search and Visualization System for TCM Clinical Data," and "Data-mining System of TCM Clinical Data" were created for the sharing system. They have been applied to more than 20 hospitals, including the National Base of Clinical Study. The "TCM Clinical Data Warehouse" contains hundreds and thousands of digital medical

records, and based on these records over 100 studies have been conducted. Underpinned by the sharing system of TCM clinical study information and guided by "Real-World Good Clinical Practice," the digitization of clinical diagnosis and treatment data and research works related to them were initiated. The methods and technology framework of a TCM clinical study have formed following the instinctual nature and logic of its development [12,13].

By using TCM data resources of ancient and contemporary scientific literature and diagnosis and treatment information of clinical practice, a solid foundation has been laid for drawing the TCM navigational chart. Time and tide wait for no man. With the exploitation of big data, the goal of TCM taking off will be realized.

Potential topics of big data research in TCM

As soon as our ideas change, we are freed from the obsession of causality and focus on correlation to seek the ways to continuously improve TCM effectiveness and boost health service quality and capacity

Changing our beliefs is important. The era of big data challenges the way we live our lives and the way we communicate. The greatest step that we can take is for society to abandon its craving for causality and focus on correlation instead, which overthrows the convention inherited from ancient times. How we make decisions and comprehend reality is challenged as well.

A TCM theoretical system involves ontology of relation that emphasizes the interaction and consistency between host and guest. In terms of effectiveness (balancing yin against yang), TCM sets up an indirect and complex correlation between a physician (the host) and his/her interventions and the functional condition of a patient (the guest). This correlation has evolved into a unique and systematic theory system that efficiently guides clinical practice. Despite being a large and complex system, the theory is still unable to elucidate why it can tell what, in terms of biology. These "what's" have greatly contributed to mankind's health, and the system has survived several millennia and remains active until today. The strong wind of the big data era, which focuses on correlations, can blow off the mist hanging over TCM caused by modern society over-pursuing causality for a long time. Every effort must be done to seek effective methods and formulas in terms of correlations to improve the capability to solve problems encountered in medical practice rather than dedicate much energy and resources to small and precise data that address the "why's." If TCM can tell both what and why, it will become new medicine that people will be eager to pursue.

The platform should be set up, the potential of big data available should be activated, and the construction of TCM data resources should be reinforced to enable innovative energy hidden in clinical practice to be maximally released

Converting the ancient and contemporary “logbooks” with inheritance and innovation characteristics to a modern TCM navigational chart efficiently serves TCM clinical practice. Aside from changing our ideas, information with potential value around us should be activated with digital information technology, and the expansion of TCM data resources should be accelerated as well. Moreover, through artificial intelligence and data mining, humans can be combined with machines. By means of the “power of understanding” of individuals, TCM big data can be turned into a powerful tool that can predict dynamic health trends and facilitate diagnosis and intervention, which boosts the health service capacity and level of TCM.

The construction of a national center for TCM data should be expedited, data-sharing mechanism established, data from critical and life-threatening diseases collected, and data from diagnoses and treatments and clinical studies aggregated

The rich clinical resources of TCM and its integration with western medicine are the treasures of China coveted by many people. If the advantages of China’s “whole nation” system can be exploited to the fullest, all data from TCM hospitals can be collected with good capacity, digitized, and aggregated. Therefore, sharable clinical data resources can be formed, and TCM’s utilization of big data for development can be accelerated. The experiences obtained by the National Base of Clinical Study can be used as an advantage in the construction of a sharing system for clinical study information. By following the form of unifying clinical practice and scientific studies and by focusing on critical and life-threatening diseases, a national center for clinical data and its sub-centers should be established. Under unified regulations and standards, data collection, aggregation, and sharing can be conducted by empowering the National Base of Clinical Study and the major specializations. Consequently, they can bring in and motivate related departments in hospitals or even the whole hospital to participate in the efforts. The national center will not only be the hub for data aggregation but will also provide the technology for data digitization, analysis, and utilization. The center will be a study center for big data projects and a training base for trainees from various disciplines.

In Oxford University, the Li Ka-Shing Center for Health

Information and Discover was launched. Sponsored by the Li Ka-Shing Foundation and the Higher Education Funding Council of England, the center will conduct big data analysis. At the launch ceremony, Prime Minister David Cameron stated that it has the potential to revolutionize medical research and health care in the country and that it would pioneer new advances in the analysis of medical data to help scientists better understand human diseases and their treatment. This event deserves to be noted [14].

