

Current status of China's critical care medicine big data platform and future prospects

Long-Xiang Su¹, Sheng-Jun Liu², Xiang Zhou¹, Mei-Li Duan³, Jian Liu⁴, Bin Ouyang⁵, Bing-Yu Qin⁶, Hong-Liang Wang⁷, Xue Wang⁸, Lei Xu⁹, Xiang-You Yu¹⁰, Fei-Hu Zhou¹¹, Yun Long¹

¹Department of Critical care medicine, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Science & Peking Union Medical College, Beijing 100730, China;

²Department of Postdoctoral Residency Program, Chinese Academy of Medical Science & Peking Union Medical College, Beijing 100730, China;

³Department of Critical Care Medicine, Beijing Friendship Hospital, Capital Medical University, Beijing 100050, China;

⁴Department of Critical Care Medicine, The First Hospital of Lanzhou University, Lanzhou, Gansu 730000, China;

⁵Department of Critical Care Medicine, The First Affiliated Hospital of Sun-YatSen Hospital, Guangzhou, Guangdong 510080, China;

⁶The Department of Critical Care Medicine, Henan Provincial People's Hospital, Zhengzhou, Henan 450003, China;

⁷Department of Critical Care Medicine, The Second Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang 150001, China;

⁸Department of Critical Care Medicine, First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi 710061, China;

⁹Department of Critical Care Medicine, Tianjin Third Central Hospital, Tianjin 300170, China;

¹⁰Department of Critical Care Medicine, The First Affiliated Hospital of Xinjiang Medical University, Urumqi, Xinjiang 830054, China;

¹¹Department of Critical Care Medicine, First Medical Center, Chinese PLA General Hospital, Beijing 100853, China.

Overview

The term “Big Data” originated at the 11th Electronic Materials Conference World Annual Conference, which originally referred to the large amount of data generated by the application of technology.^[1] Medical big data includes not only the medical history and examination data accumulated during patient hospitalization, but also patient-related follow-up data, prognostic data from outpatient, emergency, and medical insurance settlement departments as well as clinical experiment centers. So far, it has profound applications in the various specialties of medicine.^[2-4] However, intensive care medicine (ICU) is different from other medical fields. In comparison with clinical practice data, medical data in ICU have the following characteristics: large scale, rapid production, diverse dimensions, inaccuracies, heterogeneity, incompleteness, complexity, and privacy concerns.^[5] In fact, in the process of constructing major ICU databases in China and worldwide, these databases have been optimized at great length. Taking heterogeneity as an example, the possible formats for medical data include text type, digital type, and image type. Text-based data include demographic characteristics, drug use, medical history, and symptoms. Digital data include various laboratory test results, vital signs, and monitoring instrument data.

Image-type data include various imaging evaluations, such as type-B ultrasound, computerized tomography, magnetic resonance imaging, X-ray, and other examinations. In the formation of a database, standardized formats for text-type data are required. For digital data, the time of the collection is emphasized, and trends in treatment methods and various indicators are considered. For image data, in addition to conventional imaging reports, an artificial intelligence image reading model could be established by combining diagnoses and image characteristics. So far, China has not yet reached an industry-standard consensus in the field of critical care medicine. The Critical Care Medicine Branch of the China Health Information and Health Care Big Data Society was established in March 2019 and aims to establish and promote critical care big data applications. After a comprehensive investigation, it was found that quantity and quality of ICU database construction were limited. More effort should be made to improve this situation. We fully believe that the development of China's critical care databases will make utmost of big data platform, which could provide patients with better diagnosis and treatment processes, provide management decision-makers with accurate and objective diagnosis and provide medical researchers with a higher-quality data.

Access this article online

Quick Response Code:



Website:

www.cmj.org

DOI:

10.1097/CM9.0000000000001366

Long-Xiang Su and Sheng-Jun Liu contributed equally to this work.

Correspondence to: Dr. Yun Long, Department of Critical Care Medicine, Chinese Academy of Medical Science & Peking Union Medical College, Beijing 100730, China
E-Mail: ly_icu@aliyun.com

Copyright © 2021 The Chinese Medical Association, produced by Wolters Kluwer, Inc. under the CC-BY-NC-ND license. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Chinese Medical Journal 2021;134(14)

Received: 28-10-2020 Edited by: Peng Lyu

Survey on the demand for a Chinese critical care database

From March to May 2020, supervised by the Critical Care Medicine Branch of the China Health Information and Health Care Big Data Association, a survey was conducted among critical care physicians in 20 tertiary hospitals across China. A total of 58 valid questionnaires were collected. Among the physicians who filled out the questionnaire, 64.0% were chief and deputy chief physicians, 59.0% were doctors under 45 years old, and 88.0% were medical practitioners with a master's degree or above. Hence, these results represent the characteristics of China's current critical care medicine database management and demand. The survey was used to analyze two main areas, critical care database construction and critical care database demand.

