BMJ Open

Barriers and facilitators to data quality of electronic health records used for clinical research in China: a qualitative study

Kaiwen Ni, Hongling Chu,® Lin Zeng,® Nan Li, Yiming Zhao

Research Center of Clinical Epidemiology, Peking University Third Hospital, Beijing, China

Correspondence to
Dr Yiming Zhao;
yimingzhao115@163.com

ABSTRACT

Objectives There is an increasing trend in the use of electronic health records (EHRs) for clinical research. However, more knowledge is needed on how to assure and improve data quality. This study aimed to explore healthcare professionals' experiences and perceptions of barriers and facilitators of data quality of EHR-based studies in the Chinese context.

Setting Four tertiary hospitals in Beijing, China.

Participants Nineteen healthcare professionals with experience in using EHR data for clinical research participated in the study.

Methods A qualitative study based on face-to-face semistructured interviews was conducted from March to July 2018. The interviews were audiorecorded and transcribed verbatim. Data analysis was performed using the inductive thematic analysis approach.

Results The main themes included factors related to healthcare systems, clinical documentation, EHR systems and researchers. The perceived barriers to data quality included heavy workload, staff rotations, lack of detailed information for specific research, variations in terminology, limited retrieval capabilities, large amounts of unstructured data, challenges with patient identification and matching, problems with data extraction and unfamiliar with data quality assessment. To improve data quality, suggestions from participants included: better staff training, providing monetary incentives, performing daily data verification, improving software functionality and coding structures as well as enhancing multidisciplinary cooperation.

Conclusions These results provide a basis to begin to address current barriers and ultimately to improve validity and generalisability of research findings in China.

Strengths and limitations of this study

► The qualitative approach used in this study allowed for in-depth exploration of healthcare professionals' experiences and perceptions of barriers and facilitates to data quality of electronic health records used for clinical research in China.

► Purposive sampling was used to select potential participants in our study. Age, gender, education background and medical specialties were taken into account in the recruitment process in order to ensure diversity of participants.

► The participants were recruited from four hospitals in Beijing and that the generalisability of the findings may be limited because of the small sample size.

► The cultural background and experience of the authors may influence the interpretation of the data.

INTRODUCTION

China has lagged behind in the adoption of electronic health records (EHRs) in health-care institutions but the use of EHR systems is now rapidly growing as a result of the New Medical Reform.[1 2] The implementation of health information technology, especially EHRs, has been identified as a primary focus for this New Medical Reform, with a large potential impact on modern hospital management and care quality.[3 4] The widespread adoption of EHRs has made it easier for researchers to access and aggregate longitudinal patient data. There are significant opportunities to use EHR data for research purposes, including estimation of disease incidence or prevalence, comparative effectiveness studies and safety surveillance.[5–8]

Although randomised controlled trials (RCTs) are recognised as the 'gold standard' to assess the efficacy of interventions and treatments, they are costly, time-consuming and do not reflect real clinical practice.[9] The data from EHR-based research are collected under real-world circumstances, which improves representativeness as well as reducing cost and effort.[10] Evidence generated from EHR-based studies can complement that of RCTs and may provide important information in real-world settings where RCTs are not feasible.[10 11] However, a major concern about EHR-based studies is the quality of data. As EHRs are used for patient care rather than specifically for clinical research, data quality may not meet the criteria for research standards to be met and this may have a negative impact on the validity of research findings.[12 13]

A key concept in the field of data quality is 'fitness for use'. Specifically, the same database may be considered as poor quality for one purpose but high quality for another purpose, which means data quality should be assessed within the context of the intended use.[14 15] Considering the topic of this study related to clinical research, the data quality refers to the fitness for use of EHR data for specific research in the paper. Poor data quality may derive from various origins, which can occur at any or each stage of data flow, including recording, migration, extraction and translation.[16] Previous studies have identified some factors that influence data quality, particularly concerning clinical documentation and coding in EHRs.[17 18] Although the quality of EHR data has been assessed in other countries, few Chinese studies exist.[19]

The use of EHRs for clinical research should take into account the contextual features that might influence data quality in China. There are hundreds of vendors of software companies and research institutions to provide EHR systems for hospitals, resulting in partial data standardisation.[20 21] Additionally, the difference between Chinese and English languages makes some existing extraction methods for dealing with unstructured free-text inapplicable to Chinese EHRs.[22] In light of its increased use of EHRs for clinical research, more knowledge is needed on how to assure and improve data quality. The aim of this study was to explore healthcare professionals' experiences and perceptions of barriers and facilitates to data quality of EHRs used for clinical research in China.