In recent TCM clinical studies, clinical diagnosis and treatment data is a small and important portion of clinical data. To conduct the research in a larger picture, that is, from the perspectives of academic medicine, clinical medicine, and epidemiology, clinical records should be merged with related medical research data to provide theoretical bases and insights for individualized diagnosis and treatment in terms of TCM pattern differentiation. To build the database of medical and health information, data from ancient literature should be integrated with those from modern scientific literature; clinical laboratory examinations, such as HIS, RIS, PACS, pathology, and medication; medical biological studies such as gene expression, SNP, proteomics, and metabolomics; and new drug clinical trials. Research based on big data not only can answer the “what” but also the “why.”

The development of IOT should be strengthened to enhance the handiness and accuracy of massive data collection of TCM

The development of IOT can improve data collection, storage, transmission uniformity, accuracy, and scope. With IOT being widely used, various data on people’s basic needs (e.g., food, clothing, shelter, and transportation) can be automatically and timely collected. Moreover, with wearable devices for data collection, a great deal of real-time dynamic data on human health conditions can be collected, thus improving the uniformity, handiness, and accuracy of data collection and reducing the cost of data collection. IOT is important not only for data collection but also for clinical interventions. For instance, IOT may remind patients to take their medication on time to improve the effects of the medicines.

Devices and equipment should be greatly developed for TCM diagnosis by taking advantage of digital information technology and by integrating them into the TCM clinical information infrastructure. Merging devices and equipment with clinical data on diagnosis and treatment will ensure strong support for the enhancement of TCM diagnosis and treatment. Currently, TCM four-diagnosis devices are applied in the simulated

astronautic program. Collected far away from the Earth, data on symptoms, pulse diagnosis, and tongue inspection of astronauts in space are systematically analyzed. This analysis provides new data and experience for the automatic checking of health conditions in space and the necessary relevant interventions.

Technology for big data processing and analysis should be further studied, and a data mining framework suitable with the complexity of TCM data should be established. In response to the demands of TCM data analysis, intelligent information processing technology should be developed, targeting the enhancement of the efficacy of TCM information processing.

From the perspective of data analysis, TCM data can be characterized to have multiple dimensions and to be nonlinear and dynamic. Focusing on the features of TCM, intelligent information-processing technologies should be researched and developed. For instance, data pre-processing, data mining, pattern analysis, categorization, and clustering should be developed. Hidden knowledge can be discovered by mining massive clinical data sets as well as patterns and trends that usually cannot be identified without big data support. Clinical experience of experts can keep abreast with the latest knowledge from data mining. They can use the computer system for real-time and automatic monitoring of changes in the way other physicians think and then summarize them to identify novel ideas for pattern differentiation and pragmatic use of medicine. In this way, they can improve their efficiency on training their students, i.e., young TCM physicians. With intelligent information processing technology exploited, the nature and logic of individualized diagnosis and treatment can be revealed in terms of pattern differentiation of TCM, and real-time data of individuals and their environment can be collected with the integration of IOT. Therefore, the goal of individualized medicine can be realized.

A system and mechanism should be created to secure the systematic data digitization and utilization of massive data resources and be able to constantly collect valuable data from impossible sources to generate big data

To create TCM big data, datafication should be done on the rich TCM ancient literature, hundreds of thousands of contemporary research, and medical records of patients produced every day. This information can be turned into inexhaustible data resources. To do this, the following issues must be addressed: (1) capital, a large amount of manpower, and materials should be mobilized; (2) only when all data resources have been aggregated and the national center of TCM data and its

regional centers and sub-centers are built, can TCM big data take shape quickly, and its orderly management and efficient use be realized; (3) as big data have become a huge resource, the clearance of usage, methods, ethical issues, privacy protection, and intellectual property protection should be properly handled; and (4) construction and utilization of data resources are a systematic engineering project underpinned by high technology. Establishing the procedure for the engineering implementation to ensure smooth running and secure quality to achieve the engineering goals requires relevant laws, regulations or measures, and running mechanism. A professional unit at the core and a running mechanism created by combining production, academy, and research are also required for the success of big data.

