Co-construction and sharing of a Chinese medical case repository helps medical workers strengthen capacity for clinical practices

Junmin Wei, Jing Li, Xibin Shen and Qihang Tian
Chinese Medical Association Publishing House, China

Abstract
The advent of the era of big data is bound to promote the development of databases. How to bring clinicians’ focus back to clinical practice and enhance the clinical capabilities of general physicians at the primary level is a question that every member of the academic publishing profession needs to consider. If data and knowledge can be shared openly, research and medical management will develop at an accelerated pace. Given the global consensus on open science, the co-construction and sharing of a Chinese medical case repository will surely improve the practical capability of clinicians.

Keywords
Chinese medical case repository, co-construction, open sharing, application and evaluation

1. Background on open science
The 41st session of the General Conference of the United Nations Educational, Scientific and Cultural Organization was held from 9 to 24 November 2021. Following an inclusive, transparent and multi-stakeholder consultation, the meeting adopted the Recommendation on Open Science, marking a new phase of global consensus on this topic. Open science includes open access (OA), open data, open reproducible research, open science evaluation and open science policies (Paton and Kobayashi, 2019).

OA was proposed by the Open Society Institute in the Budapest Open Access Initiative published in 2001 with the aim of providing the public unrestricted and free access to academic research resources. There are two main OA forms: gold OA (journal-based, free and immediate access to the full content) and green OA (subject-based, self-archiving repository) should develop side by side. With the advancement of technology and data-driven research, open research data as well as data and text mining have attracted a great deal of interest.

Green OA allows authors to deposit their articles (pre- or post-publication) into the publisher’s archive and other repositories, such as a pre-publishing platform. The copyright of a green OA article usually belongs to the publisher or a social organization, and there are specific terms and conditions that

Corresponding author:
Jing Li, Key Laboratory of Knowledge Mining and Service for Medical Journals, Chinese Medical Association Publishing House, 69 Dongheyuan Street, Xicheng District, Beijing 100052, China.
Email: lijing@cmaph.org

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (https://us.sagepub.com/en-us/nam/open-access-at-sage).
specify how and when public access to the article can be granted in the repository. The advantage of this approach is that the author does not have to pay any fees.

In the case of gold OA, authors retain the copyright of their articles and can reuse a published article (or a revised version of the article) in a new publication without seeking prior permission; however, they are required to cite the published article. Scholars may use the articles of other authors for non-commercial purposes with the consent of the editorial office or the publisher provided that they use proper citation. This is the licence for republishing a published article in an OA publication.

2. Implications of open science

At the end of 2019, the novel coronavirus (COVID-19) emerged, and the World Health Organization declared that the outbreak was a public health emergency of international concern. COVID-19 continues to affect people’s health, the economy and our society, and the global impact of the disease has generated a great deal of interest in the development of new drugs and vaccines. Open science has been considered to be an accelerator of biomedical research and is important in the face of the pandemic. The spread of the disease forced people to work aggressively on the development of new therapeutics, and open science has the advantages of sharing data, reducing duplication of work and accelerating the dissemination of important research results. As of August 2020, more than 7000 articles on COVID-19 had been submitted to the bioRxiv15 and medRxiv16 platforms for pre-publication (Tse et al., 2020); more than 2000 papers related to COVID-19 have been made available for free and rapid public exchange on the academic exchange platform for COVID-19 research results created by the Chinese Medical Association Publishing House. The purpose of pre-publication is to give researchers an opportunity to quickly disseminate their research results prior to formal publication. In the field of rare disease (RD), the value of open science is even more significant. RD research has been constrained by a lack of patients and resources as well as the need for experienced expert teams and therapeutics. These problems have caused delays in diagnosis and treatment and may be ameliorated when open science principles are followed (Rubinstein et al., 2020).

Open data sharing and access are an inevitable result of increased requirements for research transparency and reproducibility and can help facilitate secondary research on health management, education and training, and innovation. However, researchers may be reluctant to share their data for many reasons, including time and resource constraints, privacy issues linked to patients’ data, lack of access to appropriate funding, inadequate recognition of the contributions of data sponsors, and concerns that commercial or academic competitors may benefit from the shared data. Despite all these factors, researchers and users have a strong interest in creating data resources for sharing. From this perspective, we will make a tentative analysis of the scope of ‘openness’ and ‘data access’ and the strengths and weaknesses of existing data-sharing platforms based on a Chinese example of a medical case repository.