METHODS
Study design
The research is reported according to the Consolidated Criteria for Reporting Qualitative Research guidelines.[23] A qualitative study was conducted at four tertiary hospitals in Beijing with between 970 and 1500 beds from March to July 2018. These four hospitals provide medical services across all medical specialisations. EHR systems in these hospitals are provided by four different vendors and do not have research-specific functionality. This study was based on interviews with healthcare professionals at the four hospitals in order to achieve varying perspectives of data quality of EHRs used for clinical research. Each participant was informed about the project and was free to withdraw from the study at any time.

Participant recruitment
Purposive sampling was used in this study. Healthcare professionals meeting the following criteria were eligible for the study: having at least 3 years of work experience; participating in using EHRs for clinical research, including recording, extraction and translation; having experience in publication of clinical research based on EHR data or in reporting relevant research at conferences; willing to participate this study. Age, gender, education background and medical specialties were taken into account in the recruitment process in order to ensure diversity

of participants.[24] Potential participants were known by researchers through meetings and expert recommendations. The individuals targeted for enrolment were contacted by email or WeChat, a popular social network application in China that is similar in function to Twitter.

Data collection
Face-to-face semistructured interviews were conducted by two research assistants (KN and HC), a male and a female, each with a PhD. Both interviewers received training on interview methods and have experience in conducting qualitative research. Interviewers had no previous relationship with the participants prior to recruitment. Participants did not know contact details of interviewers and the study purpose prior to this study. The semistructured interview guide was constructed to begin with a general question regarding participants' experience of using EHR data for clinical research. Participants were asked to briefly describe one or two studies using data extracted from EHRs in which they had been involved. They were then asked how they extracted and managed data from EHR systems. Finally, participants were asked about two key questions, 'In your opinion, what were barriers to data quality of EHR data used for your studies?' and 'Please provide some suggestions to improve data quality of EHRs used for clinical research.' Each interview lasted between 25 and 65 min and was conducted at participants' office or meeting room. Field notes were also taken. No repeat interviews were conducted and transcripts were not returned to participants for comments.

Data analysis
The anonymised audio records were transcribed verbatim by an external transcription company. QSR NVivo V.10 was used to assist data management and analysis. The transcripts were analysed using inductive thematic analysis.[25] The analysis consisted of six phases: familiarisation, generating initial codes, searching for themes, reviewing themes, defining and naming themes and producing the final results. To ensure reliability and validity, KN and HC independently coded the transcripts and met regularly to discuss the identification of themes. Interviews and the review of transcripts continued until no new themes emerged, at which point we determined theoretical saturation had been achieved. Final themes were reviewed by all research team members to confirm the interpretation of data in a discussion meeting.

Patient and public involvement
No patients or members of the public were involved in the design and implementation of this study. The study protocol and results are available on request.

RESULTS
Characteristics of study participants
Nineteen healthcare professionals were interviewed in the study. Two potential participants refused to take part

Table 1 Demographics of the participants

Variable	N
Gender	
Female	13
Male	6
Age	
25–30	6
31–40	10
41–50	3
Education background	
Bachelor	2
Master	9
PhD	
Years of research experience*	8
2–4	8
5–7	5
8–10	4
11–15	2

*Years of research experience mean the period from the beginning of the experience of conducting clinical research to the interviews, and conducting clinical research refers to the need for study design and/or data collection and analysis.
PhD, doctor of philosophy.