Current status of critical care medicine database construction in China

High database construction cost and low industry-standard compliance rate

Forty-three percent of the interviewed doctors claim their hospitals have not yet established an independent critical illness database. The main reason was funders cannot afford the high costs of database construction and maintenance. After visiting a large domestic medical data company, we found that quotes for intensive care database solutions for middle-sized tertiary class hospitals were at least RMB 500,000 Yuan. In addition, no matter how the databases were built, they failed to meet certain industry standard. Only approximately 1/3 of hospitals refer to ICD-10, WS445, WS364, or other common data standard processing rules while constructing databases. Only 32.0% of surveyed hospitals referenced three or more database industry standards. The use of data from the public critical care medicine database can effectively solve these problems, but only approximately 34% of physicians have used the public critical care database.

Low number of cases and uneven quality of data

Among the hospitals that have established databases, it is difficult to quickly collect a usable amount of data due to the limited number of beds in their ICUs. In the survey, it was found that 72.7% of the databases had been established for 1–3 or >3 years, but <1/5 of the databases included information on >5000 cases [Supplemental Figure 1A, <http://links.lww.com/CM9/A455>]. Nearly half of the ICUs surveyed included only 500–1000 cases in their intensive care database each year. Data collection at this rate may fail to produce the number of cases that sufficient for carrying out scientific research in a short period of time [Supplemental Figure 1B, <http://links.lww.com/CM9/A455>]. In addition, even if sufficient samples are included, due to the lack of necessary quality control, most hospitals have missing data and incomplete records in their databases, which could seriously affect data analysis.

Low data source access rate and low data security reliability

Regarding data storage methods, >85% of hospitals use intra-hospital network storage models. Although the safety of patient information can be guaranteed, the storage

efficiency of these models is relatively low, and there are hidden trouble in data storage reliability. We also surveyed the data sources to which hospitals have access. The sources of data that >50% of the hospitals' ICU databases had access to include: Hospital information system (87.9%), the nursing system (87.9%), the laboratory information system (69.7%), the electronic medical records system (66.7%), and the first page of the medical record (51.5%). However, several data sources that can provide information on patients' conditions were not well-integrated with the databases at the surveyed hospitals, including the picture archiving and communication system (36.4%), the surgical anesthesia system (33.3%), the electrocardiogram reporting system (30.3%), etc. [Supplemental Figure 2A, <http://links.lww.com/CM9/A455>]. When asked about data collection methods, 63.0% of survey respondents indicated that they needed to manually enter data into the database, while fewer than 20% of responders use platforms that provide intelligent data capture services [Supplemental Figure 2B, <http://links.lww.com/CM9/A455>].

Low penetration rate of ICU specialty databases and limited data application scenarios

In this survey, half physicians used the independent critical illness database established by the hospital (50.0%) and the independent critical illness database established by the hospital or department (46.6%). Patients in the ICU have complex illness states that change over time. Most of the non-specialist database information cannot fully cover the items necessary to describe ICU care. In addition, the update frequency, granularity, and raw data processing capabilities of non-specialist databases cannot meet clinical needs. Only one-third (31.0%) of physicians use mature public databases hosted outside of China (such as Medical Information Mart for Intensive Care).^[6]

In this survey, although up to 87.9% of physicians had used various critical medicine databases in their daily work, most physicians used these databases mainly for scientific research (81.0%) and department management (65.5%). However, with regard to resolving clinical problems, a low database utilization rate was reported for clinical decision-making applications (56.9%), disease risk prediction (50.0%), etc. [Supplemental Figure 3, <http://links.lww.com/CM9/A455>].

Requirements for a critical care medicine database in China

In this survey, 88.0% of the respondents believed that it is necessary to establish an independent critical illness database. Among the respondents, 53.5% indicated that their department has an independent critical illness database that can be used for clinical work [Supplemental Figure 4A, <http://links.lww.com/CM9/A455>], while 43.1% of the respondents said that their hospital has not yet established a database platform. Among these hospitals, approximately 2/3 are preparing to build an independent critical medicine database. For the other 1/3 hospitals, high research and development expenditure become an obstacle.

The most urgent demand of critical medicine databases is for clinical scientific research and assistance of clinical diagnosis

and treatment. Under intense work pressure, most clinicians lack time to carry out scientific research. Even when clinicians have ideas for research studies, it is difficult to complete high-quality clinical research without sufficient clinical data. The most fundamental purpose of clinical research is to put the results into clinical practice, to improve diagnosis and treatment decisions for patients. In this survey, more than half of the respondents believed that the existing critical illness database cannot fully meet the needs of scientific research, due to the uneven quality of the database, the low rate of publication of scientific research articles, and the insufficient number of scientific research topics. However, the construction of a critical care medical database requires more funding support and time investment, which negatively reinforces the insufficient rate of scientific research article publication and the low number of research topics.