3. Background on the construction of the Chinese medical case repository

Artificial intelligence (AI) is a general term that encompasses a range of technologies developed for solving problems by using human-like intelligence. Machine learning (ML) is one type of AI approach. In the medical care industry, some people suggest using ML-supported clinical decision support systems to conduct image analysis and interpretation in radiology, dermatology and pathology (Hosny et al., 2018; Fogel and Kvedar, 2018; Zakhem et al., 2018; Salto-Tellez et al., 2019) and to improve the scope and accuracy of biological signal interpretation (Orphanidou et al., 2015; Triantafyllidis et al., 2015). The database-enabled deep-learning approach provides a foundation for the creation of a Chinese medical case repository.

Against the backdrop of rapid advances in technology, AI is starting to be applied in medical care, and a number of issues have emerged that deserve people’s attention. Medical care is a critical industry involving human lives. Compared to systems that are currently using low-risk AI technologies, such as
Alpha Go (Silver et al., 2016), the medical care industry needs a higher level of openness and validation before adopting AI-based tools. By examining how open science can guide the use of AI in the medical care industry, we can recognize that new ML systems need big data sets to improve their accuracy and predictive ability. Currently, biomedical databases developed by governments, medical societies and international research collaborations are relatively mature open databases in foreign countries (Table 1).

The open databases shown in Table 1 are used to keep abreast of bioinformatics developments. For example, they can be used to improve the prediction of drug responses in cancer cell lines, and to reveal the relationship between rare variants and diseases through phenotypic association analysis of genomic data. Open ML databases that are based on clinical information, such as the CheX-pert database for chest X-rays (Irvin et al., 2019), are now being widely used. In China, the relatively mature databases, such as CNKI, Wanfang Data and Cqvip, still primarily consist of papers published in scientific journals, and there are not yet databases that are based on clinical cases and serve clinical purposes.

Propelled by the trend of open science in the academic circle, scientific journals are placing new requirements on open access to articles, open data, open practice, open collaboration and open awards, aiming to promote faster and more effective science development and improve the transparency and reproducibility of research results. Case report articles are an innovative type of medical research paper.

China’s 2021 document on reform guidelines for the professional title system for health professionals published by the Ministry of Human Resources and Social Security, the National Health Commission and the National Administration of Traditional Chinese Medicine clearly states that a system of representative works should be implemented. Clinical cases, surgical videos and nursing cases, among others, can all be used as representative works when applying for a professional title. To overcome institutional barriers, structural problems and policy obstacles in scientific and technological innovation, the China Association for Science and Technology (CAST) has proposed a ‘clinical case representative works system’, encouraging clinical practitioners to write case reports based on their clinical practices, thus gathering practical experience and improving treatment performance. This approach will also help to improve the evaluation system for health professionals and accelerate the reform of the system.

In the context of open science, following the principle of green OA and the guidance of China’s policy on professional title evaluation and CAST’s initiative of a ‘clinical case representative works system’, the Chinese Medical Association Publishing House has constructed a Chinese medical case repository with the funding support of CAST. The purpose of the database is to encourage clinical practitioners to produce standardized case reports based on their clinical practices, thus improving treatment services through case sharing, expanding the diversified evaluation of physicians and accelerating the reform of the talent evaluation system.

**Table 1.** Open databases for biomedical research in foreign countries.

| Organization | Open database |
|--------------|--------------|
| National Center for Biotechnology Information, National Cancer Institute | Gene Expression Omnibus (GEO), Genbank, Pubmed, Omics Datasets |
| European Bioinformatics Institute (EBI) | Gene-Wide Association Study, Uniprot, EBI-Pfam |
| Research Collaboratory for Structural Bioinformatics (RCSB) | The Protein Data Bank (RCSB PDB) |
| Gene Ontology Consortium | Gene Ontology |
| SIB Swiss Institute of Bioinformatics, Novo Nordisk Foundation Center for Protein Research, European Molecular Biology Laboratory, Centre for Genomics and Personalized Medicine | String Database (a database of protein interactions and outputs) |
| Stanford University | RegulomeDB (a comprehensive database linked to GEO) |
| National Cancer Institute | LIDC-LDI (a database with lung images from 1018 cases) |
4. Methods for building the Chinese medical case repository

4.1 Use case resources available in the journals published by the Chinese medical association (CMA)

The construction of the Chinese medical case database would not be possible without data, so the main task in this process is collecting clinical case reports from extensive sources and expanding the size of the database. CMA journals, most of which are at the forefront of their respective disciplines, have the reputation of being the ‘banner of Chinese scientific and technical journals’. Thus, we used the CMA’s valuable resources in building the database. However, case reports have not yet been widely recognized and promoted in China’s medical institutions. Therefore, in order to expand the size of the database, we have tried to introduce and promote the database to medical workers and editorial board members of journals at various academic conferences that take place within the medical field. We have also developed a unified template for writing based on the characteristics of medical research and have trained medical workers in how to write case reports. In order to encourage medical workers to publish their clinical research results on China’s platforms, we have also introduced an incentive mechanism that involves awards for high-quality case reports.