Table 2 Overview of key themes and subthemes

Barriers	Facilitators
Healthcare systems	
► Heavy workload	► Training of staff
► Staff rotations	► Need for monetary incentives
Clinical documentation	
► Lack of detailed information for specific research	► Performing daily data verification
► Variations in terminology	
EHR systems	
► Limited retrieval capabilities	► Improving software functionality and coding structures
► Large amounts of unstructured data	
► Challenges with patient identification and matching	
Researchers	
► Problems with data extraction	► Enhancing multidisciplinary cooperation
► Unfamiliar with data quality assessment	

EHR, electronic health record.

due to heavy workload. The sample characteristics of the study are showed in table 1. There were seven respondents in one of the hospitals, and four respondents in each of the other three hospitals. Respondents were from a broad range of medical specialties, including urology, radiology, haematology, paediatrics, orthopaedics, neurology, pneumology, geriatrics, pharmacy and general surgery. The range of clinical research undertaken by respondents included quantifying the value of certain diagnostic tests, identification of disease risk factors and assessment of outcomes of interventions.

Barriers to data quality

Factors that influenced the data quality of EHR-based clinical research were identified and organised into four themes: healthcare systems, clinical documentation, EHR systems and researchers. Table 2 shows a detailed list of themes and subthemes of the perceived barriers. Illustrative quotes are in table 3.

Healthcare systems
Heavy workload

Most participants mentioned that heavy workload would affect the quality of clinical documentation in EHR systems. They pointed out that it takes too long to input each item carefully into EHRs, which seems unlikely to ensure consistent data quality in a high-intensity work environment. For example, one orthopaedic doctor noted, 'One day, you have to take charge of three or four new inpatients, you have to go to surgery, and then you have to do some of your own things, so the quality of EHR data can be affected.' One haematologist said, 'There are about 70 patients in a morning. It is time-consuming for us to record patient information into the EHR system.'

Staff rotations

Some participants indicated that staff rotations were an important barrier to the data quality of EHR-based studies. Teaching hospitals are responsible for training medical students and clinicians from other hospitals in rotations in China. One orthopaedist mentioned, 'Many healthcare workers like me are just becoming familiar with the tasks of one department and then are transferred to the next department.' One participant noted, 'There are some routine behavior questionnaires for each new inpatient and are usually completed by medical students or visiting clinicians, but these healthcare workers are often changed. Staff rotations have a negative impact on quality of EHR data.'

Clinical documentation
Lack of detailed information for specific research

EHRs are optimised for the storage of information needed for patient care but may not be sufficient for specific clinical research. One general surgeon said, 'Given the study requires laboratory values of one day, three days and five days after surgery, these data are not routinely recorded in EHRs.' Some participants noted that EHR data cannot have as much detailed information about

Table 3 Quotes related to barriers and facilitators of data quality of EHR-based clinical research

Themes	Subthemes	Quotes
Barriers	Heavy workload	One day, you have to take charge of three or four new inpatients, you have to go to surgery, and then you have to do some of your own things, so the quality of EHR data can be affected.
	Staff rotations	Many healthcare workers like me are just becoming familiar with the tasks of one department and then are transferred to the next department.
	Lack of detailed information for specific research	Given the study requires laboratory values of one day, three days and five days after surgery, these data are not routinely recorded in EHRs.
	Variations in terminology	This drug we studied is very common in hospitals, but different physicians are accustomed to using different terms and formats to indicate the drug name in EHRs.
	Limited retrieval capabilities	Our EHRs cannot use pathological diagnosis results as search terms to obtain desired patient populations.
	Large amounts of unstructured data	Some important information for clinical research is in the form of unstructured data in EHR systems. The data has to be extracted from each individual health record manually by researchers.
	Challenges with patient identification and matching	Unfortunately the patient identification numbers for the same patient from the inpatient and outpatient settings are not the same, which makes it more difficult to combine information from these settings for clinical research.
	Problems with data extraction	The data extracted by staff in the information department often have various problems, you need to communicate with them again and again.
	Unfamiliar with data quality assessment	I am not quite sure how to assess data quality. I usually mainly consider data completeness.
Facilitators	Training of staff	Targeted training is often seen as a way to improve data standardization for research purposes.
	Need for monetary incentives	Giving a reward to resident physicians and medical students may improve the data quality of routine behavior questionnaires, so that other healthcare professionals can use these baseline data for clinical research.
	Performing daily data verification	It may be possible for a research nurse to check the EHR data of the ward on the day it is entered.
	Improving software functionality and coding structures	I suggest that data in the departments of radiology and pathology can be modularized in EHR systems, so that data screening and extraction can be facilitated.
	Enhancing multidisciplinary cooperation	Conducting a real-world study (like EHR-based clinical research) requires teamwork. The overall quality of data will improve if the team can communicate and cooperate to ensure the quality of each stage.