During the clinical diagnosis and treatment process [Supplemental Figure 4B, <http://links.lww.com/CM9/A455>], respondents said that they would like the critical care medicine database to assist them with clinical diagnosis and treatment decisions (56.9%), disease risk assessment prediction (50.0%), efficacy assessment (46.6%), and multidisciplinary treatment (25.9%). Regarding these goals, the current critical care database can only predict the patient's disease risk based on the existing clinical guidelines and related scoring systems. Due to the lack of mature machine learning modules, curative effect evaluation modules, and 360° patient diagnosis and treatment information comprehensive viewing modules, the clinical assistance functions expected by the interviewees cannot be fully realized by the current databases.

In medical management, respondents believed that their existing critical illness database could not meet the requirements of the department in terms of disease medical quality management (63.6%), operational data analysis (60.6%), medication management (57.6%), and hospital control management (51.5%) [Supplemental Figure 4C, <http://links.lww.com/CM9/A455>]. Hence, the lack of top-level design for medical management during the database construction process is closely related to the inability to fully meet the above medical management needs.

The future of the Chinese critical care medicine database

To better build the Chinese Critical Care Medicine database and maximize its data utilization, the following standards should be met as much as possible in the construction of the database^[7]:

1. Data storage: The overall structure of a two-level star network should be constructed with the China Health Information and Health Care Big Data Society as the center, and with the cooperation of various academic groups to achieve effective organization and management of data.
2. Data collection: To achieve real-time data uploading and improve data granularity, an automated extraction mode should be adopted as the leading mode, and an intelligent prediction mode should be combined with the manual input mode. Consistent standardized data collection sites should be used in each center to unify

data collection standards. Regular quality control should be carried out. Specific items and certain granular data should be collected for specific patients.

3. Data analysis: The necessary data analysis software should be embedded in the software integration platform, and online analysis with common statistical methods can be done according to the types and characteristics of the data, which could save time for clinical researchers and improve work efficiency.
4. Data display: According to the needs of policymakers, medical administrators, physicians, nurses, and other personnel involved in patient diagnosis and treatment, the original data, statistical data, and various reports should be presented to facilitate various decisions.
5. Data security: The principle of record unification and deidentification processing in the data center and limited data sharing among member units should be adopted to achieve the maximum use of data based on ensuring data security.

Acknowledgment

We thank for the assistance from Yidu Tech (Hang Peng, Kai-Tian Luo, Yan-Shan Shi, and Ruo-Bing Xia).

Funding

The study was supported by the China Health Information and Health Care Big Data Association Severe Infection Analgesia and Sedation Big Data Special Fund (No. Z-2019-1-001) and the China International Medical Exchange Foundation Special Fund for Young and Middle-aged Medical Research (No. Z-2018-35-1902).

Conflicts of interest

None.

References

1. Bailly S, Meyfroidt G, Timsit JF. What's new in ICU in 2020: big data and machine learning. *Intensive Care Med* 2018;44:1524-1527. doi: 10.1007/s00134-017-5034-3.
2. Li Q, Fan QL, Han QX, Geng WJ, Zhao HH, Ding XN, *et al.* Machine learning in nephrology: scratching the surface. *Chin Med J* 2020;133:687-698. doi: 10.1097/CM9.0000000000000694.
3. Qiu QT, Zhang J, Duan JH, Wu SZ, Ding JL, Yin Y. Development and validation of radiomics model built by incorporating machine learning for identifying liver fibrosis and early-stage cirrhosis. *Chin Med J* 2020;133:2653-2659. doi: 10.1097/CM9.0000000000001113.
4. Yuan H, Fan XS, Jin Y, He JX, Gui Y, Song LY, *et al.* Development of heart failure risk prediction models based on a multi-marker approach using random forest algorithms. *Chin Med J* 2019;132:819-826. doi: 10.1097/CM9.0000000000000149.
5. Norrie J. The challenge of implementing AI models in the ICU. *Lancet Respir Med* 2018;6:886-888. doi: 10.1016/S2213-2600(18)30412-0.
6. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, *et al.* MIMIC-III, a freely accessible critical care database. *Sci Data* 2016;3:160035. doi: 10.1038/sdata.2016.35.
7. Zeng X, Yu G, Lu Y, Tan L, Wu X, Shi S, *et al.* PIC, a paediatric-specific intensive care database. *Sci Data* 2020;7:14. doi: 10.1038/s41597-020-0355-4.

How to cite this article: Su LX, Liu SJ, Zhou X, Duan ML, Liu J, Ouyang B, Qin BY, Wang HL, Wang X, Xu L, Yu XY, Zhou FH, Long Y. Current status of China's critical care medicine big data platform and future prospects. *Chin Med J* 2021;134:1684-1686. doi: 10.1097/CM9.0000000000001366