To conduct big data analysis in TCM, talents from multiple disciplines should be fostered; a knowledge framework should be built for researchers involved in big data analysis; key issues on the innovation and development of TCM should be explicitly identified; and knowledge and technology related to computers, data structure, and data mining models should be understood and grasped. Talents should be absorbed from different disciplines, and the combination of enterprises and research institutes should be promoted to provide sufficient technical tools for TCM data analysis and data mining. Specifically, graduate courses should be set up to attract professional researchers from the fields of TCM and data study to cultivate interdisciplinary talents, who not only understand TCM but are also well versed in clinical practice, data collection, and analysis.

Conclusions

With thousands of years of development, TCM has accumulated a huge mass of big data. On the basis of the unique characteristics of TCM big data, improving the utilization ability and fully tapping the resource value of TCM big data not only require a technical support system but also clinical medicine, information, standardization research, statistical analysis, and multi-disciplinary collaboration of researchers in many fields. Organization and management development strategy can serve as a pair of invisible hands that guide and plan big data research. On the side of theory, based on “from clinic, to clinic”, the real world clinical research paradigm is proposed to build the personalized diagnosis and treatment technology based on big data. On the side of techniques, computer collection and storage, mathematical statistics, artificial intelligence, data mining, information retrieval are developed rapidly, providing the solid technical bases to realize the dream of creating a

navigational chart for TCM. Integrating the above techniques and different subjects, a TCM big data collaborative innovation center should be established; data integration standard systems should be improved and perfected; knowledge sharing and innovation should be improved; and the quality of talent and staff should be improved and expanded to lay the talent and technology foundation to innovate the TCM clinical research methodology.

Acknowledgements

We would like to thank Prof. Sulun Sun and Prof. Shusong Mao for guiding us on behalf of the State Administration of Traditional Chinese Medicine National TCM Clinical Research Base of the Chinese Medicine and Clinical Research Information Sharing System Construction Expert Group.

Compliance with ethics guidelines

Baoyan Liu declares that he has no conflict of interest. This manuscript does not involve a research protocol requiring approval from relevant institutional review boards or ethics committees.

References

1. Mervis J. Agencies rally to tackle big data. *Science* 2012; 6: 22
2. Lazer D, Kennedy R, King G, Vespignani A. The parable of Google Flu: traps in big data analysis. *Science* 2014; 343(6176): 1203–1205
3. Reichman OJ, Jones MB, Schildhauer MP. Challenges and opportunities of open data in ecology. *Science* 2011; 331(6018): 703–705
4. China Computer Federation. The establishment ceremony of task force on big data. October 26, 2012. http://www.ict.ac.cn/xwzx/jssxw/201210/t20121026_3667554.html
5. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009; 57(7232): 1012–1015
6. Butler D. Web data predict flu. *Nature* 2008; 456: 287–288
7. National Science Foundation. Solicitation 12-499: Core techniques and technologies for advancing big data science & engineering (BIGDATA). <http://www.nsf.gov/pubs/2012/nsf12499/nsf12499.pdf> (Accessed August 10, 2014)
8. Mayer-Schönberger V, Cukier K. *Big Data: a revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt, 2013
9. Maury MF. *The Physical Geography of the Sea*. Nabu Press, 2010
10. Liu CH. Contemplation on the construction of knowledge base system of traditional Chinese medicine literature digitization project of TCM ancient literature. Proceedings of Chinese TCM information research conference Second session of Congress and academic exchange conference, 2003, 1–5 (in Chinese)
11. Cui M, Xie Q, Li HY. Connotation and principles research of traditional Chinese medicine informatics. *J Zhejiang Univ Tradit Chin Med (Zhejiang Zhong Yi Yao Da Xue Xue Bao)* 2009; 33: 638–641 (in Chinese)
12. Liu BY. The real world TCM clinical scientific research paradigm. *J Tradit Chin Med (Zhong Yi Za Zhi)* 2013; 54(6): 451–455 (in Chinese)
13. Liu BY, Xie Q, Shi HX, Wag B, Zhou XZ, Zhang RS, Guo YF, Zhang XP. Construction of technology platform of clinical research in real world: organization and management strategies. *J Tradit Chin Med (Zhong Yi Za Zhi)* 2013; 54(24): 2071–2075 (in Chinese)
14. Liu SL. Open ceremony of the first big data medical institute in United Kingdom. May 4, 2013. http://news.xinhuanet.com/tech/2013-05/04/c_115638244.htm