EHR, electronic health record.

patient's family history or medical history as prospectively collected data for studies.

Variations in terminology

Almost all participants pointed out that the lack of data standardisation is a key barrier to data quality. For efficiency, healthcare workers may use shortened terminologies or acronyms to refer to diagnoses and interventions for clinical uses. One pharmacist said, 'This drug we studied is very common in hospitals, but different physicians are accustomed to using different terms and formats to indicate the drug name in EHRs.' One gerontologist said, 'The use of International Classification of Diseases is non-standard and field definitions are often modified for clinical purposes by clinicians. This meets the needs of routine clinical records, but partially affects data standardization and has challenges for research purposes.'

EHR systems
Limited retrieval capabilities

The usability of EHRs for clinical research relies on the ability to retrieve complete and accurate information. Hospitals choose from different vendors to provide EHR systems, and they have different levels of retrieval capabilities and are unlikely to be chosen with data extraction for research purposes as a factor. One ultrasonologist noted, 'Our EHRs cannot use pathological diagnosis results as search terms to obtain desired patient populations.'

Large amounts of unstructured data

EHR systems incorporate many free-text fields with unstructured data, including physician notes, images, medications and so on. Participants mentioned, 'Some important information for clinical research is in the form of unstructured data in EHR systems. The data has to be extracted from each individual health record manually by researchers.' Data quality of EHR-based studies may be affected due to human errors in these processes.

Challenges with patient identification and matching

Only one of the four hospitals could link outpatient data and hospitalisation data in our study. Outpatient data can provide significant patient follow-up information after interventions or treatments. However, patients receive a new patient identification number each time when they receive outpatient services. As data from different sources cannot be linked by matching patient identification numbers, it is difficult for researchers to optimally use this important combined information. One ultrasonologist said, 'Unfortunately the patient identification numbers for the same patient from the inpatient and outpatient settings are not the same, which makes it more difficult to combine information from these settings for clinical research.'

Researchers

Problems with data extraction

Many participants extracted data from EHRs and entered it into the Excel software by themselves. A gerontologist explained, 'It always takes one or two months for the staff in the hospital information department to assist with the extraction of data, due to time constraints concerning the operation and project development of the hospital information system, and data extraction is not their main task. I therefore prefer to extract data from EHR system by myself.' One pharmacist noted, 'The data extracted by staff in the hospital information department often have various problems, you need to communicate with them again and again.' Additionally, researchers are required to manually extract information from free-text fields and scanned documentation that stored as images in EHR systems, which is a time-consuming process and may cause errors that affect data completeness and consistency.

Unfamiliar with data quality assessment

Most participants agreed with the importance of data quality, but they were not familiar with how to systematically assess data quality of EHR-based studies. The lack of understanding of features of EHR data may have impacts on data quality improvement and result interpretation. One urologist said, 'I am not quite sure how to assess data quality. I usually mainly consider data completeness.'

Facilitators to improving data quality

All of the interviews included discussions on how to improve the data quality of EHR-based clinical research. Table 2 also presents themes and subthemes of the perceived facilitators.

Training of staff

Training of staff was a perceived facilitator to data quality of EHR-based studies. Participants mentioned that it is important for healthcare workers to receive education and training regarding the utilisation of EHR system functions and coding schemes, as systematic documentation of clinical data for regular healthcare delivery is the basic prerequisite before further enhancing data quality. One pharmacist said, 'Targeted training is often seen as a way to improve data standardization for research purposes.'