4.2 Joint construction by medical institutions across the country

The construction of the Chinese medical case repository requires the joint efforts of the CMA, journal editors and clinical medical workers. If an open database were to be built based only on the data of a single institution and if it were to collect only limited clinical cases, the goal of green OA could not be fully realized. An open database must be a regional alliance of institutions if it hopes to demonstrate the value of region-wide open sharing of science and technology information.

Construction based on different groups. Medical workers should be evaluated based on different groups according to the nature of their work. Roughly, they can be divided into three groups: clinical workers who are mainly engaged in treating patients and saving lives; scientific research workers who are engaged in academic innovation and development; and educational workers who are primarily engaged in instructing and answering questions. Clinical workers should be evaluated on the basis of their performance and practical work ability. Practical work ability is reflected through high-quality clinical cases. Case reports may appear in a unitary format but have diverse content. In the era of big data, using text alone is no longer the mainstream form of communication. Recording videos of the treatment process in clinical cases, especially in the context of rare and difficult diseases, is a new model of case reporting.

In the medical case database, we have opened a special section for video resources that focus on rare diseases and established a rare disease expert group. Using the CMA’s advantages, we have worked with international academic organizations on rare disease research and with patient organizations at home and abroad to hold seminars on rare disease cases and saved the video recordings of those meetings in the video section of the database. Our aim is to provide clinicians, especially those working in primary-level hospitals, with more vivid learning materials that improve their ability to recognize and treat rare diseases. Case reports, imaging materials, videos and other types of resources have enriched the content of the database.

Construction based on different grades. When case reports are used as an indicator for evaluation, their quality should be valued over their quantity. To further promote the development of the database, we need to engage with medical institutions and front-line clinicians at different levels to understand the real needs in the reform of the evaluation of clinicians and learn about the views of medical workers on the ‘clinical case representative works system’. We need to organize experts in the field of clinical medicine to develop a comprehensive, graded evaluation method based on some data, such as quantified review conclusions of case reports (including their clinical value, academic quality and writing quality) and the number of their downloads, reposts and
comments. Only when case reports are graded can they present their full value and better represent the authors’ qualifications when used in title promotion.

Construction based on different levels. The level and regional location of medical institutions determine their requirements for capacity building and their requirements for the abilities and responsibilities of doctors. It is thus not scientific to use case reports as the sole criterion for evaluating all hospitals and all doctors; the role and value of case reports need to be differentiated based on the level of hospitals. Case reports provided by top-class Grade A hospitals should focus on the diagnosis and treatment of rare and difficult diseases, and they can be used as instructional materials for medical workers at primary-level medical institutions. Case reports provided by primary-level medical institutions, on the other hand, should focus on the standardized treatment of common diseases. They can reflect the medical skills of primary-level doctors and can be used as a reference in title promotion. The subdivision of case sources and case types satisfies the varied needs of primary-level doctors, general practitioners and specialists and lays the foundation for the evaluation of clinicians based on different levels.

Construction based on different disciplines. The process of disease progression is complex and sometimes requires multidisciplinary consultation between clinical and medical-technical departments. Therefore, cases that originate in medical-technical departments are equally valuable for the database. However, due to their different clinical duties, medical-technical and clinical departments have different focuses, so different case report templates need to be developed according to the focus of specific departments. When the database has collected a certain number of case reports, it will be possible to highlight the characteristics of each hospital’s specialties and even analyse the disease spectrum of different geographical regions using discipline-based marks of case reports.

5. Sharing

The Chinese medical case repository publishes the full text of case reports in an OA model (under the CC-BY-NC-ND protocol). All articles published on the platform are available for free on the internet, and any user can read, download, print, retrieve and hyperlink the literature for non-profit purposes. After publishing articles in the database, authors are permitted to republish them on other media platforms. While sharing their own cases, clinicians can also study how others treat rare and difficult diseases. This could help primary-level doctors improve their professional skills and thus could promote the even distribution of medical resources. In this way, patients are able to get primary medical treatment in local hospitals, thus relieving the pressure on top-class Grade A hospitals and addressing the public’s concern over the difficulty in accessing medical services. Through the platform of the repository, an institution-based evaluation of clinical capacity could be conducted, which will in turn promote the participation of universities, hospitals and other institutions in the co-construction of the repository, thus adding new content to the sharing model. In addition, a specialty-based evaluation of clinical capacity conducted through the repository could also mobilize the specialist branches of the CMA and further enrich member services (such as case review and interactive Q&A).