Need for monetary incentives

Some participants suggested that there was a potential for monetary incentives to improve data quality. Under the premise of meeting clinical purposes, monetary incentives can stimulate healthcare workers to improve standardisation and accuracy of EHR data for research purposes. One orthopaedist said, 'Giving a reward to resident physicians and medical students may improve the data quality of routine behavior questionnaires, so that other healthcare professionals can use these baseline data for clinical research.'

Performing daily data verification

Performing data verification by research nurses was stated as a way to improve the quality of EHR data. Several participants pointed out that it may be possible for a research nurse to check the EHR data of the ward on the day it is entered. If there are obvious problems, they can perform follow-up with clinicians or patients.

Improving software functionality and coding structures

The EHR system needs to be developed to make coding options more convenient and effective to clinicians than free-text fields. Compared with unstructured data, structured data can be easily extracted electronically, which avoids human errors. One urologist mentioned, '(I) suggest that data in the departments of radiology and pathology can be modularized in EHR systems, so that data screening and extraction can be facilitated.'

Enhancing multidisciplinary cooperation

Many participants mentioned that enhancing multidisciplinary involvement is an important factor for conducting clinical research based on EHRs. The research team should include the following: clinician, nurse, IT staff, clinical epidemiologist and data professional. One neurologist said, 'Conducting a real-world study (like EHR-based clinical research) requires teamwork. The overall quality of data will improve if the team can communicate and cooperate to ensure the quality of each stage.'

DISCUSSION

In this study, we tried to identify barriers and facilitators to data quality of EHR-based clinical research as reported by healthcare professionals in China. The main themes included factors related to healthcare systems, clinical documentation, EHR systems and researchers.

Considering the increasing trend of use of EHR data for clinical research, understanding the factors that influence the quality of data may alleviate current barriers and ultimately improve validity and generalizability of research findings.

Heavy workload was identified as a key barrier to the quality of EHR data. Over 1.56 billion patient visits were made to hospitals in the first half of 2016 in China.[26] It is not surprising that some participants considered that the quality of outpatient data might not meet the needs of clinical research. This point is consistent with previous discussions of Pourasghar et al[27] that high workload was one of main barriers to the quality of documentation in EHR systems, which was improved in areas where nurses were involved. Staff rotations were considered to affect adequately record clinical information for patients in our study, which may further affect the quality of secondary use of EHR data for research purposes. Other studies have also reported that documentation habits had an impact on data quality, and have suggested that training and feedback can be adopted to mitigate the variability of clinical documentation.[28] These feelings were similar to those presented by participants in our study. Additionally, participants mentioned that providing monetary incentives may be an effective way to improve data quality, encouraging clinicians and medical students to be more responsible for recording in EHR systems. Nevertheless, incentives could influence clinical practices differently depending on the amount and business model.[29] Providing monetary incentives may be an effective way to ensure the quality of some EHR data, but may also lead to selective recording behaviours that attract higher incentives.[30]

Participants perceived that software functionality and coding schemes used in hospitals may be not comprehensive sufficiently, which affects the data quality of EHR-based clinical research. EHR data are difficult to standardise for data integration and interoperability in China and this is likely to be the case in other countries.[20 31 32] This is mainly because hospitals choose their own vendors to develop different EHR systems. The available data entry functions and coding systems built into EHR systems determine what is recorded. Previous studies demonstrated that EHR systems developed by different vendors can result in considerable differences in the quality of prescription records and episodes of care.[33 34] It also leads to problems about interoperability between different medical terminologies, affecting the potential for data sharing and assurance of consistency and quality of large-scale EHR-based studies that need to link EHR data.

Healthcare professionals usually conduct EHR-based clinical research independently in China.[26] However, using EHR data for clinical research requires multidisciplinary involvement. Clinicians should acknowledge the limitations of their expertise and cooperate with other professionals.[35] For example, relating to manual data extraction from EHRs by clinicians, collaboration with IT staff and using electronic data extraction methods can avoid potential human errors. Additionally, although participants agreed with the importance of data quality of EHR-based clinical research, they had no idea about how to assess the quality of data. Evidence showed that studies based on EHR data of unknown quality have negative impacts on care quality and patient safety.[36 37] It is important for researchers to achieve competence in the methods of assessing and reporting data quality findings, which helps EHR data users and consumers understand the limitations of data and results based on these data. However, there are still inconsistencies in the dimensions of data quality and the methods used to assess these dimensions both in China and in other parts of the world.[13 38] The lack of consistency of methods and processes of data quality assessment (DQA) makes it difficult to compare the results of DQAs.[13 39] Thus, DQA should be further studies to standardise the terminologies of data quality concepts and to improve the feasibility of task-dependent DQA by researchers.

The limitations of this study include that the participants were recruited from four hospitals in Beijing and that the generalisability of the findings may be limited because of the small sample size. There is also a possibility that we had continued to interview and other themes may have emerged. However, all of the research team felt that no new information was forthcoming and that saturation had been achieved with the nineteen participants. All the interviews were conducted by the first and second authors of our study. Their cultural background and experience may have influenced data collection and analysis, but they were experienced in conducting qualitative research and a meeting was held to discuss and ensure the rigour of interpretation. We believe that these findings can contribute to an increased understanding of barriers and facilitators of that influence data quality of EHR-based clinical research in China.

CONCLUSION

We have identified various barriers to data quality of EHR-based clinical research in the Chinese context, which may have impacted on the validity and generalisability of research findings. Staff training regarding systematic and complete documentation of clinical data for routine medical services was emphasised as a key factor, which is the basic prerequisite before further enhancing data quality of EHR-based clinical research. To overcome the time-related problems, providing monetary incentives and performing daily data verification were perceived as two important ways. In terms of EHR systems, the improvement of software functionality and coding structures may help to increase data standardisation and interoperability for research purposes. Multidisciplinary cooperation can also improve the quality of clinical research based on EHRs.

REFERENCES

1. Shu T, Liu H, Goss FR, *et al*. EHR adoption across China's tertiary hospitals: a cross-sectional observational study. *Int J Med Inform* 2014;83:113–21.
2. Lei J, Sockolow P, Guan P, *et al*. A comparison of electronic health records at two major Peking University Hospitals in China to United States meaningful use objectives. *BMC Med Inform Decis Mak* 2013;13:96.
3. Yip WC, Hsiao WC, Chen W, *et al*. Early appraisal of China's huge and complex health-care reforms. *Lancet* 2012;379:833–42.
4. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. *N Engl J Med* 2010;363:501–4.
5. Cowie MR, Blomster JI, Curtis LH, *et al*. Electronic health records to facilitate clinical research. *Clin Res Cardiol* 2017;106:1–9.
6. Shah AD, Langenberg C, Rapsomaniki E, *et al*. Type 2 diabetes and incidence of cardiovascular diseases: a cohort study in 1·9 million people. *Lancet Diabetes Endocrinol* 2015;3:105–13.
7. Ball R, Robb M, Anderson SA, *et al*. The FDA's sentinel initiative--A comprehensive approach to medical product surveillance. *Clin Pharmacol Ther* 2016;99:265–8.
8. Shiner B, Westgate CL, Gui J, *et al*. A Retrospective Comparative Effectiveness Study of Medications for Posttraumatic Stress Disorder in Routine Practice. *J Clin Psychiatry* 2018;79.
9. Kim HS, Lee S, Kim JH. Real-world Evidence versus Randomized Controlled Trial: Clinical Research Based on Electronic Medical Records. *J Korean Med Sci* 2018;33:e213.
10. Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA. Routinely collected data and comparative effectiveness evidence: promises and limitations. *CMAJ* 2016;188:E158–E164.
11. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000;342:1887–92.
12. Weiner MG, Embi PJ. Toward reuse of clinical data for research and quality improvement: the end of the beginning? *Ann Intern Med* 2009;151:359–60.
13. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc* 2013;20:144–51.
14. Kahn MG, Brown JS, Chun AT, *et al*. Transparent reporting of data quality in distributed data networks. *EGEMS* 2015;3:7.
15. de Lusignan S. The optimum granularity for coding diagnostic data in primary care: report of a workshop of the EFMI Primary Care Informatics Working Group at MIE 2005. *Inform Prim Care* 2006;14:133–7.