With the expansion of ML technology research, the database will inevitably focus on the capability to predict diseases. Through the continuous collection of high-quality cases and the accumulation of authoritative knowledge and using ML technology, machines will be able to recognize high-quality cases and understand and apply medical knowledge. They will function like a medical brain that provides assistance to clinicians in their decision-making during the clinical treatment process. However, what is described here is an ideal situation. Even if new medical technologies can work in the lab, they cannot be readily transferred to clinical practices in real-world contexts. If medical workers stop seeking self-improvement, give up making clinical decisions and rely too heavily on database analysis, that may lead to a deterioration in their clinical decision skills and adversely affect treatment results. Therefore, the Chinese medical case repository should improve its graded evaluation system and play a role in scientific evaluation and continuing education.
6. Conclusion

The co-construction of the database cannot be achieved without the support of the government, doctors and patients and coordination among hospitals. Sharing is not just about doctors improving their clinical practices. It is also about providing better medical services for patients and creating a greater latitude for the country in dealing with public health issues. The creators of and contributors to the database resources are also the owners and users of those resources, while journal editors act as intermediaries. By providing and maintaining the Chinese medical case repository, journal editors are making their contribution to the improvement of doctors’ clinical skills and the development of the open science process.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors received no financial support for the research, authorship and/or publication of this article.

References

Fogel AL and Kvedar JC (2018) Artificial intelligence powers digital medicine. NPJ Digital Medicine 1: 5.
Hosny A, Parmar C, Quackenbush J, et al. (2018) Artificial intelligence in radiology. Nature Reviews Cancer 18(8): 500–510.
Irvin J, Rajpurkar P, Ko M, et al. (2019) CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. Proceedings of the AAAI Conference on Artificial Intelligence 33: 590–597.
Orphanidou C, Bonnici T, Charlton P, et al. (2015) Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring. IEEE Journal of Biomedical and Health Informatics 19(3): 832–838.
Paton C and Kobayashi S (2019) An open science approach to artificial intelligence in healthcare. Yearbook of Medical Informatics 28(1): 47–51.
Rubinstein YR, Robinson PN, Gahl WA, et al. (2020) The case for open science: Rare diseases. JAMIA Open 3(3): 472–486.
Salto-Tellez M, Maxwell P and Hamilton P (2019) Artificial intelligence: The third revolution in pathology. Histopathology 74(3): 372–376.
Silver D, Huang A, Maddison CJ, et al. (2016) Mastering the game of go with deep neural networks and tree search. Nature 529(7587): 484–489.
Triantafyllidis A, Velardo C, Chantler T, et al. (2015) A personalised mobile-based home monitoring system for heart failure: The support-HF study. International Journal of Medical Informatics 84(10): 743–753.
Tse EG, Klug DM and Todd MH (2020) Open science approaches to COVID-19. F1000Research 9: 1043.
Wang SS (2021) Open-access journals and the ethical issues of open-access articles. Chinese Journal of Tissue Engineering Research 25(21): 3278 (in Chinese).
Zakhem GA, Motosko CC and Ho RS (2018) How should artificial intelligence screen for skin cancer and deliver diagnostic predictions to patients? JAMA Dermatology 154(12): 1383–1384.

Author biographies

Junmin Wei is general manager and editor-in-chief of the Chinese Medical Association Publishing House (CMA Publishing House). His research focuses on journal management and integrated development.

Jing Li is a journal publisher and vice director of the department of Chinese medical case repository of the CMA Publishing House. Her research focuses on digital publishing and its combination with traditional publishing. She has published 10 academic articles.

Xibin Shen is director of the department of digital publishing of the CMA Publishing House and assistant director of the Key Laboratory of Knowledge Mining and Service of Medical Journals. His research focuses on digital publishing and OA publishing. He has conducted several provincial and ministerial research projects and published more than 30 academic articles.

Qihang Tian is an assistant journal publisher of the department of Chinese medical case repository of the CMA Publishing House. His research focuses on digital publishing and its combination with traditional publishing.