16. Verheij RA, Curcin V, Delaney BC, *et al*. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. *J Med Internet Res* 2018;20:e185.
17. Coleman N, Halas G, Peeler W, *et al*. From patient care to research: a validation study examining the factors contributing to data quality in a primary care electronic medical record database. *BMC Fam Pract* 2015;16:11.
18. Lucyk K, Tang K, Quan H. Barriers to data quality resulting from the process of coding health information to administrative data: a qualitative study. *BMC Health Serv Res* 2017;17:766.
19. Yang Y, Zhou X, Gao S, *et al*. Evaluation of Electronic Healthcare Databases for Post-Marketing Drug Safety Surveillance and Pharmacoepidemiology in China. *Drug Saf* 2018;41:125–37.
20. Wang Z. Data integration of electronic medical record under administrative decentralization of medical insurance and healthcare in China: a case study. *Isr J Health Policy Res* 2019;8:24.
21. Zhao J, Zhang Z, Guo H, *et al*. E-health-oriented community health information system in china: our challenges, solution, and experience. *Telemed J E Health* 2011;17:584–8.
22. Xu Y, Li N, Lu M, *et al*. Development and validation of method for defining conditions using Chinese electronic medical record. *BMC Med Inform Decis Mak* 2016;16:110.
23. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007;19:349–57.
24. Palinkas LA, Horwitz SM, Green CA, *et al*. Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Adm Policy Ment Health* 2015;42:533–44.
25. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol* 2006;3:77–101.
26. Li P, Xie C, Pollard T, *et al*. Promoting Secondary Analysis of Electronic Medical Records in China: Summary of the PLAGH-MIT Critical Data Conference and Health Datathon. *JMIR Med Inform* 2017;5:e43.
27. Pourasghar F, Malekafzali H, Koch S, *et al*. Factors influencing the quality of medical documentation when a paper-based medical records system is replaced with an electronic medical records system: an Iranian case study. *Int J Technol Assess Health Care* 2008;24:445–51.
28. Parsons A, McCullough C, Wang J, *et al*. Validity of electronic health record-derived quality measurement for performance monitoring. *J Am Med Inform Assoc* 2012;19:604–9.
29. Ghosh A, McCarthy S, Halcomb E. Perceptions of primary care staff on a regional data quality intervention in Australian general practice: a qualitative study. *BMC Fam Pract* 2016;17:50.
30. Roland M. Linking physicians' pay to the quality of care--a major experiment in the United kingdom. *N Engl J Med* 2004;351:1448–54.
31. van Velthoven MH, Mastellos N, Majeed A, *et al*. Feasibility of extracting data from electronic medical records for research: an international comparative study. *BMC Med Inform Decis Mak* 2016;16:90.
32. Hemingway H, Asselbergs FW, Danesh J, *et al*. Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. *Eur Heart J* 2018;39:1481–95.
33. Opondo D, Visscher S, Eslami S, *et al*. Quality of Co-Prescribing NSAID and Gastroprotective Medications for Elders in The Netherlands and Its Association with the Electronic Medical Record. *PLoS One* 2015;10:e0129515.
34. Sollie A, Sijmons RH, Helsper C, *et al*. Reusability of coded data in the primary care electronic medical record: A dynamic cohort study concerning cancer diagnoses. *Int J Med Inform* 2017;99:45–52.
35. Ioannidis JP, Greenland S, Hlatky MA, *et al*. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* 2014;383:166–75.
36. Middleton B, Bloomrosen M, Dente MA, *et al*. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. *J Am Med Inform Assoc* 2013;20:e2–e8.
37. Kahn MG, Ranade D. The impact of electronic medical records data sources on an adverse drug event quality measure. *J Am Med Inform Assoc* 2010;17:185–91.
38. Kahn MG, Callahan TJ, Barnard J, *et al*. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. *EGEMS* 2016;4:18.
39. Chen H, Hailey D, Wang N, *et al*. A review of data quality assessment methods for public health information systems. *Int J Environ Res Public Health* 2014;11:5170–207.